Acta Universitatis Sapientiae
Electrical and Mechanical Engineering, 15 (2023) 14–32

DOI: 10.2478/auseme-2023-0002

A Review of Using Visual Odometery Methods in Autonomous UAV Navigation in GPS-Denied Environment

Hussam M. ROSTUM[1], József VÁSÁRHELYI[2]

[1] Institute of Automation and Info-Communication, Faculty of Mechanical engineering and Informatics, University of Miskolc, Miskolc, e-mail: hussam.rostum@uni-miskolc.hu
[2] Institute of Automation and Info-communication, Faculty of Mechanical engineering and Informatics, University of Miskolc, Miskolc, e-mail: jozsef.vasarhelyi@uni-miskolc.hu

Manuscript received November 21, 2023, revised December 10, 2023

Abstract: This review paper centers on strategies employed for location determination in regions lacking GPS signals. It primarily explores a range of vision-based methods tailored for this purpose, categorizing them accordingly. The article delves into the utilization of optical flow for feature extraction-based Visual Odometry (VO) and delves into advanced optical flow estimation methods that hinge on deep learning techniques. It compares the efficacy and practical applications of frequently utilized visual localization methods while also checking the efficiency of previous researches by reapplying the algorithms to new data and comparing the results.

Keywords: Visual odometry, Optical flow, Deep learning Methods, GPS-denied localization.

1. Introduction

According to [1], autonomous navigation systems must possess the capability to estimate, perceive, and comprehend their surroundings in order to accomplish tasks such as path tracking, motion planning, obstacle avoidance, and target detection. GNSS, or Global Navigation Satellite Systems, furnish dependable environmental data and instantaneous positioning for self-governing vehicles such as Autonomous Vehicles (AVs) and Unmanned Aerial Vehicles (UAVs). In scenarios where Unmanned Aerial Vehicles (UAVs) are operating in congested and complex environments, GNSS signals may encounter issues like fading, multipath effects, jamming, and spoofing, which can result in signal loss. Hence, GNSS may not be a reliable solution for UAVs flying at lower altitudes in terrains marked by numerous obstructions, such as forests, cities, and canyons/mountains

[2]. To tackle this problem, over the recent years, scientists have devised various localization approaches specifically for UAVs, including vision-based and LIDAR-based techniques. Incorporating multiple sensors in robotic localization studies can be an expensive undertaking, and it may also add extra weight and power consumption to the device. The challenges mentioned above have impeded the widespread implementation of several robot localization algorithms [3]. As an alternative, visual localization, a computer vision-based technology, has emerged as an appealing option. The operating principle of visual localization is based on capturing images of the surrounding environment using a visual camera, followed by determining the position and orientation of the area around it, thereby producing a map of the unknown territory. One of the major advantages of this technology is that the camera's cost is relatively low, and it has the ability to capture extensive environmental data, including visual aspects such as color and texture. Nonetheless, visual localization demands high computational power, and images require a significant amount of storage space, and software development for visual localization is relatively challenging. Moreover, the visual system is susceptible to lighting conditions, and it may not function effectively in poorly lit environments. At present, there are two primary vision-based localization methods, namely Relative Visual Localization (RVL) and Absolute Visual Localization (AVL). RVL encompasses Visual Odometry (VO) and Visual Simultaneous Localization and Mapping (VSLAM). Visual Odometry (VO) is a technique for calculating the self-motion of a robot, utilizing monocular or binocular cameras. This literature review summarizes the current state of research and challenges facing Visual Odometry localization technologies in environments where GPS is unavailable. The primary contributions are summarized in the following areas:

- The research introduces the working principle of optical flow and its algorithm, as well as the broad applications of optical flow in visual odometry. The main focus of this review is on the introduction of FlowNet and its subsequent improved algorithms, and it includes a comparison to guide the selection of optical flow estimation algorithms based on deep learning.
- The paper summarizes the main challenges in the development of localization technologies under GPS-denied conditions and proposes potential solutions.

The structure of this review paper is as follows. In Section 2, 3, 4 and 5 it introduces optical flow-based techniques used in visual odometry and examines their applications. Section 6 contains the conclusion of the review article.

2. Visual Localization Through optical flow-based visual odometry techniques

Visual odometry is a vision-based navigation method that uses visual features from a sequence of images to estimate the relative pose of a robotic platform. It is a well-established technique for autonomous navigation and localization, in which an agent estimates its location and orientation based on the visual information from its onboard cameras. In the last few years, researchers have put forth several VO techniques, which can be classified into two categories based on what type of camera is employed: monocular camera methods and stereo camera methods. A binocular camera is the most widely employed stereo camera, which can utilize the space between the two cameras to gain depth information. Using RGBD cameras, both image and depth information can be acquired concurrently; however, the scope of the acquired depth data is restricted and dependent on infrared light, in addition to consuming considerable power. The simplicity of structure and affordability that monocular cameras offer have encouraged a plethora of studies to be conducted on them. The utilization of monocular cameras as a vision sensor for the VO method is widespread due to their affordability, compactness, and energy efficiency. This makes them suitable for small platforms, including Unmanned Aerial Vehicles (UAVs) [7], [8]. VO can be categorized into two approaches - the direct method and the feature-based method. The latter is usually regarded as the predominant approach of VO due to its benefits of strong resistance to rotation, fuzziness, and scale transformation. This method functions by estimating the motion pose of the camera via the selection of specific points (like corner points) in the image and the concurrent assessment of the motion conditions of the connected feature points in the two frames preceding and succeeding it. Currently, numerous feature extraction techniques have been developed in the domain of computer vision, for example SURF (Speeded Up Robust Features) [10] (building histogram according to the magnitude of gradient value), SIFT(Scale-Invariant Feature Transform) [9] (generating features by utilizing the histogram of gradient direction and gradient magnitude),and ORB (Oriented FAST and Rotated BRIEF(Binary Robust Independent Elementary Features)) [11] (constructing histogram depending on the pixel value). Chen et al employed the SURF algorithm to identify and correlate feature points. For the purpose of matching, they utilized the Approximate Nearest Neighbor (ANN) algorithm. Initially, the SURF features drawn from each of the template flags are broken down into eight clusters and stored. These clusters are then compared with the databases of the images to be processed and the likely images are identified. Finally, the features are matched with the aid of ANN [12]. *Figure 1*. represents an example of that. Zheng et al. [13] conducted a study in which they applied various algorithms for feature

extraction on images in five different settings. The data in the *Table 1* displays that The ORB was the quickest, whereas SIFT method took the longest time to compute.

