

## Article

# Sparse Optical Flow Implementation Using a Neural Network for Low-Resolution Thermal Aerial Imaging

Tran Xuan Bach Nguyen <sup>1,\*</sup>  and Javaan Chahl <sup>1,2</sup> <sup>1</sup> School of Engineering, University of South Australia, Mawson Lakes, SA 5095, Australia<sup>2</sup> Aerospace Division, Defence Science and Technology Group, Edinburgh, SA 5111, Australia\* Correspondence: [nguty219@mymail.unisa.edu.au](mailto:nguty219@mymail.unisa.edu.au)

**Abstract:** This study is inspired by the widely used algorithm for real-time optical flow, the sparse Lucas–Kanade, by applying a feature extractor to decrease the computational requirement of optical flow based neural networks from real-world thermal aerial imagery. Although deep-learning-based algorithms have achieved state-of-the-art accuracy and have outperformed most traditional techniques, most of them cannot be implemented on a small multi-rotor UAV due to size and weight constraints on the platform. This challenge comes from the high computational cost of these techniques, with implementations requiring an integrated graphics processing unit with a powerful on-board computer to run in real time, resulting in a larger payload and consequently shorter flight time. For navigation applications that only require a 2D optical flow vector, a dense flow field computed from a deep learning neural network contains redundant information. A feature extractor based on the Shi–Tomasi technique was used to extract only appropriate features from thermal images to compute optical flow. The state-of-the-art RAFT-s model was trained with a full image and with our proposed alternative input, showing a substantial increase in speed while maintain its accuracy in the presence of high thermal contrast where features could be detected.

**Keywords:** optical flow; thermal imaging; LWIR; navigation; UAVs; deep learning



**Citation:** Nguyen, T.X.B.; Chahl, J. Sparse Optical Flow Implementation Using a Neural Network for Low-Resolution Thermal Aerial Imaging. *J. Imaging* **2022**, *8*, 279. <https://doi.org/10.3390/jimaging8100279>

Academic Editors: Alicja Kwasniewska, Hamed Mozaffari and Yudong Zhang

Received: 23 August 2022

Accepted: 7 October 2022

Published: 12 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The ability of unmanned aerial vehicles (UAVs) to navigate autonomously in unknown environments is vital for their further integration into human society. Currently, UAVs rely almost entirely on global navigation satellite systems (GNSS) for navigation applications. Nevertheless, GNSS are known to be unreliable in urban areas in urban canyons, or under forest canopies and are not available underground. Furthermore, GNSS systems do not provide any sensing capacity that might allow avoidance of unknown obstacles in the environment, thus making the solution less reliable in dynamic scenes.

Unlike other sensor-based systems, vision systems can provide real-time information about objects present in the scene. Furthermore, vision systems do not rely on signals coming from satellites such as GNSS, thus making it more resilient to conventional jamming [1]. Many researchers have demonstrated the potential of vision system for UAVs with promising results [2–6]. However, there remain challenges of using vision-based systems on UAVs due to the challenges of fusing spatial and temporal information from the sensors into a coherent model, which can be as simple as motion blur from high angular rates and lower rate of ground movement at higher altitude due to perspective [7] or lack of texture in scenes resulting in no information. Furthermore, due to high degrees of freedom of the UAVs, variations in roll, pitch and yaw of an aircraft with strap down cameras will result in different viewing angles and rates of image motion of the same scene captured from the same location [8].

A navigation system that can be deployed onto small UAVs must be small in size and light in weight. Small UAVs have limited payload capacity, which makes it difficult to utilise

more computationally demanding algorithms for better accuracy. Some researchers have tried to solve this problem by using a cloud-based computer processing system to process data transmitted from UAVs in real-time [9–12]. However, this type of solution has limited range and cannot operate very far from the ground control station. Furthermore, it also shares the same unreliability issues faced by sensor based systems where the connections are not always available. Hence, it is necessary to explore new navigation algorithms that are less expensive for UAVs.

Optical flow can be defined as the apparent motion of brightness patterns across two frames [13]. Optical flow is a computer vision technique that is often associated with insect inspired studies [2,14]. Flying insects are able to navigate in a dynamic environment with a tiny brain [3]. Furthermore, insects have been shown to rely on optical flow for takeoff and landing [2,15], obstacle avoidance [16], terrain following [17] and flight speed regulation [18]. In navigation, optical flow can be used actively as frontal obstacle avoidance and altitude control [6], or passively to collect current operating states of the aircraft such as pitch and roll [19,20], descent angles [21] and direction of travel or lateral drift [5,22,23].

Over some decades, optical flow algorithms have been dominated by spatiotemporal image processing techniques to compute optical flow. Some examples include the Horn and Shunck technique [13], the Farneback algorithm [24], the gradient based such as the Lucas–Kanade [25], correlation and block matching methods [26] and the image interpolation technique [27] to name a few.

With ease of access to more powerful graphic processing units (GPU), scientists have been experimenting with optical flow implementations based on deep learning concepts with great success. FlowNet [28] was the first model in the field but its efficacy was inferior to traditional techniques. FlowNet2 [29] was created by stacking multiple FlowNet layers, which vastly increased the efficacy and outperformed many traditional methods, but required much more memory, making this approach unsuited to current embedded systems on small drones. Later work focused on light-weight models by borrowing many popular concepts from traditional techniques. SpyNet [30] uses a coarse-to-fine approach while LiteFlow [31] relies on a brightness map to solve occlusion problem. PWC-Net [32] utilises stereo matching, feature extraction and cost volume resulting in high efficacy while having substantially smaller model size compared to FlowNet2. Most recently, the Recurrent All-Pairs Field Transform (RAFT) [33] and its lighter version, RAFT<sub>s</sub>, was introduced that achieved state-of-the-art efficacy while also having one of the lowest memory requirements. The RAFT models were inspired by optimisation-based approaches from traditional optical flow techniques.

While research activity is significant, there is a substantial gap in the literature into night operation given that this period represents approximately half of the potential operating time of a system. There are many reasons for this, such as the historically high cost of thermal sensors, the difficulty to operate after dark due to regulatory restriction and challenges to launch and retrieval of small aircraft at night.

In this paper, we explore a simple but effective technique that further enhances the performance of the RAFT<sub>s</sub> model in terms of how many frames can be processed per second with thermal imagery. This technique can potentially further decrease the computational requirement of deep learning based optical flow techniques, which makes it suitable for aerial navigation applications where a dense flow field is often unnecessary.

## 2. Related Work

Thermal imaging has various advantages over visual light spectrum in some applications, not only to aid navigation in challenging lighting scenarios but also to reveal information that is invisible to the naked eye. The physics fundamentals of thermal sensors and their advantages and disadvantages are well documented in [1,34,35].

Beside being studied for autonomous navigation, thermal imaging has been used to monitor railway infrastructure [36], monitor crops [37–39], driver monitoring [40], face

recognition [41], vital sign extraction [42] and for COVID-19 detection [43,44] just to name a few.

In the navigation domain, earlier work has demonstrated encouraging results when combining long-wave infrared (LWIR) with optical light wavelengths to detect hidden features in dark scene. Maddern et al. [45] relied on LWIR thermal sensor to enhance the tested system over long periods of time by compensating the adverse effects from solar glare on RGB images. Brunner et al. [46] used thermal data to detect and reject poor features such as dust and reflective surfaces that are visible in the visual spectrum. Mouats et al. [47] proposed a multispectral stereo odometry system for unmanned systems and later developed a purely thermal stereo odometer for UAVs [48]. In the odometer, a pair of thermal sensors were located in front of the UAV to capture data, and were later used for feature matching and thermal 3D reconstruction. The results demonstrated that a thermal odometry system could produce comparable outcomes to a standard visible spectrum system.

While these results contributed to the field, they will struggle in some scenes when significantly hotter or cooler objects enter and leave the scene, due to lack of a high dynamic range calibrated radiometric sensors. The automatic gain control (AGC) system is provided with thermal sensors to compensate for drastic changes in the overall pixel intensity of the thermal image. However, they are designed to support the display of information to humans with limited ability to resolve intensity variations in a scene. The gain control causes substantial changes in scene contrast to accommodate the range of temperatures of objects in the scene.

Modern radiometric thermal sensors produce high bit depth images to represent thermal emissivity. A radiometric sensor is calibrated against a standard black body by the manufacturer, greatly improving the consistency and accuracy of thermal data [49]. Furthermore, these sensors are small, light and have low power consumption, which makes them suit the weight and size constraints of UAVs.

Using these radiometric sensors, Khattak et al. [50] proposed a multi-spectral fusion approach combining optical light and LWIR sensors to allow small UAVs to navigate in a dark tunnel without the presence of GNSS. The drone carried a short range illumination source for the optical sensor while the thermal sensor provided thermal data at longer range. They later developed a thermal-inertial system [51] which was capable of results comparable to other state-of-the-art visual methods used in mines underground.

Khattak et al. [52] also proposed a low-cost solution using thermal fiducial markers to help UAVs navigate reliably in a laboratory environment. The team [53] later developed a purely thermal-inertial system that utilised full 14 bit radiometric data. The study demonstrated that by using direct pre-AGC 14 bit thermal data, they could not only overcome the troublesome AGC process, but also increased the resilience and consistency of signals against loss of features over time.

Most recently, they [54] proposed fusion of thermal and visual inertial with a LIDAR (light detection and ranging) sensor to improve reliability for pose estimation. The team in [55–57] attempted to construct a 3D map of the environment with thermal sensor. They used a combination of range camera and thermal camera to collect real-time indoor data, to which 3D point clouds are matched via a RANSAC-based EPnP scheme. However, LIDAR and SLAM (simultaneous localisation and mapping) techniques are generally computationally expensive to run, making it undesirable for small UAVs. In an attempt to solve this issue, Lu et al. [58] presented an unsupervised deep learning network that can construct a 3D environment from thermal images for low light conditions, as an alternative technique to LIDAR sensor. The results showed this proposed technique is capable of providing a good 3D map in tested sequences, but those sequences are relatively simple.

