

Research Article

Energy-Efficient Unmanned Aerial Vehicle (UAV) Surveillance Utilizing Artificial Intelligence (AI)

Hai T. Do,¹ Linh H. Truong,² Minh T. Nguyen ,¹ Chen-Fu Chien ,² Hoang T. Tran,^{3,4} Hoang T. Hua,¹ Cuong V. Nguyen,⁵ Hoa T. T. Nguyen,¹ and Nga T. T. Nguyen¹

¹Thai Nguyen University of Technology, Thai Nguyen 240000, Vietnam

²National Tsing Hua University, Hsinchu 30013, Taiwan

³Faculty of Electrical-Electronic Engineering, Duy Tan University, Danang, 550000, Vietnam

⁴Center of Electrical Engineering, Duy Tan University, Danang, 550000, Vietnam

⁵Thai Nguyen University of Information and Communication Technology, Thai Nguyen 240000, Vietnam

Correspondence should be addressed to Minh T. Nguyen; nguyentuanminh@tnut.edu.vn

Received 11 June 2021; Accepted 8 September 2021; Published 13 October 2021

Academic Editor: Rehmat Ullah

Copyright © 2021 Hai T. Do et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, unmanned aerial vehicles (UAVs) enhance connectivity and accessibility for civilian and military applications. A group of UAVs with on-board cameras usually monitors or collects information about designated areas. The UAVs can build a distributed network to share/exchange and to process collected sensing data before sending to a data processing center. A huge data transmission among them may cause latency and high-energy consumption. This paper deploys artificial intelligent (AI) techniques to process the video data streaming among the UAVs. Thus, each distributed UAV only needs to send a certain required information to each other. Each UAV processes data utilizing AI and only sends the data that matters to the others. The UAVs, formed as a connected network, communicate within a short communication range and share their own data to each other. Convolution neural network (CNN) technique extracts feature from images automatically that the UAVs only send the moving objects instead of the whole frames. This significantly reduces redundant information for either each UAV or the whole network and saves a huge energy consumption for the network. The UAVs can also save energy for their motion in the sensing field. In addition, a flocking control algorithm is deployed to lead the group of UAVs in the working fields and to avoid obstacles if needed. Simulation and experimental results are provided to verify the proposed algorithms in either AI-based data processing or controlling the UAVs. The results show promising points to save energy for the networks.

1. Introduction

Autonomous UAV networks have been deployed in many applications in both military and civil fields. With the ability to handle large data as well as maneuverability, UAVs are capable of completing a wide range of applications such as oil and gas facilities for security [1], surveillance [2], emergency response, and seaport [3]. They are dynamic and effective for sensing and monitoring surveillance purposes [4], and especially, they can be the core technology in Internet of Things vision in which the distributed UAVs can collect sensing data and exchange the data to each other [5, 6].

An UAV network consists of sensing devices, control algorithms, and communications. UAVs in the network

cooperatively work together to complete specific missions. Each UAV can obtain visual sensing data by its equipped camera. The sensing data is then exchanged throughout the network for mission purposes. There are two main structures of information sharing: centralized and distributed [7]. In centralized networks, a central processor performs all tasks include collecting, computing, and delivering commands to other nodes in a network. The centralized scheme has a single point failure of the central processor, and other nodes must maintain a connection with the central node. In distributed networks, information is exchanged between nodes, and the computation and decision-making strategies are performed in each UAV itself. Usually, UAV networks operate in a distributed fashion to improve robustness and

reduce communication burden as an UAV only needs to connect with its neighbors. Besides information sharing, another consideration is control algorithms for multiple UAV formations. In UAV-based surveillance systems, UAVs have to encounter numerous obstacles because they normally operate at low altitudes in urban environments due to policy restrictions [8].

Control algorithms should be able to drive an UAV formation to targeted areas without collisions with obstacles as well as other UAVs. In [9], a control algorithm for a team of micro-UAVs based on a leader-follower approach was proposed. The above-proposed methods have shown a good performance in terms of formation shape keeping and smooth maneuvering. However, the obstacle avoidance has not been considered. The Artificial Potential Field (APF) method has been investigated to deal with obstacle avoidance problems [10, 11]. In the papers, the impulsive and attractive forces are generated by the potential field for an agent to avoid collision and remain the desired distance in a formation. However, the APF method possesses limitations due to local minimal problems. At these points, the total force due to attractive and repulsive forces is zero, which prevents the UAV to reach targets. In addition, APF methods have shown poor performance in handling obstacles that have convex and concave shapes [12]. Another powerful approach for controlling swarm robots is flocking control which was first proposed by Olfati-Saber [13]. In flocking control, agents in a group only need to keep a certain distance from their neighbors, which is different from formation control algorithms where agents maintain a rigid position respecting their neighbors. Flocking control algorithms allow the formation to effectively change formation shapes when encountering obstacles. This feature makes flocking algorithms become suitable approaches for UAV-based surveillance systems.

UAV networks have been providing the most successful application for surveillance systems. An UAV-based platform for drought mapping of agricultural crops is presented in [14]. In [15], multiple UAVs are used to monitor and detect traffic congestions. A framework for wildfire monitoring based on multiple UAV system is developed in [16]. Surveillance tasks often require rapid ability to monitor multiple interested points. As UAVs operate in aerial environments, they have a broader vision and encounter few obstacles than other kinds of robots. These features make UAVs become appropriate approaches in surveillance systems.

The intelligent surveillance system (ISS) is a surveillance system with strong data analysis capabilities. An ISS can not only detect or track objects but also analyze data to anticipate behaviors of objects or upcoming events. These kinds of work have been done with minimal intervention from human. Numerous applications of ISS can be found in literature like traffic monitoring [17, 18] or home security [19]. The ISS is a modern technology that makes use of knowledge from various technical fields such as sensing devices, communications, signal processing, and artificial intelligence (AI) [20]. However, due to a large number of cameras deployed in practical surveillance systems, the collected sensing data from the cameras are also large. This leads to

numerous issues in terms of system accuracy, time, data complexity, etc.

