# INTELLI 2024

The Thirteenth International Conference on Intelligent Systems and Applications

March 10th –14th, 2024

Athens, Greece

**INTELLI 2024 Editors**

Carsten Behn, Hochschule Schmalkalden, Germany

Vítor H. Pinto, FEUP, Portugal

# INTELLI 2024

# Forward

The Thirteenth International Conference on Intelligent Systems and Applications (INTELLI 2024), held between March 10th and March 14th, 2024, continued a series of international events on advances towards fundamental, as well as practical and experimental aspects of intelligent and applications.

The information surrounding us is not only overwhelming but also subject to limitations of systems and applications, including specialized devices. The diversity of systems and the spectrum of situations make it almost impossible for an end-user to handle the complexity of the challenges. Embedding intelligence in systems and applications seems to be a reasonable way to move some complex tasks from user duty. However, this approach requires fundamental changes in designing the systems and applications, in designing their interfaces and requires using specific cognitive and collaborative mechanisms. Intelligence becomes a key paradigm, and its specific use takes various forms according to the technology or the domain a system or an application belongs to.

We take here the opportunity to warmly thank all the members of the INTELLI 2024 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to INTELLI 2024. We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the INTELLI 2024 organizing committee for their help in handling the logistics of this event.

We hope that INTELLI 2024 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the area of intelligent systems and applications.

**INTELLI 2024 Chairs**

**INTELLI 2024 Steering Committee**
Carsten Behn, Schmalkalden University of Applied Sciences, Germany
Stefano Berretti, University of Firenze, Italy
Marcin Paprzycki, Systems Research Institute, Polish Academy of Sciences – Warszawa, Poland
Gil Gonçalves, University of Porto, Portugal

**INTELLI 2024 Publicity Chairs**
José Miguel Jiménez, Universitat Politecnica de Valencia, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain

# INTELLI 2024
## Committee

**INTELLI 2024 Steering Committee**

Carsten Behn, Schmalkalden University of Applied Sciences, Germany
Stefano Berretti, University of Firenze, Italy
Marcin Paprzycki, Systems Research Institute, Polish Academy of Sciences – Warszawa, Poland
Gil Gonçalves, University of Porto, Portugal

**INTELLI 2024 Publicity Chairs**

José Miguel Jiménez, Universitat Politecnica de Valencia, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain

**INTELLI 2024 Technical Program Committee**

Azizi Ab Aziz, Universiti Utara Malaysia, Malaysia
Lounis Adouane, Université de Technologie de Compiègne, France
Leo Aguilera, 33 Technologies LLC, USA
Ari Aharari, SOJO University, Japan
Bilal Ahmad, University of Warwick, UK
Zaher Al Aghbari, University of Sharjah, UAE
Miltos Alamaniotis, University of Texas at San Antonio, USA
Antonios Alexos, University of California Irvine, USA
Sarra Alqahtani, Wake Forest University, USA
Rachid Anane, Coventry University, UK
Olugbenga Moses Anubi, Florida State University, USA
Arvind Bansal, Kent State University, USA
Suzanne Barber, The University of Texas at Austin, USA
Dariusz Barbucha, Gdynia Maritime University, Poland
Carmelo Bastos-Filho, University of Pernambuco, Brazil
Rafael Batres, Tecnologico de Monterrey, Mexico
Carsten Behn, Schmalkalden University of Applied Sciences, Germany
Fayçal Bensaali, Qatar University, Qatar
Lyes Benyoucef, Aix-Marseille University, France
Giuseppe Berio, IRISA | Université de Bretagne Sud, France
Stefano Berretti, University of Firenze, Italy
Jonathan Berrisch, University of Duisburg-Essen, Germany
Mahdis Bisheban, National Research Council Canada (NRC), Canada
José Miguel Blanco, Masaryk University, Brno, Czech Republic
Francisco Bonin Font, University of the Balearic Islands, Spain
Lucas Botoni de Souza, Federal University of Technology - Paraná, Brazil
Frederic Bousefsaf, LCOMS | Université de Lorraine, France
Simeon C. Calvert, Delft University of Technology, Netherlands
Valérie Camps, Paul Sabatier University | IRIT, Toulouse, France
Carlos Carrascosa, Universitat Politècnica de València, Spain

Cesar Castelo-Fernandez, Institute of Computing | University of Campinas, Brazil
Martin Cech, University of West Bohemia, Pilsen, Czech Republic
Coskun Cetinkaya, Kennesaw State University, USA
Chin-Chen Chang, Feng Chia University, Taiwan
Tongwen Chen, University of Alberta, Canada
Guilherme Conde, Federal University of Western Pará, Brazil
Angelo Croatti, University of Bologna, Italy
Daniela D'Auria, Free University of Bozen-Bolzano, Italy
Mohammed Dahane, Université de Lorraine, France
Robertas Damaševičius, Silesian University of Technology, Poland
Chuangyin Dang, City University of Hong Kong, Hong Kong
Andrea D'Ariano, Roma Tre University, Italy
Jos De Brabanter, KU Leuven ESAT-STADIUS, Belgium
Toon De Pessemier, imec - WAVES - Ghent University, Belgium
Angel P. del Pobil, Jaume I University, Spain
Jens Dörpinghaus, Federal Institute for Vocational Education and Training (BIBB) / German Center for Neurodegenerative Diseases (DZNE), Germany
Paweł Drąg, Wrocław University of Science and Technology, Poland
Nelson Duarte, CIICESI | ESTG | Politécnico do Porto, Portugal / IRIEM, Hong Kong
Arianna D'Ulizia, National Research Council - IRPPS, Italy
Ahmed Ewais, Arab American University, Jenin, Palestine / Vrije Universiteit Brussel, Belgium
Tullio Facchinetti, University of Pavia, Italy
Ana Fernández Vilas, School of Telecommunication Engineering | University of Vigo, Spain
Stefka Fidanova, IICT-BAS, Sofia, Bulgaria
Manuel Filipe Santos, University of Minho, Portugal
Edgar Giovanny Cuzco Silva, Universidad Nacional de Chimborazo, Ecuador
Todorka Glushkova, Plovdiv University "Paisii Hilendarski", Bulgaria
Helder Gomes Costa, Universidade Federal Fluminese (UFF), Brazil
Gil Gonçalves, University of Porto, Portugal
Sérgio Gorender, Federal University of Bahia, Brazil
Javier Gozalvez, Universidad Miguel Hernandez de Elche, Spain
Emmanuelle Grislin-Le Strugeon, LAMIH | Université Polytechnique Hauts-de-France (UPHF), France
Yousif A. Hamad, Imam Ja'afar Al-Sadiq University, Iraq / Siberian Federal University, Russia
Ibrahim A. Hameed, Norwegian University of Science and Technology (NTNU), Norway
Wahida Handouzi, Tlemcen University, Algeria
Ridewaan Hanslo, University of Pretoria, South Africa
Wladyslaw Homenda, Warsaw University of Technology, Poland
Tzung-Pei Hong, National University of Kaohsiung, Taiwan
Wei-Chiang Hong, School of Education Intelligent Technology - Jiangsu Normal University, China
Christopher-Eyk Hrabia, Technische Universität Berlin | DAI-Labor, Germany
Chih-Cheng Hung, Kennesaw State University - Marietta Campus, USA
Syed Muhammad Zeeshan Iqbal, BrightWare LLC, Riyadh, Saudi Arabia
Zahid Iqbal, University of Porto, Portugal
Ajune Wanis Ismail, Universiti Teknologi Malaysia, Malaysia
Raheleh Jafari, School of Design | University of Leeds, UK
Anubhav Jain, Telstra, India
Juergen Jasperneite, Fraunhofer IOSB-INA, Germany
Thomas Jell, Siemens Mobility GmbH, Germany

Andrés Jiménez Ramírez, University of Seville, Spain
Maria João Ferreira, Universidade Portucalense, Portugal
Mihaela Juganaru, IMT - Mines de Saint Etienne, France
Janusz Kacprzyk, Systems Research Institute - Polish Academy of Sciences, Poland
Ryotaro Kamimura, Tokai University, Japan
Keiichi Kaneko, Tokyo University of Agriculture and Technology, Japan
Mehdi Kargar, Ted Rogers School of Management | Ryerson University, Canada
Alexey Kashevnik, SPIIRAS, Russia
Okba Kazar, University of Biskra, Algeria
Alireza Khanteymoori, Universitätsklinikum Freiburg, Germany
Leoneed Kirilov, Institute of Information and Communication Technologies - Bulgarian Academy of
Sciences, Bulgaria
Sotiris Kotsiantis, University of Patras, Greece
Boris Kovalerchuk, Central Washington University, USA
Akmaral Kuatbayeva, Astana IT University, Kazakhstan
Tobias Küster, DAI-Labor / Technical University of Berlin, Germany
Victoria Lapuerta, Universidad Politécnica de Madrid, Spain
Antonio LaTorre, Universidad Politécnica de Madrid, Spain
Frédéric Le Mouël, Univ. Lyon / INSA Lyon, France
Deok-Jin Lee, Kunsan National University, South Korea
George Lekeas, City Universty - London, UK
Maurizio Leotta, University of Genova, Italy
Chanjuan Liu, Dalian University of Technology, China
Mingjie Liu, The University of Texas at Austin / Nvidia Corporation, USA
Francesco Longo, University of Calabria, Italy
Daniela López De Luise, CI2S Labs, Argentina
Majdi Maabreh, The Hashemite University, Jordan
Francesca Maridina Malloci, University of Cagliari, Italy
Telmo Matos, Porto School of Engineering (ISEP) | University of Porto (FEUP) | CIICESI (ESTG), Portugal
Harald Mayer, JOANNEUM RESEARCH Forschungsgesellschaft mbH, Austria
René Meier, Hochschule Luzern, Germany
António Meireles, GECAD - Research Group on Intelligent Engineering and Computing for Advanced
Innovation and Development, Portugal
Jérôme Mendes, Institute of Systems and Robotics (ISR-UC), Portugal
Márcio Mendonça, Federal University of Technology - Paraná (UTFPR), Brazil
Tarek Menouer, Umanis Research & Innovation, France
Jair Minoro Abe, Paulista University & Institute of Advanced Studies | University of São Paulo, Brazil
Jose M. Molina, Universidad Carlos III de Madrid, Spain
Vítor Monteiro, University of Minho, Portugal
Ceci Morales, iRobot, USA
Fernando Moreira, Universidade Portucalense Infante D. Henrique, Portugal
Paulo Moura Oliveira, UTAD University, Vila Real / INESC-TEC- Technology and Science, Porto, Portugal
Debajyoti Mukhopadhyay, Mumbai University, India
Muddasar Naeem, ICAR-CNR, Naples, Italy
Filippo Neri, University of Naples, Italy
Pranav Ajeet Nerurkar, NMIMS University, Mumbai, India
Dinh-Luan Nguyen, Michigan State University, USA

Supphachai Thaicharoen, Srinakharinwirot University, Bangkok, Thailand
Pei-Wei Tsai, Swinburne University of Technology, Australia
Berna Ulutas, Eskisehir Osmangazi University, Turkey
Paulo Urbano, Universidade de Lisboa - BioISI, Portugal
Jan Vascak, Technical University of Kosice, Slovakia
Costas Vassilakis, University of the Peloponnese, Greece
Anna-Maria Velentza, University of Macedonia, Thessaloniki, Greece
Minjuan Wang, San Diego State University, USA
Yifei Wang, Georgia Institute of Technology, USA
Kanoksak Wattanachote, Guangdong University of Foreign Study, China
Dietmar Winkler, TU Wien | CDL-SQI, Vienna, Austria
Stefanie Wuschitz, Miss Baltazar's Laboratory, Vienna, Austria
Mudasser F. Wyne, National University, USA
Maria Gabriella Xibilia, University of Messina, Italy
Peng Xu, Technical University of Munich (TUM), Germany
Wenju Xu, Amazon, USA
Longzhi Yang, Northumbria University, UK
Leila Zemmouchi-Ghomari, Ecole Nationale Supérieure de Technologie, ENST, Algiers, Algeria

**Copyright Information**

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

# Table of Contents

# Using a Dexterous Robotic Hand for Automotive Painting Quality Inspection

Bruno Santos, Gil Gonçalves, Vítor H. Pinto
*SYSTEC (DIGI2), ARISE & ECE Dept.*
*Faculdade de Engenharia, Universidade do Porto*
Porto, Portugal
{brunosantos, gil, vitorpinto}@fe.up.pt

Francisco M. Ribeiro
*INESC TEC*
Porto, Portugal
francisco.m.ribeiro@inesctec.pt

*Abstract*—**The rapid evolution of industrial automation has revolutionised industries over the past few decades, with technology replacing manual labour through the integration of industrial robots. This paper delves into the pivotal role of industrial automation in bolstering productivity, curtailing labour costs, and elevating product quality, paramount pursuits for businesses striving for competitiveness. Situated within a "Digitisation of Human Skills" project, this research focuses on automating tasks in the automotive industry, specifically automotive painting quality inspection. The proposed approach combines a Shadow Dexterous Hand and a UR5 robotic arm, controlled through teleoperation with the BioIK kinematic retargeting algorithm. Two methods, OpenPose and MediaPipe, were assessed for acquiring human hand data for teleoperation. The study demonstrates the successful replication of human movements for automotive painting quality inspection utilising BioIK and MediaPipe, underscoring the potential of automation in this critical industrial domain.**

*Keywords*—*Shadow Dexterous Hand; UR5; MediaPipe; BioIK; teleoperation.*

## I. Introduction

The automation of industrial processes has undergone significant evolution in recent decades, fundamentally transforming the operational landscape of the industries. Beginning in the 1960s, advancements in technology have led to the gradual replacement of manual labour with industrial robots, as noted in [1].

In fact, industrial automation has become imperative for companies striving to maintain competitiveness in the contemporary marketplace, owing to its inherent advantages [2]. Chief among these advantages is the substantial increase in productivity [3], attributable to the rapidity of robots compared to human operators, facilitating uninterrupted processes. Additionally, reduced reliance on human labour translates into decreased labour costs and fewer human errors, ultimately enhancing product quality.

As elucidated, this expansive domain of industrial automation holds immense potential and is poised to play a pivotal role in the future of various industries. Consequently, optimising the level of industrial automation should be a primary focus for companies.

This paper is situated within the context of the project "Digitalização da Arte Humana" (Cibertoque), translated as "Digitisation of Human Skills", which aims to develop a robotic system capable of replicating human movements for the purpose of automating certain tasks within the automotive industry, specifically at Stellantis company. The paper's particular focus lies in the realm of automotive painting quality inspection. The research presented constitutes the initial phase of a study whose ultimate goal is to fully automate this task by ideally replacing human operators with robotic systems.

For the accomplishment of the aforementioned task, the proposed approach involves controlling a Shadow Dexterous Hand [4] coupled with a UR5 [5], a collaborative robotic arm. The Shadow Hand consists of an anthropomorphic robotic hand that provides high flexibility of movements due to its 24 joints, specifically engineered to replicate, with the utmost fidelity, the kinematics and precision of the human hand. The control of this robotic set will be executed based on teleoperation, leveraging the kinematic retargeting algorithm BioIK. In order to acquire the human hand data for the teleoperation process, two methods will be explored, namely OpenPose and MediaPipe.

The paper is structured as follows: Section II presents a comprehensive literature review encompassing relevant topics; Section III elucidates the detailed implementation process of the algorithms employed; Section IV showcases the results achieved through the application of the methods discussed in Section III; and lastly, Section V offers conclusions drawn from the work developed in this study and the future work.

## II. Literature Review

In this section, pertinent subjects will be addressed within the context of this paper. For each of the topics, the respective literature review will be presented.

In order to execute the teleoperation of an anthropomorphic robotic hand, various methods have been explored. Some studies have proposed approaches for human hand motion acquisition involving the use of data gloves, as demonstrated in [6] [7] [8], or gloves equipped with passive markers, as documented in [9] [10]. In addition, certain research endeavours have focused on methods for non-anthropomorphic robotic grippers, exemplified by [11] [12].

However, the present study is dedicated to implementing a teleoperation method tailored for an anthropomorphic robotic hand and leveraging vision-based techniques to capture human hand motion, offering distinct advantages in terms of flexibility and freedom. Notably, it eliminates the necessity of wearing

restrictive gloves. This choice was motivated by the alignment of these techniques with the specif ic requirements of automotive painting inspection and their cost-effectiveness when compared to the utilisation of data gloves.

Subsequently, the methods used will be presented and studied, organised into two sections: data acquisition and kinematic retargeting.

### A. Data Acquisition

Within vision-based methods, there are two well-known algorithms, namely OpenPose [13] and MediaPipe [14].

OpenPose, founded on a convolutional neural network, stands out as the pioneering real-time, multi-person system designed to collectively detect 135 human body, hand, facial, and foot keypoints. Its selection is primarily attributed to its high accuracy, even in challenging scenarios characterised by occlusions and cluttered backgrounds. OpenPose's prowess is substantiated by its superior performance on two datasets, namely Max Planck Institute Informatik (MPII) Human Pose [15] and Common Objects in Context (COCO) [16], where it outperformed prior methods, thus attaining a state-of-the-art status. Additionally, OpenPose's real-time capabilities align seamlessly with the requirements of robotic hand teleoperation and it is an open-source method, meaning it is freely available for research utilisation. Furthermore, in [17], an evaluation of pose estimation accuracy was conducted utilising OpenPose in conjunction with a single RGB-D camera, this study reported a good performance by OpenPose, highlighting the standard deviation of the detected keypoints below 3 millimetres, indicating a high level of repeatability.

On the other hand, MediaPipe, developed by Google, also offers a high-fidelity solution for hand and finger tracking, leveraging machine learning algorithms to infer 21 hand keypoints from a single image. Unlike other methods, such as OpenPose, that rely primarily on powerful desktop environments, MediaPipe distinguishes itself by its low computational footprint, making it possible to achieve real-time performance even on a mobile phone and inclusively scaling to multiple hands. This real-time performance was, in fact, tested in three mobile devices, namely Google Pixel 3, Samsung S20 and iPhone 11, achieving impressing processing times, such as 16.1, 11.1 and 5.3 milliseconds, respectively, utilising the "Full" model [14]. The MediaPipe algorithm has demonstrated its efficacy in various robotic applications, such as in [18], where it effectively captured hand positions with good-quality results. This capability has translated into successful robot control by hand gestures, showcasing the algorithm's reliability and practical utility in real-world scenarios.

### B. Kinematic Retargeting

Transitioning to the control of the robotic hand, this subsection undertakes a comprehensive exploration of existing literature on kinematic retargeting algorithms.

Notable among these is BioIK, first introduced in [19]. Then, a clean and high-performance C++ version of the same algorithm was presented in [20].

BioIK is an open-source software package for Robot Operating System (ROS) featuring a bio-inspired optimisation algorithm adept at solving complex optimisation problems, such as inverse kinematics. This algorithm is distinguished by its user-defined weighted goals, offering a high degree of flexibility and control encompassing 18 goal types. These goals include the fundamentals, such as position, orientation, and pose (position and orientation) goals, but also include more complex goals, such as:

- Minimal Displacement Goal: tries to keep each joint angle as close as possible to the previous one;
- Centre Joints Goal: tries to keep each joint centred at the respective joint limits;
- Avoid Joint Limits: similar to Centre Joints Goal;
- Look At Goal: tries to align the orientation of a specified link to a goal position;
- Maximum Distance Goal: tries to keep the position of a specific link within a maximum range from a goal position.

To elucidate the operational efficacy of BioIK, the authors conducted a series of illustrative demonstrations, including one involving the integration of the Shadow Hand with the KUKA LWR 4+ robotic arm. This particular demonstration focused on the application of BioIK to plan the robotic hand and arm motions for turning a wheel button on an audio mixer, as visually represented in Figure 1. For the execution of this task, a tripod grasp configuration was employed, with precise control exerted over the thumb, first finger, and middle finger through the BioIK algorithm. The final trajectory was achieved by combining a set of 200 solutions generated by the BioIK algorithm. This approach resulted in the Shadow Hand's successful manipulation of the wheel button, achieving the desired rotation without any incidence of slippage or unintended disengagement.
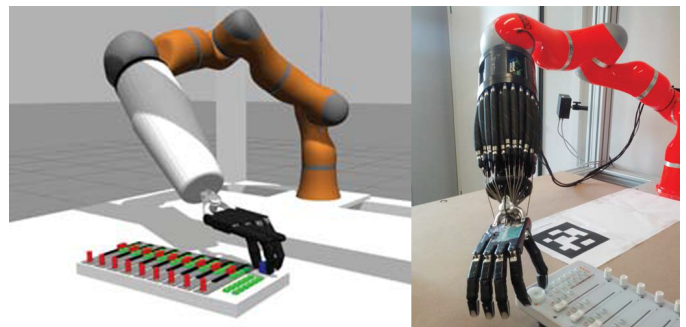


Figure 1. Shadow Hand experiment performed in Gazebo simulation and in reality [20]

In [21], Shuang Li et al. introduced a novel approach to address the complex challenge of kinematic retargeting, leveraging neural network techniques.

In this paper, the authors presented TeachNet, a novel end-to-end neural network architecture. This architecture's primary objective is the estimation of joint angles requisite for replicating the configuration of the human hand within the

robotic domain, specifically with the utilisation of a Shadow Dexterous Hand.