Further, also utilising direct 14 bit thermal data, Shin and Kim [59] proposed a direct thermal-infrared SLAM algorithm to measure up to six degrees of freedom of the UAV. The results showed that the 14 bit data increased the robustness of the system in adverse lighting conditions.

The rest of this paper is organised into nine sections. Section 3 outlines our previous works and motivations for this study. Section 4 introduces the thermal sensor and revisits our technique to solve the brightness constancy problem associated with optical flow from thermal imaging. Section 5 outlines the difference between the sparse and dense optical flow technique for aerial applications and the feature extraction algorithm used in this study. Section 6 presents our collected dataset and our method to generate ground truth from real-world data. Section 7 presents the neural network that was used in this study, the two RGB datasets, how the neural network was trained and the evaluation methodology. Sections 8 and 9 report and analyse results from the experiment. Section 10 outlines lessons learnt and possible future research directions.

### 3. Motivations and Contribution

This study continues our work [60–62] to explore aerial applications of optical flow from low-resolution thermal images, “Thermal Flow”. Thermal Flow was implemented as a downward-looking system, mounted beneath the UAV, that output 2D optical flow vectors in X and Y displacement to track the movement of the platform. Thermal Flow was designed to mimic the output of the very efficient and popular PX4Flow [20] system, which can be integrated easily into available autopilot systems, such as the PixHawk [61].

RADAR systems such as circular scanning millimetre-wave [63–66] and LIDAR [67–70] have been implemented on UAVs, both of which are high-performance range sensors that emit electromagnetic energy at different frequencies. Emissions require substantial power from the platform, and in large systems, potentially levels of energy that are dangerous to humans and might interfere with other sensors or airspace users. Active sensors usually include effective range as one of their primary performance metrics, limited by power emitted and sensitivity to signal received. Passive sensing using computer vision techniques generally need platform motion or binocular separation and require onboard processing with sophisticated algorithms. There are advantages as well, including that range is limited only by platform motion (or platform binocular separation) and reflectance or illumination of the target, not by emitted radiation levels. All active sensors run the risk of having their emissions being detected in contested environments, indicating the presence of the UAV and potentially the nature of its activities to an adversary.

Our previous work relied on traditional optical flow techniques. On the other hand, deep-learning-based optical flow networks outperform traditional techniques in various synthetic datasets in several key benchmark criteria [71]. However, these techniques are very expensive to run and require an integrated GPU system to run in real time, which is not always suitable for aerial applications on small UAVs. Hence, in this study, we want to explore a new technique to further reduce the computational requirements of deep learning optical flow models.

The state-of-the-art RAFT<sub>s</sub> model [33] was chosen due to it achieving very high accuracy with small memory footprint and fewest parameters. Nevertheless, the model is still computationally expensive compared to the very popular sparse Lucas–Kanade. One of the reasons is that the network was designed to produce a dense optical field, which is not necessary in this application of Thermal Flow, which only requires a single reliable 2D vector. Instead, a sparse technique is preferred over the dense technique due to its much less computational cost, so the onboard computer system can be smaller to satisfy physical constraints of UAVs.

To achieve this, we are inspired by the use of the Shi–Tomasi [72] algorithm by the Lucas–Kanade algorithm [25]. The features then were combined into a new image, with smaller size, to be used as an alternative input to the network. Since the input is smaller, this technique can potentially decrease the computational requirement for the task while maintaining accuracy. This study aims to bring a deep-learning-based optical flow network that outperforms traditional techniques onto UAVs by reducing the computational requirement.

#### 4. Optical Flow with Thermal Imaging

This section outlines the thermal sensor used to collect the data in this study. This section also presents the problems with optical flow estimation when re-scaling to an 8 bit image format from 14 bit raw data, and revisits the technique to improve this process.

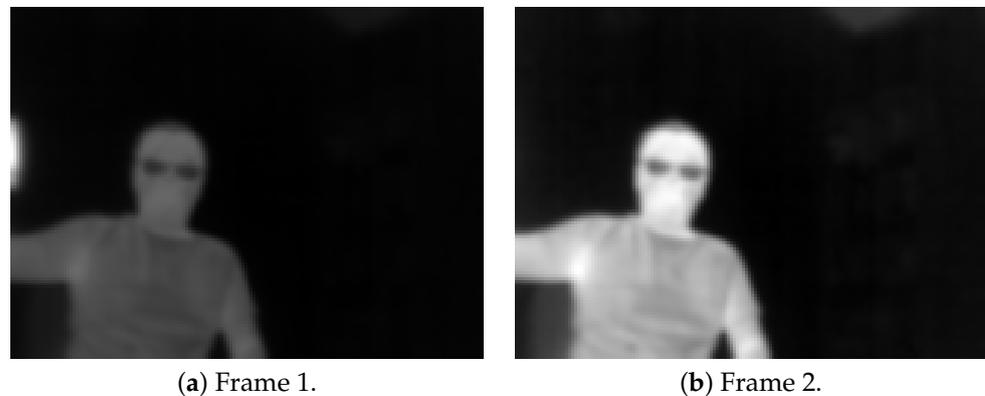
##### 4.1. Thermal Sensor

All of the images used in this study were captured by the radiometric FLIR Lepton3 [73]. The Lepton3 is a long-wave infrared sensor, and was calibrated in manufacturing against a standard black body [49]. The sensor can output  $160 \times 120$  pixels at 8.7 Hz. The sensor has a  $56^\circ$  field of view and has been shown to be adequate for airborne applications with small angular movements without further need for re-calibration [60].

##### 4.2. Automatic Gain Control

Most thermal sensors, such as the FLIR Lepton 3 that was used to collect our data, output radiometric data in 14 bit format. On the other hand, the RAFT\_s network, and available computer vision libraries, such as OpenCV [74], are designed to process 8 bit images. This is largely due to modern standard displays being designed to match the intensity discrimination of human observers [1]. The 14 bit raw data from the sensor must be converted into 8 bit to display on the screen or to work with available computer software.

Thermal sensors such as the FLIR Lepton include an automatic gain control (AGC), which improves the contrast of the image when converting to 8 bit when there is a dramatic change in the temperature range present in an image. Figure 1 displays one example of a hot cup moving out of the screen, which shows a dramatic change in image contrast between consecutive images. This contrast change is likely to cause problem for many image matching algorithms. This change between frames violates the main assumption of optical flow equation, which is brightness constancy [13].

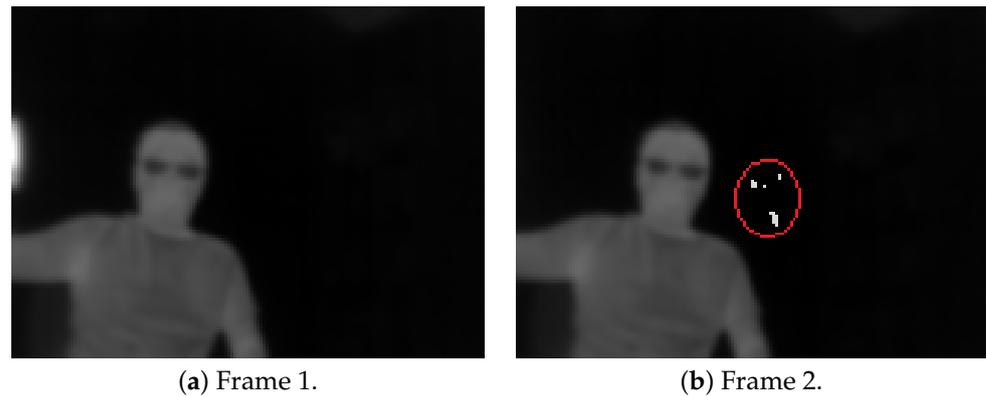


**Figure 1.** AGC changes the contrast in the images when a hot cup exits a scene: 1–2.

There were attempts to solve this problem. One approach [48] has been to greatly reduce the AGC response time so that image contrast does not change rapidly. Nevertheless, this technique only reduces the problem and does not solve it completely. Another group in [75] proposed an approach to manually set the range of the AGC. However, this approach requires prior information about the scene, which makes it less adaptable to unknown environments.

We revisit our conversion technique in [61] to re-scale two consecutive 14 bit images to 8 bit from the maximum and minimum pixel intensities found across both images. Figure 2 shows our technique. The technique, however, introduces some negligible artefacts.

Figure 2 shows the output of the sample images from Figure 1 with our technique. The contrast of the pair of images is maintained that can satisfy the optical flow requirement of brightness constancy.



**Figure 2.** A pair from figure 1 with our technique with small artefacts circled in red.

### 5. Sparse and Dense Optical Flow Technique in UAV Navigation

A broad distinction can be made between dense and sparse optical flow techniques. Dense optical flow techniques are designed to compute an optical flow vector for every pixel within the image. On the other hand, sparse techniques only output optical flow for selected parts of the image. As a result, sparse techniques will typically require less computing resources than the dense techniques [76].

Thermal Flow is designed to mimic the output of the PX4Flow device, which is a 2D vector,  $flow\_x$  and  $flow\_y$ , which indicates the movement of the aircraft in X and Y displacements. The Thermal Flow system is intended to be mounted underneath the aircraft looking straight down possibly to augment navigation, which leads to a relatively simple optical flow field, compared to looking at shallower angles that might include the horizon. A dense optical flow field is not desirable in this application due to the high computational cost, which limits its use for small UAVs. The sparse technique on the other hand, has been shown to achieve sufficient accuracy for navigation applications in real life in various studies [2,5]. Therefore, it is possible that the sparse technique can be applied to greatly reduce the size of the data feeding into the neural network to reduce processing time while maintaining the accuracy.

