

Received 11 March 2024, accepted 10 April 2024, date of publication 19 April 2024, date of current version 7 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3391371

RESEARCH ARTICLE

Economic Fruit Trees Recognition in Hillsides: A CNN-Based Approach Using Enhanced UAV Imagery

MARAL HOOSHYAR^{ID}, YUAN-SHUO LI, WEN CHUN TANG, LING-WEI CHEN,
AND YUEH-MIN HUANG^{ID}, (Senior Member, IEEE)

Department of Engineering Science, National Cheng Kung University, Tainan 70101, Taiwan

Corresponding author: Yueh-Min Huang (huang@mail.ncku.edu.tw)

This work was supported by National Science and Technology Council under Grant 110-2321-B-067F-001

ABSTRACT Managing sloping terrains worldwide presents a significant challenge due to the lack of structured management practices, particularly when integrating various fruit trees in irregular arrangements. This study addresses the complexity arising from mixed cultivation by proposing a solution utilizing Unmanned Aerial Vehicle (UAV) imagery for tree species recognition and fruit tree classification. Our approach involves equipping UAVs with multispectral and optical cameras to capture imagery over experimental sloping terrain. The collected data undergoes processing for classifying different types of fruit trees, roads, and buildings through Orthophoto images. Convolutional Neural Networks (CNN) are employed for image recognition in challenging hillside terrains, with deep neural network methods, specifically VGG-16, VGG-19, and ResNet-50, being applied and compared. VGG-16 achieved significant accuracy in multispectral imagery analysis. Subsequently, various image fusion techniques, including Brovey, Hue-Saturation-Value, Principal Components Analysis (PCA), and Gram-Schmidt, were explored, with PCA demonstrating superior performance. The study revealed that image fusion, particularly with near-infrared or red-edge bands, significantly enhanced prediction accuracy compared to standalone multispectral imagery. The combination of visible-band fused imagery with additional spectral bands yielded the highest accuracy, improving overall prediction accuracy from 0.76 to 0.92. This research provides valuable insights applicable to diverse regions grappling with challenges in managing sloping terrains and mixed fruit tree cultivation.

INDEX TERMS Convolutional neural network, fruit trees classification, image fusion, multispectral imagery, unmanned aerial vehicle.

I. INTRODUCTION

In recent decades, the utilization of image processing has expanded its reach into various sectors including education [1], healthcare [2], agriculture [3], unmanned aerial vehicles (UAVs) [4], and numerous other domains. Advancements in technology have helped the agriculture industry apply a profound transformation to prosper farming and eliminate challenges and issues [5]. Employing UAVs in agriculture has made a revolution because of providing real-time data for farmers to take action on time with more productivity (e.g., [6], [7]). UAVs are equipped with various sensors and

imaging technologies, enabling them to identify and classify plants features for different aims [8]. For instance, UAVs have been employed to optimize irrigation with remarkable precision [9]. Also, UAVs have provided significant opportunities for farmers to monitor farms in terms of crop health and detect diseases [10]. Obviously, UAV imagery is capable of providing vast amounts of effective data to help farmers enhance productivity and minimize environmental impact. Notably, many research articles emphasized that using pesticides ends up with environmental pollution in terms of soil (e.g., [11], [12]) and water (e.g., [13], [14]). Multispectral and optical cameras mounted on UAVs capture imagery aimed at detecting crop diseases and pests, ultimately enhancing agricultural productivity by fostering healthier crop growth [15].

The associate editor coordinating the review of this manuscript and approving it for publication was Jiankang Zhang^{ID}.

Multispectral imagery excels in capturing specific wavelength bands, enabling the detection of diseases and irrigation issues for targeted pest control (e.g., [16], [17]). Meanwhile, optical cameras, based on RGB technology, provide high-resolution images essential for detailed crop monitoring, growth analysis, and identifying visible signs of damage (e.g., [18], [19], [20]). The integration of these technologies offers a synergistic approach to precision agriculture, leveraging the detailed spatial information from optical cameras alongside the rich spectral data from multispectral imagery to enhance the accuracy and efficiency of pest management strategies (e.g., [21], [22]). Our research aims to leverage UAV imagery, combining multispectral and optical cameras, to address the unique agricultural challenges faced by farmers in the hillsides of Taiwan. These areas, as highlighted in prior studies (e.g., [23], [24]), present significant access difficulties for traditional agricultural machinery due to their rugged terrain and variability in elevation. Our study's focus is on a diverse agricultural field within the Nanhua District of Tainan City, characterized by its steep hillsides and a variety of fruit trees, including lychee, longan, plantain, mango, peach, and bamboo. This diversity, coupled with the visual similarity between bamboo and lychee trees, complicates pest management strategies, necessitating precise identification and treatment methods tailored to each fruit type.

Our approach utilizes UAVs equipped with advanced imaging technologies, including optical cameras and multi-channel spectrometers, to navigate the low-altitude environments unique to our experimental site. By capturing images across red, green, blue, near-infrared, and red-edge spectral bands, we aim to compute vegetation indices and texture features that are indicative of specific crop health and pest presence. The integration of high-resolution optical images with multispectral data allows for detailed analysis of tree and fruit types, enhancing our ability to differentiate between species with similar appearances, such as bamboo and lychee trees.

Recognizing the complementary nature of optical and multispectral imagery [25], our research employs image fusion techniques, including Resampling, Brovey, HSV, PCA, and GS methods, to improve the overall accuracy and resolution of the images. These enhanced images then serve as the basis for developing a multispectral image recognition system, which utilizes CNNs with architectures like VGG-16, VGG-19, and ResNet-50 for accurate classification and recognition of the diverse fruit trees present.

Thus, the ultimate goal of our study is to refine the precision of agricultural practices on hillsides, employing UAV technology to overcome the limitations imposed by the terrain. By comparing the efficacy of various CNN models and image fusion techniques, we aim to identify the most effective methods for enhancing the quality of UAV imagery. This will enable targeted pest management strategies, reduce the need for manual labor, and decrease the time and financial costs associated with traditional pest control methods.

The contributions of our proposed approach are summarized as follows:

- Innovative approach for pest management: Utilizing real-time data and UAV technology for precise identification and targeted pesticide spraying on hillsides, significantly enhancing pest control efficiency and reducing pesticide usage.
- Technological advancements in image processing: Integrating CNN with advanced image fusion techniques (Brovey, HSV, PCA, and GS) and optimizing with state-of-the-art neural network models (VGG-16, VGG-19, ResNet-50) for unparalleled accuracy in pest identification.
- Establishing new standards in precision agriculture: By developing enhanced channels from standard RGB data and leveraging the latest in UAV and image processing technologies, our research sets new benchmarks for accuracy and efficiency in precision agriculture, contributing to sustainable farming practices and improved crop health on challenging terrains.

The structure of this article is as follows: section two presents related works, section three describes materials and methods, section four deals with experimental design and results, and section five presents discussion and conclusions.

II. RELATED WORKS

The geographical constraints and dense population density in Taiwan have led farmers to seek alternative means of maximizing agricultural output, particularly through the utilization of hillside areas. However, the aging demographic of Taiwanese farmers coupled with the limitations of traditional agricultural practices have posed significant challenges in meeting modern productivity demands. In this context, the integration of UAVs into agriculture emerges as a compelling solution, offering real-time data that streamlines farming processes and enhances overall efficiency [26].

UAVs streamline agricultural production by reducing time, costs, and labor-intensive tasks [27]. Numerous studies advocate for the use of UAVs in remote sensing, facilitating the identification and classification of tree species using real-time data (e.g., [28], [29], [30], [31]). Conventional methods of interpreting remote sensing imagery for tree species classification necessitate substantial human resources and financial investments [28]. Multi-temporal imagery captures crucial data on vegetation phenology, such as temporal changes in plant growth stages and stand structure, thereby improving classification accuracy [29]. In the realm of multispectral remote sensing, variables such as spectral characteristics, vegetation indices, texture, shape, and structure are leveraged to construct robust classification models [30]. Machine learning algorithms analyze plant features from captured images, enabling precise classification [30]. Moreover, the integration of spectral images with machine learning techniques further refines plant categorization (e.g., [31], [32], [33]). Traditional aerial photography often faces limitations

due to weather conditions, notably cloud cover, which can impede visibility. In contrast, multi-axis rotor UAVs offer enhanced maneuverability, operational flexibility at low altitudes, reduced costs, and the ability to capture imagery even in adverse weather conditions such as heavy cloud cover [34]. The proliferation of high-resolution UAV imagery has democratized access to agricultural imaging, thereby amplifying the beneficial applications of UAV technology in the agricultural sector.

Several studies have employed UAV imagery coupled with machine learning algorithms to address various agricultural challenges [35]. Zhou et al. [36] utilized RGB imagery to classify wetland vegetation using object-based image analysis scenarios and machine learning algorithms, including Bayes, KNN, SVM, DT, and RF. Their findings highlighted RF's superior performance, achieving an overall accuracy of 89.76%. Similarly, Feng et al. [37] investigated weed presence in farmland using UAV images, where RF outperformed SVM, DT, and KNN algorithms after feature extraction and dimensionality reduction. Ye et al. [38] leveraged SVM, ANN, and RF algorithms with UAV-based multispectral imagery for banana Fusarium wilt detection, with SVM exhibiting the best performance.