The task of directly solving joint regression problems from human hand images is known to be a particularly challenging one. This difficulty primarily stems from the inherent distinction between the domains occupied by the robot hand and the human hand. In light of this, TeachNet comprises two distinctive branches: the robot hand branch, which assumes the role of a teacher, and the human hand branch, serving as the student. Both branches operate with depth images as input and jointly output the required joint angles. Between these two branches, there is a consistency loss strategically designed to align the corresponding features. To better understand the network architecture, it is illustrated in Figure 2.
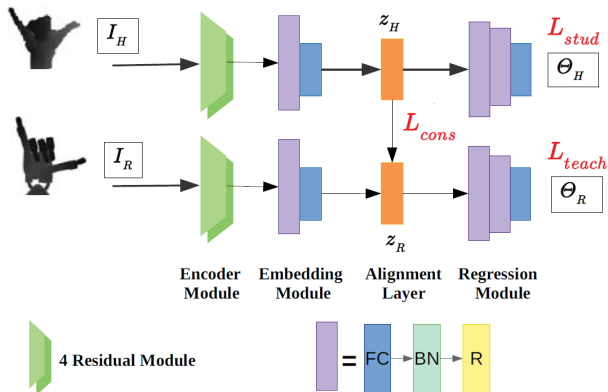


Figure 2. TeachNet architecture [21]

In order to train the developed network, the authors created a new dataset utilising the existing dataset BigHand2.2M [22] and the aforementioned BioIK solver. The process involved the extraction of depth images from the BigHand2.2M dataset and subsequently mapping the corresponding hand keypoints into the corresponding joint angles of the Shadow Hand with BioIK. The solver was set with the following goals: mapping of fingertip positions with a weight of 1, mapping of proximal interphalangeal joint positions with a weight of 0.2, and lastly, mapping proximal and distal phalanges with a weight of 0.2.
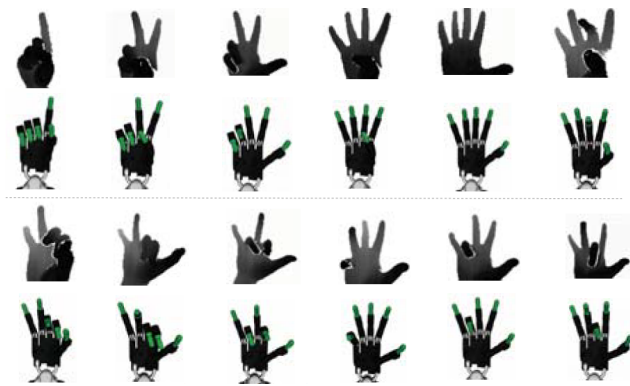


Figure 3. Successful teleoperation results [21]

Finally, to evaluate the efficacy of the proposed method, a series of hand gestures were executed by five novice teleoperators and replicated by Shadow Hand. This test resulted in a success rate of 78.26%, being some of the successful poses presented in Figure 3. Furthermore, manipulation experiments were also performed and compared regarding the necessary time to accomplish a specific task with the DeepPrior++ [23] method. The outcomes underscored the notable efficiency gains achieved through the application of TeachNet, with an average task execution time reduction of 57% compared to the utilisation of DeepPrior++.

## III. IMPLEMENTATION

This section delves into the implementation of the algorithms previously mentioned in Section II.

### A. Data Acquisition

Regarding the data acquisition phase, first, OpenPose was employed together with Microsoft Kinect v2, similarly as in [17]. The Kinect integrates a FullHD (1920 x 1080 pixels) RGB camera and a 512 x 424 pixels resolution depth camera. In addition, it utilises time-of-flight (ToF) technology to capture depth. This method measures distances based on, as the name implies, time-of-flight, i.e., the round trip time of a light signal emitted and then received by the device.

In the case of OpenPose, the Kinect had two primary purposes. Firstly, capturing 2D images of the human hand, these frames were processed by OpenPose to detect the corresponding 2D hand keypoints. Subsequently, the depth information associated with the detected keypoints was extracted also from Kinect, resorting to its depth sensor.

Afterwards, and resulting from the poor results achieved with OpenPose, detailed in Section IV, MediaPipe was tested in conjunction with a stereo-vision ZED camera. The ZED is a stereo camera developed by StereoLabs that captures high-definition images with depth in real-time. The camera utilises two sensors with a baseline of 12 centimetres to mimic the human stereoscopic vision, enabling it to generate a depth map for, ideally, each pixel in the image. The camera has a resolution of up to 2K (2208 x 1242 pixels) and a field of view of 110 degrees horizontally and 60 degrees vertically.

In this specific case, the ZED camera was utilised solely for capturing 2D raw images from both cameras, functioning as two independent cameras. However, the fixed and known baseline, i.e., the distance between the two cameras, provided an advantageous condition for applying stereo-vision techniques. The approach for 3D detection of human hand keypoints involved the independent 2D detection of keypoints for each camera with MediaPipe, followed by the conversion of these 2D positions from both cameras into depth values utilising stereo-vision techniques, following the formula:

$$Depth = \frac{f \cdot B}{d} \qquad (1)$$

where $f$ corresponds to the focal length, $B$ to baseline and $d$ to disparity, this is, the distance in pixels between corresponding keypoints in the left and right images.

### B. Kinematic Retargeting

Moving to the kinematic retargeting algorithm, based on the analysis of the algorithm realised in Section II, the BioIK algorithm was selected for mapping human movements to the robotic system composed of the robotic arm and hand. BioIK was favoured due to its ability to simultaneously control the robotic arm and hand.

Prior to running the solver itself, some processing was applied to the keypoints; namely, the acquired human hand keypoints were transformed in a way that the links (distance between two keypoints, equivalent to bones distance) had the same size as the respective ones in the Shadow Hand. For this, an iterative process was performed, starting from the wrist and ending at each fingertip. This process consisted of maintaining the direction of each link and changing its size according to the respective link in the robotic hand. Being $K$ a hand keypoint and $d_{link}$ the Shadow Hand corresponding link distance between two keypoints, the mentioned iterative process can be described by:

$$K_{i+1} = K_i + d_{link} \cdot \frac{\overrightarrow{K_i\,K_{i+1}}}{||\overrightarrow{K_i\,K_{i+1}}||} \qquad (2)$$

The final objective of applying this technique was to get better results since the BioIK solver is based on inverse kinematics. The result of the application of this technique is shown in Figure 4. Being Shadow Hand designed to be the same size as a human hand, the difference between keypoints is not substantial, except for the little finger since all the robotic hand primary fingers have the same size, but in reality, the little finger is smaller than the other primary fingers.
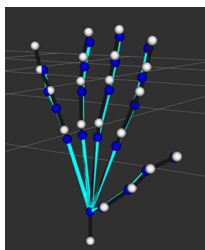


Figure 4. Keypoints before (in blue) and after (in white) the application of robotic hand mapping

Considering the implementation of BioIK itself, a total of 6 different goal types were utilised. The respective goals and their weights are detailed below:

- Position goals for fingertips mapping with weights of 1.0;
- Position goals for knuckles mapping with weights of 0.2;
- Position goals for wrist mapping with a weight of 0.25;
- Direction goals for proximal phalanges of the primary fingers with weights of 0.1;

- Direction goals for intermediate and distal phalanges of the thumb with weights of 0.1;
- Joint function goals for dealing with the coupled joints with a weight of 1.0;
- Centre joints goal with a weight of 0.1;
- Minimal displacement goal with a weight of 0.15;
- Joint function goal to ensure a maximum wrist extension with a weight of 0.5;
- Joint function goal to ensure a workspace above the base plane of UR5 with a weight of 0.15.

It should be noted that the penultimate goal mentioned aims to prevent collisions between the robotic hand's forearm and the bonnet where the paint quality inspection took place. Considering the physical properties of the Shadow Hand, if the hand and respective forearm were horizontally aligned, i.e., with null flexion of the wrist, the forearm would collide with the bonnet.

In summary, the procedure for performing this task involved capturing the keypoints of the human hand and subsequently providing them to the BioIK algorithm, which aims to determine the positions of each joint in the robotic arm and hand, thereby replicating the previously recorded movement, including the intrinsic hand movements and its spatial position. Lastly, these joints' positions were sent to Shadow Hand and UR5, resorting to SrRobotCommander, a high-level interface to control the robotic set.

The algorithms were implemented on a computer with the following specifications:

- Central Processing Unit (CPU): AMD Ryzen™ 9 7950X3D
- Random Access Memory (RAM): 32.0 GB
- Graphics Processing Unit (GPU): NVIDIA GeForce RTX 4090 24GB

### IV. RESULTS

This section presents and discusses the outcomes obtained with the deployment of the algorithm described in Section III.

Commencing with the data acquisition, MediaPipe has shown a notable superior performance relative to OpenPose.

During testing with OpenPose algorithm, it was observed that specific keypoints, in particular the fingertips, were frequently detected incorrectly, leading to significant deviations from their actual positions. After further examination, it was found that these deviations were due to the integration with Microsoft Kinect. A slight difference between the actual fingertip and its detection by OpenPose was enough for the depth value extracted from the Kinect to deviate from the fingertip and correspond, for example, to the background, thus inevitably resulting in a keypoint position measurement quite dissimilar to the actual one. To minimise the impact of this issue, a median filter was applied to the acquired keypoints position. This filter was chosen due to the low influence of outliers on the filtered result. Figure 5 illustrates the effectiveness of the median filter in removing outliers, such as the wrongly detected index fingertip by OpenPose.
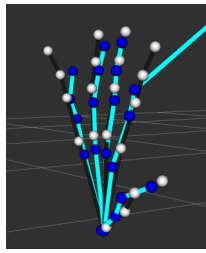
Figure 5. Hand keypoints before (in blue) and after (in white) the application of the median filter

In addition to this initial problem, OpenPose proved to be highly unstable. Frequently, this software incorrectly detected some of the fingers, as can be seen in Figure 6 by the change in the characteristic detection colour of each finger.
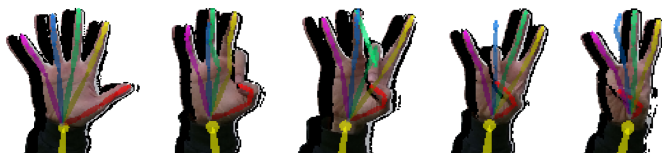


Figure 6. OpenPose misdetections (left image for reference)

Considering these results, in addiction to the fact that OpenPose is extremely computationally heavy compared with MediaPipe, this last method was chosen.

Moving on to the BioIK kinematic retargeting algorithm, its execution time has been experimentally verified as 200 milliseconds, with an end-to-end time delay of about 1 second. When practically applied, it was possible to infer that it behaved as initially expected, replicating with a fair degree of precision the movements made by the human hand.

Once both the acquisition method and kinematic retargeting algorithm were well defined, experimental tests were conducted to evaluate the developed algorithm.
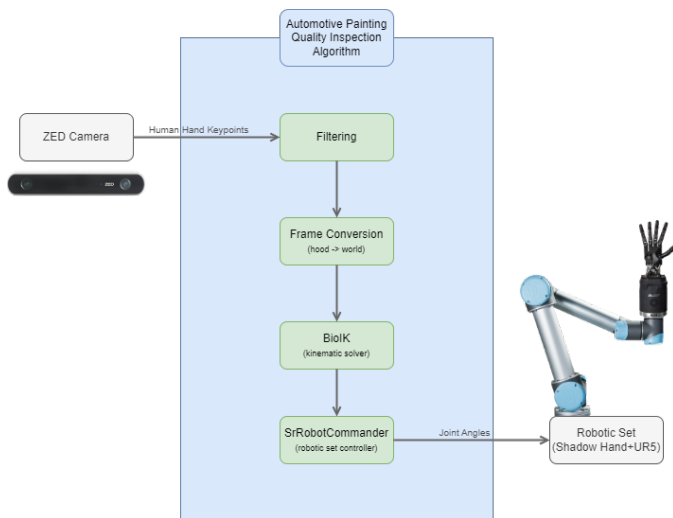


Figure 7. Automotive painting quality inspection algorithm diagram

The complete algorithm sequence, presented in Figure 7, for the accomplishment of the automotive painting quality inspection, includes the following steps:

1) Capture human hand images with a ZED camera;
2) Acquire 3D hand keypoints positions, utilising MediaPipe and stereo-vision techniques, according to a referential frame positioned on the car bonnet;
3) Change the referential frame from the car bonnet to the robot frame;
4) Calculate the robotic set joints' angles utilising BioIK;
5) Lastly, send the calculated joints to the robotic set.

Throughout the various iterations of tests carried out, slight differences (mostly less than a centimetre) were observed in the depth component, normal to the surface, of the acquired points. Consequently, in some instances, the robotic hand exhibited slight deviations from the surface.



Figure 8. Automotive painting quality inspection process

Nonetheless, despite these minor discrepancies, the algorithm proved effective in replicating human movements, as evidenced by the series of frames presented in Figure 8, extracted from a demonstration video [24] showcasing the algorithm's execution.

## V. CONCLUSION AND FUTURE WORK

The research undertaken in this study focused on advancing the automation of quality inspection in automotive painting, specifically by replicating human movements through the integration of an anthropomorphic robotic hand, Shadow Dexterous Hand, and a UR5 robotic arm. Through the exploration of kinematic retargeting techniques, BioIK emerged as the preferred algorithm due to its efficient control over both the robotic arm and hand.

In the process, the integration of OpenPose and MediaPipe for data acquisition revealed notable insights. Despite initial considerations, OpenPose faced challenges related to precision and stability, leading to its exclusion from the final methodology. These challenges served as valuable lessons in the selection of appropriate technologies and highlighted the importance of robust data acquisition methods in achieving accurate task replication.

The exclusion of OpenPose prompted a reevaluation of alternative options for data acquisition. MediaPipe, with its more stable performance, was successfully integrated.

Regarding the automotive painting quality inspection task, thanks to the utilisation of the BioIK together with MediaPipe and stereo-vision techniques, successful replication of movements resembling those performed by a human operator during the paint quality inspection task was achieved.

As future work, it would be very interesting to integrate more sophisticated and especially more suitable tactile sensors into the robotic hand, particularly at the fingertips, capable of providing detailed and nuanced feedback. This integration would elevate the precision and effectiveness of the automated automotive painting quality inspection system.

## REFERENCES

[1] Y. Zhao, "Current status and industrialization development of industrial robot technology," vol. 81, 2021, pp. 804–808, doi: 10.1007/978-3-030-79197-1_117.

[2] M. Bartoš *et al.*, "An overview of robot applications in automotive industry," *Transportation Research Procedia*, vol. 55, pp. 837–844, 2021, doi: 10.1016/j.trpro.2021.07.052.

[3] C. Rosa, F. Silva, and L. P. Ferreira, "Improving the quality and productivity of steel wire-rope assembly lines for the automotive industry," *Procedia Manufacturing*, vol. 11, pp. 1035–1042, 2017, doi: 10.1016/j.promfg.2017.07.214.

[4] Shadow Robot Company, London, GBR, *Shadow Dexterous Hand Technical Specification*, 2021, [Online]. Available from: https://www.shadowrobot.com/wp-content/uploads/2022/03/shadow_dexterous_hand_e_technical_specification.pdf 2024.02.05.

[5] Universal Robots, Odense, DNK, *UR5 Technical specifications*, 2016, [Online]. Available from: https://www.universal-robots.com/media/50588/ur5_en.pdf 2024.02.05.

[6] B. Fang, F. Sun, H. Liu, and D. Guo, "A novel data glove using inertial and magnetic sensors for motion capture and robotic arm-hand teleoperation," *Industrial Robot*, vol. 44, p. 155 – 165, 2017, doi: 10.1108/IR-07-2016-0179.

[7] J. Nassour *et al.*, "A robust data-driven soft sensory glove for human hand motions identification and replication," *IEEE Sensors Journal*, vol. 20, pp. 12 972–12 979, 2020, doi: 10.1109/JSEN.2020.3001982.

[8] Q. Zhu *et al.*, "Teleoperated grasping using a robotic hand and a haptic-feedback data glove," in *2020 IEEE International Systems Conference (SysCon)*, 2020, pp. 1–7, doi: 10.1109/SysCon47679.2020.9275927.

[9] A. Fabisch *et al.*, "A modular approach to the embodiment of hand motions from human demonstrations," in *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*, 2022, pp. 801–808, doi: 10.1109/Humanoids53995.2022.10000165.

[10] Y. Lee *et al.*, "Visual-inertial hand motion tracking with robustness against occlusion, interference, and contact," *Science Robotics*, vol. 6, 2021, doi: 10.1126/scirobotics.abe1315.

[11] C. Meeker, M. Haas-Heger, and M. Ciocarlie, "A continuous teleoperation subspace with empirical and algorithmic mapping algorithms for nonanthropomorphic hands," *IEEE Transactions on Automation Science and Engineering*, vol. 19, pp. 373–386, 2022, doi: 10.1109/TASE.2020.3035156.

[12] X. Bai *et al.*, "Kinect-based hand tracking for first-person-perspective robotic arm teleoperation," in *2018 IEEE International Conference on Information and Automation (ICIA)*, 2018, pp. 684–691, doi: 10.1109/ICInfA.2018.8812561.

[13] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, p. 172 – 186, 2021, doi: 10.1109/TPAMI.2019.2929257.

[14] F. Zhang *et al.*, "Mediapipe hands: On-device real-time hand tracking," 2020, doi: 10.48550/arXiv.2006.10214.

[15] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2d human pose estimation: New benchmark and state of the art analysis," 2014, p. 3686 – 3693, doi: 10.1109/CVPR.2014.471.

[16] T.-Y. Lin *et al.*, "Microsoft coco: Common objects in context," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8693 LNCS, p. 740 – 755, 2014, doi: 10.1007/978-3-319-10602-1_48.

[17] F. Lygerakis, A. C. Tsitos, M. Dagioglou, F. Makedon, and V. Karkaletsis, "Evaluation of 3d markerless pose estimation accuracy using openpose and depth information from a single rgb-d camera," 2020, p. 533 – 538, doi: 10.1145/3389189.3398005.

[18] S. Sreenath, D. I. Daniels, A. S. D. Ganesh, Y. S. Kuruganti, and R. G. Chittawadigi, "Monocular tracking of human hand on a smart phone camera using mediapipe and its application in robotics," 2021, doi: 10.1109/R10-HTC53172.2021.9641542.

[19] S. Starke, N. Hendrich, S. Magg, and J. Zhang, "An efficient hybridization of genetic algorithms and particle swarm optimization for inverse kinematics," 2016, p. 1782 – 1789, doi: 10.1109/ROBIO.2016.7866587.

[20] P. Ruppel, N. Hendrich, S. Starke, and J. Zhang, "Cost functions to specify full-body motion and multi-goal manipulation tasks," 2018, p. 3152 – 3159, doi: 10.1109/ICRA.2018.8460799.

[21] S. Li *et al.*, "Vision-based teleoperation of shadow dexterous hand using end-to-end deep neural network," vol. 2019-May, 2019, p. 416 – 422, doi: 10.1109/ICRA.2019.8794277.

[22] S. Yuan, Q. Ye, B. Stenger, S. Jain, and T.-K. Kim, "Bighand2.2m benchmark: Hand pose dataset and state of the art analysis," vol. 2017-January, 2017, p. 2605 – 2613, doi: 10.1109/CVPR.2017.279.

[23] M. Oberweger and V. Lepetit, "Deepprior++: Improving fast and accurate 3d hand pose estimation," vol. 2018-January, 2017, p. 585 – 594, doi: 10.1109/ICCVW.2017.75.

[24] https://youtu.be/lpxIPF_6WYc 2024.02.05.

# Generation of Captions Highlighting the Differences between a Clothing Image Pair with Attribute Prediction

Kohei Abe
*Graduate School of Information Science and Technology,*
*Hokkaido University*
Sapporo, Hokkaido, Japan
email: ko.abe@ist.hokudai.ac.jp

Soichiro Yokoyama
*Faculty of Information Science and Technology,*
*Hokkaido University*
Sapporo, Hokkaido, Japan
email: yokoyama@ist.hokudai.ac.jp

Tomohisa Yamashita
*Faculty of Information Science and Technology,*
*Hokkaido University*
Sapporo, Hokkaido, Japan
email: yamashita@ist.hokudai.ac.jp

Hidenori Kawamura
*Faculty of Information Science and Technology,*
*Hokkaido University*
Sapporo, Hokkaido, Japan
email: kawamura@ist.hokudai.ac.jp

*Abstract*—**Detailed information for comparisons between products is necessary in consumers' product purchasing process, especially during the information search and choice evaluation phases. However, conventional product descriptions, which are the main source of information, tend to focus only on the product in question, and thus do not adequately express the differences between products. To solve this problem, garments are treated as target products, and a caption-generation method that emphasizes the differences between pairs of garment images using a deep-learning model for image caption-generation is proposed and its effectiveness verified. The proposed method selects and outputs captions that express differences in features from a set of captions generated for input-garment image pairs. Subject experiments confirmed that the proposed method accurately represented the feature differences between garments and provided useful information for consumers to compare garments. In particular, the proposed method is highly effective for garment pairs with similar features.**

*Keywords-deep learning; image captioning; consumer support; information provision.*

## I. INTRODUCTION

In the field of consumer behavior, the sequence of processes involved in the purchase of a product is widely recognized as the purchase decision-making process [1]. This process comprises five stages: problem recognition, information search, alternative evaluation, purchase decisions, and post-purchase evaluation. In the problem recognition phase, consumers identify their needs and problems, and collect information to satisfy them in the information search phase. In the evaluation of alternatives, the consumer compares and evaluates products based on the collected information, and selects and purchases a specific product in the purchase decision stage. In the post-purchase evaluation, the degree of satisfaction was determined based on the results of the product use. During the information search and evaluation of alternatives phase, consumers need detailed information to understand the characteristics and differences of products and make the right choices. This

information can originate from a variety of sources, such as user reviews, expert opinions, and comparison websites; however, product descriptions are one of the most important sources of information that consumers interact with in the early stages of their purchasing decisions. Product descriptions can successfully convey the basic features of a product; however, they tend to focus only on the product in question and do not adequately describe the differences between products. This lack of information may affect consumers' final purchasing decisions and post-purchase evaluations.

Image-caption generation is a research area for generating descriptive text from images; however, it primarily generates a single sentence for a single input image. It is impossible to generate a caption for each image by considering the relationships between multiple images. Some studies have aimed to generate distinctive image captions by comparing input images with similar images in a database; however, they cannot specify the images to be compared, as was the aim of this study.

