14th International Conference on Control, Automation and Information Sciences

Book of Abstracts and Late Breaking Results Papers

October 27-29, 2025
Jeju, Republic of Korea

IEEE

Table of Contents

Preface

Part A: Regular Papers (Abstracts Only)

Part B: Late Breaking Results Papers

PREFACE

Welcome Message from the General Chair

It is my great honor and pleasure to invite you to the 14th International Conference on Control, Automation, and Information Sciences (ICCAIS) 2025, organized by the Korean Aerospace University. This prestigious gathering aims to bring together brilliant minds from around the world—scholars, researchers, and students—who are dedicated to developing cutting edge technologies and pushing the boundary of knowledge.

I extend my sincere gratitude to each of you for your commitment to research and innovation. Your endless efforts and groundbreaking discoveries are shaping a future where intelligent systems revolutionize industries and improve lives globally. This conference serves as a vital platform for knowledge exchange, collaboration, and intellectual discourse, driving the next wave of technological advancements.

We are privileged to host this conference on Jeju Island, a UNESCO World Heritage Site renowned for its breathtaking landscapes and rich cultural heritage. Known as the "Island of the Gods," Jeju is not just a paradise for nature lovers but also a hub of sustainability and innovation. I encourage you all to draw inspiration from its beauty and explore its wonders while engaging in insightful discussions.

Let us seize this opportunity to engage, exchange ideas, and foster meaningful collaborations. May this conference inspire new discoveries, strengthen academic partnerships, and pave the way for future innovations.

Looking forward to meet you and a productive and enriching conference.

Sincerely,

Professor Byungkyu Kim

Korea Aerospace University, Republic of Korea

Acknowledgement Message from the Organizers

It is our great pleasure to welcome you to the 14th International Conference on Control, Automation and Information Sciences (ICCAIS 2025), hosted on the beautiful shores of Jeju Island. This conference serves as an innovative and collaborative forum for international researchers to present and discuss cutting-edge achievements and future directions in control theory, robotics, automation systems, information sciences, and related technologies.

We extend our deepest gratitude to the Technical Program Committee members and reviewers for their outstanding efforts in curating a selection of high-quality papers. We also commend the dedication and professionalism of the Organizing Committee, local staff, and session chairs, whose hard work has been instrumental in bringing this event to fruition.

Our sincere thanks go to the plenary speakers and distinguished guests for their generous support and invaluable expertise, as well as to all presenters for their insightful contributions to ICCAIS 2025. We are particularly grateful to our sponsors and patrons for their generous financial support, which has been essential to the conference's success.

We hope ICCAIS 2025 fosters meaningful exchanges and inspires new collaborations. Thank you for joining us, and we look forward to the impactful discussions ahead.

Sincerely,

Local Arrangement Co-Chairs

Professor Seonggun Joe, Korea Aerospace University, Republic of Korea

Professor Bohyun Hwang, Korea Aerospace University, Republic of Korea

Professor Pileun Kim, Korea Aerospace University, Republic of Korea

Tutorial "Bridging Theory and Applications with Possibility Theory"

Presenter: Professor Jeremie Houssineau, Nanyang Technological University

Summary

Although it is widely accepted that there two main sources of uncertainty, that is, epistemic uncertainty corresponding to lack of knowledge and aleatoric uncertainty corresponding to random phenomena, the usual approach is to model both using the tools of probability theory. In this tutorial, I will show how a particular version of possibility theory offers an intuitive approach to model epistemic uncertainty which can be combined with probability theory to yield a general framework in which different sources of uncertainty are faithfully modelled. Such a framework provides a principled basis for many successful heuristics, allowing to understand and improve them, hence furthering the applicability of the Bayesian method. The effectiveness of this approach will be illustrated via a range of applications including distributed inference, ensemble Kalman filtering and space situational awareness. This tutorial does not require any prior knowledge of possibility theory.

Biography

Jeremie Houssineau is an Assistant Professor in the Division of Mathematical Sciences at Nanyang Technological University (NTU), Singapore, and also holds an honorary Associate Professor position in the Department of Statistics at the University of Warwick. His primary research interest lies in tackling challenges in statistical inference by integrating possibility theory and probability theory to model both epistemic and aleatoric uncertainty, with applications including robust inference, reinforcement learning, active learning, and multi-target tracking. He also serves on the Editorial Board of Statistics and Computing. Previously, he was an Assistant Professor (tenured in 2023) in the Department of Statistics at the University of Warwick (2019-2023), a Postdoctoral Fellow in the Department of Statistics and Data Science at the National University of Singapore (2016-2018), a PhD student at the School of Engineering and Physical Sciences, Heriot-Watt University (2011-2015), and a Research Engineer at Naval Group (2009-2011).

Keynote Speech “Autonomous Navigation and Localisation: A Tale of Two Viewpoints”

Presenter: Professor Girish N. Nair, The University of Melbourne

Abstract

We discuss two almost diametrically opposite stochastic perspectives on autonomous navigation and localisation: one discrete, and the other continuous and infinite-dimensional. In the first approach, we model the autonomous agent as a discrete-valued, partially observed Markov decision process, and augment it with entropy-based measures of trajectory uncertainty and complexity. These measures capture the minimum internal cost, in bits, of storing/communicating state trajectory beliefs and sensor observation sequences. In robotics, whole-of-trajectory entropies have previously been dismissed as intractable to optimise due to the nonlinear joint entropy functional acting on the entire trajectory. We show that, surprisingly, these entropies can be put into convenient stage-additive forms that enable optimisation using standard techniques, leading to principled trade-offs between exploitation and exploration. In simulations, trajectories reminiscent of the motion of some animals are produced.

In the second approach, we model the agent dynamics as a continuous-valued stochastic process with infinite-dimensional, noisy measurements. This is motivated by systems with low-dimensional dynamics but high-dimensional sensors, for instance autonomous vehicles equipped with vision or LiDAR. For linear systems, we explicitly derive the optimal linear filter in the sense of the minimum mean square error, analogous to the classic Kalman filter. We then propose an extension of this approach to handle nonlinearities, leading to an EKF-like approach with infinite-dimensional measurements. This extension provides a novel system-theoretic justification for the use of imagegradients in vision-based estimation. We demonstrate the practical utility of this filter on a real-world aerial drone dataset, showing significantly improved performance compared to an established visual odometry technique.

This is based on recent work with Dr Timothy L. Molloy (ANU) and Maxwell M. Varley:

https://doi.org/10.1109/TAC.2023.3250159

https://doi.org/10.1109/TAC.2023.3264177

https://doi.org/10.1109/TAC.2024.3464892

Biography

Girish N. Nair is a Professor in the Department of Electrical and Electronic Engineering at The University of Melbourne. He is a Fellow of the IEEE and was an Australian Research Council Future Fellow (2015 - 2019). His research focuses on the interplay between estimation, control and information theory, in both stochastic and nonstochastic settings. He has received several prizes, including the 2014 George S. Axelby Outstanding Paper Award from the IEEE Control Systems Society and a 2006 SIAM Outstanding Paper Prize. From 2019 - 2024 he was the lead Australian investigator for the Australia-US Multidisciplinary University Research Initiative on Neuro-Autonomy. From 2017 - 2022 he served as Deputy Head of Department (Research & Research Training), and from 2022 - 2024 he led the Control & Signal Processing research group in the Department. He is the General Chair of the 67th IEEE Conference on Decision and Control, to be held in Sydney in 2028, and the inaugural Chair of the Victoria/New South Wales Joint Chapter of the IEEE Information Theory Society.

Plenary Speech “Commercialization of Medical Microrobotics”

Presenter: President Jong-Oh Park, Korea Institute of Medical Microrobotics

Abstract

Microrobotics is surely one of very challenging research issues, and this impression has not changed up to now. It requires advanced technologies in various fields such as actuation, navigation, intelligence, communications, power supply, and even extinction. These high-level technologies result in high manufacturing costs. Consequently, microrobotics seems more suitable for human body-related fields, such as medical and warfare applications, emphasizing the importance of medical microrobotics.

While promising and fundamental technologies like the use of biological protein motors and enzyme functions remain hot topics for discussion, fundamental research will eventually translate into practical products. In this presentation, various commercialization efforts are explored—from applications in the digestive organ and circulatory system to anti-cancer therapy. Before discussing real disease applications, the basics of medical microrobotics are introduced.

Several detailed practical efforts are highlighted. For example, in digestive organs, actively controlled capsule endoscopes are demonstrated, focusing on precise actuation and navigation, as well as 3D reconstruction and autonomous symptom detection. In the circulatory system, diagnostic devices such as catheters—traditionally relying on surgeon skill—have been adapted to remote control systems, including a compact magnetic actuation system. Another development includes transarterial chemical embolization with precise guidance of embolic beads. These commercialization efforts are expected to lead to tangible products and further active research.

Biography

Jong-Oh Park is a renowned South Korean robotics scholar, currently serving as the president of the Korea Institute of Medical Microrobotics (KIMIRo), an organization he established in 2019 to advance medical microrobotics research and commercialization. He received his B.S. in Mechanical Engineering from Yonsei University in Seoul, followed by an M.S. in Mechanical Engineering at KAIST in Daejeon. After earning his master's degree, he completed a Ph.D. at Stuttgart University and worked as a researcher under Hans-Jurgen Warnecke at Fraunhofer-IPA in Stuttgart, Germany, from 1982 to 1987. He was associated with the Korea

Institute of Science and Technology (KIST) from 1987 to 2004, serving as the director of the 21C Frontier Project “Intelligent Microsystem Program” for five years. He was a professor at Chonnam National University from 2005 to 2022 and simultaneously led the Robot Research Initiative (RRI). Jong-Oh Park has held prominent roles in the international robotics community, including chairman of the International Federation of Robotics (IFR) in 2006 and an executive board member. He also contributed as Vice Chair of the Organizing Committee for ISR 2013, organizing chair of ICCAS 2013, and editor of IEEE Biorob in 2014. Between 2014 and 2017, he served as vice president of the Institute of Control, Robotics and Systems (ICROS) and is currently the chairman of its BioRobot Research Group. Appointed to the Presidential Committee on the 4th Industrial Revolution and the Healthcare Ad Hoc Committee in 2017, Jong-Oh Park has been a pivotal figure in advancing robotic technologies in medicine and industry.

Conference Program – Day 1 (27/10/2025)

Time	Events	Location
8:00 - 9:00	Conference Registration	Sarah 2
9:00 - 12:00	Tutorial “***Bridging theory and applications with possibility theory***” by Professor Jeremie Houssineau from Nanyang Technological University (**including 20 minutes morning tea break**).	Hallah 1
12:00 - 13:30	Lunch Break	Hotel Restaurant
13:30 - 14:50	Session L1: Late-Breaking Results 1	Hallah 1
14:50 - 15:20	Afternoon Tea Break	Sarah 2
15:20 - 17:00 (Parallel Session 1)	Session L2: Late-Breaking Results 2	Hallah 1
15:20 - 17:00 (Parallel Session 2)	Session L3: Frontier Special Session	Hallah 3
18:00 - 20:00	Special Session on Advanced Medical Technologies	Hallah 1
19:00 - Late	Conference Reception	Sarah 2

Session L1: Late-Breaking Results 1 (27/10/2025, 13:30 – 14:50, 80 minutes)

Note: Each paper has 8 minutes to present and 3 minutes for Q&A		
Order	**Paper Titles**	**Authors**
1	A Novel User Intent Prediction System for Improved Ad Targeting Strategies	Rian Khetani; Pranav Kulkarni
2	Keyframe Selection Strategies for Balancing Global Accuracy and Local Stability in Endoscopic SLAM	Hyeon Seo Kim; Seokmin Hong; Byungjeon Kang
3	A PDMS-Based Acoustic Fresnel Lens for Vortex Trapping of Microparticles	Viet Hoang Nguyen; Hiep Xuan Cao; Daewon Jung; Jong-Oh Park; Byungjeon Kang
4	Mode-Switching-Based Control Algorithm in Acoustic Actuator for 3D Microrobot Manipulation in Nonhomogeneous Medium	Hiep Xuan Cao; Daewon Jung; Viet Hoang Nguyen; Seokmin Hong; Jong-Oh Park; Byungjeon Kang
5	On-Device Vision Language Model Based Natural Language Mission Execution Framework	Seunggyu Lee; Seungjune Oh; Sangwoo Jeon; Park Youngju; Siwon Lee, Ghilmo Choi and Dae-Sung Jang
6	Situation-Aware Risk Quantification for UAV-Satellite Communications Using Semantic Anomaly Detection	Park Junbeom; Jong-Sou Park; Zizung Yoon
7	Design, fabrication and evaluation of 6R-Pantograph Antenna mechanism	Hyeongseok Kang; Byungkyu Kim

Session L2: Late-Breaking Results 2 (27/10/2025, 15:20 – 17:00, 100 minutes)

Note: Each paper has 9 minutes to present and 3 minutes for Q&A

Order	Paper Titles	Authors
1	Dynamic Modeling of a Novel R-BWB Aircraft for Altitude Control Studies	Sungmin Shim; Junseop Shin; Taejin Park; Dae-Sung Jang
2	A Perceptive Fingerless Soft Gripper via Inductive Sensing	Haneul Kim; Changbeom Shim; Seonggun Joe
3	A Preliminary Study on Programmable Soft Architectures via Surface Tessellations	Hayeon Shin; Nhan Huu Nguyen; Seonggun Joe
4	Design and Fabrication of Metamaterial Based Capacitive Sensor via 3D Printing	Seik Park; Eunji Kim; Seonggun Joe
5	Multichannel Ultrasound Driving Module-Based Drug Delivery System: Inducing Targeted Release of Doxorubicin for Liver Cancer Therapy	Daewon Jung; Hiep Xuan Cao; Viet Hoang Nguyen; You Hee Choi; Tae Il Kang; Seokmin Hong; Jong-Oh Park; Byungjeon Kang
6	Machine Learning-Based Estimation of Root-Zone Soil Moisture from Heterogeneous Data Sources	Kamilla Rakhymbek; Markhaba Karmenova
7	Bearing-based Distributed Formation Control for Unmanned Aerial Vehicles	Jaewan Choi; Younghoon Choi
8	Reinforcement Learning Based Controller for High-Precision Dual-Servo Stages	Seung Woo Seo; Yongho Jeon; Moon Gu Lee

Session L3: Frontier Special Session (27/10/2025, 15:20 – 17:00, 100 minutes)

Note: Each paper has 8-10 minutes to present and 3 minutes for Q&A		
Order	**Paper Titles**	**Authors**
1	Noise-Augmented CBAM-UNet for Robust Ultrasound Liver Tumor Segmentation	Minsun Kim; Seokmin Hong; Byungjeon Kang
2	Registration Between Physical Model and Virtual Model via Image Marker in MR Device	Jingyu Kim; Seokmin Hong; Byungjeon Kang
3	Discriminative Correlation Filter-Based Object Tracking: Recent Research Trends and Challenges	Jeongwoo Lee; Pileun Kim
4	Entity-Aware Trajectory Refinement : Next Best View Based 3D Gaussian Splatting	Namjun Park; Pileun Kim
5	Investigating the Accuracy-Latency Trade-off for Real-Time Open-Vocabulary Models on an Edge Device	Jongyoon Park; Daeil Ko; Pileun Kim
6	Robust Path Navigation via Reinforcement Learning	Daeyoel Kang; Jongyoon Park; Pileun Kim

Conference Program - Day 2 (28/10/2025)

Time	Events	Location
8:00 - 8:40	Conference Registration	Sarah 2
8:40 - 9:00	Welcoming Ceremony	Hallah 1
9:00 - 10:00	Keynote Speech "***Autonomous Navigation and Localisation: A Tale of Two Viewpoints***" by Professor Girish N. Nair from The University of Melbourne	Hallah 1
10:00 - 10:20	Morning Tea Break	Sarah 2
10:20 - 12:00 (Parallel Session 1)	Session A1: Tracking	Hallah 1
10:20 - 12:00 (Parallel Session 2)	Session A2: Radar	Hallah 3
10:20 - 12:00 (Parallel Session 3)	Session A3: Robot Control	Sarah 1
12:00 - 13:30	Lunch Break	Hotel Restaurant
13:30 - 14:50 (Parallel Session 1)	Session B1: Deep Learning for Multi-Object System	Hallah 1
13:30 - 14:50 (Parallel Session 2)	Session B2: Machine Learning 1	Hallah 3
13:30 - 14:50 (Parallel Session 3)	Session B3: Sensor Processing	Sarah 1
14:50 - 15:20	Afternoon Tea Break	Sarah 2
15:20 - 17:00 (Parallel Session 1)	Session C1: Finite Set Statistics	Hallah 1
15:20 - 17:00 (Parallel Session 2)	Session C2: Control and Navigation	Hallah 3
18:30 - Late	Conference Banquet	Sarah 2

Session A1: Tracking (28/10/2025, 10:20 – 12:00, 100 minutes)

Note: Each paper has 15 minutes to present and 5 minutes for Q&A		
Order	**Paper Titles**	**Authors**
1	Adaptive Tracking Method for Extended Target Based on Gaussian Processes	Jie Shi; Xiaomeng Cao; Weifeng Liu
2	A Labeled GM-PHD Filter with Target Amplitude Information	Shunyi Luo; Linfeng Xu; Peijie Yang
3	Track-Before-Detect with Stereo Microphone Arrays	Jiazheng Wang; Mohammad Fard; Amirali Khodadadian Gostar; Reza Hoseinnezhad
4	Covariance Intersection-Based Distributed Fusion for Heterogeneous Sensor Networks with Multi-Target Tracking in Clutter	Jiazheng Li; Ye Yuan; Yuhuan Xiong
5	MD-IPDA Combined with GP for ETT	Taek Lyul Song; Hoseok Sul; Dongsheng Yang; Yunfei Guo; Jee Woong Choi

Session A2: Radar (28/10/2025, 10:20 – 12:00, 100 minutes)

Note: Each paper has 15 minutes to present and 5 minutes for Q&A		
Order	**Paper Titles**	**Authors**
1	Radar Detection and Tracking Method for Extended Target Based on Closed-Loop Feedback	Yi Luo; Mingyu Zhong; Haiyang Liu; Qinyao Chang; Wujun Li; Wei Yi
2	Auxiliary-Guided Subspace-Aligned Tensor Completion for Radar Data Recovery	Lijun Liang; Tiancheng Li; Yan Song; Pablo Chamoso
3	Joint Data Transmission Mode Selection and Power Allocation for Multi-Target Tracking in Radar Network	Hao Jiao; Peng Zhang; Junkun Yan; Chang Gao; Hongwei Liu; Yun Zhu
4	Scanning Sector Selection Scheme in Multi-radar System under Dynamic Interference Environment	Yuan Tang; Peng Zhang; Junkun Yan; Rongrong Wang; Tianyi Jia; Bo Jiu; Hongwei Liu
5	Cartesian and Polar Formulations of the NCV Model in Radar Tracking	Mahendra Mallick; Linfeng Xu; Xiaoqing Tian

Session A3: Robot Control (28/10/2025, 10:20 – 12:00, 100 minutes)

Note: Each paper has 15 minutes to present and 5 minutes for Q&A		
Order	**Paper Titles**	**Authors**
1	Energy-Efficient Drone Base Station Deployment Using Multi-Objective Dragonfly Algorithm	Chee Wei Nea; Ying Loong Lee; Feng Ke; Kok Chin Khor
2	Harmonic-Model-Based Wiggling Correction for Robust Indirect Time-of-Flight Imaging	Jaehyeok Kim; Jaehwa Jeong
3	A Multi-rotor Drone Trajectory Planning Method Based on Energy Optimization	Jiangting Wang; Weifeng Liu; Linqing He
4	Adaptive Tracking Control for Wheeled Mobile Robots Using Sliding Mode Control and Disturbance Observer	Thi Mai Do; Nam Hoai Nguyen; Phuoc Doan Nguyen; Hoa Dinh Nguyen
5	A Low-Code Methodology for Developing AI Kiosks: a Case Study with the DIZEST Platform	SunMin Moon; Jangwon Gim; Chaerin Kim; Yeeun Kim; YoungJoo Kim; Kang Choi

Session B1: Deep Learning for Multi-Object System (28/10/2025, 13:30 – 14:50, 80 minutes)

Note: Each paper has 15 minutes to present and 5 minutes for Q&A		
Order	**Paper Titles**	**Authors**
1	ATSNet-Based Joint Track-Before-Detect and Classification for Extended Target	Qinyao Chang; Miao Li; Xiaolong Li; Yi Luo; Wujun Li; Wei Yi
2	Low-Overlap Point Cloud Registration via Pre-Trained Implicit Autoencoders	Francisco Molina; Martin Adams
3	Behavior Recognition Based on Multimodal Information of Unmanned Aerial Vehicle Swarm of DeepSeek V3	Xuanyuan Shi; Weifeng Liu; Wanyu Li
4	Multi-Agent Planning for Pursuing Multiple Objects via Proximal Policy Optimisation	Hoa Van Nguyen; Diluka Moratuwage; Tran Thien Dat Nguyen; Changbeom Shim; Ba-Ngu Vo

Session B2: Machine Learning 1 (28/10/2025, 13:30 – 14:50, 80 minutes)

Note: Each paper has 15 minutes to present and 5 minutes for Q&A

Order	Paper Titles	Authors
1	Tuberculosis Lesion Classification and Detection Method Guided by Attention Mechanism	Mingli Lu; Haojie Xie; Benlian Xu; Mingming Wang; Jing Liu; Jinliang Cong
2	Design and Implementation of an AI-Powered Robotic System for Laser Cutting Application in Nuclear Power Plant	Rajat Jayantilal Rathod; Himanshu K. Patel; Priyank Jayantilal Rathod; Meet Parmar
3	SafeDoc Agent For Facility Safety Report Evaluation	Jaehyun Kim; Gayeong Kim; Jaemoon Lee; Woosuk Choi
4	Robust Small Methane Plume Segmentation in Satellite Imagery	Khai Duc Minh Tran; Hoa Van Nguyen; Aimuni Binti Muhammad Rawi; Hareeshrao Athinarayanarao; Ba-Ngu Vo

Session B3: Sensor Processing (28/10/2025, 13:30 – 14:50, 80 minutes)

Note: Each paper has 15 minutes to present and 5 minutes for Q&A

Order	Paper Titles	Authors
1	Subband Matrix Information Geometry Based Radar Extended Target Detection in Clutter Environment	Minjing Yan; Wujun Li; Xiaolong Wan; Yunlian Tian; Wei Yi
2	3D Ground Truth Reconstruction from Multi-Camera Annotations Using UKF	Linh Van Ma; Unse Fatima; Tepy Sokun Chriv; Haroon Imran; Moongu Jeon
3	Performance Evaluation of Large Language Models for Humidity Forecasting in Shipyards	Hyunji Kim; JaeSum Lee; Daekyeom Lee; Jaemoon Lee; Kang Choi
4	Doppler-Assisted Automotive Radar Data Clustering Under Micro-Doppler Effect	Jing Zeng; Aranee Balachandran; Ratnasingham Tharmarasa

Session C1: Finite Set Statistics (28/10/2025, 15:20 – 17:00, 100 minutes)

Note: Each paper has 15 minutes to present and 5 minutes for Q&A		
Order	**Paper Titles**	**Authors**
1	Unifying the SLAM Back and Front Ends: Joint, Vector-Set SLAM	Felipe Inostroza; Martin Adams
2	CADA: Confidence-Aware Dynamic Association for Multi-Object Tracking in Complex Environments	Hao Wang; Yun Zhu; Shuang Liang; Wanying Zhang; Li Zhao
3	Multi-Sensor Multi-Object Trajectory Estimation with Lineage Inference	Tran Thien Dat Nguyen; Hoa Van Nguyen; Changbeom Shim
4	A Lightning-Fast Sensor Control Algorithm for Swarm Tracking	Tharani Rajapaksha; Amirali Khodadadian Gostar; Aidan Blair; Reza Hoseinnezhad
5	Real-Time Intrusion Detection of Ghost Attacks in Distributed Multi-Object Tracking Without Offline Pre-Training	Vahid Ghorbani; Amirali Khodadadian Gostar; Amir Ghorbani; Zahir Tari; Aidan Blair; Reza Hoseinnezhad

Session C2: Control and Navigation (28/10/2025, 15:20 – 17:00, 100 minutes)

Note: Each paper has 15 minutes to present and 5 minutes for Q&A		
Order	**Paper Titles**	**Authors**
1	A Three-Dimensional Path Planning Algorithm for Low-Altitude Transportation	Weifeng Liu; Zijun Hu; Shihao Zhao
2	Finite-Time Control Barrier Functions for Safety-Critical Control of Discrete Systems with Application to Robot Navigation	Liang Gao; Xiongchang Li; Zongyang Lv; Yuhu Wu
3	On Finite Bayesian Potential Games	Shuting Le; Yuhu Wu; Benlian Xu; Zhengtian Wu
4	Discrete Position Estimation of Mini 4WD AI Vehicle on a Course by Dempster-Shafer Theory Based State Space Modeling	Norikazu Ikoma
5	A Riemannian Geometric Approach to Space Debris Orbit Determination	Leonardo Cament; Pablo Barrios; Martin Adams

Conference Program - Day 3 (29/10/2025)

Time	Events	Location
8:20 - 9:00	Conference Registration	Sarah 2
9:00 - 10:00	Plenary Speech "***Commercialization of Medical Microrobotics***" by President Jong-Oh Park from Korea Institute of Medical Microrobotics	Hallah 1
10:00 - 10:20	Morning Tea Break	Sarah 2
10:20 - 12:00 (Parallel Session 1)	Session D1: Machine Learning 2	Hallah 1
10:20 - 12:00 (Parallel Session 2)	Session D2: AI for Healthcare	Hallah 3
10:20 - 11:50 (Parallel Session 3)	Session D3: KAU-BK21 Special Session	Sarah 1

Session D1: Machine Learning 2 (29/10/2025, 10:20 – 11:50, 90 minutes)

Note: Each paper has 12 minutes to present and 3 minutes for Q&A		
Order	**Paper Titles**	**Authors**
1	Spatially Gridded-Input LSTM for Streamflow Forecasting: An Evaluation Across Diverse Hydrologic Regimes	Kamilla Rakhymbek; Almas Alzhanov; Nurassyl Zhomartkan; Yerzhan Baiburin; Aliya Nugumanova
2	A Two-Stage Deep Learning Network for Super-Resolution Direction-of-Arrival Estimation	Xiaolong Li; Qinyao Chang; Zhen Cui; Wujun Li; Wei Yi
3	Soft Kinematics-Constrained Smooth Trajectory Prediction using a Diffusion Transformer-based Architecture	Vishnu Dev Tripathi; Vaibhav Malviya; Rahul Kala
4	GenRadar: Radar Data Generation via Diffusion Network Conditioned on Multi-Radar Fused Scenes	Zuyuan Guo; Wujun Li; Hongfu Li; Yuxuan Huang; Wei Yi
5	Korean Speech Recognition via Retrieval-Augmented Generation-Based Intent-Aware Keyword Correction	Chaerin Kim; Yeonghun Chae; Hyoungki Ahn; Changbeom Shim

Session D2: AI for Healthcare (29/10/2025, 10:20 – 11:50, 90 minutes)

Note: Each paper has 12 minutes to present and 3 minutes for Q&A		
Order	**Paper Titles**	**Authors**
1	Classification of Driver Gene Mutations from Thoracic CT Images Using Two-step Learning	Shion Watanabe; Tohru Kamiya; Takashi Terasawa; Takatoshi Aoki
2	Automated Viability Classification of Human Lymphocytes via Machine Learning-Assisted Digital In-Line Holography	Kang Choi; SunMin Moon; Huijin Rim; Sanghoon Shin; Du Yong Kim; Yeonghun Chae; Sungkyu Seo
3	Haptic Feedback for Telerobotic Ultrasound Procedures: Enhancing Safety Through Force Limiting	Meet Parmar; Himanshu K. Patel; Rajat Jayantilal Rathod
4	Real-Time Colonoscopy Simulation and Expert Data Acquisition for Inverse Reinforcement Learning	Noheun Myeong; Pileun Kim

Session D3: KAU-BK21 Special Session (29/10/2025, 10:20 – 11:50, 90 minutes)

Note: Each paper has 8-10 minutes to present and 3 minutes for Q&A

Order	Paper Titles	Authors
1	Performance Enhancement of α-Ga2O3 MSM X-ray Detectors via Asymmetric Pt/Ti Electrode Engineering	Hyeongju Cha, Sunjae Kim, Ji-Hyeon Park, Dae-Woo Jeon, and Wan Sik Hwang
2	Digital Twin-based Design and Simulation of a Self-sensing Gripper Finger for Adaptive Grasping	Geonjun Kim, Hyun-Jin Baek, Sukhwa Lee, Sungku Kang, Hae-Sung Yoon
3	Experimental Study on Switching Control Scheme of Folding-Wing UAV for Flight Mode Transition	Seunghyun Choi, Dongwon Jung
4	Flight Load Derivation of a Lift & Cruise eVTOL Aircraft Based on V–n Diagram Analysis	Myeon-Gi Kim, Jinho Shin, Yun-Sang Lee, Sang-Woo Kim, Moonsung Cho
5	Finite Element Analysis of the Thermal Response of Glass Fiber/Vinyl Ester for Drone's Outer Skin	Seong-Hun Jeon, Myeong-Gi Kim, Sang-Woo Kim
6	Ga_2O_3 Schottky Barrier Diodes as Dual Platforms for X-ray Detection and Natural Material Shielding	Wansik Hwang

Part A: Regular Papers

Session A1: Tracking

A1-1. Adaptive Tracking Method for Extended Target Based on Gaussian Processes

Authors: Jie Shi, Xiaomeng Cao, and Weifeng Liu

Abstract. To address the challenge of tracking group targets whose shapes undergo sudden changes in dynamic environments, this paper proposes an adaptive tracking method based on Gaussian Processes (GP) and Extended Kalman Filtering (EKF). By combining Gaussian Process regression with Bayesian inference, a sliding-window mechanism is designed to capture shape mutation information in real time during target motion. Using prior information, the contour is re-estimated, enabling rapid adjustment of the tracking model's parameter configuration. This achieves adaptive contour updates and effectively improves the accuracy and real-time performance of tracking targets with variable shapes.

A1-2. A Labeled GM-PHD Filter with Target Amplitude Information

Authors: Shunyi Luo, Linfeng Xu, and Peijie Yang

Abstract. Radar multi-target tracking faces significant challenges in complex environments characterized by heavy clutter and high target maneuvers. The traditional Gaussian mixture probability hypothesis density (GM-PHD) filter offers efficient multiple target tracking through bypassing data association, but it struggles to achieve target identities and often fails in scenarios with dense clutter or closely spaced targets. To address these issues, this paper proposes an improved GM-PHD filter by incorporating amplitude information into the weight update process to improve discrimination between targets and clutter in data association. In addition, a threshold-based clustering strategy based on kinematic similarity is introduced to enhance the separation of closely spaced targets and maintain label consistency. Simulation results show that the proposed method achieves improved identity recognition, tracking accuracy, and trajectory continuity in cluttered and dense-target environments compared with conventional GM-PHD and labeled GM-PHD filters.

A1-3. Track-Before-Detect with Stereo Microphone Arrays

Authors: Jiazheng Wang, Mohammad Fard, Amirali Khodadadian Gostar, and Reza Hoseinnezhad

Abstract. Situational awareness of autonomous vehicles in urban environments heavily relies on line-of-sight sensors such as cameras, lidar, and radar. However, these visual sensors may fail to detect targets when they are occluded. In contrast, audio signals have stronger penetration capabilities, enabling human to perceive the direction-of-arrival (DoA) of sound. This directional information can also be extracted using beamforming techniques applied to multi-channel microphone arrays. In this work, we utilize a stereo microphone array to track the 2D position of a single vehicle. Such a minimal array configuration is susceptible to ambient noise and vehicle sounds leading to high variance in position estimation. To address this issue, we propose a track-before-detect (TBD) based particle filter algorithm that incorporates stereo microphone measurements for 2D vehicle tracking. The proposed method is validated using a custom dataset that includes scenarios where vehicles pass at various distances from the microphone array. Using bird's-eye view camera footage as ground truth, experimental results demonstrate that our method achieves accurate and low-variance 2D sound source tracking.

A1-4. Covariance Intersection-Based Distributed Fusion for Heterogeneous Sensor Networks with Multi-Target Tracking in Clutter

Authors: Jiazheng Li, Ye Yuan, and Yuhuan Xiong

Abstract. With increasing demands on multi-target tracking (MTT) accuracy in complex environments, this paper proposes a distributed heterogeneous sensor network based on the covariance intersection (CI) fusion criterion, integrating radar and infrared (IR) sensors. First, we establish target motion and sensor measurement models, and unify the measurement dimensions and formats for cross-sensor compatibility. Unlike single-target tracking, MTT under cluttered conditions requires resolving measurement-to-track associations. To address this, the joint probabilistic data association (JPDA) algorithm is implemented locally at each sensor node for target state estimation and update. To fuse these local estimates without assuming inter-node correlation, the CI criterion is employed. Monte Carlo simulation results validate the effectiveness and accuracy of the proposed method in cluttered MTT scenarios.

A1-5. MD-IPDA Combined with GP for ETT

Authors: Taek Lyul Song, Hoseok Sul, Dongsheng Yang, Yunfei Guo, and Jee Woong Choi

Abstract. Target tracking using high-resolution sensors often suffers from multiple detections generated by a single target, leading to a typical multiple detection (MD) problem that requires distinguishing between true target detections and clutter. To address this issue, a Multiple Detection-Integrated Probabilistic Data Association (MD-IPDA) algorithm is combined with a Gaussian Process (GP) for robust single-target tracking in cluttered environments. The MD-IPDA framework utilizes the target existence probability (TEP) as a track scoring mechanism to ensure reliable track maintenance. Meanwhile, GP is employed to construct measurement model for multiple scattering points, based on predefined contour points determined by the target's shape. The performance of the proposed extended target tracking (ETT) method is validated through simulation experiments and further assessed using the information reduction factor (IRF), allowing direct comparison with existing ETT algorithms.

Session A2: Radar

A2-1. Radar Detection and Tracking Method for Extended Target Based on Closed-Loop Feedback

Authors: Yi Luo, Mingyu Zhong, Haiyang Liu, Qinyao Chang, Wujun Li, and Wei Yi

Abstract. Millimeter-wave radar systems are extensively applied in intelligent traffic systems (ITS) due to their superior sensing capabilities. As resolution improves, a single detected target may occupy several radar resolution cells simultaneously, known as the extended target (ET) effect. Conventional radar target detection and tracking systems employ the detect-before-track (DBT) framework. However, existing DBT frameworks treat detection and tracking algorithms as independent steps, and their performance degrades significantly when encountering complex scenarios with several nearby targets due to interference from both the ET's own neighboring cells and other ETs. To address this challenge, this paper proposes a closed-loop detection-tracking framework exploiting radar ET tracking information. By incorporating a feedback mechanism that utilizes the obtained target shape and motion information from tracking modules into the detection process, the proposed method effectively mitigates inter-target echo interference, achieving mutually reinforcing improvements in both detection and tracking performance. Simulation results demonstrate that the proposed framework significantly enhances detection and tracking capabilities in complex multi-extended-target scenarios compared to conventional DBT approaches.

A2-2. Auxiliary-Guided Subspace-Aligned Tensor Completion for Radar Data Recovery

Authors: Lijun Liang, Tiancheng Li, Yan Song, and Pablo Chamoso

Abstract. To address the critical data loss problem in multichannel radar systems in complex electromagnetic environments, we propose a subspace-aligned tensor completion framework that effectively incorporates auxiliary information. Conventional matrix models flatten high-dimensional data into two-dimensional representations, disregarding the intrinsic low-rank structures of multidimensional data. Furthermore, existing completion algorithms predominantly depend on the primary radar's observations, leading to information scarcity that compromises reconstruction accuracy under severe data loss. To overcome these limitations, we develop a Tucker decomposition-based joint optimization framework that simultaneously enforces low-rankness of the three-dimensional observation tensor via weighted nuclear norm regularization, integrates preprocessed auxiliary radar tensors as reconstruction priors, and establishes a factor matrix alignment mechanism to enhance structural consistency. Theoretically, we establish an upper bound on the recovery error with four components: low-rank optimization error, primary auxiliary tensor deviation, subspace alignment error, and observation noise. Simulations verify that our method consistently outperforms both standard low-rank completion techniques and naive fusion approaches across various observation ratios. Notably, the empirical error strictly complies with the derived theoretical bound, validating the efficacy of our integrated framework, which combines auxiliary information guidance with subspace alignment.

A2-3. Joint Data Transmission Mode Selection and Power Allocation for Multi-Target Tracking in Radar Network

Authors: Hao Jiao, Peng Zhang, Junkun Yan, Chang Gao, Hongwei Liu, and Yun Zhu

Abstract. To address the issues of high communication overhead and information degradation in conventional radar networks (RNs) relying on static fusion architectures, this paper proposes a joint data transmission mode selection and power allocation scheme. This scheme aims to maximize multi-target tracking performance under constraints on communication bandwidth at the fusion center and node transmission power. Firstly, we derives a Bayesian Cramér-Rao Lower Bound of mixed fusion (BCRLB-MF) to quantify the joint impact of power allocation and data transmission mode selection on tracking accuracy under detection uncertainty conditions. Secondly, based on the normalized BCRLB-MF, a mixed-integer nonlinear programming model is established to achieve coordinated optimization of data transmission modes and power. Lastly, a fast solver is designed using the L2-box alternating direction method of multipliers (ADMM) algorithm. Simulation results indicate that the proposed scheme can effectively dynamically balance resource utilization and enhance the tracking performance of RN for targets.

A2-4. Scanning Sector Selection Scheme in Multi-radar System under Dynamic Interference Environment

Authors: Yuan Tang, Peng Zhang, Junkun Yan, Rongrong Wang, Tianyi Jia, Bo Jiu, and Hongwei Liu

Abstract. This paper proposes a scanning sector selection scheme in multi-radar systems to collaboratively enhance target area coverage in the presence of dynamic jamming. Specifically, a scanning sector selection method is developed within a reinforcement learning (RL) framework, modeled as a Markov Decision Process (MDP). By employing a Deep Q-Network (DQN), the system learns an optimal strategy to guide the radar in selecting appropriate scanning sectors at each time step. Simulation results demonstrate that the proposed method effectively enables the radar to adapt its scanning sectors in response to dynamic interference, thereby significantly improving target area coverage.

A2-5. Cartesian and Polar Formulations of the NCV Model in Radar Tracking

Authors: Mahendra Mallic, Linfeng Xu, and Xiaoqing Tian

Abstract. The nearly constant velocity (NCV) model in 2-D is widely used in radar tracking using the range and azimuth measurements. The NCV model is linear and the measurement model is nonlinear in the Cartesian coordinates. In contrast, if a polar state vector is used with components range, azimuth, range-rate, and azimuth-rate, then the dynamic model for the NCV motion becomes nonlinear, and the measurement model is linear. Some existing publications adopt a polar state vector for the NCV motion but assume a linear dynamic model. This formulation actually represents a motion with a nearly constant range-rate and nearly constant azimuth-rate, rather than the NCV motion. In this paper, we examine the accuracy of this approximate model using a maritime target tracking scenario. We use a cubature Kalman filter for the filtering problem and calculate the root mean square errors and posterior Cramér-Rao lower bound for filter evaluation. Monte Carlo simulation results demonstrate that the approximate linear model using a polar state vector yields less accurate state estimates compared with the linear dynamic model and nonlinear measurement model using a Cartesian state vector.

Session A3: Robot Control

A3-1. Energy-Efficient Drone Base Station Deployment Using Multi-Objective Dragonfly Algorithm

Authors: Chee Wei Nea, Ying Loong Lee, Feng Ke, and Kok Chin Khor

Abstract. This paper presents a multi-objective dragonfly algorithm (MODA)-based drone base station (DBS) placement scheme aimed at improving energy efficiency in aerial mobile networks. Unlike conventional single-objective approaches that focus solely on maximizing coverage, the proposed method simultaneously optimizes both network coverage and power consumption. Firstly, we formulate a bi-objective optimization problem that targets to jointly maximize network capacity and minimize transmit power consumption of the DBSs. Then, we develop a DBS deployment framework, which first performs user clustering and association with DBSs based on a k-means method, followed by a MODA-based DBS placement algorithm which effectively balances network capacity and power consumption. Simulation results demonstrate that the MODA-based scheme consistently outperforms benchmark algorithms, making it a promising solution for energy-aware DBS-assisted networks.

A3-2. Harmonic-Model-Based Wiggling Correction for Robust Indirect Time-of-Flight Imaging

Authors: Jaehyeok Kim and Jaehwa Jeong

Abstract. Three-dimensional spatial information is widely used across many fields, and the need for measurement techniques that provide high accuracy, precision, and robustness to environmental variation is increasingly critical. This paper addresses wiggling caused by harmonics of the modulation waveform in Indirect Time-of-Flight (iToF) and proposes a method that simultaneously corrects distance and amplitude without hardware modification. We employ a harmonic model designed for real signals and estimate residual harmonic components via prior calibration. We then use Particle Swarm Optimization (PSO) to estimate the model parameters from the calibration data. In the online stage, we apply Newton's method to a single frame, thereby removing wiggling in both distance and amplitude. Experiments at modulation frequencies of 12, 18, and 24 MHz confirm consistency between the model and measurements. At 12 MHz the distance Mean Absolute Error (MAE) decreases from 7.25 cm to 0.85 cm after a single iteration and converges to 0.82 cm after two to three additional iterations, demonstrating convergence speed and reproducibility. Because only four parameters per pixel are stored, memory demand is lower than prior-data-based corrections, and convergence within two or three iterations per frame supports practicality. The approach is also expected to be effective when combined with Multi-Path Interference (MPI) mitigation.

A3-3. A Multi-rotor Drone Trajectory Planning Method Based on Energy Optimization

Authors: Jiangting Wang, Weifeng Liu, and Linqing He

Abstract. In the optimization of complex environmental trajectory, traditional methods are mostly based on time optimization, ignoring the constraints of energy consumption on flight endurance, which makes it difficult to balance endurance and mission efficiency in actual applications. In response to this, this paper proposes a drone trajectory planning method based on energy optimization: in the framework of traditional time optimization objective function, refine energy consumption modeling is integrated, adaptive weight adjustment strategies are innovatively introduced, and the game relationship between flight time and energy consumption is dynamically balanced. The experiment was developed based on the ROS simulation environment, and multiple sets of differentiated time-energy weight ratio parameters were set for comparison and testing, and the correlation curve between time and energy consumption was visualized with the help of Matplotlib. The results show that while ensuring the rationality of flight time, this method can effectively reduce the total energy consumption of drones, providing an optimization paradigm that takes into account timeliness and economy for drone mission planning in actual scenarios such as logistics distribution, patrol and monitoring, and expanding the application ideas of drone trajectory planning in complex environments.

A3-4. Adaptive Tracking Control for Wheeled Mobile Robots Using Sliding Mode Control and Disturbance Observer

Authors: Thi Mai Do, Nam Hoai Nguyen, Phuoc Doan Nguyen, and Hoa Dinh Nguyen

Abstract. This research introduces an adaptive tracking controller for the kinetic loop to solve the trajectory tracking problem for non-holonomic wheeled mobile robots subject to undesired factors including parameters uncertainty and unknown external disturbance. The proposed controller is based on the integral sliding mode control and disturbance observer. It is combined with an existing adaptive controller for the kinematic loop. The proof for stability of the overall system is performed based on the Lyapunov theory. Simulations with comparisons are carried

out to prove the effectiveness of the designed controller. Based on the simulation results, it can be concluded that the proposed control scheme has high capacity of anti-disturbance and anti-uncertainty, significantly reduces the chattering magnitude and improves the tracking performance.

A3-5. A Low-Code Methodology for Developing AI Kiosks: a Case Study with the DIZEST Platform

Authors: SunMin Moon, Jangwon Gim, Chaerin Kim, Yeeun Kim, YoungJoo Kim, and Kang Choi

Abstract. This paper presents a comprehensive study on enhancing kiosk systems through a low-code architecture, with a focus on AI-based implementations. Modern kiosk systems are confronted with significant challenges, including a lack of integration, structural rigidity, performance bottlenecks, and the absence of collaborative frameworks. To overcome these limitations, we propose a DIZEST-based approach methodology, a specialized low-code platform that enables intuitive workflow design and seamless AI integration. Through a comparative analysis with existing platforms, including Jupyter Notebook, ComfyUI, and Orange3, we demonstrate that DIZEST delivers superior performance across key evaluation criteria. Our photo kiosk case study further validates the effectiveness of this approach in improving interoperability, enhancing user experience, and increasing deployment flexibility.

Session B1: Deep Learning for Multi-Object System

B1-1. ATSNet-Based Joint Track-Before-Detect and Classification for Extended Target

Authors: Qinyao Chang, Miao Li, Xiaolong Li, Yi Luo, Wujun Li, and Wei Yi

Abstract. Detection and classification of extended targets in radar systems presents a critical and complex challenge. While the multi-frame track-before-detect (MF-TBD) algorithm is a powerful framework for weak target detection, its traditional design inherently struggles with extended targets. This paper addresses the challenges of detecting and classifying extended targets based on MF-TBD algorithms. Specifically, traditional MF-TBD algorithms are primarily designed for point targets, which inherently limits their detection performance for extended targets and lacks target classification capabilities. To address this, we revise the MF-TBD process. Instead of relying on merit function threshold detection, we utilize accumulated states across regions within each frame to detect targets. Additionally, we introduce a segmentation network to achieve accurate multi-frame extended target detection and classification. For the segmentation results, we design a fusion strategy based on the merit function, enabling a comprehensive target detection output. Numerical results show that the proposed algorithm exhibits good detection performance and segmentation accuracy under complex conditions for weak extended targets.

B1-2. Low-Overlap Point Cloud Registration via Pre-Trained Implicit Autoencoders

Authors: Francisco Molina and Martin Adams

Abstract. Point cloud registration is a fundamental task in computer vision and robotics, aimed at estimating the rigid transformation that aligns two 3D scans of the same scene. In real-world scenarios, this task becomes challenging due to low overlap between scans, which can result from occlusions, viewpoint changes, or inherent sensor limitations. Recent deeplearning-based methods, such as PREDATOR, CoFiNet, and GeoTransformer, have made significant progress on this problem by incorporating attention mechanisms, spatial hierarchies, and geometric priors. However, all these approaches share a common limitation: they learn point-wise descriptors from scratch without explicit guidance about global geometry. This article studies the feasibility of applying the Implicit AutoEncoders (IAE) approach as an unsupervised pre-training stage to obtain more robust geometric representations. By training an encoder to reconstruct implicit surfaces from point clouds, descriptors capturing both local and global information are obtained. The results show that, contrary to expectations, pre-training does not improve and even degrades the quality of correspondences and alignment accuracy. A thorough analysis leads to six reasons why IAEs are not suitable candidates for pre-training registration models.

B1-3. Behavior Recognition Based on Multimodal Information of Unmanned Aerial Vehicle Swarm of DeepSeek V3

Authors: Xuanyuan Shi, Weifeng Liu, and Wanyu Li

Abstract. This study proposes a method for behavior recognition of UAV swarms based on multimodal information using the DeepSeek-V3 large model, aiming to address the issues of accuracy and robustness in UAV swarm behavior recognition in complex dynamic scenarios. By fusing static attributes with dynamic radar measurement data, a unified input representation framework is constructed, and the self-attention mechanism of DeepSeek-V3 is used to model spatiotemporal dependencies. Innovatively, the DeepSeek-V3 large model is proposed for recognizing the behavior of UAV swarms, and a prompt optimization strategy is adopted instead of traditional fine-tuning methods to guide the model to output results of behavior classification, formation matching, and anomaly detection. Experiments show that in a swarm composed of 4 Matrice 300 RTK UAVs, the average accuracy of this method in behavior classification tasks reaches 89.3%, the formation pattern matching degree is 84.7%, the comprehensive F1 score for anomaly detection is 0.84, and the mean absolute error (MAE) for trajectory prediction is 1.15 meters, providing an efficient solution for intelligent collaborative control of UAV swarms.

B1-4. Multi-Agent Planning for Pursuing Multiple Objects via Proximal Policy Optimisation

Authors: Hoa Van Nguyen, Diluka Moratuwage, Tran Thien Dat Nguyen, Changbeom Shim, and Ba-Ngu Vo

Abstract. This paper addresses the challenge of efficient multi-agent planning for capturing multiple dynamic objects using deep reinforcement learning (DRL). We propose a scalable solution based on the multi-agent proximal policy optimisation (MAPPO) with three key innovations. First, we employ a dynamic state representation that adapts to a variable number of objects, ensuring scalability. Second, we introduce an optimal task allocation mechanism using the Hungarian algorithm to provide a cooperative signal for agent assignments. Finally, this is coupled with a multi-component reward function designed to incentivise both rapid and coordinated pursuit. Experiments demonstrate that our method enables agents to efficiently and

collaboratively pursue all objects across various scenarios, significantly outperforming strong multi-agent reinforcement learning (MARL) baselines in both coordination and efficiency. The results highlight a practical pathway toward developing effective, adaptive multi-agent systems for complex, real-world tasks.

Session B2: Machine Learning 1

B2-1. Tuberculosis Lesion Classification and Detection Method Guided by Attention Mechanism

Authors: Mingli Lu, Haojie Xie, Benlian Xu, Mingming Wang, Jing Liu, and Jinliang Cong

Abstract. The precise classification of the detected lesions can help doctors further determine the type of lesions and implement treatment plans, which is of great importance for the later treatment of the disease. Due to the diverse morphology and complex structure of pulmonary tuberculosis lesions, the classification accuracy of the general network is low in the classification process. Based on this problem, a method is proposed to classify and detect pulmonary tuberculosis lesions guided by a normalized attention mechanism. The NA-YOLO precise classification network for lesions was constructed. Firstly, in terms of data augmentation, a multi-scale image label synchronous data augmentation scheme is proposed to balance the number of various lesions. In terms of network construction, the attention modules of the channel and space branches guide the network to focus on the lesion feature areas, further enhancing the extraction of interchannel and local spatial information. Secondly, the loss is optimized through SIoU to reduce the total degree of freedom between images. Ultimately, achieving a precise classification of four common types of pulmonary tuberculosis lesions. The high-precision classification performance of the proposed method was further verified through ablation and comparative experiments.

B2-2. Design and Implementation of an AI-Powered Robotic System for Laser Cutting Application in Nuclear Power Plant

Authors: Rajat Jayantilal Rathod, Himanshu K. Patel, Priyank Jayantilal Rathod, and Meet Parmar

Abstract. This paper describes the functional design and implementation of an AI-powered robotic system to perform the critical laser cut operation in a nuclear power plant. There are a number of laser-cut applications for nuclear power plant operations, but our prime focus is to develop a robotic system for nuclear reactor maintenance, specifically en masse coolant channel replacement work in a pressurised heavy water reactor. Earlier, laser technology was used in various manufacturing industries and some non-working nuclear reactor core operations. Over the last few years, it has been observed that their applications are now being extended to nuclear power plant maintenance, decommissioning, and other plant operations. Recent developments have been made in robotic systems as well, still integration of laser cutting applications still needs integrated autonomous robotic-based solutions for nuclear power plants maintenance work. The traditional system still uses large semi-automated power manipulator systems that are operated manually by radiation workers. The paper has outlined the gap and implemented an Artificial Intelligence and Machine Learning algorithms-based robotic system that can be used for precise movement, path planning, and decision-making capabilities. This research will provide the design and implementation of an AI-powered robotic system focused on laser cutting applications to be used for nuclear power plant maintenance work in a hazardous nuclear environment. Moreover, it will also help to demonstrate experimental capability to collect the data for AI-powered nuclear waste management and disposal technology.

B2-3. SafeDoc Agent For Facility Safety Report Evaluation

Authors: Jaehyun Kim, Gayeong Kim, Jaemoon Lee, and Woosuk Choi

Abstract. Following two major large-scale structural collapse incidents in the 1990s, there emerged an urgent need to establish a systematic framework to ensure the safety of infrastructure and buildings. As a result, the detailed safety inspection system was introduced, and since 2008, the scope of evaluation has expanded to include comprehensive safety assessments. Although reports are generally prepared and submitted in accordance with official evaluation criteria, there are frequent instances where the content or formatting fails to align with the required standards, yet such issues often go unnoticed during submission. To mitigate these risks, the necessity of introducing AI-powered automated verification systems for report evaluation has become increasingly apparent. In this study, we propose an evaluation-guided large language model (LLM) architecture designed to verify whether safety inspection reports are properly structured and complete prior to submission. The proposed method achieved approximately 55% improvement in detecting omitted elements compared to existing approaches, highlighting its effectiveness and demonstrating its strong potential for future application in fine-grained report evaluation through specialized sLLM.

B2-4. Robust Small Methane Plume Segmentation in Satellite Imagery

Authors: Khai Duc Minh Tran, Hoa Van Nguyen, Aimuni Binti Muhammad Rawi, Hareeshrao Athinarayanarao, and Ba-Ngu Vo

Abstract. This paper tackles the challenging problem of detecting methane plumes, a potent greenhouse gas, using Sentinel-2 imagery. This contributes to the mitigation of rapid climate change. We propose a novel deep learning solution based on U-Net with a ResNet34 encoder, integrating dual spectral enhancement techniques (Varon ratio and Sanchez regression) to optimise input features for heightened sensitivity. A key achievement is the ability to detect small plumes down to 400 m^2 (i.e., for a single pixel at 20 m resolution), surpassing traditional methods limited to larger plumes. Experiments show our approach achieves a 78.39% F1-score on the validation set, demonstrating superior performance in sensitivity and precision over existing remote sensing techniques for automated methane monitoring, especially for small plumes.

Session B3: Sensor Processing

B3-1. Subband Matrix Information Geometry Based Radar Extended Target Detection in Clutter Environment

Authors: Minjing Yan, Wujun Li, Xiaolong Wan, Yunlian Tian, and Wei Yi

Abstract. In high-resolution radar systems, the target energy is frequently distributed across multiple range cells, and the presence of clutter in the detection environment further complicates the situation, leading to substantial degradation in the signal-to-clutter ratio. These problems significantly increase the challenge of detecting extended targets in clutter backgrounds. Traditional extended target detection methods predominantly assume noise-dominated environments and often neglect clutter influences. In contrast, matrix information geometry (MIG) possesses inherent advantages for clutter suppression. However, most existing

MIG detection algorithms are designed for point targets, disregarding the spatial distribution characteristics of extended targets. Accordingly, this paper proposes an extended target matrix information geometry (ET-MIG) detection algorithm, which fully exploits the feature information across multiple range cells. By constructing representative coordinates in the manifold space, the proposed method more accurately captures the distribution characteristics of extended targets. It also provides a more precise geometric measure of the difference between clutter and targets, thereby improving detection performance. Both simulation and semi-measured results demonstrate that the ET-MIG algorithm exhibits significantly superior detection performance compared to the MIG algorithm.

B3-2. 3D Ground Truth Reconstruction from Multi-Camera Annotations Using UKF

Authors: Linh Van Ma, Unse Fatima, Tepy Sokun Chriv, Haroon Imran, and Moongu Jeon

Abstract. Accurate 3D ground truth estimation is critical for applications such as autonomous navigation, surveillance, and robotics. This paper introduces a novel method that uses an Unscented Kalman Filter (UKF) to fuse 2D bounding box or pose keypoint ground truth annotations from multiple calibrated cameras into accurate 3D ground truth. By leveraging human-annotated ground-truth 2D, our proposed method, a multi-camera single-object tracking algorithm, transforms 2D image coordinates into robust 3D world coordinates through homography-based projection and UKF-based fusion. Our proposed algorithm processes multi-view data to estimate object positions and shapes while effectively handling challenges such as occlusion. We evaluate our method on the CMC, Wildtrack, and Panoptic datasets, demonstrating high accuracy in 3D localization compared to the available 3D ground truth. Unlike existing approaches that provide only ground-plane information, our method also outputs the full 3D shape of each object. Additionally, the algorithm offers a scalable and fully automatic solution for multi-camera systems using only 2D image annotations.

B3-3. Performance Evaluation of Large Language Models for Humidity Forecasting in Shipyards

Authors: Hyunji Kim, JaeSum Lee, Daekyeom Lee, Jaemoon Lee, and Kang Choi

Abstract. Shipyard painting operations are highly sensitive to ambient humidity, necessitating accurate forecasting for effective process control. Traditional forecasting methods, such as recurrent neural networks (RNNs) and long short-term memory, typically perform well under stable conditions but struggle with fluctuating environmental dynamics. This work studies large language models (LLMs), specifically Gemma variants, for hourly humidity forecasting using real-time sensor data from multiple shipyard zones. The experimental results show that Gemma-12B significantly improves forecasting accuracy under more variable autumn conditions, while the RNN-based method generally outperforms all Gemma variants during stable early-summer conditions. Specifically, Gemma-12B achieves substantially lower mean squared error, demonstrating its superior capability in capturing non-linear temporal dependencies inherent in dynamic humidity patterns. These results indicate the robustness of LLMs in such contexts and their promising applicability to industrial forecasting tasks.

B3-4. Doppler-Assisted Automotive Radar Data Clustering Under Micro-Doppler Effect

Authors: Jing Zeng, Aranee Balachandran, and Ratnasingham Tharmarasa

Abstract. Modern radars have high resolution and produce multiple measurements from each target in a measurement cycle. Clustering radar measurements is a significant and challenging problem. Utilizing Doppler information is an effective method to improve clustering accuracy, while the presence of the micro-Doppler effect could significantly affect the clustering accuracy. In this paper, a two-step clustering algorithm to group micro-Doppler measurements with the measurements from the target body is proposed by accounting for the tire position constraints and maximizing range rate likelihood. Matlab simulated results demonstrate that our proposed method significantly reduces computation time by 99% while improving overall clustering accuracy by 2% compared to our previous clustering approach.

Session C1: Finite Set Statistics

C1-1. Unifying the SLAM Back and Front Ends: Joint, Vector-Set SLAM

Authors: Felipe Inostroza and Martin Adams

Abstract. This paper introduces a mathematical distribution, based on Bayesian Random Finite Sets (RFSs), to represent the Simultaneous Localization And Mapping (SLAM) posterior. Referred to as the Joint Vector-Set (JVS) distribution, it combines a vector state modelling a vehicle's trajectory and a set state representing a feature based map. In contrast to state of the art feature based SLAM approaches, Joint Vector-Set SLAM (JVS-SLAM) does not require specific measurement to estimated feature association decisions, nor map management routines, usually considered to comprise the SLAM front end. Effectively unifying the SLAM back and front ends within a single set based Bayesian recursion, this article demonstrates that JVS-SLAM provides robust performance in the presence of motion model noise when compared with state of the art solutions.

C1-2. CADA: Confidence-Aware Dynamic Association for Multi-Object Tracking in Complex Environments

Authors: Hao Wang, Yun Zhu, Shuang Liang, Wanying Zhang, and Li Zhao

Abstract. Existing online visual Multi-Object Tracking (MOT) in complex scenes requires the effective fusion of different features. However, existing methods often employ static feature fusion strategies, which adapt to dynamic challenges with difficulty such as occlusions or erratic movements, leading to frequent identity switches. To address this limitation, we propose a novel Confidence-Aware Dynamic Association strategy built upon the Generalized Labeled Multi-Bernoulli (GLMB) filtering framework. The core of our approach is a dynamic fusion module that adaptively adjusts the weights between motion, appearance and spatial costs based on tracking confidence and motion predictability. We further introduce an Enhanced Re-activation Mechanism to reliably recover targets after prolonged occlusions. Evaluations on the MOT16 and MOT20 datasets validate the effectiveness of our approach. Our tracker achieves a significant performance boost over the baseline GLMB method, demonstrating superior identity preservation capabilities and a marked reduction in identity switches.

C1-3. Multi-Sensor Multi-Object Trajectory Estimation with Lineage Inference

Authors: Tran Thien Dat Nguyen, Hoa Van Nguyen, and Changbeom Shim

Abstract. In this paper, we propose an algorithm to estimate object trajectories and lineage information in the presence of spawning events using multiple sensors. For efficient computation, we approximate the predicted multi-object density with object spawning by a generalized labeled multi-Bernoulli (GLMB). To handle multi-sensor observation, we apply the multi-sensor GLMB filter update using Gibbs sampling with an approximate proposal to efficiently sample for significant GLMB components. The experimental results with a 3D tracking scenario demonstrate accurate estimation performance both in terms of object states and lineage information.

C1-4. A Lightning-Fast Sensor Control Algorithm for Swarm Tracking

Authors: Tharani Rajapaksha, Amirali Khodadadian Gostar, Aidan Blair, and Reza Hoseinnezhad

Abstract. This paper introduces a novel sensor control approach tailored for tracking swarms of autonomous agents, such as Unmanned Aerial Vehicles (UAVs). The aim of this proposed method is to track swarms in the presence of a controllable sensor with limited sensor resolution, improving the computational time. A key contribution of this work is the incorporation of swarm information into the formulation of the sensor control problem, thereby improving the computational performance in real-time execution. To address this, the Sequential Monte Carlo (SMC) swarm tracking-based sensor control method is proposed, along with a new likelihood function that integrates swarm characteristics into the pseudo-update step. Our proposed method combines both the SMC-based swarm tracking filter with the Labeled Multi-Bernoulli filter (LMB), enabling efficient and accurate estimation of swarm cardinality, overall dynamics, shape-related information, and individual target trajectories using controllable sensors. Numerical evaluations demonstrate that the proposed approach achieves a significantly improved computational efficiency compared to the LMB filter-based sensor control method, while providing marginal improvements in tracking accuracy.

C1-5. Real-Time Intrusion Detection of Ghost Attacks in Distributed Multi-Object Tracking Without Offline Pre-Training

Authors: Vahid Ghorbani, Amirali Khodadadian Gostar, Amir Ghorbani, Zahir Tari, Aidan Blair, and Reza Hoseinnezhad

Abstract. Leveraging the recently proposed Average Likelihood for Attack Resilient Multi-object (ALARM) filtering method, this study develops an intrusion detection scheme to counter stealthy, measurement-oriented adversarial interventions known as ghost attacks in distributed multi-object tracking systems. In this scheme, each node executes both the ALARM filter and a conventional multi-object filter in parallel; a mismatch in the cardinalities of their respective multi-object densities signals a potential attack. Once an anomaly is detected, the estimated ghost cardinality is obtained by comparing the two filter outputs. Estimated ghost counts are then evaluated against ground-truth using the ℓ_2 error (Euclidean norm of residuals), mean squared error, root mean squared error, mean absolute error, and coefficient of determination (R^2). This approach provides a unified framework for both detecting stealthy adversarial activity and estimating the number of false targets in a distributed multi-object tracking environment.

