Vision-Based 3-D Localization of UAV Using Deep Image Matching

Mansoor Khurshid, Muhammad Shahzad, *Senior Member, IEEE*, Hasan Ali Khattak, *Senior Member, IEEE*, Muhammad Imran Malik, and Muhammad Moazam Fraz, *Senior Member, IEEE*

***Abstract*—Unmanned aerial vehicles (UAVs) have revolutionized various industries by providing efficient and automated flight capabilities. However, reliance on GPS and traditional navigation systems poses challenges in scenarios where signal interference or failures occur. In this research, we present a novel computer-vision-based method to enhance UAV navigation, enabling accurate height and location estimation. Our approach utilizes a sophisticated network that leverages a pair of images to estimate UAV height. The pyramid stereo-matching network is employed to extract robust image features and generate a disparity map. Subsequently, a custom network processes and convolves these data, employing diverse computer vision techniques to achieve precise height estimation. To evaluate the effectiveness of our proposed method, we collected a comprehensive dataset by conducting flights with a Phantom 4 Pro drone over the NUST Main campus, H-12 Islamabad. The dataset encompasses images captured at 10 different heights, spanning from 100 to 280 m, with flights evenly spaced 20 m apart. In rigorous evaluations, our approach demonstrates promising results compared to existing methods. By liberating UAVs from reliance on GPS, this vision-based 3-D localization technique holds immense potential to ensure successful flights even in challenging environments.**

***Index Terms*—3-D localization of UAV, computer vision, deep image matching, feature detection, UAV height estimation, unmanned aerial vehicles (UAVs).**

I. Introduction

IMAGE matching is a fundamental task in machine vision with significant applications in depth estimation, motion detection, tracking, 3-D object reconstruction, and height estimation. It involves comparing pairs of images captured by a camera with a slight shift in camera position. Feature matching is employed to measure the similarity between image pairs and calculate the disparity, which is subsequently used to determine the depth maps.

One specific application of image matching is in the field of vision-based autonomous navigation, where unmanned aerial vehicles (UAVs) [1] utilize visual cues for self-localization. In scenarios where GPS and altimeter failures occur due to factors such as jamming, technical issues, or unforeseen circumstances, reliable localization becomes crucial for UAVs. As UAV localization heavily relies on onboard GPS and altimeter systems, malfunctions can render the UAV ineffective. Thus, alternate approaches are essential aids in autonomous flight in such scenarios.

UAVs are typically equipped with high-resolution cameras that can be leveraged to estimate the UAV height from the ground. This involves an image-matching procedure, where features are extracted from two images to establish correspondences. The resulting correspondences are then utilized to calculate the disparity and depth maps. However, this pipeline is complex due to factors such as viewpoint changes, variations in illumination, and dynamic prominent features, making image matching and depth estimation highly challenging.

Various models based on mathematical transforms (e.g., Fourier and wavelet) [2], [3] and rotation and scale-invariant feature descriptors (e.g., SIFT, SURF, and histogram of dominant gradients) [4], [5], [6], [7] have been proposed. These models work by extracting key points from two images to obtain feature maps, which then enable feature matching by establishing correspondence among those images. Approximate nearest-neighbor search algorithms, coupled with postprocessing procedures (hierarchical k-means tree, RANSAC) [8], [9], have been utilized to enhance matching results. Nevertheless, nearest-neighbor search encounters limitations as they lack local texture. To address this, researchers have recently proposed deep trainable models [primarily relying on end-to-end trainable deep convolutional neural network (CNN) architectures] [10], [11], [12], [13] that facilitate robust feature extraction and establish correspondence by incorporating larger context. Moreover, they are effective in calculating disparity efficiently, which is useful in applications such as 3-D object modeling, 3-D reconstruction, and depth estimation [14], [15], [16], [17]. In the context of depth estimation, these maps are frequently used to find the distance of objects/pedestrians from the camera. They were also used to find the height of UAVs from the ground using ground cameras and height calculation during indoor flights [1], [5], [18] but not used for aerial-to-ground distance calculation. This is the area

Manuscript received 15 August 2023; revised 23 November 2023, 9 February 2024, and 25 April 2024; accepted 19 June 2024. Date of publication 2 July 2024; date of current version 12 July 2024. *(Corresponding author: Muhammad Shahzad.)*

Mansoor Khurshid, Muhammad Imran Malik, and Muhammad Moazam Fraz are with the School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan.

Muhammad Shahzad is with the School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan, and also with the Chair of Data Science in Earth Observation, Technical University of Munich (TUM), 80333 Munich, Germany (e-mail: muhammad.shahzad@tum.de).

Hasan Ali Khattak is with the School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan, and also with the Faculty of Computing and Information Technology, Sohar University, Sohar P.C 311, Oman.

Digital Object Identifier 10.1109/JSTARS.2024.3422310

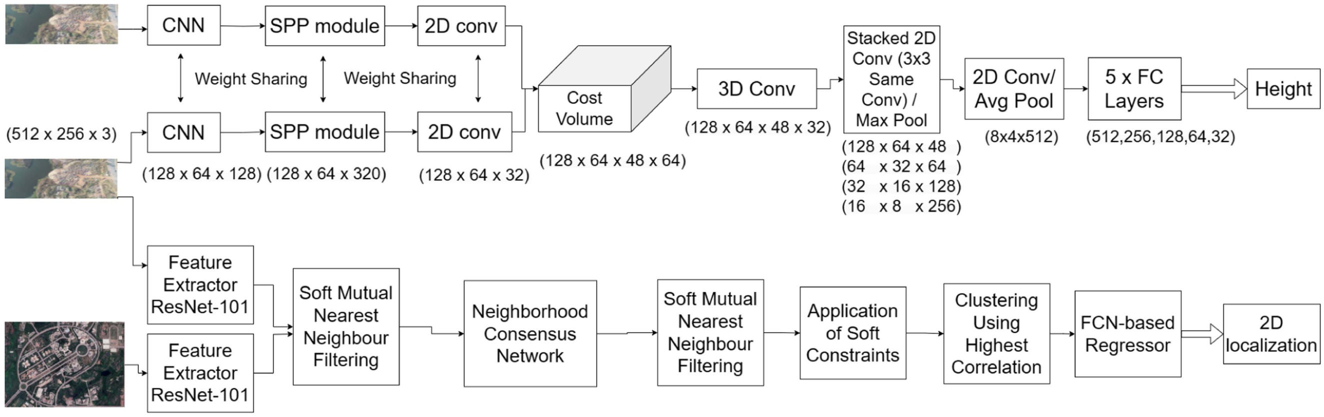


Fig. 1. Main network diagram. The upper part of the diagram represents the network that extracts the feature using weight sharing and then calculate the cost volume. This cost volume is processed by 3-D convolutions as well as 2-D convolutions. Finally, fully connected layers are used for final prediction of height. The lower part of the network diagram shows the template mating of the image taken from the drone with an orthomosaic for 2-D localization. Both the networks were combined for 3-D localization.

we focus, i.e., to compute the aerial to ground distance of the UAV using a pair of images taken by the onboard camera.

Motivated by the concepts of deep CNNs and end-to-end learning, this article presents a complete pipeline for end-to-end learning and 3-D localization of UAVs, as illustrated in Fig. 1. The primary focus is to utilize this technique for estimating the UAV height from the ground and integrating it with 2-D data to achieve 3-D localization. The proposed approach involves utilizing a pair of images (UAV camera) to extract features and establish correspondence, enabling disparity calculation and subsequent determination of the UAV depth/height from the ground level. Within the proposed architecture, the key contributions include the following.