Table 1: Point and Time comparison of three feature extraction algorithms (SIFT, SURF, and ORB) [13]

ORB		SIFT		SURF	
Point	**Time(s)**	**Point**	**Time(s)**	**Point**	**Time(s)**
168	0.00282	171	0.01889	86	0.01244
299	0.00411	253	0.01899	254	0.01621
251	0.00470	234	0.01966	187	0.01386
168	0.00221	175	0.02557	183	0.01570
19	0.00175	24	0.01669	20	0.01038

Figure 1: The SURF algorithm extracts feature points and matches them

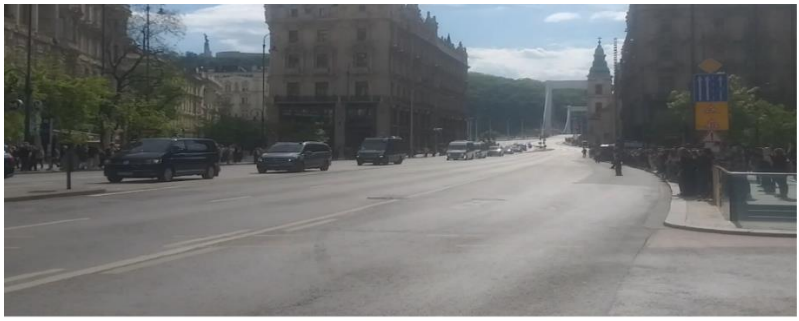
Although the feature-based approach is the most common method for Visual Odometry, it does possess some drawbacks. These include its requirement for extensive computing power and the fact that the feature extraction and descriptor computation processes are lengthy. It has been noted that when using this method to represent image motion, only a few hundred features are extracted, resulting in a significant amount of data loss compared to the hundreds of thousands of pixels that make up the image. This is especially apparent when there is a lack of texture in the image. It has been observed that the features that can be determined from the image are highly limited, making it extremely difficult for the feature-based approach to accurately estimate the camera motion in this circumstance. Consequently, scientists are attempting to utilize optical flow to enhance the feature-based approach. The navigation challenges can be tackled using optical flow techniques, which are based on the same principles of directional sensing

and localization that birds and insects employ while flying [14]. Currently, this technology has become a popular choice for robot positioning and navigation, providing dependable speed and position data. The optical flow method takes advantage of the modifications to pixels in the image sequence in the time domain and the relationship between adjacent frames in order to pinpoint the correlation between the preceding frame and the current frame, thereby computing the movement of elements between adjacent frames. It can be said that the instantaneous change rate of gray level of a certain coordinate point of the two-dimensional image plane is commonly classified as the optical flow vector. Optical flow has a great advantage as it can precisely measure and spot the location of a moving goal without any knowledge of the scene's information. Furthermore, it remains effective even when the camera is in motion. Optical flow not only furnishes insight into the unknown environment, but also assists in figuring out the direction and velocity of the robot and can detect moving subjects without any prior details about the scene. Moreover, because of the relatively advanced development of optical sensors, the price is usually low, and it is quite simple to reduce their size, which can effectively lessen costs and raise portability [4].

3. Traditional Methods of Optical Flow

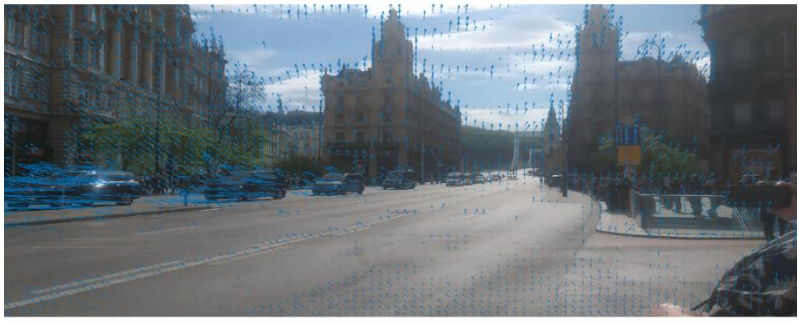
Different methods for computing optical flow have been suggested by researchers, including Lucas-Kanade algorithm [16], HornSchunck algorithm [17], image interpolation algorithm [18], block matching algorithm [19], and feature matching algorithm [9]. Optical flow estimation is typically predicated on the supposition that there is a constancy in brightness and a smoothness present. The constancy of brightness postulates that the luminance of an object does not waver between two successive frames. The idea of smoothness further expresses that the displacement is minimal; thus, the values of pixels in the vicinity are alike [17]. In real-world scenarios, the lighting conditions rarely meet the requirement of having the same intensity in adjacent frames for optical flow methods. Any changes in the lighting conditions will then have an impact on the accuracy of the optical flow measurements. Zhang et al. [20] proposed an optical flow localization technique based on ROF (Rudin-Osher-Fatemi) denoising to tackle the issue of optical flow calculation in non-uniform lighting. The convex optimization theory and duality principle were utilized to decompose the image in changing lighting and to minimize its effect. Boretti et al. [21] applied the Lucas-Kanade method for computing sparse optical flow fields and detected features with an ORB detector to calculate the optical flow of the image and further extract the motion information of the position and attitude of the MAV (Micro Aerial Vehicle) [20]. However, it was difficult to strictly adhere to the constant smoothness assumption in practicality. An image pyramid approach was

taken up to prevent the breakdown of feature detection in times of significant movement [21]. In the same way, Lou et al. [22] used image texture decomposition and image pyramid techniques to improve the Lucas-Kanade optical flow algorithm, diminishing the interference of illumination changes and large displacements on the detection of moving objects. The authors of this paper reproduced the research results of [22] and presented it in *Figure 2*.