The development of AI technologies has been rapidly increased in recent years. In [21], motion information is combined with a convolution neural network (CNN) to classify and to track a crowd of people. Sultani et al. [22] develop specific classification models to recognize events and correctly identify various activities of human. In [23], the authors propose a knowledge representation framework for describing patterns in video sequences. The proposed framework has shown more advantages in the ability to rapidly detect objects on screen compared to deep learning techniques. AI techniques have been also used in managing network traffic. Ant Colony Optimization (ACO) is applied to improve the performance of software-defined networks (SDN). The quality of experience of SDN increased 24.1% by applying ACO on the weight graph of the SDN controller. Most AI algorithms usually require powerful hardware to process a huge amount of data. This feature limits applications of advanced AI-based signal processing algorithms in practice.

The hardware constraints are more strict in UAV network surveillance systems. An UAV can only bring a finite amount of batteries. Equipping more onboard processing devices will increase the weight of UAVs that reduces their operating time. Commercial UAVs can operate within 20-40 minutes per charged cycle [4]. Most of the energy consumed comes from propulsion [24], which can be solved by optimizing total flight time in case of data collection and analysis tasks in wireless sensor networks application on a single charge [25]. In surveillance applications, the UAVs often perform tasks at a certain altitude and position until the energy runs out, optimizing the flight time may not be appropriate. The monitoring or sensing data could be images or videos that may cost a big amount of memory storage in each UAV. This also consumes a lot of energy consumption in case of UAVs transmitting data to server sides or between UAVs via wireless data transmission. As mentioned in [26], the power consumption of wireless data transmission is proportional with the package size; thus, the smaller the size of transmitted data, the smaller the energy consumption.

As shown in Figure 1, in surveillance application, each UAV monitors a certain area. Data from the UAVs can be exchanged between neighboring UAVs. The data collection in the form of a video format of UAVs may cost a big amount of memory storage in each UAV and also the transmission bandwidth. In addition, while performing surveillance, the UAVs often fly in a fixed position; hence, most of the scenes do not change over time, and only moving objects are noticeable. The transmission of redundant data such as background frame and overlapped area is a waste of resources [27]; however, further analysis tasks are only concerned with moving objects.

In this work, a framework for high-energy-efficient UAV surveillance networks is proposed. A group of UAVs is deployed to cover an area that needs to be observed. A flocking algorithm is used to drive a group of UAVs moving to sensing areas. The algorithm guarantees that the UAV team

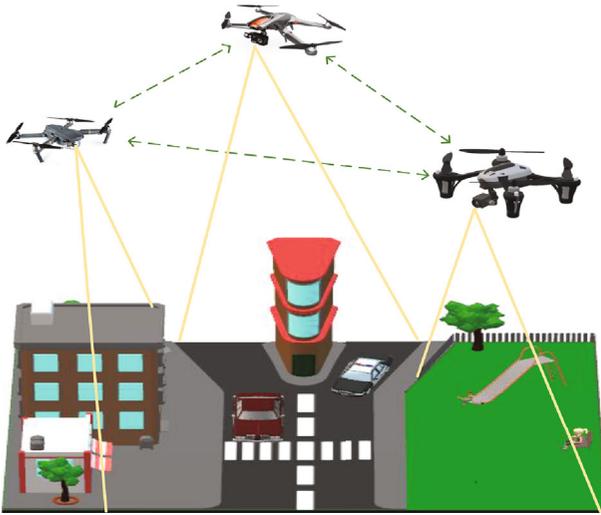


FIGURE 1: Illustration of a distributed UAV network monitoring a sensing area. The UAVs can cover a large area for surveillance purposes.

can safely travel to required locations and forms an appropriate shape to cover the sensing areas. Then, an AI-based method is proposed with the aim of reducing redundant data for the UAVs while performing surveillance tasks of collecting data. The data processing algorithm can be divided into three main steps: (i) background modeling which removes all moving objects in scenes and background stitching that combines the background modeled from each UAV, (ii) noticed object extraction of each frame captured by UAVs, and (iii) data reconstruction of combined background modeling in step (i) and noticed objects from step (ii). The methodology can be referred as a kind of compress sensing technique which is aimed at saving power consumption by removing such redundant data of sensors [28, 29].

The rest of this paper is organized as follows: Section 2 provides briefly the system models that describe either the UAV network deploying in the sensing field or the AI-based methods to process the surveillance data collected from the UAVs. In Section 3, the whole problems are addressed. The flocking control algorithm and the AI-based data processing method are provided in detail. Section 4 presents both simulation and experimental results following all the steps modeled in Section 3. Finally, conclusions and future research directions are provided in Section 5.

2. System Model

In this section, the system models are presented. First is a model of an UAV network with the ability to travel, to avoid obstacles, and to collect video streaming data. Each distributed UAV can also be able to exchange the data with neighbors to construct the completed information of sensing regions. The AI-based data processing framework for enhancing an energy-efficient approach of a UAV network is analyzed.

2.1. Multiple UAV Systems. Considering a team of UAVs, the team is deployed in a ground center. After receiving a mis-

sion request task, the UAV team will move to a target location. The target location is defined as a virtual leader to be able to lead the UAVs in a flocking control algorithm. The collaborative algorithm in [30] is chosen to drive the UAV team. The UAV formation can safely reconfigure formation shapes to avoid collisions with obstacles while migrating. When UAV team arrives at the target place, the team gradually forms quasilattice formation to fully cover sensing areas.