1) Designing an end-to-end trainable network that estimates the UAV height using pairs of images for 3-D localization. The pyramid stereo-matching network [17] is employed as the base network for feature extraction from image pairs and subsequent disparity map calculation, which is adopted and enhanced to utilize for the estimation of the UAV height.
2) Demonstrating the proposed model on real-time UAV imagery by acquiring training and testing data through multiple drone flights.
3) Collecting a unique custom dataset by flying a Phantom 4 Pro drone within the premises of the NUST Main Campus H-12 Islamabad, Pakistan. This dataset is made publicly available for the benefit of the research community.

II. Related Work

Numerous researchers have made substantial contributions to the fields of image matching and depth estimation, employing a wide array of techniques, including invariant feature descriptors and CNN-based approaches. Our research specifically focuses on two significant research areas 1) image matching to calculate disparity and 2) depth estimation.

Image matching forms the cornerstone of computer vision, playing a pivotal role in essential tasks, such as image correspondence, optical flows/disparity, and person reidentification, among others. Traditional image-matching techniques employ Siamese-based networks to find correspondences. However, recent advancements have seen the emergence of more sophisticated methods. Mughal et al. [19] explore the use of deep convolutional neural networks for real-time 2-D localization of UAVs, leveraging the UAV camera and locally stored orthomosaic images. For disparity estimation, Mayer et al. [20] contribute to the field by providing three realistic, diverse, and large-scale synthetic stereo video datasets, enabling effective convolutional network training for real-time disparity estimation. Zhang et al. [21] focus on efficient disparity estimation, incorporating squeezed cost volumes and attention-based spatial residual modules. Chang and Chen [17] utilize a pyramid stereo-matching network to establish image correspondences for disparity calculation. Their network employs shared weights to extract features from two images, and a 4-D cost volume is employed to derive the disparity.

For depth estimation from monoscopic imagery, Haseeb et al. [22] provide a machine learning configuration that gives the obstacle detection system a way to calculate the distance between the monocular camera and the item being seen by the camera in order to enhance self-supervised monocular distance estimation on fisheye and pinhole camera pictures. Kumar et al. [23] offer a unique multitask learning technique to improve self-supervised distance estimation on monocular fisheye camera images. Ciganek et al. [24] give a general overview of measuring techniques and how to determine the distance between an object and a camera. They initially identified the corners in an image and the distance between them and then used mathematical equations to find the distance of camera from the object. Pavlovic et al. [25] presented the application of two different techniques (sensor based and vision based) to find the distance between a robot and an object. They used radar/ LiDAR for the sensor-based technique and used both single and stereo pairs for the vision-based technique. Yoon et al. [26] used both single image as well as pair of images to find the depth by using a depth fusion network.

For depth estimation from a pair of images, there are many methods available for cost volume optimization and matching cost computation, which have been proposed in the literature. Wang et al. [14] introduce the pseudo-LiDAR network, which estimates depth for 3-D object generation using point cloud data. Building on this, pseudo-LiDAR++ network and image calibration data are used for depth estimation [15]. Garg et al. [16] present a novel neural network architecture utilizing a loss function derived from the Wasserstein distance to output arbitrary depth values. Maximov et al. [27] worked on the generalization of depth estimation outside the training set so that accurate results can be achieved for that they have worked on direct supervision of domain invariant defocus and used a convolution neural network to learn from a pair of images with a different point of focus.

For vision-based height estimation of the UAV, Shabayek et al. [28] carried out the comparison of various techniques including horizon-based methods, optical flow, stereoscopic-based techniques and passing by vanishing points. Gökçe et al. [5] worked on distance estimation as well as detection of microunmanned aerial vehicles in different scenarios including intruder UAVs in a protected environment and controlling UAVs for environmental surveillance and monitoring. They tested local binary patterns, histogram of gradients (HOGs), and Haar-like features using the cascade of boosted classifiers. Pan et al. [18] have used both monocular image and stereo images to estimate the approach angle and height of the UAV for autonomously landing. To calculate the approach angle, they extracted vanishing lines by using the Hough transform and RANSAC algorithm, and to calculate the height, feature-based matching is adopted by extracting Harris corner from stereo images and then using approach angle 3-D reconstruction. The Kalman filter model is built by analyzing motion characteristics of the UAV for accurate height estimation. Mondragon et al. [1] worked on estimating UAV height, the pose and motion estimation using visual aid to control the aircraft from the ground. Their objective was to demonstrate that computer vision can be successfully used in control loops and for which fast image processing algorithms were required to close the gap between real-time controls and visual control, i.e., a processing rate of 15 frames per second. They have shown that to enable UAV navigation based on visual input, traditional image-collecting techniques can be combined with ad-hoc image processing and fuzzy controllers to achieve good results. Dhahbane et al. [29] conducted a survey in the domain of engineering (guidance, navigation, and control) to determine the orientation of an aircraft in space with respect to another object. They have compared different techniques including computer vision to determine the attitude (roll, pitch, and yaw) of an aircraft. Yang et al. [30] worked to determine the position and attitude of an aircraft with respect to the horizon by integrating a visual odometer (computer vision) and GPS, which was used to take results so as to minimize the trajectory estimation error. Wan et al. [31] provide a study on UAV localization technique for its autonomous navigation based on matching between onboard UAV picture sequences and a preinstalled reference satellite image. The images compared with each other are not taken under the same illumination conditions; hence, they used illumination invariant tech of Phase Correlation through mathematical deviation and able to estimate the current and next position of the UAV and also apply self-coarse correction in case the UAV is not following the planned path. Liu et al. [32] worked to find the height of the UAV and focus on the issues in optical flow due to rotational and translational movement by using gated recurrent unit neural network. Yol et al. [33] 3-D localize the UAV using the deep image-matching technique.

It is worth noting that while many of the earlier algorithms primarily concentrate on calculating UAV height from ground sensors or during indoor flights, there has been limited exploration specifically dedicated to vision-based height estimation of UAVs using the onboard camera. This particular aspect forms a crucial aspect of our research, where we aim to extend the existing methodologies to enable precise distance estimation for UAV navigation. In the subsequent sections, we will present our approach, methodology, and experimental findings to address this gap in the existing literature.

III. Methodology

Depth calculation constitutes a comprehensive pipeline that encompasses data acquisition from a pair of images. Subsequently, these images are processed through the network to calculate feature maps, which are further integrated into a 3-D cost volume. This cost volume facilitates the computation of disparity, ultimately leading to the determination of depth from the camera's perspective. The implementation process is divided into several distinct steps, each of which plays a crucial role in comprehending the overall development of the project. In the following sections, we present further details of the proposed architecture.

A. Overview

In this research, we adopt the PSMNet as our base network and tailor it to effectively establish correspondences and calculate disparities between image pairs while simultaneously estimating the height from the ground. Our network is referred to as the Pyramid Stereo-Matching and Height Estimation Network (PSMHENet). Initially, we compare PSMHENet performance with that of the original PSMNet, gradually refining our network's design step by step to achieve improved outcomes. We conduct a comparative analysis between our results and those obtained using PSMNet. Our main contributions encompass the modification of PSMNet and the design of a network specifically dedicated to height estimation, which is appended at the end of PSMNet, effectively integrating the two functionalities.