#### Feature Extraction

There are two primary feature extraction strategies: traditional corner-detection-algorithm-based and deep-learning-based frameworks. Deep-learning-based techniques include the direct visual odometry (VO) framework [77–79] and the 3D mapping model [80,81], etc. Traditional techniques are based on grayscale changes in the images such as the Harris technique [82], and its improved version, the Shi–Tomasi algorithm [72]. In general, CNN-based algorithms can take three channel RGB images as input while the traditional techniques require the images to be converted to grayscale single channel format. Generally, the deep-learning-based algorithm have performed better than conventional techniques in challenging sequences but are also much more computationally expensive to run.

Since thermal images are in single channel grayscale format, and this study focuses on improving speed, the Shi–Tomasi technique was selected as the feature extractor in this study since it works well in practice and is much cheaper to run.

Consider a sub-window in the image located at position  $(x, y)$  and the pixel intensity at this location is  $I(x, y)$ . When the sub-window shifts to a new position with displacement  $(u, v)$ , the pixel intensity at this position can be expressed as  $I(x + u, y + v)$ . The difference in pixel intensities of the window shift can be expressed as:

$$\delta = I(x + u, y + v) - I(x, y) \quad (1)$$

For good features in the thermal image, the difference is normally high. Let  $w(x, y)$  be the weights of pixels over a window; we differentiate Equation (1) with respect to X and Y

axes. The weighted sum multiplied by the intensity difference for all pixels in a window,  $E(u, v)$  can be defined as:

$$E(u, v) = \sum_{x,y} *w(x, y) * \delta^2 \tag{2}$$

Applying Taylor series expansion to the first order, to the shift intensity, we have:

$$I(x + u, y + v) \approx I(x, y) + \frac{\partial I(x, y)}{\partial x} u + \frac{\partial I(x, y)}{\partial y} v \tag{3}$$

Let:

$$I_x = \frac{\partial I(x, y)}{\partial x} \tag{4}$$

and

$$I_y = \frac{\partial I(x, y)}{\partial y} \tag{5}$$

Equation (2) becomes:

$$E(u, v) = \sum_{x,y} *w(x, y) * (I_x u + I_y v)^2 = \sum_{x,y} *w(x, y) * [(I_x u)^2 + (I_y v)^2 + 2I_x I_y uv] \tag{6}$$

Rewriting Equation (6) in Matrix notation gives us:

$$E(u, v) \approx (u, v) M \begin{pmatrix} x \\ y \end{pmatrix} \tag{7}$$

Hence:

$$M = E(u, v) \begin{pmatrix} \sum_{x,y} I_x^2 & \sum_{x,y} I_x I_y \\ \sum_{x,y} I_x I_y & \sum_{x,y} I_y^2 \end{pmatrix} \tag{8}$$

The score for each window  $R$  can be found using the eigenvalues of the matrix, which can be expressed as:

$$R = \det(M) - K(\text{trace}(M))^2 \tag{9}$$

where:

$$\det(M) = \lambda_1 \lambda_2 \tag{10}$$

and

$$\text{trace}(M) = \lambda_1 + \lambda_2 \tag{11}$$

In the Shi–Tomasi technique,  $R$  then can be found by:

$$R = \min(\lambda_1, \lambda_2) \tag{12}$$

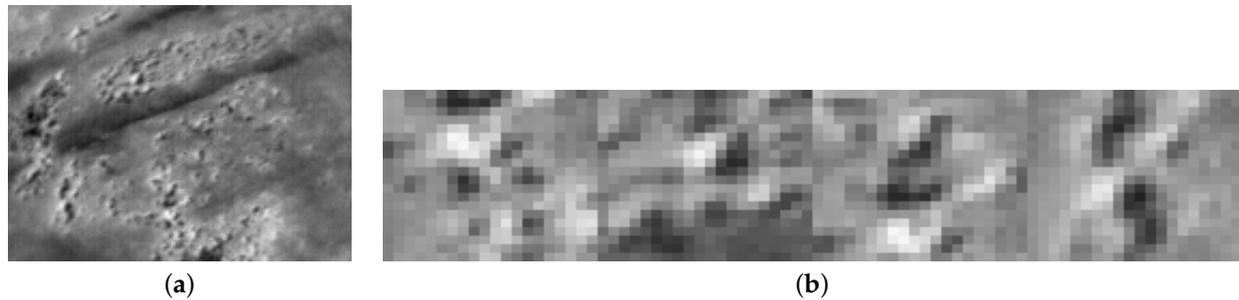
The  $R$  value represents a quality value of the correspondent corner, where a higher value indicates that the corner is a good distinct feature. We relied on the implementation of the technique in OpenCV. The parameter values are shown in Table 1.

**Table 1.** Parameter settings for the Shi–Tomasi corner detection algorithm.

Feature Detection Setting	Maximum corners	1000
	Quality level	0.02
	Minimum distance	5
	Block size	5

The returned  $R$  value of the corner will be ranked, and the highest  $R$  value corners will be chosen first. After detecting good corners, a “cropping window” parameter will be applied to crop the surrounding pixels with the chosen corner at the middle, resulting

in several sub-images. Then, these sub-images will be stitched together as an alternative image, as shown in Figure 3.



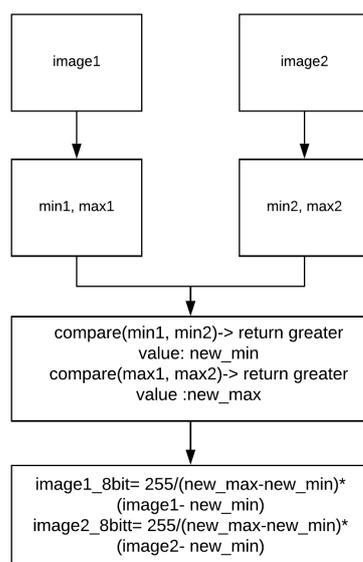
**Figure 3.** The original image (a) and the alternative input (b). The original size of (b) is  $40 \times 160$  compared to (a)  $160 \times 120$ , which is one third of the number of pixels. (a) Sample thermal frame. (b) A new image constructed from extracted features from the original thermal frame with  $40 \times 40$  as the cropping window and four features (shown at four times magnification).

In some cases, the total number of features that can be found is less than the parameter value, such as only three found compared to four needed, then the algorithm will take three instead. If there are no features to be found, the software will set the flow vector value to zero.

### 6. Thermal Dataset Availability

To the best of our knowledge, there is no currently available 14 bit thermal dataset with optical flow ground truth. All the datasets that current networks use for training and validation are generated synthetic colour dataset with known ground truth. On the other hand, obtaining real optical flow ground truth from real-world data is extremely challenging [71] due to the high degrees of freedom of UAVs.

To solve this problem, we generated ground truth from real-world thermal 14 bit data we had collected [60,62]. The 14 bit raw thermal data were downsampled to 8 bit with our technique as shown in Figure 4. After that, the traditional dense optical flow technique, Farneback [24], was used to generate ground truth from the images.



**Figure 4.** The 14 bit to 8 bit downsampling technique from [61].

Although civilian drones are restricted to daylight hours under visual flight rules, military missions must be possible in all weather at all times of day and beyond visual line of sight. Night flights were conducted with approval in military airspace described in [60].

### 6.1. Dataset 1

Dataset 1 contained images from our work from [60], which includes 12,870 images captured above a flat arid field in northern South Australia. The data were captured in late summer, during clear and hot weather with the temperature at 34 °C [83]. Figure 5 shows a colour image of the field.



Figure 5. View of the field of Dataset 1.

Figure 6 shows some 8 bit thermal images of Dataset 1.

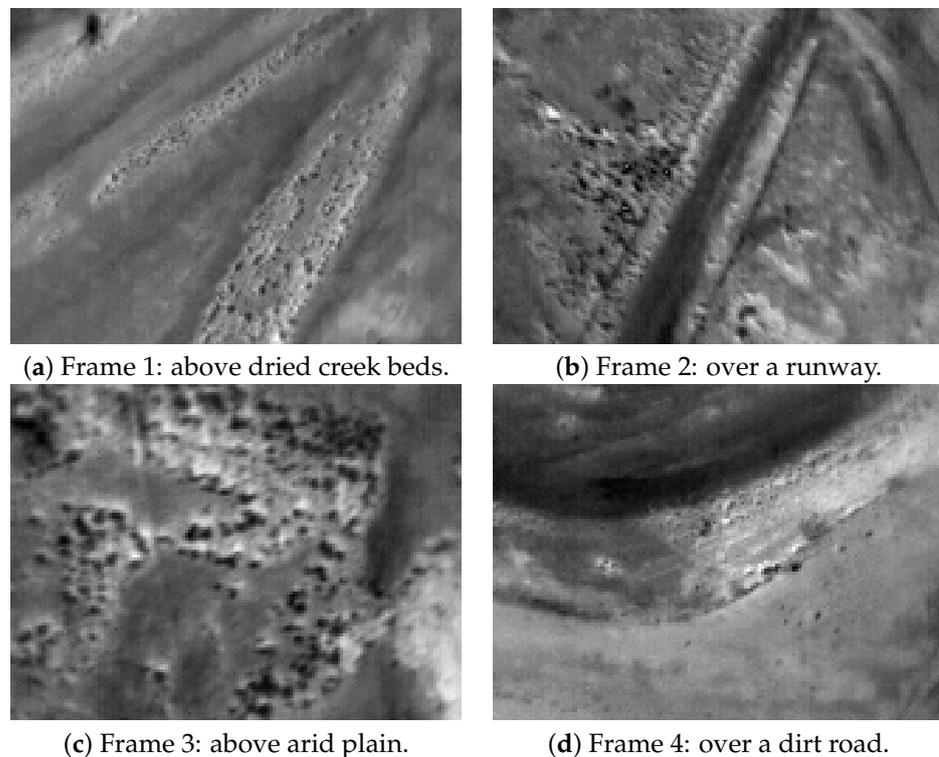


Figure 6. Some thermal frames of dataset 1, over some interesting features of the field.