Assuming each UAV can obtain its global position by sensors such as GPS. A downward camera is mounted on an UAV, which provides each UAV a constant sensing range of RS. UAVs are equipped with short-range wireless communication devices that allow them to wirelessly communicate with the others if the Euclidean distance between them is smaller than a constant R_c , noted as the communication range. Different from [30], the sensing range R_s is not required to be smaller than the communication range R_c for ensuring nonoverlapped regions. In this work, overlapped regions are acceptable to guarantee coverage performance. While processing, the overlapped data is handled by an AI-based data processing algorithm, which is proposed as follows.

2.2. AI-Based Data Processing Method. The structure of the UAV system is given in Figure 1. An UAV monitors a distinct area, and the areas handled by different UAVs might be overlapped with each other. UAVs form a distributed network and share their local sensing information with others to reconstruct global sensing data.

In the first step, background modeling is performed on the UAVs. Captured videos are processed to create backgrounds that consist of only nonmoving objects. Then, the backgrounds are sent to the neighbors at the begging and only updated as there are any changes in backgrounds. The individual backgrounds are then stitched together to form a complete background of the sensing area. In the case of overlapped images, an overlapping detection algorithm is presented. In detecting keypoints and local invariant descriptors, then matching descriptors of overlapping images, a random sample consensus (RANSAC) algorithm is utilized to obtain homography. The obtained homography matrix is then used to warp and stitch overlapped pictures.

Secondly, UAVs perform object extraction functions where moving objects are detected by comparing differences in the continuous sequence of frames. If there are motions which is detected, details of moving objects are determined by a convolution neural network (CNN). These useful data are also shared among UAV networks.

Finally, the reconstructed images are built based on extracted data sent by other UAVs. Reconstructed processes can be performed on an UAV. As sensing data are reduced by the proposed method, a burden on transmission bandwidth and computational resources is greatly diminished.

3. Problem Formulation

This section presents an overview of approaches to the problems in multiple UAV-based surveillance systems. First, the

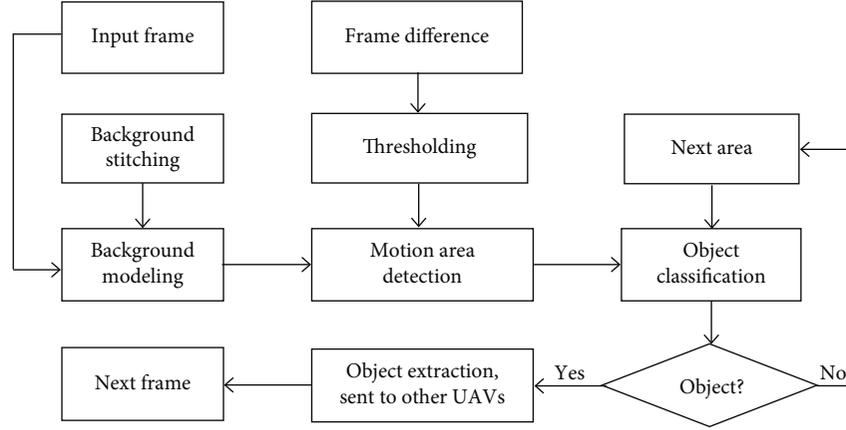


FIGURE 2: The flowchart of AI-based data processing method for UAVs that provides all the steps to reduce data for energy efficiency.

Input: Background images with overlapped areas.
 Output: Stitched background image without overlapped areas.

1. Compute the sift-key points and descriptors for images.
2. Compute distances between every descriptor in one image and every descriptor in the other image.
3. Select the top best matches for each descriptor of an image.
4. Run RANSAC to estimate homography.
5. Warp to align for stitching.
6. Finally stitch them together.

ALGORITHM 1: Overlapped area detection.

flocking algorithm to drive the UAV formations to navigate sensing areas is presented. Next, the AI-based data processing method is given. Three steps as shown in Figure 2 are presented in details.

3.1. Flocking Control for Multi-UAV Systems. In this section, the flocking algorithm for controlling a formation of UAV is presented. The network of a N-UAV team is modeled by graph $G(V, E)$, where the vertex set $V = \{1, 2, \dots, N\}$ represents UAV members. The edge set $E = \{(i, j) : i, j \in V, i \neq j\}$ is the communication link between two UAVs. Let $p_i, v_i \in \mathbb{R}^2$ denote the position and velocity vectors of the i^{th} UAV ($i = 1, 2, \dots, N$). The dynamics of an UAV is described by double integrators model as

$$\begin{cases} \dot{p}_i = v_i, \\ \dot{v}_i = u_i, \end{cases} \quad (1)$$

$(i = 1, 2, \dots, N),$

where u_i is the control input vector for the UAV i^{th} . Equation (1) can be used to model distributed UAVs having omnidirectional motion capacity.

Consider each UAV has a communication range R_c . The number of neighbors of the i^{th} UAV is defined by

$$N_i^a(t) = \left\{ j \in V : \|p_j - p_i\| \leq R_c, i \neq j \right\}, \quad (2)$$

where $\|p_j - p_i\|$ is the Euclidean distance. The superscript indicates the actual neighbors of UAV i^{th} . To provide the obstacle avoidance ability, the term “virtual neighbors” is introduced. The virtual neighbors of UAV i^{th} are defined as

$$N_i^v(t) = \{k \in V_o : \|p_{ik} - p_i\| \leq r_o, V_o = \{1, 2, \dots, K\}\}, \quad (3)$$

where r_o is the obstacle detection range, V_o is the set of obstacles, and p_{ik} is the position of UAV i^{th} projecting on the k^{th} obstacles. The virtual neighbors are utilized to generate the repulsive forces to avoid the collision between UAVs and obstacles.

