ABSTRAK

Penelitian ini mengusulkan strategi pengendalian *flocking* berbasis *deep reinforcement learning* (DRL) untuk *swarm quadcopter* yang beroperasi dalam lingkungan dengan kepadatan rintangan tinggi serta variasi jumlah anggota kawanan. Berbeda dengan metode berbasis aturan konvensional yang kurang adaptif terhadap skenario yang kompleks dan dinamis, pendekatan ini merumuskan tugas flocking sebagai suatu proses keputusan Markov dengan pengamatan parsial (*Partially Observable Markov Decision Process*/POMDP), dengan mempertimbangkan keterbatasan persepsi dan komunikasi lokal dari setiap agen. Kebijakan kendali dilatih menggunakan algoritma *Proximal Policy Optimization* (PPO) melalui skema pelatihan tersentralisasi dan pelaksanaan terdesentralisasi dalam lingkungan multiagen. Fungsi perolehan (*reward function*) dirancang dengan mengintegrasikan aturan klasik berbasis boid (*separation, cohesion, alignment*), penghindaran tabrakan secara eksplisit, serta batasan koridor virtual. Pendekatan ini memungkinkan *swarm* untuk tidak hanya mempertahankan formasi yang stabil dan menghindari rintangan, tetapi juga beroperasi secara aman dalam batas wilayah yang telah ditetapkan. Simulasi secara ekstensif telah dilakukan pada berbagai skenario, termasuk lingkungan dengan keberadaan rintangan serta jumlah agen yang bervariasi antara 5 hingga 50 *quadcopter*. Hasil kuantitatif menunjukkan bahwa metode yang diusulkan mencapai tingkat keberhasilan menuju sasaran sebesar minimal 98% pada seluruh skenario, mempertahankan tingkat *collision* dan *deconfliction* di bawah 2,5% untuk ukuran *swarm* yang tinggi, serta menjaga jarak spasial dan formasi yang stabil secara efektif.

Kata Kunci : Quadcopter, Kendali Flocking, DRL, POMDP, PPO, Multi-Agent

ABSTRACT

This research presents a deep reinforcement learning (DRL)-based flocking control strategy for quadcopter swarms operating in environments with dense obstacles and varying swarm sizes. Unlike conventional rule-based methods, which struggle to adapt to complex and dynamic scenarios, the proposed approach formulates the flocking task within a Partially Observable Markov Decision Process (POMDP), accounting for the local perception and communication constraints of each agent. The control policy is trained using the Proximal Policy Optimization (PPO) algorithm, implemented with centralized training and decentralized execution in a multi-agent setting. Classic boid-inspired rules (separation, cohesion, alignment), explicit collision avoidance, and a virtual corridor constraint are integrated into the reward function. This enables the swarm not only to maintain stable formation and avoid obstacles, but also to operate safely within predefined spatial boundaries, addressing a critical gap in existing flocking research. Extensive simulations were conducted across multiple scenarios, including environments with obstacles and swarm sizes ranging from 5 to 50 quadcopters. The quantitative results demonstrate that the proposed method achieves a target goal arrival of at least 98% across all scenarios, maintains low collision and deconfliction rates (below 2.5% for the largest swarms), and preserves effective spatial separation and stable flocking formations.

Keywords: Quadcopter, Flocking Control, DRL, POMDP, PPO, Multi-Agent.

ACKNOWLEDGMENTS

This thesis is compiled with the effort, help, and support of all the contributing elements. The author would like to express the deepest gratitude and thanks to:

1. Allah SWT, for all the love, guidance, and forgiveness in every mistake that the author has ever made, and Rasulullah SAW, as a role model who inspires the writer in living life and trying to be better.
2. My beloved mother, for her endless prayers, encouragement, and unconditional love, which have always been the greatest source of strength and motivation throughout every challenge in life and study.
3. Dr. Erwin Susanto, S.T., M.T., as my supervisor, and Dr. Ir. Sony Sumaryo, M.T., as my co-supervisor, for their invaluable guidance, insightful advice, and continuous support throughout the research and completion of this thesis.
4. All colleagues at PT Len Industri (Persero), especially the System Engineering Division, for their support, understanding, and cooperation, which have provided a stimulating and supportive working environment during this study.
5. My colleagues of the 2022 Master’s in Electrical Engineering students at Telkom University, for encouragement, and shared experiences that have enriched both my academic and personal life.
6. Finally, to all parties who have directly or indirectly contributed to this research and the completion of this thesis. Your support and encouragement are sincerely appreciated.

PREFACE

All praise and gratitude are due to Allah SWT for His infinite blessings, guidance, and mercy. With His grace, the author has successfully completed this thesis entitled "Flocking Control of Swarm Quadcopters Based on Deep Reinforcement Learning." This thesis is submitted in fulfillment of the requirements for graduation from the Master's Program of Electrical and Telecommunication Engineering, School of Electrical Engineering, Telkom University.