With the rise of deep learning, CNNs have gained prominence for crop detection in complex scenarios. Tetila et al. [39] utilized CNNs for soybean pest detection, with ResNet-50 achieving the highest accuracy of 93.82%. Li et al. [40] employed data augmentation techniques and various CNN architectures, with GoogLeNet demonstrating superior performance for crop pest recognition, albeit with increased computational requirements. While hyperspectral imagery is recognized as an advanced tool, its complexity and cost compared to multispectral imagery are acknowledged (e.g., [41], [42]). Multispectral UAV imagery has shown promise in urgent monitoring tasks such as wheat yellow rust identification [42] and sunflower lodging detection using deep learning and image fusion [43].

Despite the valuable insights provided by existing research, there remains a gap in leveraging image fusion techniques to enhance accuracy. Our study seeks to address this gap by proposing a CNN-based approach that integrates multispectral and optical cameras. This integration enables us to capture images across various spectral bands, including red, green, blue, near-infrared, and red-edge, facilitating the computation of vegetation indices and texture features indicative of specific crop health and pest presence. Furthermore, the fusion of high-resolution optical images with multispectral data enhances our ability to analyze tree and fruit types in detail, thereby aiding in the differentiation of species with similar appearances, such as bamboo and lychee trees. Ultimately, this integration of UAV imagery, optical cameras, multispectral imagery, and CNN represents a significant advancement in precision agriculture, particularly in promoting sustainable practices in hillside regions.

III. MATERIALS AND METHODS

A. STUDY AREA

The study area is nestled within the picturesque mountainous expanse of Nanhua District in Tainan City, pinpointed at coordinates $23^{\circ}0'18.2''$ N and $120^{\circ}28'15.9''$ E (Figure 1). With an area of 171.5 square kilometers and 3,602 hectares of arable land, this area is the largest in Tainan City and is located between 250 and 900 meters above sea level. It is a very hilly area with mountains on all sides. The terrain gradually slopes from the northeast to the west. At 900 meters above sea level, the highest peak in the area is Guanshan, a branch of the West Aliguan mountain range. The land in this area is hilly, and the soil is fertile and suitable for planting fruit trees. This geographical setting was deliberately chosen to embody various altitudes across the hillsides of Taiwan. Importantly, the mountains surrounding the experimental field area are likely to obscure the signal, resulting in poor signal strength when measuring the field and increasing the difficulty of this study. Our research is centered around the implementation of vertical take-off and landing for a quadrotor UAV, a strategic choice dictated by the region's challenging topographical characteristics. Particularly, the quadrotor UAV in our study was equipped with a MicaSense Altum multispectral camera and a Zenmuse X4S optical camera. These cutting-edge technologies empower our investigation, enabling the acquisition of high-quality aerial imagery essential for our research objectives.

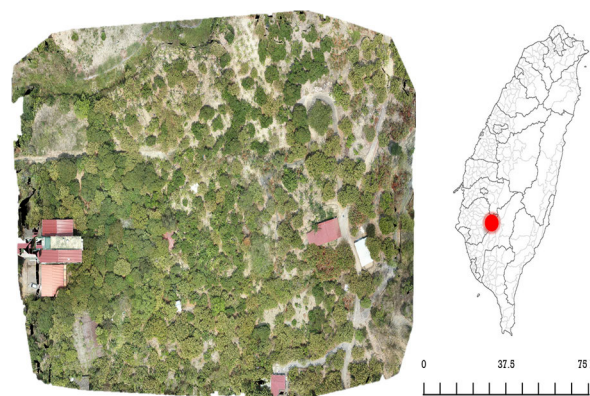


FIGURE 1. Orchard on sloped terrain at the site in Nanhua District, Tainan City.

In this research, we set up 12 aerial survey points in the experimental field with regards to the hilly terrain of the site, there is a vertical difference exceeding 10 meters between the highest and lowest points. Images reflecting this height difference impact the accurate representation of fruit trees areas in subsequent forward projection synthesis. In this study, the UAV image acquisition method involves ground flight, using the original ground height from the Digital Elevation Model (DEM), obtained by eliminating ground objects from the real ground model. Images are captured from 30 meters above the original ground height.

B. DATA COLLECTION AND LAYOUT OF UAV IMAGERY

This study employs aerial photogrammetric targets placed in an experimental field to enhance the accuracy of UAV orthophoto images for identification of fruits types. Twelve rectangular aerial survey points, measuring 33 × 33 cm, were positioned using a topographic GNSS receiver. Table 1 presents the three-dimensional coordinates measured by the receiver at 12 aerial survey points; with reference to the TWD97 geodetic datum as the utilized coordinate system. The study addresses challenges posed by hilly terrain by employing a four-rotor UAV with vertical take-off and landing capabilities, equipped with MicaSense Altum multispectral and Zenmuse X4S optical cameras. Images are captured 30 meters above the original ground height, determined from a digital elevation model. Over three periods (March 2020, July 2020, and April 2021), a total of 21600 UAV aerial images, including optical and multispectral data, were acquired and processed to generate orthophotos for the specified fruits (lychee, longan, plantain, mango, peach, and bamboo) for subsequent analysis. The study focuses on differential fruits identification using texture analysis and emphasizes the impact of image resolution on experimental results.

TABLE 1. Coordinate information on aerial survey points in the Nanhua area of Tainan City.

Serial Number	Point Number	(X, Y)	Ellipsoidal Height
1	A6	(2544944.799, 195777.582)	160.082
2	A9	(2544895.474, 195781.5)	160.169
3	B6	(2544887.872, 195748.473)	157.925
4	B5	(2544897.647, 195729.827)	154.763
5	A7	(2544910.24, 195708.652)	150.257
6	B3	(2544987.611, 195764.753)	149.278
7	A2	(2544956.114, 195742.349)	153.075
8	A1	(2544966.842, 195693.718)	142.855
9	B7	(2544967.528, 195640.472)	140.992
10	B2	(2544987.817, 195718.242)	145.294
11	B4	(2544899.63, 195675.639)	149.745
12	A5	(2544942.089, 195716.036)	145.981

The gathered images not only encapsulate general optical views but also include multispectral images, categorized into five distinct types based on differing band information: red, green, blue, near-infrared, and red edge. This categorization provides a rich dataset for our investigation, and Figure 2 visually delineates the five spectral bands captured by the multispectral camera (Altum) alongside the aerial images taken by the optical camera (X4S).

Knowing the location and coordinates of each measurement, we can calculate where the camera was when capturing

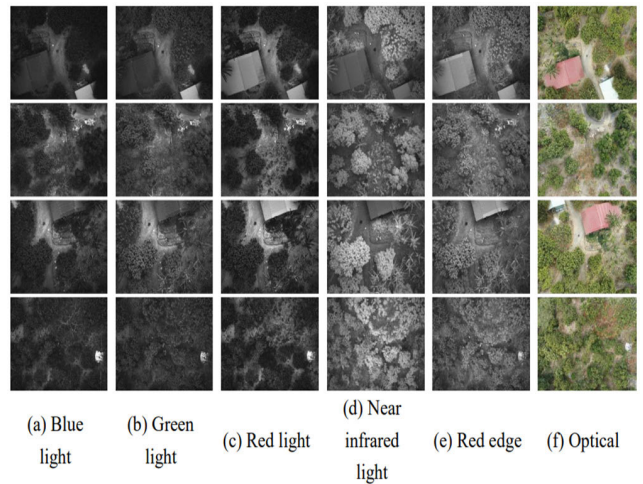


FIGURE 2. Schematic diagram of multi-spectral and high-resolution optical images captured by UAV.

the images. This calculation is essential for later creating the orthophoto, topographic map, and numerical terrain model.



The UAV flight mission was specified as follows: the UAV was assigned to fly over 12 designated points outlined in Table 1, encompassing various heights across the entire experimental site. The optical camera was utilized to capture visual imagery within the visible spectrum, providing detailed color images of the entire experimental site, including fruit trees of varying heights. Concurrently, the multispectral sensor captured imagery across multiple spectral bands beyond the visible spectrum, such as near-infrared (NIR) and red-edge, offering valuable insights into vegetation health and other related parameters.

By combining the capabilities of the optical camera and multispectral sensor, the UAV facilitated simultaneous capture of both visual and spectral information. This approach enabled thorough data collection essential for applications such as vegetation monitoring, land cover and fruit trees classification, and fruit trees health assessment. The flight plan incorporated systematic coverage with overlapping flight paths to ensure comprehensive photo capture across the experimental site. This meticulous approach to data collection facilitated accurate and detailed analysis of the terrain and vegetation, contributing to the success of the UAV mission.

C. THE WORKFLOW OF CNN

The analysis was developed and tested using a 64-bit operating system with the following specifications: an Intel® Core™ i7-8700 CPU, a GTX1080Ti graphics card, and 64 GB of RAM. In addition, our research incorporated the use of two cameras for identification purposes: MicaSense’s multispectral camera Altum, capable of receiving 5 different spectral bands, and Zenmuse’s optical camera X4S with a 20-megapixel sensor. Table 2 presents all the specifications

TABLE 2. Specifications of multispectral camera Altum and Zenmuse's optical camera X4.