To perform a comprehensive comparison with our proposed network (PSMHENet), we also employed two state-of-the-art algorithms. First, the Siamese-based image-matching technique [34] (with minor modifications) was employed to establish correspondences, calculate the disparity between the matched features, and subsequently estimate height. The Siamese-based matching network [35] employs a patch-based image-matching technique, which lacks the consideration of the full context of the image. However, the full context is crucial for accurately determining the true height of the drone. Whereas our technique

(PSMHENet) not only utilizes the entire image for comparison but also incorporates the full context of the image through two essential techniques 1) spatial pyramid pooling (SPP) and 2) 3-D convolutions, in combination with an hourglass network. This comprehensive approach enables the PSMNet to attain the global context of the image, resulting in more accurate and reliable results for height estimation.

Second, we compared our approach with a vision-based UAV localization technique proposed by Yol et al. [33]. Their method relies solely on visual information and employs a multipass localization algorithm. However, many existing approaches are susceptible to scene variations caused by changes in season or environment due to their utilization of Sum of Squared Differences (SSD). To address this issue, the authors opted for mutual information (MI) as a more robust alternative, capable of accommodating both local and global scene variations.

B. Network Diagram

The network diagram presented in Fig. 1 consists of two distinct parts. The upper part of the network showcases the input of stereo images, which are processed with shared weights to extract features from both images. These features are then amalgamated to form a 4-D cost volume, encompassing width, height of feature maps, and disparity at each location. Subsequently, this 4-D volume is convolved through various layers, reducing the feature map size while increasing the number of feature maps. Following this, the data are passed through fully connected layers to estimate the height.

On the other hand, the lower part of the network employs one image from the stereo pair, along with an orthomosaic, to conduct 2-D localization of the image within that orthomosaic. It highlights the proposed localization pipeline that fundamentally relies on feature point learning and a neighborhood consensus strategy to refine the matches between the template image patch and the prestored orthomosaic. This is accomplished by leveraging the correlation information between the convolutional features of the two images and then applying probabilistic constraints to interdetermine the point-to-point correspondences. Here, the convolutional feature maps embodying both the local and global information are extracted using a deep feature extractor and are later used to generate the correlation matrix that incorporates the feature matches for every extracted feature point. Subsequently, the probabilistic constraints followed by soft-argmax layer are applied to these established correspondences to link each feature point in the source image with the feature points in the orthomosaic. For further details, refer to our publication [19].

C. Feature Extraction

The pair of images (an example pair shown in Fig. 2) is fed into a custom feature extractor with shared weights. This feature extractor processes the images sequentially and performs convolutions on them. Subsequently, SPP is applied to the feature map, allowing the extraction of features with a broader context. By incorporating SPP with different scales, the extracted features can fully benefit from the global context of the image. Specifically, 32 features are extracted from each point of the feature map for comparison with the feature map of the other image. This approach facilitates enhanced feature matching and disparity calculation between the image pair, leading to more accurate height estimation for the UAV.

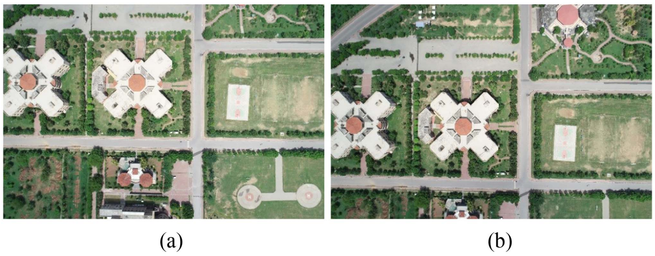


Fig. 2. Sample pair of images taken from the dataset. The pair of images is fed into the network for matching and prediction of the UAV height in meters. (a) Left image. (b) Right image.

D. Disparity Correlation Matrix

After generating the feature maps from both images, they are combined in a manner that preserves the dimensions of the feature map. This is accomplished by concatenating the 32 features from each image, resulting in a total of 64 features at each location of the feature map. The subsequent crucial step involves forming a 4-D tensor, with dimensions one-fourth that of the image, encompassing 64 features from both images and 48 disparity values. Initially, these disparity values are loaded with default values and then 3-D convolutions are employed for further processing. This process ensures the incorporation of essential information from both images and disparity values, allowing for accurate disparity estimation, which is instrumental in height calculation for the UAV.

E. Tridimensional Convolutions and Hierarchical Hourglass Framework

During this step, 3-D convolutions are employed, followed by batch normalization and ReLU activation. This process is iterated multiple times to convolve the network. Consequently, the features at each point in the feature map are reduced from 64 to 32. Subsequently, an hourglass architecture (encoder–decoder) is introduced to learn the maximum context of the feature map. The hourglass network operates with a top-down and bottom-up approach, where the feature map, already reduced to 1/4th of the image size, is further downsized to 1/8th and then 1/16th of the image size, aiming to attain a comprehensive context of the feature map. In the bottom-up process, the feature map is upsampled using transpose convolutions to achieve 1/8th and then 1/4th of the image size. This approach facilitates the acquisition of contextual information for improved height estimation.

Following the hourglass process, 3-D convolutions are performed once more, further reducing the features from 32 to 1. Consequently, the feature map is converted from 3-D to 2-D, with dimensions of 48 × 1/4 H × 1/4 W. This step enhances the feature map's representational power and prepares it for the final stage of height estimation.

F. Planar Convolutions and Dense Connections for Height Estimation

At this stage of the network, the size of the feature map is 48 × 1/4 H × 1/4 W. The objective here is to further reduce the feature map while increasing the number of planes (filters) to gather maximum information for the subsequent fully connected layers at the end. Initially, 2-D convolutions followed by MaxPool are employed to reduce the feature map to 1/8th of its original size while increasing the planes to 64.

This process of 2-D convolutions is stacked iteratively to gradually reduce the feature map to a single point and simultaneously increase the number of planes to 512. Eventually, AvgPool is applied to reduce the feature map to a single point while maintaining 512 planes, resulting in dimensions of 512 × 1/64 H × 1/64 W $\geqslant$ 512 × 1 × 1.

The next stage entails fully connected layers, where the output of 512 is connected to 256, then 128, followed by 64, and eventually to 32, culminating in a single-output value. This final output represents the estimated height of the UAV from the ground, providing valuable information for 3-D localization and navigation.

G. Integration of Planar Localization and Height Estimation for 3-D Localization

In order to achieve 3-D localization of the UAV, we have integrated a dedicated 2-D localization network with the height estimation network, as depicted in Fig. 1. To ensure optimal performance, a pretrained network on the same geographical area was utilized. Notably, the input data for these networks differ: The height estimation network employs a pair of images with identical dimensions while the 2-D localization network requires an image along with an orthomosaic. Within the 2-D localization network, only one of the images from the image pair is utilized.

Upon combining both networks, we are able to accurately predict the height of the UAV while simultaneously achieving precise localization of the image within the orthomosaic. This seamless integration facilitates comprehensive 3-D localization of the UAV, thereby aiding in its navigational capabilities.

H. Network Architecture

The overall architecture can be divided into three primary components. In the first part, we utilize a custom feature extractor to generate feature maps. Subsequently, in the second part, these feature maps are amalgamated to form a 4-D cost volume, which represents the disparity map. Finally, the third part involves the processing of this 4-D cost volume through multiple convolution layers, culminating in the accurate estimation of the UAV height. This modular approach allows for efficient and effective height determination, enhancing the overall performance of the system.

I. Implementation Details

The implementation details encompass a comprehensive set of hyperparameters utilized during the training process. Initially, the model was trained for 150 epochs. During training, the learning rate was initially set to 0.00001, which was then fine-tuned to 0.000001 for model optimization. Beta values were assigned as 0.9 and 0.999 to optimize the optimization process for AdamW, and the weight decay was set to 0.000001. Adam optimizer was employed initially, but subsequently, the AdamW optimizer was explored. AdamW, being a stochastic gradient descent method, combines adaptive estimation of first and second-order moments with a weight decay method, effectively countering overfitting and yielding favorable outcomes.