(a) Input Image

(b) L-K Optical Flow Algorithm



(c) The Improved Algorithm by Lou etc.

Figure 2: Results of moving object detection

4. Deep Learning Methods of Optical Flow

The performance of deep learning methods for estimating optical flow surpasses that of traditional image-based approaches, and these techniques do not require explicit modeling of the entire problem, making them especially promising for use in optical flow estimation projects [23]. Dosovitskiy et al. [24] were the originators of an optical flow estimation technique based on learning, which they dubbed FlowNet. This supervised learning approach was designed to address the issue of optical flow estimation. In their work, Dosovitskiy et al. [24] introduced the Flying chairs dataset for the purpose of training the FlowNet network. Tests demonstrated that the FlowNet model developed using this synthetic dataset was able to generalize to images of the real world. *Figure 3.* illustrates the structure of two FlowNets. The first, known as FlowNetSimple, consists of a series of networks with only convolutional layers, in which two consecutive frames of input images are superimposed. On the other hand, FlowNetCorr takes two frames of pictures and processes them separately, extracting their respective features through a convolution layer and then performing a matching.

Despite its advantages, one of the major drawbacks of FlowNet is its high prediction error rate, making it unable to correctly process small displacements and real-world data. As a result, FlowNet2 was developed as an improved version.

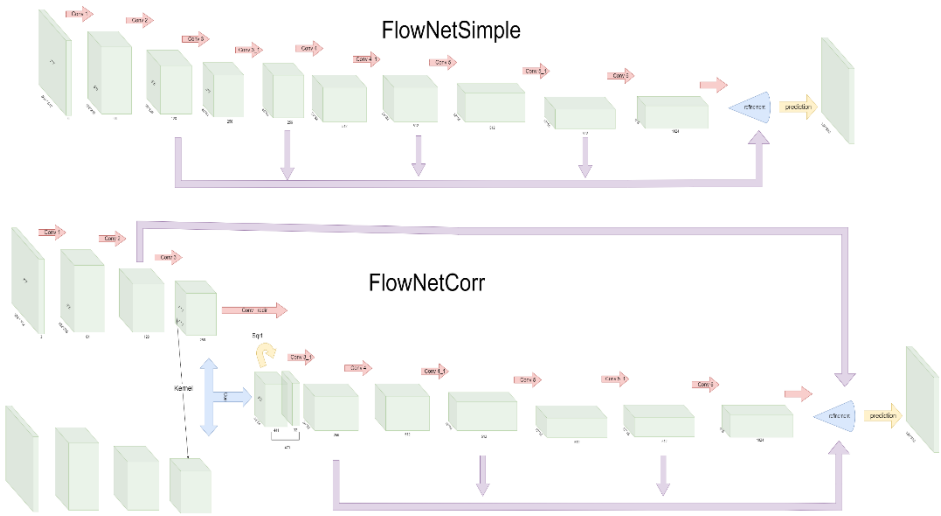


Figure 3: The network structures of FlowNetSimple (top) and FlowNetCorr (bottom)

Ilg et al. [25] demonstrated a substantial decrease in estimation errors when using a serial multiple network architecture in combination with a branch network to address optical flow estimation problems with small object displacement. Sun et al. [26] proposed an enhanced version of PWC-Net to calculate optical flow with high resolution. This network applies convolutional neural networks (CNNs) to capture image features and then employs the pyramid processing principle to predict optical flow at low resolution, while gradually progressing to the desired resolution [26]. PWC-Net has the benefit of being simpler to train than its counterpart, FlowNet2. Zhu et al. [23] proposed EV-FlowNet, based on FlowNet, as a novel self-supervised deep learning channel for estimate optical flow from event-based cameras. To do this, they first presented a novel approach for the depiction of event flow as images. A deep learning network is applied to an image with four channels, it encodes positive events with the first two channels and negative events with the other two channels. EV-FlowNet utilizes a single input of a (256×256×4) image sourced from a specific event stream. Through the use of the estimated traffic from the network, the corresponding gray level image taken from the same camera at the same time as the event is then used as a supervision signal which provides a loss function during the training process. The combination of images and self-monitoring loss is enough to enable the network to accurately predict optical flow solely from events [23].

Figure 4. shows the network structure of EV-FlowNet. The convolutional layer (in green) is responsible for downsampling (encoding), and the convolutional results from each layer are kept and linked to the upsampling (decoding) layer for use as a skip layer. The middle blue part is the residual block, which helps to further extract features.

The last part (in yellow) is the upsampling (decoding) section, which is accomplished by symmetric padding.