A team of UAVs forms a certain formation structure to navigate a large sensing field. Each UAV must avoid collisions with other members as well as obstacles. The distributed flocking algorithm consists of three components, namely, formation control f_i^f , obstacle avoidance f_i^o , and navigation f_i^n , which is given as

$$u_i = f_i^f + f_i^o + f_i^n. \quad (4)$$

The first term f_i^f is the formation control term to generate the attractive and repulsive forces for UAV members to form a formation. This term is also used to regulate the velocity matching of UAVs in group. The term is designed as [21]

Input: Two consecutive frames, t and $t+k$, of the input video.
Output: Motion area in each frame.

1. The video is segmented into frames, (consider each frame as an image).
2. Two images / frames (A & B) that is background modeled and frame t
3. Converting the images A & B into gray scale.
4. Computing a difference between these two gray scale images.
5. If, significant difference is detected between these A & B frames, it can conclude that some motion has occurred.

ALGORITHM 2: Motion area detection.

TABLE 1: A custom CNN object classification architecture.

Stage	Object classification architecture
1	Resized image $W \times H \times 3$
2	Convolution layer: $W \times H \times 64, F = 3 \times 3, S = 1$ Max-pooling layer: $W/2 \times H/2 \times 64, F = 2 \times 2, S = 2$ Dropout layer: dropout rate = 0.1
3	Flatten layer
4	Fully connected layer: dense (128)
5	Fully connected layer: dense (1)
6	Active function: sigmoid function

$$f_i^f = c_1^a \sum_{j \in N_i^a} \phi_a(\|p_{ik} - p_i\|_\sigma) n_{ij} + c_2^a \sum_{j \in N_i^a} a_{ij}(p)(v_j - v_i), \quad (5)$$

where c_1^a, c_2^a are positive constants, ϕ_a is the action function [19], n_{ij} is the vector connecting p_i and p_j , and the $\|\cdot\|_\sigma$ is defined as $\|x\|_\sigma = (1/\epsilon)[\sqrt{1 + \epsilon\|x\|^2} - 1]$ with constant $\epsilon > 0$. The $\|\cdot\|_\sigma$ is differential everywhere and is utilized to construct smooth collective potentials.

The second term f_i^o is introduced to prevent UAVs from collisions with obstacles in sensing environments. This term is designed by

$$f_i^o = c_1^o \sum_{j \in N_i^o} \phi_o(\|p_{ik} - p_j\|_\sigma) n_{ik} + c_2^o \sum_{j \in N_i^o} b_{ik}(p)(v_{ik} - v_j), \quad (6)$$

where $\phi_o(\cdot)$ is the repulsive action function [19] and n_{ik} is the vector along the line connecting p_{ik} and p_i . The adjacency matrices $A = [a_{ij}]$ and $B = [b_{ik}]$ are defined by a graph [31].

The UAVs may be deployed in a ground center, and they have to travel to certain places depending on sensing missions. The last component, navigation term, is introduced to provide navigating abilities for UAV formations. The component is given as

$$f_i^n = -c_1(p_i - p_t) - c_2(\bar{p}_{(N_i^a \cup \{i\})} - p_t), \quad (7)$$

where c_1, c_2 are the positive constants, $\bar{p}_{(N_i^a \cup \{i\})}$ is defined as $\bar{p}_{(N_i^a \cup \{i\})} = (1/|N_i^a \cup \{i\}|) \sum_{i=1}^{|N_i^a \cup \{i\}|} p_i$, and $|N_i^a \cup \{i\}|$ are the neighbors of UAV i^{th} including itself. The target location is p_t where the UAV formations have to navigate.

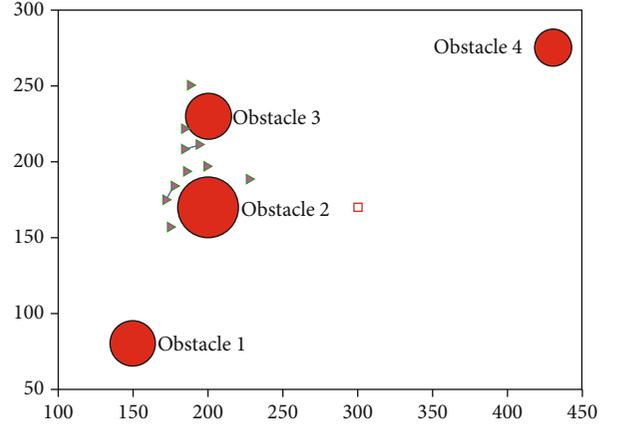


FIGURE 3: UAV formation maneuvers through obstacles in working environments.

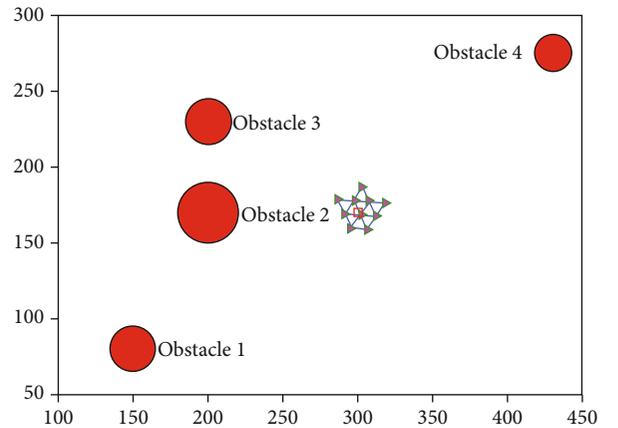


FIGURE 4: The UAV formation creates a quasilattice shape when reaching a desired location.

3.2. Background Modeling. In the background modeling process, there are two sub-steps: background modeling and background stitching captured by different UAVs.

The median filter technique is used to perform background modeling. The main idea of the median filter is to run through the signal entry by entry, replacing each entry with the median of neighboring entries. In this work, thanks to the idea of median filter technique, a number of frames are chosen randomly from the video captured by UAVs and background modeling tasks are performed. The number of frames chosen may vary through experiments to choose an appropriate number of frames to achieve better results of performance.