Session C2: Control and Navigation

C2-1. A Three-Dimensional Path Planning Algorithm for Low-Altitude Transportation

Authors: Weifeng Liu, Zijun Hu, and Shihao Zhao

Abstract. A hierarchical multi-objective optimization method is proposed for path planning in low-altitude logistics transportation. The first-level transportation uses an improved A* algorithm to achieve the shortest path planning for urban trunk transportation, while the second-level transportation employs a 3D A* algorithm combined with a risk cost function to complete obstacle avoidance path planning for end delivery. Simulation results show that the time optimization algorithm for first-level transportation can reduce flight time by 13%, and the risk optimization algorithm for second-level transportation can decrease path risk value by 41.8% while improving search efficiency by 3.85%. This research provides an efficient and safe intelligent path planning solution for low-altitude logistics transportation.

C2-2. Finite-Time Control Barrier Functions for Safety-Critical Control of Discrete Systems with Application to Robot Navigation

Authors: Liang Gao, Xiongchang Li, Zongyang Lv, and Yuhu Wu

Abstract. This paper proposes a unified finite-time control framework for safety-critical control of nonlinear discrete-time systems. First, a novel concept of finite-time discrete control barrier function (FT-DCBF) is introduced to ensure finite-time safety recovery, along with the definition of a strong safety set whose forward invariance is rigorously proven. Subsequently, the discrete control Lyapunov function (DCLF) is integrated with the proposed FT-DCBF within a quadratic programming (QP) framework to construct a unified finite-time safety controller. To demonstrate the effectiveness of the proposed method, it is applied to obstacle avoidance navigation for wheeled mobile robots (WMRs) in complex environments. Under disturbance conditions, the proposed control strategy significantly enhances the navigation safety and robustness of the WMRs. Finally, simulation results validate the efficacy of the proposed control framework.

C2-3. On Finite Bayesian Potential Games

Authors: Shuting Le, Yuhu Wu, Benlian Xu, and Zhengtian Wu

Abstract. In order to effectively solve the equilibrium problem of Bayesian game with finite policy set, the transformation of incomplete information and the establishment of Bayesian potential equation in Bayesian game are proposed in our previous reslut. On this basis, combined with the characteristics that human utility in Bayesian game is not affected by the same type, this paper establishes an improved potential equation and designs the corresponding improved algorithm which can reduce the computational complexity. Secondly, using the characteristics of matrix semi tensor product, an alternative method to solve the potential function of Bayesian potential game is given. Finally, in the framework of Bayesian game, the theoretical results are applied to the deployment of Network Functions Virtualization (NFV) system service function chain with incomplete information, and then provide the best resource allocation scheme and joint resource optimization algorithm in the sense of game equilibrium.

C2-4. Discrete Position Estimation of Mini 4WD AI Vehicle on a Course by Dempster-Shafer Theory Based State Space Modeling

Author: Norikazu Ikoma

Abstract. In a context of "Mini 4WD AI", which equips a microcomputer board to control single DC motor powered by two AA batteries to drive its four wheels simultaneously, vehicle's position estimation over the circuit course is crucial to win "Mini 4WD AI" tournament held by an academic community in Japan. Simplified modeling is mandatory to cope with to the actual Mini 4WD AI vehicle system due to high-speed driving of the vehicle and limited data transfer rate between on-board microcomputer and host PC. We propose to estimate the position in discrete space to specify a certain course-block among all the possible blocks consisting of the course. Dempster-Shafer theory based state space modeling and its state estimation have been employed to deal with uncertainty of the vehicle's position on the course more efficiently. Formulation of the problem in a state space model based on Dempster-Shafer theory, state estimation procedure being possible to work in real-time, and evaluation in simulation experiments have been demonstrated in this paper.

C2-5. A Riemannian Geometric Approach to Space Debris Orbit Determination

Authors: Leonardo Cament, Pablo Barrios, and Martin Adams

Abstract. Tracking and cataloging orbiting Space Objects (SOs) is a critical component of modern Space Situational Awareness (SSA). The unpredictable nature of orbital dynamics, driven by numerous external forces, motivates the use of probabilistic filtering techniques for accurate state estimation. Much of the literature focuses on Gaussian Mixture (GM) or computationally expensive particle based filtering in SSA, however GM representations often fall short in capturing the complex, non-Euclidean structure of orbital motion. Therefore in this work, the Probabilistic Admissible Region (PAR) approach is used for initial orbit determination and, to overcome the limitations of classical GM approaches, we introduce a manifold-based approach that models the SO state on a Riemannian manifold M, allowing for more accurate prediction while maintaining computational tractability. We implement a Cubature Kalman Filter (CKF) predictor on M and validate the approach using orbital data from four CubeSats operated by the Universidad de Chile. Initial results show that the proposed method can be comparable to linear Gaussian prediction methods, particularly in short observation scenarios, and often superior in long-observation scenarios.

Session D1: Machine Learning 2

D1-1. Spatially Gridded-Input LSTM for Streamflow Forecasting: An Evaluation Across Diverse Hydrologic Regimes

Authors: Kamilla Rakhymbek, Almas Alzhanov, Nurassyl Zhomartkan, Yerzhan Baiburin, and Aliya Nugumanova

Abstract. Reliable daily streamflow forecasts are essential for flood-risk management and water-resource planning. We evaluate Long Short-Term Memory models driven by spatially distributed gridded versus point-based ERA5-Land meteorological inputs across four catchments that span snow-dominated, rainfall-driven, and mixed hydro-climatic regimes. For each basin, LSTM-Grid and LSTM-Point models are trained with identical architectures. Gridded inputs consistently improve performance: median Nash-Sutcliffe Efficiency rises from 0.84 to 0.93 in the mountainous Middle Fork Flathead River and from 0.78 to 0.90 in the Uba River, while the rainfall-fed Queets basin shows a moderate gain from 0.62 to 0.69. In the mixed South Santiam basin, performance increases from 0.54 to 0.58, and snow-water equivalent contributes only when supplied in a gridded form. These results demonstrate that spatially distributed predictors capture topographic and climatic heterogeneity missed by single-point inputs, producing more robust forecasts across diverse settings. The findings support operational adoption of gridded-input LSTMs, especially in regions with pronounced spatial variability, while highlighting the diminishing performance in more homogeneous low-relief basins.

D1-2. A Two-Stage Deep Learning Network for Super-Resolution Direction-of-Arrival Estimation

Authors: Xiaolong Li, Qinyao Chang, Zhen Cui, Wujun Li, and Wei Yi

Abstract. Achieving high angular resolution is essential in modern radar systems to distinguish closely spaced targets. Super-resolution direction-of-arrival (DOA) estimation addresses this need by overcoming the Rayleigh resolution limit. Traditional sparse reconstruction methods suffer from high mutual coherence among dictionary columns when the angular grid is densified, resulting in degraded reconstruction accuracy. Although unrolled architectures such as learned iterative shrinkage-thresholding algorithm (LISTA) mitigate this issue to some extent, they remain burdened by a large number of parameters and slow inference. To address these issues, this paper proposes a two-stage network named TS-LM Net, which integrates LISTA and multilayer perceptron (MLP). The proposed network decomposes the DOA estimation task into two sequential stages: a coarse integer-level estimation followed by a fine fractional-level refinement. By avoiding a single-stage mapping over the full angular grid, the proposed network significantly reduces the number of parameters. Simulation results demonstrate that TS-LM Net achieves comparable high estimation accuracy while realizing lower parameter complexity and faster runtime performance.

D1-3. Soft Kinematics-Constrained Smooth Trajectory Prediction using a Diffusion Transformer-based Architecture

Authors: Vishnu Dev Tripathi, Vaibhav Malviya, and Rahul Kala

Abstract. Trajectory prediction is widely used in the navigation of social robotic systems. Model-driven approaches (or Physics-based algorithms) cannot capture the complexity of social compliance, and thus, data-driven approaches are widely used. We study a diffusion-based trajectory prediction algorithm that generates samples of trajectories from a learned distribution. We observe that while the generator keeps the trajectory close to the ground truth trajectory for at least one sample and thus has a small error, most of the samples have zig-zag shapes, sharp turns, are kinematically infeasible, and are unrealistic. These attributes are absent in the model-driven approaches. As the first objective, we force the trajectory generator to output kinematically feasible and smooth trajectories. Since humans can kinematically take sharp turns, our notion of kinematic constraints is soft; smoother turns are preferable, while the model should allow sharp turns if absolutely necessary. This results in more realistic trajectory prediction. We also compare the approach to several state-of-the-art methods. As another objective, we also critically analyse the public datasets for algorithmic ease of handling corner cases like sharp turns, high densities, high/low travel speeds, a large number of people, etc. and thus characterize datasets by complexity in corner cases. Not all publicly available datasets have high social interaction and mostly feature agents going nearly straight. High-interaction datasets have corner cases that are hard to learn, as they represent the heavy-tail problem. We evaluate the datasets specifically on those corner cases. We show that the inclusion of soft kinematics smoothness constraints is particularly valuable for high complexity datasets.

D1-4. GenRadar: Radar Data Generation via Diffusion Network Conditioned on Multi-Radar Fused Scenes

Authors: Zuyuan Guo, Wujun Li, Hongfu Li, Yuxuan Huang, and Wei Yi

Abstract. High-fidelity radar data is essential for the robust validation of radar signal processing, yet the physical collection of diverse, large-scale datasets is often prohibitively expensive, and fails to capture rare, critical scenarios. Furthermore, even simulated datasets are often static and computationally expensive to expand. To address this data scarcity and inflexibility, we propose GenRadar, a data augmentation framework for generative radar data synthesis. Our approach re-frames the challenge as a conditional view synthesis problem, learning to generate realistic Range-Azimuth (RA) radar maps. The core of our method is a two-stage architecture that decouples scene understanding from image generation. First, a Transformer-based encoder fuses contemporaneous data from multiple radars into a single, unified, and viewpoint-agnostic scene representation. This holistic latent vector then serves as a strong condition for a powerful diffusion-based decoder, which synthesizes a sharp, physically plausible RA map. Extensive experiments demonstrate that GenRadar achieves state-of-the-art performance, significantly outperforming existing generative baselines in both quantitative metrics and qualitative realism. By providing a scalable method for transforming fixed datasets into flexible, on-demand generative tools, GenRadar offers a powerful solution for data augmentation and the creation of comprehensive test scenarios.

D1-5. Korean Speech Recognition via Retrieval-Augmented Generation-Based Intent-Aware Keyword Correction

Authors: Chaerin Kim, Yeonghun Chae, Hyoungki Ahn, and Changbeom Shim

Abstract. This study proposes a novel Korean speech recognition system in Kiosks that integrates domain-specific keyword correction and visualization report generation based on Retrieval-Augmented Generation (RAG). Conventional speech recognition for Korean, characterized by its agglutinative structure, numerous homophones, and significant pronunciation variations, often exhibits high error rates, especially when recognizing specialized terminology in the cultural heritage domain. To address these challenges, the proposed system utilizes a keyword correction mechanism that combines a topic-specific keyword dictionary with semantic similarity analysis, thereby refining the recognized speech input. The system automatically generates structured HTML visualization reports based on the corrected queries. Experimental results using 70 speech data samples across seven cultural heritage topics demonstrate that the keyword correction method improved speech recognition accuracy from 45.71% to 91.43%, while the visualization report system significantly enhanced user satisfaction and learning effectiveness compared to traditional text-based responses. This work overcomes structural limitations of Korean speech recognition and demonstrates its practical applicability in cultural heritage education and guidance scenarios.

Session D2: AI for Healthcare

D2-1. Classification of Driver Gene Mutations from Thoracic CT Images Using Two-step Learning

Author: Shion Watanabe, Tohru Kamiya, Takashi Terasawa, and Takatoshi Aoki

Abstract. In recent years, the number of cancer-related deaths has been extremely high, with lung cancer accounting for the largest proportion. Therefore, early detection and treatment are important to reducing the number of lung cancer deaths. The diagnosis and treatment of lung cancer play a role in determining the progression of the disease and selecting a treatment plan. Genetic testing is particularly important because it can confirm the presence or absence of driver gene mutations that affect the growth, invasion, and metastasis of cancer cells. The results determine the course of treatment. When driver gene mutations are confirmed by genetic testing, highly effective, molecularly targeted drugs that target only those molecules become available. However, since driver gene mutations are currently difficult to identify by imaging, testing is performed by an invasive biopsy, which is burdensome to patients. In the field of medical imaging, therefore, it is considered beneficial to construct a computer-aided detection (CAD) system to reduce the burden on physicians and support decision-making by using computer-generated quantitative analysis results as a "second opinion". Thus, we propose a machine learning method to identify driver gene mutations in chest CT images to develop a noninvasive, quick, and low-cost CAD system. This method uses supervised learning with LightGBM and includes radiomics features extracted from chest CT images, as well as features from clinical information, as explanatory variables. Next, we detect the presence or absence of genetic mutations using a two-class classification. We applied the proposed method to 173 chest CT images and achieved the following classification performance: accuracy = 97.7%, true positive rate (TPR) = 97.5%, false positive rate (FPR) = 2.2%, and area under the curve (AUC) = 0.999. These results confirm the method's effectiveness.

D2-2. Automated Viability Classification of Human Lymphocytes via Machine Learning-Assisted Digital In-Line Holography

Authors: Kang Choi, SunMin Moon, Huijin Rim, Sanghoon Shin, Du Yong Kim, Yeonghun Chae, and Sungkyu Seo

Abstract. Digital In-line Holography (DIH) is a compact and label-free imaging technique that captures diffraction-based "shadow" patterns of biological cells without complex optical setups or staining. However, manual analysis of DIH datasets is challenging due to speckle noise, multiple scattering, and twin-image artifacts. To address these challenges, this study introduces an automated machine learning framework that integrates YOLOv8-based cell detection, extraction of physically meaningful diffraction features, i.e., Shadow Parameters, and viability classification through regression models and threshold-based decisions. Experimental validation with human lymphocyte samples identified the width of secondary diffraction maxima (WSM) as the key predictor for cell viability. Among tested regression algorithms, the random forest model performed best ($R^2 = 0.8517$), and using only WSM achieved a classification accuracy of 92.8%, highlighting its potential for real-time, label-free viability assessments. The complete analytical process was implemented using the dizest low-code platform, allowing rapid visual workflow creation without programming, thus offering a practical and scalable solution suitable for clinical diagnostics and industrial bioprocess monitoring.

D2-3. Haptic Feedback for Telerobotic Ultrasound Procedures: Enhancing Safety Through Force Limiting

Authors: Meet Parmar, Himanshu K. Patel, and Rajat Jayantilal Rathod

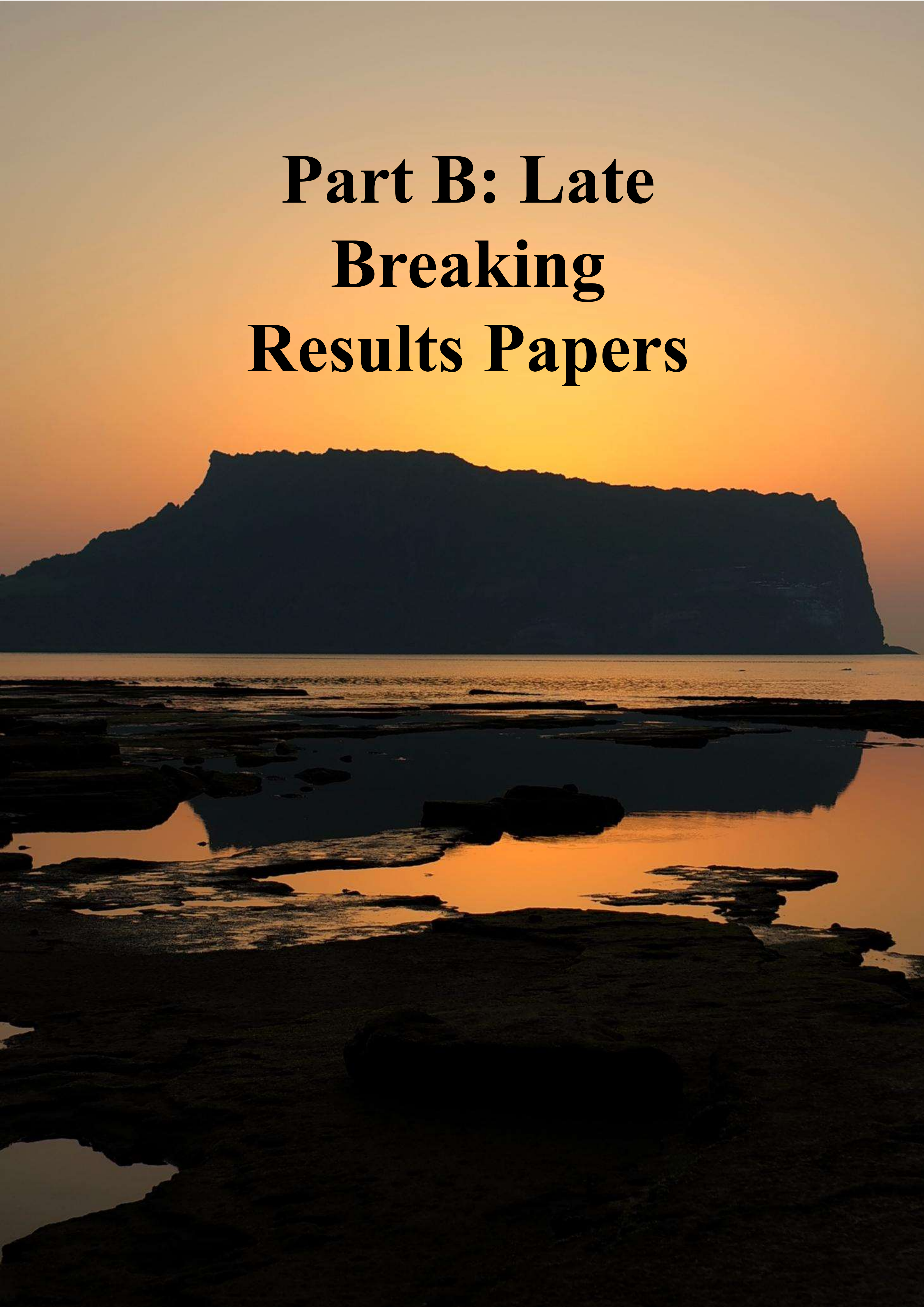
Abstract. The development of telerobotic systems for medical diagnostics has made it easier for people in remote areas and areas that don't get enough healthcare to get it. Ultrasound imaging is an important part of non-invasive diagnostics that depends on precise control of probe pressure to make sure the images are clear and the patient is safe. But because telerobotic ultrasound systems don't give direct tactile feedback, they can apply too much or too little contact force, which can make the diagnosis less effective and the patient less comfortable. This paper talks about a telerobotic ultrasound system with haptic feedback that uses a real-time force-limiting algorithm to give the operator more control and lower the risks that come with using too much force. The system has a master-slave structure. On the patient side, there is a UR5 robotic arm with an ultrasound probe and a force sensor. On the operator side, there is a haptic interface that lets the operator feel the force in both directions. Using tissue phantoms to test the system shows that adding haptic feedback to force limitation greatly cuts down on force overshoot and makes tasks more precise. This makes the system safer and better at doing remote diagnostics.

D2-4. Real-Time Colonoscopy Simulation and Expert Data Acquisition for Inverse Reinforcement Learning

Authors: Noheun Myeong and Pileun Kim

Abstract. This study proposes a real-time simulation system that replicates the physical interactions during colonoscopy using colon and endoscope models implemented in the SOFA framework. The system records expert manipulation data in real time to construct a dataset for future applications in Inverse Reinforcement Learning (IRL). Joystick inputs and endoscopic video frames are synchronized at the frame level, allowing for precise alignment of states and actions. SuperPoint-based feature extraction is used for real-time preprocessing to ensure robust state estimation, even in environments constrained to an RGB camera and lighting. This methodology facilitates the generation of reliable training datasets within anatomically realistic and real-time simulated environments, thereby establishing a foundation for learning autonomous control policies.

Part B: Late Breaking Results Papers



Session L1: Late Breaking Results 1

A Novel User Intent Prediction System for Improved Ad Targeting Strategies

Rian Khetani
Independent Researcher
rian.khetani@gmail.com

Pranav Kulkarni
Stanford University
pranavsk@stanford.edu

***Abstract*—Predicting customer intent in real time is a challenge in ecommerce. Failing to correctly identify whether a user is browsing or on the verge of purchase typically results in inefficient ad targeting and wasted impressions. Accurately determining customer intent allows for more effective ad allocation, saves costs, and improves conversion rates. In this study, we present a novel user intent prediction system that leverages customer interaction logs and item properties to classify sessions based on intent. By modeling these behaviors, the system aims to accurately determine the customer's intent which allows for better aligned ads with the customer's intent and buying stage, leading to higher conversion rates and improved ROI. Using the publicly available Retailrocket recommender system dataset that contains web events, item metadata, and category hierarchy, we employ multi-snapshot augmentation, session-level and prefix-based feature engineering, historical conversion rate features, categorical encoding, leaky feature removal, and median imputation combined with time-aware data splitting and LightGBM with class-imbalance weighting and F1-based threshold optimization. We achieve a high overall accuracy of 0.979, AUC of 0.906, and a PR-AUC of 0.296 (over 4x better than baseline of 0.07). While the high accuracy is driven by the dominance of non-purchase sessions (93% of dataset), the system is able to meaningfully distinguish purchase intent under class imbalance. These results demonstrate that the model can effectively flag valuable users for ad targeting, offering a significant improvement over baselines. The system represents a step toward practical real-time intent detection and is designed to evolve toward deep learning and sequence based models, where the order and timing of interactions (views, clicks, add-to-cart, purchases) may offer richer predictive signals than static features alone. Aside from ad targeting, our system has applications in personalized product recommendation, A/B testing, and general domains like finance or streaming platforms where behavior prediction is critical.**

***Index Terms*—Machine Learning, User Intent Prediction, Marketing Systems**

Introduction

Customer intent prediction has become a critical issue in the realm of digital commerce. With ecommerce sites competing to capture the attention of consumers, the ability to distinguish between casual browsing and serious purchasing interest sessions is critical in maximizing the effectiveness of advertising. Traditional targeting methods are mostly based on general demographic data or past buying data, which often do not capture the subtle indicators of interest that can be observed during a single browsing session. This gap therefore results in the wastage of advertising impressions, overcharging, and missed opportunities to convert high value customers.

Advancements in machine learning have enabled the modeling of customer journeys on a more granular level, using session-based features that model real-time behaviors including page views, clicks, add-to-cart interactions, and purchases. However, the issue of very skewed datasets, when only a relatively small percentage of sessions result in purchases, remains a serious problem. Common classifiers tend to overfit the majority of the non-purchase class, thus providing less optimal recall and precision to users with higher purchase intent.

In this work, we present a real-time user intent prediction system to be used in advertising targeting. Leveraging multi-snapshot augmentation to capture intent signals across different session stages, historical conversion rate features to add population-level priors, and prefix-based modeling to predict intent before a session concludes, our approach predicts customer intent (purchase likelihood) sooner than traditional methods that rely on full session data. Based on the publicly available Retailrocket dataset, containing logs of web interactions, item metadata, and a categorical hierarchy, we create a modelling pipeline integrating extensive feature engineering, categorical encoding, leaky feature removal, and median imputation. A gradient boosting classifier (LightGBM) with threshold optimization is then trained under time-aware data splits to handle the class imbalance. Figure 1 summarizes how intent scores can be leveraged to activate ad bidding/segmentation, on-site personalization, and recommendations to drive higher conversion rate and lower cost per acquisition.

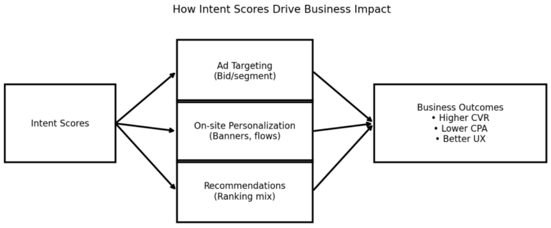


Fig. 1. How Intent Scores Drive Business Impact

Literature Review

Modern recommender systems have evolved from memory-based methods and factorization methods into deep neural architectures that are capable of modeling rich user interactions and item dynamics. Foundational surveys detail this shift and codify it by model class and by data modality highlighting how deep learning unlocks end-to-end feature learning while also displaying challenges such as sparsity, cold start, and evaluation under class imbalance [1]. These reviews also note a widening gap between offline ranking metrics versus real world impact, motivating domain specific objectives such as early intent recognition rather than only item prediction.

Within sequential and session-based recommendation, a central thread is learning signals from click sequences. Convolutional encoders treat recent interactions as a temporal image to capture local patterns [2], while self-attention models (SASRec) selectively weight relevant history and scale to both sparse and dense regimes [3]. Large-scale production systems further these design choices and emphasize pragmatic multi-stage retrieval through ranking pipelines, feature hygiene, and continuous iteration. For example, Pinterest's Related Pins recounts three years of system evolution and shows that simple, well engineered components often dominate early phases [4]. More recent work, such as transformer-based models that fuse transactional and product relations for contextual recommendations, adds context [5]. Together, these works frame the progression from sequence encoders toward context-aware ranking.

Empirical assessments specific to session-based ecommerce reinforce that architecture alone is not density; training regimes, loss design, and preprocessing choice often drive the spread in reported gains. A comprehensive evaluation across deep SBR approaches in retail settings stresses repeatable setups and careful metric selection [6]. That theme of methodological rigor and metric choice is especially salient for highly skewed purchase datasets like the Retailrocket dataset, where PR-AUC is more relevant than accuracy and AUC for minority class detection.

Two fast moving frontiers further reshape this design space. First, large language models (LLMs) are being integrated as representation learners, instruction-following recommenders, and planning agents. Surveys argue for three complementary paradigms: pre-train, fine tune, and prompt. This blueprint maps LLMs onto recommendation tasks [7]. Concrete models include LLMRec, which augments user-item graphs using signals extracted from LLMs [8]; InstructRec, which reframes recommendation as instruction following [9]; and a Recommender AI agent, which utilizes a tool-using, memory-augmented agent around a base recommender [10]. These works suggest opportunities to enrich our proposed system by pairing our prefix-time features with natural language context and combining our model with an agentic layer for intervention policy.

Second, diffusion models have entered recommendation as generative data augmenters, re-rankers, and preference samplers, A 2025 survey synthesizes applications across sequential, multimodal, and knowledge-graph settings [11]. Methods include DiffSBR for session-based sequences [12], Diff-POI for spatial preference sampling [13], and plug-in diffusion modules that apply uncertainty-aware signals into standard pipelines [14]. In parallel, robustness and multimodality advances such as DVIB's information-bottleneck distillation [15] and MTSTRec's time-aligned fusion of text, vision, and price [16] illustrate orthogonal levers our proposed system could possibly adopt for cold start and noisy signal regimes common in early session intent.

Finally, field evidence cautions that recommendation quality and targeting policies have market-level implications beyond top-N accuracy. Both classic and recent studies show that while recommenders can lift individual exploration and sales volume, they may simultaneously concentrate aggregated demand which in turn amplifies "blockbusters" [17]. For purchase intent prediction, this highlights the value of precision-focused, fairness-aware threshold. For example, using PR-AUC guided operating points so that early interventions such as discounts and outreach improve efficiency without sacrificing exposure imbalance.

In sum, prior work establishes robust sequential encoders for early behavior [18], [3]; production lessons on scalable, multi-stage recommenders [4]; the importance of task-aligned metrics under skew [6]; and two catalytic trends: LLM-powered reasoning and interaction [7] and diffusion-based generative augmentation [11]. Our proposed model naturally fits into this arc by pivoting from next-item ranking to early purchase intent detection on session prefixes, optimizing directly for PR-AUC under $\sim 7\%$ prevalence, and demonstrating that targeted, prefix-time features can recognize high-intent sessions early enough to deploy timely, real-time interventions such as targeted advertisements, discounts, or other methods.

METHODOLOGY

Dataset

We used the Retailrocket recommender system dataset, a publicly available ecommerce dataset comprising:

- User behavior logs: page views, clicks, add-to-cart actions, purchases
- Item metadata: product features such as category and properties
- Category hierarchy: hierarchical grouping of items

The dataset contains hundreds of thousands of user sessions. For this study, we applied sessionization and filtering steps to build training examples suited for intent prediction.

Preprocessing and Feature Engineering

The first step was to sessionize the data: user activity was segmented into sessions using a 30-minute inactivity threshold. Next the dataset went through multi-snapshot augmentation. Instead of waiting until sessions ended or only using one prefix cutoff, we created "snapshots" of each session at meaningful cutoffs. Empirically, purchase probability increases with session progress (Figure 2), motivating prefix snapshots so decisions can be made earlier in the session. These snapshots enable intent prediction earlier in the session and expand the training sample. The sessions were cut off when cumulative view counts reached values of 3, 5, 7 or at elapsed times of

60, 180, and 420 seconds with a maximum of three snapshots per session.

We then engineered features based on the dataset such as cumulative statistics, recency features, diversity features, item popularity signals, historical conversion rate features, and calendar and time features. Cumulative statistics included the number of “views,” “clicks,” “carts,” and “purchases.” Recency features included “time since last action” and “time since session start.” Diversity features included “unique categories and items viewed.” Item popularity signals were the global frequency of viewed items, and the historical conversion rate features were the prior conversion rates of items and categories. Calendar and time features included the “hour of day” and “day of week.”

The dataset was then categorically encoded using one-hot encoding (pd.get_dummies) for categorical variables such as “category IDs.” Any features that could directly reveal the intent and allow the model to cheat were then removed, and missing values were handled via median imputation to ensure compatibility with oversampling and LightGBM. We used median imputation rather than dropping sessions, since high-intention sessions were already rare, and dropping them may further skew the imbalance.

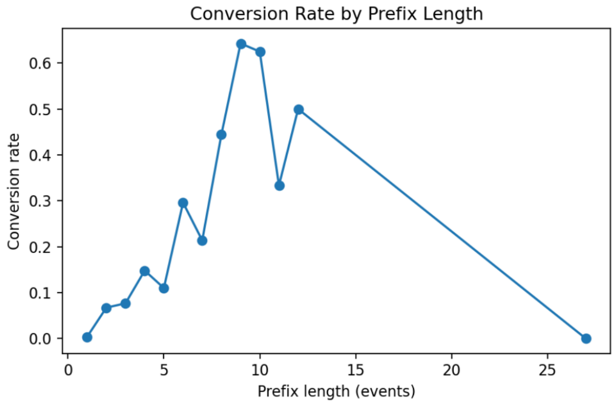


Fig. 2. Conversion Rate by Prefix Length: As the number of events within our session prefixes grows, the conversion rate generally rises, indicating stronger intent later in sessions. The dip at very long prefixes reflects sparse, high variance tail data, not a change in the overall trend. Few sessions last that long; nothing more than an outlier.

Modeling

- Split: Time-aware splitting was used to ensure realistic evaluation, with 80% of sessions for training, 10% for validation, and 10% for testing
- Classifier: LightGBM (LightGBMClassifier) with the following key parameters
  - objective = binary
  - learning_rate = 0.08
  - num_leaves = 63
  - subsample = 0.8, colsample_bytree = 0.8
  - reg_lambda = 0.5
  - n_estimators = 2000 with early stopping
- Class imbalance: Adjusted via scale_pos_weight (set to 8.75 based on imbalance ration tuning)
- Threshold optimization: Decision thresholds tuned on the validation set for maximum F1-score, rather than default 0.5. We visualize precision, recall, and F1 across thresholds and select threshold = 0.679 on the validation split (Figure 3); applied to the test set this operating point yields precision = 0.428 and recall = 0.293.

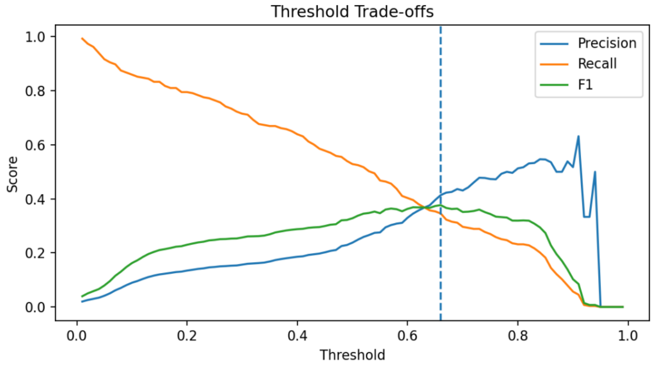


Fig. 3. Threshold Trade-offs: As we raise the decision threshold, precision increases and recall decreases. The green curve shows F1 (different from the Macro-F1 score), which shows the balance of precision and recall. The vertical line shows the threshold we deploy (0.679), which was selected on validation to balance our precision and recall.

Evaluation Metrics

- AUC: Overall ranking quality between high intent vs low intent sessions
- PR-AUC: Emphasis on the high intent class under imbalance
- Macro-F1: Balanced measure of precision and recall across both classes
- Precision & Recall: Reported at the F1-optimal threshold to capture trade-offs
- Accuracy: Reported for complete context, but interpreted cautiously due to class imbalance

The end-to-end pipeline containing sessionization, prefix snapshots with de-leakage, feature store, LightGBM scoring, and the calibrated decision rule is shown in Figure 4.

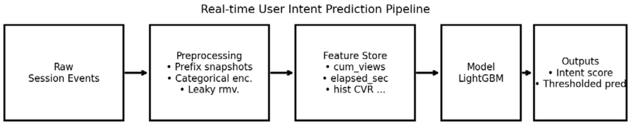


Fig. 4. System Pipeline

RESULTS

We evaluated the proposed system on the Retailrocket dataset using a time-aware train/validation/test split (80/10/10). The final configuration employed LightGBM with class imbalance weighting (scale_pos_weight = 8.75) and F1-based threshold tuning.

The PR curve highlights performance under imbalance (AP = 0.296) and marks the chosen operating point (threshold = 0.679) in Figure 5.

On the test set, the system achieved an **AUC** of **0.906**, an **overall accuracy** of **0.979**, and a **PR-AUC** of **0.296**, over four times higher than the baseline prevalence of 0.07. At the optimized decision threshold (∼0.679), the system obtained a **Macro-F1** score of **0.669**, with **precision** of **0.428** and **recall** of **0.293** for high-intent sessions.

These results demonstrate that, while accuracy is dominated by the majority non-purchase class, the system is able to meaningfully improve the detection of purchase intent compared to random or naive baselines, which can be seen in Table 1.

TABLE I
LIGHTGBM MODEL VS RANDOM GUESSING

	LightGBM Model	Guessing (p=0.07)	50/50 Guessing
AUC	0.906	∼0.5	∼0.5
Accuracy	0.979	∼0.87	∼0.5
PR-AUC	0.296	∼0.07	∼0.07
Macro-F1	0.669	∼0.5	∼0.387
Precision	0.428	∼0.07	∼0.07
Recall	0.293	∼0.07	∼0.5

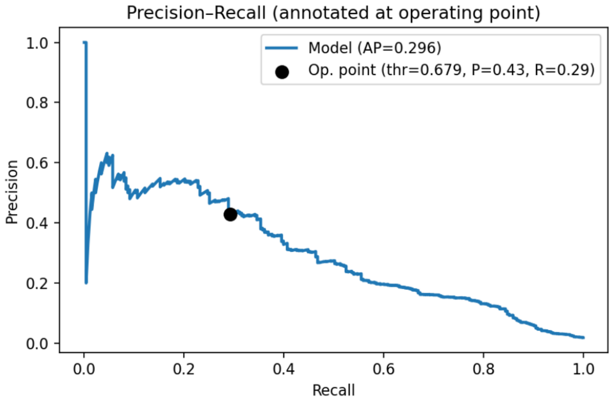


Fig. 5. Precision-Recall Curve: The x-axis is recall (how many of the true buyers we find) and the y-axis is precision (how many of our flagged sessions are actually buyers). The overall area (AP = 0.296), a method to calculate PR-AUC, summarizes performance under class imbalance; the black dot marks our chosen operating point.

DISCUSSION

The experimental findings showcase that the proposed system is effective at identifying user purchase intent, even under extreme class imbalance. The model has a test PR-AUC of 0.296 which substantially outperforms the baseline prevalence rate of 0.07. Additionally, the model achieves an AUC of 0.906. For context, a comparable study on purchase intent prediction using multimodal association rules and the same Retailrocket dataset achieves a top AUC of 0.833, indicating that our system surpasses existing approaches by a clear margin. The Macro-F1 score of 0.669 shows that the system has a reasonable balance between precision (0.428) and recall (0.293), despite the extremely skewed dataset and dominance of non-purchase sessions (∼93% of the dataset). These results indicate that the session-level features combined with the multi-snapshot augmentation strategies allow the model to learn and detect significant behavioral cues associated with purchase intent.

Feature attributions in Figure 6 showcase that deeper engagement (views_pref), recent cart context, and popular items increase predicted intent, while inactivity gaps and lower activity depress it. These findings are significant for real-time ad targeting and recommendation systems. The system is able to recognize high intent users earlier in their browsing sessions by operating on session on session prefixes instead of complete sessions, and therefore can intervene in a timely manner to provide a personalized recommendation or a special promotion. This capability to detect high intent users early is especially applicable to digital ecommerce environments where user attention spans are short and limited and conversion opportunities are easily lost. The cumulative gain shows that the top-scored 10% of sessions capture roughly 78–80% of purchasers, about 7–8x random (Figure 7).

Despite strong results, import limitations still remain. First, despite the model demonstrating a strong ranking capability (high AUC), the precision and recall on the minority purchase class remain modest. This is indicative of the natural challenge of distinguishing true purchase intent and uncertain browsing behavior, or window shopping, particularly when early session signals are sparse. Second, the system is highly accurate (0.979), but this metric is heavily influenced by the majority class and thus contributes limited insight under data imbalance. Third, we use gradient boosting using hand-crafted features, which, despite being interpretable, may not fully capture sequential dependencies and temporal dynamics within sessions.

Future research should experiment with deep learning and sequence-based models that explicitly model the order and timing of events (e.g., recurrent or transformer-based architectures). These techniques can be more effective at capturing long-range correlations and subtle behavioral cues than static features alone. Also, it would be effective to test the system in an online environment with live user interactions and feedback providing stronger evidence of its practical impact on efficient ad targeting and customer engagement.

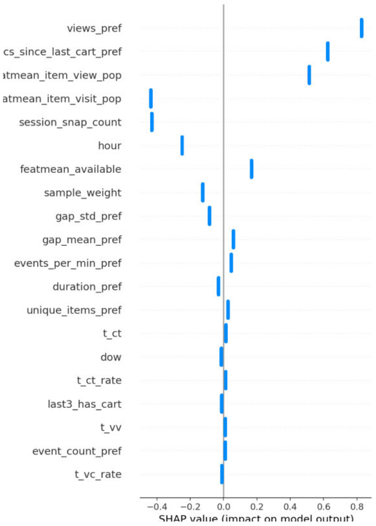


Fig. 6. SHAP Summary: Each row is a feature, and each dot is a session. Position (left/right) shows how much the feature pushes the prediction up or down. Features higher on the plot matter more overall to our model.

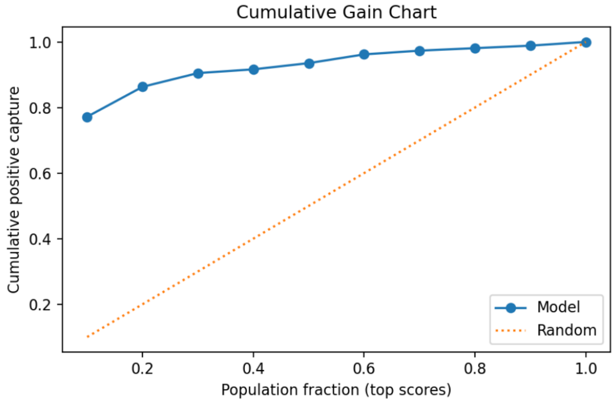


Fig. 7. Cumulative Gain Curve: Reading left-to-right by top-scored fraction of sessions, the curve shows what share of all buyers we capture if we only target the top slice. The model's curve far above the diagonal random line indicates a strong concentration of buyers. For example, the top 10% captures about 78-80% of buyers.

CONCLUSION

This study presents a real-time user intent prediction system for ecommerce, utilizing multi-snapshot session augmentation, historical conversion-rate features, extensive feature engineering, leaky feature removal, and a LightGBM classifier with threshold optimization. Using the Retailrocket dataset for evaluation, the system achieved a test AUC of 0.906, PR-AUC of 0.296 (over 4x better than baseline of 0.07), Macro-F1 of 0.669, precision of 0.428, recall of 0.293, and accuracy of 0.979. These results demonstrate a substantial improvement over baselines prevalence and showcase the system's ability to recognize high purchase intent signals under extreme class imbalance.

Although high accuracy reflects the dominance of non-purchase sessions, the results indicate that meaningful predictive signals can be obtained from early and partial user interactions. This can be applied directly to real-time ad targeting, recommendation systems, and personalization, where detecting high-value users early is essential and can provide an upper edge over competitors.

Future work will involve extending the framework to sequence-based deep learning models (e.g., recurrent or transformer architectures) that could more accurately capture the sequential order and timing of user interactions. Such models, along with enriched behavioral and contextual features, may further enhance precision and recall for purchase intent detection and broaden the applicability of this approach to other areas like finance, media streaming, online experimentation, or any live web environment.

REFERENCES

[1] J. Zhang, C. Yao, H. Sun, and X. Chen, "Deep learning based recommender system: A survey and new perspectives," *ACM Comput. Surv.*, vol. 52, no. 1, pp. 1–38, Jan. 2019. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3285029

[2] W. Tang and M. Wang, "Personalized top-N sequential recommendation via convolutional sequence embedding," in *Proc. ACM WSDM*, Marina Del Rey, CA, USA, Feb. 2018, pp. 565–573. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3159652.3159656

[3] Q. Kang and J. McAuley, "Self-attentive sequential recommendation," in Proc. IEEE ICDM, Singapore, Nov. 2018, pp. 197–206. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8594844

[4] K. Palovics, A. Benczur, D. Lukacs, and D. Siklósi, "Related pins at Pinterest: The evolution of a real-world recommender system," in *Proc. WWW*, Perth, Australia, Apr. 2017, pp. 401–410. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3041021.3054202

[5] J. Xu, Z. Lin, and F. Wang, "Contextual product recommendation using transformer-based models: Uncovering product dependencies in transactional data," *Oper. Res. Forum*, vol. 6, no. 1, pp. 1–15, Jan. 2025. [Online]. Available: https://link.springer.com/article/10.1007/s43069-025-00521-1

[6] C. Hidasi, M. Quadrana, A. Karatzoglou, and D. Jannach, "Comprehensive empirical evaluation of deep learning approaches for session-based recommendation in e-commerce," *Entropy*, vol. 24, no. 11, p. 1575, Nov. 2022. [Online]. Available: https://www.mdpi.com/1099-4300/24/11/1575

[7] Y. Sun *et al.*, "Recommender systems in the era of large language models (LLMs)," *IEEE Trans. Knowl. Data Eng.*, early access, Apr. 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10506571

[8] Y. Chen, J. Wu, and R. Tang, "LLMRec: Large language models with graph augmentation for recommendation," in *Proc. ACM WSDM*, Mérida, Mexico, Mar. 2024, pp. 1115–1123. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3616855.3635853

[9] H. Wu *et al.*, "Recommendation as instruction following: A large language model empowered recommendation approach," *ACM Trans. Inf. Syst.*, vol. 42, no. 2, pp. 1–27, Apr. 2024. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3708882

[10] X. Liu *et al.*, "Recommender AI agent: Integrating large language models for interactive recommendations," *ACM Trans. Inf. Syst.*, vol. 42, no. 4, pp. 1–22, Jul. 2024. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3731446

[11] Y. Zhang *et al.*, "Diffusion models in recommendation systems: A survey," *arXiv preprint arXiv:2501.10548*, 2025. [Online]. Available: https://arxiv.org/abs/2501.10548

[12] F. Li, Z. Ma, and L. Zhou, "DiffSBR: A diffusion model for session-based recommendation," *Inf. Process. Manage.*, vol. 62, no. 5, p. 103445, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306457325002250

[13] Y. Qin, H. Wu, W. Ju, X. Luo, and M. Zhang, "A diffusion model for POI recommendation," *arXiv preprint arXiv:2304.07041*, Apr. 2023. [Online]. Available: https://arxiv.org/abs/2304.07041

[14] S. Park, J. Kim, and Y. Lee, "Diffusion model as a base for cold item recommendation," *Appl. Sci.*, vol. 15, no. 9, p. 4784, 2025. [Online]. Available: https://www.mdpi.com/2076-3417/15/9/4784

[15] Z. Wang *et al.*, "DVIB: Towards robust multimodal recommender systems via variational information bottleneck distillation," in *Proc. ACM WWW*, Sydney, Australia, Apr. 2025, pp. 1021–1030. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3696410.3714840

[16] M. Gupta, A. Sharma, and R. Patel, "MTSTRec: Multimodal time-aligned shared token recommender," *OpenReview*, 2025. [Online]. Available: https://openreview.net/forum?id=yWDvVl9Wtp

[17] H. Lee and K. Hosanagar, "Impact of recommender systems on sales volume and diversity," 2019. [Online]. Available: https://www.semanticscholar.org/paper/Impact-of-Recommender-Systems-on-Sales-Volume-and-Lee-Hosanagar/8109f0a47f433b32979d9b3f2da9facee5eba4ad

[18] J. Tang and K. Wang, "Multi-rate deep learning for temporal recommendation," in *Proc. ACM SIGIR*, Pisa, Italy, Jul. 2016, pp. 1079–1082. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/2911451.2914726

[19] M. Alfonsi, "Real time recommendations with clickstream data," M.S. thesis, Politecnico di Milano, Milan, Italy, 2022. [Online]. Available: https://www.politesi.polimi.it/handle/10589/197048

[20] K. Kamehkhosh and D. Jannach, "A comparison of frequent pattern techniques and a deep learning method for session-based recommendation," 2017. [Online]. Available: https://www.semanticscholar.org/paper/A-Comparison-of-Frequent-Pattern-Techniques-and-a-Kamehkhosh-Jannach/5f6ff6990a1afac7e7b5616283b1d4e67a9d034f

[21] A. Sharma *et al.*, "A comprehensive review of recommender systems: Transitioning from theory to practice," *arXiv preprint arXiv:2407.13699*, 2024. [Online]. Available: https://arxiv.org/abs/2407.13699

Keyframe Selection Strategies for Balancing Global Accuracy and Local Stability in Endoscopic SLAM

Hyeon-Seo Kim
Department of Datascience
Chonnam National University
Gwangju, South Korea
kimhs9574@jnu.ac.kr

Seokmin Hong
Medical AI Lab
Korea Institute of Medical Microrobotics
Gwangju, South Korea
seokmin.hong@kimiro.re.kr

Byungjeon Kang*
Department of Datascience
Chonnam National University
Gwangju, South Korea
bjkang8204@chonnam.ac.kr

Abstract—**This study evaluates the effect of keyframe selection strategies on endoscopic simultaneous localization and mapping (SLAM) using the trans t2 b sequence of the C3VD dataset. Six strategies, including Fixed-interval, Adaptive baseline, Adaptive kf25/50/100, and Sharp-photo only, were compared. The Fixed-interval method achieved the highest global accuracy, reporting the lowest errors in ATE Sim(3) (11.43 mm) and ATE SE(3) (16.30 mm). In contrast, Adaptive kf50 exhibited the largest global error (18.38 mm and 18.63 mm, respectively). For local stability, the Sharp-photo only strategy demonstrated the best rotational precision with Rotation RPE Δ = 1: 0.36° and Δ = 5: 1.14°, whereas Adaptive kf25 showed the largest rotational deviation at Δ = 5 (1.30°). In terms of translational stability, Adaptive Kf50 achieved the lowest Translation RPE Δ = 1 (2.95 mm), while Sharp-photo only maintained the best performance at Δ = 5 (5.53 mm). Overall, these results highlight a trade-off between global trajectory accuracy and local motion stability, suggesting that while adaptive scheduling enhances frame efficiency, Fixed-interval and Sharp-photo only approaches provide more consistent and robust performance under endoscopic imaging conditions.**

Keywords—Endoscopy, SLAM, Keyframe Selection

I. INTRODUCTION

Endoscopic imaging requires navigating narrow, deformable, and dynamic anatomical pathways where accurate localization and mapping are indispensable for 3D reconstruction, surgical navigation, and intra-operative guidance. However, clinical endoscopic videos suffer from challenges such as specular highlights, fluid-induced bubbles, tissue deformation, and repetitive texture patterns which often degrade the robustness of visual tracking. Processing every frame exacerbates cumulative drift and imposes a significant computational overhead. Thus, selecting a representative subset of frames, namely keyframes, emerges as a critical design factor to balance global accuracy and computational efficiency.

Conventional visual simultaneous localization and mapping (SLAM) systems, such as ORB-SLAM and Direct Sparse Odometry (DSO), typically rely on Fixed-intervals or heuristic thresholds for keyframe insertion. However, their applicability in endoscopic environments remains uncertain [1, 2]. Previous studies have primarily addressed image preprocessing techniques, including bubble removal and contrast enhancement, as well as depth stabilization, whereas systematic evaluations of keyframe selection strategies in endoscopic SLAM are still scarce. Motivated by this limitation, the present study aims to provide a quantitative comparison between Fixed-interval and adaptive score-based approaches, highlighting their respective strengths and weaknesses in endoscopic SLAM.

II. RELATED WORK

Keyframe selection has long been a critical topic in SLAM research. The ORB-SLAM family incorporates new keyframes based on the number of feature points and tracking stability, while DSO improves robustness by minimizing photometric error. More recently, neural network-based methods for frame scoring and keyframe resampling techniques have been introduced. In the context of endoscopic imaging, however, most research has focused on preprocessing techniques such as bubble removal and contrast enhancement, along with depth stabilization, whereas systematic evaluation of keyframe selection policies remain limited.

A. ENeRF-SLAM / NeRF-based SLAM research

The advent of neural radiance fields (NeRF) has introduced new possibilities by enabling simultaneous dense reconstruction and pose estimation. Rosinol et al. demonstrated through NeRF-SLAM that real-time dense mapping and tracking can be achieved using only monocular input [3]. Luo et al. proposed ENeRF-SLAM, specially designed for endoscopic environments, which achieved high reconstruction performance by leveraging neural implicit representations in narrow and dynamic anatomical pathways [4]. Liao et al. developed VI-NeRF-SLAM, which enhanced robustness by integrating inertial measurement unit (IMU) data with visual information [5]. Li et al. introduced EC-SLAM, which combines truncated signed distance function (TSDF) encoding with bundle adjustment, achieving a balance between tracking accuracy and real-time performance [6].

B. Fixed-interval vs. Adaptive Keyframe Selection

Keyframe selection strategies can broadly be categorized into Fixed-interval and adaptive approaches. Qin et al. highlighted the limitations of Fixed-interval keyframe selection in RGB-D SLAM and proposed an adaptive method that dynamically adjusts based on environmental changes and motion [7]. Kuo et al. introduced an information-theoretic criterion based on uncertainty and entropy, demonstrating applicability to multi-camera systems [8]. Thorne et al. proposed a submodular optimization framework for systematic keyframe selection, demonstrating that it can achieve both map summarization and effective pose constraints [9]. Dias et al., through a systematic literature review, emphasized that despite recent progress, this field still lacks a solid theoretical foundation and comprehensive evaluation [10].

Therefore, the present study contributes by systematically comparing Fixed-interval and adaptive score-based keyframe strategies in endoscopic SLAM, offering empirical insights into their impact on both global accuracy and local stability.

III. METHODS

A. Dataset

The experiment used the *trans t2 b* sequence from the C3VD dataset, a publicly available dataset consisting of 103 frames. This dataset features narrow anatomical passages, non-uniform illumination, and repetitive tissue patterns. It provides RGB endoscopic images, depth maps, and ground-truth camera poses, making it suitable for evaluating SLAM performance under challenging endoscopic conditions.

B. Model (ENeRF-SLAM)

The model employed in this study is ENeRF-SLAM, an endoscopy-specific SLAM system based on neural implicit representations. It is designed to handle the tight anatomical spaces and strong light reflections commonly encountered in endoscopic imaging. The system estimates the camera pose through a tracking module and reconstructs dense maps using a NeRF-based mapping module. This architecture enables a direct evaluation of how different keyframe selection policies affect both global accuracy and local stability.

C. Keyframe Selection Strategies

Six keyframe selection strategies were compared. The first approach, Fixed-interval selection, chooses keyframes at regular frame intervals, a traditional approach widely used in feature-based SLAM systems such as ORB-SLAM [12]. The second approach, Adaptive baseline selection, addresses challenges common in endoscopic environments such as shaking, blur, strong reflections, complex texture patterns, and rapid viewpoint changes. Using all frames increases computational cost and drift, while using too few frames weakens global constraints. The adaptive strategy evaluates three normalized scores with sharpness, photometric consistency, and motion to select informative frames while excluding redundant or low-quality frames. Sharpness is computed using the Laplacian distribution, photometric consistency measures brightness and color similarity between adjacent frames, and motion quantifies viewpoint change [7-8]. The scores are normalized to the range of 0-1 and combined with a weighted summation, which ensures balanced coverage and reduces overfitting. The third approach, Adaptive kf25/50/100, restricts the number of selected keyframes to 25, 50, and 100, respectively, to investigate the impact of distribution density on performance. Within the 103-frame sequence, this strategy examines whether learning and matching remain effective when the number of frames is limited by computation or memory constraints. Following Qin et al [7], a frame count limit was applied to experimentally control sequence coverage. The fourth approach, Sharp-photo only, isolates the effect of image quality by selecting keyframes exclusively according to sharpness and optical coherence. By excluding the motion score, this ablation study establishes a lower bound for performance when motion estimation is unstable, thereby testing how much reconstruction can rely solely on visual quality [1-2].

D. Evaluation Metrics and Visualization

Performance was assessed using Absolute Trajectory Error (ATE) and Relative Pose Error (RPE). ATE evaluates global trajectory alignment. Two alignment methods were used: Sim(3) and SE(3). Sim(3) (similarity transformation in 3D) represents a transformation group that includes rotation, translation, and isotropic scaling, which is particularly important for monocular SLAM, where scale ambiguity exists. In contrast, SE(3) (Special Euclidean transformation in 3D) includes only rotation and translation with a fixed scale, representing rigid-body transformations. In both cases, the Umeyama alignment was applied to match the predicted trajectory with ground truth, and the root mean square error (RMSE) was reported. RPE evaluates local pose accuracy, evaluating rotational and translational errors between pairs of frames at fixed intervals. Specifically, errors were computed at Δ=1 and Δ=5 frames, with RMSE values reported for both rotation (degrees) and translation (meters). This dual evaluation provides a comprehensive view of global drift suppression (ATE) and local motion stability (RPE).

For visualization, ATE values under Sim(3) and SE(3) alignment were plotted to represent the RMSE of the absolute position error per frame. In addition, SE(3) alignment provides a scale factor indicating how much scaling is required to align the predicted trajectory with ground truth, and this factor was also visualized to evaluate scale stability. RPE results were plotted for Δ=1 and = 5 in both rotation and translation, providing insight into the short-term stability of each keyframe selection method. Considering the characteristics of endoscopy, the experimental results are reported in millimeters (mm) rather than meters (m).

IV. RESULTS

In this section, global accuracy and local stability were evaluated using ATE and RPE, respectively, for six keyframe selection strategies (Fixed-interval, Adaptive baseline, Adaptive kf25/50/100, Sharp-photo only). Lower values indicate better performance. ATE was computed by aligning the predicted trajectory with the ground truth using Sim(3) and SE(3) alignment, where RPE was calculated at Δ=1 and Δ=5 frames.

A. ATE Sim(3) Results

In the ATE Sim(3) alignment, the Fixed-interval method achieved the lowest RMSE error of 11.43 mm, while the Adaptive kf50 method yielded the highest error of 18.38 mm, as shown in Fig. 1. This result demonstrates that uniformly spaced global constraints are beneficial for suppressing cumulative drift. Furthermore, the Sim(3) alignment provides a scale factor that indicates the isotropic scaling required to match the predicted trajectory with the ground truth.

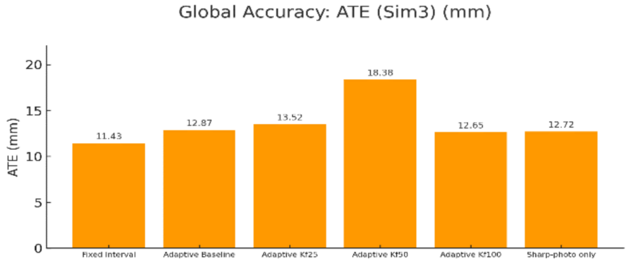


Fig. 1. Comparison of ATE Sim(3) RMSE results

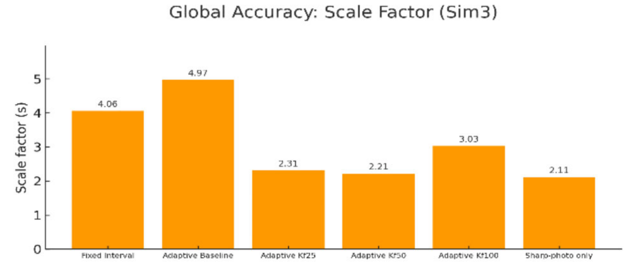


Fig. 2. Comparison Scale Factor Sim(3) results

A scale value closer to 1 implies a more stable global scale estimate. In this experiment, the Sharp-photo only method reported the most stable scale of 2.11, whereas the Adaptive Baseline method showed the largest deviation of 4.97. This suggests that certain adaptive configurations can compensate for scale mismatches through Sim(3) alignment, which can obscure differences when only the ATE values are considered.

B. ATE SE(3) Results

In the ATE SE(3) alignment, the Sharp-photo only method achieved the lowest RMSE of 14.80 mm, while the Adaptive kf50 method recorded the highest RMSE of 18.63 mm, as shown in Fig. 3. These results indicate that methods with unstable scale estimation yield higher SE(3) errors, whereas methods with stable scale estimation achieve lower SE(3) errors. Thus, SE(3) alignment highlights scale sensitivity and stability across different keyframe selection strategies.

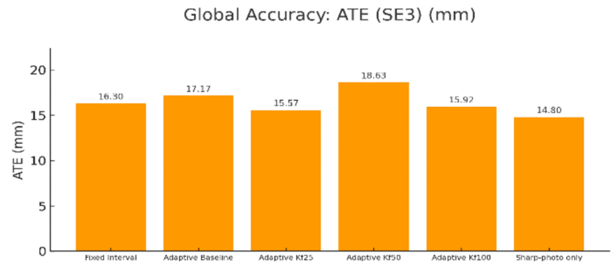


Fig. 3. Comparison ATE SE(3) RMSE results

C. Local Stability: Rotation RPE (Δ=1, 5)

Rotation RPE calculates the angular difference between the predicted and ground-truth camera orientations, expressed as RMSE. The experimental results showed that the Sharp-photo only method achieved the lowest error with Δ = 1: 0.36° and Adaptive baseline method achieved the lowest error with Δ = 5: 1.07°, while the Adaptive baseline exhibited the highest error at Δ = 1: 0.54°, and the Adaptive kf25 showed the largest error at Δ = 5: 1.30°, as shown in Fig. 4 and Fig. 5. These results indicate that fine-tuning keyframe selection strategies can improve the stability of rotational motion estimation in endoscopic SLAM.

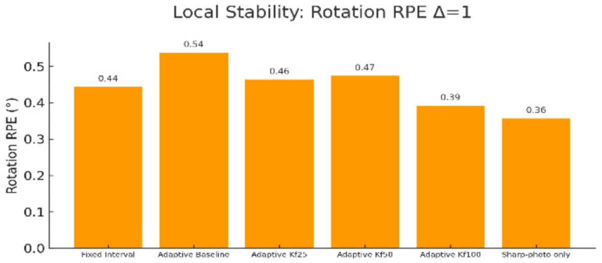


Fig. 4. Comparison of Rotation RPE (Δ=1) results

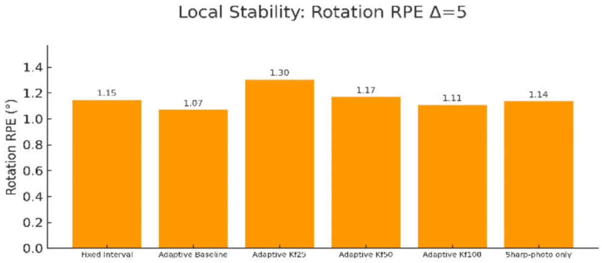


Fig. 5. Comparison of Rotation RPE (Δ=5) results

These findings suggest that when global constraints are weakened, rotational errors accumulate even over short intervals, and local rotational stability plays a critical role in visual quality under frequent reflections, blurring, and sharp turns in endoscopic imaging.

D. Local Stability: Translation RPE (Δ=1, 5)

Translation RPE evaluates the translational accuracy of camera movement between frames. The experimental results showed that the Adaptive kf50 method produced the lowest error of 2.95 mm at Δ = 1, whereas the Sharp-photo only method achieved the lowest error of 5.53 mm at Δ = 5. Conversely, the Adaptive baseline method exhibited the highest error of 6.64 mm at Δ = 1 and 10.48 mm at Δ = 5, as shown in Fig. 6 and Fig. 7. This demonstrates that, in endoscopic environments, maintaining consistent local frame-to-frame movement is essential for stable position tracking by preserving the convergence region of optimization and spatial correlation.