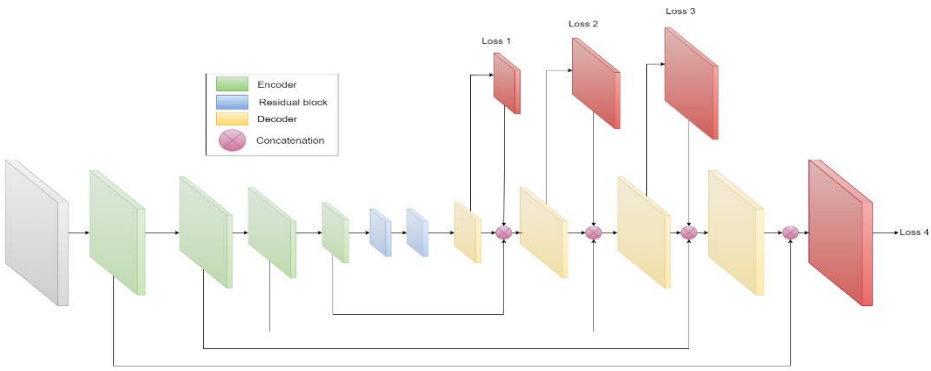


Figure 4: Illustrates the network structure of EV-FlowNet

The decoder activations from each set are passed through a depth wise convolutional layer to generate stream predictions at the resolution. The loss is applied to this stream prediction, and the prediction is also connected to the decoder activation [23]. Shi and Yin [27] introduced GeoNet, a collaborative unsupervised learning system, which was compared to EV-FlowNet for the purpose of estimating monocular depth, optical flow, and ego-motion from videos. The data acquired by the camera is comprised of rigid flow (static characteristics) and non-rigid flow (dynamic features). These are derived from not just the camera's motion, but also the movement of the target object. By taking this into account, the researchers devised a brand-new cascaded architecture with two stages to calculate both the global and fine displacement of the picture, respectively, to be able to adaptively tackle the figure of rigid and non-rigid flow. *Table 2* below presents an evaluation of FlowNet and its improved algorithms. *Figure 5* shows a schematic illustration of the UAV autonomous navigation network designed by Mumuni et al. [33] Ground plane segmentation (G-Seg) maps are also used to calibrate the metric scale. DepthNet estimates depth per frame, EgoMNet estimates relative camera pose, and OFNet predicts optical flow. The network also estimates confidence maps for each task [33].

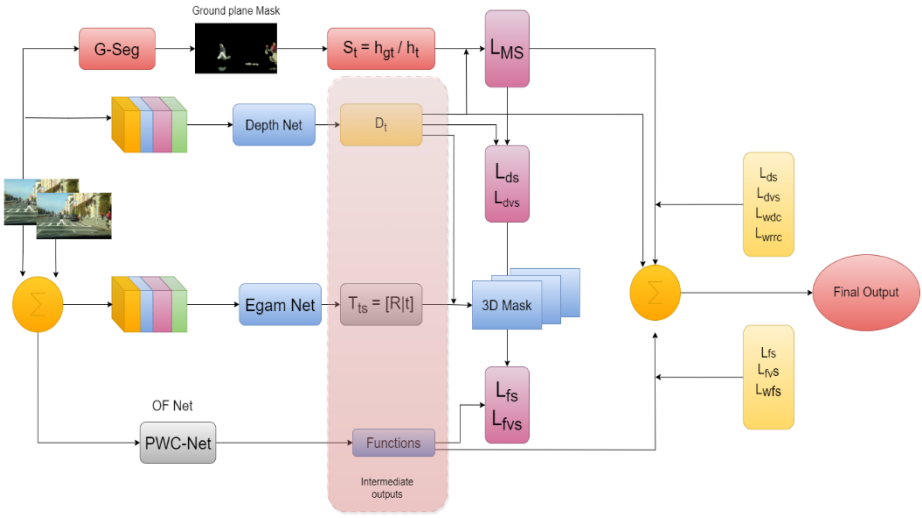


Figure 5: Schematic illustration of the UAV autonomous navigation network

5. Application of Visual odometry

The focus on Visual Odometry (VO) has grown substantially in recent years, as technology has become increasingly employed in the realms of robotics, autonomous driving vehicles, and Augmented Reality (AR). In the absence of GPS, Visual Odometry (VO) has become a popular choice due to its cost effectiveness and ease of access, acting as a supplementary tool to Inertial Navigation Systems (INS) and wheel odometers.

Table 2: A comparison of the benefits and drawbacks of FlowNet and its succeeding modified algorithms

Method	Main Contribution	Main Disadvantages	Dataset
FlowNet [24]	CNN was the first to set the example of being able to forecast OF.	Prediction accuracy not satisfactory, the system not responding to small movements or actual data.	Middlebury [28] + KITTI [29] + Sintel [30] + FlyingChairs [24]
FlowNet2[25]	FlowNet surpassed in terms of accuracy and speed.	The impact of image noise on forecast is still significant.	Middlebury + KITTI + Sintel + Flying Chairs
PWC-Net [26]	Reduce the size of the CNN, process training simpler.	Blurry estimates.	KITTI + Sintel + Flying Chairs
EV-FlowNet [23]	Use event camera data as an input for the network.	The picture capture technique is unusual and causes poor migration.	MVSEC [31]
GeoNet [27]	Evaluation difficulty of rigid and non-rigid flows.	The GeoNet model demonstrates less successful results than the direct unsupervised flow.	KITTI+ Cityscape [32]

VO techniques can be classified into three primary categories: those based on geometry, those based on deep learning, and those that combine the two. Conventional monocular VO algorithms typically involve three primary stages: tracking, optimization, and a closed-loop module. These processes exploit the geometric properties of the scene and often rely on optical flow techniques to identify image features. Despite the fact that methods that have been in use for a long time are usually more dependable and precise when it comes to determining a pose and navigational purposes, they frequently lack the ability to accurately deduce the scale without extra data. Simultaneously, the utilization of deep

learning can address the above issues by training convolutional neural networks with a considerable amount of data. As opposed to manually setting geometric limitations, DL-based approaches can get them by taking advantage of the pre-existing knowledge in the training information. Even when the shift in the apparent position of an object is not significant, it is still possible to calculate a reasonable position and depth. It is evident that online learning can be an effective tool for enhancing efficiency. Despite the advantages of deep learning-based approaches, the traditional methods still have high accuracy in estimating ego-motion [34].

