2023 8th International Conference on Robotics and Automation Engineering

A Comparison of Clustering vs YOLO for Drone Swarm Centroid Detection

Nisha Kumari
School of Information Technology
Deakin University
Melbourne, Australia
kumarinis@deakin.edu.au

Kevin Lee
School of Information Technology
Deakin University
Melbourne, Australia
kevin.lee@deakin.edu.au

Chathurika Ranaweera
School of Information Technology
Deakin University
Melbourne, Australia
chathu.ranaweera@deakin.edu.au

Jan Carlo Barca
School of Information Technology
Deakin University
Melbourne, Australia
jan.barca@deakin.edu.au

2023 8th International Conference on Robotics and Automation Engineering (ICRAE) | 979-8-3503-2765-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICRAE59816.2023.10458512

***Abstract*—The rapid advancement of drones has led to the emergence of drone swarm applications in various domains, including surveillance, search and rescue, and package delivery. Efficient coordination and formation control of drone swarms is crucial for accomplishing complex tasks. The research presented in this paper proposes a novel approach for detecting individual drones in drone swarm formations through the utilisation of the K-means clustering algorithm. The algorithm assigns drones to the nearest centroids, creating cohesive subgroups and optimizing formation quality. To assess its efficacy, a comparative analysis is conducted between the K-means clustering algorithm and you only look once (YOLO) based computer vision detection algorithm. Through extensive simulation experiments, it is found that the K-means clustering algorithm outperforms the YOLO-based detection algorithm in terms of formation quality and computational efficiency. It consistently achieves more accurate and stable swarm formations making it suitable for real-time swarm control applications. The results of this presented research open the way for the effective use of drone swarms in a wide range of real-world applications and contribute to the development of cutting-edge swarm control approaches.**

***Index Terms*—Drone Swarm Formations, K-means Clustering Algorithm, Computer Vision**

I. Introduction

As a recently emergent technology, drones have enabled an entirely new class of applications in diverse domains. The current trend in this technology is away from large remotely-controlled drones and towards networks of small autonomous drones that can collectively complete complex tasks in a timely and cost-effective manner. Developing efficient sensing, communication and control algorithms that can accommodate the needs of highly dynamic drone networks with heterogeneous mobility levels is a significant challenge. The integration of computational intelligence into drone networks enables nodes with intelligence to make smarter networking decisions by utilising the learning capabilities of these nodes.

In recent years, advancements in technology have facilitated the replication of various biological behaviours exhibited by insects, animals, birds, and other organisms through the utilisation of robots, unmanned airborne vehicles (including conventional aircraft), underwater autonomous vehicles and similar technological innovations [1] [2]. The scope of applications is continuously expanding, encompassing both military and civilian domains, frequently involving the utilisation of surveillance or exploration techniques in various regions.

Swarm robotics and formation control have gained significant attention in recent years for their potential applications in various fields. Formation control is widely utilised in various applications, including a significant domain that encompasses search and rescue operations [3]. The purpose of utilising a formation is when agents within a specific arrangement (i.e., drones) are involved in activities such as surveillance or exploration, they possess the capability to collectively create an antenna that is significantly larger in size compared to each individual agent. One advantage of this is the enhanced level of sensitivity. Another factor to consider is that various sensors may possess distinct functionalities and when combined their aggregate functionality may result in the emergence of new sensor capabilities [4].

The increasing growth of drone technology has resulted in an increased demand for precise and effective algorithms for detecting drone swarms. Manual feature extraction and traditional clustering procedures have limitations when dealing with complicated formations and overlapped drones. To tackle these challenges, deep learning-based algorithms for object detection, such as YOLO have demonstrated significant potential in diverse object recognition tasks. When dealing with drones that are outside the camera's range of view, their performance may suffer. While, K-means clustering provides a unique approach by examining positional data, regardless of visual cues. The purpose of this research is to investigate the suitability and limitations of the YOLO and K-means algorithms for detecting drone swarms grouped in a diamond formation, utilizing a stereo-vision camera setup with varying observation angles.