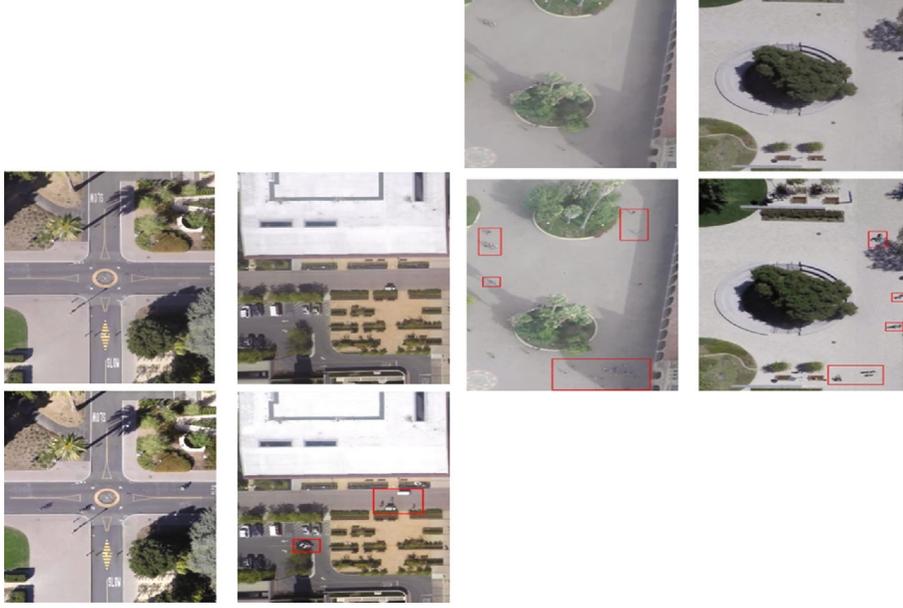


FIGURE 5: Background modelling and corresponding raw frames.

After UAVs perform their background modeling processes, the significant data will be sent to neighbors for reconstruction. However, in case of overlapped areas monitored by UAVs, it is not necessary to stitch those areas. In order to do that, the algorithm is called overlapped area detection in which the key points and its corresponding between backgrounds perform the background stitching. Algorithm 1 represents the overlapped area detection steps as follows.

3.3. Noticed Object Extraction. Object detection task in aerial images is a challenging and interesting problem. With the cost of drones or UAVs decreasing, more aerial devices could be deployed. Hence, there is a surge in the amount of aerial data being generated. It will be very useful to have models that can extract valuable information from aerial data. However, since most objects are only a few pixels wide, some objects are occluded and objects in shade are even harder to detect. Thus, a hybrid noticed object extraction system that is a combination of existing method and custom object classification model to extract a valuable information from aerial data is proposed in this work.

Firstly, frame difference and thresholding technique are applied in each frame to estimate the moving areas that can be referred as the noticed area. For surveillance tasks with UAVs, the objects are often very small, so that directly applying object detection algorithms can result in missing or incorrectly detected objects. Therefore, in this paper, the first step is to determine the motion area by comparing frame t with frame $t+k$, for $k=1, 2, 3, \dots$ to find the difference and thus is motion area as shown in Algorithm 2.

After that, for those areas, a custom convolution neural network (CNN) object classification model is built to perform the classified task whether the area contain objects. If the area contains objects that are predefined in the training

model process, object tracking task is executed and extracted the significant data.

Because of the difference of image size of training data, at the first step, all the images used for training classified model are resized into squares in size. The choice of image size that affects the performance of the model can be done through experiments. Because the training images are extracted from the dataset, they are small in size, so a lightweight classified model is built. From the resized images of size $W \times H \times 3$, W is the width, H is the height, and 3 is the number of channel layer. The first one is the convolution layer with 64 feature maps with 3×3 kernel size and the stride of 1. After that, these feature maps are downsampled by 2×2 max-pooling layer and at the end of stage 1 are downsampled by a dropout layer as shown in Table 1.

After classifying the areas with objects, the object tracking algorithm is implemented to track and extract the information of the objects towards the central UAVs or between the UAVs.

3.4. Data Reconstruction. The data reconstruction can be performed between UAVs. Objects that are extracted from previous UAVs will be sent to the central UAVs or nearby UAVs along with their locations in the frame. Based on the information received, UAVs can process and perform data reconstruction. The most important point is that each distributed UAV can achieve the whole data collected from the network with a minimal storage. This shows the effectiveness of the proposed method.

4. Simulation and Experimental Results

In this section, the proposed control algorithm known as the flocking algorithm is simulated with 10 UAVs. Then, the experimental results show all the steps mentioned in Section 3. The experiment results are performed by using Python 3.6



FIGURE 6: Results of background captured by different UAVs are stitched.