5.1 Obstacle detection

Obstacle detection in VO (Visual Odometry) is a method of analyzing a sequence of images to detect the presence of obstacles or objects in the environment. It is used in navigation and localization systems, as well as in autonomous robots. By analyzing the images, the system can identify objects and obstacles, such as walls, furniture, or other objects that could potentially block the path of the robot or other entities. The optical flow algorithm is advantageous in that it can detect not only the location of a moving object, but also its speed and direction. Furthermore, since it does not need background modelling or updating, it is a very popular choice [22]. Currently, three predominant approaches for detecting obstacles by vision are monocular cues, stereo vision, and motion parallax [53]. The motion parallax method relies heavily on optical flow and uses it to get the movement and shape of both the viewer and the objects in the scene from a set of images. As indicated by Meneses et al. [54], optical flow was adopted to spot obstacles rather than calculating the movement of the robot. According to the definition of optical flow, areas with low optical flow intensity have less relative motion and thus a lower likelihood of containing obstacles. Through robot navigation, the optical flow data is conveyed to a Support Vector Machine (SVM) classifier with a radial basis function (RBF) kernel, which is used to determine if any obstacles are present in the intended path. This allows the robot to modify its course, usually in the direction with a lower intensity of optical flow [54]. Kendoul et al. [53] made use of the dense optical flow approach to gain knowledge about the entire environment and applied the Gunnar-Farneback method to calculate dense optical flow [15]. When the Unmanned Aerial Vehicle (UAV) is in motion, the presence of a multitude of optical flow vectors in its field of view suggests that an obstacle is present ahead. Conversely, if the magnitude of the optical flow vector is less than the set threshold, it implies that there is nothing blocking the UAV's path. The authors tackle the issue of foreground and background being intertwined by utilizing different thresholds to differentiate between regions, based on the magnitude of the optical flow vector. Subsequently, a clustering process is employed to

amalgamate similar regions [53]. Zhang et al. [55] proposed an auto-localization method for scenarios wherein tall buildings stand on both sides and there is no aerial view. To evaluate the rotation of the image captured by the monocular camera, they employed an optimization algorithm with the RANSAC algorithm to fit the points that correspond to the obstructions on both sides into a straight line, thus reducing the variance of errors in the forward-looking points of the robot. Then, they used Kalman filtering algorithm to determine the vanishing point of the straight line, which is the point at which the parallel obstacles on either side converge at an infinite distance. The Ackermann steering model and singular value decomposition are utilized in the algorithm for calculating the trajectory of the car, while wheel encoders are used to figure out the rate of translation. The application of optical flow in CNNs makes it easier to process objects in motion within very active settings. Rashed et al. [56] boosted the results of semantic segmentation by taking advantage of motion and depth information from optical flow when distinguishing between roads, structures, and trees, among others. A CNN architecture that is based on semantic segmentation through the combination of multiple modes of data is mainly used in autonomous driving, where earlier knowledge is utilized to make the segmentation process better [57].

5.2 Multi-sensor fusion

Under conditions where GPS is not available, the standard approaches to localization are usually SLAM (simultaneous localization and mapping), inertial IMU (inertial measurement unit) positioning, and visual localization. SLAM is precise, but the size and price of it are considerable; IMU apparatuses tend to be erratic, causing integral mistakes; and visual localization needs considerable real-time computing power support. Currently, multi-sensor fusion localization has been demonstrated to be successful in lessening localization errors, augmenting system robustness, and decreasing expenses. Consequently, it has become a popular approach among researchers. Shen et al. [5] put forward a multi-sensor fusion localization algorithm leveraging the Extended Kalman Filter (EKF). Visual Odometry (VO) was employed to calculate the velocity and position of the UAV, and a magnetometer was utilized to measure its attitude. Subsequently, the Extended Kalman Filter [5] was employed to adjust any drift observed in the inertial navigation system (INS). Kim et al. [35] suggested a way to enhance Shen's research by devising a Feature Point Threshold Filter (FPTF) algorithm that can enhance the performance of INS + Optical Flow sensor fusion by modifying the threshold according to the elevation and speed of the UAV. Due to the inability of monocular visual odometry to precisely determine depth and distance, a significant issue that it encounters is scale blur. To address this, Yu et al. [36] use prior information from the environment to find a solution. In their

research, a GPE (Ground Point Extraction) algorithm based on Delaunay triangulation [37] and ground points to combine ground points from subsequent frames is adopted, supposing a steady camera height over the ground and fitting the ground plane employing the least squares method. The Point Aggregation (GPA) algorithm determines the true scale by calculating the ground plane based on the collected data. Subsequently, a RANSAC-based optimizer is employed to solve a least squares problem in order to finalize the scale.

The results from experiments conducted on the KITTI dataset demonstrate that the framework put forward by Yu et al. [36] is effective in terms of both translation and rotation errors. Additionally, the framework demonstrates excellent computational efficiency, achieving a performance frequency of 20 Hz on the KITTI dataset. Mostafa et al. [4] suggested a new intelligent hybrid vision-aided inertial navigation system to deal with the issue of scale ambiguity in vehicle motion estimation that optical flow provides. The system is constructed of three distinct components: a Visual Odometry (VO) module, a Gaussian Regression Process (GPR) for predicting the movements of an Inertial Navigation System (INS) and another GPR for forecasting the drift of the VO. The operation can be broken down into three phases: first, when GNSS signals are available, the monocular VO and INS drift estimator are trained; second, the monocular VO drift estimator is trained to model the errors related to the speed estimates of the monocular VO; and finally, if GNSS signals are lost, the monocular VO drift estimator is used for prediction. This scheme has the advantage of being able to build up a representation of any discrepancies caused by the monocular visual odometry drift or the inside navigation system drift when GNSS data is accessible, as well as foreseeing these errors during times when GNSS readings are not available. Previous regression techniques using Gaussian Process Regression [38] and Support Vector Machine (SVM) [39] may yield unreliable results if there are insufficient features available or when there are frequent duplicated patterns. Mostafa et al. [4] put forth a technique that efficiently resolves the major issue of not being able to address missing optical flow vectors in some regions of an image due to conflicting matching. It has been demonstrated through experiments that the algorithm is able to effectively decrease positioning error during a GNSS signal loss. In comparison to the pure VO/INS procedure and VO/INS with GPR rectification, the algorithm reduced positioning error to 47.6% and 76.3% respectively, when a GNSS signal disruption happened for one minute. Xu et al. [40] present a multi-layer methodology for multi-target tracking which combines regular optical flow restrictions with product terms and utilizes a Sequential Convex Programming (SCP) technique to tackle the ensuing nonconvex optimization issue. To enhance the precision and dependability of the autonomous navigation algorithm of the aircraft, they modelled and inspected the mistakes of each sensor in the VO/IMU

integrated navigation system and investigated the Kalman filtering-based free combination filtering equation [40]. Nevertheless, drones and other miniaturized systems have restrictions when it comes to their dimensions, the amount of cargo they can carry, and the power they possess, issues that are regularly experienced in the realms of computer vision and robotics.