To ensure robust results, multiple training strategies were adopted. While the L2 loss was initially utilized, the presence of outliers prompted a transition to L1 loss (absolute error), which proved to be more effective in generating accurate results. The model underwent extensive training with numerous epochs, consistently updating to enhance its performance. In the final iteration, the model was trained for 190 epochs, culminating in the desired outcomes for height estimation and localization of the UAV.

IV. Experimental Evaluation

In this research, comprehensive experiments were conducted to assess the performance of the proposed approach, aiming to simulate real-world conditions encountered in UAV imagery within our test dataset. Extensive investigations into algorithmic details and hyperparameters were also undertaken to identify the optimal configuration of the model concerning efficiency and accuracy. The results of this thorough analysis, referred to as the ablation study, are presented in this section.

A. Data Curation

Data collection emerged as one of the most challenging aspects of this research undertaking. The project necessitated the acquisition of a dataset comprising downward-looking drone camera images exhibiting a minimum overlapping of 75%–80% between consecutive images. Unfortunately, the existing stereo image datasets available in the market mainly consisted of street views or images of objects in confined spaces with limited variations in depth. Consequently, the creation of a custom dataset became imperative to cater to the specific requirements and objectives of this project.

1) Data Acquisition: The data gathering process necessitated the utilization of a professional drone equipped with the capability to fly at various heights, including higher altitudes. In addition, a suitable location was chosen for the drone flights, leading to the selection of the NUST Main Campus, H-12 Islamabad. A total of 10 flights were meticulously planned to cover a diverse range of 10 different heights. The drone employed for data collection was the Phantom 4 Pro, which conducted multiple flights over the NUST Main Campus to collect the necessary data. The dataset encompassed data gathered at 10

TABLE I
DETAILS OF DATASET COVERING THE TIME, HEIGHT OF UAV, AREA COVERED, IMAGE RESOLUTION, CAMERA ANGLE, ETC

Height (m)	**Date / Time**	**Area Covered (m)**	**No of Images Taken**	**Flight Time / Path**	**Camera Angle/ Resolution / Overlapping**
100	14 Aug 21 / 12:05 P.M.	1000 × 487	260	16 min / 5769 M	90°/ 5472 × 3648/ 80 %– 20 %
120	13 Aug 21 / 5:10 P.M.	1040 × 487	198	14 min / 5332 M	90° / 5472 × 3648/ 80 %–20 %
140	15 Aug 21 / 5:26 P.M.	1000 × 738	280	16 min / 8277 M	90° / 5472 × 3648 / 80 %–45 %
160	15 Aug 21 / 3:39 P.M.	1000 × 738	240	16 min / 8277 M	90° / 5472 × 3648 / 80 %– 50 %
180	15 Aug 21 / 2:05 P.M.	1000 × 738	198	15 min / 7548 M	90° / 5472 × 3648 / 80 %– 50 %
200	15 Aug 21 / 12:27 P.M.	1000 × 733	160	15 min / 6819 M	90° / 5472 × 3648 / 80 %– 50 %
220	14 Aug 21 / 3:23 P.M.	1000 × 738	126	14 min / 6091 M	90° / 5472 × 3648 / 80 %– 45 %
240	14 Aug 21 / 1:51 P.M.	1000 × 746	102	54 min / 5362 M	90° / 5472 × 3648/ 80 %– 40 %
260	14 Aug 21 / 5:03 P.M.	1000 × 757	112	14 min / 6092 M	90°/ 5472 × 3648/ 80 %– 50 %
280	14 Aug 21 / 6:34 P.M.	1000 × 736	105	14 min / 6091 M	90° / 5472 × 3648 / 80 %– 55 %
Total	3 Days	-	1781	200 min	-

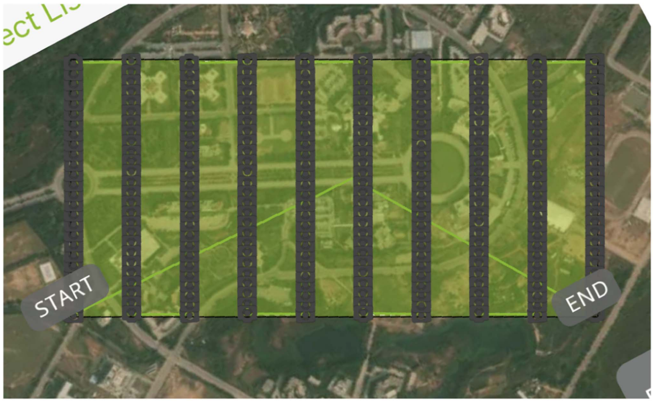


Fig. 3. Pix4dcapture mobile app [36] is used to capture the dataset. This app has the option to enter the complete flight path along with start and end points as well as the percentage of overlapping required among consecutive images. The figure shows the start, end, and the complete path of drone flight for capturing the dataset. The black lines (drone flight path) indicate the position of images taken during flight. The image pair at the end of each path and the start of the new path was ignored as they have little overlapping.

distinct heights, commencing from 100 m and ascending in increments of 20 m up to 280 m.

To optimize the drone flights, the area of interest was defined using the Pix4dcapture mobile app [36], and its Grid 2D option was utilized for setting the flight paths. This app provides different options in which if you select the area on map, start point, end point, height, and front/side overlapping, it will automatically send the drone to capture images as per the area selected, as shown in Fig. 3. The consecutive images (with overlapping) were then used as pairs. About, 80% front overlapping (consecutive images) and 20% side overlapping were used to capture maximum images from a specific area (front overlapping is more relevant in consecutive images). The consecutive images as pairs with 80% overlapping were used. As far as "base" distance is concerned, it was automatically adjusted as per the height of the UAV so that overlapping remains the same (80%), and the base distance is not used as a parameter during training. For further insights, refer to Table I for technical specifics of the dataset.

TABLE II
DETAILS OF DATASET AFTER AUGMENTATION PROCESS

Height (m)	Date / Time	No of images
100	14 Aug 21 / 12:05	1040
120	13 Aug 21 / 5:10 P.M.	792
140	15 Aug 21 / 5:26 P.M.	1120
160	15 Aug 21 / 3:39 P.M.	960
180	15 Aug 21 / 2:05 P.M.	792
200	15 Aug 21 / 12:27 P.M.	640
220	14 Aug 21 / 3:23 P.M.	504
240	14 Aug 21 / 1:51 P.M.	408
260	14 Aug 21 / 5:03 P.M.	448
280	14 Aug 21 / 6:34 P.M.	420
Total	3 Days	7142

2) Data Preparation: Upon successful completion of all flights, a total of 1781 images were collected. However, this dataset was deemed insufficient for the training of deep CNNs. Consequently, data augmentation techniques were employed to augment the dataset. During the data augmentation process, images were subjected to flipping and rotation. Careful consideration was given to ensure that the dimensions of the images remained unchanged, thus rotation was limited to 180° while horizontal and vertical flipping were also implemented. Furthermore, the direction of camera movement (shifting of camera) was maintained in a consistent manner throughout the augmentation process. The augmented dataset details are presented in Table II for reference.

3) Dataset Split: In order to achieve optimal results during the training of the network, a conventional 80%–20% split on mixed data was deemed unsuitable due to the dataset containing data of various heights. Mixing data of distinct heights in the training set could lead to suboptimal performance. Therefore, a novel approach was adopted to split the dataset based on the specific height values. By organizing the split in this manner, data of distinct heights could be segregated, ensuring a more effective training process.