This study aimed to provide adequate information to consumers when comparing products. As a concrete initial effort towards this goal, a method for generating captions that highlight the differences between two products is proposed and evaluated. Clothing is selected as the target product. Clothing is an everyday purchase for consumers and has various features, such as pattern, material, length, and collar shape. Therefore, consumers need to compare product features during product selection. In the proposed method, two different garment images are independently input into an image-caption generation model to generate multiple captions. Next, the prominence of each attribute in each image is calculated using the garment attribute estimation model and the frequency of occurrence in the caption. This is compared between images, and the caption containing more salient attributes than one image is selected from the multiple captions generated for each

Figure 1. Overview of the proposed method.

| Model | BLEU4 | METEOR |
|-------|-------|--------|
| NIC [5] | 27.7 | 23.7 |
| NICA [8] | 25.0 | 13.9 |
| SCST [9] | 31.9 | 25.5 |
| ClipCap [10] | 33.5 | 27.5 |
| OFA [13] | 44.9 | 32.5 |

## II. RELATED WORK

This section describes the main areas relevant to this study, namely image caption generation, caption generation for multiple images, garment attribute estimation, and garment image caption generation.

### A. Image Caption Generation

Image-caption generation is the task of generating an appropriate description of a single-input image. A comparison of the main image-caption generation models for the benchmark dataset Microsoft Common Objects in Context (MS COCO) [2] is presented in Table I. Bilingual Evaluation Understudy (BLEU) [3] and Metric for Evaluation of Translation with Explicit Ordering (METEOR) [4] are automatic metrics that measure the similarity between the generated and correct captions, with higher values indicating better model performance. Vinyals et al. [5] proposed a model based on a deep recurrent architecture that combines a Convolutional Neural Network (CNN) [6] and Long Short Term Memory (LSTM) [7]. Subsequently, Xu et al. [8] introduced an attention mechanism that focused on specific regions in an image when generating different words. Furthermore, Rennie et al. [9] proposed a model that incorporates reinforcement learning. Recently, image-language pre-training models that learn using large amounts of image-text pair data have achieved higher accuracy than conventional models. Mokady et al. [10] proposed a model that combines the image language pre-training model Contrastive Language–Image Pre-training (CLIP) [11] and the language model Generative Pre-trained Transformer 2 (GPT-2) [12], which reduces training time and achieves highly accurate caption generation. Wang et al. [13] also proposed a pre-training model using 20 million image-text pair data. All these models generate a single-sentence caption for a single input image. In this study, one-sentence captions are generated for each of the two input images. A one-input, one-output image caption generation model is used independently to generate multiple captions for each input image. Each caption is then scored, and the highest caption is generated one sentence at a time to generate a one-sentence caption for each of the two images.

### B. Caption Generation for Multiple Images

Several efforts have been made to generate captions for multiple images as an application of conventional image-caption generation. One example is the change in the image-caption generation initiative. This method identifies changes

image and output. The proposed caption-generation method yields captions that contain more salient features than one garment, with one sentence for each image and an average of approximately 14 words. Examples of the captions obtained are shown in Figure 1. In the subject experiment, it was evaluated whether the captions obtained using the proposed method contained obvious errors, how well they described features that were only present in one garment, and whether they were useful for comparing garments. This experiment confirmed that the generated captions adequately described the differences between products and provided useful information for product comparison.

The remainder of this paper is organized as follows. Section II describes work related to this study. Section III describes the proposed method. Section IV describes in detail the models and datasets used in the experiments. Section V describes the experiments on the comparative validation of the proposed method by employing different scoring methods. Section VI describes the experiments that qualitatively evaluate the captions generated by the proposed method. Finally, Section VII discusses the conclusions of this study and future perspectives.

between two input images and generates a one-sentence caption describing the change [14][15]. In this study, a caption is generated for each input image. In conventional image-caption generation, which tends to generate generic sentences, the distinctive parts of the input images are often ignored. To address this problem, an approach called feature-based image-caption generation is currently in progress [16][17]. In this approach, a single input image is compared to a set of similar images in a database to identify the distinctive aspects of the input image, which are then reflected in the caption. However, this approach does not specify similar images explicitly. In this study, two specified images are compared. The attribute estimates calculated for each image are compared, and a relative score is calculated. The caption score is then calculated by summing the attribute estimates that appear in the caption and is used for caption selection.

### C. Clothing Attribute Estimation

Clothing attribute estimation is the task of estimating features, such as the material, pattern, collar shape, and sleeve length of clothing in an image. Examples of the estimated attributes include cotton, floral, sleeveless, and leather. This task has been applied to garment retrieval and recommendation. Chen et al. [23] proposed a model that combines a CNN [25] trained on a large image dataset, ImageNet [24] with a multilayer perceptron for a garment image retrieval task that matches images of garments worn by a person with those from a fashion e-commerce site. Similarly, Huang et al. [26] proposed a deep model that included two CNNs to handle street images and e-commerce site images in garment image retrieval. Both models were trained using bounding boxes to identify garment regions. In contrast, Liu et al. [21] proposed a model that learns garment landmark information, such as sleeve and collar positions, estimates the landmarks during inference, and uses this information as an aid for garment attribute estimation. A comparison of the garment-attribute estimation models on the benchmark dataset, Deepfashion [21] is presented in Table II. The Top-k Recall [22] was used as an evaluation metric. This assigns the top-k attributes with the highest probability of estimation to each image and measures the number of correctly estimated attributes. By estimating landmark information, FashionNet can better recognize the shape and position of garments and perform better than models that use only bounding boxes. Here, consumer perceptions of attributes are subjective and depend on age and gender. Different consumers may consider different attributes important when comparing garments. However, as a first attempt in this study, the weighting of the attributes did not change. Only estimates objectively calculated using the model were used.

### D. Clothing Image Caption Generation

Sonoda et al. [18] proposed a method for searching for similar input images from a set of garment images they collected and applied the obtained garment information and features of similar images to a template. Yang et al. [19] proposed a framework that supports the creation of product

TABLE II
COMPARISON OF CLOTHING ATTRIBUTE ESTIMATION MODELS

| Model | Top-3 Recall | Top-5 Recall |
|---|---|---|
| WBIT [23] | 27.46 | 35.37 |
| DARN [26] | 40.35 | 50.55 |
| FashionNet [21] | 45.52 | 54.61 |

introductions on e-commerce websites. In their study, attribute- and sentence-level rewards were introduced to improve the quality of captions generated. They also adopted a method for integrating the training of the model using maximum likelihood estimation, attribute embedding, and reinforcement learning. In addition, a large dataset for garment image-caption generation containing approximately one million images was constructed. Cai et al. [20] removed noisy garment images and reconstructed a clean garment image dataset. These studies generated captions describing the salient features of a single-input garment image. They are insufficient for the purpose of this research, that is, to provide information when comparing garments, in that they cannot express the detailed differences between different garments. In this study, a caption is generated that highlights the differences between two input garment images.

## III. PROPOSED METHOD

This section describes the caption-generation method proposed in this study, which highlights the differences between garment image pairs. An overview of the method is presented in Figure 1. The method considers a pair $X = \{x_i \mid i = 1, 2\}$ of different garment images as input and outputs a caption pair $Y' = \{y_i' \mid i = 1, 2\}$ corresponding to each image, where $x_i$ is the i-th garment image, and $y_i'$ is the output caption corresponding to $x_i$. In Figure 1, the attribute set annotated to the image is displayed next to each image. This method comprises four modules: caption set generation, attribute scoring, caption scoring, and caption selection. The following sections describe these modules in detail.

### A. Caption Set Generation Module

The caption set generation module considers a pair $X$ of different garment images as input, inputs each image independently of the image caption-generation model, and outputs a caption set $Y = \{y_{ij} \mid i = 1, 2; j = 1, 2, \ldots, J\}$ corresponding to each image. Here, $y_{ij}$ represents the j-th caption for image $x_i$. The image-caption generation model used in this study is described in detail in Section IV.

### B. Attribute Scoring Module

The attribute scoring module considers a pair of different garment images $X$ and a caption set $Y$ as input and outputs a set of attribute scores $A = \{a_{ik} \mid i = 1, 2; k \in K\}$ for each image. Here, $K$ is the set of attributes to be evaluated and $a_{ik}$ is the score of attribute $k$ for image $x_i$. An attribute score is a numerical expression of the prominence of a particular attribute exhibited by a garment image; the higher the score,

the stronger the garment image that exhibits that attribute. An example of an attribute score for the garment image $x_1$ in Figure 1 is 0.20 for crewneck, 0.15 for pocket, and 0.01 for sleeveless, which were calculated to be higher when the image had the attribute prominently and lower when it did not. In this study, two methods of attribute scoring were considered: attribute scoring based on attribute estimation, and attribute scoring based on frequency of occurrence.

*1) Attribute Scoring Based on Attribute Estimation:* Attribute scoring based on attribute estimation uses a garment-attribute estimation model, whose output is the estimated probability of each attribute for an input-garment image. The estimated probability of an attribute for each image was calculated, and this value was used as the attribute score. This is illustrated in (1), where $p_{ik}$ is the estimated probability of attribute $k$ for image $x_i$. The clothing attribute estimation model used in this study is described in detail in Section IV.

$$a_{ik} = p_{ik} \tag{1}$$

*2) Attribute Scoring Based on Frequency of Occurrence:* The caption generated for each garment image using the caption set generation module reflects the garment characteristics. If a particular attribute appears frequently in a caption set, it can be regarded as one of the main features of the garment. This method calculates the frequency of occurrence of each attribute in the caption set for each image and uses this value as the attribute score. This is illustrated in (2), where $f_{ijk}$ is the number of occurrences of attribute $k$ in the caption $y_{ij}$.

$$a_{ik} = \frac{1}{J} \sum_{j=1}^{J} f_{ijk} \tag{2}$$

### C. Caption Scoring Module

The caption scoring module considers a caption set $Y$ and an attribute score set $A$ as inputs, and outputs a caption score set $C = \{c_{ij} \mid i = 1, 2; j = 1, 2, \ldots, J\}$. The caption score is a numerical expression of the extent to which the caption reflects the salient attribute differences between the garment images and attributes specific to each image; a higher score is regarded as emphasizing the differences between one image and the other. Here, $c_{ij}$ represents the score of the caption $y_{ij}$. In this study, two caption scoring methods were considered: caption scoring based on the comparison of top attributes and caption scoring based on the addition of relative scores. These methods are described in detail as follows.

*1) Caption Scoring Based on Comparison of Top Attributes:* This method first obtains an attribute set $K_i^{top-n}$ with the top $n$ attribute scores for each image. Next, the difference set $D_i$ of $K_i^{top-n}$ for each image is the difference attribute set, and the product set $T$ is the common attribute set. These are presented in (3)~(5):

$$D_1 = K_1^{top-n} \setminus K_2^{top-n} \tag{3}$$

$$D_2 = K_2^{top-n} \setminus K_1^{top-n} \tag{4}$$

$$T = K_1^{top-n} \cap K_2^{top-n} \tag{5}$$

Finally, the difference between the number of attribute occurrences in the different attribute sets and the number of attribute occurrences in the common attribute set for each caption was calculated and used as a caption score. This process is illustrated in (6), where $f_{ijk}$ is the number of occurrences of attribute $k$ in caption $y_{ij}$.

$$c_{ij} = \sum_{k \in D_i} f_{ijk} - \sum_{k \in T} f_{ijk} \tag{6}$$

This method assigns higher scores to captions containing more differentiated and fewer common attributes.

*2) Caption Scoring Based on Relative Score Addition:* This method first calculates the difference in attribute scores between images to obtain the relative attribute scores $\Delta a_{ik}$. These are given by Equations (7) and (8), respectively.

$$\Delta a_{1k} = a_{1k} - a_{2k} \tag{7}$$

$$\Delta a_{2k} = a_{2k} - a_{1k} \tag{8}$$

Next, the relative attribute scores corresponding to the attributes in the caption are added and used as the caption score. The process is described in (9), where $K_{y_{ij}}$ represents the set of attributes contained in the caption $y_{ij}$.

$$c_{ij} = \sum_{k \in K_{y_{ij}}} \Delta a_{ik} \tag{9}$$

Using this method, captions containing more attributes with relatively high attribute scores have higher scores.

### D. Caption Selection Module

The caption selection module considers the caption sets $Y$ and $C$ as input, selects the caption with the highest caption score in the caption set corresponding to each image, and outputs a set of captions $Y' = \{y_i' \mid i = 1, 2\}$ that highlights the differences. This process is represented by (10).

$$y_i' = \underset{y_{ij}}{\operatorname{argmax}} c_{ij} \tag{10}$$

## IV. MODELS AND DATASETS

This section describes the image-caption generation models, garment attribute estimation models, and garment image datasets used in the study.

### A. Image Caption Model and Clothing Attribute Estimation Model

This study is looking at reflecting different national and regional fashion cultures in captions in the future. Therefore, image-caption generation models that can handle garment image data in various languages are desirable. Among the image-caption generation models compared in Section II, ClipCap [10] is a combination of CLIP and the language model GPT-2. It is easy to handle non-English data because CLIP exists for multiple languages [29], and the language model Generative Pre-trained Transformer 4 (GPT-4) [30], which is similar to GPT-2, supports multiple languages. Furthermore, as shown in Table I, the accuracy is sufficiently high among the major image-caption generation models. Therefore, in this

TABLE III
COMPARISON OF CLOTHING IMAGE DATASETS

| Dataset | Number of images | Attributes | Captions |
|---|---|---|---|
| FACAD170K [20] | 178,849 | yes | yes |
| DeepFashion [21] | 289,222 | yes | no |
| FashionGen [27] | 325,536 | no | yes |
| iFashion [28] | 1,062,550 | yes | no |

study, ClipCap was used as the image-caption generation model in the caption set generation module. FashionNet [21] was used as the garment-attribute estimation model in the attribute-scoring module. This model estimates the landmarks of a garment and uses the obtained information for garment attribute estimation. This model can capture the fine-grained features of a garment image and is highly accurate.

### B. Clothing Image Dataset

A comparison of the main garment image datasets is shown in Table III. In this study, the FACAD170K garment image dataset [20] with both attributes and captions, which enables an attribute-based caption evaluation, was used to train the image-caption generation model. An example of the FACAD170K data is shown in Figure 2. Each garment image was crawled from a generic website, mainly Google Chrome, and was either an image of a person wearing the garment or an image of the garment alone, with a one-sentence caption from the web. The data collected using this method reflect the variety of styles and trends in clothing that real consumers interact with on a daily basis and are therefore highly suitable for simulation and analysis to mimic the context of consumers' clothing choices. The same caption is provided for garments of different colors. The bold text in the captions for Figure 2 represents multiple attributes assigned to a single garment image. FACAD170K has 990 attributes. In contrast, training the garment-attribute estimation model requires bounding boxes and landmark information to identify garment regions. However, FACAD170K did not contain these annotations. Because annotation is time-consuming, we used FashionNet's Deepfashion [21] pre-training model for garment attribute estimation. DeepFashion contains 1000 attributes, 292 of which match FACAD170K. The top ten attributes with the highest frequency of occurrence in FACAD170K and their frequencies are listed in Table IV. FACAD170K and DeepFashioin data with these attributes were used to evaluate the proposed method.

## V. COMPARATIVE VERIFICATION OF ATTRIBUTE AND CAPTION SCORING METHODS

### A. Objectives

This experiment aimed to compare and validate attribute scoring based on attribute estimation and frequency of occurrence in the attribute scoring module and caption scoring based on the comparison of top attributes and the addition of relative scores in the caption scoring module to find the best combination of methods for generating captions that highlight differences.



crisp **stripe** across the chest and **sleeve** further the **athletic** appeal of this **essential crewneck** t tee

*Bold text indicates attributes.*
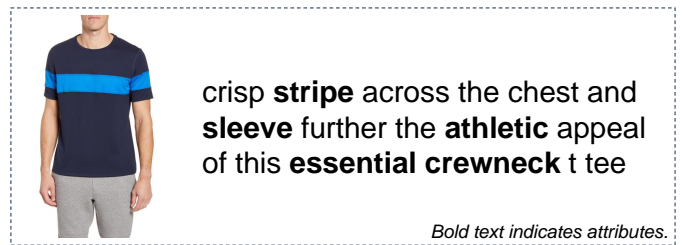
(a) Clothing image A and corresponding caption.



**patch pocket** add a dose of utilitarian **style** to a **button front** jacket that is ideal for both office and **weekend** wear

*Bold text indicates attributes.*

(b) Clothing image B and corresponding caption.

Figure 2. Examples of data from the FACAD170K dataset.

TABLE IV
HIGH-FREQUENCY ATTRIBUTES COMMON TO BOTH FACAD170K AND DEEPFASHION

| Attribute | Frequency (%) |
|---|---|
| cotton | 4.53 |
| cut | 4.41 |
| soft | 3.76 |
| sleeve | 2.98 |
| fit | 2.81 |
| leather | 2.58 |
| stretch | 2.46 |
| classic | 2.45 |
| knit | 2.31 |
| strap | 2.25 |

### B. Methods

In this experiment, the captions generated using the four proposed methods were automatically evaluated. In the caption set generation module, the image-caption generation model ClipCap was trained using 177,849 training data points from FACAD170K. The key parameters during training were set to a learning rate of $2.0 \times 10^{-5}$, a batch size of 40, and 10 epochs. These parameters were set based on the settings used in the original study [10]. $J = 100$ captions were generated for each image, based on the probability distribution of the language model. In the attribute scoring module, 292 attributes common to FACAD170K and DeepFashion were used as the attribute set $K$ to be evaluated. Caption scoring based on top attribute comparisons in the caption scoring module uses the top $n = 9$ attributes. The values were determined based on preliminary experiments that compared the estimated and correct attributes for different values of $n$. The model was evaluated by comparing the inferred results of the model against FACAD170K and DeepFashion with correct labels. The evaluation metrics are as follows. The set of attributes

annotated for a garment image $x_i$ is the overall attribute set $K_i^{GT}$, and the set of attributes with only one garment image is the differential attribute set $D_i^{GT}$. This is expressed in (11) and (12).

$$D_1^{GT} = K_1^{GT} \setminus K_2^{GT} \qquad (11)$$

$$D_2^{GT} = K_2^{GT} \setminus K_1^{GT} \qquad (12)$$

Let $K_{y_i'}$ be the attribute set contained in the caption $y_i'$. The precision, Recall, and F1 scores were calculated between $K_{y_i'}$ and the differential attribute set $D_i^{GT}$ to assess the degree of description of the attributes that differed between garments. Similar indices were calculated between $K_{y_i'}$ and the overall attribute set $K_i^{GT}$ as supplementary indices to assess the degree of description of the attributes in each garment image. Larger values of these indices are preferable. The evaluation was performed on 10,000 pairs, and the average value of each evaluation indicator was calculated.

### C. Results

The evaluation results for the captions generated by the proposed method in FACAD170K and DeepFashion are listed in Tables V and VI. A comparison of the results across datasets shows that the evaluation values for FACAD170K are higher than those of DeepFashion for all indicators. This is because the image-caption generation model ClipCap was trained on the FACAD170K data; consequently, the attribute information of FACAD170K was more appropriately reflected in the captions. For attribute scoring methods, frequency-of-occurrence-based attribute scoring tends to perform better than attribute estimation-based attribute scoring on both datasets. In particular, FACAD170K outperformed the attribute scoring based on attribute estimation for all evaluation indicators. Regarding caption scoring methods, caption scoring based on relative score addition outperformed caption scoring based on top-attribute comparisons for all evaluation indices in both datasets. These results indicate that under the experimental conditions of this study, the combination of attribute scoring based on the frequency of occurrence and caption scoring based on relative score addition is the most effective.

## VI. Qualitative Evaluation of Generated Captions

### A. Objectives

This experiment aimed to assess how accurately the captions generated by the proposed method represent the features of a single garment, how well they capture the differences between garment image pairs, and how useful they are for comparing garments.

### B. Methods

In this experiment, the captions generated using the proposed method were presented to a group of subjects for evaluation. The subject group comprised ten male and female subjects in their 20s. Clothing image pairs and captions are shown in Figure 3. Five pairs of clothing images were prepared (Pairs 1 to 5). To compare the effectiveness of the proposed method based on the similarity between garments, Pairs 1 to 3 have high similarity, whereas Pairs 4 and 5 have low similarity. They were selected based on visual confirmation and the degree of agreement between the attributes of each garment image. The individual garment images were assigned a name, such as 1A for the image on the left side of Pair 1 and 1B for that on the right side. The proposed method employs the method that achieved the best performance in the experiments described in Section V. Specifically, attribute scoring based on frequency of occurrence and caption scoring based on relative score addition were applied. The questions and options set are shown in Table VII. Q1 was designed to assess how accurately the caption represented garment characteristics. Q2 and Q3 assessed how well the captions described the features of only one garment. Furthermore, Q4 was established to test the usefulness of the caption pairs provided for comparing garments. A five-point Likert scale was used to answer each question. In addition, the subjects were asked to explain the reasons for their choice of options and any erroneous features and features not described in the caption. Wilcoxon's signed-rank test was used as the test method. This test checked whether the answers to each question were significantly biased from neutral and the significance level was set at 5%. A similar test was used to examine the difference in responses between the two questions. A significant difference between the distribution of responses to the two questions was tested to determine whether a significant difference existed. For the comparative analysis based on the similarity between clothing image pairs, the Mann-Whitney U test was employed because of the different sample sizes, and the significance level was set at 5%. This analysis examined whether significant differences existed in the distribution of responses between garment pairs with different similarities. Furthermore, Bonferroni correction was applied to account for the effects of multiple tests.