In recent times, researchers have been attempting to enhance the visual localization system, due to the challenge of needing high computing power and portability. The authors of the paper [41] by He et al. conducted an examination of the state-of-the-art edge-based visual odometry (EBVO) and developed an optimization framework known as PicoVO. This framework can effectively lower the computation and memory requirements. Santamaria et al. [43] came up with an efficient, cost-effective and high-performance approach to state estimation to allow MAVs to autonomously fly with minimal processing power requirements. This technique incorporates a smart camera with a monocular camera, an ultrasonic distance sensor and a three-axis gyroscope. The camera allows for high-frequency optical flow measurements, range of reflective surfaces, and three-axis angular rate to be taken, thus ruling out the need for the CPU to execute real-time image processing. The inventiveness of their proposed algorithm is that it does not rely on the optical flow information to determine the linear velocity, but rather directly observes the motion state with the initial optical flow information, thus separating the process and measurement noise. Pastor-Moreno et al. [6] created a system, OFLAAM, intended to be used on micro air vehicles (MAVs). The design of the system is composed of a downward-pointing optical flow camera, a forward-facing monocular camera, and an inertial measurement unit (IMU). The use of a localization module enables the mapping of these features into a specific vocabulary. This module applies the DBoW2 algorithm for the purpose of position correction by loopback detection [44] in order to address the problem of optical flow drift. By merging a high-speed optical flow localization with a low-rate positioning algorithm, an autonomous localization of the MAV can be achieved while also reducing the overall computational load [6]. Dong et al. [45] came up with a creative relative localization technique for utilization with unmanned aerial vehicles. They took photos of the ground from a camera attached beneath the drone, then used the SURF algorithm to identify points between two frames and the Fast Approximate Nearest Neighbors (FLANN) algorithm to determine the optical flow. The velocity of UAVs can be determined through the application of the optical flow motion estimation equation and known parameters. The proposed system incorporates measurements from a SINS, electronic compass, optical flow, altimeter system, and laser rangefinder to achieve relatively precise localization data. A summary of the sensor fusion VO algorithms is presented in *Table 3* below, which highlights the main issues and their respective solutions.

Table 3: Comparison of VO Algorithms for Sensor Fusion

Method	Objectives	Solutions
Shen et al. [5] Kim [35]	Implement multi-sensor fusion	- Create an EKF-based fusion algorithm. - Create a fusion algorithm using FPTF
Yu et al. [36] Mostafa [4] Xu et al. [40]	Solving scale ambiguity	- Train scale drift predictors in the presence of GNSS signals - examination of each sensor's error
PicoVO [42] Angel [43] OFLAAM [6] Dong [45]	Use navigation algorithms on low-computing platforms.	- real-time picture processing - Combining low-speed positioning techniques with high-speed OF - Using OF Motion Estimation

5.3 Estimate Speed-Distance-Position

Evaluating the state of motion is a significant area of research when it comes to localization, and the VO technique that is based on optical flow/feature matching can supply an abundance of details about self-motion. Ho et al. [46] applied an extended Kalman filter (EKF) in combination with images from a monocular camera to analyze the divergence of the temporal flow vector during the UAV's vertical touchdown, ascertain the altitude and vertical velocity of the UAV, and govern the UAV's landing [46]. For every taken picture, they used the FAST algorithm to discover corners and followed them in the subsequent frame utilizing a Lucas-Kanade tracker, which enabled them to quantify the contraction and expansion of optical flow and calculate the divergence of optical flow. Mumuni et al. [33] incorporated Structure of Motion (SFM) [47] with Optical Flow and Monocular Depth Estimation (MDE), to augment the precision of depth estimation. Generally, the dense measurements derived from MDEs are uncertain in scale, while the sparse depths derived from SFM models possess a metric scale. MDE furnishes abundant depth measurements, so only some sparse depth information from SFM is required to complete the picture. By combining these two approaches, the precision of the unmanned aerial vehicle's depth estimation can be improved. Recent years have seen impressive results from CNN-based optical flow models [48] [49], however Mumuni et al.'s [50] algorithm is more efficient in terms of memory and computation [50]. To apply optical flow for the purpose of localization on small unmanned aerial vehicles (UAVs), it is necessary to develop a more efficient optical flow algorithm. Because of that McGuire et al. [52] developed the EdgeFlow algorithm, which builds on a feature density distribution-based collision detection algorithm [51] by introducing a variable time horizon for subpixel flow detection and employing spatial edge distribution for the purpose of image motion tracking. Tests have shown that the procedure is

computationally effective enough to be executed at close to the speed of frames on restricted embedded processors, offering reliable speed and distance estimations for Unmanned Aerial Vehicles in unfamiliar environments. Zheng et al. [13] used the Lucas-Kanade sparse optical flow algorithm in their localization approach for indoor UAVs, which requires real-time optical flow calculation, and additionally implemented Forward-Backward bidirectional tracking optimization. They opted for ORB feature extraction for its real-time capabilities, KNN forward-backward bidirectional matching for feature matching, and RANSAC algorithm to filter the matching outcomes. According to the experimental results, this approach has a high degree of accuracy in predicting velocity and location [13]. A novel end-to-end network is suggested in work by Huang et al. [57] for learning optical flow and calculating camera ego motion. They utilized an autoencoder and estimated optical flow using the PWC-Net created by Sun et al [26] (CNN Encoder).