The preparation and completion of this thesis has not only expanded my technical and scientific understanding but have also shaped my character and perseverance in facing various challenges. The work reflects my dedication to advancing the field of swarm robotics, particularly in developing intelligent, adaptive control strategies for UAV swarms using reinforcement learning. It is my sincere hope that the results and insights shared in this work will contribute meaningfully to the academic community and inspire further research and innovation in the domain of multi-agent systems and autonomous aerial robotics. Suggestions for further improvement of this thesis are highly appreciated. May this work continue to be improved and provide valuable contributions to readers and to Indonesia, especially for the advancement of education and research in the field of robotics in the future.

Bandung, 24th July 2025

Bahtiar Yoga Prasetyo

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

UAV	Unmanned Aerial Vehicle
DRL	Deep Reinforcement Learning
DNN	Deep Neural Network
POMDP	Partially Observable Markov Decision Process
MAPPO	Multi Agent Proximal Policy Optimization
CTDE	Centralized Training Decentralized Execution
GAE	Generalized Advantage Estimation
FC	Fully Connected
CPU	Central Processing Unit
GPU	Graphical Processing Unit
RAM	Random Access Memory
HIL	Hardware in the Loop
SO (3)	Special Orthogonal Group 3D
PD	Proportional Derivative (controller)

LIST OF SYMBOLS

x, y, z	Position coordinate in the inertia frame
Φ, θ, ψ	Euler angles: roll, pitch, yaw
m	Quadcopter mass
T	Total thrust force
k_t	Thrust coefficient
k_q	Drag coefficient
l	Quadcopters arm length
g	Gravity
$I_{xx},\ I_{yy},\ I_{zz}$	Moment of inertia around x, y, z respectively
d_{com}	Maximum inter-agent communication distance
d_{scan}	Maximum perception range
θ_{scan}	Perception Field of View
P_i^s	Initial position of agent i
P_i^g	Goal/target position of agent i
P_j^b	Position of obstacle j
P_c	Center flock position
d_{max}^g	Maximum distance to the goal
d_{min}^b	Minimum safe distance to obstacles
d_{min}^{sep}	Minimum separation distance between agents
ϵ_{ali}	Alignment tolerance parameter
d_{max}^{coh}	Maximum distance to the flock centroid for cohesion
$\mathcal{C}$	Corridor (virtual tunnel) area
d_{lat}	Lateral distance to the corridor center
d_{ver}	Vertical distance to the corridor center
$L^{CLIP}\ (\theta)$	Clip loss function (PPO Objectives)
V_Φ	Value function (PPO Objectives)
γ	Discount factor (PPO Objectives)
λ	Generalized advantage estimation (PPO Objectives)
$\boldsymbol{a}_t$	Action (PPO Objectives)

$\boldsymbol{s}_t$	State (PPO Objectives)
π_θ	Policy (PPO Objectives)
$\boldsymbol{o}_t^s$	Self-state observation
$\boldsymbol{o}_t^{nbr}$	Neighbour state observation
$\boldsymbol{o}_t^{obs}$	Obstacle observation
$\boldsymbol{o}_t^g$	Goal observation
$\boldsymbol{r}_t^g$	Goal reaching reward
r_t^c	Collision avoidance reward
r_t^f	Flocking maintenance reward
$R_t^{sep}, R_t^{ali}, R_t^{coh}$	Reward for separation, alignment, cohesion respectively
r_t^{cor}	Corridor maintenance reward
$\omega_g, \omega_c, \omega_f, \omega_{cr}$	Weight coefficients for goal, collision, flocking, and corridor rewards respectively
$\omega_{sep}, \omega_{coh}, \omega_{ali}$	Weight coefficients for separation, cohesion, and alignment rewards respectively

CHAPTER 1
INTRODUCTION

1.1. Background

In recent years, research related to UAV swarms has been explored in several applications such as search and rescue missions [1], agriculture [2], mapping [3], military [4], and entertainment [5]. Compared to single UAVs, swarms are capable of handling more complex tasks while offering enhanced fault tolerance and robustness. The collective behavior of such UAVs relies on the implementation of flocking strategies, which serve as a high-level control mechanism and are central to swarm operations [6]. Flocking is a phenomenon observed in nature, referring to the coordinated group movement of animals such as bird flocks, fish schools, or herds of land animals. Inspired by these natural systems, Reynolds introduced the Boids model in 1986 [7] which simulates flocking behavior based on three fundamental heuristic rules: separation, alignment, and cohesion.