### C. Results

The proportions of the responses to each question are shown in Figure 4. In Q1, approximately 60% of the respondents answered 'strongly disagree' and the p-value of the Wilcoxon signed-rank test was $4.02 \times 10^{-12}$, indicating a bias towards negative opinions rather than neutrality. For Q2, all responses were 'strongly agree' or 'agree', with a Wilcoxon test p-value of $6.50 \times 10^{-22}$, indicating a bias towards positive opinions rather than neutrality. In Q3, the proportion of respondents who answered 'strongly agree' was approximately 50% lower than that in Q2, but the p-value of the Wilcoxon test was $1.25 \times 10^{-13}$, indicating a bias towards positive rather than neutral opinions. The p-value of the Wilcoxon test between the responses to Q2 and Q3 is $1.27 \times 10^{-8}$, confirming a significant difference between the two questions. In Q4, the total number of 'strongly agree' and 'agree' responses reached approximately 80%, with a Wilcoxon test p-value of $5.41 \times 10^{-5}$, indicating a bias towards more positive than neutral opinions. The percentages of responses to Q4 in clothing image pairs with high and low similarity are shown in Figure 5. The Mann-Whitney U-test results showed a p-value of $1.03 \times 10^{-3}$, confirming a significant difference.

TABLE V
RESULTS IN FACAD170K

| Attribute Scoring | Caption Scoring | Differential Attributes | | | Overall Attributes | | |
|---|---|---|---|---|---|---|---|
| | | Precision | Recall | F1 Score | Precision | Recall | F1 Score |
| Attribute Estimation | Comparison of Top Attributes | 0.145 | 0.171 | 0.144 | 0.198 | 0.174 | 0.173 |
| | Relative Score Addition | 0.157 | 0.223 | 0.172 | 0.212 | 0.225 | 0.206 |
| Frequency of Occurrence | Comparison of Top Attributes | 0.204 | 0.324 | 0.236 | 0.248 | 0.294 | 0.258 |
| | Relative Score Addition | 0.214 | 0.369 | 0.256 | 0.274 | 0.353 | 0.297 |

TABLE VI
RESULTS IN DEEPFASHION

| Attribute Scoring | Caption Scoring | Differential Attributes | | | Overall Attributes | | |
|---|---|---|---|---|---|---|---|
| | | Precision | Recall | F1 Score | Precision | Recall | F1 Score |
| Attribute Estimation | Comparison of Top Attributes | 0.051 | 0.089 | 0.058 | 0.077 | 0.091 | 0.077 |
| | Relative Score Addition | 0.070 | 0.136 | 0.084 | 0.123 | 0.164 | 0.131 |
| Frequency of Occurrence | Comparison of Top Attributes | 0.057 | 0.139 | 0.075 | 0.088 | 0.143 | 0.104 |
| | Relative Score Addition | 0.059 | 0.156 | 0.080 | 0.096 | 0.171 | 0.118 |



Figure 3. Presented clothing image pairs and captions shown to participants.

### D. Discussions

The results for Q1 suggest that the captions generated by the proposed method accurately describe the characteristics of the garment. However, several users pointed out that garment image 1B, which is made of denim fabric, was incorrectly described as a corduroy material. The corduroy attribute appeared nine times in the caption set for garment image 1B, compared to zero times for garment image 1A, resulting in a higher attribute score, and captions containing the corduroy attribute were preferentially selected. This can be attributed to the difficulty in recognizing detailed materials using CLIP. As some subjects judged this difference in material to be non-erroneous, this feature is also difficult for humans to identify.

The results for Q2 and Q3 suggest that the captions may

TABLE VII
SET QUESTIONS AND OPTIONS

| Question | | Options |
|---|---|---|
| Q1 | Do you think the caption clearly misdescribes a feature of the clothing? | • Strongly Agree |
| Q2 | Do you think the caption describes one or more feature that is unique to the item of clothing? | • Agree<br>• Neutral |
| Q3 | Do you think the caption describes all the features that are unique to the item of clothing? | • Disagree<br>• Strongly Disagree |
| Q4 | Do you think that the two captions would help you to compare the clothing if you were deciding whether to buy one of the clothing items? | |



(a) Q1: Do you think the caption clearly misdescribes a feature of the clothing?

(b) Q2: Do you think the caption describes one or more feature that is unique to the item of clothing?
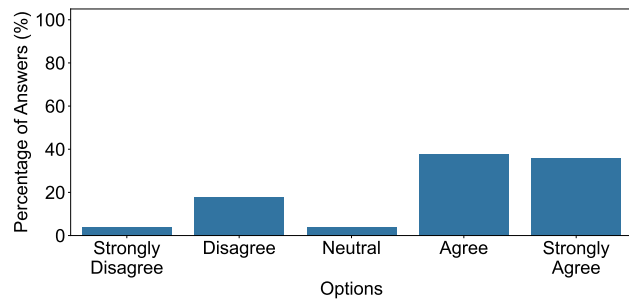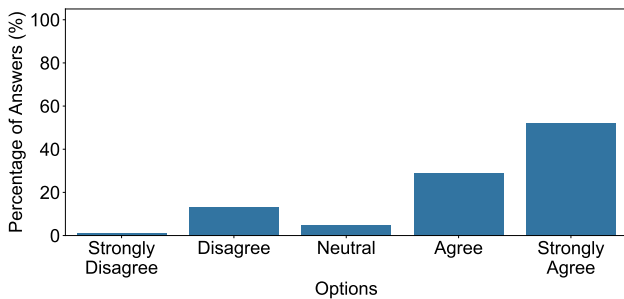
(c) Q3: Do you think the caption describes all the features that are unique to the item of clothing?

(d) Q4: Do you think that the two captions would help you to compare the clothing if you were deciding whether to buy one of the clothing items?

Figure 4. Percentage of answers to each question.
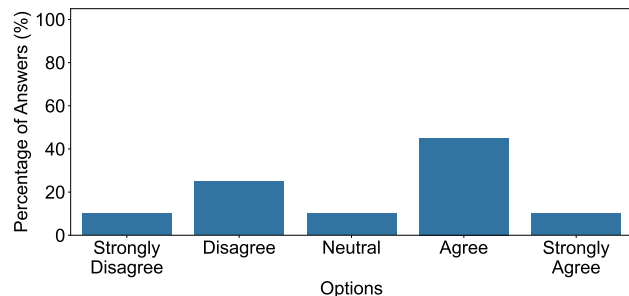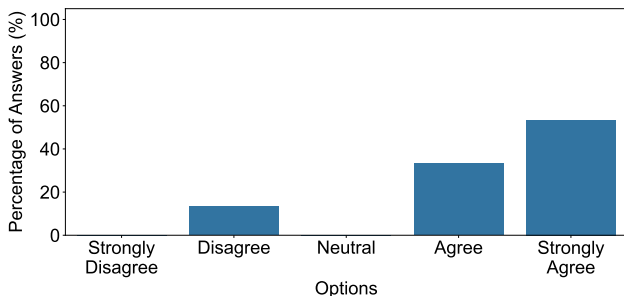


(a) Similar pairs.

(b) Dissimilar pairs.

Figure 5. Percentage of answers to Q4 for similar and dissimilar pairs.

describe at least one feature unique to a garment, but not all of them exhaustively. An example of an exhaustive description is

provided in Pair 3. Differences in length and the presence or absence of the logo were described in the caption, and many respondents indicated that all features unique to one garment were described in the caption. Conversely, as examples of non-exhaustive descriptions, it was pointed out that Pair 1 did not describe the number of pockets and Pair 2 did not describe the different colors of the garments. The attribute 'pocket' allows the presence or absence of pockets to be reflected in the caption, but it is considered difficult to reflect the number of pockets. Regarding color, although attributes for color exist in FACAD170K, the same caption is given to garment images of different colors, and there are few descriptions of color, which may be attributed to the difficulty in generating descriptions for color.

The results for Q4 showed that the generated captions provided useful information for the comparison of garments. Furthermore, they were more useful for pairs with high similarity than for those with low similarity. Some participants commented that reading the captions helped them focus on features of highly similar pairs that were not immediately noticeable in the images, such as the differences in length and graphics for Pair 1, position of the lines for Pair 2, and differences in length for Pair 3. However, in the less similar pairs, Pairs 4 and 5, the differences in the features pointed out by the captions were visually clear, and many were critical to the usefulness of the captions in the comparison. As there were no opinions that the captions described differences in features that were difficult to notice, it was considered that captions highlighting differences in pairs with low similarity were not useful.

## VII. CONCLUSION AND FUTURE WORK

In this study, a caption-generation method that highlights the differences between pairs of garment images to provide useful information for consumers when comparing products was proposed and evaluated. In this method, two different garment images are first input independently into an image caption generator to generate multiple captions. Attribute scores are then calculated for each image. A caption score is then calculated for each caption in the multiple captions generated for each image using the attribute scores. Finally, the captions are selected and output based on caption scores. Automatic evaluation experiments were conducted on attribute scoring and caption scoring, focusing on accurately describing the features of a single garment and the differences between garments. Methods employing attribute scoring based on the frequency of occurrence and caption scoring based on relative score addition were rated highly. Attribute scoring based on frequency of occurrence uses the frequency of an attribute's occurrence in the caption as the attribute score, whereas caption scoring based on relative score addition calculates the relative value of the attribute score and adds it to the number of attributes that appear. Furthermore, captions generated by a combination of methods that received high ratings in the automatic evaluation experiment were presented to the subjects, and a qualitative evaluation of their usefulness

was conducted. The results confirm that the proposed method provides useful information for comparing two garments. It was also confirmed that the proposed method is more effective for highly similar garment pairs than for less similar garment pairs. As it is assumed that consumers often compare garments with high similarity when comparing garments, an approach for garments with high similarity is planned.

The proposed method can only specify two garment images as input images. We plan to extend this approach to handle more than three garment images to better meet consumer garment comparison needs. Specifically, we believe that the relative scores can be calculated in the same manner as in the present study by subtracting the average attribute scores of the other images from the attribute score of one garment during attribute scoring. In addition, the performance of the proposed method is highly dependent on the accuracy of the image caption generation model and the diversity of the generated captions. Because the image caption generation model can be easily changed to other models because of the structure of the proposed method, there is room to verify its performance when using state-of-the-art models, such as GPT-4 Vision.

## ACKNOWLEDGMENT

## REFERENCES

[1] M. R. Solomon, "Consumer behavior : buying, having, and being," 12th ed, Pearson Education, 2016.

[2] T.-Y. Lin et al., "Microsoft coco: Common objects in context," Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, pp. 740-755.

[3] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: A method for automatic evaluation of machine translation," Proceedings of the 40th annual meeting of the Association for Computational Linguistics, pp. 311-318, 2002.

[4] S. Banerjee and A. Lavie, "METEOR: An automatic metric for MT evaluation with improved correlation with human judgments," Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization, pp. 65-72, 2005.

[5] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3156-3164, 2015.

[6] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," International conference on machine learning, pp. 448-456, 2015.

[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[8] K. Xu et al., "Show, attend and tell: Neural image caption generation with visual attention," International conference on machine learning, pp. 2048-2057, 2015.

[9] S. J. Rennie, E. Marcheret, Y. Mroueh, J. Ross, and V. Goel, "Self-critical sequence training for image captioning," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7008-7024, 2017.

[10] R. Mokady, A. Hertz, and A. H. Bermano, "Clipcap: Clip prefix for image captioning," arXiv preprint arXiv:2111.09734, 2021.

[11] A. Radford et al., "Learning transferable visual models from natural language supervision," International conference on machine learning, pp. 8748-8763, 2021.

[12] A. Radford et al., "Language models are unsupervised multitask learners," OpenAI blog, vol. 1, no. 8, pp. 9, 2019.

[13] P. Wang et al., "Ofa: Unifying architectures, tasks, and modalities through a simple sequence-to-sequence learning framework," International Conference on Machine Learning, pp. 23318-23340, 2022.

[14] H. Jhamtani and T. Berg-Kirkpatrick, "Learning to describe differences between pairs of similar images," Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 4024-4034, 2018.

[15] D. H. Park, T. Darrell, and A. Rohrbach, "Robust change captioning," Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4624-4633, 2019.

[16] J. Wang, W. Xu, Q. Wang, and A. B. Chan, "Group-based distinctive image captioning with memory attention," Proceedings of the 29th ACM International Conference on Multimedia, pp. 5020-5028, 2021.

[17] Y. Mao et al., "Rethinking the reference-based distinctive image captioning," Proceedings of the 30th ACM International Conference on Multimedia, pp. 4374-4384, 2022.

[18] A. Sonoda and G. Niina, "Apparel EC saito ni okeru setsumei bun jidou seisei (Automatic Generation of Descriptions in Apparel E-commerce Sites)," Proceedings of the Japan Society of Management Information National Conference, pp. 125-127, 2018.

[19] X. Yang et al., "Fashion captioning: Towards generating accurate descriptions with semantic rewards," Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16, pp. 1-17.

[20] C. Cai, K.-H. Yap, and S. Wang, "Attribute conditioned fashion image captioning," IEEE International Conference on Image Processing, pp. 1921-1925, 2022.

[21] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang, "Deepfashion: Powering robust clothes recognition and retrieval with rich annotations," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1096-1104, 2016.

[22] Y. Gong, Y. Jia, T. Leung, A. Toshev, and S. Ioffe, "Deep convolutional ranking for multilabel image annotation," arXiv preprint arXiv:1312.4894, 2013.

[23] M. H. Kiapour, X. Han, S. Lazebnik, A. C. Berg, and T. L. Berg, "Where to buy it: Matching street clothing photos in online shops," Proceedings of the IEEE international conference on computer vision, pp. 3343-3351, 2015.

[24] J. Deng et al., "ImageNet: A Large-Scale Hierarchical Image Database," IEEE conference on computer vision and pattern recognition, pp. 248-255, 2009.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[26] J. Huang, R. S. Feris, Q. Chen, and S. Yan, "Cross-domain image retrieval with a dual attribute-aware ranking network," Proceedings of the IEEE international conference on computer vision, pp. 1062-1070, 2015.

[27] N. Rostamzadeh et al., "Fashion-gen: The generative fashion dataset and challenge," arXiv preprint arXiv:1806.08317, 2018.

[28] S. Guo et al., "The imaterialist fashion attribute dataset," Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pp. 0-0, 2019.

[29] F. Carlsson, P. Eisen, F. Rekathati, and M. Sahlgren, "Cross-lingual and Multilingual CLIP," Proceedings of the Language Resources and Evaluation Conference, pp. 6848-6854, 2022.

[30] OpenAI, "GPT-4 Technical Report," arXiv preprint arXiv:2303.08774, 2023.

# Analysis of Weather Information and Road Surface Images for Snow Removal Dispatch Prediction

Hiroki Okura
*Graduate School of Information
Science and Technology,
Hokkaido University*
Sapporo, Hokkaido, Japan
email:hirokio@ist.hokudai.ac.jp

Soichiro Yokoyama
*Faculty of Information
Science and Technology,
Hokkaido University*
Sapporo, Hokkaido, Japan
email:yokoyama@ist.hokudai.ac.jp

Tomohisa Yamashita
*Faculty of Information
Science and Technology,
Hokkaido University*
Sapporo, Hokkaido, Japan
email:yamashita@ist.hokudai.ac.jp

Hidenori Kawamura
*Faculty of Information
Science and Technology,
Hokkaido University*
Sapporo, Hokkaido, Japan
email:kawamura@ist.hokudai.ac.jp

*Abstract*—This study focuses on addressing the challenges associated with decision-making in winter road snow removal operations, aiming to alleviate the burden on snow removal personnel. Specifically, we propose a system that collects and visualizes information on road weather data and snow conditions to support decision-making by responsible personnel. Additionally, by sharing the collected information, we aim to facilitate the sharing of premonitions about changes in decision-making among snow removal personnel, reducing the need for physical inspections. In this paper, we discuss the utilization of the collected information, presenting a module proposal that utilizes deep learning to quantify the snow coverage in images captured by fixed-point cameras. We also explore the analysis of data for the introduction of a predictive function for snow removal operations.

*Keywords*—Snow Removal, Information Sharing System, Weather Information.

## I. INTRODUCTION

Winter road snow removal operations play a critical role in maintaining road traffic and essential infrastructure in snowy regions. One significant challenge in these operations is the difficulty of decision-making regarding deployment.

The decision to deploy snow removal operations is made by responsible personnel, determining whether to conduct snow removal activities from 16:00 to 2:00 the next day. Snow removal tasks encompass two main types: "new snow removal" and "road surface leveling." New snow removal aims to eliminate accumulated snow on road surfaces when it poses a hindrance to road traffic due to heavy snowfall. Road surface leveling is performed to ensure road flatness when conditions like compacted snow growth or residual snow impede traffic.

Particularly, the decision-making for new snow removal operations scheduled from 2:00 the next day is heavily influenced by road conditions up to that point and localized weather changes after 16:00. Even veteran personnel often overturn their decisions made at 16:00, leading to frequent decision reversals.

Due to the frequent reversals in deployment decisions, snow removal personnel are burdened with the need to be prepared for deployment even if it was deemed unnecessary at 16:00. Additionally, a pre-deployment assessment called "snow patrols" involves physically traveling to the snow removal area by car just before the operation to visually confirm the snow conditions. However, driving on snow-covered roads late at night poses risks of vehicle stacking and slip accidents, necessitating alternative methods.

To alleviate the burdens associated with such snow removal deployment decisions, this study adopts an approach based on Visual Analytics, utilizing a visual interface for analytical reasoning concerning large and complex datasets [1]. We aim to develop a system that collects and visualizes information on road snow conditions and weather data to support decision-making by personnel. As part of the functionality of the proposed snow removal deployment decision support system, we implement a feature to capture real-time road images from fixed cameras in the target area and gather meteorological information.

The collected information is visualized on the system to aid decision-making by personnel, enabling consistent deployment decisions that are not affected by weather fluctuations, ultimately reducing the effort required for snow patrols. Furthermore, by making the information accessible to snow removal personnel, we aim to facilitate the sharing of premonitions about changes in deployment decisions due to weather fluctuations.

In this paper, we describe the structure of a snow-related information sharing site that has been launched as part of the snow removal deployment decision support system. We also propose a snow coverage estimation module using deep learning to automate the visual inspection of snow coverage

in fixed camera images for the site.

Moreover, we consider introducing a snow removal deployment prediction feature to assist in making advanced deployment decisions. The decision to carry out snow removal operations at 16:00 the day before requires considering the snow conditions and weather changes, making it a challenging task even for experienced personnel. To address this, we develop an algorithm to predict the necessity of snow removal based on the collected information and validate its effectiveness by introducing the system into real-world scenarios.

## II. Related Research

In this section, we discuss research on prediction systems related to snow removal deployment support systems and explore relevant technologies in image processing used for snow coverage estimation.

### A. Research on Systems Using Prediction Techniques

In tandem with the development of snow removal deployment prediction functionality, there exists research focusing on forecasting related demands to facilitate appropriate responses.

One of Japan's challenges revolves around anticipating the increase in the elderly population and the decrease in the working-age population. In light of this, it is crucial to consider the efficient optimization of emergency ambulance transport services for the long term. Addressing this, Okamoto et al.'s research [2] proposes an emergency ambulance transport demand prediction model to enhance the efficient operation of emergency medical transport services. The study validates the suitability of the emergency transport demand prediction model using data from Matsuyama City. Additionally, the research reports insights, such as significantly higher transport rates per capita for the age group of 75 and above, variations in demand based on season, day of the week, and time.

Focusing on the pressing issue of heatstroke, Inai et al.'s study [3] analyzes emergency dispatch data related to heatstroke at a granular level. The objective is to formulate sustainable measures for future heatstroke incidents. The study specifically tackles the prediction of the number of heatstroke patients for adaptation in emergency medical scenarios, demonstrating the feasibility of predicting heatstroke patient numbers from meteorological data.

Addressing the challenge of pedestrians' falls on snow-covered walkways in Sapporo, Kato et al.'s research [4] deals with predicting the occurrence of emergency transport due to falls on such walkways. Using deep learning and past daily emergency transport data, the study constructs a model to predict the number of emergency transports resulting from falls on extremely slippery surfaces formed on sidewalks and pedestrian crossings.

Moreover, research has been conducted to maintain safe and efficient traffic in snowy regions during winter. Takahashi et al.'s study [5] aims to predict the winter road conditions to ensure proper management of road infrastructure. The research, initiated in 2004, observes weather and road surface temperature and develops a prediction model, operating a prototype in 2005.

Hori et al.'s study [6] focuses on constructing a winter road surface freezing prediction system to support the pre-spraying of anti-freezing agents. Using neural networks and discriminant analysis, the study builds a model to detect the temporal variations in road surface temperature and moisture. The established prediction system proves capable of accurately forecasting road surface freezing three hours ahead.

In connection with these studies, our research aims to develop a prediction system for the necessity of snow removal deployment to support the intricate decision-making process and validate its effectiveness by introducing the system into real-world scenarios.

### B. Semantic Segmentation

Semantic segmentation refers to the image recognition task using deep learning that divides an image into pixels and assigns semantic labels to each pixel. Since the introduction of Fully Convolutional Network (FCN) by Long et al. [7], methods based on the structure of FCN have become fundamental in this field. To obtain high-resolution feature maps, an Encoder-Decoder structure is commonly employed, utilizing FCN as the Encoder and incorporating the feature maps into the Decoder to recover spatial information. Representative models within this framework include Seg-Net [8] and U-Net [9]. Seg-Net performs downsampling in the Encoder and then upsampling in the Decoder corresponding to the number of downsampling steps. It utilizes the spatial indices of the maximum values in each pooling layer during upsampling, enabling clearer inferences. U-Net, while sharing the Encoder-Decoder structure with Seg-Net, adopts skip connections that concatenate low-dimensional features with high-dimensional ones, serving as a means to recover information lost during downsampling. Additionally, other methods, such as PSP-Net [10] and DeepLab [11] have been proposed, contributing to diverse research in this domain.

Furthermore, there is research that performs segmentation using images overlooking road surfaces targeted for heating to determine the snow coverage. In Imahara's study [12], snow detection in images captured from an approximately 30-meter height overlooking parking lots in Sapporo city was conducted. The results demonstrated reliable accuracy in snow detection, indicating its suitability for controlling road heating systems.