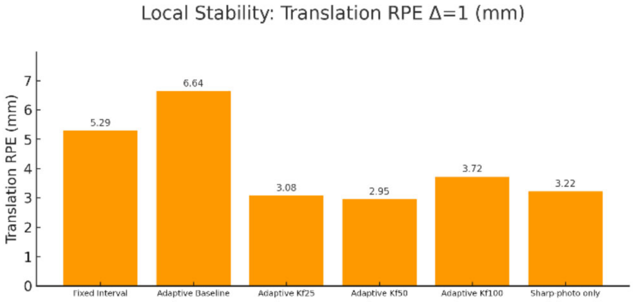


Fig. 6. Comparison of Translation RPE (Δ=1) results

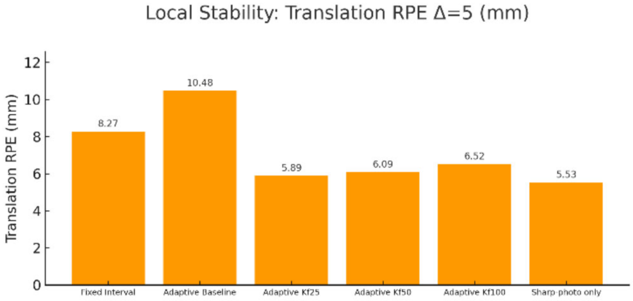


Fig. 7. Comparison of Translation RPE (Δ=5) results

In summary, the Fixed-interval strategy consistently achieved the best global accuracy, reporting the lowest ATE values in both Sim(3) (11.43 mm) and SE(3) (16.30 mm) alignments. In contrast, the Sharp-photo only demonstrated superior rotational stability at Δ = 1 (0.36°). At Δ = 5, the Adaptive Baseline method achieved the lowest rotation error (1.07°), while Adaptive kf25 exhibited the largest error (1.30°). Regarding translational stability, the Adaptive kf50 achieved the lowest Translation RPE at Δ = 1 (2.95 mm), while Sharp-photo only showed the best performance at Δ = 5 (5.53 mm). Overall, adaptive strategies tended to be more scale-sensitive, with greater variation in Scale Factor (Sim3) across configurations. These findings suggest that while adaptive scheduling improves frame efficiency, the Fixed-interval and Sharp-photo only approaches provide more consistent accuracy and stability under the same endoscopic conditions.

V. DISCUSSION AND CONCLUSION

This study demonstrates that keyframe selection strategies exert distinct and inherently opposing influences on global accuracy and local stability in endoscopic SLAM Using ATE and RPE as evaluation metrics, the Fixed-interval method consistently achieved the best global trajectory accuracy, showing the lowest errors in both ATE Sim(3) (11.43 mm) and ATE SE(3) (16.30 mm). In contrast, local stability varied depending on the evaluation interval and metric. The Sharp-photo only strategy achieved the most stable rotation performance at $\Delta = 1$ (0.36°) and the lowest translation error at $\Delta = 5$ (5.53 mm). Meanwhile, Adaptive kf50 demonstrated superior short-term translation accuracy at $\Delta = 1$ (2.95 mm), and the Adaptive baseline method achieved the lowest rotation error at $\Delta = 5$ (1.07°). These results indicate that no single strategy consistently outperforms others across all evaluation metrics; rather, each method exhibits distinct advantages under specific spatial or temporal conditions. These findings indicate that uniform spatial coverage benefits global registration, whereas selectively incorporating motion and quality cues can improve short-term pose estimation. The results also highlight the role of motion-aware scoring in controlling the spatial distribution of keyframes. When global constraints are weakened, rotational errors accumulate rapidly, particularly in environments characterized by reflections, blur, and abrupt viewpoint changes. Conversely, high-quality frames contribute to improved local stability but may fail to prevent scale drift over longer sequences. This trade-off underscores the necessity of designing hybrid policies that combine the strengths of both fixed and adaptive strategies.

These findings align with and extend prior studies in endoscopic SLAM. In particular, Teufel et al.'s OneSLAM demonstrated generalizable SLAM performance across diverse endoscopic domains without retraining, indicating that the strategies investigated in this study may also be applicable to broader endoscopic environments [11]. Similarly, Wang et al.'s EndoGSLAM highlighted the effectiveness of keyframe policies in balancing the trade-off between real-time dense reconstruction and tracking via Gaussian splatting [12]. Furthermore, Liu et al. showed that incorporating appearance and geometry priors enhances adaptability in endoscopic navigation, suggesting that future SLAM frameworks could benefit from integrating both fixed and adaptive policies, explored in this study [13].

A key limitation of this work is that the evaluation was performed on a single sequence of the C3VD dataset, which constrains the generalizability of the findings. To improve reliability, future studies should evaluate multiple sequences under varying anatomical and imaging conditions, repeat experiments across datasets, and incorporate additional evaluation metrics beyond ATE and RPE. Multi-level benchmarks that reflect both geometric accuracy and clinical usability would provide a more comprehensive validation.

In conclusion, this study provides a systematic comparison of Fixed-interval and Adaptive score-based keyframe selection strategies in endoscopic SLAM. By highlighting their respective strengths and weaknesses in terms of global accuracy and local stability, the results offer practical guidelines for designing keyframe policies in endoscopic navigation systems. This work contributes to establishing a quantitative foundation for future research on optimizing SLAM performance in medical imaging environments.

ACKNOWLEDGMENT

This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00242528). This research was supported by a grant of the Korean Health Technology R&D Project through the Korea Health Industry Development Institute(KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea(grant number : RS-2023-00302146).

REFERENCES

[1] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, “ORB-SLAM: A versatile and accurate monocular SLAM system,” *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[2] J. Engel, V. Koltun, and D. Cremers, “Direct sparse odometry,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 3, pp. 611–625, 2018.

[3] A. Rosinol, M. Abate, Y. Chang, and L. Carlone, “NeRF-SLAM: Real-time dense SLAM with neural radiance fields,” *arXiv preprint* arXiv:2210.13641, 2022.

[4] Z. Luo, C. Wang, Y. Li, and J. Wang, “ENeRF-SLAM: A Dense Endoscopic SLAM with Neural Implicit Representation,” in *Proceedings of the MICCAI Workshop on Medical Imaging with Deep Learning*, 2023.

[5] H. Liao and H. Ai, “VI-NeRF-SLAM: Visual-Inertial Neural Radiance Field SLAM,” *Journal of Real-Time Image Processing*, vol. 20, no. 4, pp. 567–582, 2024.

[6] Y. Li, X. Chen, J. Zhang, and S. Li, “EC-SLAM: Efficient Neural Implicit SLAM with TSDF Encoding and Bundle Adjustment,” *arXiv preprint* arXiv:2404.13346, 2024.

[7] T. Qin, P. Li, and S. Shen, “Adaptive keyframe selection for real-time RGB-D SLAM,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 5861–5868.

[8] C.H. Kuo, X. Chen, and G. Huang, “Information-theoretic keyframe selection for multi-camera visual odometry,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 221–228.

[9] J. Thorne, M. Li, and B. Williams, “Submodular optimization for keyframe selection in visual SLAM,” *arXiv preprint* arXiv:2410.05576, 2024.

[10] A. Dias, P. Rodrigues, R. Ventura, and J. Lima, “A systematic literature review on keyframe selection in visual SLAM,” *Robotics*, vol. 12, no. 3, 88, 2023.

[11] J. Teufel, R. Ansari, and N. Navab, “OneSLAM: A generalized monocular SLAM framework for surgical environments,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 19, no. 5, pp. 877–889, 2024

[12] C. Wang, Z. Luo, Y. Li, and J. Wang, “EndoGSLAM: Real-time dense endoscopic SLAM with Gaussian Splatting,” *arXiv preprint* arXiv:2403.15124, 2024.

[13] X. Liu, H. Zhang, Y. Song, and P. Wang, “SLAM with appearance and geometry priors for endoscopic navigation,” *Medical Image Analysis*, vol. 82, 102617, 2022

A PDMS-Based Acoustic Fresnel Lens for Vortex Trapping of Microparticles

Viet Hoang Nguyen
Korea Institute of Medical Microrobotics
College of AI Convergence
Chonnam National University
Gwangju, Korea
viethoang20151552@jnu.ac.kr

Hiep Xuan Cao
Ultrasonic Actuator Lab
Korea Institute of Medical Microrobotics
Gwangju, Korea
hiep.caoxuan@kimiro.re.kr (H.X.C)

Daewon Jung
Ultrasonic Actuator Lab
Korea Institute of Medical Microrobotics
Gwangju, Korea
jungdaewon@kimiro.re.kr

Jong-oh Park
Korea Institute of Medical Microrobotics
Gwangju, Korea
jop@kimiro.re.kr

Byungjeon Kang
Korea Institute of Medical Microrobotics
Graduate School of Data Science
College of AI Convergence
Chonnam National University
Gwangju, Korea
bjkang8204@jnu.ac.kr (B.K)

***Abstract*— Acoustic tweezers have gained increasing attention as a contactless method for particle manipulation, yet most existing implementations rely on bulky standing-wave devices or complex phased transducer arrays that restrict their practicality. In this study, we propose and demonstrate a PDMS-based acoustic lens inspired by the principle of a Fresnel zone plate. The lens was fabricated using a simple and low-cost mold-casting process, highlighting its scalability and suitability for compact acoustic systems. Numerical simulations and experimental validations confirmed the effectiveness of the proposed lens, demonstrating its ability to generate a vortex trap and to stably confine a 3D-printed spherical particle with a diameter of 750 µm. These findings establish the PDMS acoustic lens as a practical and cost-effective alternative for acoustic manipulation, while also opening opportunities for its future integration into biomedical, diagnostic, and microfluidic platforms.**

Keywords—acoustic lens, acoustic tweezers, contactless manipulation, spatial phase modulation

I. INTRODUCTION

Acoustic tweezers have emerged as a promising contactless strategy for object manipulation. By harnessing acoustic radiation forces and pressure gradients generated by ultrasonic fields, acoustic tweezers can trap, translate, rotate, and arrange a wide variety of objects, ranging from nanoscale particles to microscale structures. Unlike optical or magnetic methods, acoustic tweezers rely on mechanical waves that are inherently biocompatible, energy-efficient, and capable of operating in opaque or heterogeneous media. This unique combination of features establishes acoustic tweezers as a powerful and versatile tool for micromanipulation applications. In recent years, the applications of acoustic tweezers (AT) have expanded across diverse fields, ranging from acoustofluidics for bioparticle separation[1] and cell patterning[2] to biomedical applications such as targeted drug delivery[3], as well as in rapid prototyping[4] and advanced graphic display technologies[5].

Early studies utilized acoustic standing waves to generate radiation forces for object trapping and manipulation[6], [7]. Nevertheless, their implementation typically requires an opposing transducer configuration, which restricts their use in in vivo environments. In contrast, traveling waves utilize phase modulation to form acoustic tweezers, eliminating the need for a reflector that is essential in standing-wave configurations and thereby providing an open manipulation workspace. By controlling the phase of individual transducers, transducer arrays have demonstrated the ability to generate various types of acoustic tweezers, such as twin traps, vortex traps, and bottle traps, which can stably trap and manipulate objects[8]. Recent studies have further advanced this approach to achieve particle manipulation with up to five degrees of freedom[9] and to enable targeted particle delivery in in vivo models[10]. Although this method offers the advantages of dynamic and flexible control, it is still constrained by the complexity of the control system and the bulkiness of the setup. Therefore, in the absence of such dynamic flexibility as provided by transducer arrays, acoustic lenses have continued to be investigated as compact alternatives capable of effectively focusing acoustic waves to generate acoustic tweezers or to shape acoustic fields into specific high-resolution patterns[11]. Several previous studies have introduced diverse acoustic lens designs, ranging from solid Fresnel[12] and polymer-based lenses[13] to tunable liquid[14] and bottle-beam configurations[15]. These developments highlight the versatility of acoustic lenses in focusing acoustic waves and in shaping acoustic fields to generate acoustic tweezers. However, these lenses are still limited by the need for specially designed transducers, complex computations, or high manufacturing costs. Therefore, there remains a demand for a new type of lens that can be easily integrated with flat transducers commonly available on the market, while maintaining a simple design and manufacturing process.

In this work, we propose an acoustic lens designed based on the principles of a Fresnel zone plate and fabricated from polydimethylsiloxane (PDMS) through a simple mold-casting process. This PDMS-based lens is capable of generating a vortex trap that enables the stable confinement of particles, thereby facilitating their contactless manipulation. After designing and fabricating a prototype of the acoustic lens, the generated acoustic fields were scanned and the lens was applied for particle trapping. The obtained results clearly demonstrate the effectiveness of the proposed lens design.

II. PROPOSED ACOUSTIC LENS

A. *Design of the Acoustic Lens*

PDMS was selected as the material for fabricating the acoustic lens due to its mechanical and acoustic properties. To mimic a biological environment, our experiments were conducted in water, while PDMS exhibits an acoustic impedance of approximately 1.05×10^6 kg/m^{2}s, which is close to that of water (1.48×10^6 kg/m^{2}s). This impedance matching minimizes reflection at the interface and enables the effective transmission of acoustic waves. As an elastomeric polymer, PDMS also possesses characteristic elasticity, providing mechanical robustness against impact. In addition, its hydrophobicity makes it well suited for aqueous applications, and it can be easily molded into precise structures through a simple casting process, allowing low-cost and reproducible fabrication. These advantages highlight PDMS as a practical and versatile material for developing compact acoustic lenses.

The proposed lens is designed based on the framework of spatial phase modulation reported in prior work, which generates a vortex trap by decomposing the phase modulation of a transducer array into a focusing element and a holographic signature[8]. In our approach, the acoustic lens is first designed to focus acoustic wave at a focal point using the principle of a Fresnel zone plate, originally developed for optics, consists of a set of concentric rings alternating between opaque and transparent regions. Light incident on the zone plate diffracts around the opaque rings rather than passing directly through them, and the ring pattern can be calculated such that the diffracted waves constructively interfere at the desired focal point. Following this concept, an acoustic lens can be designed to focus acoustic waves into a specific region, with the radius of each concentric ring determined by the following equation:

$$r = \sqrt{\frac{1}{4}n^2\lambda^2 + n\lambda h} \tag{1}$$

Where n is an integer, λ is the wavelength of the acoustic wave, and h is the focal length. In a conventional Fresnel zone plate, concentric rings alternating between opaque and transparent regions are designed to retain the contributions from rings with the same phase. In this work, however, the out-of-phase rings are compensated by increasing the thickness of the PDMS lens, such that these rings acquire an additional phase shift of π radians and become in phase with the remaining rings. As a result, the focused energy is enhanced through the constructive contribution of acoustic waves from all transducer elements. The additional thickness of PDMS lens needed to introduce a phase shift of π radians can be calculated as follow:

$$\Delta h = \frac{1}{2f}\frac{c_2 c_1}{|c_2 - c_1|} \tag{2}$$

Where f denotes the frequency of the acoustic wave, and c_1 and c_2 are the speed of sound in water and PDMS, respectively. The PDMS-based lens designed for acoustic wave focusing is shown in Fig. 1a. After focusing the acoustic wave, a holographic signature of a vortex trap was applied to the lens by introducing a continuous phase increase of 2π along its azimuthal direction. To achieve this, the lens thickness must be gradually varied around one full azimuthal rotation, with the height variation corresponding to a total phase shift of 2π. Consequently, the required thickness variation equals 2Δh, , resulting in the acoustic geometry shown in Fig. 1b.

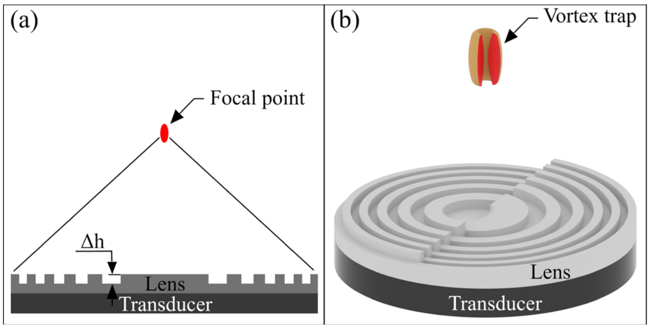


Fig. 1. Geometry of: (a) the acoustic lens for focusing acoustic waves, and (b) the acoustic lens for generating the vortex trap.

B. *Acoustic Lens Fabrication*

Once the acoustic lens had been designed, its three-dimensional geometry was modeled in CAD software, and this model was subsequently employed to create the mold. The mold files were converted into STL format and fabricated using a high-resolution 3D printer. Printed mold parts were washed with isopropyl alcohol and then sprayed with a thin coating of Easy Release 200 to facilitate demolding. PDMS (SYLGARD 184) was prepared by thoroughly mixing the base and curing agent at a weight ratio of 10:1, followed by vacuum degassing to remove any trapped bubbles. The degassed mixture was poured into the molds and cured in a convection oven at 80 °C for 4 hours. After curing, the lenses were removed from the molds and excess material around the edges was trimmed away, producing the final PDMS lens structures..

III. RESULTS

In this section, the performance of the proposed acoustic lens is evaluated through both numerical simulations and experimental validation. First, the simulated acoustic fields are presented to confirm the ability of the lens to focus acoustic waves and to generate vortex traps. Subsequently, experimental measurements of the acoustic fields are compared with the simulations to verify the accuracy of the fabrication. Finally, particle-trapping experiments are reported to demonstrate the capability of the lens in stably confining particles. For both the simulations and experiments, which were conducted in a water environment. The acoustic lens was designed with a diameter of 50 mm and coupled to an array of seven transducers, each 15 mm in diameter, as shown in Fig. 3. All seven transducers were driven by the same sinusoidal signal at a frequency of 1 MHz to ensure uniform operation.

The simulations were carried out in COMSOL Multiphysics using full three-dimensional geometries, and the results are presented in Fig. 2a. The acoustic field was shaped into the characteristic pattern of a vortex trap, with the pressure focused in a cylindrical distribution. After fabrication, the acoustic lens was coupled to the transducer array, and the acoustic field was scanned in a water tank using a 1.0-mm needle hydrophone (Precision Acoustics Ltd., Dorchester, UK) mounted on a linear stage capable of translating along two horizontal axes and the vertical axis, with a scan step of 0.35 mm. The transducers were driven with an input voltage of 20 Vpp, and the scanned results are shown in Fig. 2b. A comparison between the simulated and measured fields indicates a high degree of agreement, with minor discrepancies attributed to differences in material properties between the simulation and experiment, numerical approximations in the simulation, and small fabrication errors.

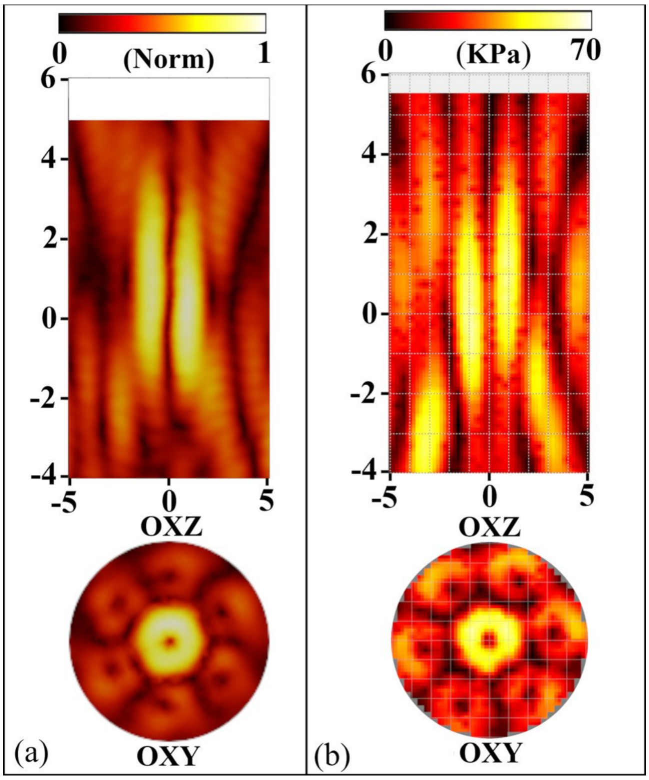


Fig. 2. Acoustic fields obtained from (a) simulation and (b) experimental scanning.

To further evaluate the performance of the proposed acoustic lens, the input driving voltage was increased to 50 Vpp for particle-trapping experiments. For this purpose, a spherical particle with a diameter of 750 µm was fabricated using 3D printing (VeroWhite Plus material). The particle was initially released into the water tank and allowed to sink under gravity toward the focal region of the acoustic field. When the particle approached the focal point, it was stably confined within the vortex trap generated by the lens, as illustrated in Fig. 3. The particle remained trapped without noticeable displacement, indicating that the acoustic forces produced by the lens were sufficient to counteract both gravitational and hydrodynamic forces. This result confirms the ability of the proposed acoustic lens to achieve stable and contactless trapping under relatively high driving voltages.

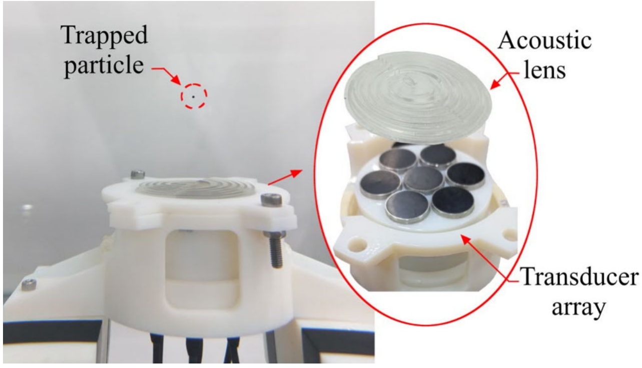


Fig. 3. Results of particle trapping experiment.

IV. CONCLUSION

In this work, we proposed and demonstrated a PDMS-based acoustic lens designed according to the principle of a Fresnel zone plate. The lens was fabricated through a simple mold-casting process and characterized through both numerical simulations and experimental validation. The results showed that the lens can generate a vortex trap, with good agreement between simulated and measured acoustic fields. Furthermore, proof-of-concept particle-trapping experiments confirmed the ability of the lens to achieve stable confinement.

These findings highlight not only the effectiveness of the PDMS acoustic lens as a compact and low-cost alternative to conventional transducer arrays or complex 3D-printed lens structures, but also its potential for practical integration into portable acoustic systems. The simplicity and scalability of the mold-casting process make the design highly suitable for lightweight and transportable devices, where space and cost constraints are critical. Furthermore, the lens can function as a modular component within larger acoustic manipulation platforms, enabling flexible incorporation into biomedical, diagnostic, or microfluidic systems. In addition, the geometry of the PDMS lens can be readily tailored by adjusting its thickness or profile, allowing versatile control over the acoustic field and enabling the generation of diverse trap configurations such as focus points, twin-traps, bottle beams, and vortices. Beyond these acoustic functions, PDMS can be combined with nanoparticles (e.g., MNPs, AuNPs) or coated with thin functional layers to impart additional optical or magnetic properties, thereby opening up new opportunities for hybrid acoustic–optical or acoustic–magnetic manipulation systems.

Future work will focus on optimizing the lens geometry to enhance trapping efficiency while reducing the required driving power. Additional studies will also investigate the scalability of the design toward smaller particles, including micro- and nano-scale objects, and explore its applicability in more complex environments, such as in vivo conditions. To address one of the major limitations of acoustic lenses—the lack of dynamic control for particle manipulation without relying on robotic arms or motorized stages—further research will aim to develop approaches that enable the lens to actively shift its focal point, thereby allowing direct and flexible manipulation of trapped particles.

ACKNOWLEDGMENT

This research was supported by a grant of the Korea Health Technology Development R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2023-00302148).

REFERENCES

[1] L. Shen *et al.*, “Acousto-dielectric tweezers enable independent manipulation of multiple particles,” 2024. [Online]. Available: https://www.science.org

[2] J. Shi, D. Ahmed, X. Mao, S. C. S. Lin, A. Lawit, and T. J. Huang, “Acoustic tweezers: Patterning cells and microparticles using standing surface acoustic waves (SSAW),” *Lab Chip*, vol. 9, no. 20, pp. 2890–2895, 2009, doi: 10.1039/b910595f.

[3] H. X. Cao *et al.*, “Micromotor manipulation using ultrasonic active traveling waves,” *Micromachines (Basel)*, vol. 12, no. 2, Feb. 2021, doi: 10.3390/mi12020192.

[4] K. Melde, E. Choi, Z. Wu, S. Palagi, T. Qiu, and P. Fischer, “Acoustic Fabrication via the Assembly and Fusion of Particles,” *Advanced Materials*, vol. 30, no. 3, Jan. 2018, doi: 10.1002/adma.201704507.

[5] Y. Ochiai, T. Hoshi, and J. Rekimoto, “Pixie dust: Graphics Generated by Levitated and Animated Objects in,” *ACM Trans Graph*, vol. 33, no. 4, Jul. 2014, doi: 10.1145/2601097.2601118.

[6] D. Foresti, M. Nabavi, M. Klingauf, A. Ferrari, and D. Poulikakos, “Acoustophoretic contactless transport and handling of matter in air,” *Proc Natl Acad Sci U S A*, vol. 110, no. 31, pp. 12549–12554, Jul. 2013, doi: 10.1073/pnas.1301860110.

[7] E. H. Trinh, “Compact acoustic levitation device for studies in fluid dynamics and material science in the laboratory and microgravity,” *Review of Scientific Instruments*, vol. 56, no. 11, pp. 2059–2065, 1985, doi: 10.1063/1.1138419.

[8] A. Marzo, S. A. Seah, B. W. Drinkwater, D. R. Sahoo, B. Long, and S. Subramanian, “Holographic acoustic elements for manipulation of levitated objects,” *Nat Commun*, vol. 6, Oct. 2015, doi: 10.1038/ncomms9661.

[9] H. X. Cao *et al.*, “Holographic Acoustic Tweezers for 5-DoF Manipulation of Nanocarrier Clusters toward Targeted Drug Delivery,” *Pharmaceutics*, vol. 14, no. 7, Jul. 2022, doi: 10.3390/pharmaceutics14071490.

[10] M. A. Ghanem *et al.*, “Noninvasive acoustic manipulation of objects in a living body”, doi: 10.1073/pnas.2001779117/-/DCSupplemental.

[11] K. Melde, A. G. Mark, T. Qiu, and P. Fischer, “Holograms for acoustics,” *Nature*, vol. 537, no. 7621, pp. 518–522, Sep. 2016, doi: 10.1038/nature19755.

[12] A. Franklin, A. Marzo, R. Malkin, and B. W. Drinkwater, “Three-dimensional ultrasonic trapping of micro-particles in water with a simple and compact two-element transducer,” *Appl Phys Lett*, vol. 111, no. 9, Aug. 2017, doi: 10.1063/1.4992092.

[13] J. Li *et al.*, “Three dimensional acoustic tweezers with vortex streaming,” *Commun Phys*, vol. 4, no. 1, Dec. 2021, doi: 10.1038/s42005-021-00617-0.

[14] J. Nan, H. X. Cao, J. O. Park, E. Choi, and B. Kang, “Tunable Acoustic Tweezer System for Precise Three-Dimensional Particle Manipulation,” *Micromachines (Basel)*, vol. 15, no. 10, Oct. 2024, doi: 10.3390/mi15101240.

[15] Q. Zhou, M. Li, C. Fu, X. Ren, Z. Xu, and X. Liu, “Precise micro-particle and bubble manipulation by tunable ultrasonic bottle beams,” *Ultrason Sonochem*, vol. 75, Jul. 2021, doi: 10.1016/j.ultsonch.2021.105602.

Mode-Switching-Based Control Algorithm in Acoustic Actuator for 3D Microrobot Manipulation in Nonhomogeneous Medium

Hiep Xuan Cao
Ultrasonic Actuator Lab
Korea Institute of Medical Microrobotics
Gwangju, Korea
hiep.caoxuan@kimiro.re.kr

Daewon Jung
Ultrasonic Actuator Lab
Korea Institute of Medical Microrobotics
Gwangju, Korea
jungdaewon@kimiro.re.kr

Hoang Viet Nguyen
Korea Institute of Medical Microrobotics
College of AI Convergence
Chonnam National University
Gwangju, Korea
viethoang20151552@jnu.ac.kr

Seokmin Hong
Medical AI Lab
Korea Institute of Medical Microrobotics
Gwangju, Korea
seokmin.hong@kimiro.re.kr

Jong-oh Park
Korea Institute of Medical Microrobotics
Gwangju, Korea
jop@kimiro.re.kr

Byungjeon Kang
Korea Institute of Medical Microrobotics
Graduate School of Data Science
College of AI Convergence
Chonnam University
Gwangju, Korea
bjkang8204@jnu.ac.kr

***Abstract*—Acoustic technologies offer a promising and versatile toolkit for tracking and manipulating both synthetic and biological objects, ranging from nanometer to millimeter sizes, within complex biological environments. This work reports a Mode-Switching-Based Control Algorithm was developed to precisely control a microrobot. We developed an integrated system employing a single acoustic actuator capable of both transmitting and receiving signals, thereby streamlining the hardware architecture. Experimental results demonstrated a 22.7% enhancement in peak acoustic pressure. Positional accuracy was validated with mean deviations of 80 ± 75 micrometers, 71 ± 70 micrometers, and 204 ± 200 micrometers along the X, Y, and Z axes, respectively well below the microrobot's 600 micrometers body length indicating high-fidelity control. This proof-of-concept highlights the potential for dual-function acoustic devices to enable simultaneous tracking and manipulation, paving the way for advanced biomedical applications.**

Keywords*— *acoustic manipulation, ultrasound imaging, automated manipulation, micro robotics, medical robotics, acoustic actuator

I. INTRODUCTION

Despite renal is the most commonly transplanted organ, mortality among patients awaiting kidney transplantation remains substantial. According to the 2023 annual data report, approximately 20 percent of patients on the renal transplant waitlist succumb during the waiting period, 11.4% of registrants experiencing prolonged waiting intervals exceeding five years with enormous medical costs during the waiting period [1].

Non-contact actuation using acoustics has been studied recently. With the ability to manipulate many kinds of material across wide length scales, acoustically-powered devices facilitate the most diverse applications, from lab-on-a-chip device, microfluidic channel to non-invasive surgeries [2-5]. The use of acoustic to manipulate targeted object within the human body has demonstrated significant potential in various medical applications[6-8]. This innovative therapeutic technology employs focused acoustic wave propagation to deliver concentrated acoustic energy to deep tissue structures with high spatial precision, while preserving the integrity of adjacent parenchymal tissue. However, the human body, when considered as a medium for acoustic propagation, is an inhomogeneous medium due to its varying tissue densities and acoustic properties, leading to complex wave interactions and reflections. Thus, the combination of transmitting and receiving function on a single acoustic actuator can help mitigate the limitation of the conventional system. Therefore, it is essential to develop a reliable acoustic actuator capable of transmitting signals to manipulate targets and receiving signals to navigate through the medium effectively.

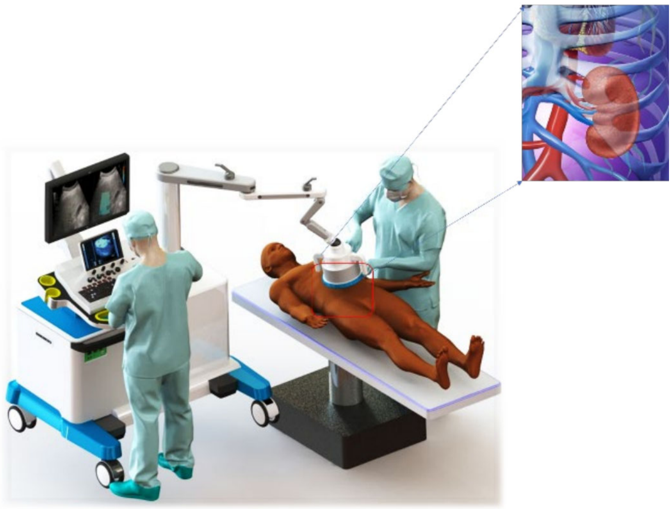


Fig. 1. The concept application for developed system

In this work, a Mode-Switching-Based Control Algorithm was developed to precisely control a microrobot. This algorithm was implemented on a novel system that utilizes a single acoustic actuator for both signal transmission and reception. The control strategy effectively optimizes acoustic radiation force and reduces undesirable side-lobes, leading to more efficient and accurate manipulation.

II. MATERIAL AND METHODS

A. *Acosutic actuator*

To manipulate the microrobots using an external acoustic field, we developed an acoustic actuator system that consists of three main subsystems: (1) a single-side ultrasound transducer (UT) array with 30 identical transducers (JAPAN PROBE, Yokohama, JP); (2) a customized amplifier; and (3) a user control interface developed on LabVIEW 2017. The 30 transducers are operated at the same frequency of 1 MHz and the same operational bipolar voltage from the P3030 power supply unit (Advantek, USA) but with independent phase control. The acoustic twin trap is generated at the focus point by using the phase modulation method, which contains the twin-trap phase (with a π-radian difference on each side of the UA array) and the focus point position information. Thus, the function to control the position of the microrobot can be expressed by the following equations:

$$F_{rad}^{control} = \overset{i=30}{\underset{i=1}{f}} (\varphi_i) \tag{1}$$

$$\varphi_i = \begin{bmatrix} 0 + 2\pi(1 - \mathrm{f}_i(\mathrm{N})) & \quad with \quad i = 1:1:15 \\ \pi + 2\pi(1 - \mathrm{f}_i(\mathrm{N})) & \quad with \quad i = 16:1:30 \end{bmatrix} \tag{2}$$

where φ_i is the phase delay in radian on the i^{th} transducer. N is the number of acoustic wave cycles. $f_i(N)$ is the decimal part of the last cycle. The customized amplifier was developed on the LabVIEW FPGA platform with FPGA PCIe- 7852R hardware. The FPGA module generated the control signal at 3.3 V containing the phase value with a resolution of 2°. Then, the amplifier magnified the signal from 10 Vpp to 100 Vpp with retention of the phase information from the control signal.

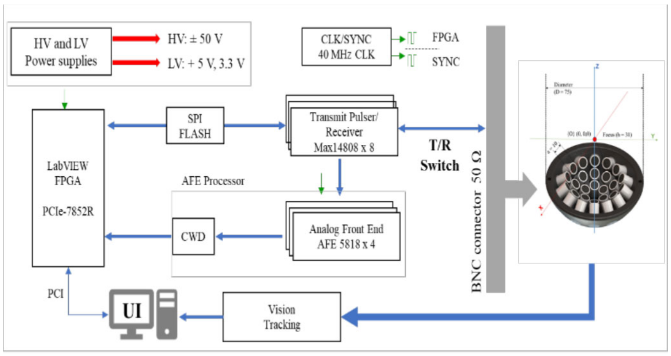


Fig. 2. System block diagram design

The key operation as below:

- The device has two operation modes: Transmit mode and receiving mode support up to 32 transmit/receive channels
- AFE processor include low noise amplifier, Programmable gain amplifier, filters and 14 bit analog to digital converter
- PC-GUI use to pass the front-end parameters to the FPGA control logic and embedded the developed switching control algorithm
- Vision system support for: tracking the target in non-homogeneous medium and focus point calibration.

B. *Mode-Switching-Based control algorithm*

A Mode-Switching-Based Control Algorithm was developed to precisely control a microrobot. This algorithm was implemented on a novel system that utilizes a single acoustic actuator for both signal transmission and reception. The algorithm flow described as below:

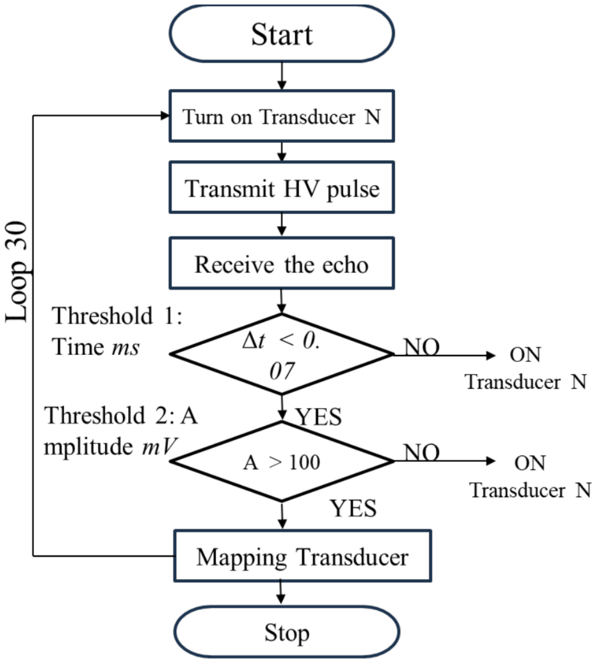


Fig. 3. Switching Algorithm flow chart

1. Initialize system. Turn on transducer N, N = 1~30
2. Generate the transmit pulse at High-Voltage 60 Vpp from system.
3. Switching to Receive mode. Wait echo back to the Receive module
4. Turn ON the input channel for the UT N
5. Threshold 1: Calculate the time travel from UT to the target. If the $\Delta t < 0.07$, The target can be blocked the UT. If NOT, the UT won't block

6. Threshold 2: Calculate the Echo amplitude from Target. If NOT, the target is partially blocked. If the A > 100, The UT is BLOCKED
7. Mapping UT in the array. Then, go back to turn on next UT and process as a loop until all UA is operated.
8. Stop the Pre-scan process.

III. RESULT AND DISCUSSION

A challenge encountered in acoustic manipulation in an ex vivo environment is the difference in permeability of the acoustic wave, which decreases as it moves through porcine fat, muscle tissue, and bone. To check the feasibility of this targeted application, an ex vivo experiment was conducted using a 30 mm thick porcine rib with an area of 120 mm × 100 mm, which was sufficient to cover the surface of the UT. The temperature at the focal point was measured by a thermometer (MTM-380SD, Lutron Instruments) during the following operating period at 5 min, 10 min, and 15 min to check the heating effect. The temperature rose from 24.4° to 27.6°. The temperature at the focal point was not over 27.6° during the operation period.

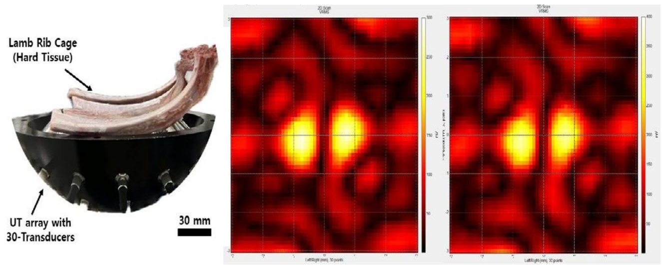


Fig. 4. Ex-vivo experiment setup

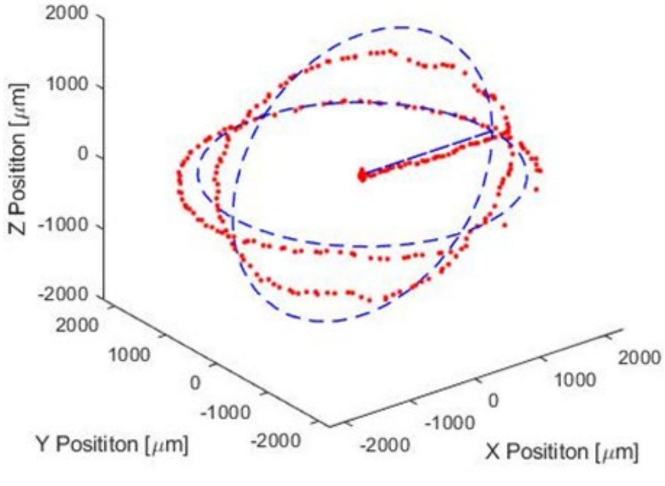


Fig. 5. Path tracking of microrobot

Evaluation of the switching-based control algorithm revealed a 22.7% increase in peak acoustic pressure (from 291kPa to 357kPa) at a constant power input. The system's manipulation accuracy was then quantified by tracking a microrobot along a spherical path. The measured mean positional deviations in the X, Y, and Z axes were 80±75μm, 71±70μm, and 204±200μm, respectively. These errors are substantially smaller than the microrobot's 600μm body length, confirming high-fidelity control.

ACKNOWLEDGMENT

This research was supported by a grant of the Korea Health Technology Development R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2023-00302153 and RS-2023-00302148).

REFERENCES

[1] K. L. Lentine et al., "OPTN/SRTR 2023 Annual Data Report: Kidney," American Journal of Transplantation, vol. 25, no. 2, 2025/02/01, doi: 10.1016/j.ajt.2025.01.020.

[2] H. X. Cao et al., "Acoustically Driven Cell-Based Microrobots for Targeted Tumor Therapy," Pharmaceutics 2022, Vol. 14, Page 2143, vol. 14, no. 10, 2022-10-09, doi: 10.3390/pharmaceutics14102143.

[3] M. Ghanem et al., "Noninvasive acoustic manipulation of objects in a living body," Proceedings of the National Academy of Sciences, pp. 202001779-202001779, 2020, doi: 10.1073/pnas.2001779117.

[4] X. Lu et al., "A Human Microrobot Interface Based on Acoustic Manipulation," ACS Nano, vol. 13, no. 10, pp. 11443-11452, 2019, doi: 10.1021/acsnano.9b04930.

[5] H. X. Cao, V. D. Nguyen, J.-O. Park, E. Choi, and B. Kang, "Acoustic Actuators for the Manipulation of Micro/Nanorobots: State-of-the-Art and Future Outlooks," Micromachines, vol. 15, no. 2, p. 186, 2024, doi: 10.3390/mi15020186.

[6] H. X. Cao et al., "Micromotor Manipulation Using Ultrasonic Active Traveling Waves," Micromachines 2021, Vol. 12, Page 192, vol. 12, no. 2, 2021-02-13, doi: 10.3390/mi12020192.

[7] A. Ozcelik et al., "Acoustic tweezers for the life sciences," Nature Methods, vol. 15, no. 12, pp. 1021-1028, 2018, doi: 10.1038/s41592-018-0222-9.

[8] M. Kaynak, A. Ozcelik, A. Nourhani, P. E. Lammert, V. H. Crespi, and T. J. Huang, "Acoustic actuation of bioinspired microswimmers," Lab on a Chip, vol. 17, no. 3, pp. 395-400, 2017, doi: 10.1039/C6LC01272H.

On-Device Vision Language Model Based Natural Language Mission Execution Framework

Seunggyu Lee
Department of Aerospace Mechanical Engineering
Korea Aerospace University
Goyang-Si, South Korea
ruga1026@kau.kr

Seungjun Oh
Department of Aerospace Mechanical Engineering
Korea Aerospace University
Goyang-Si, South Korea
awana970630@kau.kr

Sangwoo Jeon
Department of Aerospace Mechanical Engineering
Korea Aerospace University
Goyang-Si, South Korea
jsw5413@naver.com

Youngju Park
Department of Aerospace Mechanical Engineering
Korea Aerospace University
Goyang-Si, South Korea
qkrdudwngg@naver.com

Siwon Lee
Department of Electrical and Electronic Engineering
Korea Aerospace University
Goyang-Si, South Korea
leesiwon0305@kau.kr

Ghilmo Choi
Department of Aerospace Mechanical Engineering
Korea Aerospace University
Goyang-Si, South Korea
ggghilmo@naver.com

Dae-Sung Jang
School of Aerospace and Mechanical Engineering
Korea Aerospace University
Goyang-Si, South Korea
dsjang@kau.ac.kr

***Abstract*—This work presents a framework that enables robots to assist with everyday tasks by leveraging Vision-Language Models (VLMs). Our framework takes both environmental information and natural language commands from users as inputs. An on-device VLM generates and validates a behavioral plan, represented as a sequence of pseudo-functions, using an actor–critic structure. These pseudo-functions are then connected to the low-level control system of a UGV to demonstrate the feasibility of the approach. Experimental results show that the framework can reliably interpret and execute user commands, highlighting its potential as a step toward practical robot assistance in daily life.**

***Keywords*—vision language model, on-device, robot, planning**

I. INTRODUCTION

As machine learning technology advances, robots are expected to play an increasingly important role in assisting people with everyday tasks. Achieving this vision requires robots to effectively perceive their surroundings, interpret human instructions, and execute appropriate actions in real-world environments.

Recent progress in Vision-Language Models (VLMs) has demonstrated strong capabilities in grounding natural language instructions to visual inputs [1]. However, applying VLMs directly to embodied decision-making remains challenging, particularly in bridging the gap between high-level reasoning and low-level control. Existing approaches often struggle with generating reliable action sequences that can be executed on physical robotic platforms.

In this work, we present a framework that integrates on-device VLMs with an Actor–Critic structure, similar to the actor–critic design used in [2], to generate and validate behavioral plans. User commands, combined with environmental information, are translated into sequences of pseudo-functions that serve as intermediate representations. These pseudo-functions are then mapped to the low-level control system of an unmanned ground vehicle (UGV, implemented using a small RC car [3]) to evaluate the feasibility of our approach. Experimental results demonstrate that the system can interpret and execute user commands effectively, showing promise for practical robot assistance in daily scenarios.

II. SYSTEM FRAMEWORK

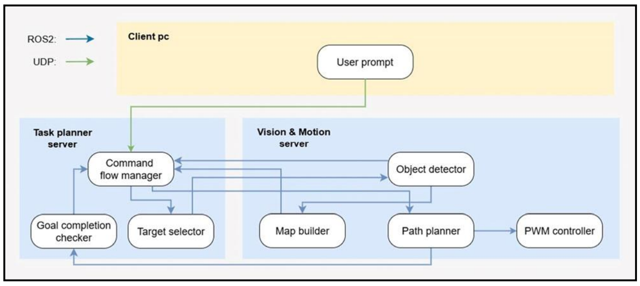


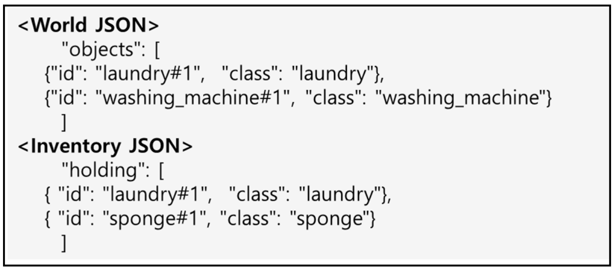
Fig. 1. Overall software architecture of the proposed framework.

A. Overview of the Framework

The proposed framework enables a mobile robot to interpret natural language instructions and execute them in real-world environments by leveraging a VLM with an actor–critic structure. Fig. 1 illustrates the overall software architecture.

B. Input Processing

The system accepts two inputs: (1) a natural language command from the user and (2) environmental data generated through Simultaneous Localization and Mapping (SLAM) [4] and object detection. The environmental data is represented as a top-view map image with detected objects as well as structured JSON file, namely *World JSON*. The *Inventory JSON* contains information about the objects the robot is holding. The *World JSON* stores spatial information about all detected objects, while the *Inventory JSON* keeps track of the items currently held by the robot. Both types of JSON files contain ID of the each object and their class, as shown in Fig.2.



```
<World JSON>
      "objects": [
   {"id": "laundry#1",   "class": "laundry"},
   {"id": "washing_machine#1",  "class":  "washing_machine"}
      ]
<Inventory JSON>
      "holding": [
   { "id": "laundry#1",   "class": "laundry"},
   { "id": "sponge#1", "class": "sponge"}
      ]
```

Fig. 2. Example of World JSON and Inventory JSON file.

Using the World JSON, we construct a top-view map image as shown in Fig.3 to make the VLM better understand the surroundings. Using the odometry data derived from SLAM, it also shows the current position and orientation of the robot.

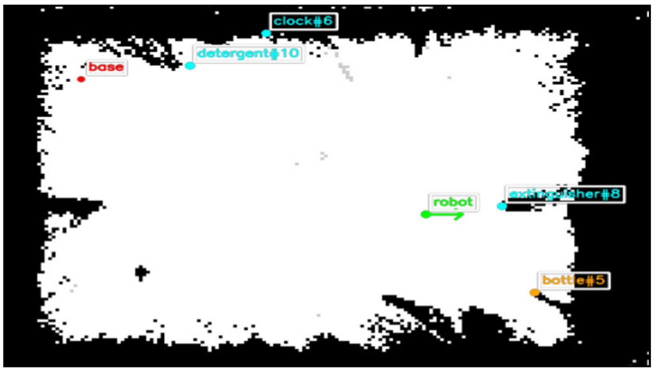


Fig. 3. Example of the top-view map image.

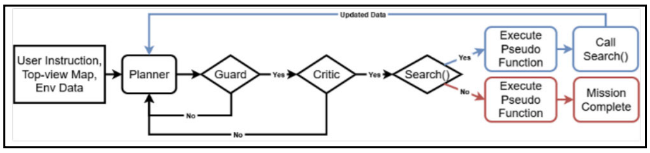


Fig. 4. Overall software architecture of the proposed framework.

C. *Planner*

Fig.4 illustrates the flow of Actor-Guard-Critic structure. The Planner, implemented with an on-device VLM, generates a behavioral plan in the form of a sequence of pseudo-functions. These pseudo-functions abstract high-level actions such as move_to(object), pick(object), place(object), search(object), and return_to_base() (Table I). By using pseudo-functions as intermediate representations, the framework bridges natural language understanding with low-level robot control.

TABLE I. AVAILABLE PSEUDO-FUNCTIONS

Function(argument)	Description
move_to(object)	Navigates the robot to the specified object's location on the map.
pick(object)	Grasps and lifts the target object; executable only when the robot is near the object.
place(object)	Places a held object at the current location or designated position.
search(object)	Actively explores the environment to locate the specified object and update the map.
return_to_base()	Returns the robot to its starting position (base station).

D. *Guard*

To ensure plan reliability, a rule-based Guard module is introduced before the Critic stage. The Guard checks whether the pseudo-functions generated meet minimal validity constraints (e.g., verifying if an object to be placed exists in the inventory). Invalid plans are rejected, and the Planner is prompted to re-generate an alternative plan with corrected reasoning.

E. *Critic*

The Critic evaluates the proposed plan with respect to the current environment and execution history. Unlike the Guard, which performs rule-based filtering, the Critic performs contextual validation to identify logical inconsistencies or hallucinated actions. If the plan is deemed invalid, feedback is sent to the Planner for replanning. This actor–critic cycle improves the robustness of action generation in uncertain or partially observable environments.

F. *Execution on UGV*

Finally, we implemented the validated pseudo-functions on a small UGV by mapping them to its low-level control system. In the ROS 2 environment, these functions were converted into motor control commands using Nav2 package [5], allowing the robot to autonomously execute navigation tasks in the test field.

III. EXPERIMENTS

A. *Experimental Setup*

To evaluate the feasibility of the proposed framework, we conducted experiments on the UGV platform described above. The robot was equipped with an onboard computer system consisting of a Jetson Orin NX 16GB [6] for running the VLM and a Jetson Orin Nano Super 8GB [7] for auxiliary processes. The platform operated under ROS2 and was integrated with object detection modules and a SLAM package to construct a two-dimensional map of the environment. Environmental information was maintained in structured data formats (World JSON and Inventory JSON), which were continuously updated during navigation.

B. *Test Scenarios*

To validate the framework, we designed scenarios that mimic everyday assistance tasks in household environments. Each scenario was given as a natural-language instruction; the Planner converted the instruction into pseudo-functions, which were then validated by the Guard and Critic before execution.

A representative case is the instruction, 'Go to the trash bin.' Here, the information about the object 'trash bin#1' was provided. The initial plan generated by the Planner was:

1) move_to("trash bin#1")
2) place("trash bin#1")
3) return_to_base()

The Guard detected that the place() action was invalid since no object was being carried in the inventory. It rejected the plan and prompted replanning. The revised plan was:

1) move_to("trash bin#1")
2) return_to_base()

We also considered multi-step instructions requiring dynamic replanning. Given "Bring the dog and the apple to the person." the system proceeded iteratively as follows:

1) search("dog") → locate the dog

2) move_to("dog#1"), pick("dog#1"), move_to("person-#1"), place("dog#1")

3) search("apple") → locate the apple

4) move_to("apple#1"), pick("apple#1"), move_to("person#1"), place("apple#1"), return_to_base()

C. *Complex Task Execution*

During mission execution, the command-flow manager maintained the World JSON and Inventory JSON while updating a top-view map built from SLAM and object detection. The Planner produced pseudo-function sequences, which were screened by the Guard before the Critic performed contextual checks using the current environment and execution history. When the Planner proposed logically inconsistent actions (e.g., place() without a held object), the Guard generated an error message and triggered replanning, preventing such plans from reaching the Critic or the robot controller.

For partially observable scenes, the Planner explicitly invoked search() to acquire additional observations, augment the map/JSON, and restructure the plan. Treating search() as a pivot step reduced the burden of generating long one-shot plans and enabled reliable multi-iteration completion of complex instructions (e.g., sequentially delivering the dog and the apple to a person). The overall Actor–Guard–Critic loop followed the command-flow design shown in the project poster and used the predefined function set (move_to, pick, place, search, return_to_base) as the execution vocabulary.

D. Results

Guard-mediated validation. Across simple “go-to” commands and multi-step delivery tasks, the Guard consistently filtered invalid plans (e.g., rejecting place() when the inventory was empty) and prompted the Planner to regenerate corrected plans, which then executed successfully on the UGV. This reduced unnecessary escalations to the Critic and improved the overall efficiency of the loop.

Adaptivity under partial observability. For longer-horizon tasks, the Planner issued search() to discover required objects, after which the map/JSONs were updated and subsequent steps executed on the UGV.

IV. DISCUSSION

A. Design choices and implications

Using pseudo-functions as an intermediate representation bridges natural-language intent and executable robot actions while enabling lightweight validity checks prior to execution. This abstraction works well with the command-flow manager that maintains a merged 2D map alongside World and Inventory JSONs, allowing plans to be reasoned about with explicit object-level semantics.

B. Role of the Guard versus the Critic

Placing a rule-based Guard before the Critic prevents trivial logical errors (e.g., placing an item not in the inventory) from reaching contextual evaluation. In our trials, this reduced unnecessary Critic invocations and guided the Planner to regenerate feasible alternatives with low computational overhead—an important property for on-device settings.

C. Information shaping via search()

Beyond enabling exploration under partial observability, specifying target object types in search() lets the system filter the surrounding objects recorded in the World and Inventory JSONs and the merged map, passing only task-relevant information to the VLM. This helps maintain focus and reduces context dilution during planning, especially in multi-object missions.

D. Current limitations and future directions

Operating with an on-device VLM constrains model size and occasionally yields hallucinated or instruction-deviating proposals; long-horizon missions may require multiple re-planning iterations. A promising direction is to leverage advanced training strategies, such as fine-tuning and reinforcement learning, to better align smaller models with task requirements and reduce failure modes. These approaches may help overcome the inherent limitations of compact architecture while maintaining deplorability on resource-constrained platforms. From a systems perspective, our end-to-end mapping from high-level specifications to low-level control on a physical UGV confirms feasibility and provides a foundation for scaling to more cluttered environments and longer missions.

V. CONCLUSION

We presented an on-device VLM-based framework that converts natural-language commands into sequences of validated pseudo-functions and executes them on a physical UGV. The Actor–Guard–Critic design improved reliability by filtering invalid plans early and enabling contextual checks before execution, while the search() pivot supported staged planning and adaptive map/JSON updates under partial observability. Experiments demonstrated successful execution of both simple “go-to” commands and multi-step delivery tasks on hardware, establishing the practicality of the approach from language to motor commands. Future work will explore advanced training strategies, such as fine-tuning and reinforcement learning, to better align compact models with task requirements and mitigate limitations while preserving deplorability on resource-constrained platforms.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016.

[2] H. Song, K. Yin, Y. Ma, D. Khashabi, and D. Roth, "Prompt with actor-critic editing for instruction tuning," arXiv preprint arXiv:2308.10088, 2023.

[3] TRAXXAS, “TRX-4 Land Rover Defender,” [Online], https://traxxas.com/82056-4-trx-4-land-rover-defender [Accessed: Jul. 18, 2025].

[4] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM: A factored solution to the simultaneous localization and mapping problem," in Proc. AAAI Conf. on Artificial Intelligence, 2002.

[5] Open Navigation Group, “Navigation2 Documentation,” [Online]. Available: https://docs.nav2.org/. [Accessed: Jul. 18, 2025].

[6] NVIDIA, *Jetson Orin NX Series Data Sheet*, v0.5, DS-10712-001, [Online]. Available: https://developer.download.nvidia.com/assets/embe dded/secure/jetson/orin_nx/docs/Jetson_Orin_NX_D S-10712-001_v0.5.pdf. [Accessed: Jul. 18, 2025].

[7] NVIDIA, "Jetson Orin Nano Developer Kit," [Online]. Available: https://www.nvidia.com/en-us/ autonomous-machines/embedded-systems/jetson-o rin/nano-super-developer-kit/. [Accessed: Jul. 18, 2025

Situation-Aware Risk Quantification for UAV-Satellite Communications using Semantic Anomaly Detection

Junbeom Park
Dept. of Computer Engineering
Korea Aerospace University
Goyang, South Korea
jbpark@kau.kr

Zizung Yoon
Dept. of Smart Drone Engineering
Korea Aerospace University
Goyang, South Korea
z.yoon@kau.ac.kr

Jongsou Park
Dept. of Computer Engineering
Korea Aerospace University
Goyang, South Korea
jspark@kau.ac.kr

***Abstract*—Unmanned Aerial Vehicle (UAV)–satellite communications are increasingly critical for mission operations yet remain vulnerable to diverse packet-level and protocol-level threats, including jamming and spoofing. Existing intrusion detection systems (IDSs), however, lack situational awareness, often assessing anomalies with static severity regardless of mission context. This study introduces a semantic detection approach that integrates DistilBERT-based packet classification with a mission-phase–aware risk score to provide situationally adaptive anomaly assessment. The proposed framework is evaluated on a custom dataset extended to six traffic classes, embedding packet-level symptoms of threats across both the packet and protocol layers. Experimental results demonstrate that DistilBERT achieves 99.0% accuracy and a 98.9% macro-F1 score, while maintaining an inference latency of 26.2ms per packet—making it suitable for on-board deployment. More importantly, the proposed Situational Risk Score effectively differentiates threat criticality across mission phases. It reveals that spoofing and privilege abuse are most critical during takeoff and landing, whereas jamming poses the greatest risk during mission execution. These findings confirm that combining semantic anomaly detection with situational context not only ensures high accuracy but also delivers operationally relevant risk quantification for UAV–satellite systems.**

Keywords—UAV-Satellite Communications, Cybersecurity, Semantic Anomaly Detection, DistilBERT, Situational Awareness

I. INTRODUCTION

UAV–satellite communications are becoming integral to mission-critical domains such as disaster response, remote sensing, and defense [1], [2]. By extending connectivity beyond line-of-sight, satellite links enable global UAV operations yet simultaneously introduce vulnerabilities to packet-level and protocol-level threats, including jamming and spoofing [3]. Incidents such as GPS spoofing or intentional jamming demonstrate how disruptions at the communication layer can escalate into mission failure, particularly during safety-critical phases such as takeoff and landing [4]. Nevertheless, securing these vital links remains a formidable challenge. Existing detection paradigms—from conventional IDSs lacking semantic context to large language models (LLMs) constrained by computational costs and unvalidated data—still fall short [5]. In UAV–satellite operations, where threat severity is highly context dependent, bridging this gap requires a robust detection paradigm that integrates semantic reasoning with situational awareness.

The semantic anomaly-detection approach proposed in this paper confronts these challenges. It serializes packets for analysis by a lightweight Transformer-based model, DistilBERT [6], enabling contextual interpretation of inter-field dependencies often missed by conventional IDSs. The classification outputs are subsequently fused with a mission-phase–aware risk score that maps threat severity to situational priorities. This combined approach is validated on a scenario-driven dataset whose six classes encompass conventional cyberattacks and packet-level symptoms of protocol-level disruptions such as jamming and spoofing. Ultimately, this work bridges the gap between low-level anomaly detection and high-level mission context, providing a foundation for integrating AI-driven security with operational decision-making in UAV–satellite networks.

II. RELATED WORK

Prior studies in UAV and satellite network security have addressed threats arising from both RF-based attacks and protocol-level intrusions. Alrefaei et al. [1], for instance, surveyed jamming and spoofing in UAV networks, demonstrating not only their critical impact on availability and integrity but also the limitations of traditional countermeasures. In parallel, large language models (LLMs) have emerged as promising tools for security applications. Chen et al. [5] presented a comprehensive survey of LLM-based cyberthreat detection, identifying key opportunities and challenges. Building on this trend, Ferrag et al. [7] proposed a lightweight BERT-based intrusion detection system for IoT/IIoT, and Hassanin et al. [8] introduced PLLM-CS for satellite networks. While these works demonstrate the feasibility of semantic detection, they primarily rely on synthetic data and lack mission-phase context. Zhao et al. [9] applied Transformer-based models for security assessment in UAV swarms, highlighting the value of contextual reasoning but without linking it to packet-level anomaly detection. Our work addresses these gaps by combining DistilBERT-based semantic packet classification with mission-phase–aware risk scoring for UAV–satellite communications.

III. PROPOSED METHODOLOGY

We propose a four-stage pipeline for situation-aware risk quantification in UAV–satellite communications. The process begins with the construction of a scenario-driven dataset using a fixed 15-field schema that reflects realistic communication flows. Each packet is subsequently transformed into a sentence-like representation to enable semantic modeling without manual feature engineering. These serialized packets are then processed by a lightweight DistilBERT model to perform context-aware anomaly classification. Finally, the classifier outputs are integrated into a mission-phase–aware risk-scoring scheme that quantifies the severity of detected anomalies across different flight stages. This pipeline

TABLE I
STRUCTURE AND DESCRIPTION OF THE 15-FIELD UAV–SATELLITE DATASET SCHEMA

Field Name	Description	Example Values
timestamp	Packet generation time in ISO 8601 format	2025-07-01T03:15:20
src / dst	Valid communication nodes from NORMAL_LINKS	gw1→iot, leo→gcs[a]
priority	Message priority determined by msg_type	LOW, MEDIUM, HIGH, CRITICAL
src_port / dst_port	Ports assigned per node from SRC_PORT_MAP / DST_PORT_MAP	gw1:1883, leo:3001[b]
src_region / dst_region	Region code from REGION_MAP	AS→AF, EQ→SP
orbit_class	Orbit category derived from ORBIT_CLASS_MAP	LEO, MEO, N/A
msg_type	Message type based on VALID_MSG_TYPES per src–dst pair	telemetry, data, command, status, ack, alert[c]
payload_type	Field type determined by msg_type	TEMP, SIZE, MOVE, SIGNAL_LOSS, NORMAL[d]
payload	Formatted content generated per payload_type	`TEMP=24.5, command=RESET`
label	Class label for anomaly detection	Normal, Injection, Replay, Privilege Abuse, Jamming, Spoofing
ttl (time-to-live)	TTL value based on src/dst role	64, 128, 200, 255
flags	Control flags by msg_type	ACK, SYN, RST[e]

[a] Examples include (gw1, iot), (rt, gw2), (gcs, meo), based on NORMAL_LINKS.