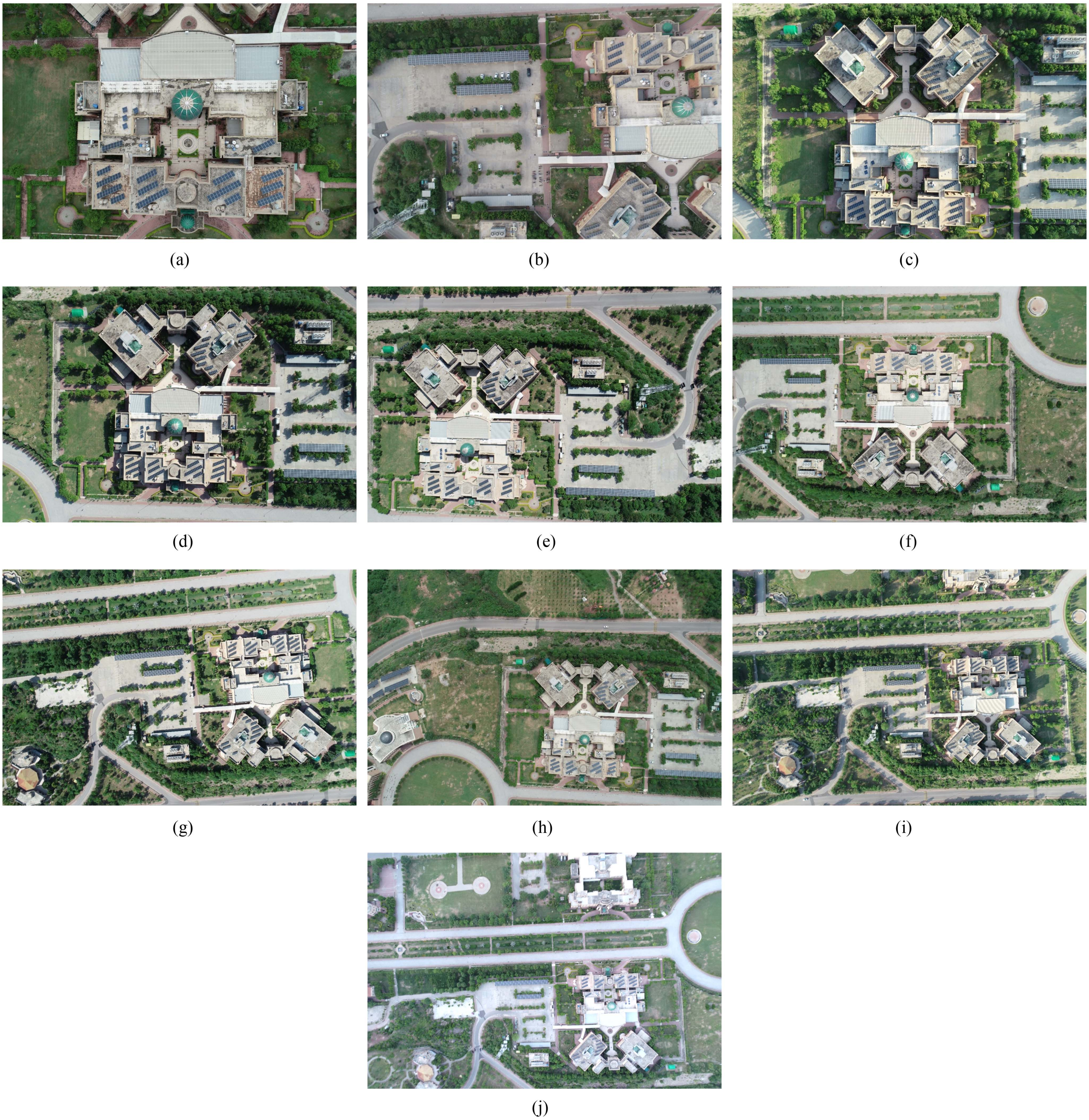


Fig. 4. Sample images from the dataset. Each image is taken from the same location with different heights. The heights of images start from 100 to 280 m with a gap of 20 m. (a) Height 100 m. (b) Height 120 m. (c) Height 140 m. (d) Height 160 m. (e) Height 180 m. (f) Height 200 m. (g) Height 220 m. (h) Height 240 m. (i) Height 260 m. (j) Height 280 m.

a) Training Set: For the training phase, data from eight distinct heights were utilized. In order to ensure a uniform distribution of data, the selected heights were nonconsecutive. The heights used for training the network were as follows: 100, 120, 140, 160, 200, 240, 260, and 280 m. This selection was made to enhance the diversity of the training set and optimize the learning process.

b) Testing Set: For validation and testing purposes, data from two distinct heights was utilized. To ensure a representative evaluation of the model's performance, the heights for the testing set were selected from within the overall distribution, rather than focusing on either the tail or head data. The selected heights for the testing set were 180 and 220 m. This approach aims to provide a comprehensive assessment of the model's generalization capabilities across a range of heights in the dataset.

4) Data Preprocessing: Prior to initiating the training of the network, several preprocessing steps were employed to optimize the training process, conserve time, and save computational resources, including processing RAM. The major steps involved in the preprocessing are outlined as follows.

a) Image Scaling: Considering the high resolution of the original images captured (5472 × 3648), which can be computationally demanding for processing in a CNN, it became essential to address the memory and processing constraints. The large image size not only consumes significant CPU resources but also limits the possibility of using large batch sizes, consequently extending the training time for the network. To mitigate these challenges, a preprocessing step involved scaling the images using interarea interpolation to a more manageable resolution of 512 × 256, aligning with the design of the network. This scaling process optimized the computational efficiency and facilitated faster training without compromising on performance.

b) Image Illumination: Given that the dataset was collected at various times of the day, significant variations in image illumination were observed. Certain images exhibited bright light with accompanying shadows while others were captured under normal lighting conditions. To address the issue of illumination invariance across the dataset, the contrast limited adaptive histogram equalization (CLAHE) technique was applied. By utilizing CLAHE, the luminance channel of the images was equalized, effectively mitigating the illumination discrepancies and enhancing the overall consistency of the dataset. This preprocessing step contributed to the improvement of the network's learning mechanism, resulting in enhanced results during subsequent stages of processing.

c) Nonoverlapping Image Pair Elimination: In order to ensure dataset uniformity and alleviate potential issues arising from insufficient overlapping, a preprocessing step was implemented to remove all image pairs captured at the edges of each flight path with less than 70% overlapping, as depicted in Fig. 3. To facilitate efficient management of the dataset, a comprehensive list of image pairs was compiled. By eliminating nonoverlapping pairs, the dataset was refined to enhance the reliability and effectiveness of subsequent stages in the research process.

d) Week Feature Elimination: To enhance the quality and relevance of the dataset, a final preprocessing step involved the removal of image pairs that lacked prominent features, specifically those captured on plain grounds with minimal distinct characteristics. By eliminating such image pairs, the dataset was refined to include only those with significant features, thus ensuring the efficacy of subsequent processing stages. In addition, all image pairs captured at the edges of each flight path were discarded, retaining only those with a substantial 70%–80% overlapping. A visual comparison of images taken from different heights but the same location is depicted in Fig. 4, further emphasizing the significance of this preprocessing step in curating a robust dataset for analysis and training purposes.

B. Results and Discussion

The proposed approach has been tested in terms of the correctness of height estimated by comparing two overlapping images taken from a UAV camera. In the subsequent paragraphs, a detailed account of the step-by-step process employed to enhance the results will be presented.