### 6.2. Dataset 2

Dataset 2 contained a total of 2800 images from our work in [62], from an empty field on a hill in South Australia. The site provides a clear view of the sky and an empty ground with minimal artificial objects. Figure 7 shows the flight path in this experiment. The UAV took off at point H, flew to point (1)-(2)-(3)-(4)-(5) and then landed at (5).



Figure 7. View of the environment of dataset 2 with its flight path in Mission Planner.

The data were captured at two different thermal contrast conditions: during a clear sunny day in late Autumn at 1600 h with high contrast in thermal data, and during a foggy rainy day during winter at 0900 h, which yields low thermal contrast. Hence, dataset 2 contains two smaller subsets of the same field that can be used to evaluate the performance of thermal flow during high and low-contrast conditions. Figure 8 shows some of the thermal images of the site under both conditions.

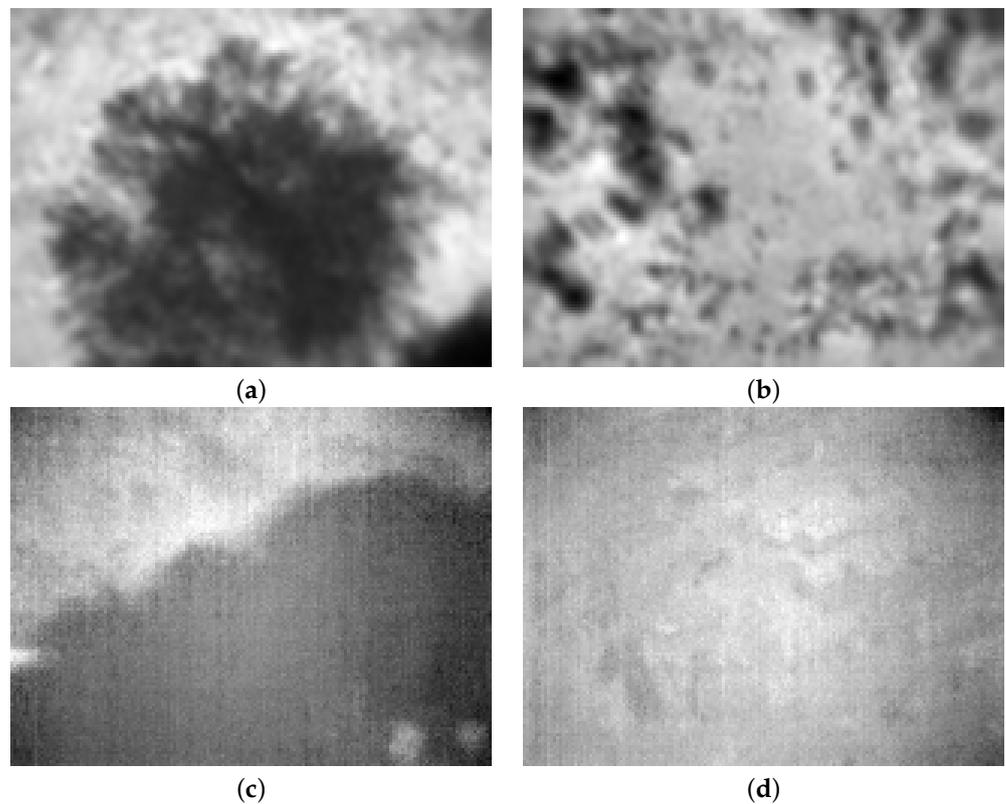


Figure 8. Some of images from dataset 2. Frame (1) and (2) shows the field during high-contrast conditions, and frame (3) and (4) shows thermal images at approximately at the same locations but under low-contrast conditions. (a) Frame 1: above a big tree during the high-contrast condition. (b) Frame 2: over an empty field during the high-contrast condition. (c) Frame 3: over a big tree during the low-contrast condition. (d) Frame 4: over an empty field during the low-contrast condition.

### 6.3. Training and Validating Sets

Table 2 shows a summary of our datasets, including how each dataset will be used for training and evaluation, number of images of each set and the conditions of the scenario.

**Table 2.** Characteristics of the collected dataset.

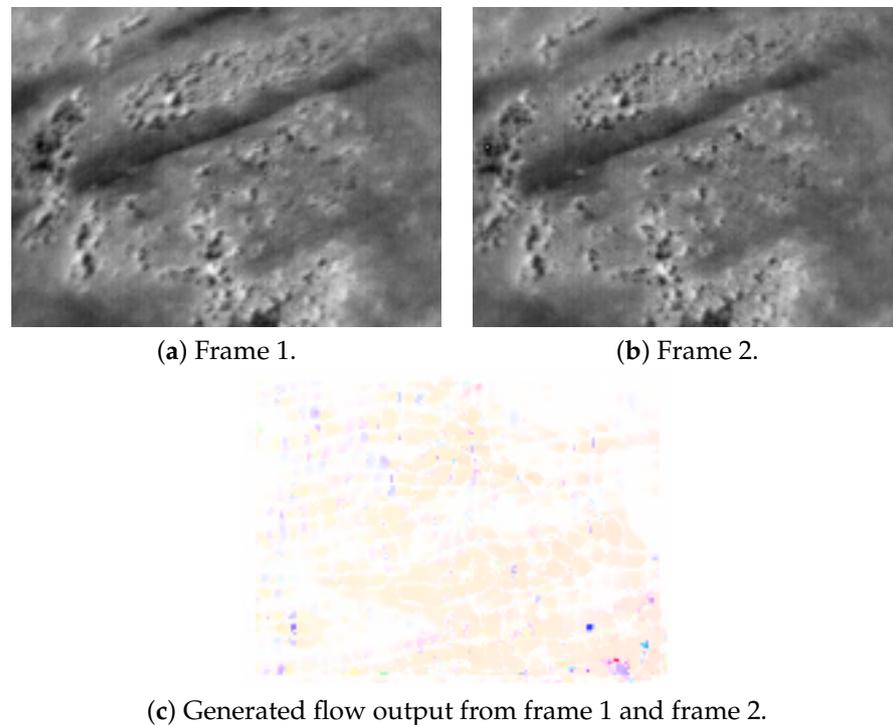
	Source	Training Set	Evaluation Set	Site Condition	Total Images
Dataset 1	[60]	Yes: 10,894	Yes: 2000	High contrast	12,894
Dataset 2	[62]	No	Yes: 2800	High and low contrast	2800
Total images		10,894	4800		15,694

In total, 10,894 images were used to train the RAFT\_s network and 4800 images were used to evaluate the data during different thermal contrast conditions.

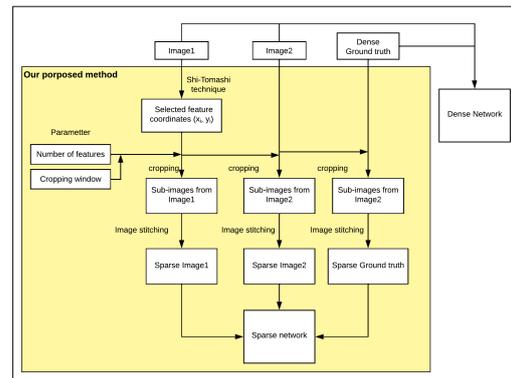
### 6.4. Generated ground truth from Thermal Dataset

Figure 9 shows a pair of images from dataset 1 and the dense optical flow ground truth generated by the Farneback technique implemented in OpenCV. To train the sparse network, the data from the dense network, including a pair of images, will be cropped at the location where coordinates of good corners were detected from Image1. Then, the respective sub-images will be aligned side by side from left to right to produce the training set for the sparse technique.

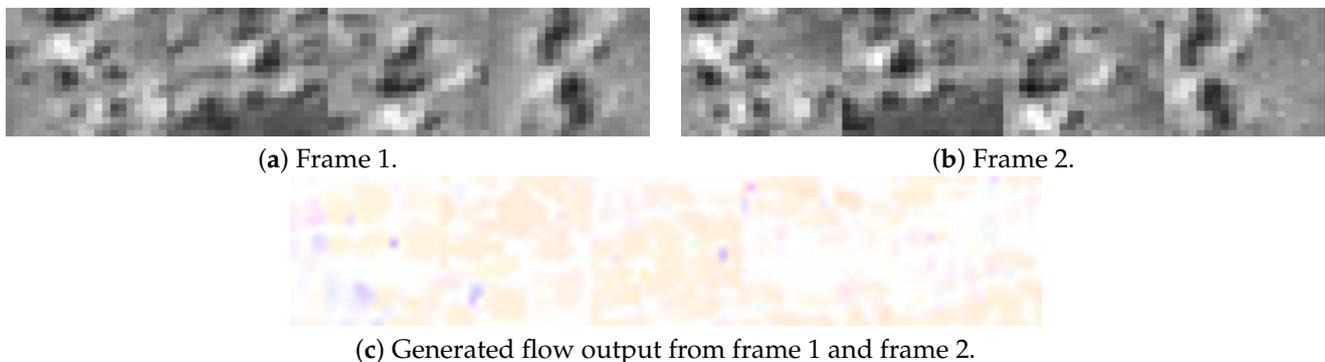
The overall process is shown in Figure 10; Figure 11 shows the result from Figure 9.



**Figure 9.** A sample sequence of thermal data and its generated ground truth from dense the optical flow Farneback algorithm in OpenCV.



**Figure 10.** Flowchart shows our proposed technique. The dense network is trained with original images and ground truth. The big yellow block shows the process of our method to select good features with Shi–Tomasi technique, combine with predefined parameters to crop original frames and reconstruct new frames from those sub-images, which are the new inputs to the neural network.



**Figure 11.** Sparse features used as input from images in Figure 9. The cropping window is  $40 \times 40$  pixels and number of features is four. All images shown here were magnified for visual purposes.