[b] Port numbers are statically assigned per node based on system design; for example, ground nodes use ports in the 1000 range (e.g., gw1:1883), while space-segment nodes use ports in the 3000 range (e.g., leo:3001).

[c] Allowed message types are predefined for each src–dst pair in the system design.

[d] Mapping: telemetry→{TEMP, HUM, POS, BATT}; data→{COORD, SIZE, DATA_TYPE, REF_ID}; command→{ACTIVATE, MOVE, RESET, ...}; ack→{RECEIVED, EXECUTED}; alert→{ANOMALY_DETECTED, ...}; status→{NORMAL, LOW_BATTERY, ...}.

[e] Flag options include ACK, PSH, ENC, SYN, and RST, as listed in FLAGS_BY_MSGTYPE.

establishes a direct link from packet semantics to mission-level risk evaluation, with the details of each stage presented in the following subsections.

A. Scenario-Driven Dataset and Threat Model

The proposed approach is evaluated on a custom dataset designed for UAV–satellite communication scenarios, which follows the fixed 15-field schema presented in Table I. These fields capture the temporal, spatial, and logical attributes of packet flows and are designed to reflect protocol families such as CSP, CCSDS, and NMEA, thereby encompassing both satellite and UAV telemetry semantics. The threat model assumes an adversary capable of executing both network-level intrusions and RF-layer disruptions, with goals ranging from simple communication denial to sophisticated manipulation of the UAV operational state. To represent these threats, the dataset's labeling taxonomy consists of one Normal class and five attack classes: Injection, Replay, Privilege Abuse, Jamming, and Spoofing. As real-world traces of such attacks are scarce, synthetic attacks were generated and modeled as observable packet-level symptoms embedded within the existing fields. For instance, Jamming is modeled as a communication disruption, manifested through irregular timing (e.g., bursty or delayed timestamp gaps), repeated status/alert messages from link-recovery attempts, and occasionally reduced ttl values. Spoofing is represented as data falsification, primarily involving manipulated POS coordinates in the payload, violations of predefined src-dst role/region constraints, and unauthorized command attempts. To prepare this structured data for semantic analysis, each packet is serialized into a sentence-like representation. This process preserves contextual dependencies among fields and translates the synthesized threat symptoms into a format that a Transformer-based model can effectively interpret.

B. DistilBERT-Based Semantic Anomaly Classification

The detection model is designed as an end-to-end semantic classification pipeline for structured UAV–satellite packet streams. Unlike traditional multi-stage IDSs that separate parsing, inference, and decision-making, the proposed approach integrates sentence-based packet construction with Transformer based embedding to enable holistic and context-aware threat detection. This integrated design minimizes error propagation often observed between disjoint stages, directly translating structured packet data into contextual security assessments. Each packet is serialized into a sentence-like representation that encodes its spatial, temporal, and logical semantics, including orbit–region associations, role-based message constraints, and consistency between flags and payloads. The sequences are tokenized using the WordPiece tokenizer and processed through the six encoder layers of DistilBERT. The final hidden state of the [CLS] token serves as an aggregate semantic representation, which is subsequently passed to a fully connected classifier that outputs one of six categories. The classifier distinguishes one normal and five attack classes:

- **Normal**: All fields comply with authorized communication flows.
- **Replay**: Packets reuse timestamp–payload pairs, violating temporal integrity.
- **Privilege Abuse**: Low-privilege nodes attempt to issue unauthorized high-level commands.
- **Jamming**: Characterized by irregular timing (e.g., bursty or delayed timestamp gaps), repetitive status/alert messages, or reduced ttl values.
- **Spoofing**: Involves falsified POS coordinates, invalid src-dst combinations, or unauthorized command payloads.

This design enables the detection of anomalies that may be structurally valid yet contextually inconsistent with mission policies. By capturing inter-field dependencies through self-attention, DistilBERT effectively identifies both conventional cyberattacks and signal-level threats, producing outputs that are directly aligned with operational security requirements in UAV–satellite networks.

C. DistilBERT-Based Semantic Anomaly Classification

Conventional IDSs typically assign a fixed severity level to each attack type—a static approach that fails to reflect the dynamic context of UAV missions. For example, a GPS spoofing attack might represent a low-grade threat during high altitude cruise but become critical during an autonomous landing phase. Therefore, a Situational Risk Score is introduced to combine classifier confidence, CIA-based intrinsic severity, and mission-phase weighting.

We first define the base risk score as

$$R_{\text{base}} = I_{\text{attack}} \times P_{\text{detect}},$$

where I_{attack} denotes the intrinsic severity of an attack with respect to the CIA triad, and P_{detect} represents the classifier confidence in the predicted class.

To incorporate mission context, this base score is extended to the situational risk score R_{SA}:

$$R_{\text{SA}} = R_{\text{base}} \times W_{\text{sit}} = \left(I_{\text{attack}} \times P_{\text{detect}}\right) \times W_{\text{sit}},$$

where W_{sit} represents the phase-specific situational weight.

Situational weights are defined for each attack type and mission phase—{Takeoff, Mission Execution, Cruise, Landing}—and are derived through scenario-driven, AHP-style pairwise judgments. The final weights are obtained through a three-step process: (i) normalization per attack to a mean value of 1.0, (ii) damping with an exponent $\gamma = 0.7$ to mitigate over-amplification, and (iii) soft clipping to the range $[0.80,\ 1.40]$ for stability. Table II summarizes these resulting values.

TABLE II
Situational Weights W_{SIT} per Attack Across Mission Phases

Phase	Inj.	Rep.	Priv.	Jam.	Spoof.
TK (Takeoff)	1.16	0.86	1.25	0.84	1.20
ME (Mission Exec.)	0.88	1.27	0.88	1.40	0.80
CR (Cruise)	0.80	1.01	0.80	0.80	0.80
LD (Landing)	1.16	0.86	1.07	0.96	1.20

The design of the weighting scheme, summarized in Table II, ensures that risk levels adapt to mission context. For instance, spoofing and privilege abuse are assigned higher weights during takeoff and landing, where integrity is most critical. In contrast, jamming receives the highest weight during mission execution, where availability is the dominant concern. These settings ensure that the same attack type is evaluated differently depending on the mission phase, thereby moving beyond static, fixed-severity scoring. The resulting score R_{SA} can also be discretized into qualitative levels (e.g., Low, Medium, High) to support mission-aware decision-making.

IV. Experimental Results and Discussion

For the base model, DistilBERT was selected owing to its efficiency and suitability for resource-constrained UAV–satellite systems. With approximately 66 million parameters, it provides a lightweight alternative to BERT while retaining strong detection capability [6]. The model achieves an average inference latency of 26.2ms and consumes approximately 789 MB of GPU memory, representing more than a 70% latency reduction and a 35–40% decrease in memory usage compared with BERT-base. These performance metrics confirm the feasibility of deploying the proposed model on board or on edge devices, thereby addressing the critical requirement for low-latency and memory-efficient solutions in such resource-constrained environments.

A. *Classification Performance Evaluation*

Table III summarizes the performance of the baseline models on the UAV–satellite six-class packet classification task.

TABLE III
Performance of Detection Models on UAV–Satellite Six-Class Packet Classification

Model	Accuracy (%)	Macro-F1 (%)	Latency (ms)
Random Forest	87.6	87.0	7.1
LSTM	92.8	92.3	16.6
BERT-base	98.5	98.3	100.0
DistilBERT	99.0	98.9	26.2

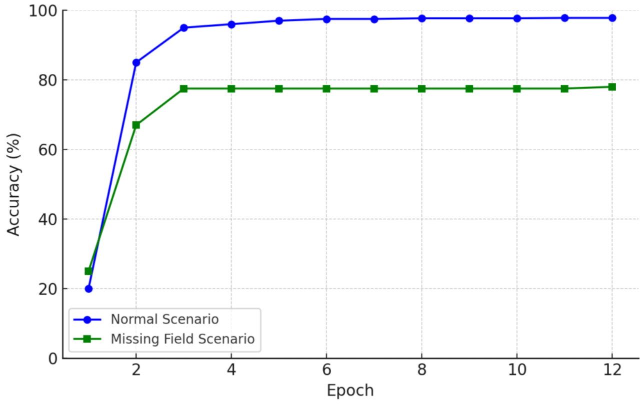


Fig. 1. Validation accuracy trends across epochs for normal and missing-field inputs in UAV–satellite six-class classification.

As shown in Table III, the Random Forest model achieves only moderate accuracy (87.6%), whereas the LSTM improves performance to 92.8% but struggles with complex cross-field dependencies. Among the Transformer-based models, BERT-base attains 98.5% accuracy but suffers from high inference latency (~ 100ms) and substantial memory requirements (>1.2 GB). DistilBERT provides the best trade-off, delivering 99.0% accuracy and a 98.9% macro-F1 score with significantly lower computational cost. Figure 1 further illustrates the model robustness under both normal and missing-field scenarios. The close alignment of the two curves indicates effective convergence and resilience even with partially degraded packet inputs. This property is critical for ensuring operational reliability in UAV–satellite networks, which often encounter bandwidth limitations and intermittent transmission faults.

TABLE IV
Situational Risk Scores R_{SA} Across Mission Phases

Attack	TK	ME	CR	LD
Injection	0.651	0.494	0.449	0.651
Replay	0.386	0.570	0.453	0.386
Privilege Abuse	0.916	0.645	0.586	0.784
Jamming	0.438	0.730	0.417	0.500
Spoofing	0.858	0.572	0.572	0.858

B. *Analysis of Situational Risk Variation Across Mission Phases*

While anomaly detection outputs identify threats, they alone do not convey their operational impact. To address this limitation, the situational risk score R_{SA} integrates classifier confidence, intrinsic CIA-based severity, and mission-phase weights. This formulation allows each attack to be evaluated relative to its potential impact across different flight phases.

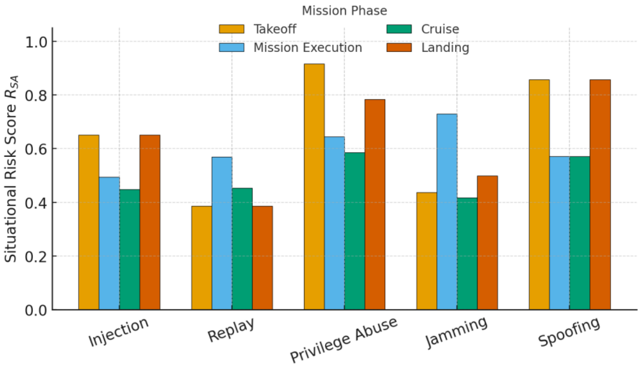


Fig. 2. Situational risk scores ($\boldsymbol{R}_{\mathrm{SA}}$) across mission phases for each attack type in UAV–satellite operations.

The analysis, illustrated in Fig. 2 and detailed in Table IV, reveals distinct risk patterns across the mission phases. Privilege abuse and spoofing exhibit the highest risk scores during takeoff and landing, where system integrity is paramount. In contrast, jamming dominates during mission execution owing to its critical impact on availability. Although injection and replay attacks show comparatively lower values, they still represent persistent risks that can accumulate over time. These phase-dependent variations demonstrate how the R_{SA} score transforms anomaly detection into mission-aware risk assessment, thereby offering practical guidance on which threats warrant priority at different stages of flight.

V. Conclusion and Future Work

This paper presented a situation-aware risk quantification approach for UAV–satellite communications based on semantic anomaly detection. The proposed methodology combines a lightweight DistilBERT classifier for six anomaly categories, including signal-level threats, with a mission-phase-dependent risk score that translates detection outputs into actionable situational awareness. On a custom six-class dataset, the model achieved 99.0% accuracy and a 98.9% macro-F1 score, outperforming conventional baselines. Furthermore, the risk score analysis revealed that spoofing and privilege abuse are most critical during takeoff and landing, whereas jamming dominates during mission execution. These findings demonstrate that the proposed approach not only delivers robust anomaly detection but also supports mission-aware decision-making in UAV–satellite systems. By linking semantic anomaly detection with situational risk scoring, this work provides a foundation for integrating AI-driven intrusion detection with mission-level decision support in aerospace networks.

While this study validates the proposed approach, several avenues for future research remain. A crucial next step is to evaluate the model robustness under real-world conditions using additional datasets and field experiments. The risk score formulation could also be enhanced by exploring alternative CIA weightings and temporal factors. Ultimately, this line of research could extend toward a unified framework for aerospace security assessment and response.

Acknowledgment

This work was supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Ministry of Trade, Industry and Energy (MOTIE), Korea Government (No. RS-2021-KI002499, HRD Program for Industrial Innovation).

References

[1] F. Alrefaei, A. Alzahrani, H. Song, and S. Alrefaei, ”A survey on the jamming and spoofing attacks on the unmanned aerial vehicle networks,” in Proc. IEEE Int. IoT, Electron. Mechatronics Conf. (IEMTRONICS), 2022, pp. 1–7. [Online]. Available: https://doi.org/10.1109/IEMTRONICS55184.2022.9795809

[2] S. Salim, N. Moustafa, and M. Reisslein, “Cybersecurity of satellite communications systems: A comprehensive survey of the space, ground, and links segments,” IEEE Commun. Surveys Tuts., vol. 27, no. 1, pp. 372–425, 2025.

[3] Y. Mekdad, A. Aris, L. Babun, A. El Fergougui, M. Conti, R. Lazzeretti, and A. S. Uluagac, ”A survey on security and privacy issues of UAVs,” Comput. Netw., vol. 224, p. 109626, 2023. [Online]. Available: https://doi.org/10.1016/j.comnet.2023.109626

[4] S. MahmoudZadeh, A. Yazdani, Y. Kalantari, B. Ciftler, F. Aidarus, and M. O. Al Kadri, ”Holistic review of UAV-centric situational awareness:, Applications,limitations, and algorithmic challenges,” Robotics, vol. 13, no. 8, p. 117, 2024. [Online]. Available: https://doi.org/10.3390/robotics13080117

[5] Y. Chen et al., “A survey of large language models for cyber threat detection,” Comput. Secur., 2024. [Online]. Available: https://doi.org/10.1016/j.cose.2024.104016

[6] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter,” arXiv:1910.01108, Oct. 2019. [Online]. Available: https://arxiv.org/abs/1910.01108

[7] M. A. Ferrag et al., “Revolutionizing cyber threat detection with large language models: A privacy-preserving BERT-based lightweight model for IoT/IIoT devices,” IEEE Access, vol. 12, pp. 23733–23750, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3363469

[8] M. Hassanin et al., “PLLM-CS: Pre-trained large language model (LLM) for cyber threat detection in satellite networks,” Ad Hoc Netw., vol. 166, p. 103645, 2024. [Online]. Available: https://doi.org/10.1016/j.adhoc.2024.103645

[9] D. Zhao, P. Shen, X. Han, and S. Zeng, ”Security situation assessment in UAV swarm networks using TransReSE: A Transformer-ResNeXt-SE based approach,” Veh. Commun., vol. 50, p. 100842, 2024. [Online]. Available: https://doi.org/10.1016/j.vehcom.2024.100842

Design, fabrication and evaluation of 6R-Pantograph Antenna mechanism

1st Hyeongseok Kang
Department of Mechanical and Aerospace Engineering
Korea Aerospace University
Goyang-si, Republic of Korea
dwtw0115@kau.kr

2nd Byungkyu Kim
Department of Mechanical and Aerospace Engineering
Korea Aerospace University
Goyang-si, Republic of Korea
bkim@kau.ac.kr

***Abstract*— Research on the development of satellite antenna technologies for rapidly transmitting large volumes of data has been actively conducted. Since the performance of an antenna is determined by the size of its aperture, the development of large antenna deployment mechanisms is essential. Accordingly, the 6R-Pantograph Antenna, which combines the conventional 6R Ring-truss antenna structure with a pantograph module is proposed. To verify the performance and feasibility of the proposed antenna mechanism, synchronization and deployment performance were evaluated with three different deployment methods (1. free deployment using torsion springs; 2. deployment speed control using a wire; 3. deployment using the pantograph module). Conclusively, experimental results showed that the pantograph module driven deployment method enables stable synchronized deployment.**

Keywords—Deployment Mechanism, Pantograph, Deployable antenna, Ring-truss

I. INTRODUCTION

With the advancement of technologies such as mobile communication, Earth observation, and space exploration, satellite antenna technology capable of rapidly transmitting large volumes of data has become essential [1-3]. To achieve high-performance antennas, large antennas with wide apertures are required [4, 5]. However, due to the spatial constraints and limited diameter of launch vehicles, it is impossible to load such large antennas in their deployed state. To address this challenge, deployment mechanisms have been developed for efficient stowage and deployment of large antennas [6, 7]. A representative antenna structure is the ring-truss, which forms a circular frame by connecting multiple unit bays [8]. The ring-truss structure has the advantages of light weight and a high deployment ratio [8]. Therefore, various studies have reported its performance under combination with deployment mechanisms [9-11].

In this study, a 6R-Pantograph Antenna was designed and fabricated by applying a pantograph mechanism to the conventional ring-truss structure. Then, its feasibility was verified experimentally. The proposed mechanism enables synchronized deployment with a single actuator and maintains a constant height during deployment. To validate its performance, case studies of three deployment methods were conducted: (1) free deployment of the 6R ring-truss using torsion springs, (2) deployment speed control using wires, and (3) deployment using the pantograph module. As the results, deployment time, synchronization performance, and structural stability were analyzed.

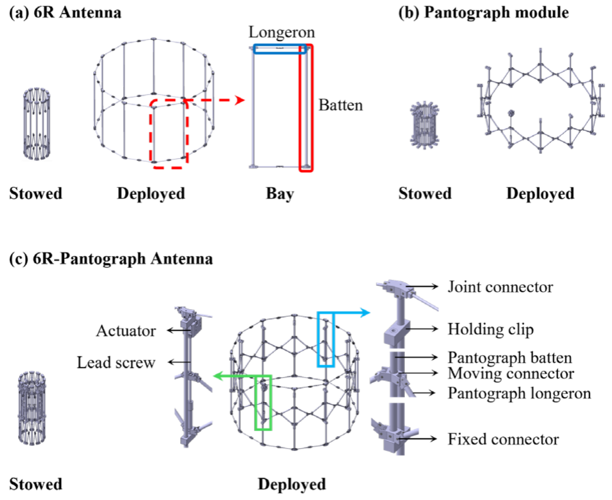


Fig 1. Conceptual design of 6R-Pantograph Antenna, (a) Basic 6R Ring-truss antenna, (b) Pantograph module, (c) 6R-Pantograph Antenna

II. DESING OF 6R-PANTOGRAPH ANTENNA

The proposed antenna adopts the 6R ring-truss structure in Fig. 1 (a) as the base frame and employs the pantograph module in Fig. 1 (b) as the driving mechanism. The design of the 6R-Pantograph Antenna, which ensures synchronized deployment and structural stability, is shown in Fig. 1 (c). In addition, torsion springs were placed at each joint of the 6R ring-truss to reduce actuator loads and improve deployment stability.

III. EXPERIMETNAL SETUP

To verify antenna performance, the effects of gravity acting on the antenna must be minimized [12, 13]. For this purpose, a gravity compensation platform was designed and fabricated as shown in Fig. 2, and experiments were conducted.

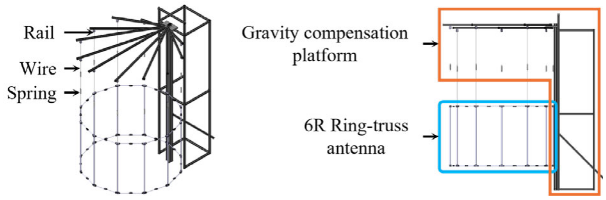
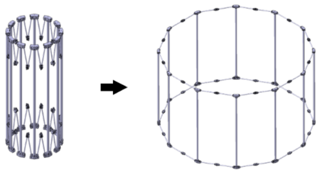
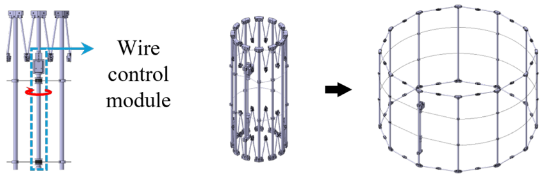


Fig 2. Gravity compensation platform experimental setup

(a) Case 1 : Free deployment using torsion spring



(b) Case 2 : Deployment speed control using wire



(c) Case 3 : Deployment using pantograph module

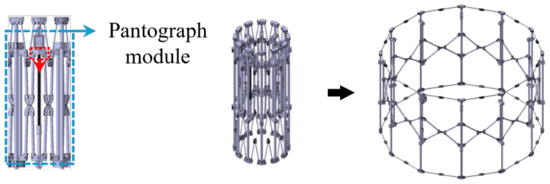


Fig 3. Experimental conditions, (a) Case 1 : free deployment using torsion spring, (b) Case 2 : Deployment speed control using wire tension, (c) Case 3 : Deployment using pantograph module

Three different deployment methods ((1) free deployment of the 6R ring-truss using torsion springs, (2) deployment speed control using wires, and (3) deployment using the proposed pantograph module) are shown in Fig.3. The same motor was used for both wire deployment and pantograph module, operating at the same speed. For each condition, deployment time, synchronization, and final geometric accuracy (including diameter) were measured to compare and analyze deployment stability and performance (i.e. diameter, time, speed).

IV. RESULTS

TABLE 1 COMPARISON OF CASE STUDIES

	Case 1	**Case 2**	**Case 3**
Deployment time [sec]	3	65	65
Target diameter [mm]	1563.16	1560.00	1565.66
Synchronization	X	X	O
Deployment speed	Fastest	Moderate	Slow
Weight [kg]	4.65	5.13	10.95

Antenna deployment experimental results are presented in Table 1. In the case of free deployment, vibrations continued for more than 10 seconds after completion due to spring preload, consequently, it delays final shape stabilization. Wire deployment allowed control of deployment speed and suppression of vibrations; however, achieving synchronized deployment between all bays was challenging due to control limitations. In contrast, the application of the pantograph module enabled synchronized deployment with minimal vibrations during the process. Although the weight increases due to the additional pantograph module, deployment performance (deployment time, target diameter) and synchronization are improved.

ACKNOWLEDGMENT

This work was supported by Korea Research Institute for Defense Technology Planning and Advancement (KRIT) grant funded by Defense Acquisition Program Administration (DAPA) (21-307-A00-013, Korean BFT Communication Satellite Large Antenna Technology).

REFERENCES

[1] X. Ma, T. Li, J. Ma, Z. Wang, C. Shi, S. Zheng, Q. Cui, X. Li, F. Liu, and H. Guo, "Recent advances in space-deployable structures in China," *Engineering,* vol. 17, pp. 207-219, 2022.

[2] A. Meguro, S. Harada, and M. Watanabe, "Key technologies for high-accuracy large mesh antenna reflectors," *Acta Astronautica,* vol. 53, no. 11, pp. 899-908, 2003.

[3] M.-J. Li, M. Li, Y.-F. Liu, X.-Y. Geng, and Y.-Y. Li, "A review on the development of spaceborne membrane antennas," *Space: Science & Technology*, 2022.

[4] M. A. Ullah, R. Keshavarz, M. Abolhasan, J. Lipman, K. P. Esselle, and N. Shariati, "A review on antenna technologies for ambient RF energy harvesting and wireless power transfer: Designs, challenges and applications," *IEEE Access,* vol. 10, pp. 17231-17267, 2022.

[5] Y. Rahmat-Samii, and A. C. Densmore, "Technology trends and challenges of antennas for satellite communication systems," *IEEE Transactions on Antennas and Propagation,* vol. 63, no. 4, pp. 1191-1204, 2014.

[6] B. Wang, J. Zhu, S. Zhong, W. Liang, and C. Guan, "Space deployable mechanics: A review of structures and smart driving," *Materials & Design,* vol. 237, pp. 112557, 2024.

[7] B. Duan, "Large spaceborne deployable antennas (LSDAs)—a comprehensive summary," *Chinese Journal of Electronics,* vol. 29, no. 1, pp. 1-15, 2020.

[8] X. Qi, H. Huang, B. Li, and Z. Deng, "A large ring deployable mechanism for space satellite antenna," *Aerospace Science and Technology,* vol. 58, pp. 498-510, 2016.

[9] M. Sattar, N. U. Rehman, and N. Ahmad, "Design, kinematics and dynamic analysis of a novel double–scissors link deployable mechanism for space antenna truss," *Results in Engineering,* vol. 22, pp. 102251, 2024.

[10] Z. Sun, Y. Zhang, and D. Yang, "Structural design, analysis, and experimental verification of an H-style deployable mechanism for large space-borne mesh antennas," *Acta Astronautica,* vol. 178, pp. 481-498, 2021.

[11] Z. Sun, N. Hu, G. Fan, D. Yang, Z. Yang, T. Liu, and Y. Zhang, "Structural design and model fabrication of a novel double-ring deployable mesh antenna," *Aerospace Science and Technology*, pp. 110351, 2025.

[12] Z. Zhao, K. Fu, M. Li, J. Li, and Y. Xiao, "Gravity compensation system of mesh antennas for in-orbit prediction of deployment dynamics," *Acta Astronautica,* vol. 167, pp. 1-13, 2020.

[13] M. Wu, T. Zhang, P. Xiang, and F. Guan, "Single-layer deployable truss structure driven by elastic components," *Journal of Aerospace Engineering,* vol. 32, no. 2, pp. 04018144, 2019.

Session L2: Late Breaking Results 2

Dynamic Modeling of a Novel R-BWB Aircraft for Altitude Control Studies

Sung Min Shim
Department of Aerospace Mechanical Engineering
Korea Aerospace University
Goyang, Korea
simsonmin6984@kau.kr

Jun Seop Shin
Department of Aerospace Mechanical Engineering
Korea Aerospace University
Goyang, Korea
john2787@kau.kr

Tae Jin Park
Department of Aerospace Mechanical Engineering
Korea Aerospace University
Goyang, Korea
qkrxowls1@kau.kr

Dae-Sung Jang
Department of Aeronautical and Astronautical Engineering
Korea Aerospace University
Goyang, Korea
dsjang@kau.ac.kr

***Abstract*— This paper proposes a dynamic modeling method for a novel Rotary-Blended Wing Body (R-BWB) aircraft that generates lift via low-speed full-body rotation. We establish a procedure for integrating aerodynamic data from Blade Element Momentum Theory (BEMT) into a 6-Degrees of Freedom (6-DOF) framework. The validity of the proposed model is verified through a PID altitude control simulation. This research provides a foundation for developing controllers for a new class of BWB aircraft with the potential for reduced noise and improved efficiency.**

***Keywords*— BEMT, BWB, PID altitude control simulation**

I. INTRODUCTION

With the advent of Urban Air Mobility (UAM), conventional Vertical Take-Off and Landing (VTOL) aircraft such as quadcopters and helicopters are gaining prominence. However, these designs face inherent limitations in noise and efficiency, prompting research into novel vehicle concepts. This paper introduces one such alternative: the Rigid-Blended Wing Body (R-BWB). The R-BWB generates lift by rotating its entire airframe at low speeds, driven by torque from small wingtip propellers. This unique mechanism presents a promising solution for achieving quieter and more efficient flight.

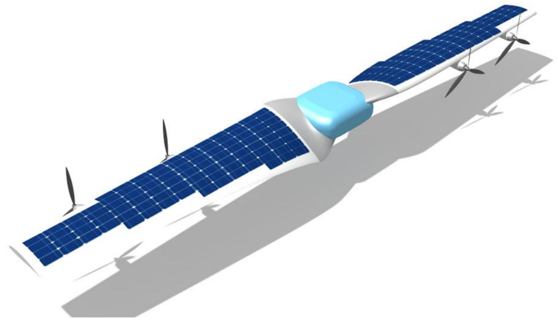


Fig. 1. concept design of R-BWB.

Before a controller for such a novel aircraft can be designed, a dynamic model that accurately predicts its motion must first be established. This paper, therefore, presents a methodology for building a simulation model of the proposed R-BWB by integrating aerodynamic data from Blade Element Momentum Theory (BEMT) into a 6-Degrees of Freedom (6-DOF) framework. The model's validity is confirmed through a PID altitude control simulation, making this study an essential first step toward designing flight controllers for this unique BWB vehicle.

II. SYSTEM MODELING

A. *Generation and Extension of 2D Airfoil Aerodynamic Data*

The initial 2D aerodynamic coefficients (C_l, C_d) for the FX63-137 airfoil used in this study were extracted using XFLR5. However, as the proposed R-BWB operates through full-body rotation, its wing sections experience the entire ±180° angle of attack range. It was therefore essential to extend this data into the post-stall region, which is not provided by XFLR5.

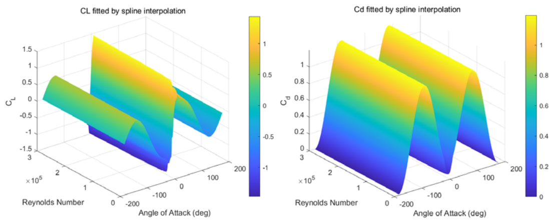


Fig. 2. Extended aerodynamic coefficients as a function of Angle of Attack and Reynolds Number.

Accordingly, we applied the Viterna-Corrigan semi-empirical model to extend the angle of attack range and then interpolated the data for Reynolds numbers (15,000 to 150,000) to match the operational speed. This process yielded the final 2D aerodynamic database, which is visualized as 3D surfaces in Fig. 1 and Fig. 2.

B. *Wing Parameter modeling and 3D BEMT Database Construction*

The primary design objective for the proposed R-BWB is to generate a target lift of 1.75 kgf at a low airframe rotational speed of approximately 80-90 RPM. To this end, a parametric study using 2D Blade Element Momentum Theory (BEMT) was conducted prior to detailed CAD modeling to determine the optimal wing geometry for satisfying this requirement.

$$\beta(r) = \frac{a}{r+b} \tag{1}$$

- β : Twist angle of Wing
- a : coefficient of twist
- b : coefficient of twist
- r : semi-span length

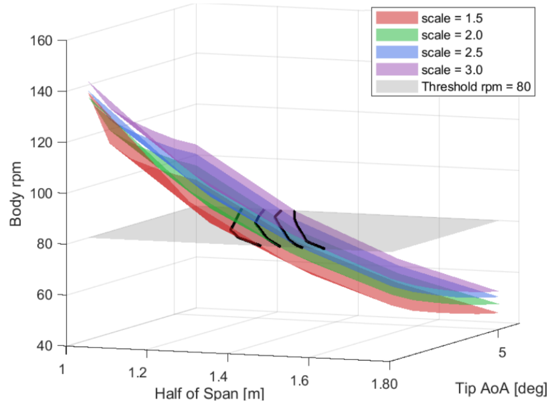


Fig. 3. concept design of R-BWB.

The key design variables for this study were the semi-span, tip twist angle, and a chord scale factor defining the tip chord length relative to the root. The wing's twist distribution, β, was set to follow the rational function shown in (1),while the root twist angle and root chord length were held constant at 25 deg and 0.35 m, respectively.

Figure 3 illustrates the required body RPM to achieve the target lift across various parameter combinations. To ensure a practical wingspan for manufacturing and handling, the chord scale factor was first fixed at 1.5. Subsequently, as depicted in Fig. 4, the final semi-span and tip twist angle were selected from this constrained design space to meet the 80 RPM target threshold. The finalized parameters are summarized in Table I.

TABLE I.

parameter	Determined Wing Parameter	
	Values	***Unit***
Semi span	1.25	[m]
Root chord length	0.35	[m]
Chord scale factor [a]	1.5	-
Root AoA	25	[Deg]
Tip AoA	4	[Deg]

[a] Scale Factor = (Root Chord Length) / (Tip Chord Length)

Based on the finalized wing geometry, a three-dimensional Blade Element Momentum Theory (3D BEMT) simulation was performed to construct a comprehensive aerodynamic database for the airframe. This method offers higher fidelity than its 2D counterpart by calculating forces and moments that account for three-dimensional effects, such as axial flow.

The simulation was conducted using the coordinate system defined in Fig. 3. The input variables included the airframe's angular velocity ($\boldsymbol{\omega}$), aileron deflection angles (δ_1, δ_2), incoming wind magnitude (v_∞) and inflow angle (λ), and the blade's current rotational position, given by the azimuth angle (ψ). Due to the aircraft's low rotational speed, aerodynamic variations at each azimuth angle significantly impact its dynamics, necessitating the calculation of instantaneous rather than were rotationally-averaged forces. Accordingly, at any given azimuth angle (ψ), the differential forces and moments on each blade element were computed and then integrated along the wingspan to find the total forces (F_x, F_y, F_z) and moments (M_x, M_y, M_z) acting on the airframe.

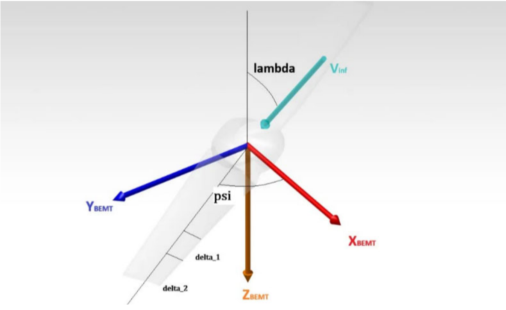


Fig. 4. Coordinate system for 3D BEMT analysis, defining inflow angle (lambda) and blade azimuth angle (psi).

This set of instantaneous forces and moments was then stored as the output for the six corresponding input variables ($\omega, \delta_1, \delta_2, v_\infty, \lambda, \psi$), creating a final 6-dimensional aerodynamic database. Within the dynamics simulator, the aerodynamic loads are retrieved in real-time by applying multi-dimensional linear interpolation to this database based on the aircraft's current stat

C. 6-DOF Equations of Motion

The dynamic behavior of the Rotary-Blended Wing Body (R-BWB) aircraft is described with respect to an Earth-fixed North-East-Down (NED) frame and a body-fixed frame, as depicted in Fig. 5. The aircraft's motion is expressed by the standard 6-Degrees of Freedom (6-DOF) rigid-body equations, which are consistent with those used for a conventional aircraft.

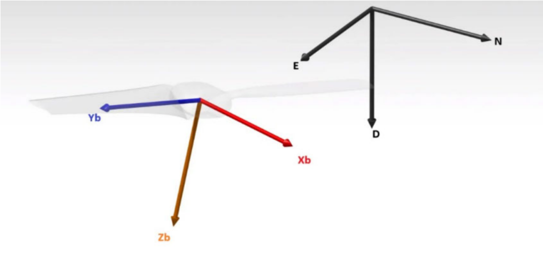


Fig. 5. Definition of the Body-fixed and Earth-fixed coordinate frames.

The translational motion of the aircraft is expressed as shown in (3). The rate of change of position ($\dot{\boldsymbol{x}}_{ned}$) is the velocity ($\boldsymbol{v}_{ned}$), and the rate of change of velocity ($\dot{\boldsymbol{v}}_{ned}$) is calculated by summing the gravitational force with the total aerodynamic force ($\boldsymbol{F}_{aero,b}$). This aerodynamic force is transformed from the body frame to the NED frame using the rotation matrix ($\boldsymbol{R}_{b\rightarrow n}$).

$$\dot{\boldsymbol{x}}_{ned} = \boldsymbol{v}_{ned} \tag{2}$$

$$\dot{\boldsymbol{V}}_{ned} = \frac{1}{m}\left(\boldsymbol{R}_{b\to n}\boldsymbol{F}_{aero,b} + \begin{bmatrix} 0 \\ 0 \\ mg \end{bmatrix}\right) \tag{3}$$

- $\boldsymbol{x}_{ned}$: Position vector in the NED frame
- $\boldsymbol{v}_{ned}$: Velocity vector in the NED frame
- m: Total mass of the aircraft
- g: Gravitational acceleration
- $\boldsymbol{F}_{aero,b}$: Aerodynamic force vector in the body frame

The rotational motion with respect to the body frame is described by the Euler equations, as shown in (4). In this equation, the angular acceleration in the body frame ($\dot{\boldsymbol{\omega}}_b$) is determined by the sum of all external moments ($\Sigma\boldsymbol{M}$) and the gyroscopic effect.

$$\dot{\boldsymbol{\omega}}_b = \boldsymbol{I}_b^{-1}(\sum\boldsymbol{M} - \boldsymbol{\omega}_b \times \boldsymbol{I}_b\boldsymbol{\omega}_b) \tag{4}$$

$$\sum\boldsymbol{M} = \boldsymbol{M}_{aero,b} + \boldsymbol{M}_{\boldsymbol{prop}} \tag{5}$$

- $\boldsymbol{\omega}_b$: Angular velocity vector in the body frame
- $\boldsymbol{I}_b$: Inertia tensor in the body frame
- $\sum\boldsymbol{M}$: Total moment vector acting on the body frame

D. Construction of Dynamics Simulation

While the state variables are integrated using the 6-DOF (Degrees of Freedom) equations of motion are defined in the body-fixed frame. However, the aerodynamic forces and moments obtained from the BEMT database are defined in the BEMT coordinate system. This creates a mismatch between the two frames that must be resolved.

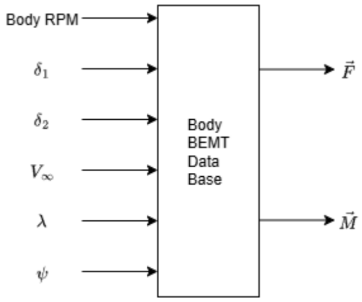


Fig. 6. Input and output structure of the Body BEMT database.

Next, the thrust and P-factor generated by the propeller must be calculated. For this, a database was constructed based on 3D BEMT simulation results. The three input variables for this database are the propeller's rotational speed (Prop RPM), the effective airspeed experienced by the propeller ($V_{\infty,BEMT\ prop,x}$), and the inflow angle (λ_{prop}).

Here, the effective airspeed ($V_{\infty,BEMT\ prop,x}$) is defined as the norm of the vector sum of three velocity components: (1) the aircraft's translational velocity, (2) the linear velocity at the propeller's location due to the aircraft's angular velocity, and (3) the induced velocity from the propeller's own rotation). The inflow angle is also calculated from this resultant velocity vector.

Figure 7. illustrates the process of obtaining the thrust and P-factor from the configured propeller BEMT database for a given flight condition.

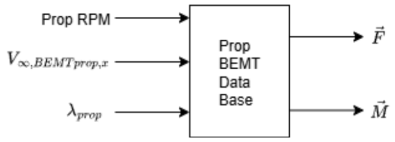


Fig. 7. Input and output structure of the Prop BEMT database.

In this study, a dynamics simulator for the entire flight system was constructed using MATLAB, based on the 6-DOF (Degrees of Freedom) equations of motion and the aerodynamic model from (3) and (4). The simulator was designed with a modular architecture, where each physical model is implemented as an independent module, allowing them to interact organically.

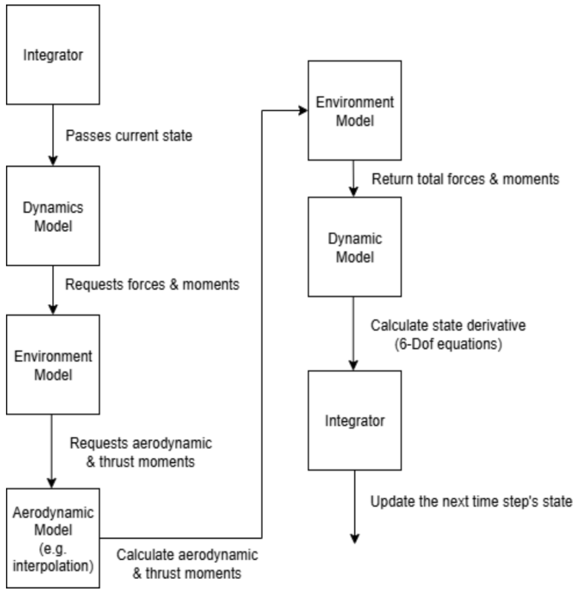


Fig. 8. Overall architecture and computational flow of the simulator.

Figure 8 presents a block diagram illustrating the main computational flow within a single time step of the simulation. The entire process is driven by a top level numerical integrator, which iteratively calculates the aircraft's next state from its current state.

III. MODEL VALIDATION AND SIMULATION

A. Aerodynamic Model Validation

To ensure the reliability of the core aerodynamic model, we experimentally validated the underlying Blade Element Momentum Theory (BEMT) model was performed. However, practical limitations in replicating axial flow conditions with our experimental setup led us to adopt a foundational validation approach.

We therefore focused on validating the static thrust performance of the 2D BEMT model, as it is the fundamental building block for the comprehensive 3D BEMT database. Verifying the model's accuracy in these controllable static conditions establishes a strong baseline of confidence in the overall aerodynamic model used in the simulation.

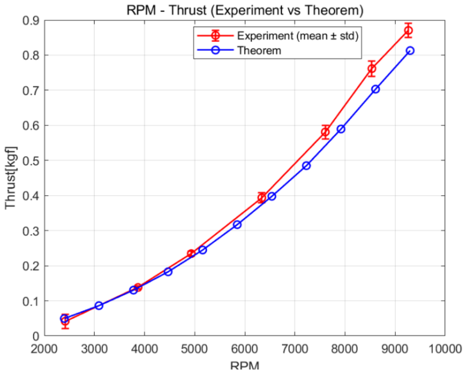


Fig. 9. Comparison of static thrust vs. RPM between BEMT simulation (Theorem) and experimental measurements.

B. *Dynamic Simulation for Altitude Control*

To verify the integrated validity of the complete dynamics simulator, a cascaded PID altitude controller with the structure shown in Fig. 10 was designed and implemented. The purpose of this controller is not to achieve high-performance flight, but rather to confirm the physical plausibility of the simulator's dynamic response by operating it in a stable closed-loop system. The controller is designed to sequentially compensate for errors in altitude, vertical velocity, and body rotation rate.

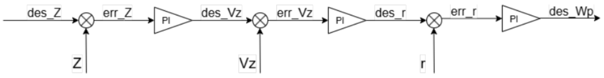


Fig. 10. Block diagram of the cascaded PID altitude controller.

To systematically tune the controller's PI gains, a linearized transfer function model around the hovering state was extracted from the nonlinear 6-dimensional aerodynamic database (from Chapter 2) using system identification techniques. Classical control theory in the Laplace domain was then applied to this linear model to determine the PI gains that ensure stable performance.

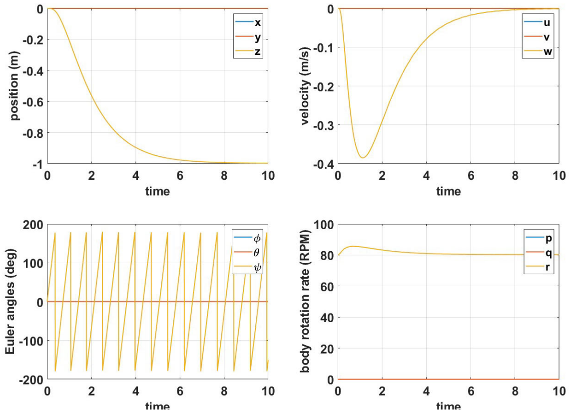


Fig. 11. Time-domain response of the aircraft state variables during the altitude control simulation.

Figure 11 presents the simulation results for a commanded descent from an initial altitude of 0 m to a target of -1 m. As seen in the position and velocity plots, the aircraft stably converges to the target altitude without overshoot. The body rotation rate, r, also stabilizes at the target value of 85 RPM. The corresponding actuator commands are shown in Fig. 12, where both propeller RPM converge uniformly and the ailerons remain neutral, consistent with a pure altitude maneuver. These stable closed-loop simulation results demonstrate that the dynamic model and aerodynamic database constructed in this study are physically plausible, verifying the simulator as a reliable platform for the future design of more advanced controllers.

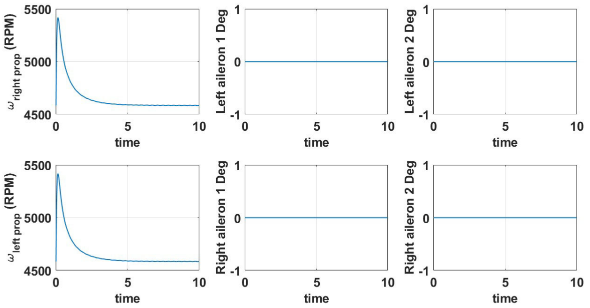


Fig. 12. Actuator command signals during the altitude control simulation.

IV. Conclusion

A. *Figures and Tables*

In this study, a complete 6-DOF dynamics simulator for a novel R-BWB aircraft, which generates lift through full-body rotation, was successfully developed. We established and presented a systematic methodology for modeling such an unconventional aircraft, beginning with the extension of 2D airfoil data, proceeding to a parametric study to define a wing geometry that satisfies the design requirements, and culminating in the construction of a comprehensive 6-dimensional aerodynamic database using 3D BEMT.

The reliability of the proposed model was substantiated through a two-stage validation process. First, the foundational 2D BEMT model was experimentally validated by comparing its static thrust predictions with physical measurements, confirming the accuracy of the core aerodynamic principles. Second, the integrated 6-DOF simulator was verified through a closed-loop PID altitude control simulation. The stable and physically plausible dynamic response observed in this simulation confirmed that the complete model serves as a valid representation of the aircraft's behavior.

This research successfully provides a validated, foundational simulation platform. It is an essential first step that enables future work to focus on the design and implementation of more advanced flight control systems for this unique class of aircraft.

References

[1] S. Park, "Blade element momentum method for propeller under nonaxial flow," *International Journal of Aeronautical and Space Sciences*, vol. 24, pp. 334–340, 2023, doi: 10.1007/s42405-022-00535-7

A Perceptive Fingerless Soft Gripper via Inductive Sensing

Haneul Kim
dept. autonomous vehicle engineering
Korea Aerospace University
Goyang-si, Republic of Korea
khn23@kau.kr

Changbeom Shim
dept. electrical engineering, computing and mathematical sciences
Curtin University
Perth, Australia
changbeom.shim@curtin.edu.au

Seonggun Joe
dept. autonomous vehicle engineering
Korea Aerospace University
Goyang-si, Republic of Korea
sjoe@kau.ac.kr

***Abstract*—Task-oriented grippers have shown reliable object manipulation through predefined strategies, particularly in structured environments. However, when they engage unknown objects (e.g., having irregular shapes, stiffness) in unstructured environments, accomplishing stable and precise prehension often requires complex perception and control. To address these open challenges, this study proposed a pneumatic adaptive fingerless gripper that can achieve reliable object grasping without prior knowledge of object. The morphology of the pneumatic adaptive fingerless gripper was made of zig-zag patterns, which plays a pivotal role in omnidirectional deformation of the gripper. To enhance the overall holding strength and lifting payload of the gripper, parametric optimization was performed. Furthermore, inductive sensing solution was seamlessly integrated with the soft gripper, enabling perceptive sensory feedback. With these in mind, we demonstrated the effectiveness of our design by successfully grasping a diverse range of objects showing potential to real-world scenarios.**

keywords—Soft gripper, Fingerless, Finite Element Methods, Inductive Sensors, Proprioceptive Sensors

I. INTRODUCTION

Unlike conventional rigid-bodied robots, soft robots have shown better safe and compliant interactions with the environment, providing adaptability and robustness simultaneously [1, 2]. Specifically in grasping and manipulation a higher degree of freedom (DoFs) is provided, allowing them to conform to object's shape without active position control [3]. Compared to rigid grippers, soft grippers have presented inherent compliance and ability to safely interact with delicate or irregularly shaped objects [4]. Soft grippers can be broadly categorized into fingered and fingerless grippers based on the presence of protruding structures. The morphological feature of the fingered and fingerless grippers is generally dependent on grasping taxonomy (i.e., pinching, enveloping, adhesion and suction). It is worth noting that determining a proper morphology of the gripper at the design phase is important in piping grasping strategies. While fingered grippers perform precise and dexterous grasping, complex control strategies and coordinated actuation are required [5-7].

In contrast, fingerless grippers enable grasping without prior knowledge of the object which leads to low computational load [8-10]. Fingerless grippers can be further divided into two main categories based on mechanical working principles. Flexible mechanics-based grippers, which utilize geometric or structural transformations. On the other hand, deformable mechanics-based grippers which rely on continuous movements of elastomeric materials.

To induce such motions, different types of actuations (e.g., Borden cable, Piezoelectric, Dielectric, Polymer or Ionic, and Pneumatic, etc.) have been exploited. Among these, pneumatic actuators have shown outstanding versatility, feasibility, and cost efficiency [11]. In particular, it is known to produce high tensile force, while achieving high strength density [12]. However, accurately estimating deformation, contact state or grasping force from actuation parameters remains challenging, due to high compliance of soft grippers.

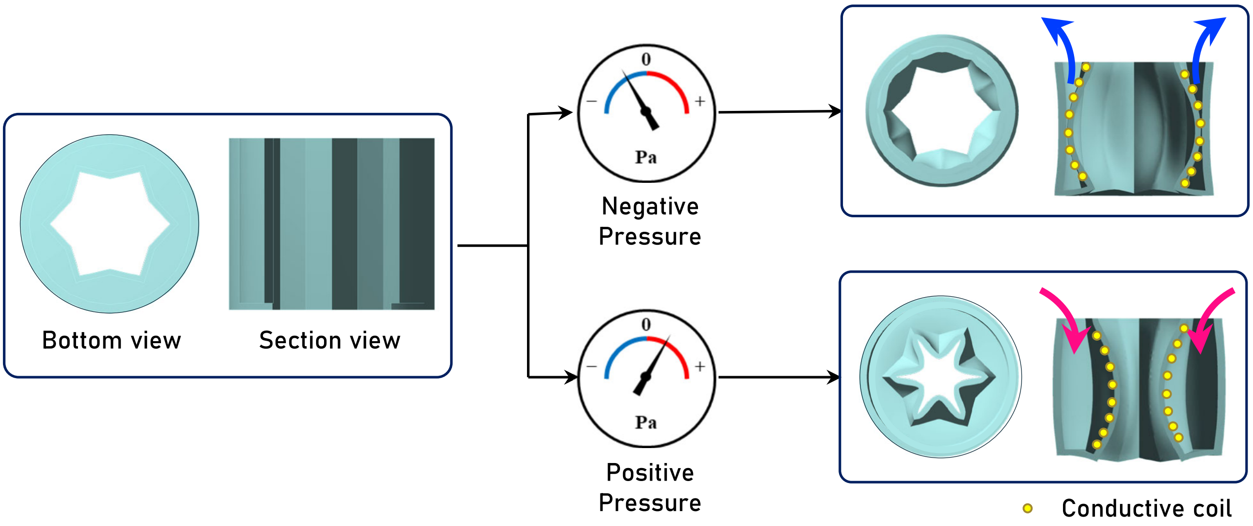


Fig. 1. Proof-of-concept of the sensorized fingerless soft gripper capable of omnidirectional deformations

Meanwhile, integrating sensors is essential for achieving closed-loop control and enhancing manipulation reliability. Herein, among possible sensing solutions, inductive sensing is remarkable for monitoring geometric variations and proximity while is barely influenced by environmental factors (i.e., temperature, or humidity) [13].

With these in mind, we proposed a novel pneumatic fingerless soft gripper for omnidirectional deformation. This study compromises flexibility and deformability, allowing for stable and compliant grasping with respect to a variety of objects having arbitrary shapes and stiffness.

II. DESIGN PRINCIPLE OF SEAMLESS INTEGRATION

A. *Mechanical Design Principle*

As presented in Fig. 1., the proposed soft gripper operates upon positive or vacuum pressure imposed at its internal chamber. Specifically, when positive pressure is applied, the aperture diameter is reduced, wrapping and gripping the object. Once the pressure is released, the aperture diameter passively restores its initial geometry, resulting in release of the object. Moreover, when negative pressure is applied, the aperture diameter increases, which enables the gripper to grasp objects larger than the aperture diameter. The variation in aperture diameter with respect to the applied pressure could be quantitatively identified through the inductive sensor.

Soft bodies made of constitutive hyperelastic materials often undergo mechanical instability such as barely predictable buckling when external pressure or load is applied. Patterns were implemented for controlling deformation to occur only in intended directions. Designed convolutions allow the gripper to move along in planned trajectories and force distributions more precisely. Additionally, the thickness asymmetry of the inner and outer wall prevent inwardly concentrated deformation and suppresses excessive expansion to the outside. The gripping force is effectively increased as the stress becomes concentrated especially along the gripping direction.

B. *Inductive Sensor Design Principle*

Conductive materials, wound along the outer surface of the patterned inner wall, operate as an inductive sensor. The inductance of the helical coil made of the conductive wire is determined by the diameter of the helix. When the gripper is pressurized, the diameter of the coil decreases, which gives higher self-inductance of the coil. Based on Wheeler's formula [14] the correlation between the inductance and the number of turns with a height can be written as:

$$L_0 = \frac{N^2D^2}{450D + 1000h_0} \tag{1}$$

where *N* and *D* indicate the number and diameter of coils. *h* is the height of helix. The given formula is valid when the height is 0.4 times greater than diameter ($h > 0.4 \cdot D$).

$$L = \frac{N^2D^2}{1000} \cdot \frac{1}{0.45D + h} \tag{2}$$

III. EXPERIMENTS AND RESULTS

The output pressure of the flow regulator (ITV0030, SMC corporation) was set to the desired pressure. The flow regulator can control the pressure, ranging from 0 to 500 kPa,

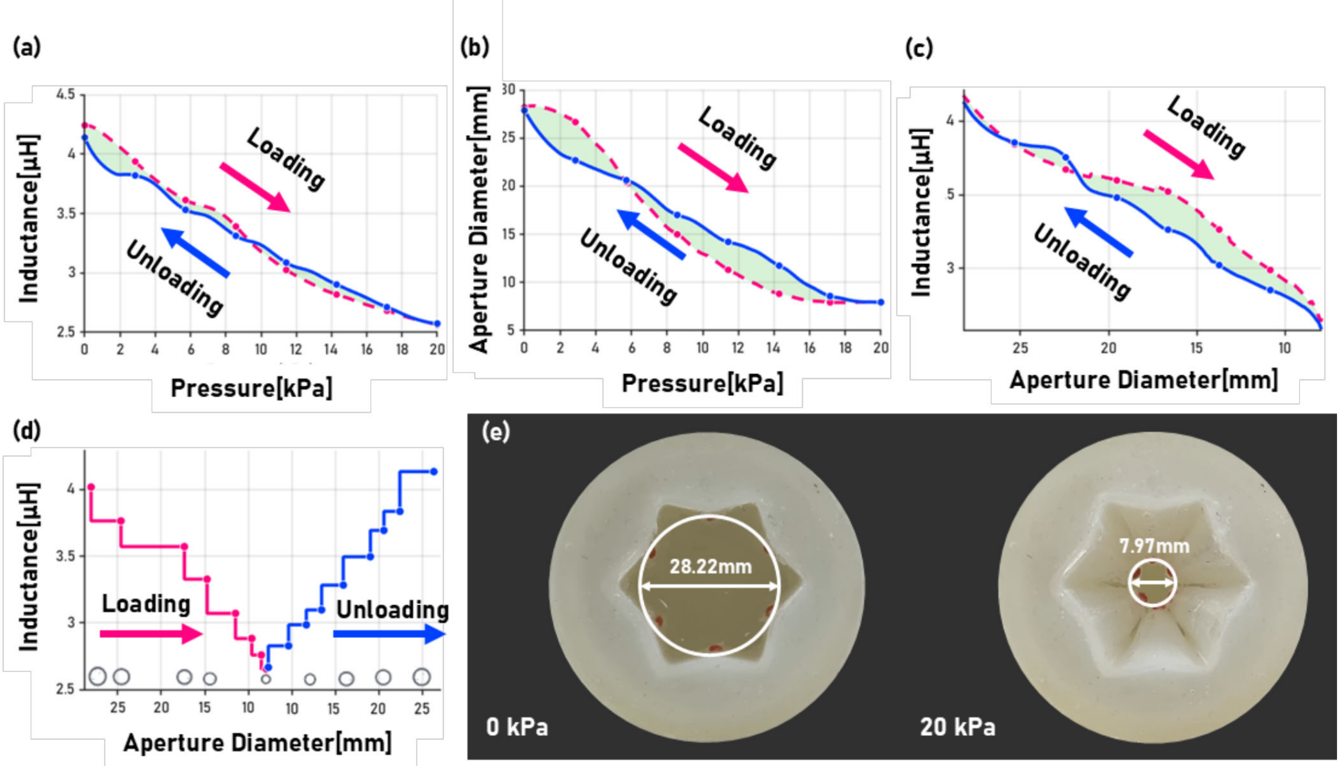


Fig. 2. Experimental results of sensorized fingerless soft gripper. (a) relation of inductance versus pressure, (b) relation of aperture diameter versus pressure, (c) relation of inductance versus aperture diameter. (d) stepwise relation of inductance to aperture diameter over time. Green colored regimes represent a hysteresis, and a positive pressure is imposed with steps of 2 kPa. (e) fingerless soft gripper performing omnidirectional movements (28.22 mm at 0 kPa and 7.97 mm at 20 kPa).

which has a built-in pressure feedback system. The pressure control of the flow regulator was realized by means of the analog input signal. The input signal is proportional to the output pressure in case flow fluctuation is absent. The pressure pump (AIRMAX A2-1800A-1030) generates a maximum pressure of 8 bar. Acquisition and generation of an analog signal for the pressure control and of the digital signal for the foam sensor were implemented by DAQ6218 and myRio 1900 (National Instruments, TX, United States), respectively.

Inductive response of the gripper exhibited a clear, consistent relationship with applied pressure and resulting aperture deformation. According to equation (2), diameter of the coils (*D*) is the only variable as the height of the inductance coil and the number of helixes is fixed. Corresponding to the equation (2), a linear decrease of inductance due to geometric contraction (upon the pressure imposed) is presented, as plotted graph in Fig. 2(a). The finding implies that the gripper's deformation relies more on flexibility than on deformability. Fig. 2(b) indicates that the aperture diameter presented similar tendency with inductance, confirming that variation of aperture diameter directly translates into measurable inductive variation. Hysteresis loops consistently appeared in both the pressure-inductance and the pressure-aperture plots. These loops are likely due to viscoelasticity of the material and friction during loading and unloading, which results in energy dissipation. As depicted in Fig. 2(c)., the relation of inductance and aperture diameter shows an excellent agreement, indicating that inductive sensing can reliably perceive the gripper deformation state, without requiring direct pressure measurement. In addition, Fig. 2(d) presents each discrete transition corresponding to time sequential deformation illustrating how inductance response to the aperture diameter contraction and recovery. As presented in Fig. 2(e)., the fingerless gripper with zigzag pattern exhibited omnidirectional movements, aperture diameter ranging from 7.97 mm (at 20 kPa) to 28.22 mm (at null phase).

These findings show potential in establishing the feasibility of an integrated, proprioceptive sensing for closed-loop control of fingerless soft grippers.

IV. DISCUSSION AND CONCLUSION

This study presented a pneumatic adaptive fingerless soft gripper capable of reliable object manipulation through seamless integration of inductive sensing. The zig-zag structure enabled omnidirectional deformation with enhanced stability, while the embedded inductive coils provided proprioceptive sensing. The experimental results confirmed that the inductive sensing could effectively percept the deformation behavior of the gripper.

Moreover, simple correlation sensing, the inductive coils offer several opportunities for extended functionality. By refining the coil geometry or adjusting the pitches of helix, spatial resolution and sensitivity could be enhanced [15]. Multi-coil arrays integrated circumferentially may lead to distributed shape estimation or contact-state detection demonstrating potential to exteroceptive sensing [16]. Moreover, by implementing higher frequencies, eddy-current based proximity sensing can be presented [14]. Another promising approach is to combine the inductive mechanism with magnetorheological elastomers (MREs). By embedding soft magnetic particles within elastomer, the stiffness of the structure can be dynamically adjusted under an external magnetic field [17]. Indeed, the same coil could serve as both an actuator and a sensing element. In other words, the integration of MREs allows the gripper to achieve seamless actuation-sensing coupling without compromising the mechanical compliance or deformation capability of the gripper.

In conclusion, seamlessly integrated sensor in soft gripper has the potential towards both proprioceptive and exteroceptive sensing, enabling autonomous and intelligent adaption to real-world applications.

ACKNOWLEDGMENT

This research was supported by the National Research Foundation (NRF) funded by the Korean government (MSIT) (RS-2024-00466888)

REFERENCE

[1] D. Rus and M. T. Tolley, "Design, fabrication and control of soft robots," Nature, vol. 521, no. 7553, pp. 467-475, 2015.

[2] C. Laschi, B. Mazzolai, and M. Cianchetti, "Soft robotics: Technologies and systems pushing the boundaries of robot abilities," Science robotics, vol. 1, no. 1, p. eaah3690, 2016.

[3] J. Shintake, V. Cacucciolo, D. Floreano, and H. Shea, "Soft robotic grippers," Advanced materials, vol. 30, no. 29, p. 1707035, 2018.

[4] M. E. Giannaccini et al., "A variable compliance, soft gripper," Autonomous Robots, vol. 36, no. 1, pp. 93-107, 2014.

[5] C. B. Teeple, T. N. Koutros, M. A. Graule, and R. J. Wood, "Multi-segment soft robotic fingers enable robust precision grasping," The International Journal of Robotics Research, vol. 39, no. 14, pp. 1647-1667, 2020.

[6] A. Pagoli, F. Chapelle, J. A. Corrales, Y. Mezouar, and Y. Lapusta, "A soft robotic gripper with an active palm and reconfigurable fingers for fully dexterous in-hand manipulation," IEEE Robotics and Automation Letters, vol. 6, no. 4, pp. 7706-7713, 2021.

[7] R. Deimel and O. Brock, "A novel type of compliant and underactuated robotic hand for dexterous grasping," The International Journal of Robotics Research, vol. 35, no. 1-3, pp. 161-185, 2016.

[8] Y. Hao et al., "A multimodal, enveloping soft gripper: Shape conformation, bioinspired adhesion, and expansion-driven suction," IEEE Transactions on Robotics, vol. 37, no. 2, pp. 350-362, 2020.

[9] S. Li et al., "A vacuum-driven origami "magic-ball" soft gripper," in 2019 International Conference on Robotics and Automation (ICRA), 2019: IEEE, pp. 7401-7408.