This experimental study focuses on achieving a diamond formation in a swarm of drones using the K-means clustering algorithm and compares its performance with the YOLO detection algorithm. The main contributions of this paper are:

1) The implementation of diamond formation shape of drone swarm.
2) Comparison of vision-based approach YOLO and K-means clustering algorithm
3) An experimental analysis of the proposed approach.

The remainder of the paper is organised as follows. Section II is describing the background that justifies the creation of details of the previous work on the K-means and YOLO approaches. Section III describes a proposed framework for the detection and tracking of diamond formation drone swarm. Section IV presents a practical implementation and experimental evaluation of the drones in a simulation environment of the gazebo simulator. Finally, section VI presents some conclusions and future work.

II. Background and Context

Various types of clustering algorithms have been proposed across diverse fields and applications. Various algorithms differ significantly in their understanding of what constitutes a cluster. The most common clustering algorithm in use is one that involves grouping objects together based on their respective distances. The definition of distance may vary across different applications. Connectivity-based, centroid-based, distribution-based, density-based, and grid-based clustering represent distinct methodologies employed in a diverse range of applications. The K-means algorithm, a widely utilised centroid-based clustering method has been employed in this study to facilitate drone swarm formation applications [5].

The K-means clustering approach is widely recognised as one of the popular methods for clustering [6] [7]. With this approach, the cluster of nearby drones will provide connectivity to each drone. The K-means algorithm is an iterative clustering method used to determine the optimal centroid, representing the position of the drone in the given problem. The K-means algorithm begins by randomly initialising the positions of the drones and subsequently iteratively updates the centroid locations and cluster boundaries [8].

The real-time object detection algorithm known as YOLO was developed in 2016 by Joseph Redmon and Ali Farhadi [9] [10]. The YOLO model exhibits numerous advantages in comparison to conventional approaches for object detection and classification. According to the findings presented in article [11] the YOLO algorithm has demonstrated the capability to analyse a range of 40 to 90 images within a single second. This implies that the processing of streaming video can occur in real-time, exhibiting a minimal latency of a few milliseconds. The YOLO algorithm has gained significant popularity and widespread usage in various domains [12] [13]. The YOLO algorithm has gained significance due to its ability to detect objects in real-time. It is important to note that this algorithm requires significant computational resources and the process of training the data can be time-consuming.

The detection method YOLO necessitates comprehensive training through the utilisation of datasets that have been accurately labelled. The convolutional neural network (CNN) architecture is trained using annotated images that contain bounding box labels. This training process enables the CNN to effectively learn how to accurately detect and classify objects. While, the K-means clustering algorithm does not necessitate explicit training as it falls under the category of unsupervised learning methods. The algorithm updates the cluster centroids iteratively, carrying on until the distance metric indicates convergence.

III. A Flexible Hybrid Approach For Drone Detection

The implementation of the YOLO object detection algorithm enables the detection of a drone swarm arranged in a diamond formation. The YOLO [14] algorithm is a cutting-edge deep learning model that has the capability to accurately identify and precisely locate objects in real-time. By subjecting the YOLO model to training using drone images and subsequently annotating them with the diamond formation, the model can acquire the capability to classify the distinct configuration of drones. In order to draw a comparison between the YOLO algorithm and the K-means clustering algorithm, it is crucial to comprehend that K-means is a type of unsupervised machine learning algorithm that is employed for the purpose of clustering. The process involves partitioning a set of data points (i.e., point cloud data) into K distinct clusters, taking into consideration their similarities within the feature space [15]. In the domain of drone swarm detection, the K-means algorithm can be utilised to analyse the spatial arrangement of identified drones, enabling the identification of clusters that correspond to the diamond formation.