	MicaSense Altum	Zenmuse X4S
Camera		
Weight	Altum + DLS 2: 406.5g Altum: 357 g DLS2: 49 g	X4S:253 g
Size	8.2 cm × 6.7 cm × 6.45 cm	12.5 cm × 10 cm × 8 cm
Spectral channel	Blue, Green, Red, Red Edge, Near-infrared (NIR)	NO
Wave length (nm)	<ul style="list-style-type: none"> Blue light (475 nm center, 32 nm bandwidth) Green light (560 nm center, 27 nm bandwidth) Red light (668 nm center, 14 nm bandwidth) Red edge (717 nm center, 12 nm bandwidth) Near-infrared light (842 nm center, 57 nm bandwidth) 	Visible light (380nm – 790nm)
Thermal image	Long wavelength infrared (8-14um)	none
Resolution	2064 x 1544 (multispectral) 162 x 120 (thermal image)	3:2, 5472×3648 4:3, 4864×3648 16:9, 5472×3078
Ground Sampling Distance (GSD)	2.1 cm/pixel (multispectral) at 60 meters 41 cm/pixel (thermal image) at 60 meters	0.865 cm/pixel at 30 meters
Catch rate	1 capture per second (all bands)	60 capture per second(4K)
Interface	USB 3.0: Connect wireless network card USB 3.0: storage device	DJI gimbal connector 2.0
Visible range	48° x 37° (multispectral) 57° x 44° (thermal image)	84°
focal length	8 mm (multispectral)	8.8 mm
Store	USB 3.0 compatible storage device	UHS-3 MicroSDXC 128GB

of both multispectral camera Altum and Zenmuse's optical camera X4S.

We implemented the CNNs workflow by a comprehensive set of software tools and technologies. Our software stack

included Windows, Anaconda, and Python 3.6, along with key libraries such as TensorFlow_gpu-1.12.0, Keras-2.2.4, CUDA9.0, cuDNNv7.6.5, python-opencv, Scikit-Learn, and Pix4Dmapper. This robust ensemble enabled seamless integration and efficient execution of our CNN-based analysis for extracting valuable insights from our data. The synergy of these tools not only streamlined our workflow but also enhanced the precision and resolution of our image processing and analysis.

Our research aims to explore three different scales for CNN workflow, namely: 100×100 , 150×150 , and 200×200 pixels, for CNN training. Importantly, as the 150×150 and 200×200 boxes may encompass various types of plants or targets, we utilized the smaller 100×100 pixel box to capture “partial” features as training data for the prediction model. This smaller window size filters out unnecessary information during training, allowing a focus on recognizing various types of fruits.

D. ORTHOPHOTO PRODUCTION

Orthophoto production for UAV imagery involves correcting distortions caused by various factors such as sensor geometry, terrain relief, and atmospheric conditions. Thus, the process of obtaining orthophotos via multispectral images includes the below key steps:

Initial Image Capture:

- UAVs capture images of the target area using optical camera and multispectral UAV images.

Atmospheric Scattering Correction:

- Grayscale values of pixels in the images are affected by both the reflection from surface objects and atmospheric scattering.
- Dark Object Subtraction (DOS) is applied to remove the influence of atmospheric scattering.
- Pixels in areas assumed to be completely dark (e.g., shadows, clear water bodies) are considered to absorb radiation and should ideally have a grayscale value of 0.
- Due to atmospheric scattering, the actual values of these dark pixels received by the sensor are not 0.
- The minimum value among the grayscale values of dark pixels is identified and subtracted from all pixels in the entire image, effectively removing atmospheric interference.

Wavelength-Dependent Processing:

- Recognizing that the degree of scattering is wavelength-dependent, each band of the multispectral image is processed individually to account for variations in atmospheric effects.

Incident Light Sensor Integration:

- An incident light sensor (DLS 2) directly connected to the Altum camera measures ambient light and sun angle during the task execution.
- This information is recorded in the data of TIFF images captured by the camera.

TABLE 3. UAV image collection and orthophoto production.

Project	2021-04-10 Optical_orthophoto	2021-04-10 MS_orthophoto
Camera Model Name(s)	FC6510_8.8_5472 X3648 (RGB)	Altum_8.0_2064x1544(Blue, Green, Red.NIR.Red Edge)
Ground Sampling Distance (GSD)	0.865 cm/pixel	1.43 cm/pixel
Area Covered	0.024 km ² /2.4011 ha/5.9363 acres	0.024 km ² /2.4384 ha/6.0286 acres
Ground Control Points	10	10
Data Set	310 images	3200 images
Output Coordinate System	TWD97/TM2 Zone 121	TWD97 / TM2 Zone 121

- Specialized processing tools, such as Pix4Dmapper, utilize this information to correct global lighting changes, particularly those caused by factors like cloud coverage.

GPS Data Integration:

- DLS 2 provides GPS data to the Altum camera, which is recorded in the photo data.
- The GPS receiver is designed to maintain low power consumption to reduce overall power usage.

Orthophoto Production with Pix4DMapper:

- Pix4DMapper, an aerial survey program, is employed to stitch UAV images together quickly and convert the aerial plan into a 3D model.
- Outputs include ortho-mosaic images, 3D point clouds, and digital surface models (DSM).
- Orthophoto images provide an accurate surface description, and the actual coordinate information can be extracted from the image content.

In summary, the described process involves mitigating atmospheric effects, utilizing incident light and GPS data for accurate corrections, and using specialized software for efficient orthophoto production from high-altitude optical and multispectral UAV images. This allows for detailed and accurate representation of the target area for further research and analysis. In this regard, Table 3 represents UAV image collection and orthophoto production of our research.

E. IMAGE COINCIDENCE MATCHING OF MULTISPECTRAL AND OPTICAL IMAGES

After testing and verifying during the development process of this research, we discovered slight errors in the spatial coordinate information attached to the multispectral and optical orthophotos. The points are slightly shifted on both the multispectral and optical images. Figure 3(a) illustrates

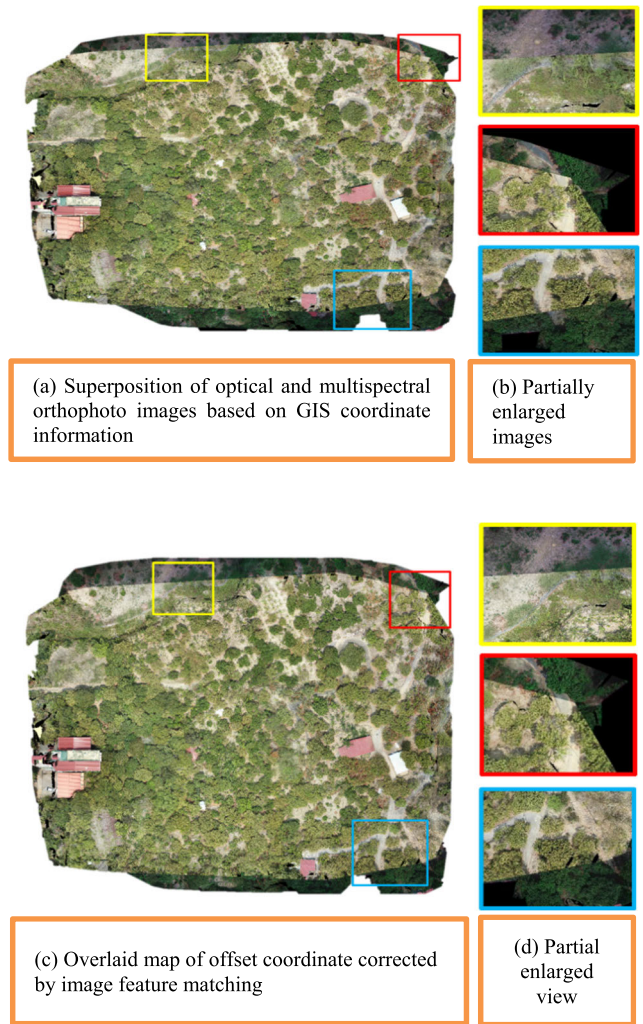


FIGURE 3. Schematic diagrams of optical and multispectral orthophoto images before and after coincidence and matching.

the result when overlapping the optical and multispectral orthophotos based on their coordinate information. In three positions, zooming in clearly shows that the two orthophotos are not completely superimposed; the roads in the images are also slightly offset (see Figure 3(b)). Such offset between orthophoto images leads to information asymmetry in subsequent multispectral image fusion. Therefore, it is necessary to correct this offset so that the two images can be aligned through the coordinates. Corrected orthophoto images will thus not introduce errors in subsequent image fusion (see Figure 3(c)). Also, Figure 3 (d) represents the corrected partial enlarged views.

This study employs the coordinate space information calculated from orthophoto images to extract two regions with the same coordinate range. Subsequently, SIFT feature matching is utilized to match the feature points. Figure 4 displays the same coordinate range. From the spectral and optical images, it is evident that the two images are slightly offset. By matching the feature points in the images, yellow

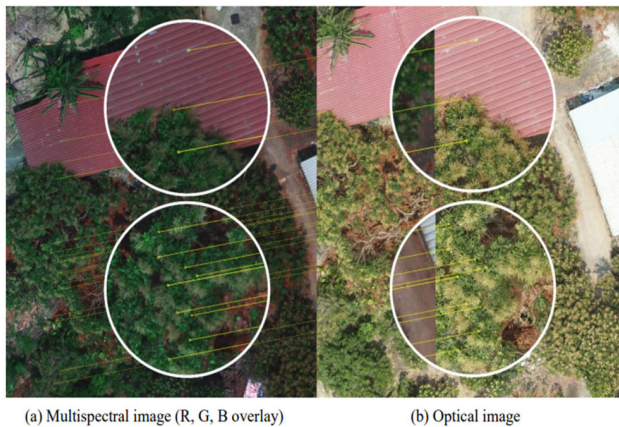


FIGURE 4. Schematic diagram of multispectral and optical image feature point matching.

line segments are used to connect the corresponding points, revealing the coordinate offset clearly. Therefore, the difference between the pixel positions of the feature points in the two images is calculated to determine the horizontal and vertical offsets. These values are then multiplied by the Ground Sampling Distance (GSD) to obtain the actual coordinate offset of the two orthophotos. By adding the offset to the orthophoto coordinate system preset as the reference point, we obtain coordinate information that is consistent with both orthophotos.