This study employs semantic segmentation for snow coverage estimation. Currently, snow removal decisions rely on manual inspection of images captured by fixed-point cameras installed at snow removal sites. To automate the assessment of snow conditions observed in these images, the study turns to semantic segmentation. Treating snow coverage estimation from images as a regression problem presents challenges, particularly in handling obstacles, such as cars, buildings, and people that may interfere with image recognition. Preprocessing is essential, and it involves addressing the significant amount of work required due to potential movements of obstacles and changes in camera angles. Attempting to estimate

obstacle regions through semantic segmentation alleviates this workload. Moreover, considering each pixel as a piece of training data allows handling a larger dataset compared to treating snow coverage estimation as a regression problem. This approach also enables leveraging information about regions prone to heavy or light snow coverage.

## III. DATA FOR SNOW REMOVAL DECISION SUPPORT SYSTEM

In this section, we will discuss the data to be visualized in the system developed for supporting decision-making in snow removal operations.

### A. Visualization Target Data

The system developed in this study collects and visualizes data from five snow removal points in Rumoi City, Hokkaido. Two fixed-point cameras are installed at each location, capturing road images from different angles to provide a detailed understanding of the snow accumulation situation. The data visualized in the system includes the following four components:

**Images from Fixed-Point Cameras:** Real-time road images captured by network cameras installed at each location.

**Snow Coverage Percentage from Image Analysis:** The percentage of snow coverage in the area of the road captured by the camera images, obtained through semantic segmentation analysis.

**Weather Information:** Weather forecast information provided by WeatherNews [13]. The short-term forecast includes hourly weather forecasts up to 72 hours in advance, and the medium-term forecast provides daily weather forecasts up to 10 days in advance. The short-term forecast includes weather codes, temperature, precipitation, atmospheric pressure, wind speed, wind direction, and relative humidity. The medium-term forecast includes weather codes, maximum temperature, minimum temperature, and precipitation probability.

**Snow Depth Information:** Snow depth information provided by WeatherNews [13]. Snow depth is presented in centimeters, assuming snow accumulation on the ground. It includes a 10-minute observation of snow depth and a 60-hour forecast with hourly intervals.

**Snow Removal Operation Prediction Probability:** The probability of the prediction for the need for snow removal operation based on the collected visualization data. This probability is generated by a logistic regression model using input features, such as estimated snow coverage percentage, weather information, and snow depth information. The algorithm for snow removal operation prediction is illustrated in Fig. 1.

Weather information and snow depth information can be obtained in detail for each snow removal point by specifying the latitude and longitude.

## IV. DEVELOPMENT OF SNOW REMOVAL DISPATCH DECISION SUPPORT SYSTEM

The proposed system aims to achieve the following objectives: i) Enable personnel responsible for deployment decisions to make decisions that are not overturned afterward, ii) Reduce the effort of snow patrols, iii) Allow snow removal personnel to access the same information as the deployment decision-makers to prepare for deployment.

### A. System Configuration

The configuration of the snow removal dispatch decision support system is illustrated in Fig. 2. The system comprises four main components: client-side, server-side, database, and external data.

External data includes images captured by fixed-point cameras, weather information, and snow depth information. The server-side retrieves and stores this external data in the database. Through API integration, the server-side enables the client-side to access the stored data consistently.

The technology stack used for developing this system includes:

- Client-side: JavaScript, React
- Server-side: Python, FastAPI
- Database: MySQL

### B. System Functions

The proposed system encompasses the following features:

1. **Real-time Display of Fixed-Point Camera Images at Snow Removal Target Locations:** A function to display real-time images from fixed-point cameras at snow removal target locations.

2. **Display of Weather Information and Snow Depth for Decision-Making in Snow Removal Operations:** A function to display weather information and snow depth data used in the decision-making process for snow removal operations.

3. **Utilization of Collected Data:**

**Snow Coverage Estimation Function:** Applies deep learning to fixed-point camera images to estimate the percentage of the snow-covered area in the road surface. - Learning Model: Utilizes Unet++, a model highly rated for semantic segmentation tasks, with Xception [14] network architecture for feature extraction in the encoder. - Pre-training: Utilizes a pre-trained model on the Imagenet dataset.

**Snow Removal Operation Prediction Function:** Applies a prediction model to collected data to forecast the necessity of snow removal operations. - Prediction Model: Utilizes logistic regression, effective for binary classification problems, providing ease of interpretation due to its simplicity.

Users can input and view information through the following screens on the client side.

**Location Selection Screen:** A screen displaying all snow removal target locations, allowing the selection of a specific location for information viewing. For each displayed
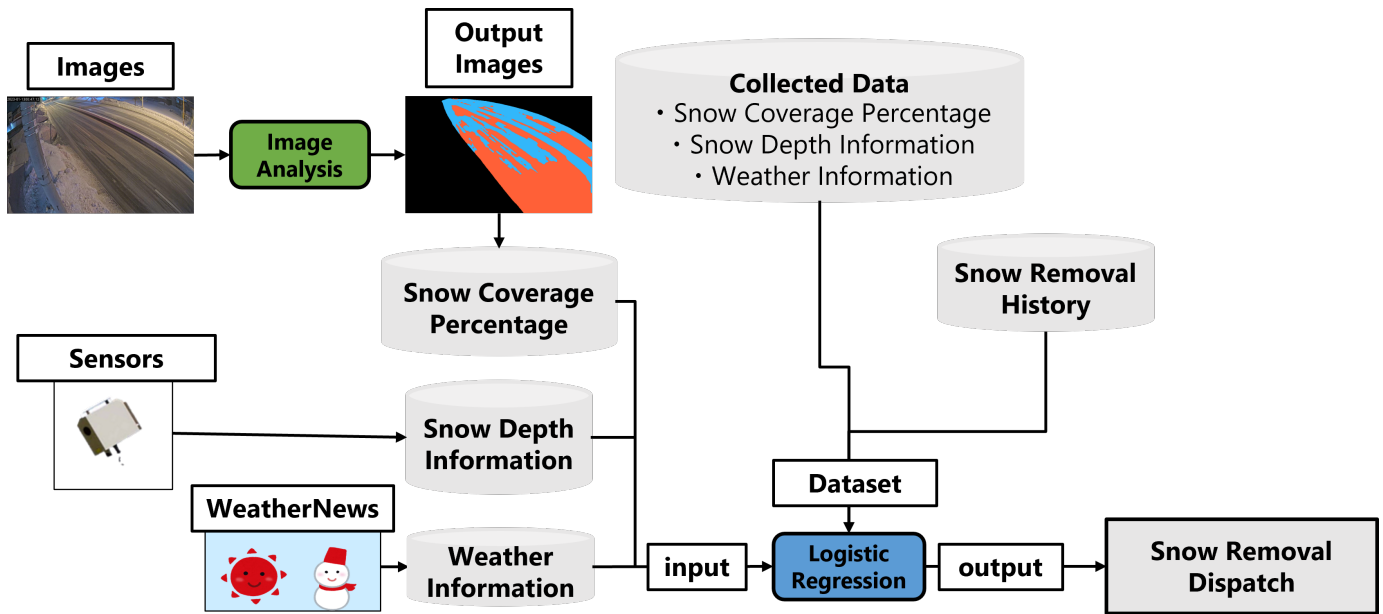
Fig. 1. Snow Removal Operation Prediction Algorithm.
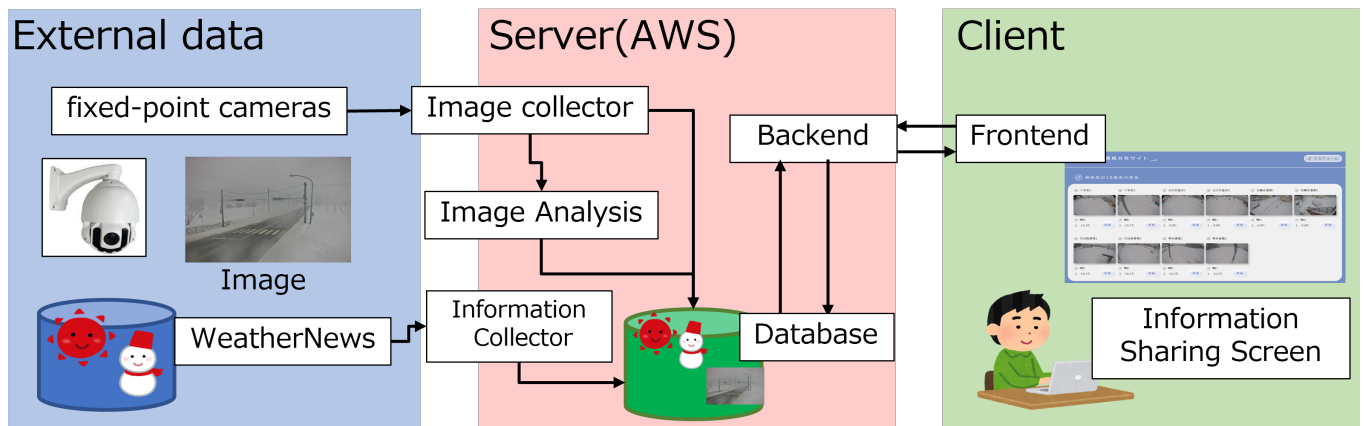


Fig. 2. Overview of System Configuration.

location, real-time images from fixed-point cameras and their update timestamps are shown. A specific layout is illustrated in Fig. 3.

**Real-time Information Screen for Each Location:** A screen presenting real-time images from fixed-point cameras, along with weather information and snow depth data for the selected location.

**Historical Information Screen for Each Location:** A screen displaying historical images from fixed-point cameras and past weather information, along with snow depth data for the chosen location. Graphical visualization of weather information and snow depth data facilitates the understanding of temporal trends.

**Forecast Information Screen for Each Location:** A screen presenting weather forecast information and snow depth forecast information for the selected location. Visualizing forecast data aids in understanding temporal patterns.

**Snow Depth Input Form Screen:** A screen for snow removal personnel to input measured snow depth values.

## V. SYSTEM EFFECTIVENESS VERIFICATION

The percentage of snow-covered areas in road images significantly influences the decision-making process for snow removal operations. Therefore, automatic analysis through semantic segmentation plays a crucial role in the utility of the proposed system. Analyzing the results of this analysis along with collected data and the historical records of operational decisions made by responsible personnel, we develop an operational prediction feature.

In Section II-B, we discussed the technology to automate and quantify the visual confirmation of snow conditions from fixed-point camera images using semantic segmentation. In this section, we apply this technology to verify the accuracy of snow coverage estimation.

Fig. 3. Site Selection Screen.

Furthermore, by analyzing the estimated snow coverage, weather information, snow depth data, historical operational decisions made by responsible personnel, and operational records, we develop an automated and quantified operational prediction feature.

### A. Accuracy Verification of Snow Coverage Estimation Using Deep Learning

For the accuracy verification of snow coverage estimation using deep learning, we evaluate the accuracy of region estimation and snow coverage percentage estimation as evaluation metrics.

*1) Dataset:* We use the fixed-point camera images mentioned in Section III-A. For this verification, we use images from 8 out of 10 locations due to labeling and system constraints. The shooting period is from December 19, 2022, to January 27, 2023. From this period, we select around 1 to 3 images from the same day, resulting in a dataset of 752 images. The image selection aims to cover different time intervals and weather conditions, including situations with no snow, heavy snow, and various other conditions.

Regarding labeling, we define the following labels for each pixel in the image: - Snow: Label for the region of the

TABLE I
TRAINING PARAMETERS.

| Epochs | 40 |
|---|---|
| Mini-Batch Size | 2 |
| Optimization | Adam |
| Learning Rate | Initial: 0.0001<br>After 25 epochs: 0.00001 |

road surface with confirmed snow. - Non-snow: Label for the region of the road surface with no confirmed snow. - Obstacle: Label for objects that may interfere with the classification of snow and non-snow regions, such as buildings, vehicles, trees, people, and poles. - Irrelevant Region: Label for all regions not classified into the above three classes.

*2) Training Configuration:* The training parameters are shown in Table I.

We distribute the dataset into training and validation images at an 8:2 ratio for each location and date. This separation allows us to evaluate the accuracy of predictions on images taken on different dates than those used for training, providing a more accurate estimation of system performance during actual operations.

*3) Evaluation Metrics:* The Intersection over Union (IoU) is used as the evaluation metric for label accuracy.

*4) Snow Coverage Estimation Method:* We utilize the snow coverage estimation proposed by ImaHara et al. [12] using the snow coverage percentage. The snow coverage percentage for each image is defined by 1.

We evaluate whether this value can be used to automate and quantify visual confirmation of snow conditions.

*5) Verification Results:* The IoU result is 0.951, indicating successful training.

For the 172 validation images, we calculate the snow coverage percentage using 1. The confusion matrix for the snow coverage percentage is shown in Table II.

In most images, the snow coverage percentages from the ground truth and predictions are in close agreement, with no significant outliers.

### B. Data Analysis for the Development of Snow Removal Dispatch Prediction Function

In this section, we apply the results of the snow coverage estimation, as verified in Section V-A, to the collected fixed-point camera images. We also analyze weather information, snow depth information, snow removal dispatch history, and the dispatch decisions made by responsible personnel. This analysis aims to explore an algorithm for predicting snow removal dispatch.

*1) Utilized Data:* We describe the data used for the analysis. The data covers the period from December 24, 2022, to February 28, 2023, for all 10 locations with fixed-point cameras mentioned in Section III-A.

For 8 out of the 10 locations, we use data for 63 days, excluding 4 days with data defects on February 15, 18, 27, and 28. For the remaining 2 locations with data defects, we use data for 36 days, excluding the same 4 days. Additionally, we use data from two laser snow depth sensors installed in different locations.

We utilize the snow coverage percentage estimates from the fixed-point camera images, as discussed in Section V-A. These estimates are obtained by applying the verification results to images taken at 4-hour intervals each day.

Weather information includes precipitation, temperature, and wind speed, provided by WeatherNews for each fixed-point camera location. Forecast information is available at 20, 24, and 28 hours.

Snow depth information is measured by laser snow depth sensors installed along the road. Two sensors are located approximately 10 km away from the fixed-point camera installation points.

Finally, we use the snow removal dispatch history in Rumoi City during the same period, as well as the recorded dispatch decisions made by responsible personnel at 16:00 on the previous day. These records are binary variables, indicating whether there was a dispatch (1) or not (0).

*2) Analysis Method:* We investigate the correlation between the snow removal dispatch records and various data collected at 16:00 on the previous day. We focus on identifying highly correlated variables to use in the prediction model.

*3) Analysis Results:* The correlation coefficients between the snow removal dispatch records and various data at 16:00 on the previous day, the change in estimated snow coverage percentage, and forecast data are presented in Tables III, IV, and V, respectively.

The tables indicate that the estimated snow coverage percentage, snow depth, temperature, and forecast data at 20:00, wind speed at 24:00, and precipitation at 28:00 exhibit strong correlations, with correlation coefficients exceeding

### C. Evaluation of Generalization Performance for Snow Removal Dispatch Prediction

Based on the data analysis presented in Section V-B, we explore an algorithm for predicting snow removal dispatch and investigate its generalization performance. The evaluation involves conducting a 5-fold cross-validation and comparing the results with the accuracy of dispatch decisions made by responsible personnel.

*1) Experimental Setup:* We describe the experimental settings, including the data used and the prediction algorithm.

Firstly, we use the data mentioned in Section V-B1, focusing on variables with an absolute correlation coefficient of 0.3 or higher, as identified in the analysis results.

Next, we consider the prediction algorithm. We apply logistic regression to generate output values for each location, and the prediction result is the average of these values for the same day.

*2) Validation Method:* We describe the validation method. We conduct a 5-fold cross-validation, ensuring that the ratios of observation locations and target variables are equal during data splitting. By specifying a seed value during the splitting, multiple datasets are created. Comparing the prediction results for these datasets allows us to investigate generalization performance independent of the dataset. Evaluation metrics, such as accuracy, precision, recall, and F1 score are examined, and the results are compared with dispatch decisions made by responsible personnel.

*3) Validation Results:* We present the validation results. Table VI shows the generalization performance obtained through 5-fold cross-validation for datasets created with seed values ranging from 0 to 4.

TABLE VI
GENERALIZATION PERFORMANCE EVALUATION THROUGH 5-FOLD CROSS-VALIDATION.

| Seed Value | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| 0 | 0.746 | 0.571 | 0.632 | 0.600 |
| 1 | 0.778 | 0.667 | 0.526 | 0.588 |
| 2 | 0.794 | 0.650 | 0.684 | 0.667 |
| 3 | 0.683 | 0.467 | 0.368 | 0.412 |
| 4 | 0.762 | 0.600 | 0.632 | 0.615 |

Additionally, Table VII displays the results when unstable model coefficients are excluded, focusing on stable model parameters.

$$\text{Snow Coverage Percentage} = \frac{\text{Number of Pixels with Snow Label}}{\text{Number of Pixels with (Snow Label + Non-snow Label)}} \tag{1}$$

TABLE II
CONFUSION MATRIX FOR SNOW COVERAGE PERCENTAGE.

| | | Actual Snow Coverage Percentage | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
| Predicted Snow Coverage Percentage | 0.0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0.1 | 9 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0.2 | 0 | 4 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0.3 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0.4 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0.5 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0.6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| | 0.7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0.8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0.9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | 1.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 126 |

TABLE III
CORRELATION COEFFICIENTS BETWEEN SNOW REMOVAL DISPATCH
RECORDS AND DATA AT 16:00 ON THE PREVIOUS DAY.

| Data Name | Correlation Coefficient |
|---|---|
| Estimated Snow Coverage Percentage | 0.334 |
| Snow Depth (Furuyama) | 0.378 |
| Snow Depth (Touge-shita) | 0.403 |
| Temperature | -0.490 |
| Precipitation | 0.182 |
| Wind Speed | 0.318 |

TABLE IV
CORRELATION COEFFICIENTS BETWEEN SNOW REMOVAL DISPATCH
RECORDS AND CHANGE IN ESTIMATED SNOW COVERAGE PERCENTAGE.

| Data Name | Correlation Coefficient |
|---|---|
| Difference from 12 hours ago | -0.218 |
| Difference from 8 hours ago | -0.199 |
| Difference from 4 hours ago | -0.082 |

TABLE VII
GENERALIZATION PERFORMANCE EVALUATION THROUGH 5-FOLD
CROSS-VALIDATION WITH STABLE MODEL COEFFICIENTS.

| Seed Value | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| 0 | 0.857 | 0.750 | 0.789 | 0.769 |
| 1 | 0.857 | 0.813 | 0.684 | 0.743 |
| 2 | 0.841 | 0.714 | 0.789 | 0.750 |
| 3 | 0.794 | 0.688 | 0.579 | 0.629 |
| 4 | 0.810 | 0.684 | 0.684 | 0.684 |

Furthermore, Table VIII presents the results of dispatch decisions made by responsible personnel during the same period.

TABLE VIII
DISPATCH DECISIONS MADE BY RESPONSIBLE PERSONNEL.

| Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|
| 0.714 | 0.520 | 0.684 | 0.591 |

The results show that the proposed dispatch prediction algorithm demonstrates superior accuracy compared to the

TABLE V
CORRELATION COEFFICIENTS BETWEEN SNOW REMOVAL DISPATCH
RECORDS AND FORECAST DATA.

| Forecast Time | Data Name | Correlation Coefficient |
|---|---|---|
| 20:00 | Temperature | -0.326 |
| | Precipitation | 0.355 |
| | Wind Speed | 0.434 |
| 24:00 | Temperature | -0.246 |
| | Precipitation | 0.284 |
| | Wind Speed | 0.323 |
| 28:00 | Temperature | -0.230 |
| | Precipitation | 0.349 |
| | Wind Speed | 0.249 |

dispatch decisions made by responsible personnel. Additionally, examining the model coefficients and refining parameters lead to higher prediction accuracy. We plan to implement this prediction algorithm and display it on the snow-related information sharing site in the future.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we provided an overview of a snow-related information sharing site, which is part of the snow removal dispatch decision support system and is already in practical use. The snow-related information sharing site automatically collects real-time road images and meteorological information for snow removal points, visualizes them on the site, and allows not only dispatch decision-makers but also snow removal personnel to access and share relevant information. We also outlined the verification of the introduction effects of the snow removal dispatch decision support system in the future.

Furthermore, we discussed our efforts in using deep learning for snow coverage estimation based on captured images from fixed-point cameras, presenting the results and potential applications of the method. Additionally, we proposed an algorithm to predict snow removal dispatch based on data collected at 16:00 on the day before the dispatch, highlighting its potential as a snow removal dispatch prediction feature.

Looking ahead, our future plans involve continued development of the snow removal dispatch prediction feature and its

implementation into the system. Subsequently, we will evaluate the accuracy of the dispatch prediction, its contribution to reducing the burden on decision-makers, alleviating the workload of snow patrols, and reducing the psychological burden on snow removal personnel through information access. These assessments will guide our ongoing efforts and contribute to the refinement of the snow removal dispatch decision support system.

## REFERENCES

[1] J. J. Thomas and K. A. Cook, "Illuminating the path: The research and development agenda for visual analytics," 2005, [retrieved: 12, 2023]. [Online]. Available: https://api.semanticscholar.org/CorpusID:60837336

[2] M. Kataoka, T. Yoshii, T. Futagami, and T. Oguchi, "A research on demand forecasting of emergency life-saving transfer," *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, vol. 71, no. 5, pp. I_407–I_414, 2015.

[3] T. Nakai, S. Saiki, and M. Nakamura, "Analyzing and predicting ambulance demand for heat stroke using fine-grained dispatch log," *IEICE Tech. Rep*, vol. 120, no. 232, pp. 13–19, 2020.

[4] H. Hoshino *et al.*, "Prediction of the number of transporting emergency patient slipping on snow roads in sapporo : Attempt using "deep learning"," *Cold Region Technology Conference*, vol. 33, pp. 153–157, 2017, [retrieved: 12, 2023]. [Online]. Available: https://cir.nii.ac.jp/crid/1522262180462753280

[5] N. Takahashi, R. Tokunaga, M. Asano, N. Ishikawa, and S. Kawakami, "Sophistication of road infrastructure management in a snowy cold region – development of a forecasting method for winter road conditions and maintenance support systems," *Studies in Regional Science*, vol. 38, no. 4, pp. 1079–1091, 2008.