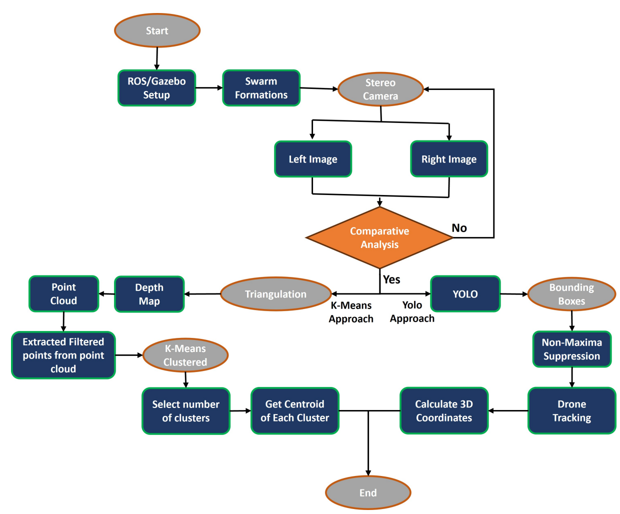


Fig. 1: Observing Drone Swarm Topology Formations Flowchart

However, there are significant differences between these two approaches. The YOLO algorithm is categorised as a supervised learning approach, as it necessitates the availability of labelled training data. Conversely, the K-means algorithm falls under the unsupervised learning paradigm and does not rely on labelled data for its operation. The YOLO algorithm exhibits the ability to detect multiple objects concurrently, whereas the K-means algorithm can only find clusters based on a fixed number of clusters (K). The YOLO algorithm offers enhanced accuracy in object localisation, while K-means clustering focuses on determining cluster centres.

The aim of the study is to evaluate the K-means clustering algorithm and compare it with the computer vision-based algorithm YOLO to detect each individual drone in drone swarm formations. The experiments are executed within the gazebo environment, with the utilisation of Rviz for the purpose of visualising the centroid of each cluster within drone swarm formations. Python is employed for the purpose of programming the simulation. Initially, a diamond drone swarm formation is simulated in gazebo as shown in Figure 1 and a stereo camera is kept at 3 m distance on various angles to observe the swarm. The main approach used in this study is to compare YOLO detection and K-means clustering algorithm. As seen in the flowchart Figure 1 a comparative analysis has been done for both approaches. On the left side of the flowchart, the main approach utilized is to use point cloud data to detect the centroid of each drone in drone swarm formations. Initially, a diamond drone swarm formation is simulated in the gazebo and a stereo-vision camera is mounted at different angles to observe the drone formation. From the stereo camera left and right images are used to extract a depth map using the triangulation method. Point cloud data is then attained using a depth map of drone diamond formation. Finally, the K-means clustering algorithm is implemented to get the centroid of each drone in drone swarm formations.

The YOLO approach is used to detect each individual drone in a diamond drone swarm formation. The YOLO algorithm is an advanced deep learning model that has the capability to accurately identify and precisely locate objects in real-time. By employing the YOLO model for training purposes on drone images and subsequently annotating them with the diamond formation, it is possible to facilitate the acquisition of knowledge by the model in recognising the distinct configuration of drones. The detection of multiple bounding boxes is achieved through the utilisation of stereo image frames. In order to refine the selection of accurate bounding boxes, a technique known as non-maxima suppression (NMS) is employed. Once accurate bounding boxes are obtained, drone tracking is performed, followed by the collection of 3D coordinates through a comparison with the actual coordinates.

A. Extracting Point Cloud Data of Diamond Swarm

Stereoscopic 3D reconstruction holds a pivotal role in the domain of computer vision. This process involves examining three-dimensional insights and spatial information about a scene by extracting a pair of stereo images. Similar to how our eyes perceive depth, these stereoscopic images communicate information about depth. Whereas a standard 2D image conveys only the dimensions of height and width, 3D reconstruction elevates this spatial understanding by introducing a third dimension depth [16]. The fundamental aim of 3D reconstruction revolves around precisely determining the corresponding location of a specific object within both the left and right images. This process serves the purpose of computing the disparity, which represents the relative positional distinction between these two object locations. This significant disparity information can then be used to arrange a full 3D scene reconstruction. Algorithms for 3D image reconstruction have extensive utility across diverse domains, encompassing 3D imaging systems, detailed 3D surface data acquisition, and the realm of medical imaging. Point clouds can be generated from various sensors and data sources, including LiDAR (Light Detection and Ranging), stereo cameras, depth sensors (e.g. Microsoft Kinect), and photogrammetry. These sensors capture points with 3D coordinates, making point clouds valuable for 3D data acquisition [17].