6. Conclusion

This review examines the optical flow-based visual odometry of localization without GPS. This research focused on the traditional approaches (such as feature extraction) and more unconventional methods (like those that involve deep learning). This review outlines the fundamental concepts for each category and, if relevant, illustrates how they are used in practice. Recent research into visual localization techniques has shown an evolving trend towards more cost-effective, compact solutions that offer greater precision. In recent years, deep learning has seen a surge in growth, prompting some scientists to explore using neural networks for visual localization tasks. This has produced noteworthy outcomes. FlowNet has been identified as the most prominent approach for extracting image motion information and providing localization support, due to its utilization of optical flow extraction and its improved algorithms.

At the end, this article checked and regenerated all the results in the previous research and reapplied all the algorithms with other data and compared the results with the old ones.

References

[1] Aguilar, Wilbert G., Verónica P. Casaliglla, and José L. Pólit, "Obstacle avoidance based-visual navigation for micro aerial vehicles", Electronics 6.1 (2017): 10, pp. 1–23.

[2] Chao, H., et al., "A comparative study of optical flow and traditional sensors in UAV navigation", in *Proc. American Control Conference. IEEE*, 2013, pp. 3858–3863.

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

[4] Mostafa, M. M., et al., "A smart hybrid vision aided inertial navigation system approach for UAVs in a GNSS denied environment", *Navigation: Journal of The Institute of Navigation* Vol. 65, no. 4, pp. 533–547, 2018.

[5] Shen, C., et al., "Optical flow sensor/INS/magnetometer integrated navigation system for MAV in GPS-denied environment", *Hindawi Publishing Corporation, Journal of Sensors*, 2016, pp. 1–10.

[6] Pastor-Moreno, D., Shin, H. S., and Waldock, A., "Optical flow localisation and appearance mapping (OFLAAM) for long-term navigation.", in *Proc. 2015 International Conference on Unmanned Aircraft Systems (ICUAS), IEEE*, 2015, pp. 980–988.

[7] Wei, W., et al., "A survey of uav visual navigation based on monocular slam", in *Proc. 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC)*, IEEE, 2018, pp. 1849–1853.

[8] Jeon, J., et al., "Run your visual-inertial odometry on NVIDIA Jetson: Benchmark tests on a micro aerial vehicle", *IEEE Robotics and Automation Letters, vol* 6, no. 3, pp. 5332–5339, 2021.

[9] Lowe, D. G., "Distinctive image features from scale-invariant keypoints", *International journal of computer vision*, vol. 60, pp. 91–110, 2004.

[10] Bay, H., Tuytelaars, T., and Van Gool, L., "Surf: Speeded up robust features", *Lecture notes in computer science* 3951, 2006, pp. 404–417.

[11] Rublee, E., et al., "ORB: An efficient alternative to SIFT or SURF", in *Proc. 2011 International conference on computer vision*, IEEE, 2011, pp. 2564–2571.

[12] Chen, L., et al., "Design of a multi-sensor cooperation travel environment perception system for autonomous vehicle", *Sensors 12.9*, 2012, pp. 12386–12404.

[13] Wenxuan, Z, Xiao, J, and Xin, T., "Integrated navigation system with monocular vision and LIDAR for indoor UAVs", in *Proc. 12th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, IEEE, 2017, pp. 924-929.

[14] Srinivasan, M., et al., "Honeybee navigation en route to the goal: visual flight control and odometry", *The Journal of experimental biology*, 199.1, 1996, pp. 237–244.

[15] Farnebäck, G., "Two-frame motion estimation based on polynomial expansion", in *Proc. 13 Image Analysis: 13th Scandinavian Conference, SCIA 2003*, Halmstad, Sweden, June 29–July 2, 2003, Springer Berlin Heidelberg, 2003, pp. 363–370.

[16] Lucas, B. D., and Kanade, T. "An iterative image registration technique with an application to stereo vision", in *Proc. IJCAI'81: 7th international joint conference on Artificial intelligence*, Vol. 2. 1981, pp. 674–679.

[17] Horn, B. K. P., and Schunck, B. G. "Determining optical flow", *Artificial intelligence* 17.1-3, 1981, pp. 185–203.

[18] Srinivasan, M. V., "An image-interpolation technique for the computation of optic flow and egomotion", *Biological cybernetics* 71.5, 1994, pp. 401–415.

[19] Farid, K., Fantoni, I., and Nonami, K. "Optic flow-based vision system for autonomous 3D localization and control of small aerial vehicles", *Robotics and autonomous systems*, 57.6-7, 2009, pp. 591–602.