[6] M. Horii and T. Fukuda, "Pavement ice prediction system in winter maintenance," *Doboku Gakkai Ronbunshu*, vol. 2001, no. 669, pp. 243–251, 2001.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[8] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2481–2495, 2017.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015.

[10] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *CVPR*, 2017.

[11] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *European Conference on Computer Vision (ECCV)*, pp. 801–818, 2018.

[12] T. Imahara, S. Yokoyama, T. Yamashita, and H. Kawamura, "Recognition of snow covering condition in road surface images by semantic segmentation for control of road heating," *Information Processing Society of Japan (IPSJ) Technical Report*, 2021.

[13] "Weathertech(wxtech®): Business utilization and analysis of weather data," https://wxtech.weathernews.com/, [retrieved: 12, 2023].

[14] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

# A Proposal of Road Network Hierarchization Method Based on Betweenness Centrality for Application to Vehicle Routing Problems

Masayuki Shimizu
*Graduate School of Information Science and Technology,*
*Hokkaido University*
Sapporo, Hokkaido, Japan
email: s.masayuki_0426@ist.hokudai.ac.jp

Soichiro Yokoyama
*Faculty of Information Science and Technology,*
*Hokkaido University*
Sapporo, Hokkaido, Japan
email: yokoyama@ist.hokudai.ac.jp

Tomohisa Yamashita
*Faculty of Information Science and Technology,*
*Hokkaido University*
Sapporo, Hokkaido, Japan
email: tomohisa@ist.hokudai.ac.jp

Hidenori Kawamura
*Faculty of Information Science and Technology,*
*Hokkaido University*
Sapporo, Hokkaido, Japan
email: kawamura@ist.hokudai.ac.jp

*Abstract*—In this study, we propose a road network hierarchization method for pathfinding that takes into account drivers' avoidance of narrow roads. The proposed method identifies nodes with high betweenness centrality, one of the centrality measures in network analysis, to connect disconnected subnetworks within each hierarchical level. To validate the effectiveness of the proposed method, multiple driver preferences regarding narrow roads are prepared, and computational experiments are conducted on a road network covering a 14 km square area in central Sapporo. Compared to performing pathfinding without hierarchical networks, the calculation time of individual pathfinding was reduced to 4-6% with the previous method and to 3-6% with the proposed method. Additionally, by using the proposed method, the average cost of routes improved compared to the previous method, and even when compared to the minimum cost paths, the worsening of route costs was about 4-9%. From the above, the superiority of the proposed method, which connects disconnected subnetworks using betweenness centrality, was confirmed.

*Keywords-Road Network Hierachization; Pathfinding; Vehicle Routing Problem.*

## I. INTRODUCTION

With the recent proliferation of online shopping, delivery companies are required to deliver goods to customers efficiently. In previous research, the problem of efficiently delivering goods to customers has been treated as a vehicle routing problem, in which the optimal route is identified to satisfy multiple constraints while minimizing delivery time as a delivery cost. The objective of the vehicle routing problem is to efficiently deliver goods to nearby customers by coordinating the timing of multiple delivery vehicles, to decrease travel distance, and to reduce the overall delivery time by reducing the number of deliveries.

When optimizing delivery plans, it is essential to have distance information for routes between a delivery base and customers. In order to efficiently compute a route between two arbitrary points, routing algorithms such as the Dijkstra algorithm [1] and the A* algorithm [2] have been developed,

but when the number of customers is large, it is difficult to compute all the necessary routing information in a practical time. For this reason, the Euclidean distance is often used because it is easy to calculate. However, as the Euclidean distance does not take into account the actual path, there can be a discrepancy with the actual travel distance and the required time, making it sometimes inappropriate for solving real-world delivery planning problems. Therefore, there is a demand for an approach that can solve a vast number of pathfinding tasks within a practical timeframe. In addition, when the vehicle routing problem is applied to actual deliveries, various factors such as weather, road conditions, and traffic congestion must be incorporated into the vehicle routing problem.

We regard the kerosene delivery planning as an inventory routing problem and formulate it as an optimization problem that reflects the actual consumption of kerosene and have advanced research on an approximate solution method using Tabu Search [3]. In pathfinding for kerosene delivery planning, it is necessary not only to reduce the computation time for pathfinding but also to take into account the delivery driver's avoidance of snow-covered narrow roads. Kerosene delivery is mainly conducted during the busy winter season in cold regions, and tank trucks are required to travel on snow-covered roads. However, many narrow roads are not cleared of snow during the winter. Tank trucks carrying kerosene may get stuck on narrow roads that have not been cleared of snow, and delivery drivers may choose routes that avoid narrow roads depending on the road conditions. Therefore, it is necessary to reflect the behavior of delivery drivers in route planning, taking into account their tendency to avoid narrow roads under conditions of snow accumulation and snowfall.

In this paper, an application to a vehicle routing problem is assumed, where delivery routes are coordinated and planned over a long period of time in a specific region. This requires performing hundreds of thousands of pathfinding tasks, taking into consideration the driver preferences.

One way to speed up pathfinding is to improve the efficiency of existing pathfinding algorithms. However, in cases where pathfinding needs to be performed hundreds of thousands of times, it becomes challenging to process all pathfinding tasks within a practical time frame, even with an efficient pathfinding algorithm.

While improving the efficiency of existing pathfinding algorithms aims for exact solutions, there are also developments in methods that preprocess graphs as an approximate solution approach. These methods accelerate pathfinding without guaranteeing the optimality of solutions. The vehicle routing problem assumed in this study involves long-term delivery planning in a specific region, requiring repeated use of the road network in that area. For problems with these characteristics, it is believed that a road network hierarchization method is effective. The reasons for this are as follows:

- Once a hierarchical network is constructed, it can be reused as long as the network remains unchanged.
- The use of simplified networks is expected to reduce the computation time for pathfinding.
- By setting the link costs during pathfinding individually according to the driver preferences, pathfinding that takes into account various driver preferences can be performed on the same hierarchical network.

Based on the prior method that implements the degree of avoidance of narrow roads in the form of driver preferences in a hierarchical pathfinding algorithm [4], this paper proposes a road network hierarchization method aimed at optimizing the delivery planning problem.

This paper is structured as follows. In Section 2, we review related work on road network hierarchization methods to clarify the positioning of this paper. In Section 3, we propose a method for connecting subnetworks based on betweenness centrality in road network hierarchization. In Section 4, focusing on actual urban road networks, the computational time for pathfinding and the cost of routes obtained using both the proposed method and previous methods are compared to confirm the effectiveness of our approach. Finally, in Section 5, the paper is concluded.

## II. RELATED WORK

### A. Hierarchical Pathfinding

In hierarchical pathfinding algorithms, there are primarily two methods of road network hierarchization: one is classification-based hierarchization, and the other is aggregation-based hierarchization. In this Section, we will overview the previous methods for these two types of hierarchization.

*1) Classification-Based Hierarchization:* Classification-based hierarchization uses the attributes of each road, such as road type (expressway, national road, prefectural road, city road, etc.) and the number of lanes, to divide the hierarchical level. Fukuda et al. [4] define fixed attributes, which are the same for every driver, and variable costs, which allow for differences among drivers. They propose a hierarchical

pathfinding algorithm that employs fixed attributes for the hierarchization of the road network and uses variable costs for pathfinding. Fukuda et al., based on the findings of previous research [5], which suggests that general drivers tend to prefer roads with a greater number of lanes, have assigned lane count as a fixed attribute. In this hierarchization based on number of lanes, the initially classified upper-level networks often become disconnected. However, their proposed hierarchical pathfinding algorithm prohibits searching from the upper-level networks to the lower-level networks. This leads to failures in pathfinding when any of the networks at the upper levels are disconnected. Therefore, it is necessary to correct these upper-level networks to resolve their disconnectedness. In the method by Fukuda et al. (hereinafter referred to as the 'previous method'), they define a threshold $H_{max}$ to control the extent to which disconnected subnetworks are connected during corrections. To validate the effectiveness of the hierarchization method, they performed pathfinding by representing the driver's avoidance of narrow roads as a preference. As a result, it was confirmed that effective pathfinding results could be outputted using the same hierarchical network for three types of drivers with varying degrees of avoidance of narrow roads. However, a limitation of this method is that the cost of the routes outputted can significantly worsen compared to the minimum cost routes.

*2) Aggregation-Based Hierarchization:* Aggregation-based hierarchization [6][7][8] applies community detection methods to road networks, treating each detected community as a new node to construct a upper-level network. Mahyar et al. [7] employ the Louvain method [9], known for its speed and accuracy, for community detection as part of their hierarchization process. Additionally, in their work, to realize hierarchization that takes into account congestion conditions, they assign travel times or traffic conditions to the link costs of the network. This allows for grouping nodes with similar congestion conditions into the same community, making it possible to achieve hierarchization that takes into account the state of congestion.

Classification-based hierarchization has the advantage of allowing pathfinding that takes into account the road structure, but it has the disadvantage of difficulty in reflecting dynamic conditions during the hierarchization process. On the other hand, aggregation-based hierarchization has the advantage of being able to reflect dynamic conditions during hierarchization, but it has the disadvantage of making pathfinding that considers road structure difficult due to changes in topology. Therefore, it is necessary to choose the method of hierarchization based on the objective. However, this paper focuses on driver preferences based on static attributes such as distance and road width and does not consider dynamic changes in the network (such as road closures due to construction) or traffic conditions. Consequently, we examine a classification-based hierarchization method.

## III. OUR PROPOSED METHOD

### A. Positioning of the Proposed Method

For application to the assumed delivery planning problems, this paper aims to develop a road network hierarchization method that reduces the sum of the construction time of a hierarchical road network and the computation time for hundreds of thousands of pathfinding tasks and takes into account various driver preferences.

We propose a road network hierarchization method that consider the importance of individual nodes in the network base on our previous work [10]. This method is based on the hypothesis that by constructing the upper-level networks around nodes of high importance, pathfinding using the hierarchical road network can effectively pass through these important nodes, and, as a result, calculate routes with less deterioration in the cost of routes.

When constructing a hierarchical road network, the upper-level networks are extracted based on road attributes such as the number of lanes, but these networks often become disconnected. However, The hierarchical pathfinding algorithm proposed by the previous method [4] is adopted in our approach, and this algorithm prohibits searching from upper-level networks to lower-level networks. As a result, pathfinding failures can occur when networks at upper levels are disconnected. To address this, the disconnections are resolved by adding nodes and links to the upper-level networks. The major difference between our method and the previous method for constructing the hierarchical network is that the nodes to be added are selected based on their importance in our approach.

### B. Problem Formulation

In this paper, a problem that extends the general shortest path problem is addressed, but first, the formulation of the general shortest path problem is explained. Given a graph $G = (V, E)$ and the weights of each link $e \in E$ denoted as $w_e \in \mathbb{R}_+$, a sequence of vertices $P = (v_1, v_2, ..., v_k)$ satisfying $e_i = (v_i, v_{i+1}) \in E, i = 1, ..., k-1$ is called a path. The variable $x_e$ indicates whether a link $e$ is included in the path: $x_e = 1$ if it is included, and $x_e = 0$ otherwise. The problem of finding the shortest path from a given start node $s \in V$ to an target node $t \in V$ can then be formulated as follows.

$$\min \sum_{e \in E} w_e x_e \tag{1}$$

$$\text{s.t.} \sum_{e \in \delta^+(v)} x_e - \sum_{e \in \delta^-(v)} x_e = 0, \quad \forall v \in V \setminus \{s, t\} \tag{2}$$

$$\sum_{e \in \delta^+(s)} x_e = 1 \tag{3}$$

$$\sum_{e \in \delta^-(t)} x_e = 1 \tag{4}$$

$$x_e \in \{0, 1\}, \quad \forall e \in E \tag{5}$$

Here, $\delta^+(v)$ denotes the set of links that have vertex $v$ as their starting point, and $\delta^-(v)$ denotes the set of links that

have vertex $v$ as their endpoint. The constraint in (2) represents that, for each visited vertex $v$, exactly one incoming link and one outgoing link are selected. The constraints in (3) and (4) signify that exactly one link leaving the start node $s$ and one link entering the target node $t$ are selected.

In this paper, a classification-based hierarchical road network is used. In constructing the hierarchical road network, a function $L(e)$ is defined to indicate the hierarchical level to which each link $e$ in the original network $G$ belongs. Then, based on the set of links $E^n$ belonging to hierarchical level $n(n = 1, 2, ..., N)$ and the set of nodes $V^n$ that are endpoints of links $e \in E^n$, the network $G^n$ for hierarchical level $n$ is constructed.

In pathfinding using a classification-based hierarchical road network, starting from the original network $G$, a progressive transition between hierarchical levels is made. At each level, it can be considered that the shortest path problem specific to that level is being solved. Finally, the paths calculated at each level are concatenated and outputted. Here, the path at each hierarchical level is denoted as $P^n$. Below, a detailed formulation of the extended pathfinding problem is presented.

$$\min \sum_{n=1}^{N} \sum_{e \in E^n} w_e^n x_e^n \tag{6}$$

$$\text{s.t.} \sum_{e \in \delta^+(v)} x_e^n - \sum_{e \in \delta^-(v)} x_e^n = 0, \tag{7}$$
$$\forall v \in V^n \setminus \{s^n, t^n\}, \forall n \in \{1, \ldots, N\}$$

$$\sum_{e \in \delta^+(s^n)} x_e^n = 1, \quad \forall n \in \{1, \ldots, N\} \tag{8}$$

$$\sum_{e \in \delta^-(t^n)} x_e^n = 1, \quad \forall n \in \{1, \ldots, N\} \tag{9}$$

$$x_e^n \in \{0, 1\} \quad \forall e \in E^n, \quad \forall n \in \{1, \ldots, N\} \tag{10}$$

Here, $N$ represents the maximum hierarchical level, $w_e^n$ is the weight of link $e$ at hierarchical level $n$, and $x_e^n$ is a variable that takes the value of $x_e^n = 1$ if link $e$ is included in the path $P^n$, and $x_e^n = 0$ otherwise. Additionally, $s^n$ and $t^n$ denote the start and target nodes, respectively, at hierarchical level $n$. Equations (7) to (9) represent the application of the general pathfinding constraints at each hierarchical level.

### C. Hierarchization of Road Network

*1) Extraction of Upper-Level Networks:* The road network addressed in this paper is implemented as an undirected graph, which serves to reduce the complexity of the network and enhance computational efficiency. As an example of driver preference, avoidance of narrow roads is adopted, and the upper-level networks are extracted based on the number of lanes, which is closely related to this preference. This driver preference was chosen based on the needs of delivery drivers in our kerosene delivery planning research, who want to take routes that avoid narrow roads not well-cleared of snow.

For a link $e$, let $l_e$ be the smaller number of lanes on one side of the road. However, if link $e$ is a one-way link with zero lanes on one side, then $l_e = 1$. When $n$ is the hierarchical level of the network, define the set of links satisfying $l_e \geq n$ as $E^n$, and the set of nodes at both ends of each link in $E^n$ as $V^n$. The subnetwork composed of $V^n$ and $E^n$ is denoted as $G^n = (V^n, E^n)$, and $G^n$ is the network at level $n$. The original road network is $G^1$, and the upper-level networks $G^n (2 \geq n \geq N)$ are extracted sequentially from $G^1$ ($V^1 \supseteq V^2 \supseteq \cdots \supseteq V^N, E^1 \supseteq E^2 \supseteq \cdots \supseteq E^N$). It is noted that $N$ is a parameter that sets the highest hierarchical level.

*2) Extraction of Representative Node Set:* In the pathfinding method using a hierarchical road network in this paper, the search is started from the original road network and progressively transitions to the upper-level networks. With each transition to an upper-level network, the route candidates are narrowed down. Since the route candidates heavily depend on the nodes included in the upper-level networks, if many nodes of low importance, which are rarely used in routes between any Origin-Destination pair (commonly referred to as OD pair), are included in the upper-level networks, routes that pass through these less important nodes may be output, potentially leading to a significant deterioration in route cost. Therefore, adding nodes of high importance to the upper-level networks is crucial for ensuring the accuracy of approximate solutions in pathfinding using the hierarchical road network.

Centrality measures are useful in measuring the importance of nodes. These measures evaluate how central elements are within a network, with common types including degree centrality, closeness centrality, and betweenness centrality [11]. Especially, betweenness centrality, which evaluates each node based on the frequency of its use in the shortest paths between any OD pair, is valuable as a measure of node importance when constructing hierarchical road networks.

While betweenness centrality typically evaluates each node individually, it is important to consider the interdependencies and cooperative relationships among nodes when taking betweenness centrality into account, as nodes are mutually dependent on each other.

Therefore, Fushimi et al. [12] have proposed what is known as set betweenness centrality, which considers the centrality of a set of nodes.

Betweenness centrality and set betweenness centrality are calculated by (11) and (12), respectively.

$$bwc(v) = \sum_{s \in V} \sum_{t \in V} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \tag{11}$$

$$SB(R) = \sum_{s \in V} \sum_{t \in V} \frac{\sigma_{s,t}(R)}{\sigma_{s,t}} \tag{12}$$

Here, $\sigma_{s,t}$ represents the number of shortest paths between nodes $s$ and $t$, $\sigma_{s,t}(v)$ is the number of these paths passing through node $v$, and $\sigma_{s,t}(R)$ denotes the number of shortest paths from node $s$ to node $t$ that pass through $\forall r \in R$. The set of nodes $R$ that maximizes (12) is considered the representative node set.

It should be noted that calculating the exact betweenness centrality requires computing the shortest paths for every OD pair in the target road network. However, this computation requires $O(V^3)$ time complexity and is challenging for large-scale networks. Therefore, a method has been proposed that approximates betweenness centrality by randomly sampling nodes from the network, performing single-source shortest pathfinding from these sampled nodes, and using the obtained paths information for the approximation [13]. This method has been shown to have good approximation accuracy in the paper by Wandelt et al. [14], and it is also adopted in this paper.

While betweenness centrality can be calculated by counting the number of times each node in the network based on the obtained paths information, for set betweenness centrality, it is necessary to consider which combination of nodes maximizes the objective function in (12). Therefore, in this paper, the greedy algorithm proposed by Fushimi et al. is used to find the set $R$ that maximizes (12). The number of elements in the representative node set $R$, denoted as $|R|$, is adjusted by the parameter $K$. For specific procedures, refer to the paper by Fushimi et al. [12].

*3) Correction of Upper-Level Networks:* The correction is applied to the upper-level networks from $G^2$ to $G^N$. This correction process mainly consists of two stages: resolving disconnections and adding the Representative node set $R$ identified in Section III-C2.

When extracting the upper-level networks based on the attributes of links, these networks may be disconnected. In such cases, an operation to resolve this by adding nodes and links is necessary. Therefore, our proposed method first resolves the disconnection in the upper-level networks. However, merely resolving the disconnection in the upper-level networks might lead to an excessive simplification, potentially resulting in the output of routes from pathfinding using the hierarchical road network with significantly worse costs compared to the minimum cost routes. To improve the accuracy of the approximate solutions in pathfinding using the hierarchical road network, the representative node set $R$ is added.

*4) Pathfinding Using Hierarchical Network:* In this paper, the pathfinding method proposed by the previous method [4] is adopted. This method's overview involves bi-directional searching from both the start and target nodes, and the transition to an upper-level network occurs only when searches from both directions reach nodes that exist in the network of the next upper hierarchical level. It should be noted that, the transition to an upper-level network is restricted only to cases that meet the above condition, preventing transitions from upper-level to lower-level networks. While any pathfinding algorithm can be used, this paper employs the A* algorithm.

The details are described below. The input consists of a hierarchical network $\{G^1, G^2, ..., G^N\}$, a start node $s$, and a target node $t$, with the output being the path from $s$ to $t$. In the the following procedure, 'advancing the search by one step' refers to extracting the node with the minimum cost from the queue, calculating and updating the costs for all nodes adjacent to the extracted node in $G^n$, and then adding the updated nodes

**Algorithm 1** Correction of Hierarchical Network

---

**Require:** Hierarchical network $\{G^1, G^2, \ldots, G^N\}$, Representative node set $R$
**Ensure:** Corrected hierarchical network
    **for** $n = 2$ to $N$ **do**
        **if** $G^n$ is disconnected **then**
            Select node with highest betweenness centrality in each subnetwork
            Form pairs of selected nodes
            **for** each pair **do**
                Search for a minimum cost path on $G^{n-1}$
                Add nodes and links from the path to $G^n$
            **end for**
        **end if**
        Select node $v^*$ with highest betweenness centrality in $G^n$
        Add all nodes from $R$ to $G^n$
        Form pairs of each $v \in R \cup \{v^*\}$
        **for** each pair **do**
            Search for a minimum cost path on $G^{n-1}$
            Add nodes and links from the path to $G^n$
        **end for**
    **end for**

---

Figure 1.  Algorithm for correction of hierarchical network.

to the queue.
Step 1:  Start the search with the level of the network to be searched set to $n = 1$.
Step 2:  Place the start node $s$ in the forward search queue, and the target node $t$ in the backward search queue.
Step 3:  If the forward search has not yet reached a node contained in $G^{n+1}$, advance the forward search by one step.
Step 4:  If the backward search has not yet reached a node contained in $G^{n+1}$, advance the backward search by one step.
Step 5:  If there are nodes already searched from both directions, conclude the search and determine the route.
Step 6:  If searches from both directions have reached nodes contained in $G^{n+1}$, use these nodes as new start and target nodes, clear the queues, update $n \leftarrow n+1$, and return to Step 2. Otherwise, maintain $n$ as it is and return to Step 3.