F. IMAGE FUSION OF MULTISPECTRAL AND OPTICAL IMAGES

Image fusion serves to enhance information from the same scene captured by diverse sensors, thereby improving image visibility, resolution, and the characteristic details of the analyzed object. This study focuses on UAV-captured images, categorized into two types: multi-spectral and optical images. Multi-spectral images, while having lower resolution, provide richer spectral information (e.g., [44], [45]), whereas optical images offer higher resolution with fewer band details (e.g., [46], [47]).

To enhance our analysis and elevate the precision of our results, we meticulously align optical and multispectral images before applying a suite of sophisticated image fusion techniques. These include PCA, Brovey, HSV, and the GS process. It is essential to acknowledge the robust support from the academic community for these methods, with numerous publications validating the effectiveness of Brovey image fusion (e.g., [48], [49]), HSV image fusion (e.g., [50], [51]), PCA image fusion (e.g., [52], [53]), and GS image fusion (e.g., [54], [55]) in improving image resolution and clarity. Unlike the conventional approach of accessing these advanced fusion methods through commercial GIS software at an additional cost, our strategy is to directly incorporate these techniques into our Python-based framework. This approach not only aims to bolster the integrity and efficiency

of our proposed recognition system but also enhances our ability to precisely interpret and utilize the environmental data captured in our study.

1) BROVEY IMAGE FUSION

In this investigation, we adapt the Brovey image fusion technique [56] by implementing adjustments. Grayscale processing is applied to the optical image, resulting in a single-band image that replaces the panchromatic image. Concurrently, the multispectral image undergoes resampling to align with the dimensions of the optical image, producing a high-resolution multispectral image. Notably, the optical image is represented numerically in three channels: R, G, and B. Grayscale processing transforms this into a single-band grayscale image. The multispectral image is further resampled to match the resolution of the optical image. We then calculate the average of the corresponding resampled multispectral image bands to generate a single-band image result. The next steps involve dividing the two single-band images to produce a new ratio result. Finally, the resampled multispectral image is multiplied by the corresponding ratio value within the image, facilitating the fusion of high-resolution multispectral imagery results. This study employs a modified Brovey image fusion approach for both optical and multispectral images (R, G, B). Figure 5 illustrates three perspectives of the site, corresponding to the optical, multispectral, and fusion images.

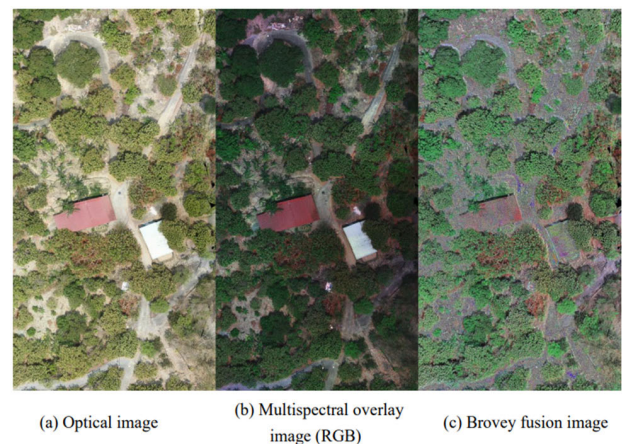


FIGURE 5. Optical and multispectral orthophoto images before and after Brovey fusion.

2) HSV IMAGE FUSION

Hue-saturation-value (HSV) fusion capitalizes on the high spatial resolution features of panchromatic or optical images along with the rich spectral information of multispectral images, creating a synergy that complements the overall information [57].

The process involves resampling low-resolution multispectral images to a higher resolution. Subsequently, the brightness (value) of the multispectral image is replaced with

the brightness of the optical image following HSV format conversion. This step aims to restore image detail after the multispectral image has been resampled. The hue and saturation of the multispectral image are then applied.

The resulting image, as illustrated in Figure 6, not only preserves the spectral characteristics of the original channels but also exhibits a notable enhancement in image detail.

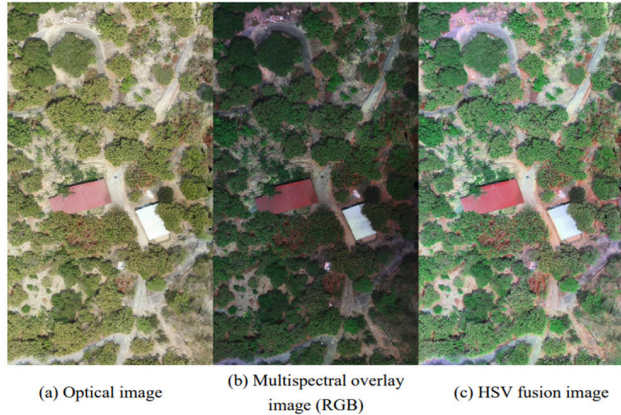


FIGURE 6. Optical and multispectral orthophotos before and after HSV fusion.

3) PCA IMAGE FUSION

This method identifies a set of comprehensive variables to substitute the original variables, aiming to capture the maximum information from the original variables (e.g., [58], [59], [60]). This set of variables is referred to as the first principal component. In the context of image fusion, the first principal component of a high-resolution image takes the place of the first principal component of the low-resolution multispectral image. Employing this new set of variables, the data undergo an inverse transformation through PCA to recover the high-resolution multispectral image. The restored image, depicted in Figure 7 as the outcome of the PCA fusion process.

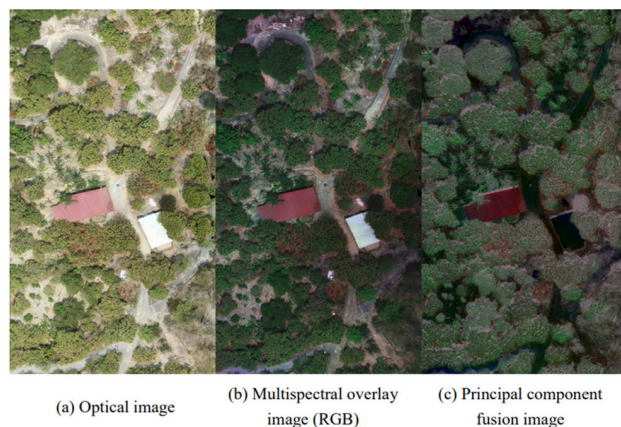


FIGURE 7. Optical and multispectral orthophotos before and after PCA image fusion.

4) GS IMAGE FUSION

To make a multispectral digital image clearer, we use the GS method. This process involves combining lower-quality color images to create a sharper and more detailed images [61].

Represents a widely employed multi-dimensional linear orthogonal transformation in statistics, particularly for orthogonalizing multi-dimensional data in remote sensing images. Beyond the reduction of redundant information within the image, it effectively eliminates strong correlations between adjacent bands. In this transformative process, the multispectral low spatial resolution image serves to simulate the single-band low-resolution image. Subsequently, the simulated single-band image is incorporated as the initial component prior to the GS transformation, applied to the low-resolution multi-dimensional image.

Following the GS transformation, the average and standard deviation of the high-resolution single-band image and the first component are computed. This calculation results in a modified high-resolution single-band image. Subsequently, a new cube is generated by substituting the first component after the GS transform with the modified high-resolution single-band image. In the final step, the dataset undergoes an inverse transformation by GS, culminating in a multispectral image featuring enhanced spatial resolution. This algorithm accommodates images with any number of bands, ensuring that the resulting fused image preserves the spectral characteristics of low spatial resolution bands, minimizing information distortion, as illustrated in Figure 8.

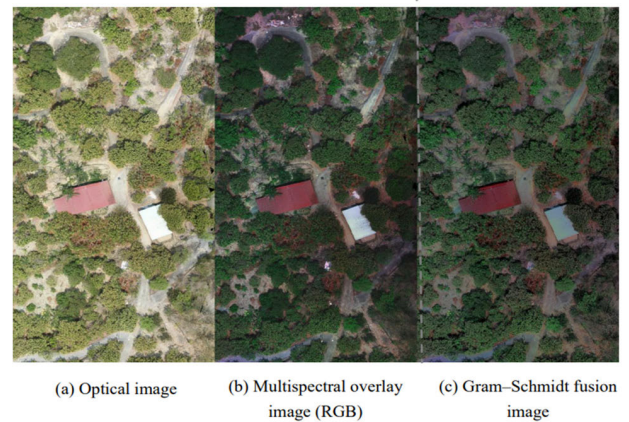


FIGURE 8. Optical and multispectral orthophotos before and after GS fusion.

G. TRAINING DATA SAMPLES

The study area, illustrated in Figure 9, is partitioned into four equal sections. Considering the challenging topography and limited manpower, our focus narrowed to the fruits species situated in the lower right corner of the site. On-site investigations were conducted to support the manual labeling of image data and the creation of the UAV high-altitude image dataset. A total of 1,693 data samples were meticulously labeled,



FIGURE 9. Area of interest for dataset production and capture.

comprising both fused and cropped multispectral and optical images.