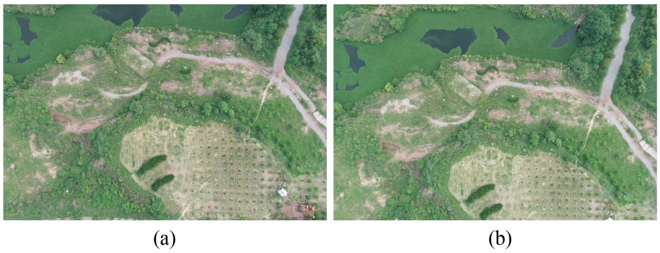


Fig. 5. Sample pair of images. This pair of images shows that due to the absence of prominent features, the prediction of height is not very accurate. (a) Left image. (b) Right image.

1) Luminance Disparities: The dataset comprises images captured at different times of the day, exhibiting varying illumination and lighting conditions. To address these variations and enhance the learning process of the network, CLAHE is employed to equalize the luminance channel of the color images. Equalizing solely the luminance channel is found to be superior to equalizing all the channels of the BGR image. As a result, all the images in the dataset are equalized before being fed into the network, leading to improved results.

2) Model Overadaptation: The primary challenge lies in the network's complexity due to the relatively limited size of the dataset. As a consequence, the network initially demonstrates signs of overfitting to the dataset, leading to unsatisfactory results on the validation dataset. To mitigate this issue, data augmentation techniques were employed. Considering the fixed input dimensions of our network, the images were subjected to 180° rotation and vertical and horizontal flipping while ensuring that the image dimensions remained unaltered. In addition, careful consideration was given to the direction in which the UAV was moving. Following the rotation and vertical flipping, the image pairs were arranged in reverse order to maintain a consistent direction (UAV moving in the forward direction) across all image pairs in the complete dataset. Subsequently, the initial results were compared with those obtained using the PSMNet approach after implementing the aforementioned modifications.

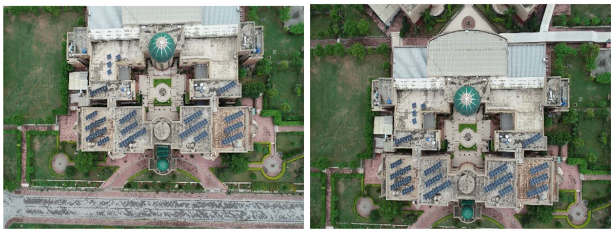
3) Preliminary Findings: The initial results, as presented in Table III, exhibited an improvement compared to PSMNet, yet they remained unsatisfactory due to the substantial average error of 10 m during height estimation. The primary objective of this research was to achieve an average error below 5 m. Further analysis revealed inconsistencies in the dataset, attributed to the imagery being captured at particular locations with varying heights during drone flights at the NUST Main Campus. Several issues were identified as follows.

1) *Variation in Heights:* Certain areas within the dataset exhibited varying heights, contributing to the overall inconsistency.
2) *Variability in Image Features:* Some image pairs lacked prominent features (see Fig. 5) while others contained significant features, introducing challenges in accurate matching (see Fig. 6).
3) *Insufficient Overlapping:* In instances where the drone changed its flight direction at specific locations, the image

TABLE III
COMPARISON OF PSMNET WITH OUR NETWORK (PSMHENET)

Milestones	Mean Absolute Height Error (m)	Min Max Error Value (m)	Ground Truth for Model Validation (m)
PSMNet	19.7	15.05–25.83	180
Proposed PSMHENet Initial Result	10.9	5.92–16.44	180
Proposed PSMHENet Final Result	**4.6**	**0.12–7.32**	180

The steps shows the modification in our network as well as data cleaning to improve results. Results obtained during initial traning process containing Average Height which is obtained by using Mean Absolute Height Error evaluation mechanism. Variations shows the minimum and maximum error in calculation the height. Initial results were obtained before data preprocessing (image scaling, illumination, removing non overlapping and weak images) whereas the final results were after data preprocessing



(a) (b)

Fig. 6. Sample pair of images. This pair of images shows that due to the presence of prominent features, the prediction of height is more accurate. (a) Left image. (b) Right image.

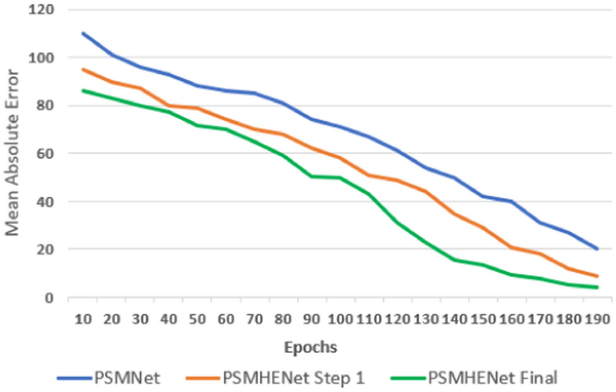


Fig. 7. Comparison shows mean absolute error (m) for 190 epochs. The blue line shows the PSMNet, which is used for comparison with our network (PSMHENet). The two versions of our network show the network after modification and data processing.

pair overlapping proved inadequate for effective comparison.

4) *Conclusive Outcomes:* Addressing these identified issues would be essential for enhancing the performance of the proposed approach and achieving the desired accuracy in UAV height estimation. Consequently, a data-cleaning process was employed to identify images lacking prominent features (as depicted in Fig. 5). These identified images, along with their corresponding augmented versions, were subsequently removed from the dataset. Following the completion of this data-cleaning procedure, a notable improvement in overall results was observed, as illustrated in Table III. The implementation of the data-cleaning process not only facilitated faster network convergence but also led to significant improvements in overall results, as depicted in Figs. 7 and 8.

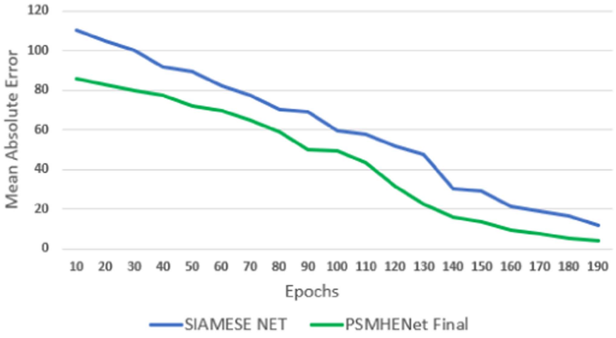


Fig. 8. Comparison of the Siamese network [35] with our network (PSMHENet) and how both networks converged to improve mean absolute error (m) after 190 epochs.

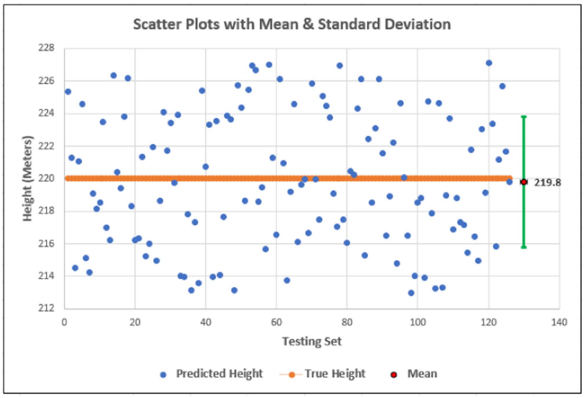


Fig. 9. Scatter plots of predicted and true heights along the Y-axis and the testing dataset along the X-axis. The testing set consists of 126 image pairs fed into the network having 220 m height. The red dot shows the mean of all predicted heights and the black line shows the standard deviation.