[10] Y. Yang, K. Vella, and D. P. Holmes, "Grasping with kirigami shells," Science Robotics, vol. 6, no. 54, p. eabd6426, 2021.

[11] X. Zeng and H.-J. Su, "A high performance pneumatically actuated soft gripper based on layer jamming," Journal of Mechanisms and Robotics, vol. 15, no. 1, p. 014501, 2023.

[12] M. S. Xavier et al., "Soft pneumatic actuators: A review of design, fabrication, modeling, sensing, control and applications," 2022.

[13] H. Wang, M. Totaro, and L. Beccai, "Toward perceptive soft robots: Progress and challenges," Advanced science, vol. 5, no. 9, p. 1800541, 2018.

[14] C. Bartoletti, R. Buonanni, L. Fantasia, R. Frulla, W. Gaggioli, and G. Sacerdoti, "The design of a proximity inductive sensor," Measurement Science and Technology, vol. 9, no. 8, p. 1180, 1998.

[15] S. Hamaguchi et al., "Soft inductive tactile sensor using flow-channel enclosing liquid metal," IEEE Robotics and Automation Letters, vol. 5, no. 3, pp. 4028-4034, 2020.

[16] Y. Mei et al., "Simultaneous shape reconstruction and force estimation of soft bending actuators using distributed inductive curvature sensors," IEEE/ASME Transactions on Mechatronics, vol. 29, no. 4, pp. 2849-2857, 2024.

[17] H. Wang, M. Totaro, A. A. Blandin, and L. Beccai, "A wireless inductive sensing technology for soft pneumatic actuators using magnetorheological elastomers," in 2019 2nd IEEE international conference on soft robotics (RoboSoft), 2019: IEEE, pp. 242-248.

A Preliminary Study on Programmable Soft Architectures via Surface Tessellations

Hayeon Shin
dept. autonomous vehicle engineering
Korea Aerospace University
Goyang, South Korea
shinhy9@kau.kr

Nhan Huu Nguyen
dept. material science
Japan Advanced Institute of Science and Technology
Nomi, Japan
nhnhan@jaist.ac.jp

Seonggun Joe
dept. autonomous vehicle engineering
Korea Aerospace University
Goyang, South Korea
sjoe@kau.ac.kr

***Abstract*—Kinematic synthesis enabling to multi-dimensional movements of soft bodies have been performed, in a way that either considering actuation policies (e.g., multiple actuators) or seamlessly merging different regimes that exhibit a single modal movement. This work introduces a design strategy that allows soft bodies to exhibit desired kinematic movements while ensuring reliability upon cyclic loading/unlading conditions. To this aim, parametric modeling for unit cells designed by surface tessellation is performed, identifying design parameters (including width, height, ligament, and thickness) of the unit cell. The entire soft body mechanics is analyzed by employing constitutive hyperelastic modeling. As a result, it is identified that deformability is strongly influenced by the width-to-heigh ratio, yet the relevant stiffness of the entire soft body is dependent on both thickness of ligament and of unit cell. In light of these findings, programmable deformation of soft bodies is achievable by distributing different unit cells (having different unit cell density), which reveals that inherent stiffness of conventional soft bodies can be enhanced.**

Keywords—Programmable deformation, Topological tessellation, Rhombus unit cell, Hyperelastic modeling, Finite element analysis, Stiffness-flexibility transition

I. INTRODUCTION

Over the last few years, soft robots have shown strong potential in enabling compliant interaction with environment [1]. Remarkably, their potential in embodied intelligence opened new possibilities to mimic the nature of living organisms, as well as to breakthrough bottlenecks that conventional robotics have been suffered [2]. Some of pioneered research have indeed successfully accomplished not only in demonstrating bioinspired robots, but also in addressing industrial scenarios [3]. Note that these achievements are likely due to inherent constitutive hyperelasticity of the material used.

In general, as the hyperelasticity increases (higher orders), the entire soft body undergoes large deformation, which is mainly attributed to high elongation at break, deformability, and relevant stiffness at the material level [4]. Herein, ensuring both scalability and mechanical continuity of the soft body remains at open challenge. As regards the former, the robot bodies made of hyperelastic materials are strongly influenced by square-cube law, meaning that the force-to-weight ratio is aways differentiated by its dimension (1/D) [5]. For this reason, the soft body is vulnerable to mechanical instability (so-called column squirm or buckling), resulting in understanding of kinematic trajectories barely possible [6]. As regards the latter, many researchers have employed structural enhancements to reinforce their poor mechanical performance (i.e., stiffness in kPa range), yet this approach often results in discontinuous soft bodies (similar to articulated robots). Therefore, it is worth mentioning that such discontinuity in soft bodies does not truly meet a compliance that constitutive hyperelastic materials can offer.

To address these challenges, this work presents a novel design strategy that harnesses a surface tessellation to create programmable soft architectures, enabling soft bodies to exhibit desired kinematic movements while ensuring structural reliability under cyclic loading/unloading conditions and preserving mechanical continuity. Exploiting tessellated 2D surfaces, which leverage topological transformations *via* cutting patterns, we propose a rhombus-shaped unit cell, as fundamental building blocks for programmable soft architectures. These rhombus unit cells inherently offer a direct and explicit means for achieving angular deformation, making them particularly suitable for tailoring desired movements through kinematic synthesis.

Therefore, the primary objective of this paper is to identify the optimal parameters of the rhombus unit cell, ensuring they possess both appropriate stiffness and flexible design capabilities suitable for diverse soft robotic applications. To achieve this goal, we perform parametric modeling and comprehensive analysis of key geometric parameters of the unit cell, utilizing constitutive hyperelastic modeling coupled with finite element methods (FEM). In parallel, this work also establishes a systematic pipeline to investigate the correlation between geometric parameters and mechanical characteristics (i.e., deformability and axial stiffness) of the rhombus unit cell.

II. MATERIALS AND METHODOLOGY

A. Materials

Considering the fabrication and computation load to analyze mechanical performance *via* FEM, Thermoplastic Polyurethane (TPU) was selected as the base material. Indeed, TPU has been widely exploited in soft robotics together with additive manufacturing, due to its inherent hyperelastic properties, large deformation capacity, and ease of creating complex geometries *via* 3D printing [7]. In this work, TPU behavior was analyzed by constitutive hyperelastic modeling, and the material parameters were obtained by using the 2nd order Ogden model [8].

B. Unit Cell Design and Parametrization

Given that the programmable soft architecture is made of hierarchically distributed unit cells, a rhombus shape is beneficial, in sense that the entire soft body does not undergo mechanical instability while offering high design flexibility

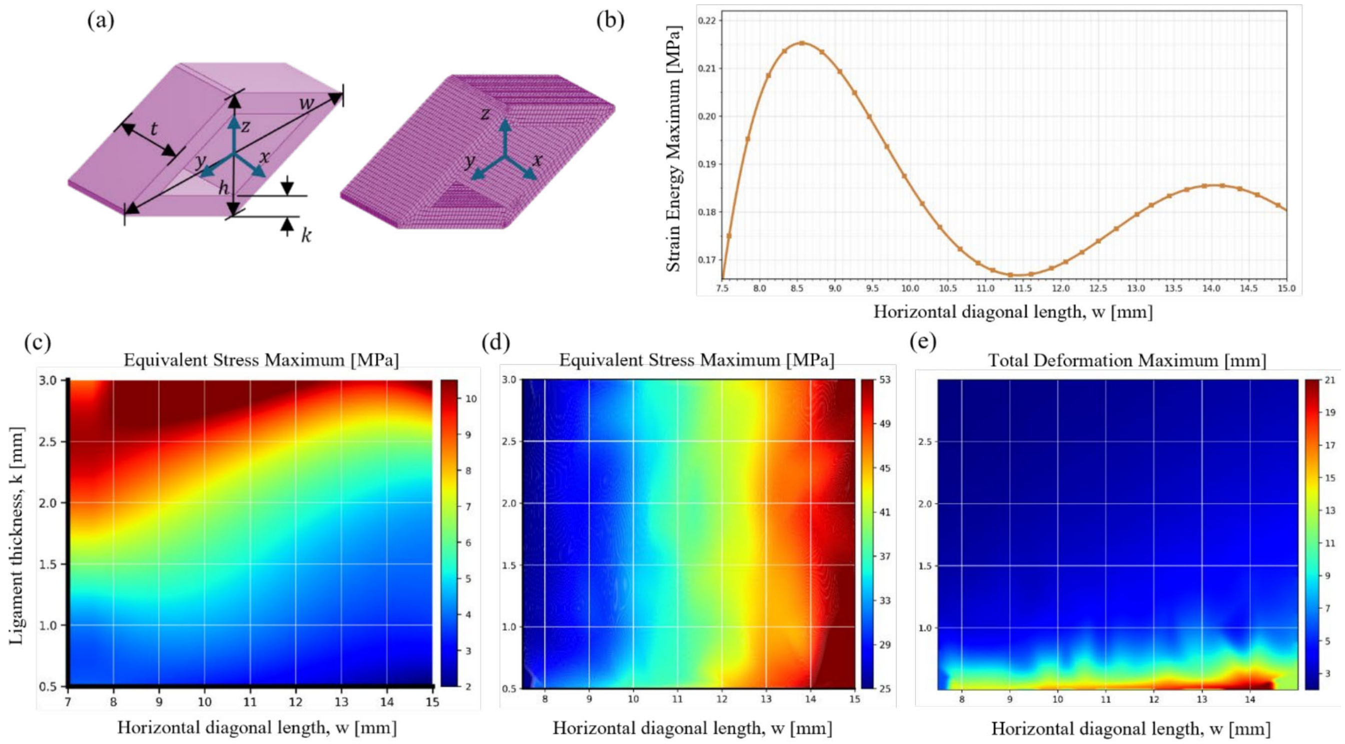


Fig. 1. Finite Element Method (FEM) analysis of the surface tessellated unit cell featuring rhombus morphology. (a) Parametric modeling of the rhombus unit cell and its global coordinate; (b) Relation of strain energy versus horizontal diagonal length (w), representing optimal flexibility; (c) Distribution of equivalent stress with respect to variations in ligament thickness and horizontal diagonal length, upon displacement-controlled loading; (d) Distribution of equivalent stress with respect to variations in ligament thickness and horizontal diagonal length, upon force-controlled loading; (e) Distribution of total deformation with respect to variations in ligament thickness and horizontal diagonal length, upon force-controlled loading.

to distribute along the soft body. As illustrated in Fig. 1(a)., mechanical characteristics of the rhombus unit cell can be identified by four geometric parameters. Specifically, the vertical diagonal height (h) was fixed at 10 mm. The horizontal diagonal length (w) varied from 7.5 to 15 mm. The thickness (t) of the entire unit cell varied from 1 to 5 mm, and the ligament thickness (k) varied from 0.5 to 3 mm to investigate their influence on the mechanical performance.

The rhombus inclination angle is then defined by the angle $\theta = \arctan\, h/w$, implicitly relating to geometry and internal force. Its distinct geometry allows for tailored mechanical responses essential for multi-dimensional movements.

C. Finite Elements Analysis

The FEM simulations were conducted using commercial software ANSYS 16.1 (ANSYS Inc., Canonsburg, PA, USA). A quasi-static simulation was performed with static structural module, upon tensile elongation (up to 120%). The 3D rhombus unit cell models were meshed with local refinement at ligament and corner regions. The unit cell assumed to be made of TPU exhibited constitutive hyperelastic behaviors, which was modeled by the 2nd order Ogden hyperelastic model. The governing equation of the simulation study is as follows:

$$\sigma_j = -p + \lambda_j \frac{\partial W}{\partial \lambda_j} = -p + \sum_{i=1}^{2} \mu_i {\lambda_j}^{\alpha_i - 1} \tag{1}$$

The right and left nodes were fully constrained to emulate boundary fixation, while deformation was induced by either prescribed displacement ($u(t) = 0.1t$) or a constant tensile load ($F = 5N$) applied to the top and bottom faces. In principle, all models can accommodate a large strain up to 300%, which is due to TPU behavior. Through simulation study, mechanical performances of the unit cell (i.e., maximum equivalent stress, total deformation, strain energy, and safety factors) were identified.

III. RESULTS AND DISCUSSION

A. Design strategy for stiffness reinforcement

Under displacement-controlled condition ($u(t) = 0.1t$, $h = 10$ mm, $t = 1$ mm), the stiffness of the rhombus unit cell is strongly influenced by its geometric configuration. As shown in Fig. 1(c), smaller horizontal diagonal length (w) and thicker ligaments (k) result in substantially higher equivalent stresses, meaning that higher reaction forces can be appreciated under the same displacement imposed. Smaller horizontal diagonal lengths lead to stretching-dominated deformation, which increases the local stretch ratio (λ_j) and thereby enhances the reaction force under the same imposed displacement. In contrast, thicker ligament thickness (k) increases the sectional stiffness (along diagonal direction of rhombus) through an enlarged cross-sectional area ($A \propto k \cdot t$), which results in a stiffer overall response [9].

Consequently, configurations combining small horizontal diagonal lengths and thick ligaments yield the highest overall stiffness due to their stretching-dominated response and enlarged cross-sectional area, but they also induce pronounced local stress concentrations near the ligament junctions. To mitigate this, increasing the unit cell thickness (t) significantly improves the minimum safety factor, strengthening fatigue reliability without reducing strain energy.

B. Design strategy for complaint and flexibility

When a constant tensile force is applied, the mechanical response reverses. As shown in Fig. 1(b), the strain energy reaches a peak near intermediate values of w ($\approx 8.5 - 9.2$ mm), representing the optimal flexibility range where the structure maximizes energy absorption under the fixed load. This trend can be interpreted from the general mechanical relations of stress versus deformation. The applied load determines an average tensile stress and a resulting strain along each ligament direction.

As it increases, the inclined ligaments become more horizontal relative to the loading direction, transforming the deformation mode from stretching-dominated to bending- and shear-dominated, as illustrated in Fig. 1(d). This geometric transition allows the structure to deform more easily under the same load, leading to higher strain energy and more distributed deformation. Consequently, larger horizontal diagonals produce more compliant behavior, while smaller horizontal diagonals promote stiffness through direct tensile loading of the ligaments.

In contrast, Fig. 1(e)., shows that ligament thickness (k) strongly influences the magnitude of deformation. Since the global stiffness is proportional to cross-sectional area ($K_{\text{global}} \propto A \propto k \cdot t$), the total deformation decreases inversely with k ($\delta \propto 1/k$). This inverse proportionality aligns with the reduction in total deformation observed in Fig. 1(e). Therefore, horizontal diagonal length and (k) independently modulate two orthogonal aspects of the mechanical response: the former determines how the structure deforms (stress-strain pattern), while the latter controls how much it deforms (global amplitude).

IV. CONCLUSION

In summary, this work presented a preliminary parametric study on geometry-dependent mechanical behaviors of surface tessellated rhombus unit cells. Smaller horizontal diagonals and larger (k) enhanced stiffness under displacement control, while intermediate w ($\approx 8.5 - 9.2$ mm) maximized flexibility and energy absorption under force control. Stress and deformation were primarily governed by horizontal diagonal length and (k), respectively. However, the analysis revealed an inherent trade-off between stiffness and flexibility – while increasing ligament thickness (k) improves load bearing capacity and fatigue reliability, it simultaneously suppresses global deformability. Likewise, decreasing its length increases stiffness but leads to higher local stress concentrations.

Therefore, rather than pursuing a single optimum, programmable soft architectures emerge from deliberate coordination of parameters within balanced ranges that achieve task dependent stiffness-flexibility transitions. Such a parametric interaction enables local reinforcement without sacrificing overall compliance, offering a rational basis for geometry-guided design of soft actuators and prosthetic systems capable of tunable, multi-dimensional deformation while maintaining structural integrity.

ACKNOWLEDGMENT

This research was supported by the National Research Foundation (NRF) funded by the Korean government (MIST) (RS-2024-00440436)

REFERENCES

[1] D. Rus and M. T. Tolley, "Design, fabrication and control of soft robots," *Nature,* vol. 521, no. 7553, pp. 467-475, 2015.

[2] J. Li *et al.*, "AI-embodied multi-modal flexible electronic robots with programmable sensing, actuating and self-learning," *Nature Communications,* vol. 16, no. 1, p. 8818, 2025.

[3] R. F. Shepherd *et al.*, "Multigait soft robot," *Proceedings of the national academy of sciences,* vol. 108, no. 51, pp. 20400-20403, 2011.

[4] S. Gómez-Garraza, R. de Santos, D. Infante-García, and M. Marco, "Visco-hyperelastic material model fitting to experimental stress–strain curves using a genetic algorithm and its application to soft tissue simulants," *Scientific Reports,* vol. 14, no. 1, p. 18026, 2024.

[5] L. Xu *et al.*, "Locomotion of an untethered, worm-inspired soft robot driven by a shape-memory alloy skeleton," *Scientific reports,* vol. 12, no. 1, p. 12392, 2022.

[6] A. Pal, V. Restrepo, D. Goswami, and R. V. Martinez, "Exploiting mechanical instabilities in soft robotics: control, sensing, and actuation," *Advanced Materials,* vol. 33, no. 19, p. 2006939, 2021.

[7] S. F. del Toro, J. Crespo-Sanchez, J. Ayllon, A. Rodriguez-Prieto, and A. M. Camacho, "Mechanical performance of 3D-printed TPU auxetic structures for energy absorption applications," *Polymer Testing,* vol. 143, p. 108669, 2025.

[8] C. Ursini and L. Collini, "FDM layering deposition effects on mechanical response of TPU lattice structures," *Materials,* vol. 14, no. 19, p. 5645, 2021.

[9] H. Wyatt, A. Safar, A. Clarke, S. L. Evans, and L. A. Mihai, "Nonlinear scaling effects in the stiffness of soft cellular structures," *Royal Society open science,* vol. 6, no. 1, p. 181361, 2019.

Design and Fabrication of Metamaterial Based Capacitive Sensor via 3D Printing

Seik Park
Artificial Intelligence Department
Interdisciplinary Program in Space Systems Engineering
Korea Aerospace University
Gyeonggi-do, Korea
seik2121@kau.kr

Eunji Kim
Artificial Intelligence Department
Interdisciplinary Program in Space Systems Engineering
Korea Aerospace University
Gyeonggi-do, Korea
yell1218@naver.com

Seonggun Joe
Department of Autonomous Systems Engineering
Korea Aerospace University
Gyeonggi-do, Korea
sjoe@kau.ac.kr

***Abstract*— This paper presents the design, fabrication, and experiment of a miniaturized capacitive tactile sensor fabricated using Stereolithography (SLA) 3D printing. Capacitive sensors rely on the distance of two electrodes, and the deformation of the dielectric layer occurs the capacitance changes, which makes sensors possible to detect changes of displacement, force, or pressure. Conventional fabrication approaches have limitation of flexibility and difficulties in achieving microscale structures, limiting their integration to wearable, minimized applications. To address these challenges, a lattice structure was embedded within the dielectric layer to enable linear compression and mechanical compliance. Finite Element Modeling (FEM) was conducted to analyze electromechanical behaviors, and experimental measurements validated the predictions. The proposed sensor demonstrates strong potential for tactile applications such as Braille detection, wearable human–machine interfaces, and immersive VR/AR systems.**

Keywords*— *Capacitive sensor, Tactile interface, Miniaturized sensor design, 3D printing fabrication, Lattice dielectric, Human-inspired sensing, Assistive technology

I. INTRODUCTION

Capacitive sensors operate by detecting variations in capacitance, which is defined as permittivity of dielectric layer material, the distance between two electrodes, and the area of the electrodes [1]. External mechanical input occurs measurable changes in capacitance that can be correlated with displacement, pressure, or shear. With the increasing demand for compact and flexible sensors in wearable electronics, biomedical instruments, and artificial skin systems, conventional PCB-based sensors present major limitations in terms of flexibility and resolution [2]. Recent advances in additive manufacturing provide opportunities to overcome these challenges by enabling high-resolution, customized structures [3]. In this paper, a lattice-embedded dielectric fabricated through Stereolithography (SLA) printing is proposed, providing compressibility and scalability while maintaining structural stability. Moreover, combining with conductive fabric achieves both mechanical characteristics and suitable for wearable human interfaces.

II. DESIGN AND ANALYSIS

The principle of the proposed sensor is based on a parallel plate capacitor model. The capacitance (*C*) is described, as follows:

$$C = \frac{\varepsilon_0 \varepsilon_r A}{d} \tag{1}$$

ε_0 means the vacuum permittivity, and ε_r represents a relative permittivity of the dielectric material. *A* indicates an area of the electrodes, and *d* refers to the distance between them. In our sensor, the dielectric layer consists of composite of the polymeric material and air gaps within the lattice. Therefore, effective permittivity must consider both, the mechanical characteristics of the polymers, and air gaps. With these thoughts in mind, the overall capacitance of metamaterial-based sensors can be written by:

$$C = \frac{\varepsilon_0 A}{d_{air} + \frac{d_{silicon}}{\varepsilon_r}} \tag{2}$$

wherein, d_{air} indicates the thickness of the air layer, and $d_{silicon}$ refers to the thickness of the silicon. When the sensor is compressed, total equivalent air gap directly reduces, resulting in increase of the total capacitance *C*.

The dielectric layer made of a lattice structure can undergo compressive deformations. Indeed, unlike a solid (bulk) composition of the material (which generally features nearly incompressible), the metamaterials are remarkable, ensuring extraordinary behaviors, e.g., highly deformable and compressible. To analyze and predict the mechanical behavior of the lattice structure upon compression, a Finite Element Model (FEM) was developed using Ansys software. The primary goal of this analysis was to understand how the applied displacement translates to a change in the effective thickness of the dielectric layer.

A 3D model, which placed lattice structure in a 4x4x2 array, was analyzed using ANSYS. The 14 mm-thick model was compressed in 14 steps of 0.625mm The simulation results, shown fig 1, demonstrate that the lattice geometry effectively permits compression, transforming vertical displacement into

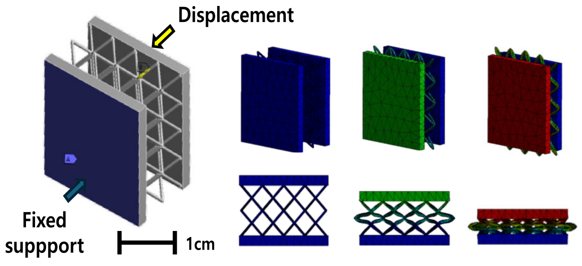


Fig 1. FEM of lattice structures

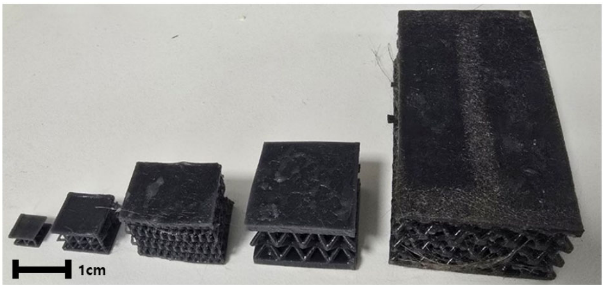


Fig 2. Lattice-embedded structures

a uniform reduction in the structure's height. This analysis confirms that the designed structure behaves as intended, providing the necessary mechanism for capacitance change.

III. FABRICATION

Early prototypes were made with SLA, and conductive fabric. As shown Fig. 2, Lattice-embedded structures are placed in array between conductive layers. The SLA printing using photocurable resin provided high-resolution fabrication of the dielectric, while conductive fabric layers were applied as electrodes [4]. To improve durability, the electrode edges were reinforced with silicone coating cured under UV exposure [5]. The silicone coating preserves the inherent softness of the conductive fabric while preventing fraying of the fibers [6]. In addition, it serves as a primary protective layer that reduces capacitance measurement errors by shielding the conductive fabric from unintended contact with external conductive objects.

IV. RESULTS AND DISCUSSION

To characterize the performance of the fabricated capacitive sensor, a quasi-static compression experiment was conducted. The experimental setup consisted of a mechanical clamp system for controlled displacement applications and a capacitance measurement module with Arduino, which connected to computer.

The experiment was designed to simulate simple displacement-driven compression, which represents the primary mode of tactile input in applications such as pressure detection sensing. The mechanical clamp system compresses slowly the dielectric block, with displacement values ranging from 13.75 mm to 5.625 mm, in 14 steps. Each displacement step, the capacitance was recorded on computer.

Fig. 3 shows comparison between the numerical analysis capacitance (Orange line), the measured experimental capacitance (Blue line), and the absolute difference. Both simulation and experiment demonstrate an increase in capacitance as the distance between the electrodes decreases This correlation in their trends validates the fundamental design principle of the sensor. Furthermore, the difference between experiment and simulation shows non-linear behaviors. It gradually increases as the sensor is compressed, peaking around 6.875 mm before slightly decreasing at the point of maximum compression. Simulated capacitance did not include parasitic capacitance, such as Electromagnetic Interference and the internal lattice structures begin to make contact, causing the dielectric material to be extruded through the open sides of the sensor.

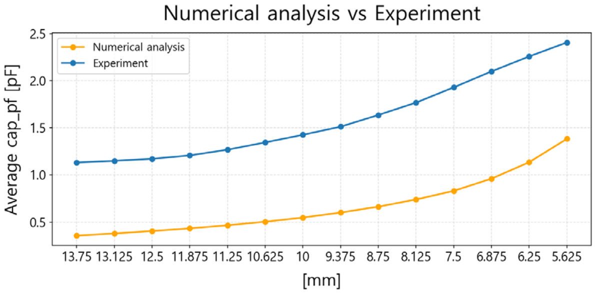


Fig. 3 Comparison of experiment and numerical analysis

To address these sources of error in future work, elaborating on analytical models is necessary, e.g., accounting material extrusion under high compression. Furthermore, the experimental setup needs to be optimized by minimizing the length of connecting wires, in order to reduce parasitic capacitance and susceptibility to external noise.

V. CONCLUSION

This study introduced a lattice-embedded capacitive sensor fabricated through stereolithography printing with a lattice embedded dielectric and conductive fabric electrodes. The fabrication process demonstrated high repeatability, producing sensors with consistent geometry and capacitance response. Quasi-static experimental test validated the theoretical predictions, showing capacitance variations calculated values.

While the sensors exhibited nonlinearities at large displacements, these can be addressed through refined lattice geometry design and advanced signal analysis methods. Moreover, this study demonstrates that 3D-printed capacitive sensors with lattice dielectrics and fabric electrodes provide a tactile interface for wearable technology, assistive devices, and immersive VR/AR applications. To achieve this, our future work relies on development of tactile sensors and better signal processing techniques.

VI. ACKNOWLEDGEMENTS

This research was supported by the National Research Foundation (NRF) funded by the Korean government (MSIT) (RS-2024-00440436)

VII. REFERENCES

[1] L. K. Baxter, "Capacitive sensors," Design and Applications, 1997.

[2] A. R. P. Mathai, T. Stalin, and P. V. y Alvarado, "Flexible fiber inductive coils for soft robots and wearable devices," IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 5711-5718, 2022.

[3] T. Someya, T. Sekitani, S. Iba, Y. Kato, H. Kawaguchi, and T. Sakurai, "A large-area, flexible pressure sensor matrix with organic field-effect transistors for artificial skin applications," Proceedings of the National Academy of Sciences, vol. 101, no. 27, pp. 9966-9970, 2004.

[4] R. L. Truby and J. A. Lewis, "Printing soft matter in three dimensions," Nature, vol. 540, no. 7633, pp. 371-378, 2016.

[5] W.-H. Jang et al., "High-temperature corrugated-core radar stealth Ni-plated glass/bismaleimide sandwich composites," Composite Structures, vol. 365, p. 119185, 2025.

[6] W. T. Lee, D. J. Hong, Y. W. Nam, and R. S. Myong, "Numerical and experimental investigation of multifunctional high-efficiency anti-icing nickel-plated carbon fiber heating elements for wing-shaped composite airfoils," Aerospace Science and Technology, p. 110249, 2025.

Multichannel Ultrasound Driving Module-Based Drug Delivery System: Inducing Targeted Release of Doxorubicin for Liver Cancer Therapy

Daewon Jung
US Actuators Lab
Korea Institute of Medical Microrobotics
Gwangju, Korea
jungdaewon@kimiro.re.kr

Hiep Xuan Cao
US Actuators Lab
Korea Institute of Medical Microrobotics
Gwangju, Korea
hiep.caoxuan@kimiro.re.kr

Hoang Viet Nguyen
College of AI Convergence
Chonnam National University
Gwangju, Korea
viethoang20151552@jnu.ac.kr

You Hee Choi
Biomedical Lab
Korea Institute of Medical Microrobotics
Gwangju, Korea
youheechoi@kimiro.re.kr

Tae Il Kang
Biomedical Lab
Korea Institute of Medical Microrobotics
Gwangju, Korea
kti@kimiro.re.kr

Seokmin Hong
Medical Ai Lab
Korea Institute of Medical Microrobotics
Gwangju, Korea
seokmin.hong@kimiro.re.kr

Jong-Oh Park
Korea Institute of Medical Microrobotics
Gwangju, Korea
jop@kimiro.re.kr

Byungjeon Kang
Graduate School of Data Science
College of AI Convergence
Chonnam National University
Gwangju, Korea
bjkang8204@jnu.ac.kr

***Abstract*— Acoustic approaches for non-invasive manipulation of microparticles have emerged as a promising strategy to enhance the efficiency of drug delivery in biomedical applications. Ultrasound-driven systems, in particular, offer distinct advantages by generating well-defined pressure nodes at precise spatial locations, thereby enabling controlled particle manipulation. In this study, we present an ultrasound driving system composed of 86 transducers arranged in a hemispherical configuration. The acoustic field distribution generated by the system was characterized using an acoustic scanning device to ensure accurate performance evaluation. The primary objective of this work is the development of an ultrasound-based acoustic driving platform capable of precise microparticle control in clinically relevant scenarios. To assess the feasibility of the proposed system under biologically relevant conditions, validation experiments were conducted using agarose phantoms embedded with doxorubicin. Experimental results demonstrate that the system can achieve an enhancement of more than 30% in targeted drug release at the focal region, thereby highlighting its potential as a non-invasive and effective tool for particle control and drug delivery in medical applications.**

Keywords— ***Ultrasound Actuator, Targeted Microparticle Delivery, Drug Targeting Efficiency***

I. INTRODUCTION

Non-contact manipulation of microparticles has been extensively investigated in the field of targeted drug delivery. Such systems are typically based on acoustic, electromagnetic, or optical fields, among which ultrasound-based platforms have been reported to exhibit superior targeting efficiency for microparticles [1]. Owing to these advantages, numerous studies employing ultrasound transducers have been conducted, with traveling-wave approaches in particular being considered highly applicable for in vivo environments [2].

In this study, a multichannel ultrasound driving module was developed, consisting of 86 transducers arranged in a hemispherical configuration. The modulation of traveling-wave phases and the resulting acoustic field patterns were characterized using a hydrophone (ONDA Corporation, USA) to scan the acoustic distribution. To further assess the biomedical applicability, experimental validation was performed with doxorubicin, a chemotherapeutic agent, to evaluate the capture efficiency at the targeted location induced by the acoustic patterns generated from the ultrasound driving module.

II. METHODS

A. Principles for Acoustic Radiation force

In an acoustic pressure field, Rayleigh microparticles with radius smaller than the wavelength are influenced by the acoustic radiation force (FRF_RFR). This force can be quantified through the Gor'kov potential field, as expressed in equations (1)–(3) [3]:

$$F_R = -\nabla U \tag{1}$$

$$U = V_P[f_1 \frac{1}{2} k_0 \langle |p_1|^2 \rangle - f_2 \frac{3}{4} \rho_0 \langle |v_1|^2 \rangle \tag{2}$$

$$f_1 = 1 - \frac{k_p}{k_0} \ \ and \ \ f_2 = \frac{2(\rho_p/\rho_0 - 1)}{2\rho_p/\rho_0 + 1} \tag{3}$$

Where 'V_P' and 'v_1' are the density and velocity of the particle, 'p_1' is the pressure value in the beam generated point. Monopole 'f_1' and Dipole 'f_2' are decided by compressibility of particle and host medium ('k_p', 'k_0') with density ('ρ_p', 'ρ_0').

B. Multichannel Ultrasound Driving Module

To maximize the acoustic pressure concentrated at the focal position, 86 ultrasound transducers (1 MHz resonance frequency) were arranged in a hemispherical configuration, as illustrated in Fig. 1 (L). The distance from the transducer surface to the focal point was determined by the radius of curvature of the hemispherical model. Each transducer was connected to a pulser–receiver controller capable of signal amplification and phase modulation, allowing adjustment of the focal position and alteration of the wave patterns. The resulting acoustic field distribution, including the pattern shape and generated pressure, was characterized using a three-dimensional acoustic scanning system with a hydrophone, as shown in Fig. 1 (R).

In this experiment, among the various acoustic patterns generated by the ultrasound driving module, a vortex configuration was selected, as it is more suitable for drug trapping over a wider region and for inducing acoustic streaming within the channel.

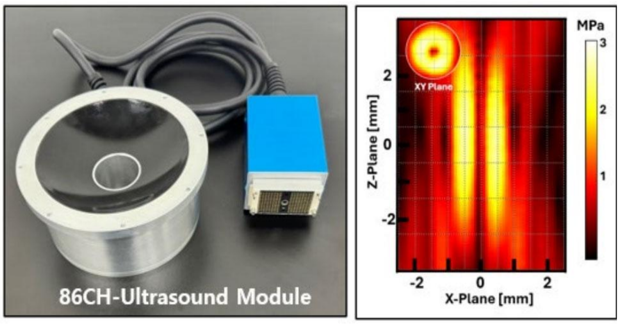


Fig.1 86CH-Ultrasound Module Prototype (L), Results of Vortex Scanning of the Ultrasonic Field (R)

C. Fabrication of a Single-Channel Agarose Phantom

To evaluate the potential medical applications of the ultrasound system, agarose phantom channels were fabricated. The concentration of agarose was set to 1% (g/100 mL) to mimic the acoustic properties of human soft tissues and accommodate the experimental conditions, and the channel diameter was designed as 2 mm, corresponding to the portal vein of the rat used as the in vivo target [4].

D. Experiment setup

Doxorubicin solution was loaded into the channels of a tissue-mimicking phantom chamber, and a micro-camera was mounted on the side of the channel to enable real-time monitoring of drug release at the target location of the ultrasound-driven module. Based on the recorded videos, the extent of drug release in samples treated with the ultrasound module was quantitatively compared to that in untreated samples using ImageJ software.

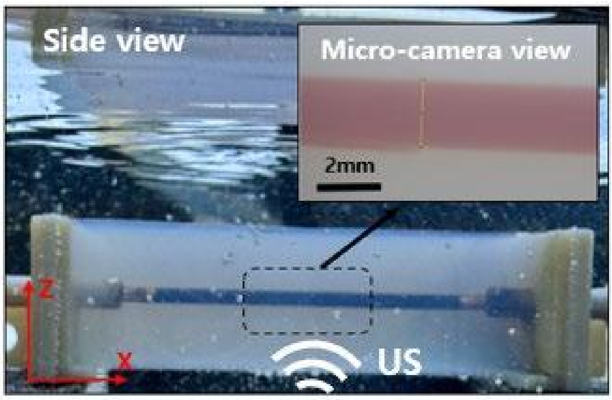


Fig.2 Experimental setup for evaluating drug release efficiency

III. RESULT

In the targeted drug release experiments using the multichannel ultrasound-driven module, an ultrasound vortex pattern was applied at the drug release target site for 20 minutes, followed by an additional 40 minutes of incubation. Compared to samples that were not exposed to the ultrasound waves, the drug release area was observed to increase by more than 30%.

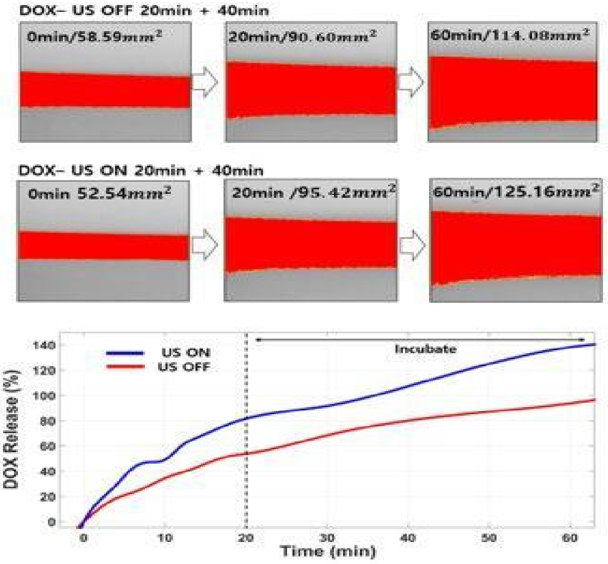


Fig.3 Results of Targeted Drug Release via Ultrasound-Driven Module: Analysis of Drug Release Area Using ImageJ (TOP), Time-Dependent Drug Release Graph (Bottom)

IV. CONCLUSIONS

In this study, generating an ultrasound vortex pattern at the target site using a multichannel module comprising 86 transducers was shown to enhance the release efficiency of doxorubicin. These results indicate the potential applicability of this approach for targeted drug delivery technologies.

ACKNOWLEDGMEN

This research was supported by a grant of the Korea Health Technology Development R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2023-00302153 and RS-2023-00302151).

REFERENCES

[1] H. X. Cao et al., “Micromotor Manipulation Using Ultrasonic Active Traveling Waves,” Micromachines, vol. 12, no. 2, p. 192, Feb. 2021, doi: 10.3390/mi12020192

[2] M.A. Ghanem; A.D. Maxwell; Y.N. Wang ; B.W. Cunitz; V.A. Khokhlova; O.A. Sapozhnikov; M.R. Bailey, “Noninvasive acoustic manipulation of objects in a living body.” Proc. Natl. Acad. Sci. USA 117, 16848–16855, 2020.

[3] B.W. Drinkwater, “Dynamic-field devices for the ultrasonic manipulation of microparticles.” Lab Chip, 16, 2360–2375. 2016.

[4] J. A. Rodríguez Ramírez, M. Ibrahín Gutiérrez, A. V. Hernández, C. Alther Negreira Casares, A. R. Fernandez and L. LEija Salas, "Acoustic Cavitation Detection Produced by Ultrasound in an Agarose-gel Phantom Analyzed with Frequencies Spectrogram," 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Dubrovnik, Croatia, 2020, pp. 1-5, doi: 10.1109/I2MTC43012.2020.9129484.

Machine Learning-Based Estimation of Root-Zone Soil Moisture from Heterogeneous Data Sources

1st Kamilla Rakhymbek
Laboratory of Digital Technologies and Modeling
Sarsen Amanzholov East Kazakhstan University
Oskemen, Kazakhstan
rakhymbekkamilla@gmail.com

2nd Markhaba Karmenova
Department of Computer Modeling and Information Technologies
Sarsen Amanzholov East Kazakhstan University
Oskemen, Kazakhstan
mkarmenova@vku.edu.kz

***Abstract*—Reliable Root-Zone Soil Moisture (RZSM) estimation is a persistent challenge, particularly in data-sparse regions. This study introduces an input-efficient framework to estimate RZSM at 50 cm and 100 cm depths using station meteorological data and a surface soil moisture from AMSR2 and ERA5-Land. An XGBoost model was trained and tested on sites in Kazakhstan and the United States to evaluate both local performance and cross-country generalizability. The proposed models significantly outperformed a baseline derived from reanalysis data. For example, at the Shemonaiha site (0-50 cm), the model reduced RMSE from 0.273 to 0.048. Results showed that using AMSR2 surface moisture predictor was superior for the shallower 0-50 cm layer, while using ERA5-Land surface moisture predictor was more stable for the 0-100 cm depth. Although the models generalized well within regions, their performance degraded significantly in cross-country applications, highlighting the necessity of regional adaptation. This study provides a practical framework for RZSM estimation, with future work aimed at enhancing model efficiency by incorporating static predictors such as soil properties.**

Keywords—root-zone soil moisture, machine learning, AMSR2, ERA5-Land, model generalizability

I. BACKGROUND

Root-Zone Soil Moisture (RZSM), the water held in the top meter of soil, is a critical variable for agriculture, hydrology, and climate regulation. It governs crop growth, influences drought and flood development, and is recognized as an essential climate variable [1]. Despite its importance, obtaining reliable measurements of RZSM is a persistent challenge. In-situ monitoring networks are often sparse, particularly in regions like Central Asia, and are frequently limited to the top layer of soil depth [2]. While satellite remote sensing provides broad spatial coverage, it can only measure moisture on the surface of soil at a coarse resolution [3], and reanalysis products such as ERA5-Land have been shown to be less reliable at deeper soil layers where they are not constrained by direct observations [4][5]. This leaves a critical gap in our ability to monitor the subsurface moisture conditions vital for decision-making.

Recent research has begun to address this gap by exploring both data assimilation and machine learning approaches. For instance, the SMAP mission provides root-zone product through assimilation of surface soil moisture into land surface models, though its coarse resolution limits field-scale use [6]. At the same time, data-driven methods such as SoMo.ml have demonstrated that global daily soil moisture estimates down to 50 cm can be generated by training recurrent neural networks on in-situ observations [7]. Other studies have shown that ensemble tree-based models can fuse multiple predictors such as satellite indices, meteorological forcing, and soil maps for 0-100 cm root-zone estimation [8]. Region-specific work has further applied downscaling and filtering models to refine spatial resolution or temporal dynamics of root-zone moisture [9], [10]. In addition to serving as an input source, ERA5-Land is often used as a reference in soil-moisture evaluations [4], and we adopt it here as a baseline at 50 cm and 100 cm using a weighted average of layers.

In this study, we estimate root-zone soil moisture at 50 and 100 cm using meteorological inputs together with surface soil moisture products of ERA5-Land reanalysis or AMSR2 satellite data. Our primary contribution is a practical, input-efficient framework that demonstrates how these available predictors can support subsurface moisture estimation at operational depths in data-limited regions. Using only shared predictors aligned in temporal resolution, we test cross-region generalization by applying models trained in the United States to Kazakh sites and vice versa. The results provide evidence on where performance is retained or degrades and clarify realistic expectations for input selection and cross-climate use.

II. PROBLEM DEFINITION

Accurately estimating root-zone soil moisture remains to be a challenge, especially where in-situ monitoring is sparse and remote sensing is limited to the surface. While advanced data assimilation and machine learning methods show promise, practical frameworks that use minimal set of easily available

---

This work has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant Number AP23489072).

inputs are needed and their generalizability across different soil types and climates is not well established. This study addresses that gap by framing the problem as follows: Can we accurately estimate RZSM at 50 cm and 100 cm using station meteorological data combined with common surface soil moisture products such as ERA5-Land or AMSR2? We evaluate this input-efficient approach by testing both its local performance and its cross-region applicability, while using ERA5-Land estimates normalized to target depths as a benchmark for comparison.

III. MATERIALS AND METHODS

A. Data Description

Four sites were selected for this study: two in East Kazakhstan (Shemonaiha and Dmitrievka) and two in Montana, United States (Conrad Ag RC and Violett). Key site characteristics are summarized in Table I.

TABLE I. SITE CHARACTERISTICS SUMMARY

Site	Country	Lat, Lon	Climate	Data source
Shemo-naiha	KZ	50.6191°N, 81.8609°E	Continental, snowmelt-driven hydrology, cold winters, warm summers	Local monitoring station
Dmitri-evka	KZ	50.7622°N, 80.8248°E		Local monitoring station
Conrad Ag Rc (2117)	U.S.	48.3000°N, 111.9167° W	Temperate continental, semi-arid, drier relative to Kazakhstan	USDA NRCS SCAN
Violett (2118)	U.S.	48.4333°N, 111.1833° W		USDA NRCS SCAN

For the Kazakhstani sites, daily air temperature and precipitation along with dekadal soil-moisture reserves at 0–50 cm and 0–100 cm (mm) data is obtained from Kazhydromet [11]. For the U.S. sites, hourly meteorological and multi-depth soil moisture data from USDA NRCS SCAN [12] stations were aggregated first to daily and then to dekadal values. To complement these in-situ records, the study also incorporated reanalysis and satellite products for both regions. ERA5-Land soil moisture (0-7 cm) and AMSR2 SSM (ascending/descending passes) were sampled near each site and averaged to the dekadal time step. All soil moisture data, including those from Kazhydromet, were converted to fractional volumetric content for comparability.

The target variables included productive soil moisture in Kazakhstan and volumetric soil moisture in the U.S., each at 0-50 cm and 0-100 cm depths. For the U.S. sites, root-zone values were derived using depth-weighted aggregation and taken last value of the dekad. All variables were aligned to a dekadal time step and spatially aligned to station coordinates. Predictor variables included mean dekadal air temperature, total precipitation, active temperature sums (>10 °C), two-dekad rolling precipitation sum, AMSR2 or ERA5-Land surface soil moisture, site coordinates, and calendar month.

As a reference, we formed an ERA5-Land baseline by adjusting to the target depths using weighted averaging: the top two layers for 0-50 cm and the top three layers for 0-100 cm.

B. Modeling Approach and Experiments

As the modeling method, we employed eXtreme Gradient Boosting (XGBoost). The model was optimized using mean squared error (MSE) loss with a fixed random seed (42) to ensure reproducibility. Hyperparameters were kept constant across all runs, with a learning rate of 0.005, maximum tree depth of 6, number of trees of 600, and 0.8 subsampling for both rows and features.

Models were trained separately for each site and soil depth using non-overlapping year splits (Table II). To focus on agriculturally relevant conditions, we restricted all experiments to the vegetative season (April-September).

TABLE II. TRAINING, VALIDATION AND TEST SPLITS FOR EACH SITE

Sites	Train	Val	Test	Season
Shemonaiha, Dmitrievka	2012-2018	2019-2020	2021-2024	Apr–Sep
Conrad Ag Rc (2117), Violett (2118)	2013-2019	2020-2021	2022-2024	Apr–Sep

To isolate the effect of the remote input, we trained two otherwise-identical model variants that differed only in the surface soil surface predictor: (i) AMSR2 SSM (0-5 cm) and (ii) ERA5-Land soil moisture (0-7 cm). All meteorological features and hyperparameters are held fixed.

We evaluated the following experimental setups:

- Intra-station – training and testing on the same site with non-overlapping year splits.
- Within-region – training on one site and testing on the other site in the same region.
- Cross-country – training on the Kazakhstani sites and evaluating on the U.S. sites, and vice versa.

Together, these setups test model generalization from within-station to cross-country scales, isolates the effect of the remote input, and enables a fair, reproducible comparison.

We report performance using two complementary metrics: Root Mean Squared Error (RMSE) to quantify absolute accuracy in the native units, and Kling-Gupta Efficiency (KGE) to summarize key aspects for hydrologic relevance which are correlation, variability, and bias.

IV. RESULTS AND ANALYSIS

The results section presents the prediction performance of XGBoost models for RZSM under multiple experimental setups, including intra-station, within-region, and cross-country evaluations.

For the intra-station setup, detailed results are reported in Tables III and IV, which summarize RMSE and KGE values for depths of 0-50 cm and 0-100 cm using AMSR2 and ERA5-Land as surface soil moisture predictors.

TABLE III. RZSM (0-50 CM) PREDICTION PERFORMANCE COMPARISON

Site	Baseline		AMSR2 (SSM predictor)		ERA5-Land (SSM predictor)	
	RMSE	***KGE***	***RMSE***	***KGE***	***RMSE***	***KGE***
Shemonaiha	0.273	-2.102	0.048	0.318	0.050	0.298
Dmitrievka	0.135	-0.565	0.057	0.141	0.052	0.273
Conrad Ag Rc (2117)	0.115	0.473	0.026	0.447	0.027	0.581
Violett (2118)	0.036	0.599	0.029	0.755	0.036	0.679

For the 0-50 cm depth, XGBoost models consistently outperformed the baseline across all sites, increasing KGE values and substantially lowering RMSE. AMSR2 generally gave slightly stronger results than ERA5-Land at this depth. This reflects AMSR2's ability to capture rapid responses to precipitation events, which is critical in the shallow root zone.

TABLE IV. RZSM (0-100 CM) PREDICTION PERFORMANCE COMPARISON

Site	Baseline		AMSR2 (SSM predictor)		ERA5-Land (SSM predictor)	
	RMSE	***KGE***	***RMSE***	***KGE***	***RMSE***	***KGE***
Shemonaiha	0.247	-1.096	0.067	0.155	0.066	0.188
Dmitrievka	0.111	-0.057	0.072	0.030	0.067	0.183
Conrad Ag Rc (2117)	0.124	-0.183	0.025	0.276	0.022	0.316
Violett (2118)	0.085	0.191	0.045	-0.671	0.038	0.277

At the 0-100 cm depth, XGBoost again improved predictions compared to the baseline, reducing RMSE and increasing KGE at nearly all sites. However, at this depth ERA5-Land often provided more stable performance than AMSR2 and appears more suitable for deeper root-zone prediction, while AMSR2 remains advantageous for capturing near-surface variability.

To assess model generalizability, the results of within-region and cross-country evaluation setups are illustrated in Figures 1 to 4. The results are expressed as RMSE heatmaps for depths of 0-50 cm and 0-100 cm using AMSR2 and ERA5-Land surface soil moisture predictors.

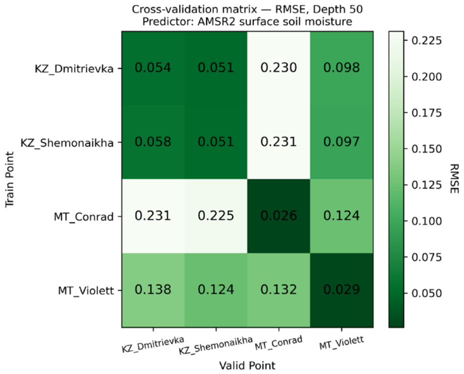


Fig. 1. Generalizability evaluation (0-50 cm): AMSR2.

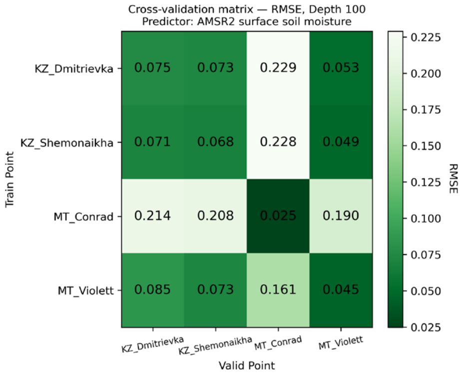


Fig. 2. Generalizability evaluation (0-100 cm): AMSR2.

At the depth of 0-50 cm (Fig. 1), within-region transfers show low RMSE, while cross-country transfers (KZ to US and vice versa) are markedly higher.

At the deeper layer of 0-100 cm (Fig. 2), within-region transfers remain adequate, though errors increase slightly compared to the 0-50 cm depth, as expected given the slower dynamics of subsurface moisture. Cross-country transfers again perform poorly, with substantially higher RMSE values.

These findings demonstrate that while AMSR2-based models are transferable within the same region, their portability across distinct climatic and soil regimes is limited, particularly at greater depths.

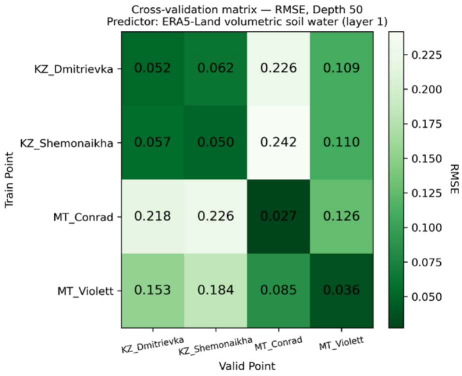


Fig. 3. Generalizability evaluation (0-50 cm): ERA5-Land.

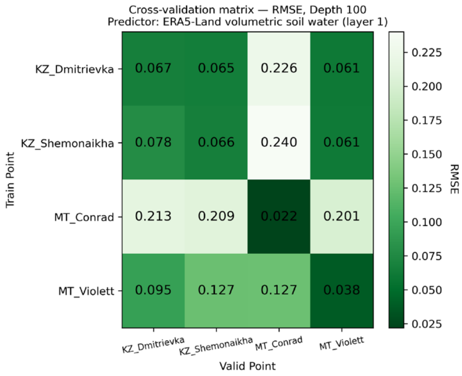


Fig. 4. Generalizability evaluation (0-100 cm): ERA5-Land.

Results at 0-50 cm using ERA5-Land (Fig. 3) are broadly consistent with those from AMSR2. At 0-100 cm, ERA5-Land (Fig. 4) maintains relatively stable within-region generalization, suggesting that its smoothed representation aligns more closely with deeper soil moisture dynamics. However, cross-country setup continues to exhibit large errors.

V. CONCLUSION

The study evaluates an input-efficient framework for estimating root-zone soil moisture at depths of 50 cm and 100 cm using meteorological data from weather station together with surface soil moisture predictors from AMSR2 and from the 0-7 cm layer of ERA5-Land. We evaluated these models against an ERA5-Land reanalysis reference normalized to target depths and used as a baseline. Across all experimental setups, XGBoost models provided superior performance compared to reanalysis reference, indicating that common meteorological inputs with surface soil moisture can meaningfully recover subsurface variability at operational depths.

Two main insights stand out: (1) using AMSR2 as the predictor better captures event-scale variability at 0-50 cm, while using ERA5-Land surface soil moisture yields more stable performance at 0–100 cm, consistent with slower deep-soil dynamics; (2) generalization is reliable within a region but degrades across countries, highlighting climatic and soil differences and the need for regional adaptation.

Future work will focus on improving model performance by integrating explicit physical information. Our primary next steps include enriching the models with static environmental predictors, such as soil texture and land cover data, to better account for landscape differences. We also plan to implement targeted transfer learning strategies, like model fine-tuning, to adapt models from data-rich to data-sparse regions.

REFERENCES

[1] M. Li, "A review of root zone soil moisture estimation methods: Progress and perspectives," *Remote Sensing*, vol. 15, no. 22, p. 5361, 2023, doi: 10.3390/rs15225361.

[2] M. Schönauer et al., "Soil moisture modeling with ERA5-Land retrievals, topographic indices, and in situ measurements and its use for predicting ruts," *Hydrology and Earth System Sciences*, vol. 28, no. 12, pp. 2617–2633, 2024, doi: 10.5194/hess-28-2617-2024.

[3] S. A. Sahaar and J. D. Niemann, "Estimating rootzone soil moisture by fusing multiple remote sensing products with machine learning," *Remote Sensing*, vol. 16, no. 19, p. 3699, 2024, doi: 10.3390/rs16193699.

[4] J. Muñoz-Sabater et al., "ERA5-Land: A state-of-the-art global reanalysis dataset for land applications," *Earth System Science Data*, vol. 13, no. 9, pp. 4349–4383, 2021, doi: 10.5194/essd-13-4349-2021.

[5] A. van Gent, *Evaluation of the ERA5-Land dataset for estimation of soil moisture in the permafrost region*, Dissertation, Stockholm Univ., Stockholm, Sweden, 2023. [Online]. Available: https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-218534

[6] R. H. Reichle, J. V. Ardizzone, G. K. Kim, R. A. Lucchesi, E. B. Smith, and B. H. Weiss, *Soil Moisture Active Passive (SMAP) Mission Level 4 Surface and Root Zone Soil Moisture (L4_SM) Product Specification Document*, Tech. Rep., NASA Goddard Space Flight Center, Greenbelt, MD, USA, 2025.

[7] R. Orth, "Global soil moisture data derived through machine learning trained with in-situ measurements," *Scientific Data*, vol. 8, no. 1, pp. 1–14, 2021, doi: 10.1038/s41597-021-00964-1.

[8] L. Karthikeyan and A. K. Mishra, "Multi-layer high-resolution soil moisture estimation using machine learning over the United States," *Remote Sensing of Environment*, vol. 266, p. 112706, 2021, doi: 10.1016/j.rse.2021.112706.

[9] T. Mahmood, J. Löw, J. Pöhlitz, J. L. Wenzel, and C. Conrad, "Estimation of 100 m root zone soil moisture by downscaling 1 km soil water index with machine learning and multiple geodata," *Environmental Monitoring and Assessment*, vol. 196, no. 9, p. 823, 2024, doi: 10.1007/s10661-024-12969-5.

[10] C. Carranza, C. Nolet, M. Pezij, and M. van der Ploeg, "Root zone soil moisture estimation with Random Forest," *Journal of Hydrology*, vol. 593, p. 125840, 2021, doi: 10.1016/j.jhydrol.2020.125840.

[11] Kazhydromet, *Hydrological Database*. [Online]. Available: https://meteo.kazhydromet.kz/database_hydro_en/. [Accessed: May 20, 2025].

[12] G. L. Schaefer, M. H. Cosh, and T. J. Jackson, "The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN)," *J. Atmos. Oceanic Technol.*, vol. 24, no. 12, pp. 2073–2077, Dec. 2007.

Bearing-based Distributed Formation Control for Unmanned Aerial Vehicles

Jaewan Choi
Department of Artificial Intelligence
Korea Aerospace University
Gyeunggi-do, Republic of Korea
jwmc_1118@kau.kr

Younghoon Choi
Department of Smart Drone Engineering
Korea Aerospace Univerity
Gyeunggi-do, Republic of Korea
younghoon.choi@kau.ac.kr

Abstract— **The utilization of unmanned aerial vehicles (UAVs) for reconnaissance and surveillance has been rapidly increasing in recent years. To effectively monitor wide areas, significant research efforts have focused on the simultaneous operation of multiple UAVs, including formation flight and task allocation. This study proposes a distributed formation-flight system in which UAVs maintain formation through inter-vehicle information exchange. The system utilizes bearing-based data to preserve the formation geometry and addresses a key limitation of conventional bearing-based approaches, which require two or more leaders. The proposed method enables formation maintenance with a single leader. Furthermore, a collision-avoidance algorithm is designed, allowing each UAV to autonomously avoid obstacles while maintaining the overall formation as much as possible. The proposed system is validated through numerical simulations in MATLAB, demonstrating its effectiveness in both formation maintenance and safe maneuvering.**

Keywords—Distributed Formation Control, Leader-Follower, Bearing-based Control, Unmanned Aerial Vehicle

I. INTRODUCTION

In modern warfare, Unmanned Aerial Vehicle (UAVs) are increasingly used for reconnaissance and target detection[1-2]. To achieve wide-area surveillance or high-precision detection, multiple UAVs are often required to operate simultaneously. This requires multi-agent technologies such as path planning, task allocation, and formation control [3]. Among them, formation control is a core technology that enables multiple UAVs to fly in coordinated shapes while maintaining a structured group formation.

Distributed formation control [4–5] refers to a strategy in which each UAV maintains the formation through continuous information exchange with neighboring agents and collective decision-making. Due to its scalability, reliability, and fault tolerance, this approach is particularly well-suited for modern battlefields characterized by attritional warfare. Bearing-based formation control [6–7] represents a representative approach in distributed formation control, utilizing relative bearing information between agents to form and maintain a desired configuration [6–7]. However, conventional bearing-based methods require at least two leaders to maintain the formation geometry. Moreover, obstacle avoidance by follower UAVs depends on the intentional movement of the leaders, which limits flexibility and operational efficiency.

To overcome these limitations, this study proposes a novel bearing-based formation system that enables the entire formation to be maintained with only a single leader. Additionally, each UAV is capable of autonomously avoiding obstacles, thereby enhancing the flexibility and efficiency of multi-agent operations.

II. BEARING-BASED FORMATION CONTROL

A. Preliminary

Let the set of agents be defines as $V = \{1, 2, \dots, n\}$, and the connectivity between agents be denoted as $E \subseteq V \times V$. The communication topology among n agents can then be represented as $G = (V, E)$. If agent i receives information from agent j, this relationship can be expressed as $(i, j) \in E$. The set of neighboring agents from which agent i receives information is defined as $N_i = \{j \in V : (i, j) \in E\}$.

The number of leaders and followers are defined as n_l and n_f, respectively, and assume that each agent has position and velocity vectors p_i and v_i. The position vectors of the leaders and followers can be collectively represented as $\boldsymbol{p}_l = \left[\boldsymbol{p}_1^T, \dots \boldsymbol{p}_{n_l}^T\right]^T, \boldsymbol{p}_f = \left[\boldsymbol{p}_{n_l+1}^T, \dots \boldsymbol{p}_n^T\right]^T$. Similarly, the velocity vectors of the leaders and followers are expressed as $\boldsymbol{v}_l = \left[\boldsymbol{v}_1^T, \dots \boldsymbol{v}_{n_l}^T\right]^T, \boldsymbol{v}_f = \left[\boldsymbol{v}_{n_l+1}^T, \dots \boldsymbol{v}_n^T\right]^T$.

The bearing vector that agent i obtains from its neighbor agent j is defined as follows in (1).

$$\boldsymbol{g}_{ij} = \frac{\boldsymbol{p}_j - \boldsymbol{p}_i}{\|\boldsymbol{p}_j - \boldsymbol{p}_i\|}, \quad \forall (i, j) \in E \tag{1}$$

Using the bearing vector $\boldsymbol{g}_{ij}$, the orthogonal projection matrix $\boldsymbol{P}_{\boldsymbol{g}_{ij}}$ can be denoted as in (2).

$$\boldsymbol{P}_{\boldsymbol{g}_{ij}} = \boldsymbol{I}_d - \boldsymbol{g}_{ij}\boldsymbol{g}_{ij}^T \tag{2}$$

The desired formation $G(\boldsymbol{p}^*)$ satisfy the constraints given in (3) and (4).

$$\frac{\boldsymbol{p}_j^*(t) - \boldsymbol{p}_i^*(t)}{\|\boldsymbol{p}_j^*(t) - \boldsymbol{p}_i^*(t)\|} = \boldsymbol{g}_{ij}^*, \quad \forall (i, j) \in E \tag{3}$$

$$\dot{\boldsymbol{p}}_i^*(t) = \dot{\boldsymbol{p}}_i(t), \quad \forall i \in V_l \tag{4}$$

Based on the target bearing vector $\boldsymbol{g}_{ij}^*$, the bearing Laplacian matrix $\boldsymbol{B}\big(\boldsymbol{G}(\boldsymbol{p}^*)\big)$ can be defined as follows in (5).

$$[\boldsymbol{B}(\boldsymbol{G}(\boldsymbol{p}^*))]_{ij} = \begin{cases} \boldsymbol{0}_{d\times d}, & i \neq j, \ (i,j) \notin E \\ -\boldsymbol{P}_{\boldsymbol{g}^*_{ij}}, & i \neq j, \ (i,j) \in E \\ \sum_{k\in N_i} \boldsymbol{P}_{\boldsymbol{g}^*_{ik}}, & i = j, \ i \in V \end{cases} \tag{5}$$

The bearing Laplacian matrix can be partitioned into block form as shown in (6), separating the leader-follower structure.