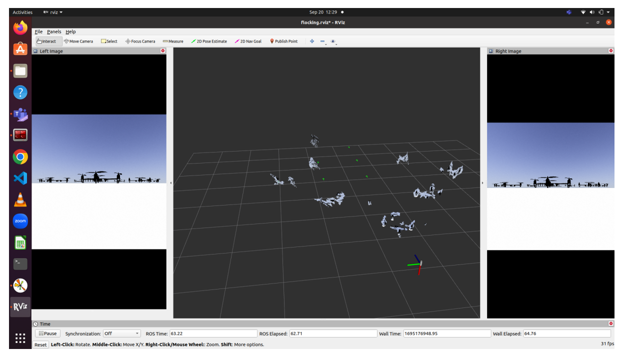


Fig. 2: Point Cloud of Stereo Images

A point cloud provides a more comprehensive representation of the 3D scene, taking depth information from both visible and partially hidden drones. It can capture drones that may not be apparent in the 2D images due to occlusion [18]. Stereo vision relies on disparities between corresponding points in the left and right images to calculate depth. Even if a drone is partially visible in one image but not in the other due to occlusion, the stereo-matching algorithm can still estimate its depth by identifying matching features between the images. This can result in the inclusion of the drone's 3D position in the point cloud [19]. Figure 2 depicts the simulation of a diamond drone swarm at 0° angle where in left and right images only 3 drones are visible but the point cloud shows the depth data of these images and the drones that are hidden can also be seen in point cloud image. Stereo vision relies on finding correspondences between pixels in the left and right images to calculate depth [20].

In this scenario, the depth map is used to know the depth of every drone in the frames. The formulas 1, 2, and 3 show the basic ideas behind the disparity-based method for estimating the three-dimensional spatial properties of a point (XL, YL) or (XR, YR) based on the camera system's properties, such as the focal length, baseline, and image resolution.

- **Focal length** ($\mathcal{F_L}$) **:** Camera's focal length in (*pixel*)
- **Baseline** ($\mathcal{B}$) **:** Distance between the two cameras in (*meter*)
- **Image size (resolution) :** height * width in (*pixel*)
- The point we locate is ($\mathcal{X_L}$, $\mathcal{Y_L}$) , ($\mathcal{X_R}$, $\mathcal{Y_R}$) (unit: *pixel*)

where, $\mathcal{X_L}$ = Left Camera at X-axis, $\mathcal{Y_L}$ = Left Camera at Y-axis, $\mathcal{X_R}$ = Right Camera at X-axis and $\mathcal{Y_R}$ = Right Camera at Y-axis

$$\mathcal{X} = \frac{(\mathcal{X}_L - \mathcal{C}_X) * \mathcal{B}}{\mathcal{D}} \tag{1}$$

$$\mathcal{Y} = \frac{(\mathcal{Y}_L - \mathcal{C}_Y) * \mathcal{B}}{\mathcal{D}} \tag{2}$$

$$\mathcal{Z} = \frac{\mathcal{F}_L * \mathcal{B}}{\mathcal{D}} \tag{3}$$

$$\mathcal{D} = abs(\mathcal{X_L} - \mathcal{X_R})$$

IV. Results and Discussion

The objective of this study is to present practical results that illustrate the advantages and drawbacks of utilising YOLO and K-means algorithms for the purpose of detecting drone swarms arranged in a diamond formation. It is anticipated that the YOLO algorithm will demonstrate exceptional performance in identifying visible drones within the camera's visual range. Conversely, the K-means algorithm is expected to yield significant centroid data pertaining to drones, even in instances where they may be partially obstructed. Furthermore, this study aims to determine the most suitable application scenarios for each technique, thereby offering significant insights into the detection of drone swarms in practical contexts.