Due to the uneven distribution of fruits across the region, certain categories had limited labels. To address this, data enhancement techniques were applied to augment the number of samples in specific categories to 350, mitigating data imbalance during model training. Subsequently, 80% of the labeled data was allocated for model training, with 20% reserved for both validation and testing post-model training.

The area exhibits hybrid planting, with lychee and longan trees intermixed and cultivated over a larger expanse, along with diverse fruits dispersed throughout. The field encompasses six different fruits: mango, plantain, bamboo, peach, lychee, and logan. Given the utilization of the sliding window method for large-scale orthophoto image recognition, determining an appropriate window size is crucial. However, since various fruits occupy differently sized areas based on vertical ground observation, a one-size-fits-all window is impractical. The GSD for the orthoimage in this study is 0.865 cm/pixel, resulting in orthophoto divisions into three boxes sized 100×100 , 150×150 , and 200×200 pixels.

Recognizing that the larger 150×150 and 200×200 boxes may encompass different fruit types or targets, we opted for the smaller 100×100 pixel box to capture “partial” features as training data for the prediction model. This choice helps filter out unnecessary information during training, as not all fruits align with the same-sized window. Training exclusively involved small-scale texture extraction from fruits, contributing to a reduction in the model’s overall size. Fig. 10 displays images from the dataset employing a window size of 100×100 pixels.

IV. EXPERIMENTAL DESIGN AND RESULTS

The UAV was outfitted with dedicated multispectral and optical cameras, each undertaking scanning flights over the research site to capture vertical ground photos. Rigorous correction of the geometry and optics of the radiation image was performed to obtain a comprehensive image of the entire

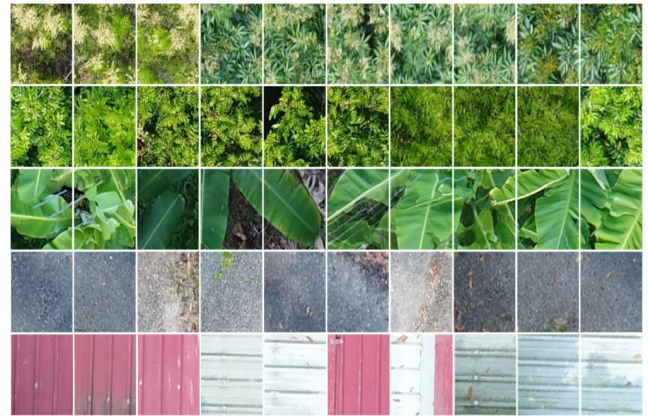


FIGURE 10. Preview of dataset cut from optical image dataset.

research area. Subsequently, with the orthophoto image in hand, a thorough examination and analysis of both optical and multispectral images were conducted.

The experimental phase of this study encompassed distinct stages: data collection and compilation, multispectral image fusion, network model training, and orthophoto image processing. In the preceding section, we elucidated the training steps involved in prior work, with a detailed introduction to the dataset displayed after. The lower limit of the training steps is outlined, providing a comprehensive overview. Appendix A and Figure 11 provide the pseudocode of our proposed approach, and a visual representation of the data processing procedures essential for training.

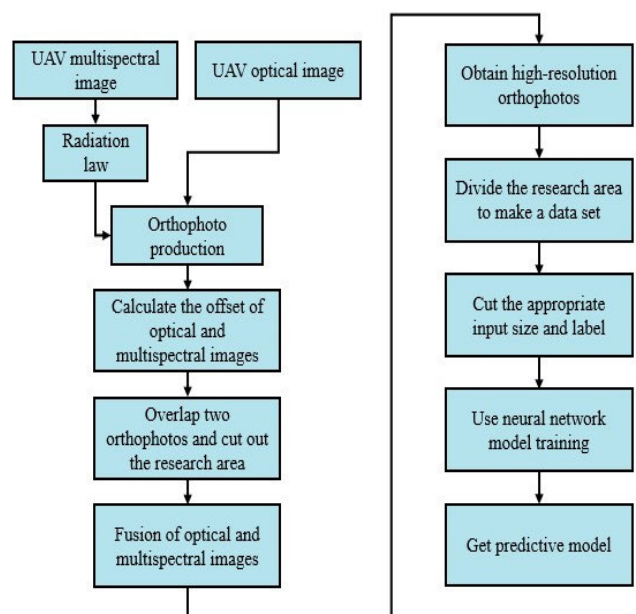


FIGURE 11. UAV image preprocessing and model training.

A. NETWORK MODEL TRAINING

CNNs have demonstrated remarkable success in image classification, with structures like VGG and ResNet

achieving notable results in the ImageNet competition [62]. CNN's strength lies in its multi-layered architecture that autonomously learns features, encompassing various content at different levels. For instance, deeper convolutional layers possess a broader receptive field, facilitating the assimilation of more abstract feature content. Conversely, shallower layers focus on local area features due to their smaller receptive fields. The abstraction of features at different levels enhances the network's ability to handle variations in object direction, size, and position, thereby contributing to improved classification performance.

The primary goal of image classification is to categorize different images into distinct classes and minimize classification errors. Our emphasis is on single-label classification, where each image is assigned to a sole category. Accuracy serves as the evaluation metric for single-label classification, with positive sample data determined based on a probability threshold. If the probability exceeds the threshold, it is considered positive; otherwise, it is negative, with the default threshold set at 0.5. Adjusting the threshold produces varying statistical outcomes.

In single-label classification, each sample pertains to only one definite category. Correct prediction of the category results in accurate classification, while an incorrect prediction indicates a classification error. Accuracy, quantified as $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{TN} + \text{FN})$, serves as the most intuitive indicator. Here, TP and TN denote true positives and true negatives, respectively, while FP and FN represent false positives and false negatives. Accuracy reflects the probability of correctly classifying all samples, and different thresholds yield distinct accuracy values.

The hyperparameters for the VGG-16, VGG-19, and ResNet-50 model architectures are configured for classification tasks involving seven classes. Each model consists of two dense layers, with 4096 nodes each, and concludes with an output layer comprising seven nodes to match the number of classes. These models process images with dimensions of 100×100 pixels, encompassing three color channels (RGB), resulting in an input size of (100, 100, 3). They are trained for 100 epochs with a batch size of 64. For training, the Adam optimizer is employed with a learning rate of 0.00001, beta_1 of 0.9, beta_2 of 0.999, and an epsilon value of $1e-08$. This setup aims to optimize the models by minimizing the loss, measured by the categorical crossentropy function, and to enhance performance as assessed by accuracy metrics.

B. ORTHOPHOTO RECOGNITION

Our approach involves training an identification model specifically designed to recognize orthophoto images of the site. The model is configured to assign distinctive colors to each type of fruit based on identification outcomes, facilitating the differentiation of various fruits within the images.

Upon completing the model training, the orthophoto image undergoes classification, and corresponding colors are applied to the identified fruits. In this experimental context, the colored results are visualized to showcase the diverse

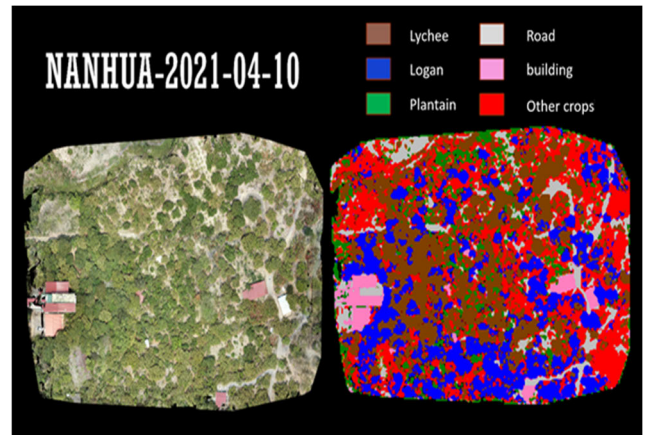


FIGURE 12. Classification results of fruits in orthophoto image of site.

categories of fruits, as illustrated in Figure 12: lychees are depicted in brown, longans in blue, plantains in green, soil and roads in gray, buildings or roofs in pink, and other fruit categories are represented in red. This color-coded representation enhances the interpretability and clarity of the orthophoto image, aiding in the effective distinction of different fruits types.

C. DIFFERENT CHANNEL IMAGE COMBINES OF MULTISPECTRAL IMAGES

This section outlines various comparison processes employed in our study. Initially, we utilized optical images with three distinct neural network models (VGG-16, VGG-19, ResNet-50) for training, aiming to identify the most suitable model for our specific objectives. Subsequently, the selected model was employed to assess and compare the outcomes of different multispectral image fusion methods, enabling the determination of the most effective fusion technique. Finally, the fusion result was utilized to incorporate additional band information from multispectral images, where single-channel near-infrared, red edge, and NDVI (Normalized Difference Vegetation Index) images served as supplementary channel images for further training. This comprehensive approach allows for an in-depth exploration and comparison of diverse processes to enhance the overall performance of our study.

1) OPTICAL IMAGE DATASET TRAINING

Following the augmentation of optical images, we employed three distinct neural network models for dataset training. VGG-16, VGG-19, and ResNet-50, known for their effectiveness in CNNs, were selected to achieve exceptional results. To further enhance identification accuracy, we engaged in deeper network model training.