### 7. The RAFT<sub>s</sub> Model

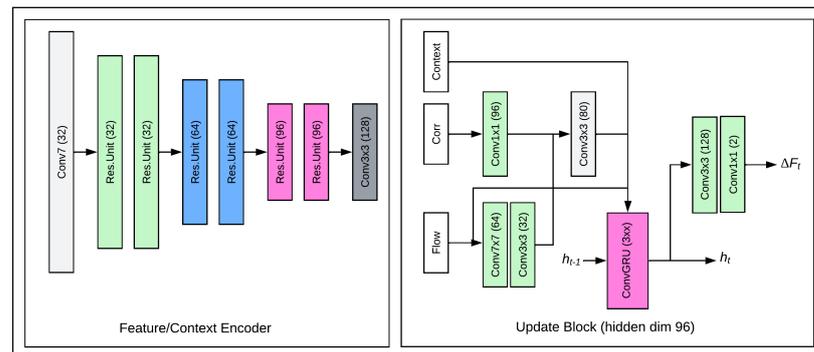
The RAFT (recurrent all-pairs field transform) deep learning model consists of a composition of convolutional layers and recurrent layers in three main blocks: a feature context encoder, a convolution layer and a recurrent gated recurrent unit (GRU)-based layer.

The model extracts per pixel features and updates the flow field iteratively through a recurrent unit from the correlation volumes. In the feature extraction layer, two frames are taken as input where features are extracted separately similar to the FlowNetCorr model. The convolution layer consists of six residual layers with resolution halved on every second layer while the number of channels are increased. The model uses a single GRU with  $3 \times 3$  filter in the GRU-based layer.

Figure 12 shows the RAFT<sub>s</sub> model used in this study.

#### 7.1. RGB Optical Flow Dataset

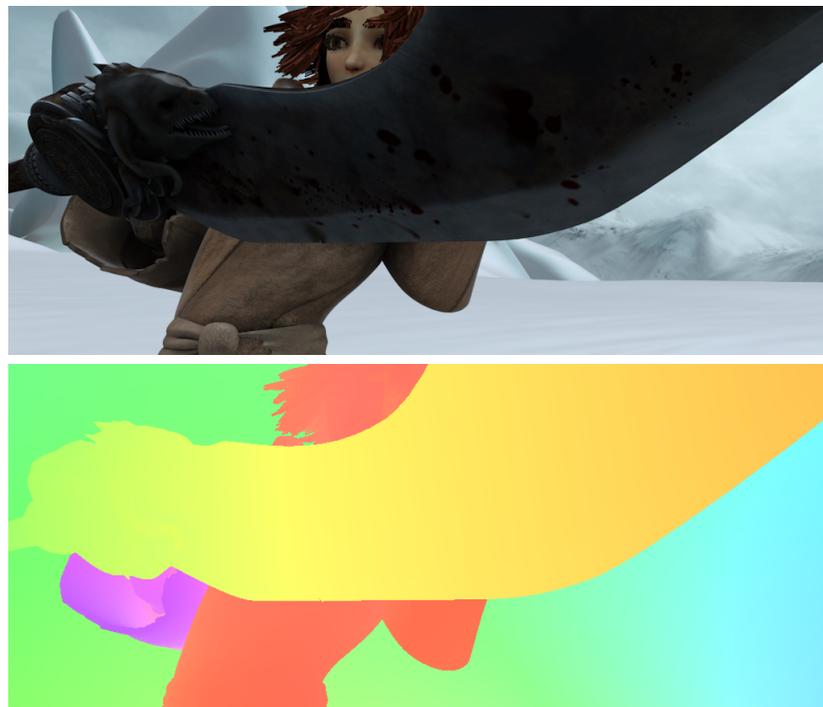
We transferred learning with pre-trained weights from the MPI-Sintel final and flying chairs dataset. The flying chairs dataset was selected due to it containing a large number of images that represent 2D motion, while the MPI-Sintel final introduces more complex 3D motion under more challenging lighting conditions. While the dominant motion of the thermal dataset is 2D, there are still 3D motion effects in some scenes due to the large number of degrees of freedom of UAVs. Hence, introducing a dataset with some 3D motion is essential. The details of the two datasets are presented in the following sections.



**Figure 12.** The model was used in this study [33].

### 7.1.1. MPI-Sintel

Prior to 2015, MPI-Sintel [84] was the largest dataset for optical flow and disparity measurement. The frames within the dataset were extracted from open source 3D animated movies, so the MPI-Sintel is entirely synthetic. With high resolution at  $1024 \times 436$  pixels, the frames included added effects from nature such as motion blur, fog or sun glare to make them more realistic. The training set consists of 1064 frames, which were divided into 23 training sequences. The evaluation set consists of 564 frames with 12 testing sequences. Dense optical flow ground truth is only available with the training set. The dataset provides three version: Albedo, clean and final. Albedo is the simplest set without any added effects, the clean version introduces small variation to the illumination between sequences and the final version adds more drastic effects. Researchers have been commonly using the clean and final versions over the Albedo. Figure 13 shows one example from the MPI-Sintel final dataset.



**Figure 13.** An example of an image (up) and its dense optical flow ground truth (down) from the MPI-Sintel final dataset.

### 7.1.2. Flying Chairs

The dataset were introduced along side the first deep-learning-based model, FlowNet [28]. It was designed specifically for training the deep learning network. The

frames were constructed by placing 3D chair models above random backgrounds from Flickr. The dataset contains 22,872 frames with 22,232 for training and 640 for evaluation sets. The dataset is entirely synthetic and does not contain 3D motion, so it is limited to a single view only. Figure 14 shows one example from the dataset.



**Figure 14.** An example of an image (**up**) and its dense optical flow ground truth (**down**) from the flying chairs dataset.

### 7.2. Train the Model

We transferred learning with pre-trained weights from the MPI-Sintel and flying chairs dataset, and our generated ground truth thermal data from the dataset as described in Sections 6 and 6.4. The dense technique was trained with the ground truth from the whole image, as shown in Figure 9, and the sparse technique was trained with the cropped data, as shown in Figure 11.

For the technique that utilises the whole image, we labelled it “dense” to differentiate it from the model that uses the proposed technique, we call “sparse”.

The model was trained with the batch size of 10, with 160,000 steps, learning rate of 0.0001 and weight decay of 0.0001.

The network was trained on a computer with an Intel Core i7-7700 CPU, 64GB of RAM and Nvidia GTX 1080 Ti GPU. The operating system was Ubuntu 20.04, other programs including: pytorch version 1.6.0, torchvision 0.7.0, cudatoolkit 10.1, python 3.8 and OpenCV 4.5.5.

### 7.3. Evaluation Methodology

The dense and sparse models are evaluated on two criteria: Accuracy and speed. The speed is measured on how many frames per second (FPS) the network can process. The speed produced by the dense network is labelled as “dense FPS”, and the speed produced by our method is labelled as “sparse FPS”. The “difference” parameter is the percentage difference between the “sparse FPS” and the “dense FPS”. The “dense FPS” rate was 11 FPS from our experiment.

Accuracy is measured based on the normalised cross-correlation between the output signals from each model of the same image sequence. The normalised cross-correlation value is in the range  $[-1;1]$ , with a value close to 1 indicating the two signals are the same, and vice versa.

There is a relationship between speed, cropping window and the number of features. A bigger cropping window and a larger number of features will likely return better accuracy but with lower speed. This relationship was investigated.

We learned that the cropping window at  $40 \times 40$  pixels and the total of four features works reliably, and provide a balanced of speed and accuracy. Hence, these parameters are applied to the presented signals.

## 8. Result

This section presents the signals of the dense and sparse technique from the evaluation set. The parameter for the sparse technique is: “cropping window” at  $40 \times 40$  pixels and four features.

### 8.1. Signals Accuracy

In this section, the overlay of dense and sparse signals over X and Y displacements and the normalised cross-correlation value are presented.

#### 8.1.1. Dataset 1

Figure 15 shows the dense and sparse signals from the sequence of dataset 1. A very high normalised cross-correlation value of 0.988 in the X and 0.968 in the Y displacements indicating a strong relationship between the two signals, which means that the sparse technique are capable of maintaining accuracy with significantly less input data.

The average number of features found in this case matched the set value, which is four.

#### 8.1.2. Dataset 2 during High Thermal Contrast Conditions

Figure 16 shows the dense and sparse signals from the sequence of dataset 2. A very high normalised cross-correlation value of 0.989 in X and 0.94 in Y displacements indicate a strong relationship between the two signals; thus, the sparse technique performs comparatively well in this scenario.

The average number of features found in this case also matches the set value, which is four.

#### 8.1.3. Dataset 2 during Cold-Soaked Low Thermal Conditions

The number of features drops significantly; Figure 17 shows the dense and sparse signals from the sequence of dataset 2. A low normalised cross-correlation value of 0.06 in X and 0.16 in Y displacements indicating a very weak relationship between the two signals. Furthermore, most of the normalised correlation values in both X and Y displacements are negative, which indicates the sparse technique may not work well in low contrast environments. This result is consistent with our findings with other techniques in [61]. Since the Shi–Tomasi operator relies on contrasting features to detect good corners, this technique falls short in the the low contrast thermal condition.

Furthermore, since the Shi–Tomasi operator was not very effective, the number of features found was less than the set value of four. The average value of features found in this case was 1.83.