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_{ll} & \boldsymbol{B}_{lf} \\ \boldsymbol{B}_{fl} & \boldsymbol{B}_{ff} \end{bmatrix} \tag{6}$$

where $\boldsymbol{B}_{ll}$ represents the bearing Laplacian block corresponding to the leaders, $\boldsymbol{B}_{lf} = \boldsymbol{B}_{fl}^T$ denotes the off-diagonal block between leaders and followers, and $\boldsymbol{B}_{ff}$ corresponds to the followers' bearing Laplacian block.

If the desired formation includes two or more leaders, the positions of the follower agents can be determined using the bearing Laplacian as follows:

$$\boldsymbol{p}_f^*(t) = -\boldsymbol{B}_{ff}^{-1}\boldsymbol{B}_{fl}\boldsymbol{p}_l^*(t) \tag{7}$$

B. Bearing-based Target Formation Design

Conventional bearing-based formation control methods require two or more leaders, which limits their applicability to real-world systems. To overcome this limitation, a virtual leader-based bearing formation system is proposed, as illustrated in Fig. 1.

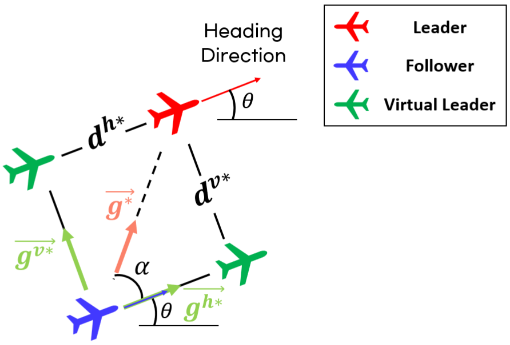


Fig. 1. Virtual leader-based formation system

In this approach, two virtual leaders are generated: one positioned at a distance d^{h*} along the leader's direction of motion, and the other at a distance d^{v*} in the orthogonal direction. These virtual leaders serve as reference points that determine the follower's position based on bearing constraints with two virtual leaders. The desired bearing vectors from the follower to the virtual leaders are given by $\boldsymbol{g}^{h*} = [1, 0]^T$, $\boldsymbol{g}^{v*} = [0, 1]^T$. Under these constraints, the virtual leader aligned with the motion direction determines the forward position of the follower, while the perpendicular virtual leader determines its lateral position.

Although the leader typically generates an obstacle-free path, however, collisions may still occur due to the size or configuration of the formation. To enhance robustness of the system, each follower can temporarily adjust its vertical distance d^{v*} upon detecting a potential collision. This allows autonomous obstacle avoidance without disrupting the overall formation.

A V-shaped formation is designed using the proposed virtual leader-based system, as depicted in Fig. 2 and Table. 1.

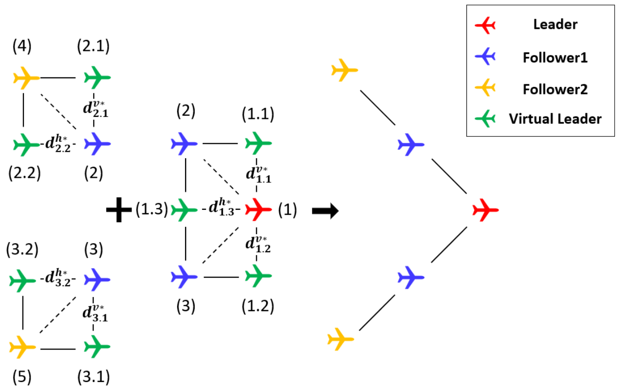


Fig. 2. Desired V-shaped formation

TABLE I. PARAMETER OF V-SHAPED FORMATION

$\boldsymbol{d^{v*}_{1.1}}, \boldsymbol{d^{v*}_{1.2}}, \boldsymbol{d^{h*}_{1.3}}, \boldsymbol{d^{v*}_{2.1}}, \boldsymbol{d^{h*}_{2.2}}, \boldsymbol{d^{v*}_{3.1}}, \boldsymbol{d^{h*}_{3.2}}$	2 m

C. Controller Design

To control follower UAVs modeled by a nonholonomic system, the controller is designed as follows in (8).

$$\begin{aligned} \dot{p}_i &= -k_p R_l \sum_{j\in N_i} P_{g^*_{ij}} R_l^T (p_i - p_j) - k_I R_l q_i \\ \dot{q}_i &= \sum_{j\in N_i} P_{g^*_{ij}} R_l^T (p_i - p_j) \\ \dot{\theta}_i &= -k_\omega (\chi_{\theta_i}^{\perp})^T R_l \sum_{j\in N_i} P_{g^*_{ij}} R_l^T (p_i - p_j) \end{aligned} \tag{8}$$

where k_p and k_ω are proportional gains for position and heading control respectively, while k_I is the integral gain for position. The matrix R_l denotes the rotation matrix of the leader.

III. SIMULATION RESULTS

A. Simulations

To validate the proposed virtual leader-based bearing formation system, numerical simulations were conducted in MATLAB. The control gains used for the follower agents in the simulation were set to $k_p = 7, k_I = 0.5$ and $k_\omega = 10$.

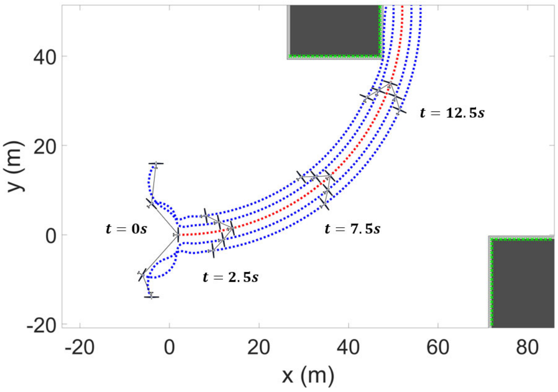


Fig. 3. V-shaped formation control simulation result ($0 \le t \le 15$)

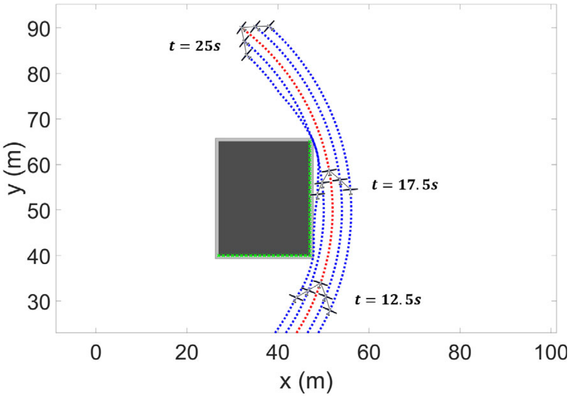


Fig. 4. V-shaped formation control simulation result ($15 \le t \le 25$)

As shown in Fig. 3, the followers successfully form a V-shaped formation within approximately 2.5 seconds. Subsequently, at around 12.5 seconds, follower 4 detects a potential collision with an obstacle and autonomously adjusts its vertical distance $d_{2.1}^{v*} \to 0$ to perform an avoidance maneuver. In Fig. 4, it is observed that follower 4 maintains $d_{2.1}^{v*} = 0$ for approximately 5 seconds to avoid the collision, and then successfully returns to the original formation.

B. *Conclusion*

In this paper, a virtual leader-based bearing formation system is proposed for the coordinated operation of multiple UAVs. The proposed method addresses the limitations of conventional bearing-based formation control that require two or more leaders, by enabling formation maintenance with only a single leader. Additionally, a collision avoidance algorithm suitable for distributed formation flight is developed.

To validate the proposed system, numerical simulations are performed in a V-shaped formation scenario. The results show that the UAVs, starting from arbitrary initial positions and orientations, converge reliably to the desired formation while tracking a moving leader. In addition, when potential collisions with obstacles are detected, each UAV autonomously adjusts its position within the formation, demonstrating the system's capability to avoid obstacles while preserving formation integrity.

ACKNOWLEDGMENT

This work was supported by the GRRC Program of Gyeonggi province [GRRC-KAU-2023-B01, Fusion Technology Research Center for Advanced Air Mobility].

REFERENCES

[1] W. Liu, T. Zhang, S. Huang, and K. Li, "A hybrid optimization framework for UAV reconnaissance mission planning," Comput. Ind. Eng., vol. 173, p. 108653, 2022.

[2] W. Y. H. Adoni, S. Lorenz, J. S. Fareedh, R. Gloaguen, and M. Bussmann, "Investigation of autonomous multi-UAV systems for target detection in distributed environment: Current developments and open challenges," Drones, vol. 7, no. 4, p. 263, 2023.

[3] Y. Alqudsi, and M. Makaraci, " UAV swarms: research, challenges, and future directions", Journal of Engineering and Applied Science, vol. 72, no. 1, p. 12, 2025.

[4] S. Vargas, H. M. Becerra, and J. B. Hayet, "MPC-based distributed formation control of multiple quadcopters with obstacle avoidance and connectivity maintenance," Control Engineering Practice, vol. 121, p. 105054, 2022.

[5] S. Huang, T. Wang, Y. Tang, Y. Hu, G. Xin, and D. Zhou, "Distributed and scalable cooperative formation of unmanned ground vehicles using deep reinforcement learning," Aerospace, vol. 10, p. 96, 2023.

[6] S. Zhao and D. Zelazo, "Bearing-based formation maneuvering," in Proc. IEEE Int. Symp. Intelligent Control (ISIC), Sep. 2015, pp. 658–663.

[7] M. H. Trinh, S. Zhao, Z. Sun, D. Zelazo, B. D. O. Anderson, and H. S. Ahn, "Bearing-based formation control of a group of agents with leader–first follower structure," IEEE Trans. Autom. Control, vol. 64, pp. 598–613, 2018.

Reinforcement Learning Based Controller for High-Precision Dual-Servo Stages

1st Seung Woo Seo
School of Mechanical Engineering
Ajou university
Suwon, South Korea
seosw99@ajou.ac.kr

2nd Yongho Jeon
School of Mechanical Engineering
Ajou university
Suwon, South Korea
princaps@ajou.ac.kr

3rd Moon Gu Lee*
School of Mechanical Engineering
Ajou university
Suwon, South Korea
*moongulee@ajou.ac.kr

Abstract— **In semiconductor manufacturing, ultra-precision stages that transfer wafer or reticle are widely used not only in the frontend process such as photolithography but also in backend process such as packaging and inspection. Design requirements of these stages are the ability to move long stroke with high speed and high positioning precision. To meet these demands, dual-servo stages with a coarse and fine structure are widely employed. However, the traditional control methods such as PID (Proportional - Integral - Differential) controller and MPC (Model Predictive Controller) have shown limitations in achieving fast and precise control due to systems' nonlinearities, disturbances, and structural uncertainties. To improve the control performance, this work proposes a reinforcement learning based controller for the aforementioned dual-servo stage. This enabled faster and more precise control despite the various disturbances, nonlinearities, and friction in the dual servo stage mechanism. It can contribute to the control system design of next-generation ultra-precision process equipment such as lithography scanners.**

Keywords— Reinforcement Learning, Dual Serro Stage, High Precision Control, Semiconductor Manufacturing

I. INTRODUCTION

In semiconductor manufacturing processes, motion systems that transport wafers or masks are essential not only for preprocessing, such as photolithography, but also for postprocessing, such as packaging and inspection [1]. These motion systems, called stages, are Cartesian robots that move moving parts in a planar coordinate system. The stage's design requirements include high positioning accuracy, high-speed movement, and long strokes. To meet these requirements, dual-servo stages with coarse and fine motion stages are widely used [2]. Conventional PID controllers and MPCs have been commonly used to control dual-servo stages, but their performance is often inadequate due to friction, nonlinearity, structural uncertainty, and electrical disturbances inherent in the stage mechanism.

In this study, a reinforcement learning-based controller is proposed to improve the control performance of the dual-servo stages. This controller enables faster and more precise control despite mechanical friction, nonlinearity, uncertainty, and electrical disturbances. This technology is expected to contribute to improving the control system performance of next-generation ultra-precision manufacturing equipment, such as semiconductor and display equipment.

II. TARGET SYSTEM AND PERFORMANCE

The target system of this work is the dual-servo stage system composed of a coarse stage based on a BLDC (Brushless Direct Current) linear motor and a fine stage based on a VCM (Voice Coil Motor), as shown in Fig. 1. The coarse stage is driven by a BLDC linear motor and guided by an LM guide, enabling large travel ranges and high-speed motion. However, due to the characteristics of the ball-type LM guide, the system is subject to nonlinearities, disturbances, and uncertainties, including stiffness variations induced by stick-slip phenomena, frictional forces caused by slipping, and changes in lubricant properties influenced by environmental factors such as temperature and humidity. Figure 2 shows a top view of the fine stage. The fine stage employs a VCM as its actuator and a leaf spring as its guiding mechanism, providing high precision positioning with minimal mechanical contact and effectively eliminating backlash [3]. For position sensing, a high-resolution linear encoder and a laser interferometer are utilized to achieve both relative and absolute displacement measurements with nanometer-level accuracy. While the proposed dual stage system combines the advantages of large-range, high-speed motion and fine, high-precision adjustment, it also requires advanced control strategies to address the complex nonlinearities and external disturbances inherent in the system.

To address these challenges, a reinforcement learning (RL) based controller is proposed for the system. The RL-based controller is a type of system that learns to make decisions by trial and error, like how humans learn through experience.

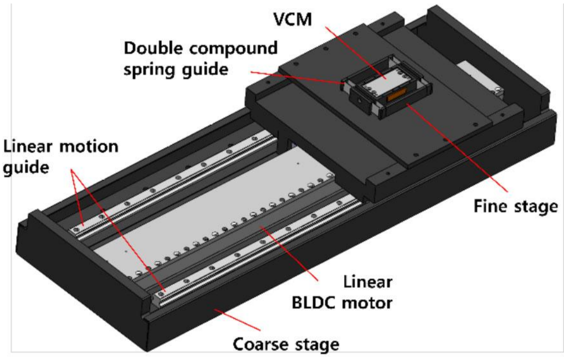


Fig. 1. Schematics of the dual servo stage.

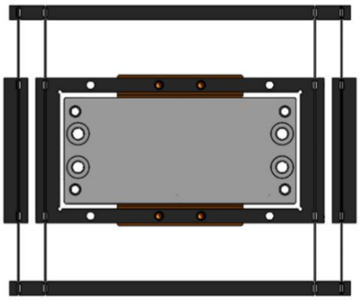


Fig. 2. Schematic top view of the fine stage.

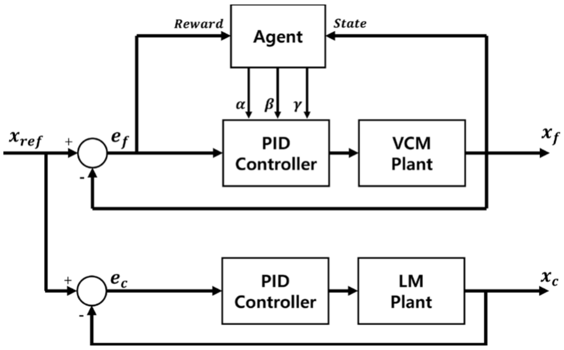


Fig. 3. Parallel type control strategy.

Instead of following a pre-set formula, it improves over time by getting rewards for good action and penalties for bad ones. The controller continuously improves adaptive control policy. Good action is an enhancement in speed and precision, and bad action is the opposite. The controller is more efficient in nonlinear and stochastic situations than the conventional methods. Figure 3 presents the control block diagram designed for this system. To implement dual servo control of the coarse and fine stages, a parallel-type control strategy is adopted, where reinforcement learning is independently applied to each control path. A simulation environment was built to quantitatively compare the RL controller's performance—such as position tracking accuracy, settling time, and overshoot—against conventional controllers. The conventional controllers are based on mathematical modeling and employing PID controllers or MPC. The RL-based controller demonstrated superior precision and faster response across a range of operating conditions. It successfully maintained nanometer-level accuracy even during high-speed movements.

III. CONTROLLER DESIGN

The proposed RL-based controller was designed as a hybrid scheme combining reinforcement learning with a PID controller. Specifically, the RL agent learned scaling factors that are multiplied by the PID gains, thereby adaptively tuning the controller in real time. Design of reward function has been shown to be a key factor not only for improving learning performance but also for determining learning stability, sample efficiency, and practical applicability [5, 6]. Therefore, recognizing the importance of reward function design, we formulated the reward function as a combination of performance-related and stability-related terms.

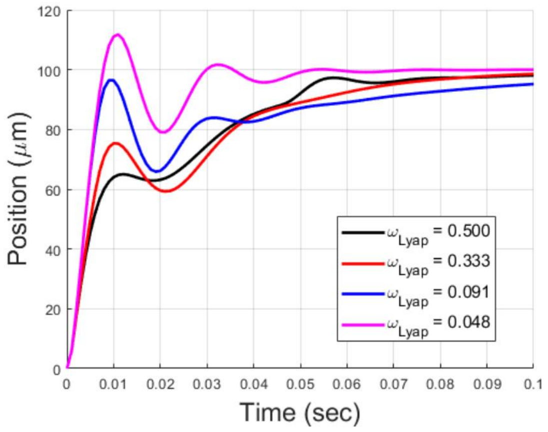


Fig. 4. Comparison of learning curves under different weighting factor settings.

Among various RL algorithms, the Deep Deterministic Policy Gradient (DDPG) was adopted. Because DDPG employs a deterministic policy, it has been reported that the algorithm generates precise control actions in continuous spaces. Moreover, it is effective when the amount of data is small or the data collection is expensive and time-consuming [7, 8].

IV. RESULTS

To validate the effectiveness of the proposed RL-based controller, a simulation environment was developed. Its performance was quantitatively compared against conventional PID controllers. Figure 5 presents the step response of both controllers, clearly showing that the RL-based controller achieves faster settling time and smaller overshoot compared to the conventional PID controller.

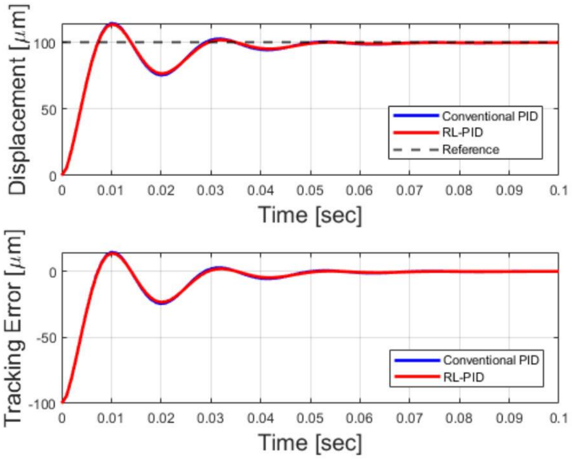


Fig. 5. Plots of step response and tracking error.

V. CONCLUSION

This study shows that reinforcement learning-based control can effectively overcome the limitations of the conventional control approaches and serve as a viable alternative for high-precision systems with complex dynamics. It can contribute to the control system design of next-generation ultra-precision process equipment such as lithography scanners.

Future work will further validate the stability and generalization capability of the proposed controller, and extend its application to dual servo systems. Experimental implementation on actual dual servo platform will demonstrate its effectiveness in achieving both fast response and high-precision tracking.

ACKNOWLEDGEMENTS

This work was supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE, South Korea), RS-2025-02304112, HRD program for Industrial Innovation.

REFERENCES

[1] A. A. R. Tonoy, M. Ahmed, and M. R. Khan, "Precision mechanical systems in semiconductor lithography equipment design and

development," Am. J. Adv. Technol. Eng. Solut, vol. 1, no. 1, pp. 71–97, 2025.

[2] Z. Zhou, J. Gao, and L. Zhang, "A positioning and tracking performance-enhanced composite control algorithm for the macro-micro precision stage," Actuators, vol. 13, no. 11, p. 433, 2024.

[3] D. J. Lee, K. Kim, K. N. Lee, H. G. Choi, N. C. Park, Y. P. Park, and M. G. Lee, "Robust design of a novel three-axis fine stage for precision positioning in lithography," Proc. Inst. Mech. Eng. Part C: J. Mech. Eng. Sci, vol. 224, no. 4, pp. 877–888, 2010.

[4] D. J. Lee, K. Kim, K. N. Lee, H. G. Choi, N. C. Park, Y. P. Park, and M. G. Lee, "Robust design of a novel three-axis fine stage for precision positioning in lithography," Proc. Inst. Mech. Eng. Part C: J. Mech. Eng. Sci., vol. 224, no. 4, pp. 877–888, 2010.

[5] Y. Dong, X. Tang, and Y. Yuan, "Principled reward shaping for reinforcement learning via Lyapunov stability theory," Neurocomputin*, vol. 393, pp. 83–90, 2020.

[6] Y. Chow, O. Nachum, E. Duenez-Guzman, and M. Ghavamzadeh, "A Lyapunov-based approach to safe reinforcement learning," Adv. Neural Inf. Process. Syst., vol. 31, 2018.

[7] R. S. Bhourji, S. Mozaffari, and S. Alirezaee, "Reinforcement learning DDPG–PPO agent-based control system for rotary inverted pendulum," Arab. J. Sci. Eng., vol. 49, no. 2, pp. 1683–1696, 2024.

[8] Y. Ding, X. Ren, X. Zhang, X. Liu, and X. Wang, "Multi-phase focused PID adaptive tuning with reinforcement learning," Electronics, vol. 12, no. 18, p. 3925, 2023.

Session L3: Frontier Special Session

Noise-Augmented CBAM-Unet for Robust Ultrasound Liver Tumor Segmentation

Minsun Kim
Graduate School of Data Science
Chonnam National University
Gwangju, South Korea
kimminsunbona@jnu.ac.kr

Seokmin Hong
Medical AI Lab
Korea Institute of Medical Microrobotics
Gwangju, South Korea
seokmin.hong@kimiro.re.kr

Byunjeon Kang
Graduate School of Data Science
Chonnam National University
Gwangju, South Korrea
bjkang8204@jnu.ac.kr

***Abstract*— Ultrasound (US) imaging is widely used for liver disease diagnosis and monitoring, but segmentation performance is often limited by speckle noise and acquisition artifacts. This study proposes a Convolutional Block Attention Module (CBAM)-based U-Net framework with Gaussian noise augmentation, using 635 B-mode liver US images with tumor masks. Models were trained under clean-only, noise-only, and mixed-domain regimes and evaluated with Dice, IoU, and pixel accuracy. Mixed-domain training achieved the best cross-domain performance, with CBAM-UNet reaching a Dice of 0.89 and IoU of 0.82. In contrast, clean-only training suffered a Dice drop of over 30% on noise data, while noise-only training maintained Dice ≈ 0.85 on noise sets but reduced performance on clean sets. CBAM-UNet surpassed the baseline U-Net only when domain diversity was present, showing that attention mechanisms are most effective with heterogeneous augmentation. These results demonstrate that noise-augmented training provides a simple and effective approach for robust ultrasound segmentation without explicit domain adaptation.**

Keywords—Ultrasound, liver tumor segmentation, CBAM-UNet, Gaussian noise augmentation

I. INTRODUCTION

Ultrasound imaging is widely used for diagnosing and monitoring liver disease due to its cost-effectiveness and real-time capability. However, segmentation performance is hindered by variability in operator skill, patient anatomy, and device quality. In addition, speckle noise and acquisition artifacts further obscure tumor boundaries, resulting in domain shifts that reduce the generalizability of segmentation models [1,2]. To address these challenges, this study investigates the use of Gaussian noise augmentation in conjunction with a Convolutional Block Attention Module (CBAM)-based U-Net framework for liver tumor segmentation. A total of 635 publicly available B-mode US images with tumor annotations were utilized, and models were trained under three distinct regimes: clean-only, noise-only, and mixed-domain training. Performance was evaluated using the Dice similarity coefficient, Intersection over Union (IoU), and pixel accuracy. Experimental results indicate that mixed-domain training achieves robust cross-domain performance without explicit domain adaptation. Furthermore, CBAM-UNet outperforms the baseline U-Net only when sufficient domain diversity is introduced, highlighting the importance of coupling attention mechanisms with training data for reliable segmentation in ultrasound imaging.

II. RELATED WORK

Speckle noise in ultrasound imaging is a multiplicative and non-Gaussian artifact caused by coherent wave interference [1,2]. It reduces image resolution and contrast while masking subtle structural details. To mitigate these effects, various denoising strategies have been explored, including speckle-reducing anisotropic diffusion (SRAD) [2,7], variational approaches such as total generalized variation (TGV), and statistical filtering methods based on Rayleigh distribution models [3,4]. While these methods can enhance contrast and boundary visibility, overly aggressive filtering may also remove diagnostically relevant textures, underscoring the importance of balanced techniques that preserve critical features for segmentation accuracy [3-7].

Beyond denoising, augmentation strategies tailored to ultrasound have emerged to improve segmentation robustness. Physics-consistent transformations account for fan-shaped imaging geometry, acoustic attenuation, and acoustic shadowing artifacts. Other methods introduce speckle parameter remapping, synthetic shadow generation, and scale variation to replicate acquisition variability [8]. Noise injections represent another effective augmentation pathway. Previous studies have shown that the choice of noise distribution and the diversity of training data are crucial determinants of segmentation performance, emphasizing the role of domain variability in model generalization.

III. EXPERIMENTS

The proposed framework is designed to achieve robust liver tumor segmentation across both clean and noise-corrupted ultrasound domains without explicit domain adaptation. To this end, experiments were to systematically evaluate the effect of Gaussian noise augmentation and the contribution of the CBAM module in improving segmentation robustness.

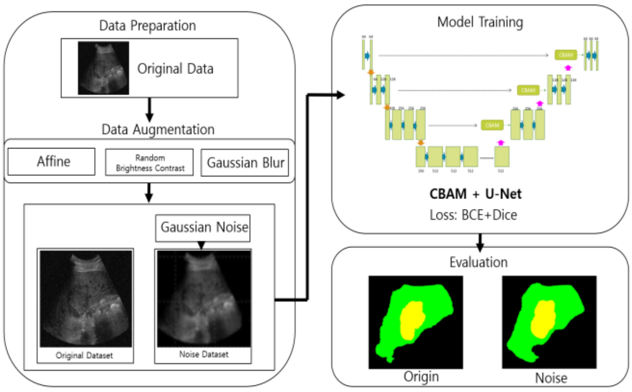


Fig. 1 Overview of the experimental pipeline, from dataset acquisition to the final segmentation results

The entire process was organized into three stages to ensure reproducibility and clarity of evaluation.

(1) Data Preparation, in which both clean and Gaussian noise–corrupted US images are preprocessed and split into training and evaluation sets;

(2) Model Training, where a CBAM-UNet is optimized using a hybrid loss function combining of Binary Cross-Entropy (BCE) and Dice losses; and

(3) Evaluation, which measures model performance on both clean and noise domains.

An overview of this pipeline, from dataset acquisition to the final segmentation results, is illustrated in Fig. 1. The hyperparameters used in this study are as follows. The learning rate was set to 3×10^{-4}, and the batch size was 8. The network was trained for a total of 200 epochs with a dropout rate of 0.3 to prevent overfitting. Additionally, Gaussian noise with a standard deviation σ set to 0.25 was applied during data augmentation to enhance model robustness. Experiments were conducted on a publicly available liver ultrasound dataset from Zenodo [9]. The dataset consists of B-mode liver US images and corresponding tumor masks. All images were resized to 256×256 pixels. To construct the noise domain, Gaussian noise with standard deviation $\sigma=0.25$ was added to the original images. The training-validation split was set to 8:2, and separate original and noise test sets were used for evaluation.

Three evaluation metrics were employed in this study. First, the IoU quantified segmentation performance as the ratio of the overlapping area between the predicted and ground-truth (GT) masks to their union. Second, the Dice Similarity Coefficient was calculated as twice the area of overlap divided by the sum of the predicted and GT, providing a balanced measure of segmentation accuracy, particularly for imbalanced classes. Finally, pixel accuracy was computed as the proportion of correctly classified pixels over the total number of pixels, offering an overall measure of classification performance.

IV. RESULTS

TABLE I. PERFORMANCE COMPARISON OF CBAM-UNET AND U-NET ON CLEAN AND NOISE TEST SETS.

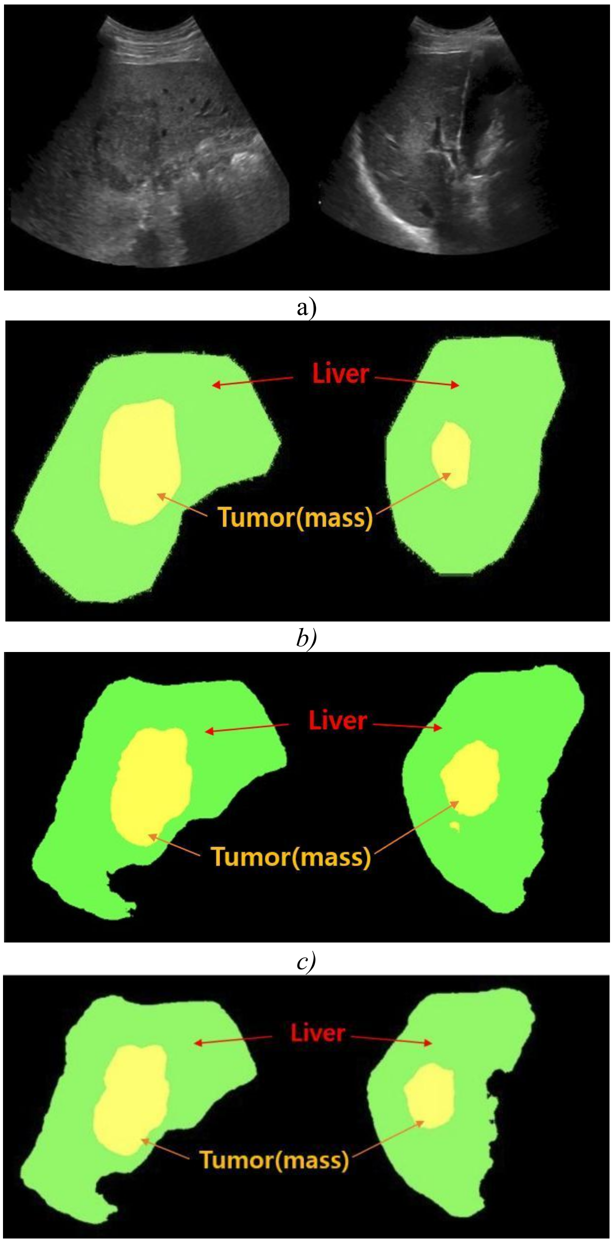
		Original GT		**Noise GT**	
Model	**Data**	***Dice***	***IoU***	***Dice***	***IoU***
CBAM + U-Net	Mixed domain	**0.88**	**0.82**	**0.88**	**0.81**
	Clean	0.85	0.80	0.55	0.40
	Noise	0.84	0.74	0.85	0.76
U-Net	Mixed domain	0.85	0.75	0.85	0.76
	Clean	0.84	0.77	0.76	0.67
	Noise	0.85	0.76	0.86	0.78

Table 1 summarizes the segmentation outcomes across different training regimes. When trained exclusively on clean images, both U-Net and CBAM-UNet demonstrated strong performance on clean test sets but suffered significant degradation on noise-corrupted test data, with Dice scores dropping by more than 30%. This finding indicates that models trained only on clean data lack robustness to noise-induced domain shift.

In contrast, noise-only training improved cross-domain generalization, maintaining Dice scores above 0.85 on noise test sets while sustaining moderate performance on clean test sets. This suggests that training exclusively on corrupted data can partially improve robustness but at the expense of balanced performance across domains.

The best results were achieved with mixed-domain training. Under this regime, CBAM-UNet consistently outperformed U-Net, achieving Dice ≈ 0.89 and IoU ≈ 0.82 across both clean and noise test sets. This demonstrates that balanced domain diversity during training provides the most stable and reliable segmentation performance. Importantly, the integration of CBAM into the U-Net backbone provided clear benefits only in the presence of domain diversity. When trained on either clean-only or noise-only datasets, CBAM-UNet matched or only slightly exceeded U-Net performance, and in some cases risked overfitting. However, under mixed-domain training, the attention mechanism synergized with data variability to refine feature representations and improve boundary delineation, leading to consistently superior results.

Fig. 2 provides qualitative comparisons between U-Net and CBAM-UNet predictions. In clean-only training, tumor boundaries appeared fragmented and less distinct, particularly in noise-corrupted test images. Conversely, CBAM-UNet trained under mixed-domain conditions produced smoother and more accurate tumor contours, closely resembling the ground-truth annotations. These visual results support the quantitative findings, highlighting the necessity of both noise augmentation and attention-based refinement in improving segmentation robustness.



a)

b)

c)

d)

Fig2. Qualitative comparison of segmentation results:
a) Input image, b) Ground-truth mask, c) Prediction by U-Net trained on original and noise datasets, d) Prediction by CBAM+U-Net trained on original and noise datasets

Overall, the results confirm two key observations:
(1) domain diversity, introduced through noise augmentation, is essential for robust ultrasound segmentation; and
(2) CBAM enhances feature discrimination and segmentation accuracy only when trained with heterogeneous data. Together, these findings validate the effectiveness of combining attention mechanisms with noise-based augmentation for cross-domain generalization in ultrasound liver tumor segmentation.

V. CONCLUSION

This study evaluated Gaussian noise augmentation for liver tumor segmentation in ultrasound imaging using 635 B-mode images and a CBAM-UNet architecture under clean-only, noise-only, and mixed-domain training regimes. Results confirmed that domain diversity is the key factor: mixed-domain training consistently achieved robust performance across both clean and noise test sets. CBAM-UNet outperformed U-Net only when sufficient variability was present, indicating that attention-based refinement is most effective when combined with heterogeneous data. These findings suggest that noise augmentation offers a practical alternative to complex domain adaptation methods, improving segmentation robustness in clinical ultrasound environments. Future work will explore varying noise levels and validation on multi-center datasets to further examine generalizability.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) under grant No. RS-2023-00242528, and Korea Health Industry Development Institute(KHIDI), funded by the Ministry of Health & Welfare under grant No. RS-2023-00302153.

REFERENCES

[1] S. Liu, Y. Wang, X. Yang, et al., “Deep learning in medical ultrasound analysis: A review,” *Engineering*, vol. 5, no. 2, pp. 261–275, 2019.
[2] B. Luijten, R. Cohen, C. Brune, et al., “Ultrasound signal processing: From models to deep learning,” *Medical Image Analysis*, vol. 79, p. 102470, 2022.
[3] P. Pramanik, A. Roy, E. Cuevas, M. Perez-Cisneros, and R. Sarkar, “DAU-Net: Dual attention-aided U-Net for segmenting tumor in breast ultrasound images,” *PLoS One*, vol. 19, no. 5, p. e0303670, 2024.
[4] W. Qiu, E. Hamburg, Y. Zhou, and Y. E. Salehani, “Lightweight U-Net for breast ultrasound image segmentation,”2025.
[5] A. Sharma, A. Aich, et al., “Hybrid attention network for accurate breast tumor segmentation in ultrasound images,”2025.
[6] M. Tirindelli, R. Annunziata, and S. Calderara, “Rethinking ultrasound augmentation: A physics-inspired approach,” in Proc. MICCAI, 2021, pp. 123–133.
[7] J. Jiang, et al., “Noise-robustness test for ultrasound breast nodule neural network classifiers,” *Frontiers in Oncology*, vol. 13, 2023.
[8] S. Vesal, et al., “Domain generalization for prostate segmentation in transrectal ultrasound images: A multi-center study,” 2022.
[9] A. Orvile, “Annotated ultrasound liver images dataset,” *Zenodo*, 2022. doi: 10.5281/zenodo.7272660

Registration Between Physical Model and Virtual Model via Image Marker in MR Device

Jingyu Kim
Medical AI Lab
Korea Institute of Medical Microrobotics
Gwangju, South Korea
kjg@kimiro.re.kr

Seokmin Hong
Medical AI Lab
Korea Institute of Medical Microrobotics
Gwangju, South Korea
seokmin.hong@kimiro.re.kr

Byungjeon Kang*
Department of Datascience
Chonnam National University
Gwangju, South Korea
bjkang8204@chonnam.ac.kr

***Abstract*—Mixed Reality(MR) is a technology that combines Virtual Reality(VR) and Augmented Reality(AR). The feature of MR is the ability to interact with both physical and virtual models in real time. This capability has led to extensive research on MR simulations in fields such as architecture, engineering, healthcare, and education. Meanwhile, in the case of effective interaction, it is very important to register between virtual and physical models accurately. Traditional registration methods involve manual registration by users or the use of image markers with patterns to obtain physical model's coordinates. However, there is an issue to acquire specific position of image markers, which is affected by given image pattern, number, their distance, and angle from the MR device. To find the markers accurately, we performed several registration experiments with various image marker-based patterns and numbers. Specifically, we conducted experiments with simple and complex image marker patterns, while varying the number of markers, distance, angle between the image markers and the MR device. Our experimental results show our effective method for the registration between image markers and MR device.**

***Keywords*—registration, MR device, image maker**

I. INTRODUCTION

XR(Extended reality) is categorized into AR(augmented reality), VR(virtual reality), and MR(mixed reality). Among these, MR is being studied and utilized in various fields due to its advantage of enabling real-time interaction between physical models of the real world and virtual models of virtual world[1]. To effectively utilize this MR environment, it is crucial to accurately register the virtual model with the physical model's location in the real world [2, 3]. Previously, manual registration was performed by users or using image markers. However, manual registration methods often resulted in errors during registration, and image marker methods affected by multi-factors. In this study, we experimented and proved the four factors that affects the registration accuracy of image markers.

II. REGISTRATION METHOD

In MR device, there are two primary methods for the registration between physical and virtual models, in which manual registration by the user and registration using image markers. The manual registration experiment was performed to evaluate the difference of the registration accuracy between automatic and manual matching methods.

A. Manual Registration

Manual registration is a method in which the user aligns the physical model with the virtual model on the MR device through three-dimensional interaction with the virtual model.

B. Image Marker Registration

Image marker registration is a method that automatically registers the physical and virtual models using image markers. When the MR device detects the image marker, it calculates the spatial relationship between the image marker and the device. And then, the MR device visualizes the virtual model at a predetermined distance from the center coordinates of the image marker. Finally, the physical model is positioned in a pre-defined location and two models are registered using this method properly.

III. AFFECT FACTORS

In order to achieve high-accuracy registration result via image markers, we performed several registration experiments using four core factors that influence the performance, that are image pattern, number of image markers, distance to the MR device, and angle relative to the MR device, respectively.

A. Pattern

To compare the registration results based on image patterns, we created two image markers with both complex and simple patterns. The complexity was defined by the number of feature points in the image pattern. Simple patterns contain 100 or fewer feature points, and complex patterns contain more than 100 feature points. The simple pattern had 52 feature points, and the complex pattern had 286 feature points in our experiment. The images of each pattern are shown in Fig. 1.

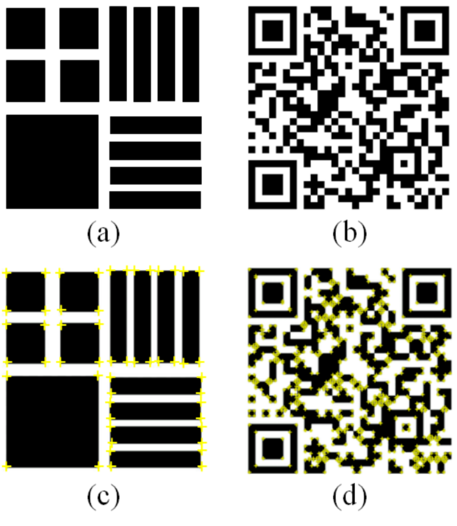


Fig. 1. Image marker pattern (a). Simple pattern (b). Complex pattern (c). Feature points of simple pattern (d). Feature points of complex pattern

B. *Numbers*

To check and compare the registration errors, we created various numbers of markers and analyzed between them. As shown in Fig. 2, our experiments involve multiple image markers. Note that the midpoint between the markers was used as the reference center point of the image markers. In the case of Fig. 2 (a), however, it is not necessary to check the registration error in single marker, so that there is no exist the center point (See Fig. 2. for further details).

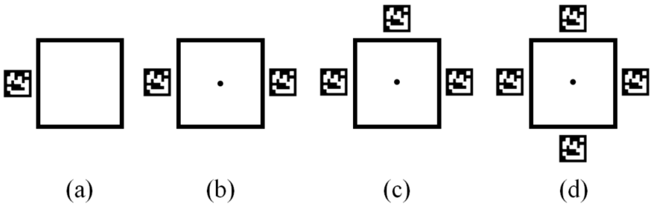


Fig. 2. Composed various numbers of image markers. (a). One image marker (b). Two image markers (c). Three image markers (d). Four image markers

C. *Distance*

To compare the registration results based on the distance between the XR device and image marker, we put them from 30 cm to 120 cm with 30 cm intervals. The distance is measured from the MR device's camera to the center point of the image marker.

D. *Angle*

To compare the registration results based on the angle between the image markers and the MR device, the MR device and image markers were positioned at multiple angles. Experiments were conducted from 0° to 60° at 15° intervals, using the center point of the image marker facing forward as the reference. as shown in Fig. 3.

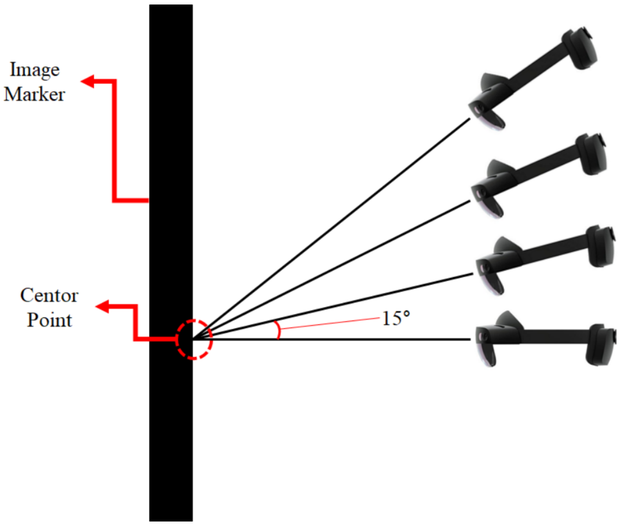


Fig. 3. Angle of Image Marker.

IV. ASSESSMENT OF REGISTRATION

Evaluate the registration result was conducted to determine the optimal method based on changes in image marker factors. The evaluation was performed using TRE (Target Registration Error), which is one of the most common methods for calculating the registration error between physical and virtual models in MR devices[4]. In our experiment, we mainly adopted Microsoft's HoloLens 2 as our MR device, and Vega VT from NDI was employed as the optical tracker.

$$TRE = \sqrt{(A_x - B_x)^2 + \left(A_y - B_y\right)^2 + (A_z - B_z)^2} \qquad (1)$$

Equation (1) calculates the registration error TRE (Target Registration Error) where A and B denoted the 3D coordinates of the physical and virtual models.

V. EXPERIMENTAL RESULT AND SUMMARY

In this study, we explored the optimal approach for achieving accurate registration between physical and virtual models with the use of image markers. The results of the registration evaluation are presented in Tables 1–3, where the registration error is expressed in millimeters. The values in the tables represent the average results acquired from five registration assessments. For all factors except the one being varied the number of markers was set to 1, the distance to 30 cm, and the angle to 0° during the assessment. The results in the tables show that higher registration accuracy was achieved with more complex image marker patterns or a greater number of feature points, a larger number of image markers, shorter distances to the MR device, and angles closer to 0°. In addition, using image markers resulted in higher accuracy than manual registration, although manual registration may yield better accuracy under ideal conditions. Manual registration is highly affected by distance and angle. In future studies, we aim to develop image marker patterns that provide stable registration performance from various angles.

TABLE I. SIMPLE PATTERN RESULTS

<table>
<tr><th>Affect Factors</th><th colspan="5">Simple Pattern Image Marker</th></tr>
<tr><th rowspan="2">Numbers</th><th>One</th><th>Two</th><th>Three</th><th colspan="2">Four</th></tr>
<tr><td>5.86mm</td><td>5.77mm</td><td>5.62mm</td><td colspan="2">5.58mm</td></tr>
<tr><th rowspan="2">Distance</th><th>30cm</th><th>60cm</th><th>90cm</th><th colspan="2">120cm</th></tr>
<tr><td>5.86mm</td><td>5.76mm</td><td>6.23mm</td><td colspan="2">8.89mm</td></tr>
<tr><th rowspan="2">Angle</th><th>0°</th><th>15°</th><th>30°</th><th>45°</th><th>60°</th></tr>
<tr><td>5.86mm</td><td>6.13mm</td><td>6.60mm</td><td>6.53mm</td><td>7.89mm</td></tr>
</table>

TABLE II. COMPLEX PATTERN RESULTS

<table>
<tr><th>Affect Factors</th><th colspan="5">Complex Pattern Image Marker</th></tr>
<tr><th rowspan="2">Numbers</th><th>One</th><th>Two</th><th>Three</th><th colspan="2">Four</th></tr>
<tr><td>4.03mm</td><td>3.24mm</td><td>3.11mm</td><td colspan="2">3.12mm</td></tr>
<tr><th rowspan="2">Distance</th><th>30cm</th><th>60cm</th><th>90cm</th><th colspan="2">120cm</th></tr>
<tr><td>4.03mm</td><td>4.24mm</td><td>4.53mm</td><td colspan="2">6.24mm</td></tr>
<tr><th rowspan="2">Angle</th><th>0°</th><th>15°</th><th>30°</th><th>45°</th><th>60°</th></tr>
<tr><td>4.03mm</td><td>4.13mm</td><td>4.12mm</td><td>4.54mm</td><td>5.12mm</td></tr>
</table>

TABLE III. MANUAL RESULTS

<table>
<tr><th>Affect Factors</th><th colspan="5">Manual Registration</th></tr>
<tr><th rowspan="2">Numbers</th><th>One</th><th>Two</th><th>Three</th><th colspan="2">Four</th></tr>
<tr><td>3.03mm</td><td>3.32mm</td><td>3.12mm</td><td colspan="2">3.22mm</td></tr>
<tr><th rowspan="2">Distance</th><th>30cm</th><th>60cm</th><th>90cm</th><th colspan="2">120cm</th></tr>
<tr><td>3.03mm</td><td>5.88mm</td><td>6.89mm</td><td colspan="2">9.89mm</td></tr>
<tr><th rowspan="2">Angle</th><th>0°</th><th>15°</th><th>30°</th><th>45°</th><th>60°</th></tr>
<tr><td>3.03mm</td><td>3.28mm</td><td>4.22mm</td><td>4.52mm</td><td>8.87mm</td></tr>
</table>

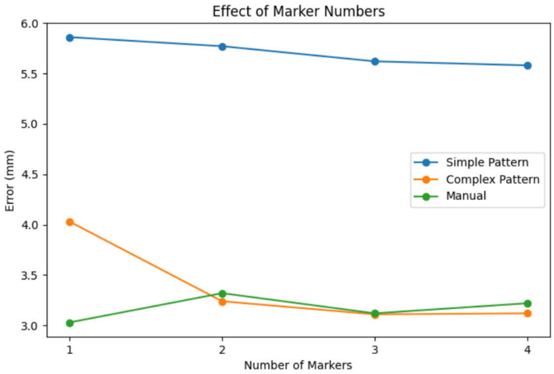


Fig. 4. Result of Number

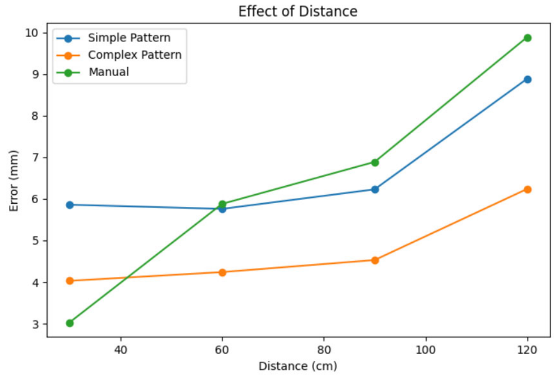


Fig. 5. Result of Distance

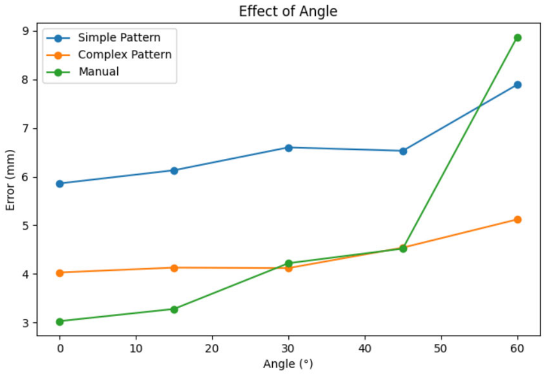


Fig. 6. Result of Angle

ACKNOWLEDGMENT

This research was supported by a grant of the Korean Health Technology R&D Project through the Korea Health Industry Development Institute(KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea(grant number : RS-2023-00302149, RS-2023-00302153)

REFERENCES

[1] M. Selvamanikkam, M. Noga, N. S. Khoo, K. Punithakumar, “High-Resolution Stereoscopic Visualization of Pediatric Echocardiography Data on Microsoft HoloLens 2” IEEE Access, vol. 12, pp. 9776-9783, 2024.

[2] S. Sirilak, P. Muneesawang, “A New Procedure for Advancing Telemedicine Using the HoloLens” IEEE Access vol. 6, pp. 60224-60233, 2018.

[3] Q. Sun, Y. Mai, R. Yang, T. Ji, X. Jiang, X. Chen, “Fast and accurate online calibration of optical see-through head-mounted display for AR-based surgical navigation using Microsoft HoloLens” International Journal of Computer Assisted Radiology and Surgery vol. 15. pp. 1907-1919. August. 2020.

[4] T. Jiang, D. Yu, Y. Wang, T. Zan, S. Wang, Q. Li, “HoloLens-Based Vascular Localization System: Precision Evaluation Study With a Three-Dimensional Printed Model” Journal of Medical Internet Research. vol. 22, no. 4, April. 202

Discriminative Correlation Filter-Based Object Tracking: Recent Research Trends and Challenges

Jeongwoo Lee
Department of Electronics and Information Engineering
Korea Aerospace University
Republic of Korea
jeongul1018@kau.kr

Pileun Kim
Department of Artificial Intelligent line
Korea Aerospace University
Republic of Korea
pkim@kau.ac.kr

Abstract—**Tracking in unmanned aerial vehicles (UAVs) is an essential technology in various applications such as defense and surveillance. Discriminative Correlation Filters (DCF)-based trackers have been extensively researched due to their high computational efficiency and real-time applications. Research in DCF-based trackers has evolved through various approaches, such as enhancing feature representations and improving the filter learning process. This paper compares and analyzes representative DCF-based trackers, examining how they have overcome the limitations of earlier models.**

Keywords—correlation filter, object tracking, spatial-temporal regularization, adaptive response, ADMM

I. INTRODUCTION

With the rapid increase in the utilization of unmanned aerial vehicles (UAVs), the importance of visual target tracking technology has grown significantly. One of the challenges in UAV tracking is the need to perform real-time tracking under the limited computational resources and battery life.

To meet these requirements, discriminative correlation filters (DCF)-based trackers have been extensively studied. DCF-based trackers facilitate tracking by detecting a target in the initial frame and then updating its position consistently across subsequent frames. However, DCF-based trackers suffer from various environments :

- the small size of objects
- the rapid movements and scale variations of both the UAV and the target
- sensor noise from limited resolution of UAV-mounted camera

This paper aims to analyze the research trends of DCF-based methods and discuss future directions for development.

.

II. REALATED WORK

To overcome these limitations, many studies have been conducted to improve the performance of DCF-based trackers. Previous research has generally focused on enhancing target features or refining the filter learning process. The evolution of these filters can be categorized into four main directions:

- Spatially-Temporally Regularized Filters
- Adaptive Filters
- Using Background and Response Information Filters
- 2nd-Order Difference Filters

Furthermore, to solve the complex objective functions that arise from these methods, the Alternating Direction Method of Multipliers (ADMM) is widely employed. This method improves convergence and computational efficiency by decomposing the complex objective functions into several sub-problems.

A. Feature Preprocessing and Enhancement

Recent studies have focused not only on filter design but also on enhancing feature representation. DCF-based models typically extract hand-crafted features like Histograms of Oriented Gradients (HOG) and Color Names (CN) to represent objects appearance. To mitigate issues arising from low-resolution and sensor noise in UAV applications, various denoising techniques such as Gaussian denoising and Non-Local Means (NLM) algorithm.

Also, to better distinguish the target from its background, the CACF[1] model employs the fuzzy C-means (FCMs) algorithm to classify feature channels into a target cluster and a background cluster. Meanwhile, to address scale variations, the MSEFCF[2] uses multi-scale enhanced features at different scales to create a more adaptive representation.

B. Baseline Model

The baseline for DCF tracking model is the Minimum Output Sum of Squared Error (MOSSE)[3] model. The MOSSE[3] optimizes the filter to minimizes the sum of squared error in the Fourier domain between the actual correlation output and ideal response. To update the filter, the objective function is minimized by finding the filter that sets its derivative to zero.

C. Spatially-Temporally Regularized Filters

The MOSSE[3] filter's high computational efficiency is derived from the use of fast Fourier Transform (FFT), which assumes that the image repeats circularity. This assumption can lead to the boundary effects. To address the boundary effects, the SRDCF[4] introduced a spatial regularization that penalizes filter coefficients in boundary regions to mitigate the boundary effect. In STRCF[5], a temporal regularization term is added to penalize abrupt differences between filters of adjacent frames, thereby enhancing stability in tracking

D. Adaptive Filters

A limitation of STRCF[5] using a fixed spatial weight across all feature channels may lead to performance degradation. To address this, the ASRCF[6] model introduces an adaptive spatial weight that adjusts according to the target's appearance changes in each frame. This approach mitigates abrupt filter variations. Furthermore, the RCFL[7] model incorporates temporal information by utilizing a continuously weighted dynamic response (CWDR) from several past frames. To enhance its adaptability, exponential smoothing is applied, assigning greater importance to more recent responses.

E. Using Background and Response Information Filters

Early DCF-based models that relied solely on target features had limited capability to discriminate the target from the background. The BiCF[8] model addresses this issue by introducing a bidirectional inconsistency error, defined as the difference between the forward and backward tracking errors. Ideally, appearance transitions across consecutive frames should remain consistent; however, factors such as illumination changes introduce discrepancies. By utilizing this error, the BiCF[8] model enhances robustness against appearance variations.

F. 2nd-Order Difference Filters

Traditional correlation filter models that rely on first-order information often exhibit relatively low discriminative capability. To address this limitation, several trackers have incorporated second-order difference terms. The CACF[1] model applies the second-order difference of features representations to help the model adapt more effectively to abrupt appearance changes. The MSEFCF[2] tracker incorporates the second-order difference on the filter, encouraging consistency with that of the previous frame. This allows the filter to update in accordance with the object's motion, ensuring smoother updates and stability. Moreover, the second-order difference is extended to multi-scale enhanced features. The LDECF[9] model advances this idea by exploiting three consecutive frames to compute a Dynamic Sensitivity Error, which measures the second-order difference in the response map. It also introduces a second-order difference term into the spatial regularization weights. These model suppresses abrupt variations induced by noise and reduces the risk of overfitting, ultimately improving both robustness and generalization.

G. Optimization Through ADMM

Most DCF-based models have an objective function in the form of a convex function. As models become more complex, the Alternating Direction Method of Multipliers (ADMM) is widely adopted as an effective optimization framework. The strategy of ADMM is to decompose a complex objective function into several simpler sub-problems by introducing auxiliary variables and enforcing equality constraints. These constraints are incorporated into the objective through an augmented Lagrangian form, which combines the original objective with a penalty term and Lagrange multipliers. Since each sub-problem is a convex function, the solution of each sub-problem is guaranteed to be unique optimal. The optimal solution is obtained by iteratively performing alternating optimization over these sub-problems. In each iteration, variables are updated sequentially to find the solution where the derivative of each sub-problem's objective function is zero.

Moreover, when convolution operations are involved, the sub-problems can be efficiently addressed in the Fourier domain using the FFT. Also, Leveraging the FFT enables rapid convergence of ADMM through parallel computations. Therefore, the application of ADMM enables the realization of real-time performance in UAV environments.

III. EXPERIMENTS

The UAV123_10fps dataset, which includes 123 video sequences with 12 attributes, was used to evaluate the tracker's performance. These attributes are Scale Variation (SV), Aspect Ratio Change (ARC), Low Resolution (LR), Fast Motion (FM), Full Occlusion (FOC), Partial Occlusion (POC), Out-of-View (OV), Background Clutter (BC), Illumination Variation (IV), Viewpoint Change (VC), Camera Motion

TABLE I. ATTRIBUTES BASED DCF BASED TRACKERS ON THE UAV123_10FPS DATASET.

Trackers	**ARC**	**BC**	**CM**	**FM**	**FOC**	**IV**	**LR**	**OV**	**POC**	**SOB**	**SV**	**VC**
SRDCF	0.472/0.387	0.389/0.333	0.527/0.476	0.427/0.354	0.418/0.270	0.436/0.383	0.431/0.275	0.492/0.432	0.504/0.423	0.585/0.509	0.531/0.465	0.474/0.398
STRCF	0.524/0.413	0.477/0.374	0.602/0.526	0.488/0.377	0.426/0.261	0.493/0.406	0.509/0.334	0.523/0.450	0.559/0.454	0.630/0.555	0.580/0.494	0.537/0.438
BiCF	0.578/0.466	0.515/0.427	0.625/0.555	0.484/0.376	0.469/0.305	0.581/0.494	0.534/0.371	0.533/0.481	0.591/0.502	0.705/0.609	0.619/0.534	0.581/0.496
CACF	0.579/0.461	0.522/0.378	0.595/0.529	0.416/0.352	0.434/0.271	0.560/0.445	0.519/0.368	0.518/0.455	0.570/0.479	0.692/0.585	0.608/0.521	0.574/0.467
RCFL	0.602/0.477	0.525/0.397	0.632/0.548	0.513/0.393	0.424/0.267	0.551/0.447	0.544/0.379	0.537/0.485	0.585/0.497	0.691/0.587	0.631/0.536	0.597/0.481
MSEFCF	0.565/0.444	0.446/0.336	0.607/0.541	0.499/0.384	0.414/0.279	0.528/0.411	0.521/0.373	0.526/0.486	0.565/0.483	0.638/0.565	0.613/0.524	0.595/0.476

LDECF	0.619/0.491	0.518/0.388	0.662/0.576	0.515/0.411	0.448/0.288	0.590/0.476	0.565/0.402	0.548/0.484	0.594/0.502	0.693/0.589	0.653/0.556	0.604/0.488

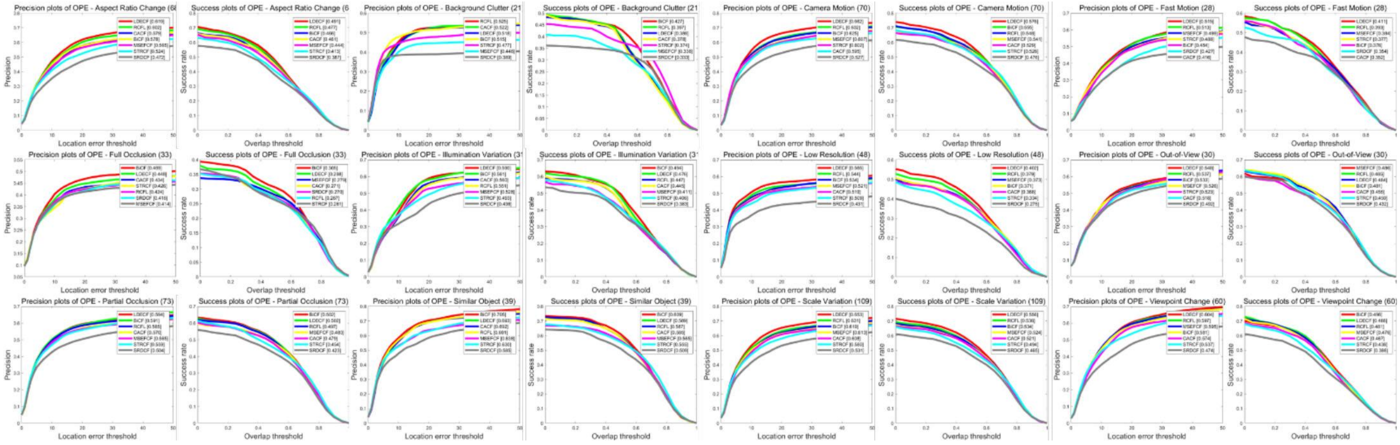


Fig. 1. Success plots and precision plot of the DCF-based trackers on the UAV123_10fps.

(CM), and Similar Object (SOB). As demonstrated in Table I, the progression of trackers indicates a consistent improvement in both precision and success rate across most evaluation attributes. However, occlusion (OC) scenarios remain the primary limitation for all trackers.

IV. CONCLUSION

DCF-based approaches are well-suited for UAV tracking because they can rapidly adapt to new scenes through online updates without requiring offline training process. Their reliance on the Fast Fourier Transform (FFT) also enables real-time operation even on limited hardware, only CPU. To mitigate the boundary effects and instability issues observed in the early MOSSE[3] model, spatial-temporal regularization was introduced. Subsequent advancements, including higher-order data fitting, multi-scale features, and response map weighting, have significantly enhanced their discriminative capacity and robustness. Furthermore, ADMM-based optimization has enhanced computational efficiency and stability.

Although recent works have achieved performance comparable to deep learning-based trackers, several challenges persist. Especially, severe occlusion scenario is still a major problem. In occlusion scenario, expanding the search area to find the object increases computational cost, and when the object reappears with a drastically changed appearance, previously learned information may no longer be useful. To resolve these issues, integrating a re-detection strategy or a long-term tracking module into the DCF framework could be an effective solution. Future research is emphasize robustness against full occlusion, abrupt rotations and illumination variations, which are common in UAV environments.