In conjunction with the data-cleaning procedure, continuous updates were made to the network parameters, encompassing shapes and other relevant parameters. As a result of these enhancements, the average error on the validation dataset notably reduced to 4.4 m while on the test dataset, it approached approximately 4.6 m. In Table IV, a comparative analysis of our network, denoted as PSMHENet, was conducted against the Siamese-based image-matching network [35] and the vision-based 3-D localization technique employed by Yol et al. [33]. Moreover, in Fig. 8, the convergence of both networks is illustrated, displaying their improvement in mean absolute error after 190 epochs.

TABLE IV

RESULTS OBTAINED WHEN OUR PROPOSED NETWORK (PSMHENET) IS COMPARED WITH TWO OTHER TECHNIQUES OF IMAGE MATCHING AND HEIGHT ESTIMATION, I.E., SIAMESE NETWORK *(OWN IMPLEMENTATION OF NETWORK) AND VISION-BASED 3-D LOCALIZATION USING MUTUAL INFORMATION

Method	Mean Absolute Height Error (m)	Min Max Error Value (m)	Ground Truth for Model Validation (m)
Siamese Networks *[35]	11.6	8.9 – 14.3	180
MI [33]	7.44	-	180
SSD [33]	6.68	-	180
Proposed PSMHENet	**4.6**	**0.12 – 7.32**	180

Moreover, Fig. 9 also illustrates the scatter plot to visualize the spread and standard deviation of the testing set.

5) Hardware, Running Times, and Evaluation Mechanism: The training of the model was conducted on a Linux machine equipped with a 22-GB GPU. The initial training process extended over several weeks, during which continuous updates were made to the model, along with parameter tuning to optimize its performance. Subsequently, the final epochs required approximately one week to complete. To evaluate the model's performance, the mean absolute height error was calculated.

V. CONCLUSION

In this research, we propose a novel approach for estimating the height of UAVs by leveraging a pair of images. We have devised a comprehensive pipeline to process these image pairs and accurately determine the UAV height. The process commenced with a stereo image dataset, where the image pair was input into our network to extract essential features. These features were then utilized for comparison and matching, allowing us to compute the disparity at each location on the feature map. This comparison was conducted using a 4-D tensor to calculate the cost, enabling accurate disparity estimation. Subsequently, we focused on refining the network by applying consecutive convolution and ReLU operations to enhance the output, reduce the feature size, and increase the number of feature maps. Through these improvements, we successfully estimated the height of the UAV from the ground. Following the training phase, we evaluated our network's performance on a separate dataset (test dataset), which consisted of heights not used during training and validation. The achieved results were promising, demonstrating the efficacy of our proposed approach.

Our research offers promising avenues for further enhancing the results and incorporating various improvements. First, ongoing efforts in refining methods and techniques for feature extraction and disparity calculation hold the potential to yield better and more efficient outcomes. Second, utilizing a dataset that encompasses diverse areas, including densely populated regions like the plains of Punjab with minimal height variations, could significantly contribute to improved results. Such areas would provide richer features for matching and disparity calculation, thus enhancing the accuracy and reliability of the estimation process.

REFERENCES

[1] I. F. Mondragón, M. A. Olivares-Méndez, P. Campoy, C. Martínez, and L. Mejias, "Unmanned aerial vehicles UAVs attitude, height, motion estimation and control using visual systems," *Auton. Robots*, vol. 29, pp. 17–34, 2010.

[2] Y. Ye, L. Bruzzone, J. Shan, F. Bovolo, and Q. Zhu, "Fast and robust matching for multimodal remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9059–9070, Nov. 2019.

[3] J.-F. Dou and J.-X. Li, "Robust image matching based on wavelet transform and SIFT," *Optik*, vol. 8009, pp. 437–444, 2011.

[4] A. Sedaghat and N. Mohammadi, "High-resolution image registration based on improved SURF detector and localized GTM," *t. J. Remote Sens.*, vol. 40, pp. 2576–2601, 2019.

[5] F. Gökçe, G. Üçoluk, E. Şahin, and S. Kalkan, "Vision-based detection and distance estimation of micro unmanned aerial vehicles," *Sensors*, vol. 15, pp. 23805–23846, 2015.

[6] Y. Ou, Z. Cai, J. Lu, J. Dong, and Y. Ling, "Evaluation of image feature detection and matching algorithms," in *Proc. 5th Int. Conf. Comput. Commun. Syst.*, 2020, pp. 220–224.

[7] M. Rodríguez, G. Facciolo, R. G. V. Gioi, P. Musé, and J. Delon, "Robust estimation of local affine maps and its applications to image matching," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2020, pp. 1331–1340. [Online]. Available: https://rdguez-mariano.github.io/

[8] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *Proc. 4th Int. Conf. Comput. Vis. Theory Appl.*, 2009, pp. 331–340.

[9] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[10] J. Cheng, Y. Wu, W. Abd-Almageed, and P. Natarajan, "QATM: Quality-aware template matching for deep learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11553–11562.

[11] J. Cheng, Y. Wu, W. Abd-Almageed, and P. Natarajan, "Image-to-GPS verification through a bottom-up pattern matching network," in *Proc. 14th Asian Conf. Comput. Vis.*, 2019, pp. 546–561.

[12] I. Rocco, M. Cimpoi, R. Arandjelović, A. Torii, T. Pajdla, and J. Sivic, "Neighbourhood consensus networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 546–561.

[13] A. Dosovitskiy, P. Fischer, J. T. Springenberg, M. Riedmiller, and T. Brox, "Discriminative unsupervised feature learning with exemplar convolutional neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 9, pp. 1734–1747, Sep. 2016, doi: 10.1109/TPAMI.2015.2496141.

[14] Y. Wang, W. L. Chao, D. Garg, B. Hariharan, M. Campbell, and K. Q. Weinberger, "Pseudo-LiDAR from visual depth estimation bridging the gap in 3D object detection for autonomous driving," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 8437–8445.

[15] Y. You et al., "Pseudo-LiDAR++: Accurate depth for 3D object detection in autonomous driving,," in *Proc. Int. Conf. Learn. Representations*, 2020.

[16] D. Garg, Y. Wang, B. Hariharan, M. Campbell, K. Q. Weinberger, and W.-L. Chao, "Wasserstein distances for stereo disparity estimation," in *Proc. Conf. Workshop Neural Inf. Process. Syst.*, 2020, pp. 22517–22529.

[17] J. R. Chang and Y. S. Chen, "Pyramid stereo matching network," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 5410–5418.

[18] X. Pan, D. Q. Ma, L. L. Jin, and Z. S. Jiang, "Vision-based approach angle and height estimation for UAV landing," in *Proc. 1st Int. Congr. Image Signal Process.*, 2008, pp. 801–805.

[19] M. H. Mughal, M. J. Khokhar, and M. Shahzad, "Assisting UAV localization via deep contextual image matching," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2445–2457, 2021, doi: 10.1109/JSTARS.2021.3054832.

[20] N. Mayer et al., "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation,," in *Proc. Comput. Vis. Pattern Recognit.*, vol. 12, 2016, pp. 4040–4048, doi: 10.1109/CVPR.2016.438.

[21] S. Zhang, Z. Wang, Q. Wang, J. Zhang, G. Wei, and X. Chu, "EDNet: Efficient disparity estimation with cost volume combination and attention-based spatial residual,," in *Proc. Comput. Vis. Pattern Recognit.*, 2021, pp. 5429–5438.

[22] M. A. Haseeb, J. Guan, D. Ristić-Durrant, and A. Gräser, "Disnet—A novel method for distance estimation from monocular camera," in *Proc. Int. Conf. Intell. Robots Syst.*, 2018, pp. 1–6.