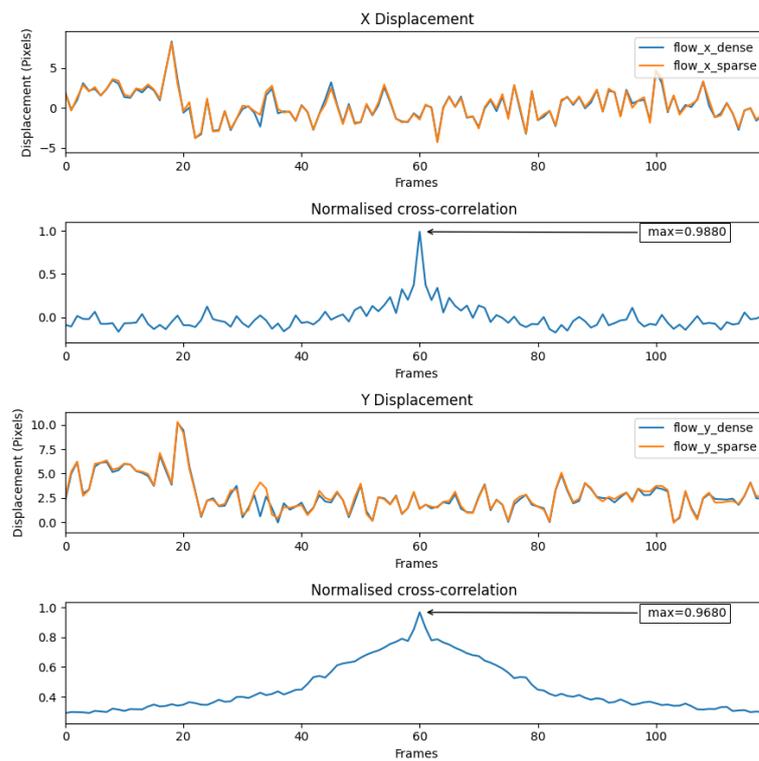


Figure 15. Dense and sparse technique signals on dataset 1.

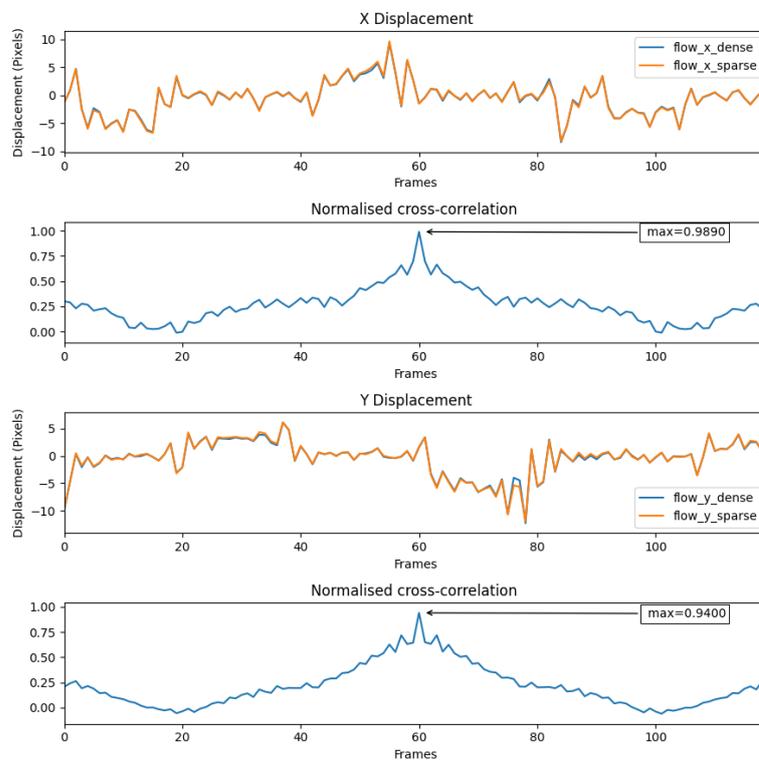
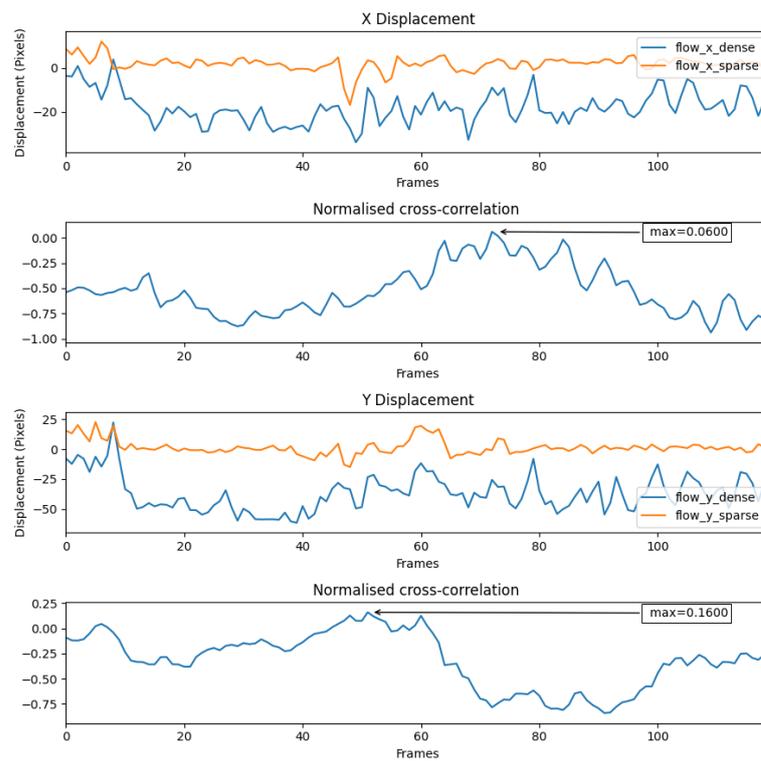


Figure 16. Dense and sparse technique signals on dataset 2 during high thermal contrast.



**Figure 17.** Dense and sparse technique signals on dataset 2 during low thermal contrast.

### 8.2. Effect of Cropping Window and Number of Features on Accuracy and Processing Time

This section outlines how changing the cropping window and number of features effects the accuracy and processing time of the model. These two separated cases are changing the cropping window while keeping the number of features constant and vice versa. Both cases were tested on dataset 1.

#### 8.2.1. Case 1: Constant Cropping Window

Table 3 shows the normalised cross-correlation values in X and Y displacements, processing time difference compared to the dense technique as a percentage. The results show that the accuracy in both X and Y displacement increase exponentially with the number of features in the image, until it reaches four features. After that, accuracy does not change significantly with higher numbers of features. The speed difference decreases steadily for each increase in number of features.

**Table 3.** The effect of “number of features” on accuracy and processing time of the model.

No Features	Cropping	Cross-Correlation X	Cross-Correlation Y	Sparse FPS	Dense FPS	Difference
1	40 × 40	0.381	0.219	29.2	11	+165.45%
2	40 × 40	0.412	0.371	24.3	11	+120.9%
3	40 × 40	0.741	0.673	22.5	11	+104/54%
<b>4</b>	<b>40 × 40</b>	<b>0.988</b>	<b>0.969</b>	<b>19.2</b>	<b>11</b>	<b>+74.54%</b>
5	40 × 40	0.991	0.983	16.3	11	+48.18%
6	40 × 40	0.967	0.931	12.3	11	+11.81%
7	40 × 40	0.961	0.981	8	11	−27.27%

### 8.2.2. Case 2: Constant Number of Features

Table 4 presents the results from changing the cropping window while maintaining four features. The sparse network achieves very low accuracy when the cropping window is less than  $30 \times 30$ . The accuracy increases significantly at  $35 \times 35$  then peaks at  $40 \times 40$ . After that, the accuracy does not change significantly and even decreases at  $55 \times 55$  pixels. The speed, meanwhile, steadily decreases with larger cropping window.

**Table 4.** The effect of cropping window on accuracy and processing time of the model.

No Features	Cropping	Cross-Correlation X	Cross-Correlation Y	Sparse FPS	Dense FPS	Difference
4	$20 \times 20$	0.123	0.07	34.5	11	+213.64%
4	$30 \times 30$	0.126	0.05	26.9	11	+144.54%
4	$35 \times 35$	0.642	0.694	21.7	11	+97.27%
4	<b><math>40 \times 40</math></b>	<b>0.988</b>	<b>0.969</b>	<b>19.2</b>	<b>11</b>	<b>+74.54%</b>
4	$45 \times 45$	0.963	0.953	17.5	11	+55.09%
4	$50 \times 50$	0.983	0.943	14.5	11	+31.81%
4	$55 \times 55$	0.921	0.953	10.2	11	−7.27%

## 9. Discussion

The results show that the our proposed technique can be applied to a dense optical flow neural network for airborne applications with thermal imaging for faster processing time, while maintaining its accuracy. This technique can potentially reduce the computational demand of the network, which translates to a lighter payload and longer operating time for small UAVs.

The proposal relied on the Shi–Tomasi feature-based technique to detect good corners in the image, which does not work well in cold-soaked conditions. This is because these features rely on the difference in temperature between parts of the environment. We experimented with lower threshold values to detect low contrast features but the algorithm also picks up image noises at random sections within the image. The dataset in question was collected during winter, early in the morning of a rainy and foggy day. Hence, the temperature difference was low, which leads to lower thermal contrast. This is consistent with our findings in [61].

We learned that the sparse network can output comparable signals to the dense network in both X and Y displacements with four selected features combined with a cropping window at  $40 \times 40$  pixels. The processing speed increased by 74.54% compared to the dense network. The sparse network does not work well with smaller value of these parameters, while larger values also do not increase the accuracy at the cost of higher computational cost.

## 10. Conclusions

This study showed that only using some good regions in a thermal image is enough to obtain a good 2D optical flow vector for certain airborne applications. Our study showed that this technique can decrease processing time by 74.54% while maintaining accuracy of the output.

However, due to the Shi–Tomasi technique relying on high contrast to detect good corners in thermal images, it does not work well under low contrast conditions. This issue potentially limits its use in some circumstances.

Future studies will look at other, versatile feature extraction techniques to solve the problem of thermal flow during cold-soaked low contrast condition.

**Author Contributions:** T.X.B.N. designed the payload, conducted the fly trials, analysed the data and wrote the draft manuscript. J.C. reviewed the work and contributed with valuable advice. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This research was supported by an Australian Government Research Training Program (RTP) Scholarship. The authors would like to thank Asanka and Sam, the University of South Australia, Philip and Kent from the Defence Science and Technology group for their assistance during the previous dataset collection.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Sample Availability:** Not applicable.