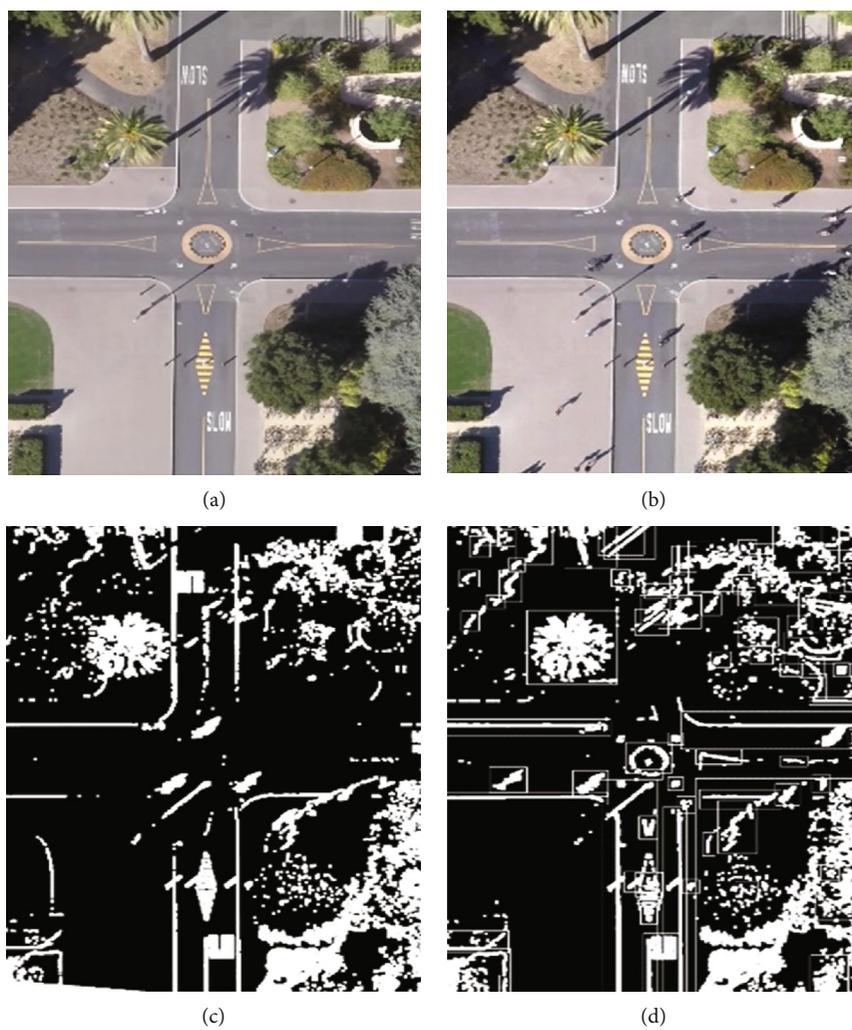


FIGURE 7: Moving area estimation: (a) background, (b) raw frame, (c) different frame, and (d) motion area detection.

with TensorFlow and Keras library on Intel Xeon E3, 16GB RAM, Windows 10 platform. The dataset is selected from Stanford Drone Dataset [32].

4.1. Simulation Results with Flocking Algorithm. In this section, a group of 10 UAVs is deployed to perform surveillance tasks in a vast area. We define a unit to calculate Euclidean distances among UAV and obstacles in the sensing field. Each UAV has a constant communication range $R_c = 10$ units. A virtual leader represents a target location for the group. The group is led by a control algorithm as mentioned in Equation (4).

In Figure 3, as the UAVs in the group encounter obstacles, they separate and move around obstacles. Connections among members may be interrupted without affecting control performance. When the group reaches the target, UAV members gradually reconnect to form a quasispherical lattice shape. The shape covers an area around a target point. Figure 4 illustrates the quasispherical lattices shape formed by the UAV group. All 10 UAVs will stay around the virtual leader and start their mission as surveillance tasks. Based on a fixed communication range R_c , they keep their connections to their neighbors as a grid network as shown in the figure.

4.2. Stanford Drone Dataset. Stanford Drone Dataset is a massive dataset of aerial images collected by drone over the Stanford Campus. The dataset is ideal for object detection and tracking problems. It contains about 60 aerial videos. For each video, we have bounding box coordinates for the 6 classes as ‘‘Pedestrian,’’ ‘‘Biker,’’ ‘‘Skateboarder,’’ ‘‘Cart,’’ ‘‘Car,’’ and ‘‘Bus’’. The dataset is very rich in pedestrians and bikers with these 2 classes covering about 85%-95% of the annotations.

4.3. Background Modeling. By using multiple frames received from UAV surveillance, the backgrounds are modelled by using a median filter technique. The number of frames needed to perform the background can be selected depending on the scene and location of the monitoring task. In this paper, because the UAVs perform the task of supervising the school staff, the traffic volume of vehicles and people is low. Therefore, 20 frames were randomly selected to perform the background as shown in Figure 5.

After the background from different scenes is captured and sent by UAVs to other UAVs, in case of overlap, the overlapping areas will be handled by the algorithm presented in the previous section as shown in Figure 6.

4.4. Noticed Object Extraction

4.4.1. Moving Area Estimation. After background modelling, moving area detection can be done through the background itself and following frames that may contain moving objects. The different area based on the predefined threshold will transform into $[0, 1]$. By using this technique instead of the traditional differencing frame method, the result is better as shown in Figure 7.