## IV. EXPERIMENT

Hierarchical road networks for Japan's heavy snowfall areas are constructed using three different methods, each with a partially different process after the extraction of the upper-level networks. the construction times of the hierarchical road networks and the sizes of the networks at each hierarchical level are compared. Pathfinding using the hierarchical networks of each method in the road network is also performed for OD pairs. the sum of the construction time of the hierarchical networks and the total computation time for individual

pathfinding are then evaluated, as well as the increase in cost relative to the minimum cost paths.

### A. Experimental Setup

A 14 km square area centered on Sapporo Station was extracted from OpenStreetMap [15] as an area where kerosene delivery is routinely conducted due to heavy snowfall in winter, and was used as the road network for the experiment (Figure 2). Note that approximately 85% of this road network consists of links with only one lane. Therefore, hierarchization based solely on the number of lanes might lead to excessively simplified upper-level networks, and some ingenuity is required to correct the upper-level networks. Lane information and link lengths are acquired from the OpenStreetMap road network for use in the hierarchization of the network and the calculation of route costs.

In this paper, three different hierarchization methods are evaluated, each differing in how disconnections are resolved, and a representative node set is added to the upper-level networks, which are extracted based on the number of lanes. Linkage method 1 resolves the disconnections using the previous method and does not add a representative node set. Linkage method 2 first resolves the disconnections using the previous method and then adds the representative node set using the proposed method. Linkage method 3, which is the method proposed in this paper, uses the proposed method to both resolve the disconnections and add the representative node set.

The parameter settings for each method are shown in Table II. Within Table II, $H_{max}$ is a threshold value that controls the extent to which disconnected subnetworks are linked during the correction process. For linkage methods 1 and 2, $H_{max} = 100$ was chosen because it yielded the highest accuracy of approximate solutions in preliminary experiments. $K$ represents the number of elements in the representative node set to be added to the upper-level networks. The upper limit of hierarchical levels for each method was set to $N = 3$.

When adding the representative node set in linkage methods 2 and 3, an approximation of betweenness centrality is necessary. The number of nodes randomly sampled from the network for the approximation of betweenness centrality was set to 1,000. This number can vary depending on the network used and its scale and is not strictly required to be this value.

Pathfinding is performed using the hierarchical networks constructed by each method, and each method is evaluated based on the results. Additionally, as an exact solution method, pathfinding without using a hierarchical network is also performed. When performing pathfinding using a hierarchical network, the method described in Section III-C4 is adopted, and when performing pathfinding without hierarchical networks, the unidirectional search A* algorithm is adopted.

OD pairs are classified based on Euclidean distance and divided into seven intervals at every 2.5 km. From each interval, 5,000 pairs are randomly selected, and pathfinding is performed for each OD pair, with the results compared.

TABLE I
VALUES OF $w_{e,i}$ FOR THE LINK COSTS ACCORDING TO EACH DRIVER
PREFERENCE TYPE FOR NARROW ROADS

| Weight | Number of lanes | | |
|---|---|---|---|
| | $l_e = 1$ | $l_e = 2$ | $l_e \geq 3$ |
| $w_{e,1}$ | 10 | 5 | 1 |
| $w_{e,2}$ | 4 | 2.5 | 1 |
| $w_{e,3}$ | 2.768 | 1.607 | 1 |
| $w_{e,4}$ | 1 | 1 | 1 |

TABLE II
PARAMETER SETTINGS FOR EACH LINKAGE METHOD

| Algorithm | Resolution of disconnection | Addition of representative node set | $H_{max}$ | K |
|---|---|---|---|---|
| Linkage method1 | previous method | - | | - |
| Linkage method2 | previous method | proposed method | 100 | 10 20 : 100 |
| Linkage method3 | proposed method | proposed method | - | 10 20 : 100 |

Next, multiple types of driver preferences are prepared, each with different degrees of avoidance of narrow roads. As the degree of avoidance of narrow roads increases, the cost of such roads also increases, which may change the total cost of the candidate routes and thus the final route selected. In this experiment, four types of driver preferences $A_1, A_2, A_3, A_4$ are prepared in order of increasing the degree of avoidance of narrow roads. For each type of driver preference $A_i(i = 1, 2, 3, 4)$, the link cost $\tilde{c}_{e,i}$, considering the number of lanes $l_e$ of link $e$, is given by (13). Here, $c_e$ is the original link cost, which in this paper is given as the Euclidean distance between the endpoints of the link. $w_{e,i}$ is the weight applied to the link cost according to the driver's avoidance of narrow roads. The values of link cost used in this experiment for $A_1, A_2, A_3$ are adopted from those calculated in [4]. The weight $w_{e,3}$ was set based on actual probe data in [4], and $w_{e,1}$ and $w_{e,2}$ were values inferred by the authors of [4] intended to represent varying degrees of avoidance to narrow roads. Additionally, this experiment incorporates a driver preference type $A_4$, representing drivers with no avoidance of narrow roads. This driver type was added to evaluate the performance of the proposed method with and without avoidance to narrow roads.

$$\tilde{c}_{e,i} = w_{e,i}c_e \tag{13}$$

This experiment was conducted in a computing environment equipped with an AMD EPYC 7402 24-Core CPU and 128GB of memory.

### B. Size and Construction Time of Hierarchical Network

The sizes and construction times of the hierarchical networks constructed by each linkage method are shown in Table III. For each method, $n = 1$ is the original road network,
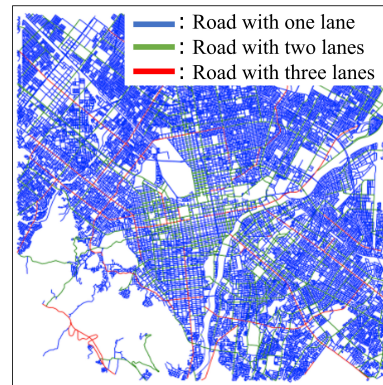


Figure 2. Road network around Sapporo Station.

and for $n \geq 2$, networks of different sizes are constructed depending on the method and parameters.

As shown in Table III, the network at hierarchical level 2 in linkage method 3, even with $K = 100$, has approximately 43% fewer nodes and about 52% fewer links compared to the network at hierarchical level 2 in linkage method 1. On the other hand, no significant differences were observed in the number of nodes and links in the hierarchical networks for levels $n = 2, 3$ between linkage method 1 and linkage method 2.

While the construction of the hierarchical network using linkage method 1 takes about 1 minute, the construction time using linkage methods 2 and 3 increases by approximately 600 times for $K = 100$. However, the hierarchical networks constructed in this paper are intended for application in the vehicle routing problem where hundreds of thousands of pathfinding tasks occur, and where delivery routes are repeatedly adjusted and planned in specific regions. Therefore, once a hierarchical network is constructed, it can be reused as long as there are no changes to its topology or target area. Considering this, the construction time for the hierarchical networks in linkage methods 2 and 3 is acceptable.

### C. Results of Pathfinding Using Hierarchical Network

The results of pathfinding using the hierarchical networks constructed by each method were compared with the results of pathfinding using the unidirectional search A* algorithm, which does not use hierarchical networks. Table IV shows the average costs and computation times for each section for the driver preference type $A_1$, which has the highest degree of avoidance of narrow roads among four driver preference types. Note that only a portion of the parameter $K$ results for linkage methods 2 and 3 are presented. In Table IV, $T_{ave}$ and $C_{ave}$ represent the average computation time and cost of the routes for each section, respectively, with bold figures indicating the lowest average costs among the routes using hierarchical networks in each section. Uni-A* refers to the unidirectional search A* algorithm.

The results in Table IV show that compared to the unidirectional search A* algorithm, the calculation time of individual

TABLE III
SIZE AND ELAPSED TIME OF HIERARCHICAL NETWORKS CONSTRUCTED BY EACH METHOD

| Algorithm | $H_{max}$ | $K$ | $n$ | number of nodes | number of links | elapsed time |
|---|---|---|---|---|---|---|
| - | - | - | 1 | 31,139 | 49,967 | - |
| Linkage method1 | 100 | - | 2 | 12,691 | 15,765 | 57s |
| | | | 3 | 5,267 | 5,790 | |
| Linkage method2 | 100 | 10 | 2 | 12,691 | 15,765 | 6h28m21s |
| | | | 3 | 5,315 | 5,849 | |
| | | 100 | 2 | 12,946 | 16,065 | 9h40m33s |
| | | | 3 | 6,648 | 7,319 | |
| Linkage method3 | - | 10 | 2 | 7,219 | 7,618 | 6h26m51s |
| | | | 3 | 2,798 | 2,865 | |
| | | 100 | 2 | 7,724 | 8,211 | 9h38m7s |
| | | | 3 | 5,076 | 5,321 | |

pathfinding is reduced to 4-6% with linkage method 1 or 2, and to 3-6% with linkage method 3, confirming that the hierarchization of road networks contributes significantly to reducing the calculation time of pathfinding.

Regarding the average cost of routes for each section, the case of $K = 100$ in linkage method 3 had the minimum average route cost in all sections among the three linkage methods. The worsening of route costs compared to the average cost of routes from the unidirectional search A* algorithm was about 4-9%. This characteristic was consistent for other driver preference types as well, suggesting that linkage method 3 could be an effective approach for constructing hierarchical networks that consider the different preferences for each driver.

### D. Discussion

The construction time of the hierarchical road network in linkage method 3 was approximately 9 hours and 40 minutes, about 600 times longer than that of linkage method 1. The processes involved in constructing the hierarchical network in linkage method 3 are: (1) extraction of the upper-level networks, (2) sampling of nodes and single-source shortest pathfinding starting from the sampled nodes, (3) approximation of set betweenness centrality and extraction of the representative node set, and (4) correction of the upper-level networks. The most time-consuming process among these is process (3), which takes approximately 8 hours and 40 minutes, or 90% of the construction time. This is because the paths obtained in process (2) are used as references to calculate the representative node set. However, as the number of paths referenced increases, the computation time of process (3) also increases. Therefore, the computation time of process (3) depends on the number of paths obtained from process (2). Single-source shortest pathfinding was performed, starting from sampled nodes (for this paper, 1,000 nodes were randomly sampled), with paths being computed to all other nodes in the network. However, it may not be necessary to compute paths to all other nodes, and reducing the number of paths to be calculated could potentially reduce the construction time of the hierarchical network.

Figure 3 shows the cost ratio compared to the minimum cost route for the driver preference type $A_1$. Note that Figure 3 is a histogram of the cost ratios for all OD pairs. From Figure 3, it can be observed that using linkage method 3 allows for the calculation of routes with cost ratios closer to 1.0 for more OD pairs compared to linkage methods 1 and 2. However, the number of OD pairs with a cost ratio exceeding 2.0 was 47 for linkage method 1, 15 for linkage method 2, and 30 for linkage method 3, indicating that even with linkage method 3, some OD pairs experienced routes with significantly worse costs compared to their respective minimum cost routes. The worsening of route costs occurred in cases where the OD pairs were closely located. In response to this issue, hierarchical networks are currently used regardless of the OD pair distance, but we will consider improvements, such as deciding whether to use hierarchical networks based on the distance of the OD pairs.

Moreover, there is potential for improvement in the method of extracting the representative node set. In the proposed method, the set betweenness centrality is calculated from the entire target network, and the representative node set is extracted based on this calculation. However, road networks have regional characteristics, and the nodes that are frequently traversed should differ depending on the movement between and within regions. The proposed method extracts the representative node set without taking this into account, and thus may miss frequently traversed nodes, especially within a specific region. We will consider an approach that divides the road network into multiple regions, calculates the set betweenness centrality for each movement between and within regions, and extracts the representative node set.

### V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a road network hierarchization method aimed at optimizing vehicle routing problems. Our method involves constructing a hierarchical road network that incorporates the concept of set betweenness centrality, an extension of betweenness centrality, which is one of the centrality measures. Both the previous method and the proposed method were applied for pathfinding on the road network around Sapporo Station, and for various types of driver preference with different degrees of avoidance of narrow roads. Compared to performing pathfinding without hierarchical networks, the calculation time of individual pathfinding was reduced to 4-6% with the previous method and to 3-6% with our method. Additionally, by using our method, the average cost of routes improved in all sections compared to the previous method,

TABLE IV

RESULTS OF PATHFINDING BY EACH METHOD (DRIVER PREFERENCE TYPE $A_1$)

| Algorithm | $H_{max}$ | $K$ | Section[km] | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0-2.5 | | 2.5-5.0 | | 5.0-7.5 | | 7.5-10.0 | | 10.0-12.5 | | 12.5-15.0 | | 15.0-17.5 | |
| | | | $T_{ave}$ [sec] | $C_{ave}$ [km] | $T_{ave}$ [sec] | $C_{ave}$ [km] | $T_{ave}$ [sec] | $C_{ave}$ [km] | $T_{ave}$ [sec] | $C_{ave}$ [km] | $T_{ave}$ [sec] | $C_{ave}$ [km] | $T_{ave}$ [sec] | $C_{ave}$ [km] | $T_{ave}$ [sec] | $C_{ave}$ [km] |
| Uni-A* | - | - | 0.50 | 11.66 | 1.42 | 16.92 | 2.74 | 22.16 | 4.15 | 27.34 | 5.76 | 32.75 | 6.95 | 37.33 | 7.62 | 39.18 |
| Linkage method1 | 100 | - | 0.03 | 13.00 | 0.07 | 18.53 | 0.13 | 23.96 | 0.21 | 29.43 | 0.30 | 35.13 | 0.38 | 40.11 | 0.48 | 42.15 |
| Linkage method2 | 100 | 20 | 0.03 | 12.97 | 0.07 | 18.49 | 0.13 | 23.91 | 0.21 | 29.38 | 0.31 | 35.11 | 0.40 | 40.10 | 0.51 | 42.14 |
| | | 60 | 0.03 | 12.95 | 0.07 | 18.38 | 0.14 | 23.72 | 0.22 | 28.98 | 0.33 | 34.71 | 0.42 | 39.76 | 0.54 | 42.15 |
| | | 100 | 0.03 | 12.83 | 0.07 | 18.22 | 0.14 | 23.59 | 0.23 | 28.81 | 0.35 | 34.51 | 0.45 | 39.68 | 0.56 | 41.83 |
| Linkage method3 | - | 20 | 0.03 | 12.88 | 0.05 | 18.21 | 0.09 | 23.67 | 0.13 | 29.14 | 0.18 | 34.69 | 0.22 | 39.34 | 0.28 | 41.49 |
| | | 60 | 0.02 | 12.82 | 0.05 | 18.12 | 0.09 | 23.42 | 0.14 | 28.71 | 0.20 | 34.28 | 0.25 | 39.06 | 0.30 | 41.32 |
| | | 100 | 0.02 | **12.67** | 0.05 | **18.03** | 0.09 | **23.31** | 0.15 | **28.54** | 0.21 | **34.13** | 0.27 | **39.05** | 0.32 | **40.76** |
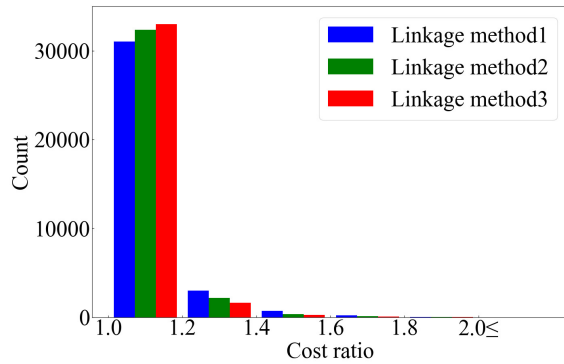


Figure 3. Cost ratio to the minimum cost paths (driver preference type $A_1$).

and even when compared to the minimum cost paths, the worsening of route costs was about 4-9%. From the above, our method demonstrated superior performance in terms of pathfinding computation time and the cost of the routes obtained, surpassing the previous method.

## REFERENCES

[1] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 269, p. 271, 1959.

[2] G. R. Jagadeesh, T. Srikanthan, and K. Quek, "Heuristic techniques for accelerating hierarchical routing on road networks," *IEEE Transactions on intelligent transportation systems*, vol. 3, no. 4, pp. 301–309, 2002.

[3] H. Ooe, S. Yokoyama, T. Yamashita, H. Kawamura, and M. Tada, "Optimization of kerosene delivery planning using tabu search," no. 15, 2022, pp. 1–8.

[4] S. Fukuda, K. Abe, H. Fuji, T. Yamada, and S. Yoshimura, "Layered route search method for large-scale multi-agent-based traffic simulation," *IPSJ Journal*, vol. 59, no. 7, pp. 1435–1444, 2018.

[5] T. Nakamura, T. Yoshii, and R. Kitamura, "Development of a simplified network which represents all roadway links," *Infrastructure planning review*, vol. 23, pp. 441–446, 2006.

[6] Q. Song and X. Wang, "Efficient routing on large road networks using hierarchical communities," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 132–140, 2011.

[7] M. M. Nejad, L. Mashayekhy, and R. B. Chinnam, "Effects of traffic network dynamics on hierarchical community-based representations of large road networks," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*, 2012, pp. 1900–1905.

[8] D. N. A. Mensah, H. Gao, and L. W. Yang, "Approximation algorithm for shortest path in large social networks," *Algorithms*, vol. 13, no. 2, 2020.

[9] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 10, p. P10008, 2008.

[10] M. Shimizu, S. Yokoyama, T. Yamashita, and H. Kawamura, "A proposal of subnetwork linkage method based on betweenness centrality in road network hierarchization," vol. 2023-ITS-95, no. 18, 2023, pp. 1–8.

[11] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[12] T. Fushimi, K. Saito, T. Ikeda, and K. Kazuhiro, "Proposing set betweenness centrality measures focusing on nodes' collaboratative behaviors and its application," *The IEICE transactions on information and systems*, vol. J96-D, no. 5, pp. 1158–1165, May 2013.

[13] U. Brandes and C. Pich, "Centrality estimation in large networks," *International Journal of Bifurcation and Chaos*, vol. 17, no. 7, pp. 2303–2318, 2007.

[14] W. Sebastian, S. Xing, and S. Xiaoqian, "Scalability of betweenness approximation algorithms: An experimental review," *IEEE Access*, vol. 7, pp. 104 057–104 071, 2019.

[15] OpenStreetMap contributors, "Openstreetmap," Available: https://www.openstreetmap.org, 2004, [retrieved: January, 2024].

# Evaluation of Request Order Decision Strategy in the Selection of Substitute Employees for Shift Management Tasks

Tomoya Chisaka
*Graduate School of Information Science and Technology, Hokkaido University*
Sapporo, Hokkaido, Japan
chisaka-t@ist.hokudai.ac.jp

Soichiro Yokoyama
*Faculty of Information Science and Technology, Hokkaido University*
Sapporo, Hokkaido, Japan
yokoyama@ist.hokudai.ac.jp

Tomohisa Yamashita
*Faculty of Information Science and Technology, Hokkaido University*
Sapporo, Hokkaido, Japan
yamashita@ist.hokudai.ac.jp

Hidenori Kawamura
*Faculty of Information Science and Technology, Hokkaido University*
Sapporo, Hokkaido, Japan
kawamura@ist.hokudai.ac.jp

*Abstract*—In workplaces with shift-based schedules, managers are burdened with the task of modifying work schedules when absences occur. In this process, known as schedule adjustment due to employee absenteeism, the manager must select a replacement employee (substitute employee) from those originally scheduled to be off-duty on the day of the absence. Within this methodology, a task arises where the manager needs to request substitute attendance from employees who were originally scheduled to be off-duty. The order in which these requests are made to employees is crucial as it directly impacts the burden on both the manager and the employees. This paper proposes a strategy for determining the order of requests based on the probability of acceptance by employees. A simulation model is constructed, and evaluations are conducted for multiple parameter sets. The results of the verification indicate that, when prioritizing the reduction of the understaffed workforce, an ascending request strategy is effective. On the other hand, if prioritizing the manager's burden is essential, a descending request strategy proves to be effective.

*Index Terms*—nurse scheduling, rescheduling, substitute attendance request.

## I. Introduction

Shift management, crucial in workplaces with variable work hours like hospitals and call centers, encompasses two main tasks: shift schedule generation and modification. Schedule generation involves creating schedules based on constraints like maximum workdays [1]–[3]. Schedule modification, on the other hand, adjusts schedules due to absences. Conventional methods involve regenerating entire schedules for the absence period, but frequent absences complicate communication and shift changes.

In contrast to this approach, there is a method where an employee (substitute) who does not have a scheduled shift on the day of the absence is selected to fill in for the absentee, and only a partial modification of the schedule is made. In this method, since the majority of the schedule remains unchanged, it does not impose a significant burden on the manager or the employees. However, the manager needs to perform the task of requesting the selected employee to work as a substitute on the day of the absence, and the order in which these requests are made (request sequence) becomes crucial. For instance, if

requests are made to specific employees only, those employees may feel a sense of unfairness. On the other hand, an increase in the number of requests from the manager to the employees may result in a significant burden for the manager. Moreover, if a substitute cannot be found, it may disrupt the execution of the task itself.

This paper proposes and evaluates request order determination strategies for selecting substitute employees in shift work systems with frequent absenteeism. Metrics include the insufficient number of employees (task stability), the number of requests (manager's burden), the number of times employees worked as substitutes, and The number of requests received from managers. The insufficient number of employees is prioritized, especially in shift-based workplaces. Using a simulation model, the paper assesses proposed strategies across various parameter sets, identifying characteristics leading to high insufficient employee counts and manager-initiated requests. Differences in these metrics between strategies are discussed.

The simulation model incorporates employees' days off requests and utilizes mathematical optimization solvers to generate shift schedules based on various constraints. Absentees are introduced probabilistically on the generated schedules, and simulations of manager-initiated substitute attendance requests are conducted. The goal is to conclude with effective request order determination strategies depending on the situation, considering the number of times employees acted as substitutes, the number of requests received by employees, and other factors. The paper concludes by summarizing the research in the fifth section.

The structure of this paper is outlined as follows: Section 2 covers related research, section 3 describes the simulation model, section 4 presents the experimental setup, experimental results, and their discussion, and section 5 provides the conclusion of this paper.