### Abbreviations

The following abbreviations are used in this manuscript:

LWIR	Long Wavelength Infrared
AGC	Automatic Gain Control
FFC	Flat Field Correction
$I^2A$	The Image Interpolation Algorithm
UAV	Unmanned Aerial Vehicle
LK	The Lucas–Kanade Algorithm

### References

1. Nguyen, T.X.B.; Rosser, K.; Chahl, J. A Review of Modern Thermal Imaging Sensor Technology and Applications for Autonomous Aerial Navigation. *J. Imaging* **2021**, *7*, 217. [[CrossRef](#)] [[PubMed](#)]
2. Chahl, J.S.; Srinivasan, M.V.; Zhang, S.W. Landing strategies in honeybees and applications to uninhabited airborne vehicles. *Int. J. Robot. Res.* **2004**, *23*, 101–110. [[CrossRef](#)]
3. Chahl, J.; Mizutani, A. Biomimetic attitude and orientation sensors. *IEEE Sens. J.* **2010**, *12*, 289–297. [[CrossRef](#)]
4. Conroy, J.; Gremillion, G.; Ranganathan, B.; Humbert, J.S. Implementation of wide-field integration of optic flow for autonomous quadrotor navigation. *Auton. Robot.* **2009**, *27*, 189. [[CrossRef](#)]
5. Rosser, K.; Chahl, J. Reducing the complexity of visual navigation: Optical track controller for long-range unmanned aerial vehicles. *J. Field Robot.* **2019**, *36*, 1118–1140. [[CrossRef](#)]
6. Miller, A.; Miller, B.; Popov, A.; Stepanyan, K. Optical Flow as a navigation means for UAV. In Proceedings of the 2018 Australian & New Zealand Control Conference (ANZCC), Melbourne, Australia, 7–8 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 302–307.
7. Gan, S.K.; Sukkarieh, S. Multi-UAV target search using explicit decentralized gradient-based negotiation. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 751–756.
8. Perera, A.G.; Law, Y.W.; Chahl, J. Human pose and path estimation from aerial video using dynamic classifier selection. *Cogn. Comput.* **2018**, *10*, 1019–1041. [[CrossRef](#)]
9. Luo, F.; Jiang, C.; Yu, S.; Wang, J.; Li, Y.; Ren, Y. Stability of cloud-based UAV systems supporting big data acquisition and processing. *IEEE Trans. Cloud Comput.* **2017**, *7*, 866–877. [[CrossRef](#)]
10. Wang, J.; Jiang, C.; Ni, Z.; Guan, S.; Yu, S.; Ren, Y. Reliability of cloud controlled multi-UAV systems for on-demand services. In Proceedings of the GLOBECOM 2017–2017 IEEE Global Communications Conference, Singapore, 4–8 December 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6.
11. Itkin, M.; Kim, M.; Park, Y. Development of cloud-based UAV monitoring and management system. *Sensors* **2016**, *16*, 1913. [[CrossRef](#)]
12. Lee, J.; Wang, J.; Crandall, D.; Šabanović, S.; Fox, G. Real-time, cloud-based object detection for unmanned aerial vehicles. In Proceedings of the 2017 First IEEE International Conference on Robotic Computing (IRC), Taichung, Taiwan, 10–12 April 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 36–43.
13. Horn, B.K.; Schunck, B.G. Determining optical flow. In *Techniques and Applications of Image Understanding*; International Society for Optics and Photonics: Bellingham, WA, USA, 1981; Volume 281, pp. 319–331.

14. Srinivasan, M.V.; Zhang, S.W.; Chahl, J.S.; Stange, G.; Garratt, M. An overview of insect-inspired guidance for application in ground and airborne platforms. *Proc. Inst. Mech. Eng. Part J. Aerosp. Eng.* **2004**, *218*, 375–388.
15. Srinivasan, M.V.; Zhang, S.; Altwein, M.; Tautz, J. Honeybee navigation: Nature and calibration of the “odometer”. *Science* **2000**, *287*, 851–853. [[CrossRef](#)]
16. Srinivasan, M.V. Honey bees as a model for vision, perception, and cognition. *Annu. Rev. Entomol.* **2010**, *55*, 267–284. [[CrossRef](#)] [[PubMed](#)]
17. Garratt, M.A.; Chahl, J.S. Vision-based terrain following for an unmanned rotorcraft. *J. Field Robot.* **2008**, *25*, 284–301. [[CrossRef](#)]
18. Esch, H.; Burns, J. Distance estimation by foraging honeybees. *J. Exp. Biol.* **1996**, *199*, 155–162. [[CrossRef](#)]
19. Chahl, J.; Mizutani, A.; Strens, M.; Wehling, M. *Autonomous Navigation Using Passive Sensors and Small Computers*; Infotech@Aerospace: Arlington, Virginia, 2005; p. 7013.
20. Honegger, D.; Meier, L.; Tanskanen, P.; Pollefeys, M. An open source and open hardware embedded metric optical flow cmos camera for indoor and outdoor applications. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 1736–1741.
21. Barrows, G.L.; Chahl, J.S.; Srinivasan, M.V. Biomimetic visual sensing and flight control. In Proceedings of the Seventeenth International Unmanned Air Vehicle Systems Conference, Bristol, UK, 30 March–1 April 2002; pp. 1–15.
22. Franz, M.O.; Chahl, J.S.; Krapp, H.G. Insect-inspired estimation of egomotion. *Neural Comput.* **2004**, *16*, 2245–2260. [[CrossRef](#)]
23. Garratt, M.; Chahl, J. Visual control of an autonomous helicopter. In Proceedings of the 41st Aerospace Sciences Meeting and Exhibit, Reno, NV, USA, 6–9 January 2003; p. 460.
24. Farneback, G. Two-frame motion estimation based on polynomial expansion. In Proceedings of the Scandinavian Conference on Image Analysis, Halmstad, Sweden, 29 June–2 July 2003; Springer: Berlin/Heidelberg, Germany, 2003; pp. 363–370.
25. Lucas, B.D.; Kanade, T. An iterative image registration technique with an application to stereo vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI), Vancouver, BC, Canada, 24–28 August 1981.
26. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising with block-matching and 3D filtering. In *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*; International Society for Optics and Photonics: Bellingham, WA, USA, 2006; Volume 6064, p. 606414.
27. Srinivasan, M.V. An image-interpolation technique for the computation of optic flow and egomotion. *Biol. Cybern.* **1994**, *71*, 401–415. [[CrossRef](#)]
28. Dosovitskiy, A.; Fischer, P.; Ilg, E.; Hausser, P.; Hazirbas, C.; Golkov, V.; Van Der Smagt, P.; Cremers, D.; Brox, T. FlowNet: Learning optical flow with convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 2758–2766.
29. Ilg, E.; Mayer, N.; Saikia, T.; Keuper, M.; Dosovitskiy, A.; Brox, T. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2462–2470.
30. Ranjan, A.; Black, M.J. Optical flow estimation using a spatial pyramid network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4161–4170.
31. Grover, N.; Agarwal, N.; Kataoka, K. liteflow: Lightweight and distributed flow monitoring platform for sdn. In Proceedings of the 2015 1st IEEE Conference on Network Softwarization (NetSoft), London, UK, 13–17 April 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–9.
32. Sun, D.; Yang, X.; Liu, M.Y.; Kautz, J. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8934–8943.
33. Teed, Z.; Deng, J. Raft: Recurrent all-pairs field transforms for optical flow. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 402–419.
34. Kwasniewska, A.; Szankin, M.; Ruminski, J.; Sarah, A.; Gamba, D. Improving Accuracy of Respiratory Rate Estimation by Restoring High Resolution Features with Transformers and Recursive Convolutional Models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 3857–3867.
35. Zhou, H.; Sun, M.; Ren, X.; Wang, X. Visible-Thermal Image Object Detection via the Combination of Illumination Conditions and Temperature Information. *Remote Sens.* **2021**, *13*, 3656. [[CrossRef](#)]
36. Stypułkowski, K.; Gołda, P.; Lewczuk, K.; Tomaszewska, J. Monitoring system for railway infrastructure elements based on thermal imaging analysis. *Sensors* **2021**, *21*, 3819. [[CrossRef](#)]
37. Gonzalez-Dugo, V.; Zarco-Tejada, P.; Nicolás, E.; Nortes, P.A.; Alarcón, J.; Intrigliolo, D.S.; Fereres, E. Using high resolution UAV thermal imagery to assess the variability in the water status of five fruit tree species within a commercial orchard. *Precis. Agric.* **2013**, *14*, 660–678. [[CrossRef](#)]
38. Meron, M.; Sprintsin, M.; Tsipris, J.; Alchanatis, V.; Cohen, Y. Foliage temperature extraction from thermal imagery for crop water stress determination. *Precis. Agric.* **2013**, *14*, 467–477. [[CrossRef](#)]
39. Alchanatis, V.; Cohen, Y.; Cohen, S.; Moller, M.; Sprintsin, M.; Meron, M.; Tsipris, J.; Saranga, Y.; Sela, E. Evaluation of different approaches for estimating and mapping crop water status in cotton with thermal imaging. *Precis. Agric.* **2010**, *11*, 27–41. [[CrossRef](#)]
40. Weiss, C.; Kirmas, A.; Lemcke, S.; Böshagen, S.; Walter, M.; Eckstein, L.; Leonhardt, S. Head tracking in automotive environments for driver monitoring using a low resolution thermal camera. *Vehicles* **2022**, *4*, 219–233. [[CrossRef](#)]