A. Experiment 1: Drone Detection in Diamond Formation using K-means Clustering

This experiment shows that K-means clustering can be used to detect drones in a flat diamond. In this experiment, the results demonstrate the performance of the K-means clustering algorithm for diamond drone swarm formation. The aim is to evaluate the comparison between the vision-based approach YOLO and K-means for detecting drones at varying angles. The simulation provides position data for the drones. The positions of the simulated drones were systematically recorded at consistent intervals, effectively capturing their coordinates within the virtual environment. The point cloud data that was gathered is utilised as input for the purpose of assessing the effectiveness of the K-means clustering algorithm in achieving the formation of a diamond. Figure 4 shows the 3D view of the actual location and estimated location of the diamond using the K-means clustering approach.

When using a stereo vision camera, where the camera is situated at a distance of 3 m from the origin and capturing images at various angles ranging from $0°$ to $90°$, it can be seen in the Figure 3 that the YOLO algorithm exhibit limitations in detecting drones that lie outside the camera's range. But,

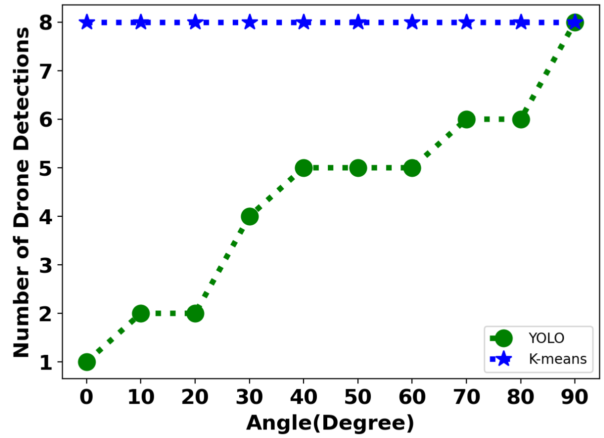


Fig. 3: Diamond Swarm Formation K-means v/s YOLO

K-means clustering can still generate centroids even in cases where the drones are not visible in the camera. As can be seen in Figure 5 centroid of each cluster (i.e., drone) is shown using different colours. The results of the experiments show that the K-means clustering algorithm is better than the YOLO detection algorithm for forming swarms of drones in diamond shape. The K-means algorithm was developed specifically for formation control, making it easier to establish and keep the desired formation. Despite its superiority in object recognition and tracking, the YOLO detection technique might not indirectly help achieve the formation control goal.

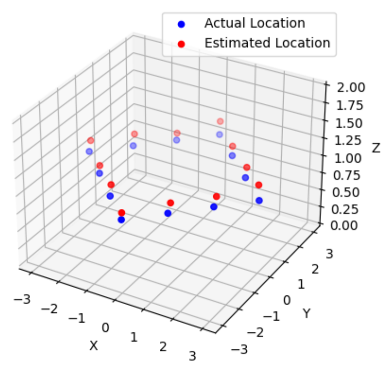
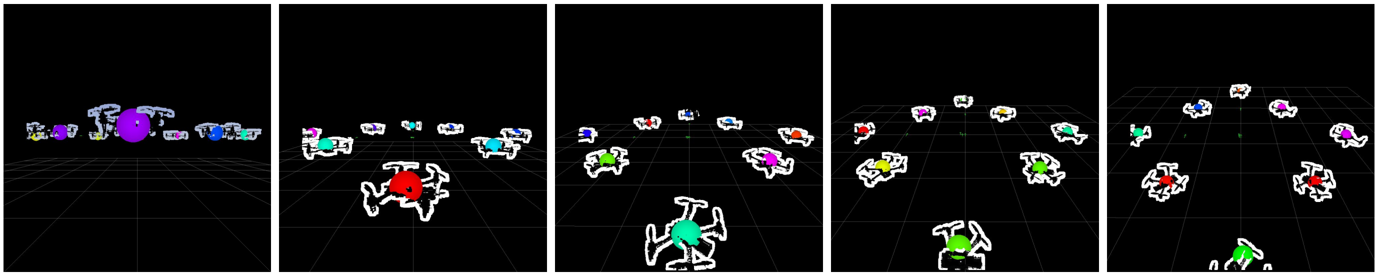