ACKNOWLEDGMENT

Following are results of a study on the "Convergence and Open Sharing System (NCCOSS)" Project, supported by the Ministry of Education and National Research Foundation of Korea.

REFERENCES

[1] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, “Visual object tracking using adaptive correlation filters,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2010, pp. 2544–2550

[2] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, “Learning spatially regularized correlation filters for visual tracking,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2015, pp. 4310–4318.

[3] F. Li, C. Tian, W. Zuo, L. Zhang, and M.-H. Yang, “Learning spatial-temporal regularized correlation filters for visual tracking,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 4904–4913.

[4] K. Dai, D. Wang, H. Lu, C. Sun, and J. Li, “Visual tracking via adaptive spatially-regularized correlation filters,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 4670–4679.

[5] F. Lin, C. Fu, Y. He, F. Guo, and Q. Tang, “BiCF: Learning bidirectional incongruity-aware correlation filter for efficient UAV object tracking,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), 2020, pp. 2365–2371

[6] Y. Zhang, Y.-F. Yu, K.-K. Huang, and Y. Wang, “Channel attentional correlation filters learning with second-order difference for UAV tracking,” IEEE Geosci. Remote Sens. Lett., vol. 20, no. 8001905, pp. 1–5, 2023

[7] Y.-F. Yu, C. Fu, and G. Lu, “Robust correlation filter learning with continuously weighted dynamic response for UAV visual tracking,” IEEE Trans. Intell. Transp. Syst., vol. 24, no. 11, pp. 11512–11524, Nov. 2023 .

[8] Y.-F. Yu, C. Fu, and G. Lu, “Multi-scale enhanced features correlation filters learning with dual second-order difference for UAV tracking,” IEEE Trans. Image Process., vol. 33, pp. 1435–1448, 2024 .

[9] Y.-F. Yu, C. Fu, and G. Lu, “Learning dynamic-sensitivity enhanced correlation filter with adaptive second-order difference spatial regularization for UAV tracking,” IEEE Trans. Image Process., vol. 34, pp. 1180–1194, 2025 .

Entity-Aware Trajectory Refinement : Next Best View Based 3D Gaussian Splatting

Namjun Park
Department of Electronics and Information Engineering
Korea Aerospace University
Republic of Korea
namjun2079@kau.kr

Pileun Kim
Department of Artificial Intelligent
Korea Aerospace University
Republic of Korea
pkim@kau.ac.kr

***Abstract*—With the rapid growth of drone applications, efficient 3D rendering has become increasingly important. This article focuses on entity-aware importance weighting to guide optimal view selection when photographing high-complex objects such as buildings. To reduce the required image set and improve reconstruction quality, we propose 3D Gaussian Splatting (3DGS) instead of conventional photogrammetry and additionally investigate how to maximize coverage at each drone waypoint. Images collected from an arbitrary flight trajectory were processed using segmentation and Next-Best-View (NBV) analysis to adaptively generate optimized waypoints for 3DGS. Through this process, 3DGS confirmed significant improvements in rendering quality compared to arbitrary trajectories. These results demonstrate that object-aware trajectory refinement enables more efficient and higher-quality reconstruction. Future work includes enabling real-time drone posture control with advanced UAV platforms.**

Keywords — Next-Best-View, 3D Gaussian Splatting, Building reconstruction, UAV photogrammetry, Object awareness

I. INTRODUCTION

Recently, periodic inspections of aircraft and bridges using drones have been conducted experimentally. Since most cracks appear microscopically, higher accuracy and completeness than conventional reconstruction methods are required. To achieve this, both the performance of the rendering tool and the fidelity of input image sets must be improved.

While recent research on 3D rendering has mainly focused on developing rendering tools for higher-quality visualization, this work emphasizes the fundamental step of capturing high-fidelity image sets to increase reconstruction efficiency. Traditional photogrammetry-based reconstruction requires many images and significant rendering time, which limits its practicality in real-world applications. As an alternative, 3D Gaussian Splatting (3DGS) enables efficient reconstruction with fewer images.

However, most commercial UAVs, such as DJI drones, typically follow pre-defined altitudes and simple waypoint trajectories, which fail to account for the diverse and complex structures of real-world objects. Moreover, no standardized user guideline currently exists for addressing such variations.

Other studies on UAV trajectory planning [3], [4] focus primarily on very large objects such as the Eiffel Tower and the Sydney Opera House and demonstrated that dividing altitude layers can be effective.

To address these limitations, this study proposes a variation in way pointing based on object segmentation, starting from an arbitrary trajectory, to adaptively refine flight paths for complex structures.

II. METHODOLOGY

Algorithm 1 Object-aware Waypoints Refinement and 3DGS

```
Require: Images $I_1, \dots, I_n$; poses $P_1, \dots, P_n$
Ensure: Refined waypoints $W^*$; 3DGS results $R_{\text{arb}}, R_{\text{ref}}$
for $k = 1$ to $n$ do
    $S_k \leftarrow Segment(I_k)$
    $C_k \leftarrow CornerDetection(S_k)$
    $U_k \leftarrow EulerRotation(P_k)$
    $q_k \leftarrow NBVScore(C_k, U_k)$
    $W^* \leftarrow RefineWaypoints(U_k, C_k, q_k)$
end for
$I^* \leftarrow Capture(W^*)$                                ▷ 2nd flight
$R_{\text{arb}} \leftarrow 3DGS(I_1, \dots, I_n)$
$R_{\text{ref}} \leftarrow 3DGS(I_1^*, \dots, I_n^*)$
Compare $(R_{\text{arb}}, R_{\text{ref}})$
```

A. Image Segmentation

Due to complexity of many objects, motion trajectory should be different regarding the object. And there are no such simple mathematical equations to express the sophisticated buildings. So, firstly we should detect and simplify what is ahead of a drone. So, we proposed instance Segmentation using YOLOv8. Classification and segmentation are well fitted, especially in JAIST circumstances. Total 533 images, 20 valid, 21 test images are used.

B. Euler Angle Rotation & Drone Body Coordinate Calibration

Using Euler angle rotation, rotation from 2D image pixel to Drone coordinate system (roll,pitch,yaw). Correcting the drone body coordinates. Because there are slight differences between drone's center coordinate and camera coordinates.

For Drone body calibration, if the values of yaw, pitch, and roll were less than 30 degree each, they were ignored, and if between 30 and 45 degrees, they were corrected by subtracting -10 degree.

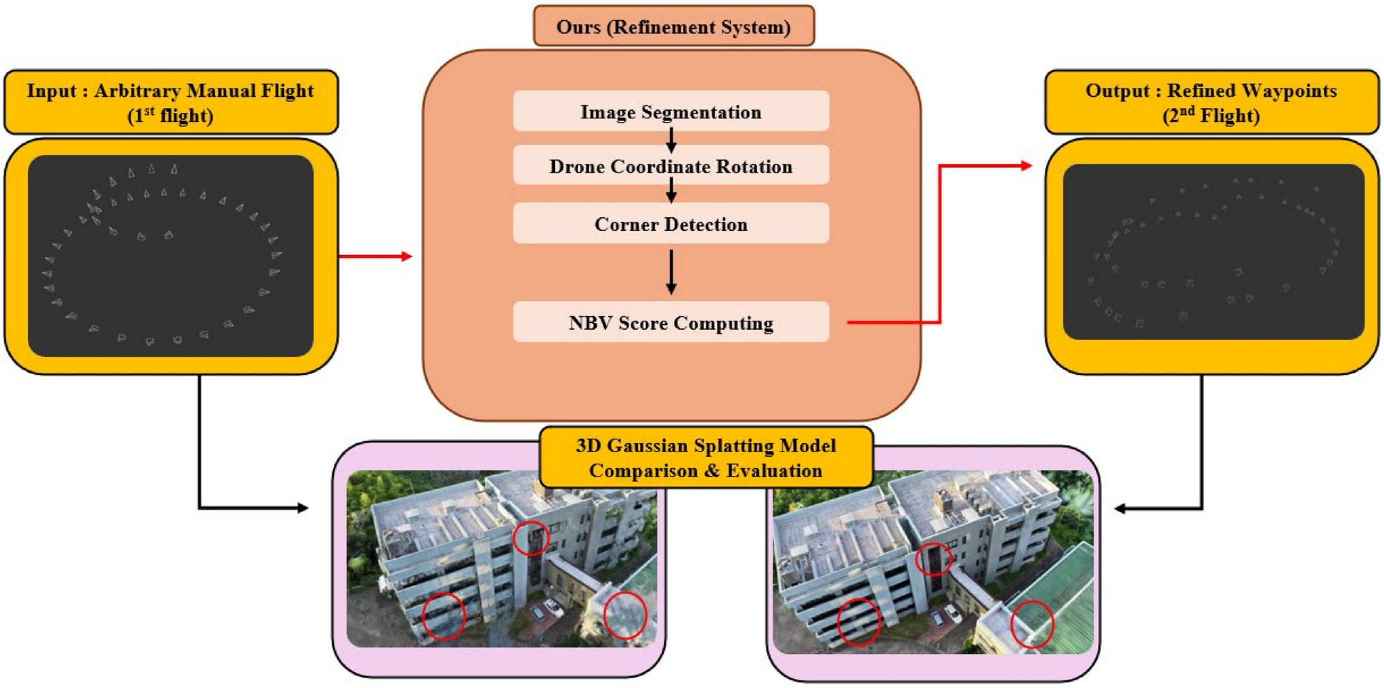


Fig.1. **Overall Workflow**

C. Corner Detection

In order to express the situation more simply, it was assumed that the building was reconstructed by rotating only “clockwise”. When the corner of the building is detected in the image taken by the drone, the yaw value is positive(+) and the roll value is hard coded as a negative number(-), smoothly surrounding the building.

We detect corners only within the segmentation masks of structural classes (e.g. building facades). A multi-scale Harris/Shi–Tomasi detector is applied after light smoothing; To avoid oversampling repetitive textures, we enforce a minimum spacing and cap the number of corners per ROI. Each retained corner is associated with the current pose to obtain a bearing, and the set of bearings C_k is passed to the NBV module as geometric evidence of High-fidelity views.

D. NBV Score Computation

For getting the next better view based on the current state, inspired by potential-field methods, we treat each YOLO-detected bounding box as a region of interest (ROI) that exerts an attractive force proportional to its semantic importance.

For building-complexity levels w{LV1–LV5}, we assigned normalized weights (LV5 is most complex).

$$w \in \{0.15, 0.25, 0.45, 0.75, 1.00\}$$

For a candidate view, the NBV score combines semantic weight, coverage, visibility, and distance attenuation: it was intended to indicate the vector direction and the induction in the sum direction regarding the given score.

III. EXPERIMENTS AND RESULTS

We compare two 3DGS reconstructions under a two-flight protocol: (I) a baseline capture along a randomly selected (arbitrary) route and (II) a revisited capture along the refined route predicted by our model using the images from (I). To ensure fairness, we fix the image budget to 40 frames for both conditions. Experiments are conducted on the most complex JAIST building, using a DJI Mavic 3 Pro at an altitude of 35–40 m under clear weather.

With this, drone users can build a trajectory which fits in their target objects by ideally modifying the trajectory by taking a random picture and then putting it in this model without thinking about how to take a waypoint even in a complex object such as a building.

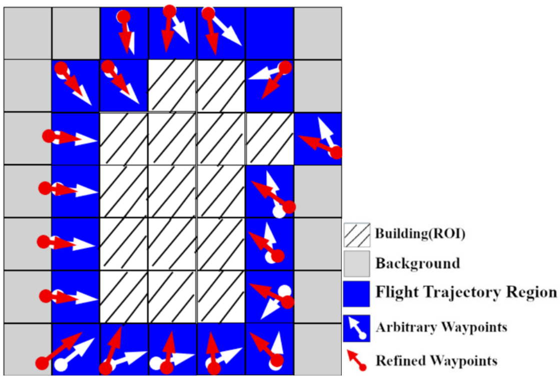


Fig.2. **Refinement FOV at each-Waypoint**

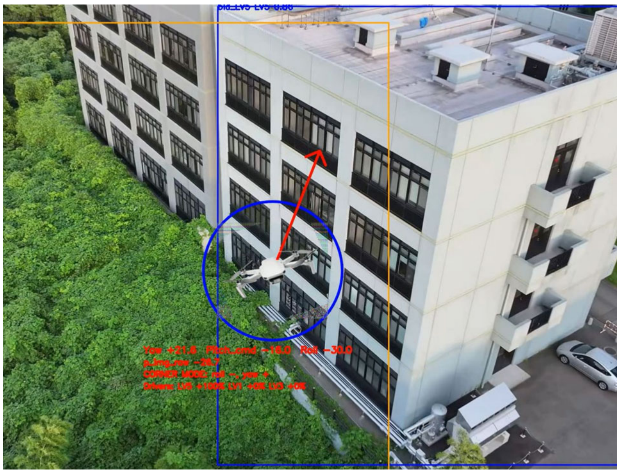


Fig.3. **Modifying drone orientation by our model**

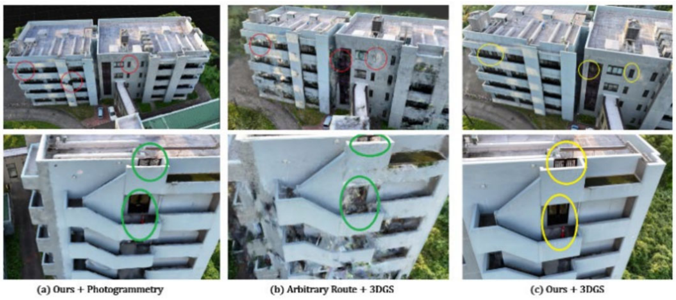

(a) Ours + Photogrammetry (b) Arbitrary Route + 3DGS (c) Ours + 3DGS

Fig.4. **Building Rendering Comparison**

	(a) Ours + Photogrammetry	(b) Arbitrary + 3DGS	**(c) Ours + 3DGS**
PSNR↑	9.23	10.67	**11.33**
SSIM↑	0.2564	0.2609	**0.2941**
LPIPS↓	0.5292	0.4734	**0.4565**

Table. I. **Evaluation for 3DGS**

As shown in the qualitative model in Fig.4. and the quantitative summary in Table.I.,"Ours+3DGS" achieves clearly higher coverage than the baselines. Fine structures—such as the iron door, the traffic cone near the aisle, facade textures, and the lateral wall—are more sharply reconstructed.

Although the absolute scores are modest since the value itself is a comparison based on the captured image of the 3D model, the relative gains are consistent: coverage +6.2%(PSNR), fidelity +13%(SSIM) (e.g., PSNR, SSIM, higher is better), and perceptual error −15.9% (e.g., LPIPS, lower is better).

IV. CONCLUSION

In this paper, we presented refining UAV waypoints to improve 3D reconstruction quality with 3D Gaussian Splatting. The current model is limited to offline processing of video or image data, but future work will extend this into a real-time route refinement protocol. Furthermore, this study expands the discussion on how flight trajectories can adapt to object complexity, opening possibilities for adjusting the number of images according to semantic importance. We expect that, as more rendering datasets with various complexities are accumulated, adaptive strategies for image acquisition can be further optimized. Ultimately, this approach shows practical potential in infrastructure for periodic aircraft and bridge inspections.

V. ACKNOWLEDGMENTS

Following are results of a study on the "Convergence and Open Sharing System (NCCOSS)" Project, supported by the Ministry of Education and National Research Foundation of Korea.

REFERENCES

[1] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," ACM Transactions on Graphics (SIGGRAPH), vol. 42, no. 4, Aug 2023.

[2] A. Bircher, M. Kamel, K. Alexis, H. Oleynikova, and R. Siegwart, "Receding Horizon 'Next-Best-View' Planner for 3D Exploration," in Proc. IEEE Int. Conf. on Robotics and Automation (ICRA), May 2016.

[3] Ultralytics, "Ultralytics YOLO (YOLOv8)," Zenodo (software release), version v8.3.120, Apr. 2025.

[4] Y. Ham, M. Michalkiewicz, and G. Balakrishnan, "DRAGON: Drone and Ground Gaussian Splatting for 3D Building Reconstruction," in 2024 IEEE International Conference on Computational Photography (ICCP), 2024, pp. 1–12.

[5] Cheng Peng, Volkan Isler, "Adaptive View Planning for Aerial 3D Reconstruction" in Proc. IEEE Int. Conf. on Robotics and Automation (ICRA), Montreal, Canada, 2019, pp. 2981–2987

[6] Zhu, S., Wang, G., Kong, X., Kong, D., and Wang, H., "3D Gaussian Splatting in Robotics: A Survey," arXiv preprint arXiv:2410.12262, 2024.

[7] Karakontis, K., Petsanis, T., Kapoutsis, A. Ch., Kapoutsis, P. Ch., and Kosmatopoulos, E. B., "Terrain-Aware Adaptation for Two-Dimensional UAV Path Planners," in Proc. 33rd Mediterranean Conf. on Control and Automation (MED), Tangier, Morocco, 2025, pp. 636–641

[8] Bisson-Larrivée, A., and Le Moine, J.-B., "Photogrammetry and the Impact of Camera Placement and Angular Intervals Between Images on Model Reconstruction," Digital Applications in Archaeology and Cultural Heritage, 2022, Art. no. e00224.

Investigating the Accuracy-Latency Trade-Off for Real-Time Open-Vocabulary Models on an Edge Device

Jongyoon Park
Department of Artificial Intelligence
Korea Aerospace University
Goyang, Republic of Korea
jongyoon0621@kau.kr

Daeil Ko
Department of Artificial Intelligence
Korea Aerospace University
Goyang, Republic of Korea
dikokau.kr@kau.kr

Pileun Kim
Department of Autonomous Vehicle Engineering
Korea Aerospace University
Goyang, Republic of Korea
pkim@kau.ac.kr

***Abstract*— This study systematically analyzes the trade-offs involved in optimizing real-time robotic perception pipelines on a resource-constrained edge device, the NVIDIA Jetson AGX Orin 64GB platform. We measured the performance of pipelines by evaluating combinations of two open-vocabulary detection models (NanoOWL and YOLO-World) and two prompt-based segmentation models (NanoSAM and EfficientViT-SAM). Each pipeline was optimized using NVIDIA TensorRT across various precision levels (FP32, FP16, and Best). We present a quantitative comparison of the latency for each configuration, providing practical insights and benchmarks for researchers and developers deploying these advanced models on embedded systems.**

Keywords—Edge Device, Zero-Shot, Real-time, Optimization, Human-Robot Interaction

I. INTRODUCTION

As the paradigm of robotics expands beyond highly controlled industrial environments into the daily lives of humans, a key challenge emerges: enabling non-expert users to communicate and collaborate with robots intuitively. Vision-Language Models (VLMs), pre-trained on massive datasets, have emerged as a powerful solution to this language-vision grounding problem [9]. VLMs offer open-vocabulary recognition, allowing robots to identify objects described by arbitrary text in a zero-shot manner [7]. However, the immense computational cost of VLMs directly conflicts with the inherent resource constraints of edge devices embedded in real-world robotic systems. Ensuring a response time of at least 10 FPS, a minimum requirement for seamless human interaction, presents a significant optimization challenge for deploying VLMs in edge environments.

This study presents a systematic approach to bridging this critical "real-time gap." To this end, we conducted a comprehensive benchmarking study of four distinct open-vocabulary instance segmentation pipelines on the NVIDIA Jetson AGX Orin platform. Our goal is not merely to accelerate a single model but to explore the multi-dimensional design space of model architectures, optimization techniques, and precision levels to find practical solutions that enable flexible, real-time human-robot interaction on resource-constrained platforms.

II. RELATED WORK

A. Open-Vocabulary Object Detection

Open-vocabulary object detection aims to detect and classify objects corresponding to arbitrary natural language text queries, moving beyond a restricted set of predefined classes [4]. This paper investigates two representative approaches:

- **NanoOWL:** An optimization framework that uses NVIDIA TensorRT to convert the powerful but computationally heavy OWL-ViT model into a highly efficient inference engine [8].
- **YOLO-World:** This model integrates open-vocabulary capabilities into the highly efficient YOLOv8 CNN framework [3, 10]. Its 'prompt-then-detect' paradigm eliminates the need for a text encoder during inference, significantly reducing latency.

B. Prompt-Based Image Segmentation

Prompt-based segmentation is the task of generating precise masks for specific objects within an image [5]. This study evaluates two lightweight models:

- **NanoSAM**: This model uses knowledge distillation to enable a lightweight ResNet-18 image encoder to mimic the output of a larger 'teacher' model [12].
- **EfficientViT-SAM**: This model is an efficient transformer architecture that uses multi-scale linear attention to reduce the complexity of the attention mechanism from quadratic to linear [2].

C. Model Optimization with NVIDIA TensorRT

NVIDIA TensorRT is a high-performance optimization and runtime library for deep learning inference. A key optimization strategy is low-precision inference, which reduces memory usage and leverages specialized hardware like Tensor Cores to maximize inference throughput, though it may carry some risk of performance degradation.

III. METHODOLOGY

This study conducted a comprehensive benchmarking investigation to systematically analyze the performance of a real-time robotic perception pipeline on a resource-constrained edge device, the NVIDIA Jetson AGX Orin 64GB platform. The perception pipeline framework used in this study follows a two-stage 'detect-then-segment' architecture, which sequentially performs object detection [3, 8] and segmentation [2, 12]. For this, we evaluated a total of four pipelines by combining two open-vocabulary detection models (NanoOWL and YOLO-World) with two prompt-

---

This paper was supported by the Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (No.RS-2021-KI002499, HRD Program for Industrial Innovation)

based segmentation models (NanoSAM and EfficientViT-SAM). Each pipeline was optimized using NVIDIA TensorRT at various precision levels (FP32, FP16, and Best), and performance was measured in FPS (frames per second). This research aimed to identify the decisive impact of model architecture and hardware-specific optimization on real-time performance.

IV. RESULTS

A. *Overall Performances Component*

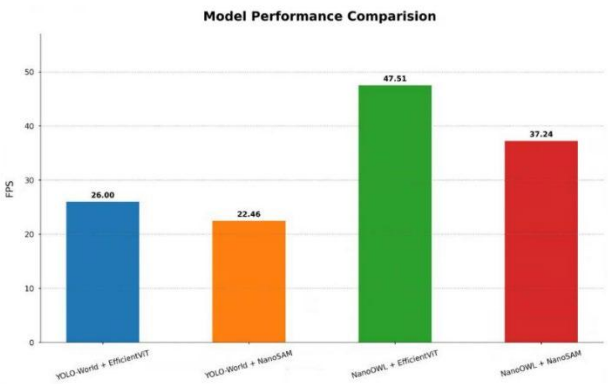


Fig. 1. end-to-end performance of the four primary pipeline architectures

Fig. 1 shows the end-to-end performance of the four primary pipeline architectures. Pipelines based on the NanoOWL detector exhibited a significant speed advantage over their YOLO-World counterparts. The highest throughput was achieved by the NanoOWL + EfficientViT-SAM combination, reaching 47.51 FPS. This suggests that the end-to-end latency is predominantly determined by the choice of the detection model, and converting models into TensorRT engines has a decisive impact on performance.

B. *Component-Level Performance Analysis*

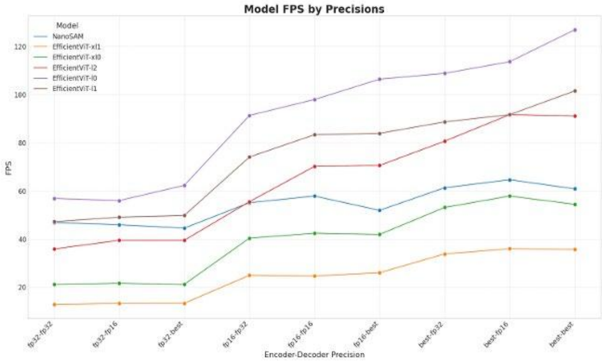


Fig. 2. Segmentation Model Performance Across Precision Levels

The differences in end-to-end pipeline performance originate from the individual characteristics of each component. This section independently evaluates the performance of the two core components that determine overall pipeline performance: the open-vocabulary object detector and the prompt-based segmenter.

Fig. 2 shows the FPS of each segmentation model based on its encoder and decoder precision. The graph clearly shows that using the 'best' mode is the fastest. The fastest segmentation model, EfficientViT-SAM-l0, showed performance exceeding 120 FPS, whereas EfficientViT-SAM-xl1 recorded the lowest FPS. NanoSAM was positioned in the middle.

V. CONCLUSION AND FUTURE WORK

A. *Conclusion*

This study presents a comprehensive performance benchmark of open-vocabulary perception pipelines on the NVIDIA Jetson AGX Orin, confirming the critical importance of hardware optimization. The NanoOWL-based pipeline, optimized with TensorRT, achieved a significantly faster processing speed than the YOLO-World-based pipeline. This demonstrates that model architecture and hardware-specific optimization play a decisive role in achieving real-time robotic perception performance on edge devices. Our research provides practical guidance for deploying Vision-Language Models (VLMs) on resource-constrained platforms.

B. *Future Work*

This study has the limitation that the evaluation was conducted on a static image dataset and did not consider the temporal complexity of real-world video streams. Furthermore, it focused solely on the perception module without integrating it into a full robotic control stack. Future research will focus on improving the performance of linguistically expressive models like YOLO-World to achieve real-time performance in edge environments. This will be a crucial research direction for developing a dominant solution that possesses both speed and linguistic understanding.

REFERENCES

[1] Brohan, A., Brown, N., Carbajal, J., Chebotar, Y., Chen, X., Choromanski, K., et al. (2023) "RT-2: vision-Language-Action models transfer web knowledge to robotic control," in Proceedings of the 7th Conference on Robot Learning (CoRL 2023).

[2] Cai, Y., Gan, Z., and Liu, Z. (2022). EfficientViT: Multi-Scale Linear Attention for High-Resolution Dense Prediction. arXiv preprint arXiv:2205.14756.

[3] Cheng, T., Zhang, X., Zhang, H., Lin, S., Ge, Y., Tong, Z., and Shan, Y. (2024) "YOLO-World: Real-Time open-vocabulary object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR 2024).

[4] Gu, X., Lin, T., Kuo, W., and Cui, Y. (2021). Open-Vocabulary Object Detection Via Vision and Language Knowledge Distillation. arXiv preprint arXiv: 2104.13921.

[5] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., et al. (2023) "Segment anything," in Proceedings of the IEEE/CVF international conference on computer vision (ICCV 2023).

[6] Lee, S., Nguyen, H. H., Beck, F., Vu, M. N., Ebmer, G., Nguyen, A., et al. (2024) Language-Driven closed-loop grasping with model-predictive trajectory replanning. arXiv Prepr. arXiv:2406.09039. doi: 10.48550/arXiv.2406.09039.

[7] Liu, Y., Li, K., Li, Y., and Wang, L. (2023). Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection. arXiv preprint arXiv:2303.05499.

[8] Minderer, M., Gritsenko, A., Stone, A., Neumann, M., Weissenborn, D., Dosovitskiy, A., et al. (2022) "Simple open-vocabulary object detection with vision transformers," in European conference on Computer Vision (ECCV 2022).

[9] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., et al. (2021) "Learning transferable visual models from natural Language supervision," in Proceedings of the 38th international conference on machine learning (ICML 2021).

[10] Yaseen, M. (2024). What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector. arXiv Prepr. arXiv:2408.15857. doi: 10.48550/arXiv.2408.15857

[11] Yu, L., Poirson, P., Yang, S., Berg, A. C., and Berg, T. L. (2016) "Modeling context in referring expressions," in European Conference on computer vision (ECCV 2016).

[12] Zhang, C., Han, D., Qiao, Y., Kim, J. U., Bae, S.-H., and Lee, S. (2023). MobileSAM: a fast segment anything model for Mobile devices. arXiv Prepr. arXiv:2306.14289. doi: 10.48550/arXiv.2306.14289.

Robust Path Navigation via Reinforcement Learning

Daeyeol Kang
Artificial Intelligence
Korea Aerospace University
Goyang, Gyeongido
kangdy1997@gmail.com

Jongyoon Park
Artificial Intelligence
Korea Aerospace University
Goyang, Gyeongido
jongyoonpark0621@gmail.com

Pileun Kim
Artificial Intelligence
Korea Aerospace University
Goyang, Gyeongido
pkim@kau.ac.kr

***Abstract*—Deep Reinforcement Learning (DRL) for mobile robot autonomous navigation confronts several inherent limitations that hinder its real-world applicability. The stochastic nature of actions generated by DRL policies can undermine performance consistency, while inefficient exploration often delays the learning process or prevents the discovery of an optimal solution. Furthermore, high-dimensional sensor inputs, such as raw LiDAR point clouds, lead to the "curse of dimensionality," making it computationally infeasible for an agent to learn effectively. This research aims to enhance the robustness of path planning by addressing these challenges. To achieve this, a hybrid approach is proposed, integrating the flexible decision-making capabilities of DRL with the stability of traditional path planning. The proposed model adopts the Twin Delayed Deep Deterministic Policy Gradient (TD3) network as its base, incorporates a low-dimensional state representation from LiDAR data, applies a reward-driven prioritized experience replay (PER), and integrates a classical path planner to guide the learning process, thereby enhancing both stability and efficiency. Experimental results in simulation environments demonstrate that the proposed model achieves faster convergence and higher mission success rates, confirming a significant improvement in path planning robustness.**

Keywords—Reinforcement Learning, Robot Navigation, Path Planning, TD3, Artificial Potential Field

I. INTRODUCTION

Deep Reinforcement Learning (DRL) holds immense promise for autonomous robot navigation, offering a framework where an agent learns an optimal policy through direct interaction with its environment. However, the practical application of DRL is fraught with challenges. The stochastic nature of DRL policies can lead to inconsistent performance [1] and high-dimensional sensor data like 3D LiDAR point clouds often trigger the "curse of dimensionality," making effective learning computationally intractable [2].

While traditional path planners like A* guarantee optimal paths, they can be computationally expensive in dynamic environments. Simpler methods like the Artificial Potential Field (APF) are fast but notoriously suffer from local minima issues. This research confronts these challenges by proposing a hybrid approach that synergistically combines the flexible, learned decision-making of DRL with the stability of traditional path-planning algorithms. The goal is to create a framework where the strengths of each method compensate for the weaknesses of the other, ultimately enhancing the overall robustness of the navigation system.

II. PROPOSED METHODOLOGY

Our model is built upon the Twin Delayed Deep Deterministic Policy Gradient (TD3) network [3] and integrates three key strategies to enhance learning stability and efficiency. The overall architecture is depicted in Fig. 1

A. Dimensionality Reduction for LiDAR State Representation

To tackle the curse of dimensionality, we preprocess the 3D LiDAR point cloud into a low-dimensional state vector. The 2D plane in front of the robot is partitioned into 20 uniform angular sectors. From each sector, only the single closest point is selected. This method dramatically reduces dimensionality to just 20 data points while preserving the most critical semantic information needed for obstacle avoidance. Compared to methods like increasing voxel size, which can lose crucial details about nearby obstacles, our angular segmentation approach proves far more effective for robust navigation.

B. Reward-driven Prioritized Experience Replay (PER)

Standard Prioritized Experience Replay (PER) [5] uses the TD-error to gauge an experience's importance, but this can be unreliable in early training stages when the Q-function is inaccurate. A high TD-error might reflect a genuinely surprising event or simply be noise from an untrained network. To overcome this, we redefine the prioritization metric to be the reward value itself. Experiences that result in **low rewards**, such as collisions, are assigned a higher sampling priority. This is an unambiguous signal of a critical event and forces the agent to learn intensively from its failures, significantly boosting learning efficiency. The sampling probability $P(i)$ is defined as:

$$P(i) = \frac{(-r_i+|min(r)|)^{\alpha}}{\sum_k(-r_k+|min(r)|)^{\alpha}+\epsilon} \tag{1}$$

where r_i is the reward of experience i, and α is a hyperparameter controlling the prioritization level.

C. Hybrid Control with Path Planning (TD3 + APF)

To mitigate the instability from DRL's stochastic exploration, we integrate the Artificial Potential Field (APF) algorithm [4] into the learning loop. APF generates a stable path by calculating an attractive force towards the goal and repulsive forces from obstacles. A key weakness of APF, however, is its tendency to get trapped in local minima (e.g., in U-shaped obstacles). Our framework leverages the DRL agent's exploratory nature to escape these traps.

The waypoint generated by APF is converted into a "guiding action"($a_{planner}$). During the training update, the model calculates target Q-values for both the actor's action (a_{actor}) and the planner's action ($a_{planner}$). The final target Q-value is the maximum of these two, guiding the policy toward a stable trajectory without stifling exploration

$$Q_{target} = max\begin{cases} min_{j=1,2}(Q'_{\phi j}(s', a_{actor}) \\ min_{j=1,2}(Q'_{\phi j}(s', a_{planner}) \end{cases} \tag{2}$$

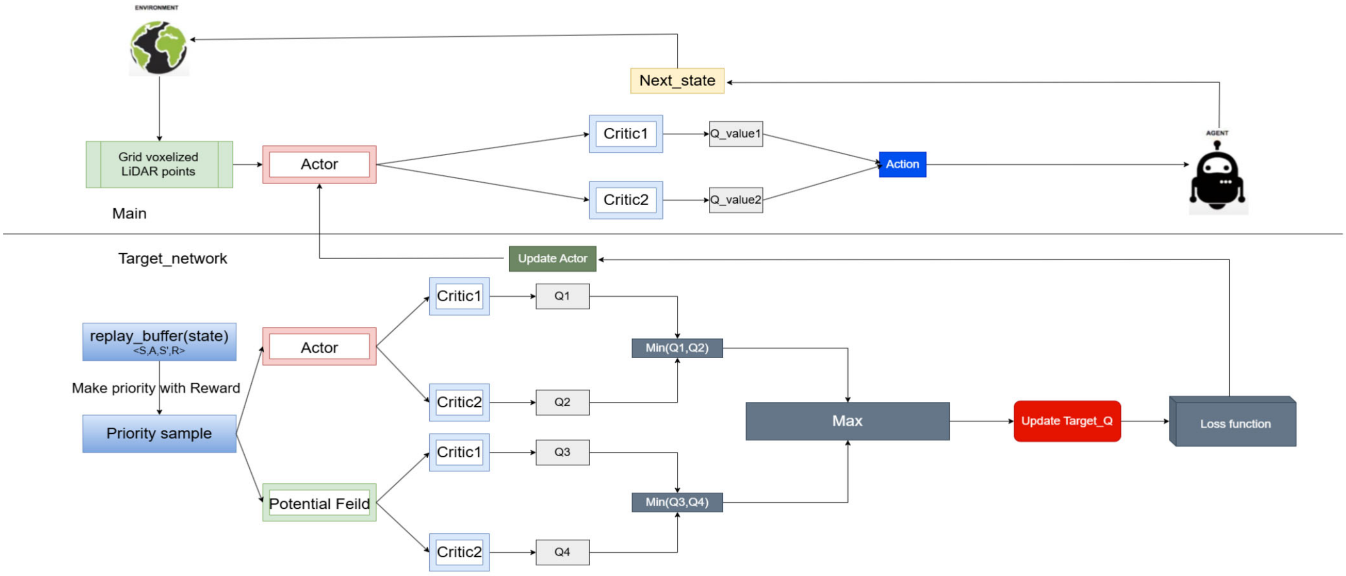


Fig. 1. The proposed Network Architecture

When the APF action leads to stagnation (low Q-value), the actor's exploratory actions can discover a way out, which is then reinforced through the learning process.

III. EXPERIMENTS AND RESULTS

A. Experimental Setup

The proposed model's efficacy was rigorously tested in NVIDIA Isaac Sim. All experiments were conducted on a system with an NVIDIA RTX 3060 GPU. The mobile robot platform was a Tracer Mini base with a 3D LiDAR. To ensure generalization, goal points were randomized for each episode. Performance was evaluated using average and maximum rewards.

B. Isaac Sim Simulation Results

As seen in Table 1, the fully integrated model (Grid+PER+PP) achieved a high average reward of 705.1, consistently succeeding in its navigation tasks. This stands in stark contrast to the baseline Grid model, which failed in the vast majority of episodes, resulting in a negative average reward of -35.22. These results unequivocally demonstrate that while each proposed component is effective individually, their synergy creates the most robust and high-performing framework.

TABLE I. REWARD COMPARISON IN ISAAC SIM ENVIRONMENT

Model (State: Grid)	Avg reward	Max reward
TD3	-35.22	973.6
TD3+PER+PP(ours)	**705.1**	**1078**

IV. CONCLUSION

This research proposed a hybrid approach to enhance the stability and efficiency of DRL-based mobile robot navigation. By integrating a low-dimensional state representation, a reward-driven PER mechanism, and a traditional path planner (APF) into the TD3 algorithm, we significantly improved the learning process's robustness.

Experiments validated that these components create a powerful synergy, leading to superior mission success rates and learning stability compared to baseline models. The most significant contribution is a novel control hierarchy where the RL agent is the primary controller and the path planner serves as a guiding supplement. This paradigm effectively addresses chronic DRL challenges and establishes a solid technical foundation for future work, including extending this framework to more complex, dynamic environments and focusing on sim-to-real transfer to physical robotic systems.

ACKNOWLEDGMENT

This work was supported by the Institute for Information and Communications Technology Promotion (IITP) grant funded by the Korean government (MSIT) (ICT Consilience Master & Doctoral Program for Cultivating Core Talent) (IITP-2024-RS-2024-00437857).

REFERENCES

[1] S. Jang and H. I. Kim, "Entropy-aware model initialization for effective exploration in deep reinforcement learning," Sensors, vol. 22, no. 15, p. 5845, 2022.
[2] R. Bellman, R. Kalaba, and D. Middleton, "Dynamic programming, sequential estimation and sequential detection processes," Proceedings of the National Academy of Sciences, vol. 47, no. 3, pp. 338-341, 1961.
[3] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing Function Approximation Error in Actor-Critic Methods," in International Conference on Machine Learning (ICML), 2018, pp. 1587-1596.
[4] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," in Autonomous robot vehicles, Springer, 1986, pp. 396-404.
[5] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," arXiv preprint arXiv:1511.05

Session D3: KAU-BK21 Special Session

Performance Enhancement of α-Ga_2O_3 MSM X-ray Detectors via Asymmetric Pt/Ti Electrode

Hyeongju Cha
Department of Materials Science and Engineering, Department of Smart Air Mobility
Korea Aerospace University
Goyang, 10540, Republic of Korea.
kauhyeongju@gmail.com

Sunjae Kim
Department of Materials Science and Engineering
Korea Aerospace University
Goyang, 10540, Republic of Korea
kau.sjkim@gmail.com

Ji-Hyeon Park
Korea Institute of Ceramic Engineering and Technology
Jinju, 52851, Republic of Korea
jhp5511@kicet.re.kr

Dae-Woo Jeon
Korea Institute of Ceramic Engineering and Technology
Jinju, 52851, Republic of Korea
dwjeon@kicet.re.kr

Wan Sik Hwang
Department of Materials Science and Engineering, Department of Smart Air Mobility
Korea Aerospace University
Goyang, 10540, Republic of Korea.
wansikhwang@gmail.com

***Abstract*— α-Ga_2O_3 has attracted considerable attention as a next-generation X-ray detection material due to its ultra-wide bandgap, high breakdown field, and excellent radiation tolerance. However, most reported α-Ga_2O_3-based metal–semiconductor–metal (MSM) photodetectors employ symmetric electrode structures, which limit charge separation efficiency and response speed. In this work, we demonstrate that asymmetric electrode design is highly effective in enhancing X-ray detection performance. High-quality α-Ga_2O_3 thin films were grown on c-plane sapphire substrates via hydride vapor phase epitaxy (HVPE), exhibiting excellent crystallinity and a smooth surface morphology with an RMS roughness below 5 nm. MSM devices with symmetric (Ti/Ti) and asymmetric (Pt/Ti) electrodes were fabricated and systematically evaluated under soft X-ray irradiation. The Pt/Ti device exhibited a markedly reduced dark current, an SNR improved by more than three orders of magnitude compared to the Ti/Ti counterpart and achieved a maximum sensitivity of ~99 μC/mGy·cm^2 at −10 V, along with a rapid decay time of ~0.1 s, thereby outperforming the symmetric device. Such device-level performance enhancements are of great significance for applications requiring real-time feedback in control and automation systems. In particular, the high SNR and fast response speed provide substantial advantages for systems and applications where real-time signal processing is critical.**

Keywords—Ga_2O_3 asymmetric MSM structure, X-ray detection, sensitivity, Signal-to-noise ratio, wide-bandgap semiconductors

I. INTRODUCTION

X-ray detectors play a critical role in various fields, including medical imaging, security screening, industrial nondestructive testing, and radiation dosimetry[1-4]. As the demand for real-time and low-dose diagnostics under high-temperature and high-radiation environments increases, the development of next-generation detectors that simultaneously achieve high sensitivity, low noise, fast response, and long-term stability has become essential. Conventional detectors suffer from limitations such as high leakage current and low sensitivity under low-dose conditions due to their narrow bandgaps. In contrast, α-Ga_2O_3, with its ultra-wide bandgap ($E_g \approx 5.3$ eV), low intrinsic carrier concentration, high breakdown field, and excellent radiation and thermal stability, has emerged as a promising material for next-generation X-ray photodetectors [5]. However, conventional symmetric MSM structures have inherent constraints in improving charge separation efficiency and response speed. In this work, we introduce an asymmetric electrode (Pt/Ti) configuration that generates a built-in electric field, thereby enabling efficient separation and collection of photogenerated carriers, suppression of leakage current, rapid response, and significantly enhanced sensitivity and signal-to-noise ratio. Based on these results, we propose a promising design strategy to advance the performance of α-Ga_2O_3-based X-ray detectors.

II. EXPERIMENTAL METHODS

A. Device Fabrication

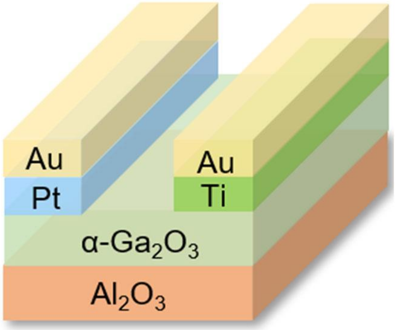


Figure 1. Schematic Image of Asymmetric Structure X-ray detector

The detectors were fabricated using HVPE-grown α-Ga_2O_3 epitaxial layers (thickness: 740nm) deposited on c-plane Al_2O_3 substrates. Ohmic contacts (Ti/Au, 20/100 nm) were annealed at 500 °C in a nitrogen (N2) atmosphere, while Schottky contacts (Pt/Au, 20/100 nm) were deposited via electron-beam evaporation and subsequently patterned into interdigitated electrodes using a conventional lift-off process, with finger widths and spacings of 4 μm and 16 μm.

B. X-ray Exposure Setup

Soft X-ray source operating at a tube current of 0.6 mA and a tube voltage of 5.0 kVp (dose rate: 3.84 mGy/s, AweXome Ray Inc.).

C. Measurement and Analysis

Electrical responses were measured using a Keithley 4200A-SCS system. The X-ray dose rate was 3.84 mGy/s at 2 cm from the source (5 kVp, measured with a Fluke 451B ion chamber). Photocurrent dynamics, sensitivity, SNR, and temporal response were characterized under varying bias conditions at room temperature.

III. RESULTS AND DISCUSSION

A. Sensitivity

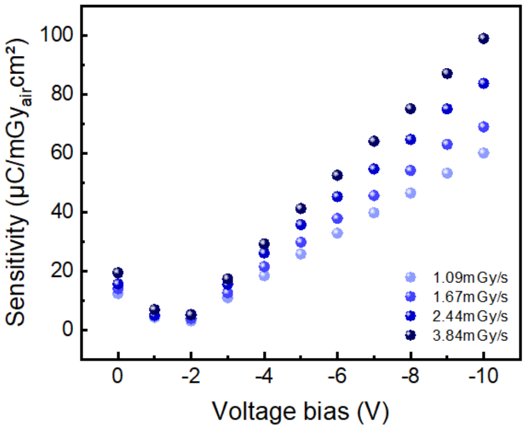


Figure 3. Sensitivity of Asymmetric Structure X-ray detector

For all dose rates, the sensitivity increases linearly with bias voltage, indicating efficient carrier collection and excellent response linearity. The highest sensitivity appears at the lowest dose rate (1.09 mGy/s) but decreases as the dose rate rises, mainly due to trap filling and field screening effects. At higher electric fields, additional mechanisms such as carrier multiplication, trap-assisted charge release, and self-trapped hole formation enhance sensitivity [6]. At −10 V bias, the device achieves a maximum sensitivity of ~99 μC/Gyair·cm². This performance is comparable to, or even surpasses, other oxide semiconductor–based X-ray detectors [7].

B. Signal-to-noise ratio

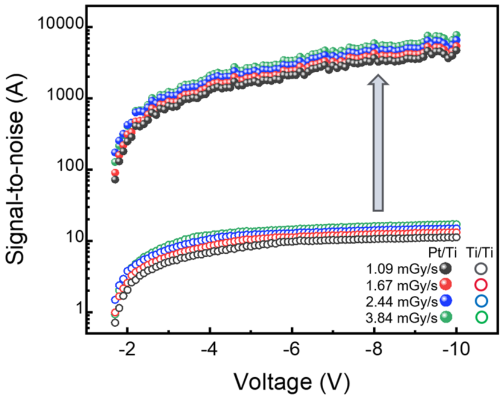


Figure 2. Comparison of Signal-to-Noise Ratio in Asymmetric and Symmetric X-ray Detector Structures

The Pt/Ti device exhibited an SNR more than three orders of magnitude higher than that of the Ti/Ti device, which is attributed to the suppression of dark current by the high Schottky barrier of Pt and the enhanced carrier separation induced by the built-in electric field of the asymmetric electrode structure [8]. As the bias increased, the Pt/Ti device showed a continuous monotonic rise in SNR, whereas the Ti/Ti device exhibited saturation at higher biases due to increased dark current associated with its lower barrier. Under a bias of −10 V, both devices demonstrated increasing SNR with dose rate; however, the Pt/Ti device reached values exceeding 7,000, while the Ti/Ti device remained below 2.0 even at the highest dose rate.

C. Decaytime

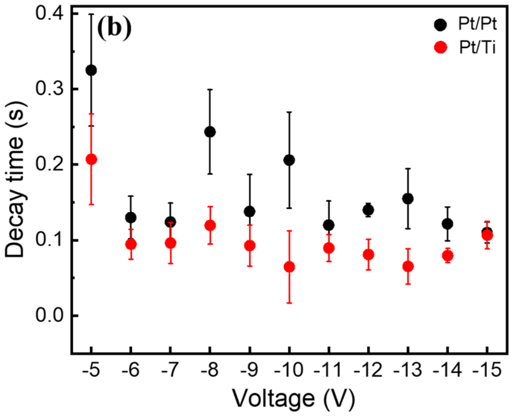


Figure 3. Decay Time Comparison of Asymmetric and Symmetric X-ray Detectors

In terms of decay time, the Pt/Ti device exhibited a fast response of approximately 0.1s, whereas the Ti/Ti device showed a relatively slower response of about 0.3s. This distinct performance difference is attributed to the high Schottky barrier of the Pt electrode, which effectively suppresses dark current, and the built-in electric field generated by the asymmetric electrode configuration, which facilitates efficient separation of photogenerated carriers.

IV. CONCLUSION

This study demonstrated that asymmetric electrode design is highly effective in enhancing the performance of α-Ga_2O_3 MSM X-ray detectors. The Pt/Ti device achieved a sensitivity of approximately 99 μC/mGy$_{air}$·cm^2, exhibited an improvement in the signal-to-noise ratio (SNR) by more than three orders of magnitude compared to the Ti/Ti device, and showed a shortened decay time of about 0.1 seconds, indicating a much faster response. These performance improvements are attributed to the high Schottky barrier of the Pt electrode, which effectively suppresses dark current, and the built-in electric field generated by the asymmetric electrode structure, which facilitates efficient separation and collection of photogenerated carriers. In particular, the high SNR and fast response speed achieved with the Pt/Ti asymmetric configuration offer significant advantages for control and automation systems where real-time signal processing is essential. Therefore, the results confirm that α-Ga_2O_3-based asymmetric electrode structures provide a promising technological foundation for next-generation X-ray detection, even under challenging conditions such as low-dose operation and high-speed imaging.

ACKNOWLEDGMENT

The author thank AweXome Ray Inc. for providing X-ray source equipment. This work was supported by the BK21 FOUR program through the National Research Foundation of Korea (NRF) funded by the Korean government (grant number: 5199990714521). This work was supported by the K-Sensor Development Program(no. RS-2022-00154729) funded by the Ministry of Trade, Industry and Energy (MOTIE, Republic of Korea).

REFERENCES

[1] B. Hou, Q. Chen, L. Yi, P. Sellin, H. T. Sun, L. J. Wong, and X. Liu, "Materials innovation and electrical engineering in X-ray detection," *Nat. Rev. Electr. Eng.*, vol. 1, no. 10, pp. 639–655, 2024.

[2] G. Harding, H. Strecker, D. Kosciesza, and J. Gordon, "Detector considerations relevant to X-ray diffraction imaging for security screening applications," in *Opt. Photon. Global Homeland Secur. V Biometr. Technol. Human Identif. VI*, vol. 7306, pp. 193–203, May 2009

[3] A. du Plessis, S. G. le Roux, and A. Guelpa, "Comparison of medical and industrial X-ray computed tomography for non-destructive testing," *Case Stud. Nondestruct. Test. Eval.*, vol. 6, pp. 17–25, 2016.

[4] T. Kron, L. Duggan, T. Smith, A. Rosenfeld, M. Butson, G. Kaplan, and K. Hyodo, "Dose response of various radiation detectors to synchrotron radiation," *Phys. Med. Biol.*, vol. 43, no. 11, pp. 3235–3247, 1998.

[5] H. Liang, S. Cui, R. Su, P. Guan, Y. He, L. Yang, and X. Du, "Flexible X-ray detectors based on amorphous Ga_2O_3 thin films," *ACS Photonics*, vol. 6, no. 2, pp. 351–359, 2018.

[6] F. Mady, M. Benabdesselam, J.-B. Duchez, Y. Mebrouk, and S. Girard, "Global view on dose-rate effects in silica-based fibers and devices damaged by radiation-induced carrier trapping," *IEEE Trans. Nucl. Sci.*, vol. 60, no. 6, pp. 4341–4348, 2013.

[7] M. Kim, S. Kim, J.-H. Park, H.-G. Cho, S.-H. Gihm, D.-W. Jeon, and W.-S. Hwang, "Heteroepitaxial α-Ga_2O_3 thin-film-based X-ray detector with metal–semiconductor–metal structure," *Phys. Status Solidi RRL – Rapid Res. Lett.*, vol. 19, no. 1, 2400193, 2025.

[8] D. Zhang, W. Zheng, R. Lin, Y. Li, and F. Huang, "Ultrahigh EQE (15%) solar-blind UV photovoltaic detector with organic–inorganic heterojunction via dual built-in-fields-enhanced photogenerated carrier-separation efficiency mechanism," *Adv. Funct. Mater.*, vol. 29, no. 26, 1900935, 2019

Digital Twin-based Design and Simulation of a Self-sensing Gripper Finger for Adaptive Grasping

1st Geonjun Kim
School of Aerospace and Mechanical Engineering
Korea Aerospace University
Goyang-si, Gyeonggi-do, Korea
mktg09@kau.kr

1st Hyun-Jin Baek
Department of Smart Air Mobility
Korea Aerospace University
Goyang-si, Gyeonggi-do, Korea
hjbaek@kau.kr

2nd Sukhwa Lee
School of Aerospace and Mechanical Engineering
Korea Aerospace University
Goyang-si, Gyeonggi-do, Korea
conner9403@kau.kr

3rd SungKu Kang
School of Aerospace and Mechanical Engineering
Korea Aerospace University
Goyang-si, Gyeonggi-do, Korea
kang.sungku@kau.ac.kr

4th Hae-Sung Yoon*
School of Aerospace and Mechanical Engineering
Korea Aerospace University
Goyang-si, Gyeonggi-do, Korea
hsyoon7@kau.ac.kr

Abstract—**This paper presents the digital twin-based design and simulation of a self-sensing gripper finger capable of adapting its grasping mode according to the size and stiffness of target objects. Unlike conventional grippers that rely on electric sensors and active control, the proposed design utilizes torsion spring-based mechanical sensing to detect objects' mechanical properties and adjust the internal structure of the finger. Multiple grasping modes are defined to handle various object conditions, and the self-sensing mechanism enables immediate responses. In addition, finite element simulations demonstrate that the finger achieves both higher rigidity and greater compliance compared to conventional compliant fingers. To complement the FEA results and optimize the new rib structure, a digital twin framework was implemented. The developed digital twin framework provides a foundation for future optimization of self-sensing grippers.**

Keywords—digital twin, passive self-sensing, adaptive grasping, Fin-ray finger, soft finger, simulation-based optimization

I. INTRODUCTION

Soft robotic grippers have attracted significant attention in recent years due to their ability to handle objects with diverse shapes, sizes, and mechanical properties. Unlike rigid grippers that provide strong grasping forces but struggle to maintain stable contact with soft or irregular objects, soft robotic grippers offer high adaptability. However, they often suffer from limited load capacity, revealing a fundamental trade-off between rigidity and flexibility.

To overcome this limitation, many studies have introduced sensor-based and control-based strategies, including tactile sensors, force feedback systems, and vision-based classification. Although these approaches enhance adaptability, they typically increase the overall system complexity, delay the response time, and consume additional energy. Therefore, conventional approaches restrict their applicability in lightweight and compact robotic systems.

The proposed study first classifies objects according to their stiffness and size, and then designs a gripper finger capable of mode transition to generate a sufficient grasping force and fingertip flexibility for each type. The finger employs a torsion spring-based mechanical sensing mechanism that detects variations in object size and stiffness and passively adjusts the timing of mode transition. In addition, finite element analysis (FEA) was conducted to evaluate the contact force and fingertip displacement, confirming that the proposed finger simultaneously improves both rigidity and flexibility compared with conventional designs. Furthermore, a digital twin framework was established to complement the FEA results and explore the optimal internal rib configuration. This virtual environment enables simulation-based evaluation of key design parameters, such as rib thickness and internal infill density, providing a foundation for future optimization of mechanically self-sensing grippers. The objective of this work is to develop a gripper that can hold objects of various stiffness regardless of their size.

II. METHODOLOGY

A. Design Concept

Based on the gripper finger structure proposed in the previous study [1], the gripper finger proposed in this study adopts a Fin-ray-inspired compliant structure with three disconnected ribs and a single V-shaped rib positioned between two disconnected ribs.

Grasping an object causes the rib to engage with the gripper's contact surface in sequence, with the fingers smoothly transitioning from a flexible state to a rigid state. Before each rib makes contact, the structure behaves compliantly and conforms to the object surface; as additional ribs engage, the overall stiffness increases. This mechanism forms the foundation for adaptive grasping of objects with varying stiffness and size.

Previous research [1] indicated that this structure improves both contact force and fingertip displacement (flexibility) of the finger compared to conventional Fin-ray finger. Finite element analysis (FEA) was conducted in ANSYS to investigate the contact force and displacement of the finger while grasping a circular object.

Conventional Fin-ray fingers tend to exhibit a trade-off between rigidity and flexibility; an increased number of

---

This work was supported by the BK21 FOUR program through the National Research Foundation of Korea (NRF) funded by the Korean government (Ministry of Education) (No.: 5199990714521).

internal ribs enhances stiffness but reduces fingertip displacement. In contrast, the proposed design improves both contact force and flexibility simultaneously, achieving a more balanced and efficient grasping performance.

B. *Mechanical Sensing Mechanism*

The grasping process is divided into multiple modes according to the sequential engagement of the disconnected ribs, *i.e.*, no rib, bottom rib, middle rib, and top rib.

The structure induces a continuous progression of multiple modes during deformation. Depending on object size and stiffness, a different mode becomes preferable, and the design objective is to extend the dwell time in the mode appropriate to the object for a more stable grasp.

Two torsion spring–based mechanisms are employed. For size sensing, the torsion springs are connected to the distal ends of the bottom rib and the top rib. For stiffness sensing, another torsion spring is mounted on the V-shaped rib.

When grasping a larger or stiffer object, the size-sensing and stiffness-sensing springs are compressed by the external load, pressing the corresponding ribs. This action advances the contact timing of rib engagement and extends the dwell time of the relevant modes, resulting in improved grasping force and flexibility.

C. *Digital Twin-based Simulation*

A digital twin environment was constructed to complement the finite element analysis and support optimization of the proposed gripper finger. Key design parameters were defined to evaluate their influence on contact force and fingertip displacement. The digital twin enables iterative simulation and visualization of the finger's deformation behavior (mode transition) under various grasping conditions, providing a virtual platform for future optimization of the internal rib structure.

III. RESULTS AND DISCUSSION

The FEA results demonstrated that the conventional Fin-ray finger shows a trade-off between rigidity and flexibility. In contrast, the proposed finger improves both contact force and flexibility simultaneously, confirming the effectiveness of the disconnected rib and V-shaped rib configuration.

Conceptually, the grasping process can be divided depending on rib engagement, and each mode corresponds to a specific range of object stiffness and size. The proposed sensing mechanism is designed to maintain a longer dwell time in the most appropriate mode for each object, thereby improving grip stability.

Because the sensing and mode switching are achieved purely through the mechanical deformation of the torsion springs, the finger responds instantaneously to contact without any electrical sensing or control processing. This super-fast response and absence of external sensors or actuators significantly simplify the gripper system, reducing both control complexity and hardware requirements.

To complement the FEA results and further explore the gripper's design performance, a digital twin environment was developed. The virtual model replicates the material properties of the finger, enabling real-time visualization and simulation of its deformation behavior (mode transition) under different grasping conditions. Parametric evaluation provides a foundation for future optimization of the internal rib configuration.

IV. CONCLUSION

This study presented a self-sensing gripper finger that adapts its grasping behavior to various object sizes and stiffnesses through a mechanical mode-transition mechanism. The proposed design achieved both higher contact force and greater flexibility compared with conventional Fin-ray fingers.

A digital twin environment was implemented to support design evaluation and parameter analysis, providing a basis for future optimization of the internal rib structure. Future work will focus on experimental verification of the mode-switching behavior and quantitative performance analysis of the optimized finger.

REFERENCES

[1] G. Kim, S. Yoon, H. Jung, H. Bae, S. Kang, and H.S. Yoon, "A soft gripper with an adaptive grasp control system allowing stable and rapid re-posturing," unpublished.

Experimental Study on Switching Control Scheme of Folding-Wing UAV for Flight Mode Transition

1st Seunghyun Choi
Department of Smart Air Mobility
Korea Aerospace University
Deogyang-gu, Goyang-si, Gyeonggi-do, Republic of Korea
tmdgus6513@kau.kr

2nd Dongwon Jung
Department of Smart Air Mobility
Korea Aerospace University
Deogyang-gu, Goyang-si, Gyeonggi-do, Republic of Korea
djung@kau.ac.kr

Abstract— **Recent advances in vertical take-off and landing (VTOL) technologies have led to the development of morphing configurations such as tilt-rotor, tilt-wing, and folding-wing vehicles. These variable-geometry systems can achieve both efficient cruise and vertical mobility, but their nonlinear dynamics and drastic aerodynamic coupling make stable transition control challenging.**

This paper presents a practical switch control methodology for a variable folding-wing VTOL UAV implemented on a Pixhawk autopilot. The proposed system synchronizes the wing morphing mechanism with attitude and thrust control loops to enable smooth and stable transition between multicopter and fixed-wing modes. An analytical blending function is formulated to gradually redistribute control authority across roll, pitch, and yaw axes throughout the transition phase.

The control logic is integrated into the PX4 flight stack and validated through both high-fidelity simulation and real flight-test data. The proposed method achieved attitude tracking errors below 2° and demonstrated consistent switching behavior under nonlinear aerodynamic interactions. The results confirm that the proposed approach provides a reliable and reproducible methodology for real-time transition control of variable-geometry UAVs, contributing to the broader application of morphing aerial platforms.

I. INTRODUCTION

Unmanned aerial vehicles (UAVs) with variable or morph- ing geometries have gained considerable attention over the past decade, as researchers seek designs that combine vertical take-off and landing (VTOL) capability with efficient forward flight [1]–[3]. Among these, hybrid configurations such as tiltrotor, tiltwing, and tailsitter systems integrate the hovering stability of multirotors with the aerodynamic efficiency of fixed-wing aircraft. These platforms are increasingly employed for urban air mobility (UAM), long-range logistics, and reconnaissance missions, offering operational flexibility and endurance that conventional UAVs cannot achieve. However, the morphing mechanisms that enable this flexibility also introduce considerable control challenges.

During the transition phase between hover and cruise flight, the aerodynamic and inertial properties of such aircraft vary dramatically, leading to strong nonlinear coupling and rapid shifts in stability derivatives. The center of lift and control effectiveness change continuously as lift generation transitions from rotors to wings. These dynamics often cause attitude oscillations or loss of control authority if not managed properly.

To address this issue, previous research has explored model predictive, gain-scheduled, and adaptive nonlinear control approaches for tiltrotor and tiltwing UAVs [4], [5]. Although these methods have improved transition smoothness, they typically rely on complex parameter tuning and high computational cost, which limits practical onboard implementation.

A particularly challenging class of morphing aircraft is the folding-wing VTOL, which physically folds or unfolds its wings during flight to switch between compact rotor-borne and extended wing-borne configurations [6]. This geometry change induces drastic variations in aerodynamic forces, moments, and moments of inertia, producing strong coupling among roll, pitch, and yaw dynamics. Consequently, a control system for a folding-wing UAV must continuously adjust control authority and blending ratios in real time according to the wing's folding angle. Analytical studies have shown that wing morphing can enhance overall efficiency, yet experimental implementations remain rare due to the inherent control complexity of such systems.

Recent advances in open-source autopilot architectures, particularly the Pixhawk platform and its PX4 middleware, have made it feasible to develop and deploy customized control logic for unconventional UAVs. Pixhawk provides real-time computing capability, a modular flight control structure, and open interfaces for actuator and sensor integration. These features enable researchers to prototype and validate morphing control algorithms directly on hardware, without reliance on proprietary systems. Such flexibility is essential for morphing UAVs, where the flight controller must tightly couple mechanical motion and aerodynamic behavior.