The training dataset comprised a total of 1,760 images, with an additional 440 images allocated for the testing set. Employing 100 training epochs ensured a comprehensive training process. The accuracy of various models

TABLE 4. UAV image collection and orthophoto production.

Model	VGG-16	VGG-19	ResNet-50
Train/test/total	1,354 / 339 / 1,693		
Epoch	100		
Model size (MB)	590	590	653
Training Cost	4 min	5 min	4 min
Accuracy	0.852	0.831	0.831

TABLE 5. Confusion matrixes for optical image dataset using VGG-16.

Predicted label	0	1	2	3	4	5	6
0	54	7	0	0	0	1	0
1	6	59	1	0	0	5	0
2	0	0	71	2	0	0	0
3	0	1	0	55	0	2	0
4	0	0	1	0	71	0	0
5	5	11	1	9	0	58	0
6	0	0	0	0	0	0	20

was assessed, and as shown in Table 4, the VGG-16 network emerged as the top performer on this specific dataset, showcasing the highest accuracy among the considered models.

In this study, the target objects in the field were categorized into seven classes: 0: lychee, 1: longan, 2: plantain, 3: road, 4: building, 5: other fruits, and 6: black landscape. The confusion matrix in Table 5-7 illustrates each model’s predictions on the expanded test dataset, revealing that different models exhibit strong prediction capabilities in distinct categories. Specifically, VGG-16 demonstrates superior identification performance in categories 1, 2, 4, 5, and 6, remaining acceptable in other categories. On the contrary, the VGG-19 model, prone to gradient loss due to its deep neural network structure, results in excessive information loss during down sampling. Consequently, the results obtained with VGG-16 are deemed more suitable for our requirements. Our evaluation was based on the combination of multispectral image fusion with various RGB imagery channels, all utilizing the VGG-16 model.

2) MULTISPECTRAL IMAGE FUSION DATASET TRAINING

We conducted a comparative analysis of model training accuracy before and after image fusion for the R, G, and B channels of multispectral images. The original multispectral image, when trained directly without preprocessing, yielded the lowest accuracy. Even after applying bicubic interpolation to the original multispectral image, there was

TABLE 6. Confusion matrixes for optical image dataset using VGG-19.

Predicted label	0	1	2	3	4	5	6
0	57	0	5	0	0	0	0
1	13	54	0	0	0	4	0
2	1	0	67	1	0	4	0
3	0	0	2	54	0	2	0
4	0	0	3	1	68	0	0
5	4	5	11	6	0	58	0
6	0	0	0	0	0	0	20

TABLE 7. Confusion matrixes for optical image dataset using ResNet-50.

Predicted label	0	1	2	3	4	5	6
0	60	0	2	0	0	0	0
1	10	57	0	0	0	4	0
2	1	0	69	2	0	1	0
3	0	0	3	54	0	1	0
4	0	0	3	0	69	0	0
5	6	7	11	7	0	53	0
6	0	0	0	0	0	0	20

TABLE 8. The accuracy results for each fusion method.

Model	VGG-16					
Train/Test /Total	1354 / 339 / 1693					
Band Selection	Red+Green+Blue					
Fusion Method	None	Resam-pling	Brovey	HSV	PCA	GS
Accuracy	0.761	0.764	0.758	0.823	0.89	0.825

no significant improvement in accuracy. The Brovey, HSV, PCA, and GS image fusion methods were employed to enhance the quality of the original dataset. The accuracy results for each fusion method are presented in Table 8 (for the VGG-16 model as it was found to be the best performing model; see Table 3), indicating that the latter three fusion methods contribute to improved accuracy compared to the original Multispectral image. Among these, PCA image fusion demonstrated the highest accuracy, followed by GS and then Brovey fusion. Concerning spectral characteristics, Brovey resulted in severely distorted images leading to identification errors, while HSV, PCA, and GS effectively retained spectral information. However, HSV and GS exhibited slight color distortion, resulting in lower accuracy.

TABLE 9. Comparative accuracy of PCA image fusion on different channel.

Training Modules	VGG-16			
Train/ Test /Total	1,760 / 440 / 2,200			
Band Selection	R+G+B	R+G+B+NIR	R+G+B+RED EDGE	R+G+B+NDVI
Accuracy	0.89	0.923	0.921	0.821
Band Selection	R+G+B+NIR+ RED EDGE	R+G+B+NIR+ NDVI	R+G+B+RED EDGE +NDVI	R+G+B+NIR+ RED EDGE +NDVI
Accuracy	0.917	0.851	0.845	0.911

3) DIFFERENT CHANNEL IMAGE COMBINES OF MULTISPECTRAL IMAGES

Our research intended to develop RGB channels to find out which channel has the capability to create high-resolution imagery with greater accuracy. Indeed, various image channels were amalgamated from multispectral images through the utilization of a multispectral camera and optical images that are based on RGB. As it can be seen in Table 9, the extracted channels encompassed red, green, blue, red edge, near-infrared (NIR), and the NDVI images, the latter being computed from red and near-infrared light. The process involved employing PCA image fusion, with the VGG-16 model serving as the training model. The comparison of results was conducted post-training on the fused images derived from different channels. In the course of this experiment, the foundational channels consisted of multispectral red, green, and blue bands, while the supplementary channels comprised red edge, near-infrared light, and NDVI. The ensuing accuracy outcomes are presented in Table 9.

Analysis of the results indicates a marginal reduction in prediction accuracy when incorporating the NDVI band into the data fused by the PCA image. This could potentially be attributed to the diverse representation of objects in NDVI images. Specifically, NDVI images delineate distinctions between green fruits and man-made structures such as buildings and roads (based on light and dark intensities). Nevertheless, they may concurrently obscure texture features among green plants. Given that the identification model in use relies on texture features for object recognition, the incorporation of NDVI images resulted in diminished accuracy.

Conversely, the inclusion of NIR or red edge individually demonstrated an enhancement in model accuracy. This

improvement is likely due to the augmented channel information, which positively influences overall accuracy.

V. DISCUSSION AND CONCLUSION

Our study delves into the intricate challenges prevalent in hillside agriculture, particularly concerning pest management across diverse fruit tree varieties. With traditional agricultural lands becoming increasingly scarce, the utilization of hillside areas for farming has gained prominence. Our research identified three primary challenges: the limited availability of arable land conducive to traditional farming practices, concerns regarding pesticide pollution, and the impracticality of employing ground-based agricultural machinery on sloping terrains. To address these challenges, we proposed an innovative solution in the form of an intelligent UAV imagery system, representing a cutting-edge machinery generation designed to efficiently identify various fruit types and enhance pest control measures. By leveraging CNNs and advanced image processing techniques, we developed a robust classification model capable of extracting distinctive features from economically significant fruit types, thus advancing agricultural applications. Our approach utilizes UAVs equipped with advanced imaging technologies, including optical cameras and multi-channel spectrometers, to navigate the low-altitude environments unique to our experimental site. By capturing images across red, green, blue, near-infrared, and red-edge spectral bands, we aim to compute vegetation indices and texture features that are indicative of specific crop health and pest presence. The integration of high-resolution optical images with multispectral data allows for detailed analysis of tree and fruit types, enhancing our ability to differentiate between species with similar appearances, such as bamboo and lychee trees. Initially, our approach employed a CNN-based strategy, utilizing architectures such as VGG-16, VGG-19, and ResNet-50. Based on our results, VGG-16 emerged as the superior model due to its simplicity and uniformity, particularly in the context of multispectral imagery analysis. Contrary to prior findings, our investigation demonstrated VGG-1's seamless integration with RGB image classification, making it adaptable for multispectral imagery analysis through fine-tuning on multispectral datasets. Moreover, VGG-16's convolutional layers efficiently captured both spatial and spectral features across all bands, outperforming alternative architectures (e.g., [39], [40]).

Subsequently, we utilized PCA, GS, HSV, and Brovey fusion techniques, with PCA yielding the highest accuracy. PCA proved superior for multispectral image fusion, capturing significant data variability effectively and leading to enhanced spatial resolution, improved spectral fidelity, and robust feature extraction. Our findings contradict previous studies, demonstrating the outperformance of PCA over alternative fusion methods [54] indicating that the GS method had been known to be superior in terms of accuracy.

In the final phase, our methodology integrating multispectral imagery, optical camera data, and eight newly developed

RGB channels achieved a high accuracy of 92 percent. This surpasses the accuracy achieved by prior studies [43], which utilized similar data but fell short of the 90 percent threshold. Our research underscores the efficacy of CNN-based models incorporating both RGB and multispectral imagery, offering superior capabilities in discerning intricate patterns and features.

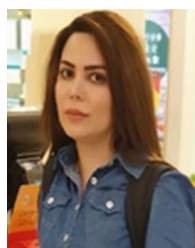
Accordingly, our study introduced the innovative concept of generating eight developed RGB channels to enhance accuracy and identify feature patterns more effectively. By incorporating multispectral image fusion techniques, we succeeded in matching the ground sample distance of multispectral images with that of optical images, resulting in enhanced image resolution and texture characteristics. This optimization, coupled with the integration of promoted RGB channel information, significantly improved accuracy from 0.76 to 0.92 percent.