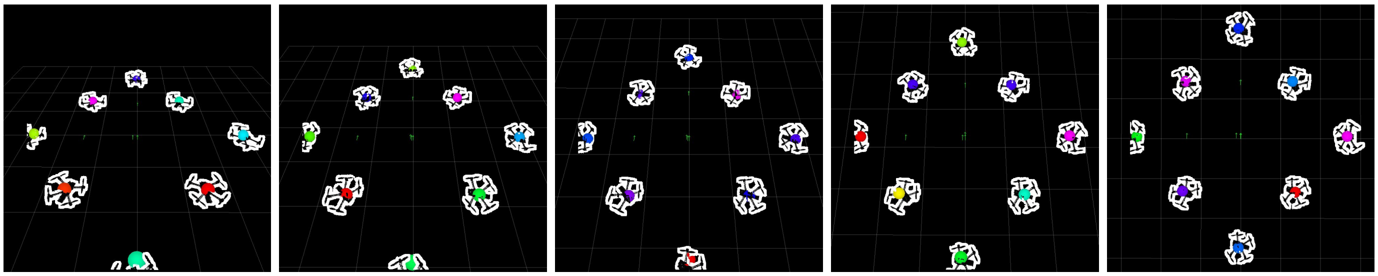
Fig. 4: Detection Centroid Diamond Swarm

B. Experiment 2: Individual Drone Detection in Diamond Formation using K-means Clustering

This experiment is a refinement of a previous one that showed how K-means clustering can be used to find drones in a flat diamond. It works well, but in this study, stereo cameras are used at different angles, from $0°$ to $90°$, so the results only apply to drones that could be seen from each angle. This experiment presents the findings of an algorithmic approach to detect individual drones using varying individual



(a) 0° (b) 10° (c) 20° (d) 30° (e) 40°



(f) 50° (g) 60° (h) 70° (i) 80° (j) 90°

Fig. 5: Drone Centroids at Various Angles

drone separation distance, with the aim of excluding centroid values of drones that are not within the camera's field of view. The algorithm 1 examines each angle to determine whether an individual drone is located within separation distance from the camera. If two centroids are below the threshold value, then taking the mean of those two centroid a new centroid value is calculated. As depicted in Figure 5 at 0° angle, only 5 drones are observable due to overlapping. Therefore, employing an individual separation distance produces a more precise outcome. Figure 6 depicts drone detection at varying separation distances. As can be seen, when the distance is less than $0.8\ m$ it shows the detection of 8 drones but when the distance is greater than $0.6\ m$ it again gives 5 centroid detections. The same approach was employed at each angle, achieving results shown in Figure 7 that are highly practical. In this graph, it is evident that the K-means clustering algorithm performs better at each angle for finding the centroid of individual drones in comparison to the YOLO algorithm.

V. Discussion

The research presented in this paper addresses a critical challenge in the realm of drone swarm applications the efficient detection and coordination of individual drones within swarm formations. As drone technology advances, its applications diversify into surveillance, search and rescue operations, and package delivery, necessitating sophisticated swarm control techniques. This study proposes a novel approach employing the K-means clustering algorithm for detecting individual drones within swarm formations. The algorithm's unique ability to assign drones to the nearest centroids results in cohesive subgroups, optimizing the overall formation quality.