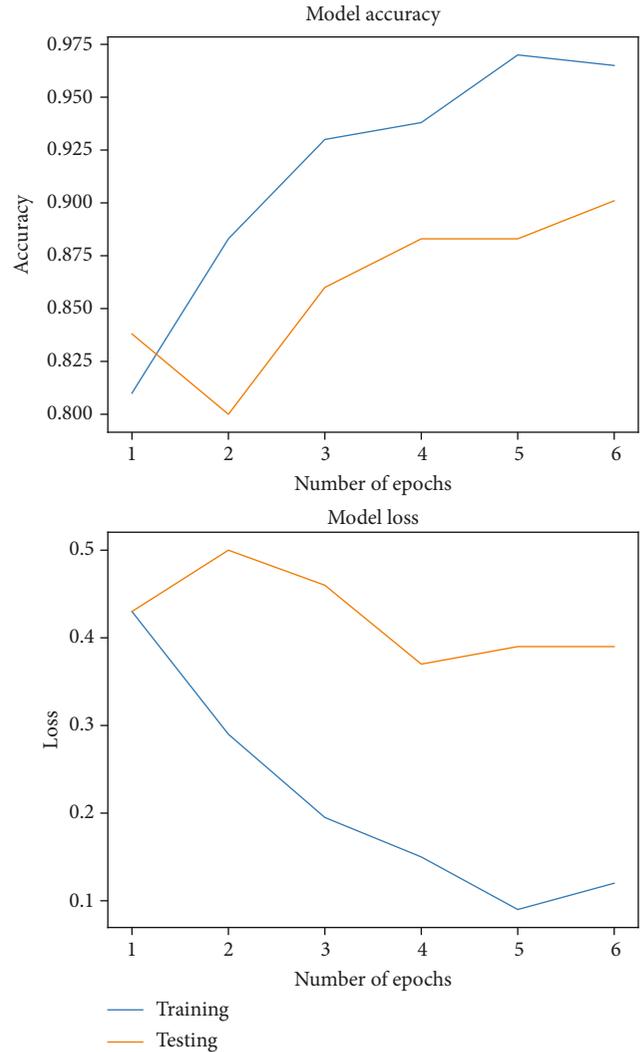


FIGURE 8: Training accuracy and loss of custom CNN object classification.

TABLE 2: Performance measurement of custom CNN object classification model.

	Precision	Recall	F1-score
Object	0.49	0.52	0.51
Noise	0.52	0.49	0.51

4.4.2. Object Classification Training Performance. Once the motion areas are detected, the object classification model classifies those areas into different predefined classes. The object classification model trained by the Stanford Drone Dataset includes objects such as vehicles, pedestrians, and cyclists. The model consists of the stages and layers shown in Table 1. After many experiments select the best hyperparameter, the model classifies the object for the best performance with the number of epochs = 6 and steps per epoch = 1000. The training accuracy at epoch = 6 is 0.965, training loss = 0.11, testing accuracy = 0.90, and testing loss = 0.39, as shown in Figure 8.



FIGURE 9: Nonobjects and objects extracted from Stanford Drone Dataset.

To evaluate the performance of the custom CNN object classification model, F1-score, accuracy, and recall metrics are used, as shown in Table 2. Due to the smaller size of the object of dataset, the performance of the proposed network could be enhanced in future improvement. In contrast, the proposed network shows a very good computational efficiency with a smaller number of operations in both training and classifying stages.

4.4.3. Moving Object Tracking and Extraction. The model object classification can be classified into two types of objects or nonobjects. If the moving area is an object, the object tracking algorithm will be implemented, and the system will be ignored. Figure 9 shows some images extracted from a categorized frame including object and nonobject with a different size.

4.4.4. Data Reconstruction. After moving objects are tracked and extracted, based on position information, x, y, w, h , where x, y are the starting location and w, h is the width and height of the size of rectangle bounded region of interest, are achieved. To measure the effectiveness of the proposed method, the videos in the Stanford Drone Dataset are split into frames and compared to the total capacity of the background and the moving object extracted. The videos in the Stanford Drone Dataset have 1422×1945 , 30 fps, and data rate = 50890 kbps. 300 frames are taken from any video with a capacity of 138 MB. After processing, the total

remaining capacity is approximately 14 MB. Thus, the percentage reduction is 90% but still ensures the quality of the image. Indeed, the proportion of objects in each frame is extremely small compared to that of whole frames. If we can eliminate most of the unnecessary data, this can significantly reduce a huge amount of data. The results either save energy for data transmission among UAVs or save storage capacity for UAVs.

Overall, the distributed network of UAVs can significantly reduce a huge data transmission for the video surveillance purposes based on the AI data processing methods. In addition, the flocking control algorithm also helps the UAVs working in the fields that are suitable to the working tasks. The energy-efficient approach is presented and solved completely.

5. Conclusions and Future Developments

This paper proposes new methods either to control multiple UAVs or to process video surveillance data based on AI techniques with CNN. The flocking control algorithms are applied into distributed UAVs to lead the UAVs travelling on the working fields and avoiding collision and obstacles. The AI-based data processing method that reduces significant redundant data streaming among UAVs is proposed. The method also reduces the training time and classification time compared to existing methods, such as YOLO detection. The overall proposed methods help reducing the

storage capacity, transmission bandwidth, and performance in surveillance application of UAVs. Indeed, the proportion of objects in each frame is extremely small, and the transmission of redundancy in each frame is not necessary. The application of the method helps to reduce approximately 90% of the excess data capacity but still ensures the quality of the image. This significantly reduces the energy consumption for UAVs in their tasks.

Future research can be done to enhance the proposed solution. In order to improve the system performance, the crucial process is an object classification task that will classify the wrong area detected from the previous step, thereby improving the efficiency of the method. Moreover, when applied to more complex applications such as traffic surveillance and agriculture, more types of object should be considered.

Data Availability

All the data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to thank Thai Nguyen University of Technology (TNU) and Ministry of Education and Training (MOET), Vietnam, for the support.