[20] Zhang, L., Xiong, Z., Lai, J., and Liu, J., "Research of optical flow aided MEMS navigation based on convex optimization and ROF denoising", *Optik*, vol. 158, pp. 1575–1583, 2018.
[21] Boretti, C., et al. "Visual Navigation Using Sparse Optical Flow and Time-to-Transit", in *Proc. 2022 Intern. Conf. on Robotics and Automation (ICRA)*, IEEE, 2022, pp. 9397–9403.
[22] Li, L., Liang, S., and Zhang, Y., "Application research of moving target detection based on optical flow algorithms", *Journal of Physics: Conference Series*, vol. 1237., no. 2., IOP Publishing, 2019, p. 022073.
[23] Zhu, A. Z., et al., "EV-FlowNet: Self-supervised optical flow estimation for event-based cameras", arXiv preprint arXiv: pp. 1802.06898 (2018).
[24] Dosovitskiy, A., et al., "Flownet: Learning optical flow with convolutional networks", in *Proc. of the IEEE international conference on computer vision*, 2015, pp. 2758–2766.
[25] Ilg, E., et al. "Flownet 2.0: Evolution of optical flow estimation with deep networks", in *Proc. of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2462–2470.
[26] Sun, D., et al., "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume." in *Proc. of the IEEE conference on computer vision and pattern recognition*, 2018.
[27] Zhichao, Y., and Shi, J., "Geonet: Unsupervised learning of dense depth, optical flow and camera pose", in *Proc. of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1983–1992.
[28] Scharstein, D., et al., "High-resolution stereo datasets with subpixel-accurate ground truth", in *Proc. Pattern Recognition: 36th German Conference, GCPR 2014*, Münster, Germany, September 2-5, 2014, Proceedings 36, Springer International Publishing, 2014, pp. 31–42.
[29] Geiger, A., et al., "Vision meets robotics: The kitti dataset", *The International Journal of Robotics Research* 32.11, 2013, pp. 1231–1237.
[30] Butler, D. J., et al., "A naturalistic open source movie for optical flow evaluation." in *Proc. Computer Vision–ECCV 2012: 12th European Conference on Computer Vision*, Florence, Italy, October 7-13, 2012, Proceedings, Part VI 12. Springer Berlin Heidelberg, 2012, pp. 611–625.
[31] Zhu, A. Z., et al. "The multivehicle stereo event camera dataset: An event camera dataset for 3D perception", *IEEE Robotics and Automation Letters*, 3.3, 2018, pp. 2032–2039.
[32] Cordts, Marius, et al., "The cityscapes dataset for semantic urban scene understanding" in *Proc. of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213-3223.
[33] Mumuni, F., Mumuni, A., and Amuzuvi, C. K., "Deep learning of monocular depth, optical flow and ego-motion with geometric guidance for UAV navigation in dynamic environments", *Machine Learning with Applications*, 10, 2022, p. 100416.
[34] Zhang, J., et al., "Deep online correction for monocular visual odometry", in *Proc. 2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021, pp. 14396–14402.
[35] Taegyun, K., et al., "Improved optical sensor fusion in UAV navigation using feature point threshold filter", *International Journal of Aeronautical and Space Sciences*, 2022, pp. 1-12.
[36] Yu, T., et al., "Accurate and robust stereo direct visual odometry for agricultural environment", in *Proc. 2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021, pp. 2480–2486.
[37] Pinggera, P., et al., "Know your limits: Accuracy of long range stereoscopic object measurements in practice", in *Proc. Computer Vision–ECCV 2014: 13th European Conference*, Zurich, Switzerland, Sept. 6-12, 2014, Proceedings, Part II 13. Springer International Publishing, 2014, pp. 96–111.
[38] Guizilini, V., and Ramos, F., "Visual odometry learning for unmanned aerial vehicles", in *Proc. 2011 IEEE International Conference on Robotics and Automation*, IEEE, 2011, pp. 6213–6220.

[39] Ciarfuglia, T. A., et al. “Evaluation of non-geometric methods for visual odometry”, *Robotics and Autonomous Systems* 62.12, 2014, pp. 1717–1730.
[40] Xu, Q., et al., “An Optical Flow Based Multi-Object Tracking Approach Using Sequential Convex Programming”, in *Proc. 16th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, IEEE, 2020, pp. 1216–1221.
[41] Schenk, F., and Fraundorfer, F., “Robust edge-based visual odometry using machine-learned edges”, in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2017, pp. 1297–1304.
[42] He, Y., et al., “Picovo: A lightweight rgb-d visual odometry targeting resource-constrained iot devices”, in *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021, pp. 5567–5573.
[43] Santamaria-Navarro, A., et al., “Autonomous navigation of micro aerial vehicles using high-rate and low-cost sensors”, *Autonomous robots*, 42, 2018, pp. 1263–1280.
[44] Gálvez-López, D., and Tardos, J. D., “Bags of binary words for fast place recognition in image sequences”, *IEEE Transactions on Robotics*, 28.5, 2012, pp. 1188–1197.
[45] Zhuoning, D., Li, W., and Zhou, Y. “An autonomous navigation scheme for UAV in approach phase”, in *Proc. IEEE Chinese Guidance, Navigation and Control Conference (CGNCC)*, IEEE, 2016, pp. 982–987.
[46] Ho, H. W., de Croon, G., and Chu, Q. P., “Distance and velocity estimation using optical flow from a monocular camera”, *International Journal of Micro Air Vehicles*, 9.3, 2017, pp. 198–208.
[47] Ioannou, P. and Fidan, B., “Adaptive control tutorial”, SIAM, 2007.
[48] Liu, L., et al. “Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation.” in *Proc. of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 6489-6498.
[49] Jonschkowski, R., et al., “What matters in unsupervised optical flow”, in Proc. *Computer Vision–ECCV 2020: 16th European Conference*, Glasgow, UK, Aug. 23–28, 2020, Proceedings, Part II 16. Springer International Publishing, 2020, pp. 557–572.
[50] Mumuni, F, and Mumuni, A., “Bayesian cue integration of structure from motion and CNN-based monocular depth estimation for autonomous robot navigation”, *International Journal of Intelligent Robotics and Applications*, 6.2, 2022, pp. 191–206.
[51] Lee, D.-J., et al., “See and avoidance behaviors for autonomous navigation”, *Mobile Robots* Xvii, vol. 5609, SPIE, 2004, pp. 23–34.
[52] McGuire, K., et al., “Efficient optical flow and stereo vision for velocity estimation and obstacle avoidance on an autonomous pocket drone”, *IEEE Robotics and Automation Letters*, 2.2, 2017, pp. 1070–1076.
[53] Farid, K., “Survey of advances in guidance, navigation, and control of unmanned rotorcraft systems”, *Journal of Field Robotics*, 29.2, 2012, pp. 315–378.
[54] Meneses, M. C., Matos, L. N., and Prado, B. O., “Low-cost Autonomous Navigation System Based on Optical Flow Classification”, arXiv preprint arXiv:1803.03966 (2018).
[55] Zhang, J., et al. “Monocular visual navigation of an autonomous vehicle in natural scene corridor-like environments”, in *Proc. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2012, pp. 3659–3666.
[56] Rashed, H., et al., “Motion and depth augmented semantic segmentation for autonomous navigation”, in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 364–370.
[57] Huang, Y., et al., “Learning optical flow with R-CNN for visual odometry”, *IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 14410–1441.