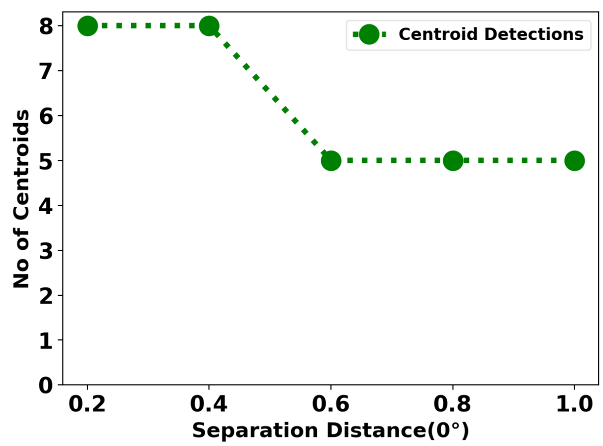


Fig. 6: Individual Drone Detection After Separation Distance

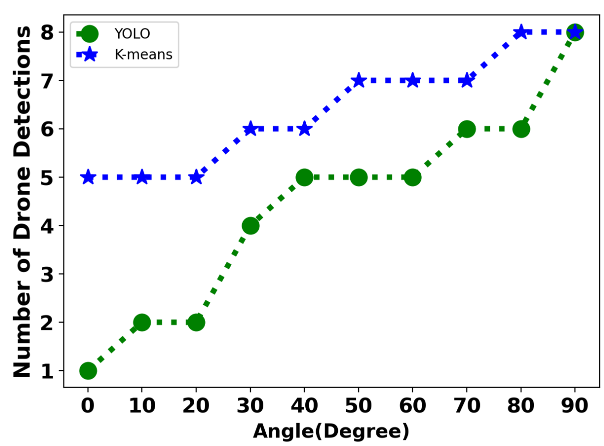


Fig. 7: Comparison of Diamond Swarm Formation

Algorithm 1 Example Algorithm

```
Input: Centroids (C), Points (P), d_min = 0.8
Output: Final Centroids (FC)
matched_centroids = []
FC = []
K = |C|                                   ▷ Number of Centroids
for i ← 1 to K do
    D = Euclidean_distance (C(i), P)
    if D <0.8 then
        Calculate Mean of matched_centroids
        matched_centroids.append (C(i))
    end if
end for
FC = sort(matched_centroids)              ▷ Sort in Ascending Order
```

The comparative analysis conducted between the K-means clustering algorithm and the widely used YOLO computer vision detection algorithm sheds light on the advantages of the proposed method. Through extensive simulation experiments, the research reveals that the K-means clustering algorithm surpasses the YOLO-based detection algorithm in both formation quality and computational efficiency. The consistent achievement of more accurate and stable swarm formations highlights the robustness of the K-means approach, making it particularly suitable for real-time swarm control applications.

One of the primary benefits of this research lies in its practical implications. By enhancing the precision and stability of drone swarm formations, the proposed K-means clustering algorithm opens avenues for diverse real-world applications. In surveillance, it ensures more effective coverage, improving the quality of data collection. In search and rescue operations, the optimized formations enable quicker response times and more area coverage, potentially saving lives in critical situations.

This research contributes to the development of cutting-edge swarm control approaches. The study not only establishes the superiority of the K-means clustering algorithm but also provides valuable insights into its potential extensions and adaptations for specific application domains. By optimizing the coordination and formation control of drone swarms, this research paves the way for the effective utilization of drone technology in a wide array of real-world scenarios, fostering innovation and efficiency across various industries.

VI. Conclusion

In Conclusion, this research introduces new techniques for detecting drone swarm formations by making use of the K-means clustering algorithm. The advantages of the K-means clustering algorithm in terms of formation quality and computing efficiency are highlighted by a comparison with the YOLO-based computer vision detection system. The proposed approach consistently produces swarm formations that exhibit enhanced precision and stability, thereby enhancing the collective coordination and cohesive movement of the swarm. Clustering is a powerful tool in numerous applications, including drone swarm coordination. The need to determine the number of drones upfront remains a significant disadvantage. Due to its reduced processing overhead, this technology exhibits potential for application in real-time swarm control. This research enhances the practicality of drone swarms and contributes to the progress of state-of-the-art swarm control techniques. The results of this study show the effective use of drone swarms in a wide range of real-world contexts and contribute to the development of cutting-edge swarm control approaches.