References

- [1] J. Cho, G. Lim, T. Biobaku, S. Kim, and H. Parsaei, "Safety and security management with unmanned aerial vehicle (UAV) in oil and gas industry," *Procedia Manufacturing*, vol. 3, pp. 1343–1349, 2015.
- [2] Y. Noh, D. Koo, Y. Kang, D. Park, and D. Lee, "Automatic crack detection on concrete images using segmentation via fuzzy C-means clustering," in *2017 International Conference on Applied System Innovation (ICASI)*, pp. 877–880, Sapporo, 2017.
- [3] F. M. Raimondi, M. Trapanese, V. Franzitta, A. Viola, and A. Colucci, "A innovative semi-immersible USV (SI-USV) drone for marine and lakes operations with instrumental telemetry and acoustic data acquisition capability," in *OCEANS 2015- Genova*, pp. 1–10, Genova, 2015.
- [4] M. T. Nguyen, C. V. Nguyen, L. H. Truong et al., "Electromagnetic field based WPT technologies for UAVs: a comprehensive survey," *Electronics*, vol. 9, no. 3, p. 461, 2020.
- [5] A. le, L. H. Truong, T. V. Quyen, C. V. Nguyen, and M. T. Nguyen, "Wireless power transfer near-field technologies for unmanned aerial vehicles (UAVs): a review," *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 7, no. 22, 2020.
- [6] H. T. Nguyen, M. T. Nguyen, H. T. Do, H. T. Hua, and C. V. Nguyen, "DRL-based intelligent resource allocation for diverse QoS in 5G and toward 6G vehicular networks: a comprehensive survey," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 5051328, 21 pages, 2021.
- [7] S. Goel, A. Kealy, and B. Lohani, "Development and experimental evaluation of a low-cost cooperative UAV localization network prototype," *Journal of Sensor and Actuator Networks*, vol. 7, no. 4, p. 42, 2018.
- [8] C. Xu, X. Liao, J. Tan, H. Ye, and H. Lu, "Recent research progress of unmanned aerial vehicle regulation policies and technologies in Urban low altitude," *IEEE Access*, vol. 8, pp. 74175–74194, 2020.
- [9] M. Turpin, N. Michael, and V. Kumar, "Trajectory design and control for aggressive formation flight with quadrotors," *Autonomous Robots*, vol. 33, no. 1-2, pp. 143–156, 2012.
- [10] Y. Qu, Y. Zhang, and Y. Zhang, "A UAV solution of regional surveillance based on pheromones and artificial potential field theory," in *2015 International Conference on Unmanned Aircraft Systems (ICUAS)*, pp. 380–385, Denver, CO, USA, 2015.
- [11] J. Liu and H. Li, "Artificial potential function safety and obstacle avoidance guidance for autonomous rendezvous and docking with noncooperative target," *Mathematical Problems in Engineering*, vol. 2019, Article ID 3451864, 17 pages, 2019.
- [12] A. D. Dang, H. M. La, T. Nguyen, and J. Horn, "Formation control for autonomous robots with collision and obstacle avoidance using a rotational and repulsive force-based approach," *International Journal of Advanced Robotic Systems*, vol. 16, no. 3, article 172988141984789, 2019.
- [13] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: algorithms and theory," *IEEE Transactions on Automatic Control*, vol. 51, no. 3, pp. 401–420, 2006.
- [14] P. J. Zarco-Tejada, V. González-Dugo, and J. A. J. Berni, "Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera," *Remote Sensing of Environment*, vol. 117, pp. 322–337, 2012.
- [15] M. Elloumi, R. Dhaou, B. Escrig, H. Idoudi, and L. A. Saidane, "Monitoring road traffic with a UAV-based system," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, Barcelona, Spain, 2018.
- [16] S. Rabinovich, R. E. Curry, and G. H. Elkaim, "Toward dynamic monitoring and suppressing uncertainty in wildfire by multiple unmanned air vehicle system," *Journal of Robotics*, vol. 2018, Article ID 6892153, 12 pages, 2018.
- [17] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: survey and experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2179–2195, 2008.
- [18] Y.-L. Chen, B.-F. Wu, H.-Y. Huang, and C.-J. Fan, "A real-time vision system for nighttime vehicle detection and traffic surveillance," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 2030–2044, 2011.
- [19] Ying-Wen Bai, Zi-Li Xie, and Zong-Han Li, "Design and implementation of a home embedded surveillance system with ultra-low alert power," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 1, pp. 153–159, 2011.
- [20] M. T. Nguyen, L. H. Truong, and T. T. H. Le, "Video surveillance processing algorithms utilizing artificial intelligent (AI) for unmanned autonomous vehicles (UAVs)," *Methods X*, vol. 8, article 101472, 2021.
- [21] H. Xue, Y. Liu, D. Cai, and X. He, "Tracking people in RGBD videos using deep learning and motion clues," *Neurocomputing*, vol. 204, pp. 70–76, 2016.

- [22] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6479–6488, Salt Lake City, UT, USA, 2018.
- [23] L. Caruccio, G. Polese, G. Tortora, and D. Iannone, "EDCAR: a knowledge representation framework to enhance automatic video surveillance," *Expert Systems with Applications*, vol. 131, pp. 190–207, 2019.
- [24] Y. Zeng and R. Zhang, "Energy-efficient UAV communication with trajectory optimization," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3747–3760, 2017.
- [25] J. Gong, T. Chang, C. Shen, and X. Chen, "Flight time minimization of UAV for data collection over wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 9, pp. 1942–1954, 2018.
- [26] Y. Xiao, Y. Cui, P. Savolainen et al., "Modeling energy consumption of data transmission over Wi-Fi," *IEEE Transactions on Mobile Computing*, vol. 13, no. 8, pp. 1760–1773, 2014.
- [27] M. T. Nguyen, L. H. Truong, T. T. Tran, and C.-F. Chien, "Artificial intelligence based data processing algorithm for video surveillance to empower industry 3.5," *Computers & Industrial Engineering*, vol. 148, article 106671, 2020.
- [28] M. T. Nguyen, "An energy-efficient framework for multimedia data routing in internet of things (IoTs)," *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 6, no. 19, 2019.
- [29] M. T. Nguyen and H. R. Boveiri, "Energy-efficient sensing in robotic networks," *Measurement*, vol. 158, article 107708, 2020.
- [30] M. T. Nguyen, H. M. La, and K. A. Teague, "Collaborative and compressed mobile sensing for data collection in distributed robotic networks," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 4, pp. 1729–1740, 2018.
- [31] G. Agnarsson and R. Greenlaw, *Graph Theory: Modeling, Applications, and Algorithms*, Prentice-Hall Inc., 2006.
- [32] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: human trajectory prediction in crowded scenes," in *European Conference on Computer Vision (ECCV)*, Amsterdam, The Netherlands, 2016.