In addressing the evolving landscape of agricultural technology and responding to the existing complexities, our future work will concentrate on expanding the analytical horizon of UAV imagery. This will involve exploring a broader spectrum of vegetation indices and advancing model training methodologies to accurately tailor channel inputs for diverse fruit types and agricultural scenarios. Emphasizing adaptability, we aim to refine our approach to image fusion and channel optimization, ensuring it remains effective amidst the rapid advancements in dataset complexity and neural network architectures. This forward-looking strategy is not only poised to enhance the accuracy and applicability of UAV imagery in mixed-agriculture terrains but also holds promise for addressing the unique challenges faced by regions with limited arable land, offering a scalable solution to improve pest control and crop monitoring across different geographical and environmental contexts.

REFERENCES

- [1] H.-Y. Lee, M. Hooshyar, C.-J. Lin, W.-S. Wang, and Y.-M. Huang, "Image4Assess: Automatic learning processes recognition using image processing," in *Proc. 38th ACM/SIGAPP Symp. Appl. Comput.*, Mar. 2023, pp. 11–16.
- [2] W. Rahane, H. Dalvi, Y. Magar, A. Kalane, and S. Jondhale, "Lung cancer detection using image processing and machine learning HealthCare," in *Proc. Int. Conf. Current Trends Towards Converging Technol. (ICCTCT)*, Mar. 2018, pp. 1–5.
- [3] K. K. Sarma, K. K. Das, V. Mishra, S. Bhuiya, and D. Kaplun, "Learning aided system for agriculture monitoring designed using image processing and IoT-CNN," *IEEE Access*, vol. 10, pp. 41525–41536, 2022.
- [4] E.-S.-M. El-Kenawy, N. Khodadadi, S. Mirjalili, T. Makarovskikh, M. Abotaleb, F. K. Karim, H. K. Alkahtani, A. A. Abdelhamid, M. M. Eid, T. Horiuchi, A. Ibrahim, and D. S. Khafaga, "Metaheuristic optimization for improving weed detection in wheat images captured by drones," *Mathematics*, vol. 10, no. 23, p. 4421, Nov. 2022.
- [5] I. H. Syeda, M. M. Alam, U. Illahi, and M. M. Su'ud, "Advance control strategies using image processing, UAV and AI in agriculture: A review," *World J. Eng.*, vol. 18, no. 4, pp. 579–589, Jul. 2021.
- [6] M. Hooshyar and Y.-M. Huang, "Meta-heuristic algorithms in UAV path planning optimization: A systematic review (2018–2022)," *Drones*, vol. 7, no. 12, p. 687, Nov. 2023.
- [7] S. Khan, M. Tufail, M. T. Khan, Z. A. Khan, J. Iqbal, and A. Wasim, "Real-time recognition of spraying area for UAV sprayers using a deep learning approach," *PLoS ONE*, vol. 16, no. 4, Apr. 2021, Art. no. e0249436.
- [8] A. Bouguettaya, H. Zarzour, A. Kechida, and A. M. Taberkit, "Deep learning techniques to classify agricultural crops through UAV imagery: A review," *Neural Comput. Appl.*, vol. 34, no. 12, pp. 9511–9536, Jun. 2022.
- [9] S. Khanal, J. Fulton, and S. Shearer, "An overview of current and potential applications of thermal remote sensing in precision agriculture," *Comput. Electron. Agricult.*, vol. 139, pp. 22–32, Jun. 2017.
- [10] T. B. Shahi, C.-Y. Xu, A. Neupane, and W. Guo, "Recent advances in crop disease detection using UAV and deep learning techniques," *Remote Sens.*, vol. 15, no. 9, p. 2450, May 2023.
- [11] P. Bokade, V. K. Gaur, V. Tripathi, S. Bobate, N. Manickam, and A. Bajaj, "Bacterial remediation of pesticide polluted soils: Exploring the feasibility of site restoration," *J. Hazardous Mater.*, vol. 441, Jan. 2023, Art. no. 129906.
- [12] P. Ngin, P. Haglund, S. Proum, and J. Fick, "Pesticide screening of surface water and soil along the Mekong river in Cambodia," *Sci. Total Environ.*, vol. 912, Feb. 2024, Art. no. 169312.
- [13] F. Hüesker and R. Lepenies, "Why does pesticide pollution in water persist?" *Environ. Sci. Policy*, vol. 128, pp. 185–193, Feb. 2022.
- [14] V. Aparicio and E. De Gerónimo, "Pesticide pollution in Argentine drinking water: A call to ensure safe access," *Environ. Challenges*, vol. 14, Jan. 2024, Art. no. 100808.
- [15] K. Neupane and F. Baysal-Gurel, "Automatic identification and monitoring of plant diseases using unmanned aerial vehicles: A review," *Remote Sens.*, vol. 13, no. 19, p. 3841, Sep. 2021.
- [16] D. Lamb, "The use of qualitative airborne multispectral imaging for managing agricultural crops—a case study in south-eastern Australia," *Austral. J. Experim. Agricult.*, vol. 40, no. 5, pp. 725–738, 2000.
- [17] Y. Huang, Y. Lan, and W. Hoffmann, "Use of airborne multi-spectral imagery in pest management systems," *Agricult. Eng. Int., CIGR J.*, vol. X, pp. 1–14, Feb. 2008.
- [18] H. Takano, M. Nakahara, K. Suzuoki, Y. Nakayama, and D. Hisano, "300-meter long-range optical camera communication on RGB-LED-equipped drone and object-detecting camera," *IEEE Access*, vol. 10, pp. 55073–55080, 2022.
- [19] F. Norouzi, S. Olyaei, and M. M. Rad, "Bit error rate improvement in optical camera communication based on RGB LED," in *Proc. 29th Iranian Conf. Electr. Eng. (ICEE)*, May 2021, pp. 47–50.
- [20] C.-J. Chen, Y.-Y. Huang, Y.-S. Li, Y.-C. Chen, C.-Y. Chang, and Y.-M. Huang, "Identification of fruit tree pests with deep learning on embedded drone to achieve accurate pesticide spraying," *IEEE Access*, vol. 9, pp. 21986–21997, 2021.
- [21] S. Zhang, C. Bi, Y. Tan, Y. Luo, Y. Liu, J. Cao, M. Chen, Q. Hao, and X. Tang, "Direct optical lithography enabled multispectral colloidal quantum-dot imagers from ultraviolet to short-wave infrared," *ACS Nano*, vol. 16, no. 11, pp. 18822–18829, Nov. 2022.
- [22] S. Thiruchittampalam, B. P. Banerjee, N. F. Glenn, and S. Raval, "Geotechnical characterisation of coal spoil piles using high-resolution optical and multispectral data: A machine learning approach," *Eng. Geol.*, vol. 329, Feb. 2024, Art. no. 107406.
- [23] G. Matthews, "Droplet sizing—A retrospective view on importance of knowing spray quality," *Outlooks Pest Manage.*, vol. 28, no. 3, pp. 127–130, Jun. 2017.
- [24] Y. Tang, C. J. Hou, S. M. Luo, J. T. Lin, Z. Yang, and W. F. Huang, "Effects of operation height and tree shape on droplet deposition in citrus trees using an unmanned aerial vehicle," *Comput. Electron. Agricult.*, vol. 148, pp. 1–7, May 2018.
- [25] C. Marais Sicre, R. Fieuzal, and F. Baup, "Contribution of multispectral (optical and radar) satellite images to the classification of agricultural surfaces," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 84, Feb. 2020, Art. no. 101972.
- [26] D. C. Tsouros, S. Bibi, and P. G. Sarigiannidis, "A review on UAV-based applications for precision agriculture," *Information*, vol. 10, no. 11, p. 349, Nov. 2019.
- [27] T. Kattenborn, J. Lopatin, M. Förster, A. C. Braun, and F. E. Fassnacht, "UAV data as alternative to field sampling to map woody invasive species based on combined Sentinel-1 and Sentinel-2 data," *Remote Sens. Environ.*, vol. 227, pp. 61–73, Jun. 2019.
- [28] K. Y. Peerbhaya, O. Mutanga, and R. Ismail, "Commercial tree species discrimination using airborne AISA eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA) in KwaZulu-natal, South Africa," *ISPRS J. Photogramm. Remote Sens.*, vol. 79, pp. 19–28, May 2013.