Building upon the PX4 autopilot platform, this study presents a practical switch control methodology for a folding- wing VTOL UAV that enables smooth transition between multicopter and fixed-wing flight. The proposed system employs analytical logistic blending functions to gradually redistribute control authority among roll, pitch, and yaw axes as the wing folds, thereby minimizing actuator discontinuities and attitude coupling during transition. The control logic is fully implemented in the PX4 firmware and validated through both simulation and real flight-test data. Overall, the proposed method provides a unified and experimentally verified method for real-time morphing control, applicable to various hybrid UAV configurations such as tiltrotors and tiltwings.

II. METHODOLOGY

The folding-wing UAV modifies its wing geometry in flight via servo actuation, allowing smooth transition between multicopter and fixed-wing modes. This enables both VTOL operation and improved cruise efficiency. Fig 1. shows the transition process, and Table I. summarizes the key specifications.

Fig. 1. Folding-Wing Vehicle

TABLE I. FOLDING-WING UAV SPECIFICATIONS

Total Weight	8.05 kg
Fuselage (FW)	
Length	1.646 m
CG_x / CG_z	0.0716 m / 0.0154 m
Xnp	0.175
Main Wing (Clark Y)	
Chord (root/tip)	0.35 m / 0.30 m
Span	2.355 m
Reference area (Sref)	0.79 m^2
Wing loading	99.96 N/m^2
Taper ratio	0.857
MAC	0.347 m
Aspect ratio	7.01
Incidence	0 deg
V-Tail (NACA 0015)	
Chord (root/tip)	0.25 m / 0.20 m
Span	0.972 m
Area	0.219 m^2
Aspect ratio	4.32
Tail volume	0.447
Incidence	-2.4 deg
Motor/Control (MC)	
Rotor spacing	1.1737 m
Cant angle	5 deg
Prop spacing	0.099 m
Motor	AT 3520 550kv
Battery	6S 5200 × 2
Disk loading	173.17 N/m^2
propeller	15×6 (wood)
Disk loading	173.17 N/m^2

A. Transition Control Architecture

The vehicle transitions between multicopter (MC) and fixed- wing (FW) modes through intermediate phases, as shown in Fig. 2. The design blends the MC and FW controllers across transition using a normalized tilt state $\tau \in [0, 1]$, where the physical tilt angle is $\theta = \tau\,\theta_{\max}$ with $\theta_{\max} = 117.5^{\circ}$. Each phase (P1, P2, B) corresponds to forward and back transition segments determined by the tilt angle τ and calibrated air speed V. A fixed-wing request triggers the forward sequence (MC→FW) when τ and V exceed the preset thresholds, whereas releasing the request reverses the process (FW→MC) as the wing tilts back.

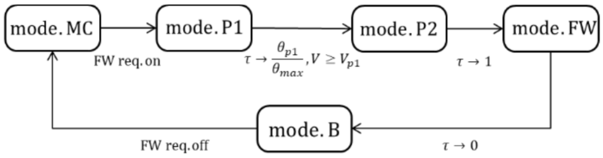


Fig. 2. Mode transition logic between MC, P1, P2, FW, and B phases.

B. Tilt Scheduling (Front/Back Transitions)

Front transition uses two time-based linear ramps; back transition ramps down to zero. Let T_{p1}, T_{p2}, T_{back} be phase durations, and θ_{p1}, $\theta_{max} = \theta_{p2}$ the phase targets. The variable t denotes the current time, and $t_{0,p1}$, $t_{0,p2}$ and $t_{0,back}$ represent the respective start times of each transition phase:

$$\tau_{p1}(t) = \frac{\theta_{p1}}{\theta_{max}} sat\left(\frac{t - t_{0,p1}}{T_{p1}}, 0{,}1\right) \tag{1}$$

$$\tau_{p2}(t) = \frac{\theta_{p1} + (\theta_{max} - \theta_{p1}) sat\left(\frac{t - t_{0,p2}}{T_{p2}}, 0{,}1\right)}{\theta_{max}} \tag{2}$$

$$\tau_{back}(t) = \tau_{max}\left(1 - sat\left(\frac{t - t_{0,back}}{T_{back}}, 0{,}1\right)\right) \tag{3}$$

safety ramp is triggered upon pilot override to ensure smooth mode switching. The first phase (τ_{p1}) tilts the wing to an intermediate multicopter mode, while the second phase (τ_{p2}) completes the transition to the fixed-wing mode. During the back transition, τ_{back}symmetrically reduces the tilt state, ensuring a stable return to the multicopter mode.

C. Weighting function for switching control

For each attitude axis i ∈ {ϕ, θ, ψ}, virtual control moments from MC and FW controllers are weighted as

$$M_i(\tau) = M_i^{MC} w_{MC,i(\tau)} + M_i^{FW} W w_{FW,i(\tau)} \tag{4}$$

Weights are computed using per-axis parameters k_i(steepness), s_i(shift), and an optional reversal:

$$w_{FW,i}(\tau) = \frac{\sigma(z_i(\tau)) - \sigma(z_{i,min})}{\sigma(z_{i,max}) - \sigma(z_{i,min})} \tag{5}$$

$$\sigma(z) = \frac{1}{1+e^{z}} \tag{6}$$

$$z_i(\tau) = (1 - 2\tau - s_i)10k_i \tag{7}$$

$$w_{MC,i}(\tau) = \begin{cases} 1 - w_{FW,i}(\tau) & if\ REV = 1 \\ w_{FW,i}(\tau) & if\ REV = 0 \end{cases} \tag{8}$$

With the normalization, $w_{MC,i}(0) = 1$ and $w_{MC,i}(1) = 0$, yielding a smooth authority transfer.

III. RESULTS

Fig. 3. illustrates the transition of switching control between multicopter (MC) and fixed-wing (FW) attitude commands as a function of the tilt angle $\theta \in [0, \theta_{max}]$ (with $\theta_{max} = 117.5^{\circ}$ in our implementation). For each attitude axis $i \in \{\phi$ (roll), θ (pitch), ψ (yaw)$\}$, the MC and FW weighting functions, $w_{MC,i}(\theta)$ and $w_{FW,i}(\theta)$, are generated using the same normalized smooth weighting formulation employed in the onboard implementation, with identical parameter values. The shaded regions indicate which controller dominates ($w_{MC,i} > w_{FW,i}$ on the left, $w_{FW,i} > w_{MC,i}$ on the right). The vertical dashed line marks the crossover angle θ^{*} where both weights are equal,

representing the point of balanced switching control between the MC and FW modes.

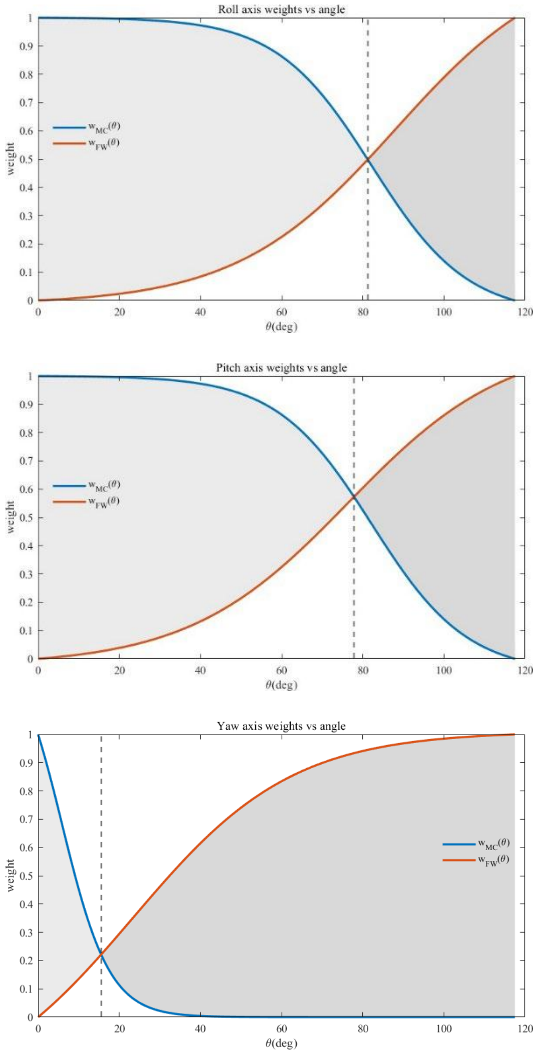


Fig. 3. Switching control weights for roll, pitch, and yaw

Interpretation.: (i) *Roll*: The roll axis maintains dominant multicopter control authority through most of the transition, with a delayed crossover designed to prevent premature wing- induced roll coupling. Since roll stability is the most safety-critical axis during transition—loss of roll control can lead to immediate instability—the system prioritizes roll preservation by retaining multicopter control longer and disabling yaw control earlier to avoid cross-axis interference. (ii) *Pitch*: The pitch axis gradually shifts control to the fixed-wing side as the wing unfolds. Because onboard components such as batteries are mounted within the wings, the center of gravity and moments of inertia vary during morphing. Thus, a smooth pitch transition is required to compensate for these inertial shifts and ensure longitudinal stability. (iii) *Yaw*: Yaw authority transitions earlier to the fixed-wing controller, as the aerodynamic tail surfaces begin to contribute. Early yaw transfer minimizes interference with roll control and supports coordinated turn stability.

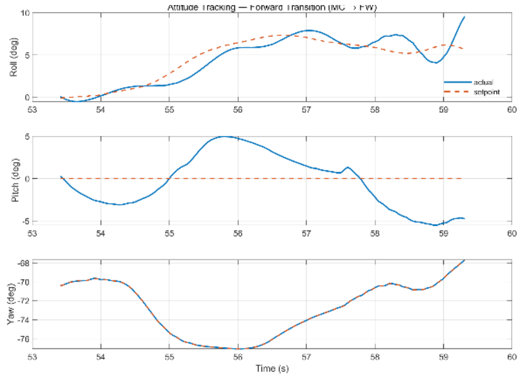


Fig. 4. Measured attitude responses from flight log during forward transition (MC→FW)

Fig. 4 presents the attitude response obtained from the flight log during the forward transition from multicopter (MC) to fixed-wing (FW) mode. The highlighted segment (53–60 s) corresponds to the period when the wing tilt angle increased from 0 to 1 (normalized setpoint), as detected from the PX4 mode flag.

As shown in Fig. 3, the attitude control is governed by a switching control scheme that adjusts the contribution of the MC and FW controllers according to the tilt angle θ. During the forward transition, this switching control gradually modifies the control behavior on each attitude axis, ensuring a smooth mode transition. The MC yaw controller is intentionally disabled earlier than the other axes to prevent roll–yaw coupling, which can occur as the wings start to tilt forward. If the yaw control were kept active during this phase, the cross- coupled moment could destabilize the roll response. Therefore, the yaw channel is switched off in advance, resulting in the FW yaw signal representing the current heading angle rather than a commanded setpoint; this explains the coincidence of the setpoint and actual yaw curves in Fig. 4.

The roll shows a stable response throughout the transition, maintaining coordination despite the removal of yaw input. This indicates that the early yaw switching effectively suppresses coupling effects and allows the roll controller to stabilize the lateral attitude. In contrast, the pitch exhibits a transient overshoot, primarily due to variations in the center of gravity (CG) and moment of inertia (MOI) as heavy components such as the battery are mounted on the tilting wings. To achieve a stable transition despite these inertial and aerodynamic changes, the pitch setpoint was fixed at 0° (level attitude), allowing the aircraft to gradually settle while compensating for the shifting mass distribution.

IV. CONCLUSION

This paper has presented a control strategy for a folding-wing UAV that enables smooth transitions between multicopter and fixed-wing flight. The proposed analytical blending technique—derived from actual PX4 autopilot implementation—dynamically redistributes control authority across the roll, pitch, and yaw axes through axis-specific logistic weighting functions. Both numerical simulations and flight tests confirmed stable attitude tracking during forward transition, maintaining RMS tracking errors below 2° across all axes.

The integrated control framework ensures seamless mode transition and demonstrates feasibility for real-time deployment by embedding morphing logic within an open-source autopilot (PX4). This approach promotes reproducibility, scalability, and broader accessibility for future morphing-wing UAV research and development.

Future work will focus on validating closed-loop flight performance under wind disturbances and varying payload conditions. To enhance robustness in off-nominal scenarios, adaptive gain scheduling and real-time aerodynamic parameter estimation will be explored. Additionally, high-fidelity Hardware-in-the-Loop (HIL) simulations and model-based flight dynamics identification will be employed to refine the aerodynamic–control integration, paving the way toward a scalable platform for next-generation morphing UAVs.

ACKNOWLEDGMENT

This work was supported by the BK21 FOUR pro- gram through the National Research Foundation of Korea (NRF) funded by the Korean government (grant number: 2120240715457).

REFERENCES

[1] M. Osman, Y. Xia, M. Mahdi, and A. Ahmed, "Hybrid VTOL UAV technologies: Efficiency, customization, and sector-specific applications," *Alexandria Engineering Journal*, vol. 120, pp. 13–49, 2025. doi: 10.1016/j.aej.2024.12.087.

[2] A. S. Saeed, A. B. Younes, S. Islam, J. Dias, L. Seneviratne, and G. Cai, "A review on the platform design, dynamic modeling and control of hybrid UAVs," in *Proc. 2015 International Conference on Unmanned Aircraft Systems (ICUAS)*, Denver, CO, USA, 2015, pp. 806–815. doi: 10.1109/ICUAS.2015.7152365.

[3] Y. Alqudsi, H. Sulaiman, and E. Pehlu¨l, "A comprehensive review of VTOL UAVs: Design, technologies, and applications," *Preprints*, 2025. doi: 10.20944/preprints202504.2524.v1.

[4] M. Hsu, AquaFly: A tilt-rotor vertical take-off and landing aquatic unmanned aerial vehicle, M.S. thesis, Univ. of Toronto, Toronto, Canada, 2020.

[5] D. Milz and G. Looye, "Tilt-wing control design for a unified control concept," in *AIAA SciTech 2022 Forum*, San Diego, CA, USA, 2022, p. 1084. doi: 10.2514/6.2022-1084.

[6] Z. Lin, B. Yan, T. Zhang, S. Li, Z. Meng, and S. Liu, "Multi-level switching control scheme for folding-wing VTOL UAV based on dynamic allocation," *Drones*, vol. 8, no. 7, 303, 2024. [Online]. Available: https://doi.org/10.3390/drones8070303.

Flight Load Derivation of a Lift & Cruise eVTOL Aircraft Based on V–n Diagram Analysis

Myeong-Gi Kim
Department of Smart Air Mobility
Korea Aerospace University
Gyeonggi-do, Republic of Korea
myeonggi07@kau.kr

Jinho Shin
Department of Smart Air Mobility
Korea Aerospace University
Gyeonggi-do, Republic of Korea
jhshin976@kau.kr

Yun-Sang Lee
Department of Smart Air Mobility
Korea Aerospace University
Gyeonggi-do, Republic of Korea
lys5472@kau.kr

Sang-Woo Kim*
Department of Aeronautical and
Astronautical Engineering
Korea Aerospace University
Gyeonggi-do, Republic of Korea
swkim@kau.ac.kr

Moonsung Cho
Department of Aeronautical and
Astronautical Engineering
Korea Aerospace University
Gyeonggi-do, Republic of Korea
mscho@kau.ac.kr

***Abstract*— This study presents the derivation of flight loads for an electric vertical take-off and landing (eVTOL) aircraft based on a V–n diagram analysis in accordance with FAA 14 CFR Part 23 Appendix A and ASTM F3396/F3396M standards. Using the KUAM-301 Rev. 02 Lift-and-Cruise configuration as a reference model, the minimum design speeds and load factors were obtained to construct the flight envelope. The derived positive and negative limit load factors were 3.0 and −1.5, and the corresponding gust limit load factors were 3.41 and −1.86. From these parameters, the maximum flight loads acting on the main structural components were calculated as 45,160 N for the main wing, 8,249.7 N for the horizontal stabilizer, and 5,579.3 N for the vertical stabilizer. The results verify the applicability of fixed-wing load estimation procedures to eVTOL cruise conditions and provide fundamental reference data for the preliminary structural design and safety evaluation of future urban air mobility vehicles.**

Keywords—Urban Air Mobility, Electric Vertical Take-Off and Landing, Flight Envelope Diagram, Load Analysis, Flight Load, FAA Part 23, ASTM F3396

I. INTRODUCTION

Urban Air Mobility (UAM) has emerged as a next-generation air transportation system aimed at alleviating urban traffic congestion and enabling short-range, high-speed operations within metropolitan areas [1]. The core vehicle of UAM, the electric vertical take-off and landing (eVTOL) aircraft, utilizes distributed electric propulsion systems to achieve low-noise and zero-emission flight, offering an environmentally sustainable solution suitable for complex urban environments [2].

The Lift & Cruise type of eVTOL aircraft separates the lift rotors for vertical take-off and landing from the cruise propellers, enabling more efficient performance in each flight phase. During cruise flight, its aerodynamic and structural behavior is comparable to that of small fixed-wing aircraft [3]. Since an independent certification for eVTOL aircraft has not yet been fully established, a standardized approach based on conventional aircraft certification criteria remains necessary. Accordingly, existing standards such as FAA 14 CFR Part 23 Appendix A and the simplified ASTM F3396/F3396M-23 have been adopted as certification standard for preliminary load estimation in eVTOL design [4–5].

In this study, FAA Part 23 Appendix A and ASTM F3396/F3396M-23 standards were applied to the Lift & Cruise eVTOL to derive the flight envelope and component loads. The results verify the applicability of fixed-wing load estimation procedures to eVTOL cruise conditions and provide fundamental data for future structural design and standardization efforts.

II. MODEL

The flight load analysis model was based on the KUAM-301 Rev. 02 aircraft, proposed under the research project "Demonstration and Verification Study for UAM Aircraft Certification and Compliance," supported by the Ministry of Land, Infrastructure and Transport (MOLIT) and the Korea Agency for Infrastructure Technology Advancement (KAIA). Fig. 1 shows the external configuration of the KUAM-301 Rev. 02 aircraft.

This vehicle is a Lift & Cruise-type eVTOL, equipped with eight lift rotors and two cruise propellers. It is designed to separate the functions of vertical take-off and landing (VTOL) from those of cruise flight, thereby achieving efficient aerodynamic performance in each flight regime. The principal geometric specifications of the KUAM-301 Rev. 02 are summarized in Table 1. The fuselage of the KUAM-301 Rev. 02 has a length, width, and height of 6.684 m, 0.994 m, and 1.208 m, respectively, with a maximum take-off weight of 1,350 kg. The main wing has a span of 13.53 m, a chord of 1.502 m, and a total planform area of 15.03 m^2. These geometric specifications represent typical aspect ratios and proportions for a small-scale UAM platform designed for short-range urban flight.

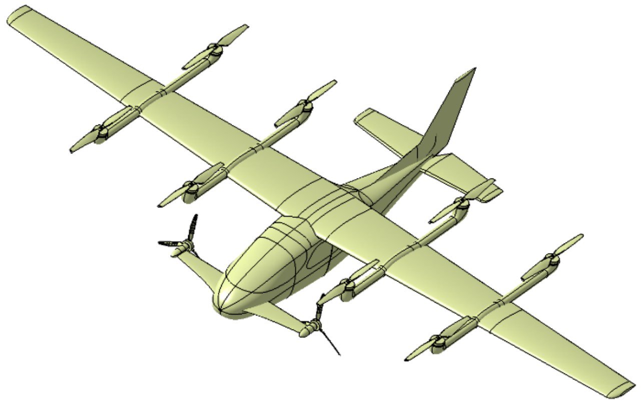


Fig. 1. Configuration of the KUAM–301 Rev. 02 aircraft

TABLE I. SPECIFICATIONS OF THE KUAM–301 REV. 02 AIRCRAFT

Design Parameters	Value
Fuselage length (m)	6.684
Fuselage width (m)	0.994
Fuselage height (m)	1.208
Maximum take-off weight (kg)	1,350
Wing loading (kg/m^2)	89.88
Wing span (m)	13.53
Wing chord (m)	1.502
Wing area (m^2)	15.03
Aspect ratio	12.19

III. FLIGHT ENVELOPE

The flight envelope (V–n diagram) visually represents the relationship between flight speed and load factor, and is used to evaluate the structural safety and flight performance of an aircraft under various operating conditions. Such evaluation enables designers to ensure structural stability during the conceptual design phase and to anticipate potential risks that may occur during flight operations. Fig. 2 illustrates a representative example of a flight envelope as defined in FAA 14 CFR Part 23 Appendix A [4].

$$V_{A,\min} = 15\times\sqrt{n_1\times(W/S)}\,, \tag{1}$$

$$V_{C,\min} = \sqrt{\frac{2\times W}{\rho\times S\times C_L}}\,, \tag{2}$$

$$V_{D,\min} = 24\times\sqrt{n_1\times(W/S)}\,, \tag{3}$$

Here, W, S, ρ, and C_L denote the maximum take-off weight, wing area, air density, and lift coefficient during cruise flight, respectively. (1)-(3) are presented in FAA 14 CFR Part 23 Appendix A and ASTM F3396/F3396M, while (2) represents the minimum flight speed at which the aircraft can operate without stalling [6]. In addition, the positive and negative maneuvering limit load factors for the Normal Category aircraft defined in FAA Part 23 Appendix A are expressed by (4)-(5), respectively.

$$n_1 \le 3.8\,, \tag{4}$$

$$n_2 = -0.5\times n_1\,. \tag{5}$$

The positive and negative gust limit load factors are obtained from the intersection between the curve of wing loading that accounts for load factor effects and the ratio of cruise speed to gust velocity.

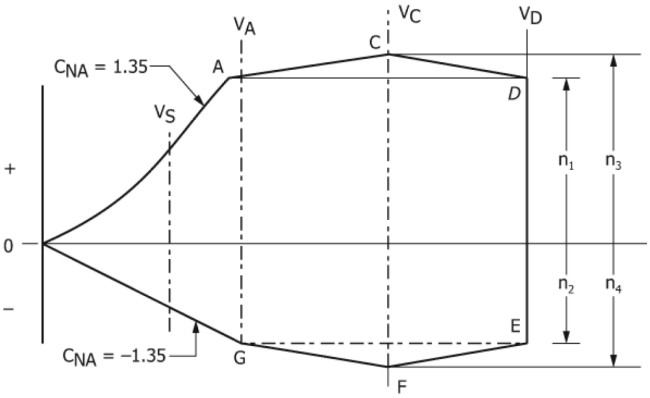


Fig. 2. Flight envelope presented in *FAA Part 23 Appendix A.*

IV. FLIGHT LOADS

A. Main wing

The main wing is one of the most critical components of an aircraft, responsible for generating lift that enables sustained flight. The load acting on the wing is calculated as the product of the maximum load factor obtained from the flight envelope and the aircraft's maximum take-off weight, as expressed in (1) [7].

$$L_w = n\times W \tag{6}$$

Here, Lw denotes the load acting on the wing, n represents the maximum load factor obtained from the flight envelope, and W is the maximum take-off weight of the aircraft.

B. Horizontal stabilizer

The horizontal stabilizer maintains the aircraft's pitch stability and controls altitude through the elevator. The load per unit area acting on the horizontal stabilizer is determined using (7).

$$\bar{w} = 4.8 + 0.534\times(n_1\times W/S) \tag{7}$$

Here, $\bar{w}$ represents the load per unit area acting on the stabilizer. During flight operations, the effective load per unit area on the horizontal stabilizer is calculated by multiplying by the higher value of the load factors obtained from (8)-(9).

$$n = (V_A/V_{A,\min})^2 \tag{10}$$

$$n = (V_C/V_{C,\min})^2 \tag{11}$$

Here, V_A and V_C denote the maneuvering and cruising speeds, respectively, while $V_{A,\min}$ and $V_{C,\min}$ represent the minimum maneuvering and minimum cruising speeds, respectively.

C. Vertical stabilizer

The vertical stabilizer maintains the directional stability of the aircraft and controls its yaw motion through the rudder. The load per unit area acting on the vertical stabilizer is determined using (12)-(13).

$$\bar{w} = 3.66\times(n_1\times W/S)^{\frac{1}{2}}\ for\ n_1\times W/S < 47 \tag{12}$$

$$\bar{w} = 0.534\times(n_1\times W/S)\ for\ n_1\times W/S > 47 \tag{13}$$

During flight operations, the effective load per unit area on the vertical stabilizer is obtained by multiplying the unit load by the higher value of the load factors derived from (10)-(11), as in the case of the horizontal stabilizer.

V. RESULTS

A. Flight envelopment

The flight envelope of the KUAM-301 Rev. 02 aircraft was derived based on the design speeds and load factors in accordance with FAA 14 CFR Part 23 Appendix A and ASTM F3396/F3396M standards. The design speeds of the KUAM-301 Rev. 02 were determined using the Sling 4 HW aircraft developed by TAF as a reference model. The maneuvering, cruising, and dive speeds were set to 206.3, 231.0, and 287.0 km/h, respectively [8]. The minimum maneuvering, cruising, and dive speeds obtained from (1)–(3) were 206.3, 157.6,

TABLE II. MINIMUM DESIGN SPEED AND LOAD FACTOR FOR KUAM–301 REV. 02 AIRCRAFT

Design Parameters	Value
Minimum maneuvering speed (km/h)	206.3
Minimum crusing speed (km/h)	157.6
Minimum dive speed (km/h)	287.0
Positive maneuvering limit load factor	3.0
Negative maneuvering limit load factor	-1.5
Positive gust limit load factor	3.41
Negative gust limit load factor	-1.86

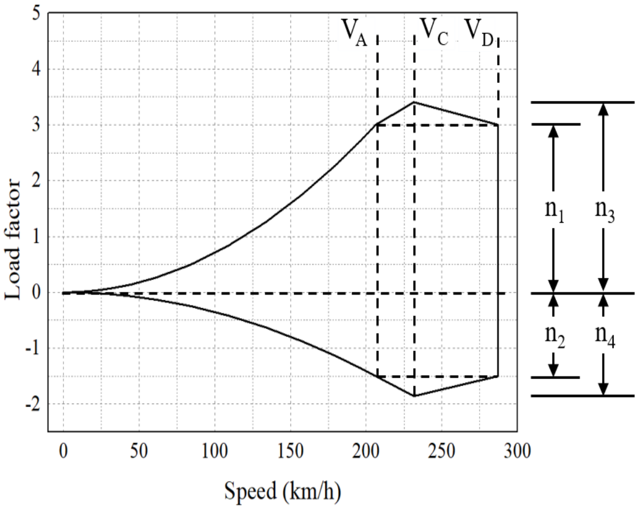


Fig. 3. V-n diagram for KUAM–301 Rev. 02 aircraft

and 287.0 km/h, respectively. At this condition, the maximum take-off weight was 1,350 kg, the wing area was 15.03 m^2, the air density was 1.225 kg/m^3, and the lift coefficient during cruise was 0.75. The cruise lift coefficient was selected based on the design parameters of the NASA X-57 aircraft, which has a comparable configuration and performance class [9]. The positive maneuvering limit load factor was set to 3.0, considering both the operational environment of the KUAM-301 Rev. 02 aircraft in urban areas and the reference data from the OPPAV (Optionally Piloted Personal Air Vehicle) program developed by the Korea Aerospace Research Institute (KARI). The negative maneuvering limit load factor was set to −1.5. The ratio of cruise speed was calculated to be 1.47, and the positive and negative gust limit load factors, obtained from the intersection of the wing loading curve and the cruise speed ratio curve considering the limit load factors, were 3.41 and −1.86, respectively. Table 2 summarizes the minimum design speeds and load factors of the KUAM-301 Rev. 02 aircraft derived from FAA Part 23 Appendix A, ASTM F3396/3396 and the corresponding aircraft specifications. Fig. 3 shows the flight envelope of the KUAM-301 Rev. 02 aircraft, which was derived based on the design speeds and load factors. Using the derived flight envelope, the flight loads acting on the main wing, tailplane, and control surfaces of the KUAM-301 Rev. 02 were subsequently calculated.

B. *Main wing*

Based on the V–n diagram, the positive maneuvering and gust limit load factors were determined to be 3.0 and 3.41, respectively. Accordingly, the higher value, the positive gust limit load factor, was applied to calculate the load acting on the main wing. Given the maximum take-off weight of the KUAM-301 Rev. 02 aircraft as 1,350 kg and the positive gust limit load factor of 3.41, the load acting on the main wing, calculated using (1), was 45,160.3 N. This value represents the maximum load that can be applied to the main wing, indicating that the wing structure must be designed to withstand this load during the aircraft's structural design process.

C. *Horizontal stabilizer*

For the KUAM-301 Rev. 02 aircraft, the positive maneuvering limit load factor was 3.0, the maximum take-off weight was 1,350 kg, and the surface area of the horizontal stabilizer was 2.33 m^2. According to (7), the load per unit area acting on the horizontal stabilizer was calculated to be 1,640.9 N/m^2. The corresponding airspeeds, V_A, $V_{A,min}$, V_C, and $V_{C,min}$, were 206.3, 206.3, 231.0, and 157.6 km/h, respectively. The ratios $(V_A/V_{A,min})^2$ and $(V_C/V_{C,min})^2$, obtained from (10)-(11), were 1.0 and 2.16, respectively. Accordingly, the higher value of 2.16 was applied as the load factor acting on the horizontal stabilizer. Consequently, the total load acting on the horizontal stabilizer, calculated as the product of the unit area load, load factor, and surface area, was determined to be 8,249.7 N.

D. *Vertical stabilizer*

The wing loading of the KUAM-301 Rev. 02 aircraft is 89.9 kg/m^2, and the load per unit area acting on the vertical stabilizer was derived using (12). The resulting load per unit area of the vertical stabilizer was 1,411.1 N/m^2. The surface area of the vertical stabilizer is 1.83 m^2, and the total load acting on the vertical stabilizer, calculated as the product of the unit area load, load factor, and surface area, was determined to be 5,579.3 N.

VI. CONCLUSION

In this study, the minimum design speeds and load factors of the KUAM-301 Rev. 02 aircraft were derived, and the corresponding flight envelope was established. Based on the derived flight envelope, the loads acting on the main structural components—the main wing, tailplane, and control surfaces—were analyzed. The maximum flight load acting on the main wing was determined to be 45,160 N, while the horizontal and vertical stabilizers experienced loads of 8,249.7 N and 5,579.3 N, respectively. It should be noted that this analysis was performed under the assumption that the aircraft operates in fixed-wing mode, with the primary objective of defining the structural boundary conditions for the conceptual design stage. Therefore, loads associated with vertical take-off, landing, and transition flight were not considered in this study. Future work should include high-fidelity analyses incorporating various operational scenarios and unsteady aerodynamic conditions to further improve the reliability of the structural design. Despite these limitations, the results obtained in this study are expected to serve as fundamental reference data for structural safety assessment and preliminary design of the KUAM-301 Rev. 02 aircraft.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. RS-2022-NR070875). This work is supported by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport(Grant No. RS-2024-00444205). This work was supported by the BK21 FOUR program through the National Research

Foundation of Korea (NRF) funded by the Korean government (grant number: 2120240715457).

REFERENCES

[1] Arafat, M. Y, and Pan, S, “Urban Air Mobility Communications and Networking: Recent Advances, Techniques, and Challenges,” Drones, vol. 8, pp. 702, 2024.

[2] Guo, J., Chen, L., Li, L., Na, X., Vlacic, L. and Wang, F. Y., “ Advanced Air Mobility: An Innovation for Future Diversified Transportation and Society, ” IEEE Transactions on Intelligent Vehicles, vol. 9, pp. 3106-3110, 2024.

[3] Su, W., Qu, S., Zhu, G., Swei, S. S. M., Hashimoto, M. and Zeng, T., "Modeling and control of a class of urban air mobility tiltrotor aircraft," Aerospace Science and Technology, Vol. 124, pp. 107561, 2022

[4] Federal Aviation Administration, Department of Transportation., Appendix A to Part 23, Title 14, 2017.

[5] ASTM F3396/F3396M-23, “Standard Practice for Aircraft Simplified Loads Criteria, “ ASTM International, 2023.

[6] Yoon, S. J., Aerodynamics, 2nd Ed., Sungandang, Seoul, 2012.

[7] Kim, S. J. and Lee, S. G., “Structural Sizing for Optionally Piloted PAV Preliminary Design,” Journal of the Korean Society for Aviation and Aeronautics, Vol. 28, pp. 83-89, 2020.

[8] Sling 4 HW Pilot Operating Handbook - Rev 1.0, 2023.

[9] Viken, J. K., Viken, S., Deere, K. A., & Carter, M. “Design of the cruise and flap airfoil for the X-57 Maxwell distributed electric propulsion aircraft, ” In 35th AIAA Applied Aerodynamics Conference, pp. 3922, 2017.

Finite Element Analysis of the Thermal Response of Glass Fiber/Vinyl Ester for Drone's Outer Skin

Seong-Hun Jeon
Department of Smart Air Mobility
Korea Aerospace University
Gyeonggi-do, Republic of Korea
jsh1101@kau.kr

Myeong-Gi Kim
Department of Smart Air Mobility
Korea Aerospace University
Gyeonggi-do, Republic of Korea
myeonggi07@kau.kr

Sang-Woo Kim*
Department of Aeronautical and
Astronautical Engineering
Korea Aerospace University
Gyeonggi-do, Republic of Korea
swkim@kau.ac.kr

Abstract— **Drones that perform missions in high-temperature environments such as structural fires, high-temperature facilities, and volcanic activity are greatly affected by the thermal environment; therefore, selecting thermal-protection materials for the drone outer skin is essential. Glass fiber/vinyl ester composite has been reported in studies evaluated on the premise of application to drone structures, making it a valid candidate material for the outer skin. This study implemented a heat-transfer analysis code by coupling a Finite Element Analysis (FEA) program with user subroutines and performed code verification against literature data for a glass fiber/vinyl ester plate. The verification was conducted under conditions where a 100 mm × 100 mm × 9 mm model was subjected to a through-thickness incident heat flux of 50 kW/m² for 3,000 s. As a result, at 3,000 s the surface temperature approached the literature-reported value of 615.89 °C, and the back surface temperature converged to approximately 412.55 °C. The density began to decrease due to pyrolysis at approximately 70 s at 0.0 mm and immediately reached the char density; at 4.5 mm it started at approximately 570 s and reached full char at about 890 s, which is similar to the depth-wise delay trend reported in the literature. Thus, it was confirmed that the coupled analysis reproduces the heat-transfer behavior of the glass fiber/vinyl ester composite at the literature level and provides a verification basis applicable to evaluating the thermal response of thermal-protection materials for drone outer skins for high-temperature missions.**

Keywords—drone, drone skin, thermal protection, glass-fiber/vinyl ester, Finite Element Analysis(FEA)

I. INTRODUCTION

Drones are employed for diverse missions in high-temperature environments such as interior reconnaissance of structural fires, inspection inside high-temperature facilities, and sampling of volcanic gases. A tethered micro-drone designed for structural-fire environments has been reported and field-demonstrated [1], and the safety and efficiency of UAV visual inspection in high-temperature facilities such as industrial boilers have been summarized [2]. An offline path-planning method for close-range inspection inside enclosed structures has also been proposed [3], and case studies have presented of using a drone to collect volcanic gases and simultaneously analyze volcanic ash and gas [4], [5]. For such drones to accomplish missions reliably under external high temperatures, it is essential to select thermal-protection materials for the outer skin that can shield the drone's temperature-sensitive core components from the external heat..

In particular, glass fiber/vinyl ester composites have reported fabrication cases predicated on application to drone structures, confirming their potential as candidate materials for the drone outer skin [6]. glass fiber/vinyl ester composites has also been evaluated as a candidate for aircraft fuselage skin structures [7].

This study implements a heat-transfer analysis code for glass fiber/vinyl ester composites by coupling the commercial Finite Element program Abaqus with the user subroutines UMATHT and DFLUX, and performs code verification against literature data for a glass fiber/vinyl ester plate. The model computes temperature and density variations under heating to evaluate whether it reproduces the literature-reported through-thickness trends of temperature rise and density reduction.

II. THEORETICAL BACKGROUND

Heat Transfer Equation

Heat-transfer simulation is used to compute the temperature-field distribution of a material. In this study, it is assumed that pyrolysis gas does not accumulate inside the material and that thermochemical expansion does not occur. The heat-transfer behavior of a composite laminated plate at high temperature was established on the basis of the conservation of energy [8].

$$\frac{\partial}{\partial t}\left(\rho h_s\right) = \nabla\left(k_1\frac{\partial T}{\partial x}\bar{i} + k_2\frac{\partial T}{\partial y}\bar{j} + k_3\frac{\partial T}{\partial z}\bar{k}\right) + \frac{\partial}{\partial z}\left(m_g^{'} h_g\right) + Q\frac{\partial \rho}{\partial t} \tag{1}$$

The left-hand side of (1) represents the rate of change of energy per unit volume of the material. On the right-hand side, the first term denotes internal heat conduction of the material, the second term denotes heat diffusion associated with the generation of pyrolysis gas, and the third term denotes the pyrolysis of the material per unit time. Here, T and t are the temperature and time, respectively; ρ is the density; $m^{'}_{g}$ is the mass flux of the pyrolysis gas. h_s and h_g are the enthalpies of the solid and the gas. k_1, k_2, k_3 denote the material's x, y, z directional thermal conductivities.

Material Decomposition

As proposed by Goldstein, the virgin material decomposes into char and gas, and this process is expressed by (2). The generated gas is released from the material while decreasing the material density.

$$virgin = gas + char \tag{2}$$

Because the material does not undergo thermal expansion during pyrolysis, its volume remains unchanged; therefore, it can be computed as in (3), and the formation of the char layer can be judged by not undergone pyrolysis(= 0) and fully pyrolyzed(= 1).

$$F = \frac{\rho_{\mathrm{v}} - \rho}{\rho_{\mathrm{v}} - \rho_{\mathrm{c}}} \begin{cases} F = 0, & \text{not undergone pyrolysis} \\ F = 1, & \text{fully pyrolyzed} \end{cases} \tag{3}$$

The density reduction can be calculated using the Arrhenius equation as in (4).

$$\frac{d\rho}{dt} = -A\left(\rho_{\mathrm{v}} - \rho_{\mathrm{c}}\right)\left(\frac{\rho - \rho_{\mathrm{c}}}{\rho_{\mathrm{v}} - \rho_{\mathrm{c}}}\right)^{\psi} e^{-E/RT} \tag{4}$$

Here, A is the pre-exponential factor, E is the activation energy, ψ is the reaction order, R is the ideal gas constant, and ρ_{v} and ρ_{c} are the virgin and char densities [9].

According to Henderson's model, as pyrolysis progresses, the effective properties can be calculated based on the rule of mixtures using (5) and (6) [10].

$$k = k_{\mathrm{v}} F + k_{\mathrm{c}}(1 - F) \tag{5}$$

$$C_P = C_{P_{\mathrm{v}}} F + C_{P_{\mathrm{c}}}(1 - F) \tag{6}$$

Here, k_{v} and k_{c} are the thermal conductivities of the virgin and char states, respectively, and $C_{P\mathrm{v}}$ and $C_{P\mathrm{c}}$ are the specific heats of the virgin and char states.

When thermochemical expansion inside the material is neglected, the mass flux of the pyrolysis gas can be expressed by (7) [11].

$$\frac{\partial \dot{m}_g'}{\partial z} = -\frac{\partial \rho}{\partial t} \tag{7}$$

III. NUMERICAL VERIFICATION

Numerical Implementation of Heat Transfer Analysis

This study performed heat-transfer analysis by coupling the commercial finite element analysis software Abaqus with user subroutines(UMATHT, DFLUX). The temperature and density data obtained at the final increment of each step were carried over as the initial conditions for the subsequent step. UMATHT subroutines was used to define internal heat generation and the material's thermal behavior during heat transfer, and its implementation referred to the method proposed by Feng et al. [12]. DFLUX subroutine is generally used to model a moving heat source, but in this study it was applied to impose a constant heat flux without considering heat-source motion.

Model Setup for Code Verification

For code verification, the analysis results for a glass fiber/vinyl ester composite reported by Feng et al. were referenced. The material properties used in the analysis are listed in Table I. The plate was modeled with the same dimensions(= 100 mm × 100 mm × 9 mm). The analysis model consisted of eight-node linear brick (DC3D8); the total number of elements was 8,712, and the number of nodes was 10,404. The total simulation time was set to 3,000 s, and a normal (through-thickness) incident heat flux of 50 kW/m² was applied to the front surface of the specimen. Fig. 1 shows the model geometry and boundary conditions.

TABLE I. THERMAL PROPERTIES OF GLASS FIBER/VINYL ESTER

Glass fiber/vinyl ester		
Properties	***Value***	***Units***
Rate constant	5.59×10^{13}	1/s
Activation energy	2.12×10^{5}	J/(kg·mol)
Reaction order	1.00	-
Heat of decomposition	3.78×10^{5}	J/kg
Virgin density	1.90×10^{3}	kg/m³
Char density	1.40×10^{3}	kg/m³
Thermal conductivity glass	1.09	W/(m·K)
Thermal conductivity glass/vinyl ester	0.43	W/(m·K)
Specific heat glass	7.60×10^{2}	J/(kg·K)
Specific heat galss/vinyl ester (at 45 °C)	9.60×10^{2}	J/(kg·K)
Specific heat galss/vinyl ester (at 140 °C)	1.21×10^{3}	J/(kg·K)
Specific heat galss/vinyl ester (at 290 °C)	1.36×10^{3}	J/(kg·K)
Specific heat gas	2.3865×10^{3}	J/(kg·K)

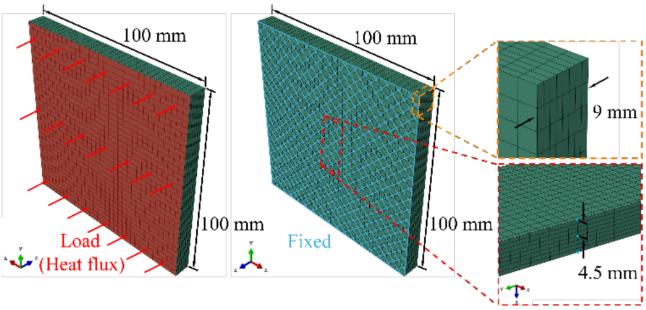


Fig. 1. Geometry, boundary conditions and mesh of the verification glass fiber/vinyl ester model

IV. RESULTS

Temperature Histories

Fig. 2 compares, under the 50 kW/m² condition, the temperature–time histories at the model top surface(= 0 mm), mid-plane(= 4.5 mm), and bottom surface(= 9 mm) with the results of Feng. The initial rapid rise followed by a gradual leveling-off matches, and the through-thickness temperature gradient likewise diminishes with time. At 3,000 s, the surface temperature in the present analysis approaches the literature value of 615.89 °C, and the back-surface temperature converges near 412.55 °C, reproducing the reported trend. The through-thickness temperature profile exhibits a pronounced nonlinear gradient in the near-surface region at early times; as heat accumulates internally with time, it transitions to a quasi-linear gradient.

After 200 s, the temperatures measured in the experiment are much higher than the calculated values. This is because, after about 200 s of heating, the glass fiber/vinyl ester composite undergoes spontaneous surface flaming, which increases the radiative heat flux incident on the material surface and alters both radiative and convective boundary conditions, making it difficult for the model to accurately predict the temperature distribution beyond 200 s [13]. Consequently, the shapes of the temperature histories, the asymptotic levels, and the through-thickness transitions are

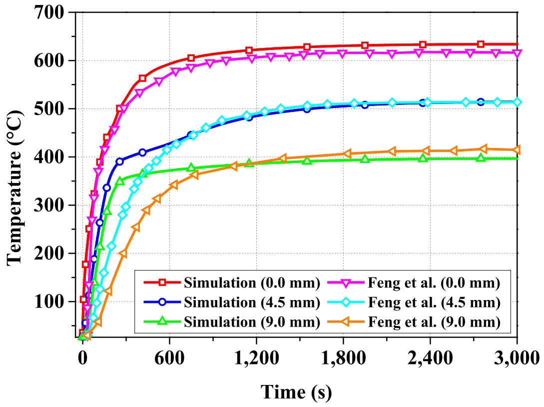


Fig. 2. Comparsion of temperature-time histories at 0.0 mm, 4.5 mm, 9.0 mm

consistent with the literature trends; quantitative agreement is achieved over the interval without boundary condition changes, whereas qualitative agreement is obtained over the interval where flaming initiates.

Density Histories

Fig. 3 compares the density–time histories computed at the model top surface(= 0.0 mm) and at the mid-plane(= 4.5 mm) with those of Feng et al. The density decreased rapidly first at the surface, and the decrease was delayed toward the interior. At the top surface, pyrolysis began at about 70 s and immediately reached the char density(= 1,400 kg/m³). At the mid-plane, pyrolysis began at about 570 s and reached the char density at about 890 s. At the mid-plane, although a difference in the onset time of pyrolysis from the reference was observed, the time to reach the char density was confirmed to be within about 30 s of the reference.

The cause of the depth-wise delay in density reduction is that as the material thickness increases, the time required for complete pyrolysis becomes longer and the pyrolysis process slows down. The pyrolysis reaction of the resin is affected not only by temperature but also by the rate of temperature rise; the closer it is to the imposed heat flux, the higher the rate of temperature rise, which accelerates the pyrolysis rate and shortens the time to complete pyrolysis [14].

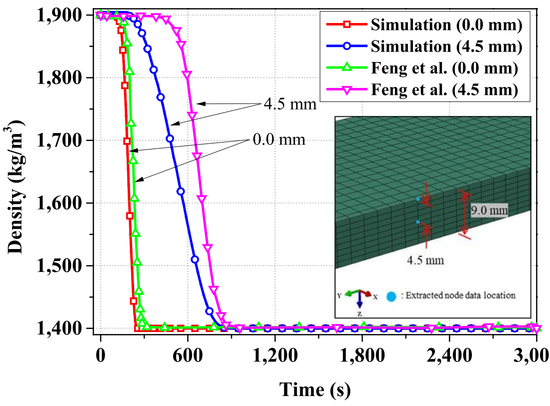


Fig. 3. Comparison of density-time histories at 0.0 mm and 4.5 mm

V. CONCLUSION

This study reproduced a heat-transfer analysis and performed code verification for a glass fiber/vinyl ester composite plate(= 100 mm × 100 mm × 9 mm) by coupling a Finite Element Analysis program(Abaqus) with user subroutines(UMATHT, DFLUX) under an incident heat flux of 50 kW/m² for a simulation time of 3,000 s. The computed temperature histories exhibited a rapid rise at the beginning of heating followed by a gradual convergence, and at the end of the analysis the temperatures at the model top surface and mid-plane reproduced the tendency to converge to the literature values of 615.89 °C at the top surface and 412.55 °C at the mid-plane. The density change followed a typical depth-wise delay behavior in which early char formation occurs at the surface and, after a delay at 4.5 mm, complete char is reached; the present analysis reproduced, similar to the literature, the trends in the onset of pyrolysis and the time to complete pyrolysis at the specimen top surface. However, a delay relative to the literature was observed for the onset time of pyrolysis at the mid-plane. This delay is consistent with the literature's explanation that resin pyrolysis is influenced by both temperature and the rate of temperature rise, and that the reaction slows as the distance from the heated face increases because the temperature rise rate becomes lower.

Therefore, the analysis presented in this study reproduces the temperature histories and density-reduction rates of the glass fiber/vinyl ester composite at the literature level, thereby securing numerical validity, and it provides a verification basis for screening the heat-transfer behavior of thermal-protection materials for high-temperature drone outer skins and for comparing thickness options.

ACKNOWLEDGMENT

This work was supported by the BK21 FOUR program through the National Research Foundation of Korea (NRF) funded by the Korean government (grant number: 2120240715457).

REFERENCES

[1] D. Aláez, M. Prieto, J. Villadangos, and J. J. Astrain, "Towards a heat-resistant tethered micro-aerial vehicle for structure fire sensing," *Appl. Sci.*, vol. 15, no. 5, Art. no. 2388, 2025.

[2] J. V. Palomba and M. MScEcon, "Unmanned aerial vehicle inspections and environmental benefits," in *Proc. 15th Asia Pacific Conf. for Non-Destructive Testing*, 2017, pp. 13–17.

[3] M. Rizia, J. A. Reyes-Munoz, A. G. Ortega, A. Choudhuri, and A. Flores-Abad, "Autonomous aerial flight path inspection using advanced manufacturing techniques," *Robotica*, vol. 40, no. 7, pp. 2128–2151, 2022.

[4] U. Tsunogai, R. Shingubara, Y. Morishita, M. Ito, F. Nakagawa, S. Yoshikawa, et al., "Sampling volcanic plume using a drone-borne SelPS for remotely determined stable isotopic compositions of fumarolic carbon dioxide," *Frontiers in Earth Science*, vol. 10, Art. no. 833733, 2022.

[5] S. Thivet, G. Bagheri, P. M. Kornatowski, A. Fries, J. Lemus, R. Simionato, et al., "In-situ volcanic ash sampling and aerosol-gas analysis based on UAS technologies (AeroVolc)," *Atmospheric Measurement Techniques Discussions*, 2024, pp. 1–33.

[6] K. Abdurohman, N. L. Muzayadah, A. Nugroho, T. Satrio, and R. S. Aritonang, "A comparison between polyester and vinylester matrice toward e-glass WR 185 polymer composites for unmanned aerial vehicle (LSU 05 NG)," in *AIP Conf. Proc.*, vol. 2226, no. 1, Art. no. 040005, Apr. 2020.

[7] T. M. McBride, "Materials for the general aviation industry: Effect of environment on mechanical properties of glass fabric/rubber toughened vinyl ester laminates," *Langley Aerospace Research Summer Scholars Program*, 1995.

[8] G. Chen, X. Bai, Z. Mei, Y. Yu, and H. Li, "Study on thermal response of glass fiber reinforced composite laminates considering material damage," *Thermochimica Acta*, vol. 738, Art. no. 180055, 2025.

[9] H. E. Goldstein, "Pyrolysis kinetics of nylon 6–6, phenolic resin, and their composites," *J. Macromol. Sci., Part A: Chem.*, vol. 3, no. 4, pp. 649–673, 1969.

[10] J. B. Henderson, J. A. Wiebelt, and M. R. Tant, "A model for the thermal response of polymer composite materials with experimental verification," *J. Composite Mater.*, vol. 19, no. 6, pp. 579–595, 1985.

[11] H. Li, N. Wang, X. Han, B. Fan, Z. Feng, and S. Guo, "Simulation of thermal behavior of glass fiber/phenolic composites exposed to heat flux on one side," *Materials*, vol. 13, no. 2, Art. no. 421, 2020.

[12] Z. Feng, B. Fan, N. Wang, X. Han, H. Li, and J. Wu, "Numerical simulation of thermal response of glass fiber/vinyl ester based on UMATHT subroutine," *Materials Reports*, vol. 35, no. 2, pp. 2191–2198, 2021. (in Chinese)

[13] S. Feih, Z. Mathys, A. G. Gibson, and A. P. Mouritz, "Modelling the tension and compression strengths of polymer laminates in fire," *Compos. Sci. Technol.*, vol. 67, nos. 3–4, pp. 551–564, 2007.

[14] M. R. Tant, J. B. Henderson, and C. T. Boyer, "Measurement and modelling of the thermochemical expansion of polymer composites," *Composites*, vol. 16, no. 2, pp. 121–126, 1985.

Ga_2O_3 Schottky Barrier Diodes as Dual Platforms for X-ray Detection and Natural Material Shielding

Wan Sik Hwang
Department of Materials Science
and Engineering, Department of
Smart Air Mobility
Korea Aerospace University
Goyang, South Korea
whwang@kau.ac.kr

***Abstract*—This work demonstrates the versatility of β-Ga_2O_3 Schottky barrier diode (SBD) detectors through two complementary applications: (i) discrimination of stacked A4 paper sheets under low-dose X-ray irradiation, and (ii) evaluation of soft X-ray shielding performance of natural wood films with different densities. The fabricated Ga_2O_3 SBDs exhibited stable diode characteristics with low leakage current and high rectification ratios (>10^6). In the A4 sheet study, the detectors achieved a maximum sensitivity of 820 $\mu C\cdot Gy^{-1}\cdot cm^{-2}$ at −10 V and successfully distinguished up to seven stacked sheets, with attenuation behavior following the Beer–Lambert law25apl_Xray A4 Ga2O3. In the wood study, shielding efficiency increased monotonically with density, with the densest sample (0.91 g/cm³) achieving >97% attenuation at room temperature and retaining >85% efficiency at 423 K. These results highlight Ga_2O_3 SBDs not only as high-sensitivity detectors but also as robust platforms for evaluating eco-friendly, lightweight shielding materials suitable for aerospace and radiation-sensitive environments.**

***Keywords*—*Ga_2O_3* Schottky diode, X-ray detection, cellulose attenuation, wood shielding, wide-bandgap semiconductors**

I. INTRODUCTION

Reliable X-ray detection and shielding are essential for applications spanning medical imaging, security inspection, and aerospace environments. Conventional semiconductor detectors such as Si and Ge face limitations due to their narrow bandgaps, which result in high leakage currents and reduced sensitivity under low-dose or high-temperature conditions. GaO, an ultra-wide bandgap semiconductor (Eg ≈ 4.8 eV), offers intrinsic advantages including low intrinsic carrier concentration, strong radiation hardness, and excellent thermal stability. These characteristics make Ga_2O_3-based Schottky barrier diodes promising candidates for next-generation X-ray detection. In this study, we present two complementary approaches to highlight the capabilities of GaO detectors: Thin cellulose discrimination through stacked A4 sheets, relevant to document authentication and low-dose security screening. Radiation shielding evaluation of natural wood films, targeting sustainable aerospace applications where weight and thermal resilience are critical

II. EXPERIMENTAL METHODS

A. Device Fabrication

The detectors were fabricated using HVPE-grown β-Ga_2O_3 epitaxial layers (11 μm, Si-doped ~1.9×10^{16} cm^{-3}) on Sn-doped substrates (~5×10^{18} cm^{-3}). Ohmic contacts (Ti/Au, 20/100 nm) were annealed at 500 °C in N_2, while Schottky contacts (Pt/Ti/Au, 20/10/100 nm) were patterned into circular electrodes (220–320 μm) which is shown in Fig. 1.

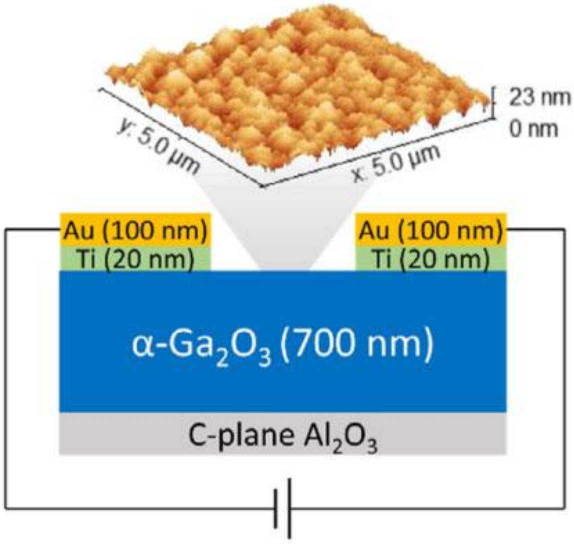


FIG. 1. Schematic structure of the vertical β-Ga_2O_3 Schottky barrier diode (SBD). The device consists of an Si-doped β-Ga_2O_3 epitaxial layer (~11 μm thick) grown on an Sn-doped β-Ga_2O_3 substrate (~650 μm thick). The top Schottky contact is formed with a multilayer metal stack of Au/Ti/Pt, while the bottom ohmic contact is composed of Ti/Au.

B. *X-ray Exposure Setup*

Soft X-rays were generated at 5.0 kVp and 0.6 mA (dose rate: 3.84 $\mathrm{mGy_{air}\cdot s^{-1}}$, AweXome Ray Inc.). Cellulose test: commercial A4 paper sheets (~100 μm thick, 80 gsm) stacked up to 10 layers. Wood test: thin films (1 mm) of four wood species with bulk densities ranging from 0.29 to 0.91 g/cm^3.

C. *Measurement and Analysis*

Electrical responses were measured using a Keithley 4200 SCS system. Attenuation efficiencies were derived from photocurrent suppression and modeled using the Beer–Lambert law for cellulose, and density-dependent linear fitting for wood films. Fig. 2 illustrates the generation and impact of low-energy X-ray radiation in enclosed aerospace environments and compares the corresponding radiation dose levels encountered under various orbital conditions. As shown in Fig. 2, electronic components onboard spacecraft can emit low-energy X-rays as a result of power switching, high-voltage discharge, or interaction with high-energy particles. These radiations, even at milligray-level intensities, can induce photocurrent noise in highly sensitive sensors and semiconductor devices, thereby degrading the signal-to-noise ratio and long-term reliability of onboard instruments.

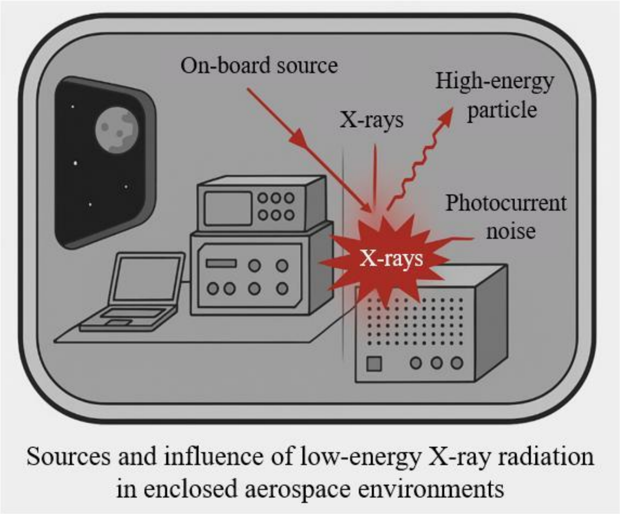


FIG. 2. Illustration of low-energy X-ray radiation effects in aerospace environments. Schematic showing possible sources of X-rays and high-energy particles inside enclosed aerospace systems, which can induce photocurrent noise in onboard electronic instrument.

III. RESULTS AND DISCUSSION

A. *Discrimination of A4 Sheets*

The Ga_2O_3 SBD exhibited well-behaved rectification and low dark current. X-ray sensitivity increased with reverse bias, peaking at 820 $\mu C\cdot Gy^{-1}\cdot cm^{-2}$ at −10 V. Attenuation efficiency followed an exponential decay with sheet thickness, enabling reliable discrimination up to seven layers. Temporal response remained stable, with rise and fall times of 0.15 and 0.13 s, respectively, regardless of attenuation level.

B. *Shielding Effectiveness of Wood Films*

Shielding efficiency strongly correlated with density: photocurrent suppression increased linearly from 0.29 to 0.91 g/cm^3. The densest wood sample reached shielding performance comparable to lightweight metallic materials. At elevated temperatures (423 K), high-density wood retained >85% efficiency, while Ga_2O_3 detectors themselves remained thermally stable. The observed degradation at high temperatures was attributed to microstructural changes (e.g., pore formation, delamination), not solely to density reduction.

C. *Combined Insights*

The A4 experiments validated Ga_2O_3 SBDs as precise probes of thin cellulose attenuation, while the wood study demonstrated their utility in evaluating sustainable shielding materials. Together, these findings show that Ga_2O_3 devices can function as both high-performance detectors and characterization platforms for novel materials.

Figure 3 shows the prototype measurement setup developed to evaluate the photoresponse of the Ga_2O_3 SBDs-based detector under X-ray exposure. The SBD was connected to an Arduino Uno microcontroller through a breadboard interface that enabled current signal acquisition and LED-based visualization. Two 9 V batteries provided the bias voltage and power supply for the Arduino system. When the detector was illuminated by X-rays light, a distinct change in the LED intensity was observed, corresponding to the photocurrent generated in the Ga_2O_3 SBD. This setup demonstrates the feasibility of using a compact Arduino-based circuit for real-time detection, calibration, and educational demonstration of semiconductor radiation sensors.

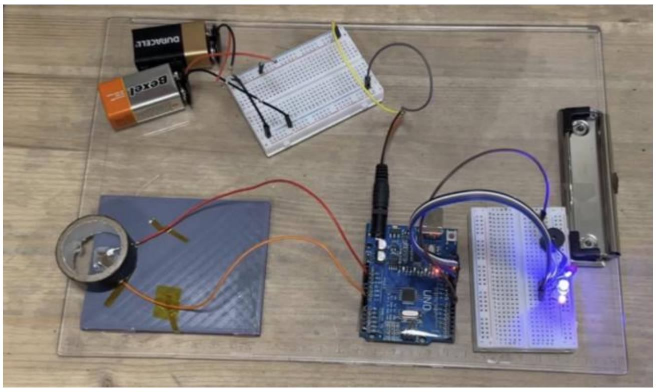


FIG. 3. Experimental setup for the Arduino-based X-ray detector interface. The Ga_2O_3 SBD (left) was connected to an Arduino Uno microcontroller (center) through a signal conditioning circuit on the breadboard. The circuit was

powered by two 9 V batteries and used to measure photocurrent variations induced by incident X-ray illumination. The Arduino output was displayed through LED indicators (right) for visual confirmation of signal response, demonstrating a simple, low-cost detection and monitoring platform.

IV. CONCLUSION

This integrated study demonstrated the dual capability of β-GaO Schottky barrier diodes: High-sensitivity detection of low-dose X-rays and discrimination of thin, low-Z cellulose layers. Robust evaluation of lightweight, density-tunable natural materials as potential secondary shielding layers for aerospace applications. These outcomes suggest broad potential in fields such as document verification, security screening, medical imaging, and radiation management in aerospace systems. Future efforts will focus on optimizing device structures and combining Ga_2O_3 detectors with hybrid shielding composites to further enhance performance.

ACKNOWLEDGMENT

The author thank AweXome Ray Inc. for providing X-ray source equipment. This work was supported by the BK21 FOUR program through the National Research Foundation of Korea (NRF) funded by the Korean government (grant number: 5199990714521). This work was supported by the K-Sensor Development Program(no. RS-2022-00154729) funded by the Ministry of Trade, Industry and Energy (MOTIE, Republic of Korea).

14th ICCAIS 2025
October 27-29,
Jeju, Republic of Korea

ISPL Intelligent Sensing & Perception Laboratory
season
IEEE
EZMEDIBOT
Curtin University
frontiers
KAU ROBOT RESEARCH CENTER
AIBIZ
KAU Korea Aerospace University
BrainKorea21 FOUR