[23] V. R. Kumar, M. Klingner, S. Yogamani, S. Milz, T. Fingscheidt, and P. Mäder, "SynDistNet: Self-supervised monocular fisheye camera distance estimation synergized with semantic segmentation for autonomous driving," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2021, pp. 61–71.

[24] F. Valocký, P. Drahoš, and O. Haffner, "Measure distance between camera and object using camera sensor," in *Proc. Cybern. Inform.*, 2020, pp. 1–4.

[25] M. Pavlović, I. Ćirić, M. Simonović, and V. Nikolić, "Application of different methods for distance estimation," in *Proc. Int. Conf. Path Knowl. Soc.-Managing Risks Innov.*, vol. PaKSoM, 2019, pp. 103–107.

[26] J. S. Yoon, K. Kim, O. Gallo, H. S. Park, and J. Kautz, "Novel view synthesis of dynamic scenes with globally coherent depths from a monocular camera," in *Proc. Comput. Vis. Pattern Recognit.*, 2020, pp. 5336–5345.

[27] M. Maximov and L. Leal-Taixé, "Focus on defocus: Bridging the synthetic to real domain gap for depth estimation," in *Proc. Comput. Vis. Pattern Recognit.*, 2020, pp. 1071–1080.

[28] A. E. R. Shabayek, C. Demonceaux, O. Morel, and D. Fofi, "Vision based UAV attitude estimation: Progress and insights," *J. Intell. Robot. Syst.: Theory Appl.*, vol. 65, pp. 295–308, 2012.

[29] D. Dhahbane, A. Nemra, and S. Sakhi, "Attitude determination and attitude estimation in aircraft and spacecraft navigation. A survey," in *Proc. 2nd Int. Workshop Human-Centric Smart Environ. Health Well-Being*, 2021, pp. 187–219.

[30] Y. Yang, Q. Shen, J. Li, Z. Deng, H. Wang, and X. Gao, "Position and attitude estimation method integrating visual odometer and GPS," *Sensors*, vol. 20, no. 7, pp. 1–16, 2020, doi: 10.3390/s20072121.

[31] X. Wan, J. Liu, H. Yan, and G. L. Morgan, "Illumination-invariant image matching for autonomous UAV localisation based on optical sensing," *ISPRS J. Photogrammetry Remote Sens.*, vol. 119, pp. 198–213, 2016.

[32] X. Liu, X. Li, Q. Shi, C. Xu, and Y. Tang, "UAV attitude estimation based on MARG and optical flow sensors using gated recurrent unit," *t. J. Distrib. Sensor Netw.*, vol. 17, 2021, Art. no. 15501477211009814.

[33] A. Yol, B. Delabarre, A. Dame, J.-É Dartois, and E. Marchand, "Vision-based absolute localization for unmanned aerial vehicles," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2014, pp. 3429–3434.

[34] J. Zbontar et al., "Stereo matching by training a convolutional neural network to compare image patches," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2287–2318, 2016.

[35] D. A. Beaupre and G. A. Bilodeau, "Domain Siamese CNNs for sparse multispectral disparity estimation," in *Proc. 25th Int. Conf. Pattern Recognit.*, 2020, pp. 3667–3674.

[36] Pix4D, "PIX4Dcapture Pro is a free planning app optimized for 3D mapping," 2020. [Online]. Available: https://www.pix4d.com/product/pix4dcapture/

Mansoor Khurshid received the master's degree in computer science from the National University of Sciences and Technology, Islamabad, Pakistan, in 2023.

He is currently a Project Manager and is responsible for managing IT and artificial intelligence projects. His research interests include computer vision and deep learning.

Muhammad Shahzad (Senior Member, IEEE) received the B.E. degree in electrical engineering from the National University of Sciences and Technology, Islamabad, Pakistan, in 2004, the M.Sc. degree in autonomous systems (robotics) from the Bonn Rhein Sieg University of Applied Sciences, Sankt Augustin, Germany, in 2011, and the Ph.D. degree in radar remote sensing and image analysis from the Department of Signal Processing in Earth Observation (SiPEO), Technische Universität München (TUM), Munich, Germany, in 2016. His Ph.D. topic was the automatic 3-D reconstruction of objects from point clouds retrieved from spaceborne synthetic-aperture-radar image stacks.

He has also attended two weeks of a professional thermography training course twice with Infrared Training Center, North Billerica, MA, USA, in 2005 and 2007. He was a Guest Scientist with the Institute for Computer Graphics and Vision, Technical University of Graz, Austria, from 2015 to 2016. He is currently a Guest Professor with TUM and a tenured Associate Professor with the School of Electrical Engineering & Computer Science, National University of Sciences & Technology, Islamabad. His research interests include applications of deep learning to solve remote sensing and computer vision problems, especially processing both unstructured/structured 3-D point clouds, optical RGBD images, and very-high-resolution radar data.

Hasan Ali Khattak (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering degree from Politecnico di Bari, Bari, Italy, in 2015.

He has been an Associate Professor with Sohar University, Sohar, Oman, and an Associate Professor with the School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Pakistan, since 2020. His current research interests include future internet architectures such as the web of things and leveraging data sciences and social engineering for future smart cities. Along with publishing in good research venues and completing successfully funded national and international projects, he is also a reviewer in reputed venues. His perspective research interests include applications of machine learning and data sciences for improving and enhancing quality of life in smart urban spaces especially in the healthcare and transportation domain through knowledge management, predictive analysis, and visualization.

Muhammad Imran Malik received the master's and Ph.D. degrees in artificial intelligence from the University of Kaiserslautern, Germany, in 2011 and 2015 respectively. His Ph.D. topic was automated forensic handwriting analysis on which he focused on both the perspectives of forensic handwriting examiners and pattern recognition researchers.

He was with the German Research Center for Artificial Intelligence GmbH (DFKI), Kaiserslautern. He is currently an Associate Professor and the Head of the Computer Science Department, School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad, Pakistan. He has authored more than 80 publications including high-ranked international conference papers, book chapters, and top-tier journal papers, which acquired more than 2500 citations at his Google Scholar profile. His research interests include machine learning and pattern recognition with a special emphasis on applications in image analysis and understanding.

Muhammad Moazam Fraz (Senior Member, IEEE) received the B.S. degree in software engineering from Foundation University, Islamabad 44000, Pakistan, in 2003, and M.S. degree in software engineering from National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan, in 2008, and the Ph.D. degree in computer science from the Faculty of Science Engineering and Computing, Kingston University London, Kingston, U.K., in 2013.

In 2003, he was a Software Development Engineer with Elixir Technologies Corporation, a California-based software Company. He was with Elixir until 2010 at various roles and capacities including software developer, development manager, and program manager. After completing the Ph.D. degree, he was a Postdoc Research Fellow with Kingston University in collaboration with the St George's University of London and U.K. BioBank on development of an automated software tool for epidemiologists to quantify and measure retinal vessels morphology and size, determine the width ratio of arteries and veins as well as the vessel tortuosity index on very large datasets, and enable them to link systemic and cardiovascular disease to the retinal vessel characteristics. He is currently a Full Professor with the School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad, Pakistan. He is also a Rutherford Visiting Fellow with The Alan Turing Institute, U.K. His research interests include applications of machine learning/computer vision techniques for diagnostic retinal image analysis.

Dr. Fraz's Ph.D. dissertation was nominated for IET Excellence Awards in 2013. He was the recipient of two Gold Medals for "Best Graduate Award" and "Securing Top Position in the batch." He is a PMI (www.pmi.org) certified Project Management Professional.