41. Szankin, M.; Kwasniewska, A.; Ruminski, J. Influence of thermal imagery resolution on accuracy of deep learning based face recognition. In Proceedings of the 2019 12th International Conference on Human System Interaction (HSI), Richmond, VA, USA, 25–27 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
42. Kwasniewska, A.; Rumiński, J.; Rad, P. Deep features class activation map for thermal face detection and tracking. In Proceedings of the 2017 10th international conference on human system interactions (HSI), Ulsan, Korea, 17–19 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 41–47.
43. Khaksari, K.; Nguyen, T.; Hill, B.Y.; Quang, T.; Perrault, J.; Gorti, V.; Malpani, R.; Blick, E.; Cano, T.G.; Shadgan, B.; et al. Review of the efficacy of infrared thermography for screening infectious diseases with applications to COVID-19. *J. Med. Imaging* **2021**, *8*, 010901. [[CrossRef](#)]
44. Khanam, F.T.Z.; Chahl, L.A.; Chahl, J.S.; Al-Naji, A.; Perera, A.G.; Wang, D.; Lee, Y.; Ogunwa, T.T.; Teague, S.; Nguyen, T.X.B.; et al. Noncontact sensing of contagion. *J. Imaging* **2021**, *7*, 28. [[CrossRef](#)]
45. Maddern, W.; Vidas, S. Towards robust night and day place recognition using visible and thermal imaging. In *Proceedings of the RSS 2012 Workshop: Beyond Laser and Vision: Alternative Sensing Techniques for Robotic Perception*; University of Sydney: Camperdown, NSW, Australia, 2012; pp. 1–6.
46. Brunner, C.; Peynot, T.; Vidal-Calleja, T.; Underwood, J. Selective combination of visual and thermal imaging for resilient localization in adverse conditions: Day and night, smoke and fire. *J. Field Robot.* **2013**, *30*, 641–666. [[CrossRef](#)]
47. Mouats, T.; Aouf, N.; Sappa, A.D.; Aguilera, C.; Toledo, R. Multispectral stereo odometry. *IEEE Trans. Intell. Transp. Syst.* **2014**, *16*, 1210–1224. [[CrossRef](#)]
48. Mouats, T.; Aouf, N.; Chermak, L.; Richardson, M.A. Thermal stereo odometry for UAVs. *IEEE Sens. J.* **2015**, *15*, 6335–6347. [[CrossRef](#)]
49. FREE Teledyne FLIR Thermal Dataset for Algorithm Training. FREE Teledyne FLIR Thermal Dataset for Algorithm Training. Available online: <https://www.flir.eu/oem/adas/adas-dataset-form/> (accessed on 30 March 2022).
50. Khattak, S.; Papachristos, C.; Alexis, K. Visual-thermal landmarks and inertial fusion for navigation in degraded visual environments. In Proceedings of the 2019 IEEE Aerospace Conference, Big Sky, MT, USA, 2–9 March 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–9.
51. Khattak, S.; Papachristos, C.; Alexis, K. Keyframe-based direct thermal–inertial odometry. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 3563–3569.
52. Khattak, S.; Papachristos, C.; Alexis, K. Marker based thermal-inertial localization for aerial robots in obscurant filled environments. In *International Symposium on Visual Computing*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 565–575.
53. Khattak, S.; Papachristos, C.; Alexis, K. Keyframe-based thermal–inertial odometry. *J. Field Robot.* **2020**, *37*, 552–579. [[CrossRef](#)]
54. Khattak, S.; Nguyen, H.; Mascarich, F.; Dang, T.; Alexis, K. Complementary multi-modal sensor fusion for resilient robot pose estimation in subterranean environments. In Proceedings of the 2020 International Conference on Unmanned Aircraft Systems (ICUAS), Athens, Greece, 1–4 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1024–1029.
55. Weinmann, M.; Leitloff, J.; Hoegner, L.; Jutzi, B.; Stilla, U.; Hinz, S. THERMAL 3D MAPPING FOR OBJECT DETECTION IN DYNAMIC SCENES. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *2*, 53–60. [[CrossRef](#)]
56. Brook, A.; Vandewal, M.; Ben-Dor, E. Fusion of optical and thermal imagery and LiDAR data for application to 3-D urban environment and structure monitoring. *Remote Sens.—Adv. Tech. Platforms* **2012**, *2012*, 29–50.
57. O’Donohue, D.; Mills, S.; Kingham, S.; Bartie, P.; Park, D. Combined thermal-LIDAR imagery for urban mapping. In Proceedings of the 2008 23rd International Conference Image and Vision Computing New Zealand, Christchurch, New Zealand, 26–28 November 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 1–6.
58. Lu, Y.; Lu, G. An alternative of lidar in nighttime: Unsupervised depth estimation based on single thermal image. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2021; pp. 3833–3843.
59. Shin, Y.S.; Kim, A. Sparse depth enhanced direct thermal-infrared SLAM beyond the visible spectrum. *IEEE Robot. Autom. Lett.* **2019**, *4*, 2918–2925. [[CrossRef](#)]
60. Rosser, K.; Nguyen, T.X.B.; Moss, P.; Chahl, J. Low complexity visual UAV track navigation using long-wavelength infrared. *J. Field Robot.* **2021**, *38*, 882–897. [[CrossRef](#)]
61. Nguyen, T.X.B.; Rosser, K.; Perera, A.; Moss, P.; Teague, S.; Chahl, J. Characteristics of optical flow from aerial thermal imaging, “thermal flow”. *J. Field Robot.* **2022**, *39*, 580–599. [[CrossRef](#)]
62. Nguyen, T.X.B.; Rosser, K.; Chahl, J. A Comparison of Dense and Sparse Optical Flow Techniques for Low-Resolution Aerial Thermal Imagery. *J. Imaging* **2022**, *8*, 116. [[CrossRef](#)]
63. Sheen, D.M.; McMakin, D.L.; Hall, T.E. Cylindrical millimeter-wave imaging technique and applications. In *Passive Millimeter-Wave Imaging Technology IX*; SPIE: Bellingham, WA, USA, 2006; Volume 6211, pp. 58–67.
64. Sheen, D.M.; McMakin, D.L.; Collins, H.D. Circular scanned millimeter-wave imaging system for weapon detection. In *Law Enforcement Technologies: Identification Technologies and Traffic Safety*; SPIE: Bellingham, WA, USA, 1995; Volume 2511, pp. 122–130.
65. Zhang, R.; Cao, S. 3D imaging millimeter wave circular synthetic aperture radar. *Sensors* **2017**, *17*, 1419. [[CrossRef](#)]
66. Smith, J.W.; Yanik, M.E.; Torlak, M. Near-field MIMO-ISAR millimeter-wave imaging. In Proceedings of the 2020 IEEE Radar Conference (RadarConf20), Florence, Italy, 21–25 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.

67. Li, R.; Liu, J.; Zhang, L.; Hang, Y. LIDAR/MEMS IMU integrated navigation (SLAM) method for a small UAV in indoor environments. In Proceedings of the 2014 DGON Inertial Sensors and Systems (ISS), Karlsruhe, Germany, 16–17 September 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 1–15.
68. Chen, D.; Gao, G.X. Probabilistic graphical fusion of LiDAR, GPS, and 3D building maps for urban UAV navigation. *Navigation* **2019**, *66*, 151–168. [[CrossRef](#)]
69. Kumar, G.A.; Patil, A.K.; Patil, R.; Park, S.S.; Chai, Y.H. A LiDAR and IMU integrated indoor navigation system for UAVs and its application in real-time pipeline classification. *Sensors* **2017**, *17*, 1268. [[CrossRef](#)]
70. Chiang, K.W.; Tsai, G.J.; Li, Y.H.; El-Sheimy, N. Development of LiDAR-based UAV system for environment reconstruction. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1790–1794. [[CrossRef](#)]
71. Shah, S.T.H.; Xuezhi, X. Traditional and modern strategies for optical flow: An investigation. *SN Appl. Sci.* **2021**, *3*, 1–14. [[CrossRef](#)]
72. Shi, J.; Tomasi, C. Good features to track. In Proceedings of the 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 21–23 June 1994; IEEE: Piscataway, NJ, USA, 1994; pp. 593–600.
73. FLIR Corp. *FLIR Lepton Engineering Data Sheet*. FLIR Corp., Wilsonville, OR, USA, 2018.
74. Bradski, G. The openCV library. *Dr. Dobb's Journal: Softw. Tools Prof. Program.* **2000**, *25*, 120–123.
75. Papachristos, C.; Mascarich, F.; Alexis, K. Thermal-inertial localization for autonomous navigation of aerial robots through obscuring. In Proceedings of the 2018 International Conference on Unmanned Aircraft Systems (ICUAS), Dallas, TX, USA, 12–15 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 394–399.
76. Lu, Y.; Xue, Z.; Xia, G.S.; Zhang, L. A survey on vision-based UAV navigation. *Geo-Spat. Inf. Sci.* **2018**, *21*, 21–32. [[CrossRef](#)]
77. Engel, J.; Koltun, V.; Cremers, D. Direct sparse odometry. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 611–625. [[CrossRef](#)]
78. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 898–916. [[CrossRef](#)]
79. Dollár, P.; Zitnick, C.L. Fast edge detection using structured forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 1558–1570. [[CrossRef](#)]
80. Endres, F.; Hess, J.; Sturm, J.; Cremers, D.; Burgard, W. 3-D mapping with an RGB-D camera. *IEEE Trans. Robot.* **2013**, *30*, 177–187. [[CrossRef](#)]
81. Poma, X.S.; Riba, E.; Sappa, A. Dense extreme inception network: Towards a robust cnn model for edge detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2022; pp. 1923–1932.
82. Harris, C.; Stephens, M. A combined corner and edge detector. In Proceedings of the Alvey Vision Conference. Citeseer, Manchester, UK, 31 August–2 September 1988; Volume 15, pp. 10–5244.
83. Bureau of Meteorology. Woomera Weather. 2020. Available online: <http://www.bom.gov.au/places/sa/woomera/> (accessed on 22 August 2022)
84. Butler, D.J.; Wulff, J.; Stanley, G.B.; Black, M.J. A naturalistic open source movie for optical flow evaluation. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 611–625.