- [29] A. Burkart, V. L. Hecht, T. Kraska, and U. Rascher, "Phenological analysis of unmanned aerial vehicle based time series of barley imagery with high temporal resolution," *Precis. Agricult.*, vol. 19, no. 1, pp. 134–146, Feb. 2018.
- [30] S. H. Lee, C. S. Chan, S. J. Mayo, and P. Remagnino, "How deep learning extracts and learns leaf features for plant classification," *Pattern Recognit.*, vol. 71, pp. 1–13, Nov. 2017.
- [31] B. Ghimire, J. Rogan, and J. Miller, "Contextual land-cover classification: Incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic," *Remote Sens. Lett.*, vol. 1, no. 1, pp. 45–54, Mar. 2010.
- [32] G. Omer, O. Mutanga, E. M. Abdel-Rahman, and E. Adam, "Performance of support vector machines and artificial neural network for mapping endangered tree species using worldview-2 data in Dukuduku forest, South Africa," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 10, pp. 4825–4840, Oct. 2015.
- [33] E. Raczko and B. Zagajewski, "Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images," *Eur. J. Remote Sens.*, vol. 50, no. 1, pp. 144–154, Jan. 2017.
- [34] C. R. Radford, *Best Practices When Using Multi-Rotor Consumer UAVs for Photogrammetric Mapping: Limitations and Possible Solutions*, 2020.
- [35] G.-H. Kwak, C.-W. Park, K.-D. Lee, S.-I. Na, H.-Y. Ahn, and N.-W. Park, "Potential of hybrid CNN-RF model for early crop mapping with limited input data," *Remote Sens.*, vol. 13, no. 9, p. 1629, Apr. 2021.
- [36] R. Zhou, C. Yang, E. Li, X. Cai, J. Yang, and Y. Xia, "Object-based wetland vegetation classification using multi-feature selection of unoccupied aerial vehicle RGB imagery," *Remote Sens.*, vol. 13, no. 23, p. 4910, Dec. 2021.
- [37] C. Feng, W. Zhang, H. Deng, L. Dong, H. Zhang, L. Tang, Y. Zheng, and Z. Zhao, "A combination of OBIA and random forest based on visible UAV remote sensing for accurately extracted information about weeds in areas with different weed densities in farmland," *Remote Sens.*, vol. 15, no. 19, p. 4696, Sep. 2023.
- [38] H. Ye, B. Cui, S. Huang, Y. Dong, W. Huang, A. Guo, Y. Ren, and Y. Jin, "Performance of support vector machines, artificial neural network, and random forest for identifying banana fusarium wilt using UAV-based multi-spectral imagery," in *Proc. 6th China High Resolution Earth Observ. Conf.*, 2020, pp. 261–270.
- [39] E. C. Tetila, B. B. Machado, G. Astolfi, N. A. D. S. Belete, W. P. Amorim, A. R. Roel, and H. Pistori, "Detection and classification of soybean pests using deep learning with UAV images," *Comput. Electron. Agricult.*, vol. 179, Dec. 2020, Art. no. 105836.
- [40] Y. Li, H. Wang, L. M. Dang, A. Sadeghi-Niaraki, and H. Moon, "Crop pest recognition in natural scenes using convolutional neural networks," *Comput. Electron. Agricult.*, vol. 169, Feb. 2020, Art. no. 105174.
- [41] T. B. Shahi, C.-Y. Xu, A. Neupane, and W. Guo, "Machine learning methods for precision agriculture with UAV imagery: A review," *Electron. Res. Arch.*, vol. 30, no. 12, pp. 4277–4317, 2022.
- [42] J. Su, C. Liu, M. Coombes, X. Hu, C. Wang, X. Xu, Q. Li, L. Guo, and W.-H. Chen, "Wheat yellow rust monitoring by learning from multi-spectral UAV aerial imagery," *Comput. Electron. Agricult.*, vol. 155, pp. 157–166, Dec. 2018.
- [43] Z. Song, Z. Zhang, S. Yang, D. Ding, and J. Ning, "Identifying sunflower lodging based on image fusion and deep semantic segmentation with UAV remote sensing imaging," *Comput. Electron. Agricult.*, vol. 179, Dec. 2020, Art. no. 105812.
- [44] L. Dong, Q. Yang, H. Wu, H. Xiao, and M. Xu, "High quality multi-spectral and panchromatic image fusion technologies based on curvelet transform," *Neurocomputing*, vol. 159, pp. 268–274, Jul. 2015.
- [45] F. Liu, G. Li, S. Yang, W. Yan, G. He, and L. Lin, "Detection of heterogeneity on multi-spectral transmission image based on multiple types of pseudo-color maps," *Infr. Phys. Technol.*, vol. 106, May 2020, Art. no. 103285.
- [46] J. B. Hutchings and R. D. McClure, "High-resolution optical imaging of three QSOs," *Publications Astronomical Soc. Pacific*, vol. 102, p. 48, Jan. 1990.
- [47] V. Ferraris, N. Dobigeon, Q. Wei, and M. Chabert, "Detecting changes between optical images of different spatial and spectral resolutions: A fusion-based approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1566–1578, Mar. 2018.
- [48] R. A. Mandhare, P. Upadhyay, and S. Gupta, "Pixel-level image fusion using Brovey transform and wavelet transform," *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.*, vol. 2, no. 6, pp. 2690–2695, 2013.
- [49] H. R. Shahdoosti, "Improved adaptive brovey as a new method for image fusion," 2018, *arXiv:1807.09610*.
- [50] L. Ding and H. Li, "A new improved HSV image fusion method," in *LiDAR Imaging Detection and Target Recognition*, vol. 2017. Bellingham, WA, USA: SPIE, 2017, pp. 426–438.
- [51] P. Li, Y. Huang, and K. Yao, "Multi-algorithm fusion of RGB and HSV color spaces for image enhancement," in *Proc. 37th Chin. Control Conf. (CCC)*, Jul. 2018, pp. 9584–9589.
- [52] J.-L. Chiang, "Knowledge-based principal component analysis for image fusion," *Appl. Math. Inf. Sci.*, vol. 8, no. 1L, pp. 223–230, Apr. 2014.
- [53] S. Yang, M. Wang, and L. Jiao, "Fusion of multispectral and panchromatic images based on support value transform and adaptive principal component analysis," *Inf. Fusion*, vol. 13, no. 3, pp. 177–184, Jul. 2012.
- [54] L. Kumar, P. Sinha, and S. Taylor, "Improving image classification in a complex wetland ecosystem through image fusion techniques," *J. Appl. Remote Sens.*, vol. 8, no. 1, Jun. 2014, Art. no. 083616.
- [55] Y. Kong, F. Hong, H. Leung, and X. Peng, "A fusion method of optical image and SAR image based on dense-UGAN and Gram-Schmidt transformation," *Remote Sens.*, vol. 13, no. 21, p. 4274, Oct. 2021.
- [56] T.-M. Tu, "Adjustable intensity-hue-saturation and brovey transform fusion technique for IKONOS/QuickBird imagery," *Opt. Eng.*, vol. 44, no. 11, Nov. 2005, Art. no. 116201.
- [57] A. Kamble, C. Maisheri, and P. Upadhyay, "HSV, IHS and PCA based image fusion: A review," *Int. J. Eng. Manage. Res.*, vol. 6, no. 1, pp. 164–167, 2016.
- [58] P. Wang, L. Li, and C. Yan, "Image classification by principal component analysis of multi-channel deep feature," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2017, pp. 696–700.
- [59] M. Grasso, B. M. Colosimo, and M. Pacella, "Profile monitoring via sensor fusion: The use of PCA methods for multi-channel data," *Int. J. Prod. Res.*, vol. 52, no. 20, pp. 6110–6135, Oct. 2014.
- [60] R. Bashir, R. Junejo, N. N. Qadri, M. Fleury, and M. Y. Qadri, "SWT and PCA image fusion methods for multi-modal imagery," *Multimedia Tools Appl.*, vol. 78, no. 2, pp. 1235–1263, Jan. 2019.
- [61] H. Dibs, H. Sabah Jaber, and N. Al-Ansari, "Multi-fusion algorithms for detecting land surface pattern changes using multi-high spatial resolution images and remote sensing analysis," *Emerg. Sci. J.*, vol. 7, no. 4, pp. 1215–1231, Jul. 2023.
- [62] D. Han, Q. Liu, and W. Fan, "A new image classification method using CNN transfer learning and Web data augmentation," *Expert Syst. Appl.*, vol. 95, pp. 43–56, Apr. 2018.



MARAL HOOSHAYAR received the bachelor's and master's degrees in industrial engineering. She is currently pursuing the Ph.D. degree in engineering science with National Cheng Kung University, Taiwan. Her research interests include strategy, optimization of supply chains and logistics, optimization of UAV path planning, designing mathematical models, and reliability.



YUAN-SHUO LI was born in Tainan, Taiwan, in 1996. He received the B.S. degree in computer science from National Pingtung University, Pingtung, in 2018. He is currently pursuing the master's degree with the Engineering Science Department, National Cheng Kung University, Tainan. His research interests include remote sensing imagery in multispectral, application of embedded systems, and LoRa wireless transmission technology, applications of machine learning in aerial imagery, image processing, machine learning applied to computer vision, and image processing in robotics applications.



WEN CHUN TANG received the bachelor's degree in electrical engineering. He is currently pursuing the master's degree with National Cheng Kung University. His research interests include deep learning, generative AI, digital chip design, and the application of digital technology in education



LING-WEI CHEN received the B.S. degree from the Department of Biomedical Engineering and Environmental Sciences, National Tsing Hua University, Taiwan, in 2018. She is currently pursuing the master's degree with the Department of Engineering Science, National Cheng Kung University, Taiwan. She is working on machine learning applied to computer vision and image processing in robotics applications. Her research interests include machine learning in aerial imagery and image processing.



YUEH-MIN HUANG (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from The University of Arizona, in 1988 and 1991, respectively. He is currently a Chair Professor with the Department of Engineering Science and Institute of Education, National Cheng Kung University, Taiwan. He has completed over 60 Ph.D. and 300 M.S. thesis students. He has coauthored three books and has published more than 280 refereed journal research articles.

His research interests include e-learning, multimedia communications, and artificial intelligence. He is a fellow of the British Computer Society, in 2011. He is the Founding Chair of the International Symposium of Emerging Technologies for Education (SETE) and the International Conference of Innovative Technologies and Learning (ICITL). He has received many research awards, such as Taiwan's National Outstanding Research Award, in 2011/2014, and the 2017 Taiwan Outstanding IT Elite Award. He is on the editorial board of several international journals in the areas of educational technology, computer communications, and web intelligence.

• • •