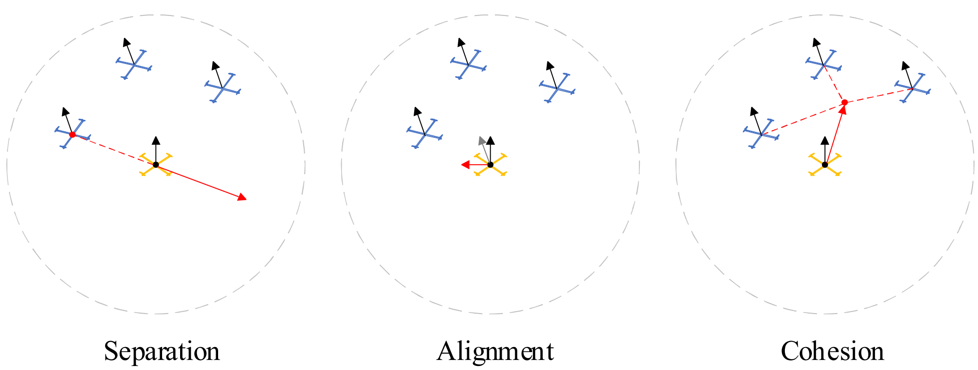

Figure 1 Boids rules

The Boids model has since formed the foundation for various studies on flocking control. For instance, in [8] a combination of Linear Quadratic Regulator (LQR) and Genetic Algorithm (GA) was applied to achieve optimal swarm movement. Although effective in maintaining tracking, aggregation, and velocity performance, the study did not address performance in obstacle-dense environment. Similarly, [9] proposed a distributed integral control method for stable and uniform navigation in large UAV swarms. While effective, this approach was limited to 30 UAVs and required considerable time to achieve convergence. Another notable contribution is the EGO-swarm algorithm introduced in [10], which utilizes vision-based navigation through spatiotemporal joint optimization. While it demonstrates effective performance in complex environments, its implementation is heavily dependent on hardware capabilities and is constrained to a limited number of UAVs.

Although these methods have shown promise, they generally rely on complex and predefined mathematical rules, limiting their adaptability to highly dynamic and uncertain environments. This study proposes a novel approach to flocking control using Reinforcement Learning (RL) for swarm quadcopter operations. The primary focus is to address scalability in scenarios characterized by dense obstacles and varying swarm sizes.

1.2. Problem Identification and Objectives

Recent advancements have highlighted RL as a powerful framework for swarm robotics due to its generalization capability, flexibility, and learning efficiency [11]. For example, [12] demonstrated the application of RL with Deep Neural Networks (DNN) combined with Force-based Motion Planning (FMP), achieving a 75% success rate in point-mass agents. However, this setup simplifies the agent's physical characteristics. In another study, [13] applied Q-Learning to a real-world quadcopter swarm in constrained spaces with high obstacle density. Despite the method's success, it utilized a virtual leader-follower scheme, limiting the overall flexibility as the swarm behavior was dependent on a single UAV.

To overcome these limitations, this research proposes an RL-based flocking control method that considers both kinematic and dynamic properties of quadcopters. The specific objectives of the study include:

1) Develop a reinforcement learning-based high-level control method for swarm quadcopter operations.
2) Analyze the performance of the proposed method across various simulated environments with different configurations.

1.3. Scope of Works

The scope of work and limitation of the research problem are as follows:

1) The Quadcopter will be used as UAV agent model incorporates kinematic and dynamic properties, excluding detailed aerodynamic modeling [14], [15].
2) The interaction among quadcopters is formulated under a Partially Observable Markov Decision Process (POMDP), considering limited communication and perception range.
3) Flocking control is implemented using deep reinforcement learning (DRL).
4) The training algorithm used is Proximal Policy Optimization (PPO) [16], with implementation as multi agent scenarios [17].
5) Performance evaluation focuses on the model's flexibility and robustness through various reward settings and environment configurations, assessing:
   a. Target reaching capability,
   b. Obstacle avoidance efficiency,
   c. Maintenance of stable inter-agent distances.

1.4. Hypothesis

The expected outcome of this research is that an RL-based flocking control model will provide a generalized and flexible solution for quadcopter swarm operations, particularly in obstacle-rich environments with varying swarm sizes. The trained model is expected to outperform traditional rule-based approaches in terms of adaptability and scalability.

1.5. Research Method

The methodology used in this research refers to the following structure:

1) Literature Study

   Conduct literature studies related to the basic concepts of quadcopter modeling (kinematic and dynamic), Reinforcement Learning along with the application of Partially Observed Markov Decision Process (POMDP) as a formulation and Proximal Policy Optimization (PPO) as a training method.

2) Flocking Control Model Design

Designing Flocking Control using Deep Reinforcement Learning.

2) Experiment and Analysis

Conducting simulations based on designed scenarios, followed by performance analysis.