References

[1] Z. Zhang, T. Yang, T. Zhang, F. Zhou, N. Cen, T. Li, and G. Xie, "Global vision-based formation control of soft robotic fish swarm," *Soft Robotics*, vol. 8, no. 3, pp. 310–318, 2021.

[2] Y. Xie, L. Han, X. Dong, Q. Li, and Z. Ren, "Bio-inspired adaptive formation tracking control for swarm systems with application to uav swarm systems," *Neurocomputing*, vol. 453, pp. 272–285, 2021.

[3] J. Lwowski, A. Majumdar, P. Benavidez, J. J. Prevost, and M. Jamshidi, "Bird flocking inspired formation control for unmanned aerial vehicles using stereo camera," *IEEE Systems Journal*, vol. 13, no. 3, pp. 3580–3589, 2018.

[4] B. D. Anderson, B. Fidan, C. Yu, and D. Walle, "Uav formation control: Theory and application," in *Recent advances in learning and control*, pp. 15–33, Springer, 2008.

[5] A. Mirzaeinia, M. Mirzaeinia, M. Shekaramiz, and M. Hassanalian, "Withdrawal: Placement of uav-mounted mobile base station through user load-feature k-means clustering," in *AIAA AVIATION 2020 FORUM*, pp. 2925–c1, 2020.

[6] U. R. Raval and C. Jani, "Implementing & improvisation of k-means clustering algorithm," *International Journal of Computer Science and Mobile Computing*, vol. 5, no. 5, pp. 191–203, 2016.

[7] Y. Meng, J. Liang, F. Cao, and Y. He, "A new distance with derivative information for functional k-means clustering algorithm," *Information Sciences*, vol. 463, pp. 166–185, 2018.

[8] H. Hu, J. Liu, X. Zhang, and M. Fang, "An effective and adaptable k-means algorithm for big data cluster analysis," *Pattern Recognition*, vol. 139, p. 109404, 2023.

[9] T. Ahmad, Y. Ma, M. Yahya, B. Ahmad, S. Nazir, and A. u. Haq, "Object detection through modified yolo neural network," *Scientific Programming*, vol. 2020, pp. 1–10, 2020.

[10] X. Wang, H. Li, X. Yue, and L. Meng, "A comprehensive survey on object detection yolo," *Proceedings http://ceur-ws. org ISSN*, vol. 1613, p. 0073, 2023.

[11] S. Shinde, A. Kothari, and V. Gupta, "Yolo based human action recognition and localization," *Procedia computer science*, vol. 133, pp. 831–838, 2018.

[12] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of yolo algorithm developments," *Procedia Computer Science*, vol. 199, pp. 1066–1073, 2022.

[13] F. Sultana, A. Sufian, and P. Dutta, "A review of object detection models based on convolutional neural network," *Intelligent computing: image processing based applications*, pp. 1–16, 2020.

[14] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[15] Z. Huang, Y. Xiang, B. Zhang, D. Wang, and X. Liu, "An efficient method for k-means clustering," 2023.

[16] Y. Liu, H. Wu, C. Wang, Y. Wei, M. Ren, and T. Feng, "Real-time dense construction with deep multi-view stereo using camera and imu sensors," *IEEE Sensors Journal*, 2023.

[17] S. A. Bello, S. Yu, C. Wang, J. M. Adam, and J. Li, "Deep learning on 3d point clouds," *Remote Sensing*, vol. 12, no. 11, p. 1729, 2020.

[18] M. Forsman, "Point cloud densification," 2010.

[19] Y. Zhang and M. Rabbat, "A graph-cnn for 3d point cloud classification," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6279–6283, IEEE, 2018.

[20] S. Chen, B. Liu, C. Feng, C. Vallespi-Gonzalez, and C. Wellington, "3d point cloud processing and learning for autonomous driving: Impacting map creation, localization, and perception," *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 68–86, 2020.