## II. Related Work

### A. Modification of work schedules

This study explores heuristic solutions to dynamic nurse scheduling problems arising from sudden absenteeism, particularly in the context of modifying work schedules in shift management [4]–[7]. Focusing on the hospital setting as a representative workplace with shift work, the research treats the adjustment of schedules due to nurse absenteeism as a dynamic scheduling challenge. The approach involves selecting a substitute attendance and simultaneously revising the entire schedule for the subsequent period to ensure feasibility when a nurse is absent on a given day.

This method addresses the challenge of modifying a nurse's schedule due to another nurse's absenteeism, which can pose a significant burden. The paper discusses a method for selecting substitute attendance by making requests on the day of absenteeism to employees not originally scheduled to work but who can accept substitute attendance without violating schedule constraints established at the time of generation.

### B. Substitute attendance request

In research on the development of methods for requesting substitute attendance, a study has been conducted using real-world data from a call center to examine the relationship between the order of requests to substitute workers and the percentage of understaffed employees [8]. Building upon this study, our paper goes beyond real-world data and conducts simulations of substitute attendance requests for multiple parameter sets. We evaluate the impact of changing the order of requests in these simulations. The various parameter sets verified in our study were constrained to realistic scenarios based on the real-world data from the aforementioned study. Unlikely scenarios, such as all employees being absent on every date within a given period or the existence of employees who always accept substitute attendance requests with a 100% probability, were excluded from the verification.

## III. Simulation model

In the real world, various methods such as phone calls, messaging apps, and oral communication are used for substitute attendance requests. Messaging apps allow simultaneous requests to multiple employees, but there's a concern about securing more substitutes than needed. Oral requests offer prompt responses but can only be made in person, limiting requests to present employees. This paper focuses on "phone calls," a common method for sudden absences, and conducts simulations. We assume one manager making individual substitute attendance requests to each employee.

### A. Generation of a shift schedule through the nurse scheduling problem

With $m$ employees, a duration of $n$ days, and three work shifts each day, the symbols used for the formulation of schedule generation are defined as follows. Additionally, each employee is assumed to submit their preferred days off in

TABLE I: PARAMETERS AND SYMBOLS OF THIS MODEL

| Parameter | Symbol |
|---|---|
| Number of employees | $X$ |
| Number of dates | $d$ |
| Probability of absence per person | $q$ |
| Maximum request acceptance count per person | $m$ |
| Acceptance probability of low acceptance level | $p_{low}$ |
| Acceptance probability of high acceptance level | $p_{high}$ |
| Number of employees with low acceptance level | $n_{low}$ |
| Number of employees with high acceptance level | $n_{high}$ |

advance. We utilized the general-purpose mathematical optimization solver, CPLEX [9].

$M = \{1, 2, ..., m\}$ : Set of employees

$N = \{1, 2, ..., n\}$ : Set of dates

$W = \{1, 2, 3\}$ : Set of time slots

$A = \{(i, j), i \in M, j \in N |$

Day off request for employee $i$ on date $j\}$

:Set of employees and day pairs for which day-off requests have been submitted.

$a_k, b_k$ : Minimum (Maximum) number of employees on duty for time slot $k$ per day

$e$ : Maximum number of working days

$r$ : Maximum consecutive working days

$s$ : Maximum consecutive working days for time slot 3.

The decision variable that takes the value 1 when employee i works during time slot $k$ on day $j$, and 0 when not working, is denoted as $x_{i,j,k}$. A feasible work schedule, satisfying all the following constraints for $X = \{x_{i,j,k}, i \in M, j \in N, k \in W\}$ is considered as one work schedule plan.

$$\sum_{k \in W} x_{i,j,k} = 1 \qquad i \in M, j \in N \tag{1}$$

$$a_k \le \sum_{i \in M} x_{i,j,k} \le b_k \quad j \in N, k \in W \tag{2}$$

$$\sum_{j \in N} \sum_{k \in W} x_{i,j,k} \le e \qquad i \in M \tag{3}$$

$$\sum_{l=0}^{r} \sum_{k \in W} x_{i,j+l,k} \le r \quad i \in M, j \in \{1, ..., n-r\} \tag{4}$$

$$\sum_{l=0}^{s} x_{i,j+l,2} \le s \qquad i \in M, j \in \{1, ..., n-s\} \tag{5}$$

$$x_{i,j,2} + x_{i,j+1,0} \le 1 \quad i \in M, j \in \{1, ..., n-1\} \tag{6}$$

$$x_{i,j,k} = 0 \qquad i \in M, j \in N, k \in W, (i,j) \in A \tag{7}$$

### B. Occurrence of absence and substitute attendance request

In the schedule generated in the previous section, absence occurrences and the process of securing substitute workers are performed day by day from the first day to the last day. In this model, to represent the sudden absence of employees, it is assumed that employee absence notifications are received on the day before the absence.

TABLE II: VERIFIED PARAMETER SETS

| | $q$ | $m$ | $m$:increment | $p_{low}$ | $p_{high}$ | $p_{high}$:increment | $n_{low}$ | $n_{high}$ |
|---|---|---|---|---|---|---|---|---|
| $q = 0.05$ | 0.05 | $2 \leq m \leq 3$ | 1 | 0.05 | $0.40 \leq p_{high} \leq 0.50$ | 0.10 | 40 | 10 |
| | 0.05 | $2 \leq m \leq 6$ | 1 | 0.05 | $0.50 \leq p_{high} \leq 0.90$ | 0.10 | 45 | 5 |
| $q = 0.10$ | 0.10 | $2 \leq m \leq 4$ | 1 | 0.05 | $0.50 \leq p_{high} \leq 0.70$ | 0.20 | 35 | 15 |
| | 0.10 | $2 \leq m \leq 6$ | 1 | 0.05 | $0.70 \leq p_{high} \leq 0.90$ | 0.20 | 40 | 10 |
| | 0.10 | $2 \leq m \leq 6$ | 1 | 0.10 | $0.50 \leq p_{high} \leq 0.90$ | 0.20 | 40 | 10 |
| | 0.10 | $2 \leq m \leq 4$ | 1 | 0.10 | 0.50 | 0 | 35 | 15 |
| | 0.10 | $2 \leq m \leq 13$ | 1 | 0.15 | $0.50 \leq p_{high} \leq 0.90$ | 0.20 | 45 | 5 |
| | 0.10 | $2 \leq m \leq 6$ | 1 | 0.15 | 0.50 | 0 | 40 | 10 |
| | 0.10 | $2 \leq m \leq 4$ | 1 | 0.15 | 0.50 | 0 | 35 | 15 |
| | 0.10 | $2 \leq m \leq 13$ | 1 | 0.20 | $0.50 \leq p_{high} \leq 0.70$ | 0.20 | 45 | 5 |
| $q = 0.15$ | 0.15 | $2 \leq m \leq 6$ | 1 | 0.05 | 0.90 | 0 | 35 | 15 |
| | 0.15 | $2 \leq m \leq 6$ | 1 | 0.10 | $0.70 \leq p_{high} \leq 0.90$ | 0.20 | 35 | 15 |
| | 0.15 | $2 \leq m \leq 10$ | 1 | 0.10 | 0.90 | 0 | 40 | 10 |
| | 0.15 | $2 \leq m \leq 6$ | 1 | 0.15 | $0.70 \leq p_{high} \leq 0.90$ | 0.20 | 35 | 15 |
| | 0.15 | $2 \leq m \leq 10$ | 1 | 0.15 | $0.70 \leq p_{high} \leq 0.90$ | 0.20 | 40 | 10 |
| | 0.15 | $2 \leq m \leq 6$ | 1 | 0.20 | $0.50 \leq p_{high} \leq 0.90$ | 0.20 | 35 | 15 |
| | 0.15 | $2 \leq m \leq 10$ | 1 | 0.20 | $0.50 \leq p_{high} \leq 0.90$ | 0.20 | 40 | 10 |
| | 0.15 | $2 \leq m \leq 20$ | 1 | 0.20 | 0.90 | 0 | 45 | 5 |
| | 0.15 | $2 \leq m \leq 10$ | 1 | 0.25 | $0.50 \leq p_{high} \leq 0.90$ | 0.20 | 40 | 10 |
| | 0.15 | $2 \leq m \leq 20$ | 1 | 0.25 | $0.50 \leq p_{high} \leq 0.90$ | 0.20 | 45 | 5 |

For each $i,k$ pair that satisfies $x_{i,j,k} = 1$ for date $j$ and $k \in W$, the absence is triggered by updating $x_{i,j,k} = 0$ with a probability $p$. Substitute attendance candidates for time slot $k$ on date $j$ are employees who satisfy $\sum_{k \in W} x_{i,j,k} = 0$ and do not violate the constraints in Equations (3), (4), (5), and (6) in Section III-A.

Substitute attendance candidates are determined in the order of requests, and requests are made one by one in order. The employee who receives the request responds with acceptance or rejection based on a probability, which is referred to as the acceptance probability. When the request is accepted, we update it to $x_{i,j,k} = 1$. The request is concluded when the number of acceptances matches the number of occurred absences. Otherwise, we proceed to the next substitute worker for the request. When responses are obtained from all substitute employees, if the number of acceptances does not match the number of occurred absences, we consider the substitute for date $j$ as insufficient. This process is repeated sequentially for all dates.

To prevent a specific employee from disproportionately accepting substitute attendance requests, set an upper limit on the number of times each employee can accept requests within a given period. This upper limit is referred to as the maximum acceptance count.

In this simulation model, we categorize employees into two groups based on their acceptance probabilities: those with a low acceptance probability, denoted as $p_{low}$, and those with a high acceptance probability, denoted as $p_{high}$. Additionally, we represent the assigned number of employees in each category as $n_{low}$ and $n_{high}$, respectively.

## C. Request Order Decision Strategy

The evaluation in this paper focuses on acceptance probability-based request order determination strategies: descending request strategy, ascending request strategy, and random reuest strategy. The descending request strategy prioritizes employees with high acceptance probabilities, likely re-ducing the overall number of requests. Conversely, the ascending request strategy targets employees with low acceptance probabilities first, potentially conserving high-acceptance employees with high acceptance levels for later requests and mitigating the insufficient number of employees. The random request strategy serves as an intermediate approach between descending and ascending request strategies.

## IV. EXPERIMENT

### A. Experimental purpose

Simulations will evaluate three request order determination strategies, as discussed in Section III-C, using multiple parameter sets presented in Table I. Notably, $X = n_{low} + n_{high}$, and $p_{low} < p_{high}$. For this study, the number of employees is set to $X = 50$, and the observation period is $d = 28$ days (4 weeks). The goal is to analyze, for each parameter set, the occurrence and extent of the anticipated properties of the strategies outlined in Section III-C.

### B. Validation range of parameter sets

In the simulation of multiple parameter sets, specific conditions are imposed to define the validation range. This study focuses on workplaces with prevalent absenteeism, necessitating a significant occurrence of absences. Therefore, verified absence probabilities, represented by $q$, are set to 0.05, 0.10, and 0.15. Additionally, to address scenarios where an excessively large maximum acceptance count $m$ might lead to only employees with high acceptance levels, resulting in minimal the insufficient number of employees and task execution instability, conditions are added for the number $n_{high}$ of employees with high acceptance levels. Here, $F$ denotes the total number of absences over the entire period.

$$n_{high}m < F \tag{8}$$

TABLE III: PARAMETER SETS WITH THE LARGEST INSUFFICIENT NUMBER OF EMPLOYEES

| Rank | 1st | 2nd | 3rd | 4th | 5th |
|---|---|---|---|---|---|
| $X$ | 50 | 50 | 50 | 50 | 50 |
| $d$ | 28 | 28 | 28 | 28 | 28 |
| $q$ | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 |
| $m$ | 2 | 2 | 2 | 3 | 2 |
| $p_{low}$ | 0.05 | 0.10 | 0.10 | 0.05 | 0.10 |
| $p_{high}$ | 0.90 | 0.90 | 0.90 | 0.90 | 0.70 |
| $n_{low}$ | 35 | 40 | 35 | 35 | 35 |
| $n_{high}$ | 15 | 10 | 15 | 15 | 15 |
| $\zeta$ | 1.821 | 1.472 | 1.354 | 1.352 | 1.339 |

Furthermore, when $p_{low} \approx p_{high}$ it becomes challenging to express the difference in acceptance probabilities between employees. Therefore, the following equation must hold:

$$p_{low} \ll p_{high} \qquad (9)$$

Furthermore, to address situations where employees with both high and low acceptance levels are reasonably mixed, the following equation must hold for $n_{low}$ and $n_{high}$:

$$n_{low} \gg 0 \qquad (10)$$

$$n_{high} \gg 0 \qquad (11)$$

Lastly, it is crucial to consider the relationship between the expected number of substitute attendances and the number of absences. If the expected number significantly surpasses the number of absences, addressing the insufficient number of employees may improve, even without altering the request order. Conversely, if the expected number is considerably lower than the number of absences, mitigating the insufficient number of employees becomes challenging, even with a change in request order. Thus, the verification focused on a range where the following equation holds:

$$1 < \frac{\frac{n_{low}}{n_{low}+n_{high}}cp_{low} + \frac{n_{high}}{n_{low}+n_{high}}cp_{high}}{\frac{F}{d}} < \beta \qquad (12)$$

Here, $\beta$ is the threshold, and $c$ is the average number of substitute attendance candidates per time slot per day. In this study, $c$ is determined based on the average number of candidates derived from pre-generated shift schedules with no absences.

### C. Experimental setup

In Section III-A, we set the minimum (maximum) number of employees on duty for time slot $k$, $a_k$, $b_k$ uniformly as $a_1 = a_2 = a_3 = b_1 = b_2 = b_3 = 8$ for a straightforward comparison. Under these conditions, the total number of absences $F$ over the entire period, considering absence probabilities $q = 0.05, 0.10, 0.15$, is approximately $F = 33.6, 67.2, 100.8$, respectively. The maximum working days $e$ for each employee is set to 20, the maximum consecutive working days $r$ is set to 4, and the maximum consecutive working days for time slot 3 $s$ is set to 3. Table II provides details for the parameter sets explored under the conditions in Section IV-B. It is important to note that in this experiment, $\beta = 1.5, c = 14$ were set based on the conditions in Section IV-B. For each of the 340

parameter sets listed in Table II, a simulation of substitute attendance requests over the entire period was conducted 300 times, measuring the average daily insufficient personnel and the average daily number of requests from managers to employees.

Subsequently, we calculated the difference in the daily insufficient number of employees and the difference in the number of requests from managers to employees for each parameter set among the three validated request order determination strategies. The difference among strategies refers to the gap between the highest and lowest average values of the daily insufficient number of employees or requests. We will now discuss the top 5, bottom 5, and 5 around the median parameter sets where the difference in the insufficient number of employees or requests among strategies is the most significant.

In particular, for the top 5 and bottom 5, we will analyze the relationship between the daily insufficient number of employees and the number of requests. Based on the trials for all parameter sets, we will also discuss the trend in the evolution of the difference in the insufficient number of employees (requests) among strategies when arranged in descending request strategy, and verify the percentage of parameter sets where a significant difference occurs. Additionally, we will analyze the distribution of each parameter at that time and determine an appropriate degree of the polynomial approximation curve representing the overall trend through cross-validation.

Finally, for the parameter sets where the difference in the insufficient number of employees(requests) among strategies is the largest, the smallest, and around the median, we will measure and compare the average number of times employees attended as substitutes (accepted requests) and the average number of requests received by employees from managers over the entire period.

### D. Result

*1) Insufficient number of employees:* For the top 5 parameter sets with largest average the daily insufficient number of employees among 300 simulations, details are presented in Table III, where $\zeta$ represents the average daily insufficient number of employees. Notable characteristics leading to higher daily insufficient number of employees include a high absence probability $q$ (indicating more absences $F$), a small maximum acceptance count $m$, and relatively small $p_{low}$ with large $p_{high}$. The value of $n_{low}$ is 35 for all sets except the 2nd set.

*2) Difference in insufficient number of employees among strategies:* Table IV presents the top 5 parameter sets(1st-5th), the 5 sets around the median(169th-173rd) with the largest difference in the daily insufficient number of employees among strategies and the 5 parameter sets(336th-340th) with the smallest difference. The symbol $\delta$ represents the difference in the daily insufficient number of employees among strategies.

From Table IV, the trends for the top 5 parameter sets include a significant difference between $p_{low}$ and $p_{high}$, with the number of employees $n_{high}$ having high acceptance level set at 15, except for the 4th set. Additionally, the maximum

TABLE IV: PARAMETER SET WITH THE MOST SIGNIFICANT DIFFERENCE IN THE INSUFFICIENT NUMBER OF EMPLOYEES BETWEEN STRATEGIES, THE LEAST SIGNIFICANT DIFFERENCE, AND AN INTERMEDIATE LEVEL OF DIFFERENCE

| Rank | 1st | 2nd | 3rd | 4th | 5th | 169th | 170th | 171st | 172nd | 173rd | 336th | 337th | 338th | 339th | 340th |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X$ | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 |
| $d$ | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 |
| $q$ | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.10 | 0.15 | 0.15 | 0.15 | 0.10 | 0.05 | 0.10 | 0.15 | 0.10 | 0.15 |
| $m$ | 3 | 4 | 3 | 4 | 3 | 2 | 4 | 8 | 2 | 7 | 6 | 7 | 20 | 11 | 18 |
| $p_{low}$ | 0.15 | 0.10 | 0.10 | 0.10 | 0.20 | 0.15 | 0.10 | 0.10 | 0.10 | 0.20 | 0.15 | 0.25 | 0.20 | 0.25 | 0.15 |
| $p_{high}$ | 0.90 | 0.90 | 0.90 | 0.90 | 0.90 | 0.50 | 0.70 | 0.70 | 0.50 | 0.90 | 0.60 | 0.50 | 0.50 | 0.50 | 0.50 |
| $n_{low}$ | 35 | 35 | 35 | 40 | 35 | 45 | 45 | 40 | 40 | 45 | 45 | 45 | 45 | 45 | 45 |
| $n_{high}$ | 15 | 15 | 15 | 10 | 15 | 5 | 5 | 10 | 10 | 5 | 5 | 5 | 5 | 5 | 5 |
| $\delta$ | 0.490 | 0.474 | 0.465 | 0.427 | 0.419 | 0.117 | 0.115 | 0.1149 | 0.1145 | 0.1145 | 0.014 | 0.012 | 0.0109 | 0.0102 | 0.0101 |

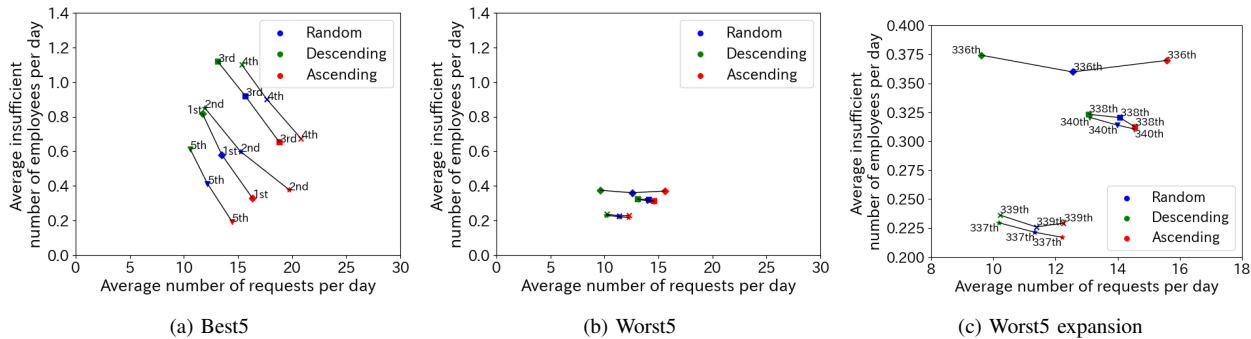

(a) Best5      (b) Worst5      (c) Worst5 expansion

Fig. 1: Relationship between insufficient number of employees and number of requests per day(Table IV).

TABLE V: VARIANCE(TABLE IV)

| | random | des | asc |
|---|---|---|---|
| 1st | 0.041 | 0.050 | 0.032 |
| 171st | 0.020 | 0.0185 | 0.013 |
| 340th | 0.016 | 0.0187 | 0.016 |

TABLE VI: AVERAGE(TABLE IV)

| | random | des | asc |
|---|---|---|---|
| 1st | 0.579 | 0.818 | 0.328 |
| 171st | 0.302 | 0.354 | 0.239 |
| 340th | 0.313 | 0.320 | 0.310 |

TABLE VII: MEDIAN(TABLE IV)

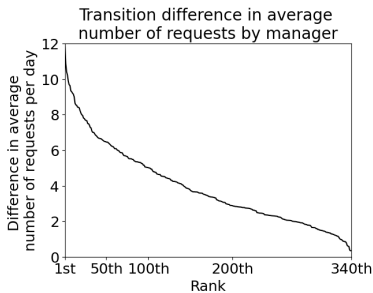| | random | des | asc |
|---|---|---|---|
| 1st | 0.571 | 0.821 | 0.303 |
| 171st | 0.285 | 0.321 | 0.214 |
| 340th | 0.285 | 0.321 | 0.285 |



Fig. 2: Transition of difference in average insufficient number of employees.

acceptance count $m$ is relatively small compared to other parameter sets in the validation.

Conversely, Table IV shows trends for the bottom 5 parameter sets, indicating a smaller difference between $p_{low}$ and $p_{high}$" compared to other sets in the validation. Additionally, the number of employees $n_{low}$ with low acceptance probabilities is consistently 45 for all 5 parameter sets. Regarding the maximum acceptance count $m$," particularly in the 338th, 339th, and 340th sets, larger values are observed compared to other parameter sets in the validation. However, when looking at the values of $\delta$," the difference between the 336th and 340th sets is extremely small, with only 0.0039.

For the 1st set, 171st set, and 340th set (Table IV), the distribution of the daily insufficient number of employees in 300 trials for each strategy (random, descending, ascending request strategy) is presented in Table V for unbiased variance, Table VI for mean, and Table VII for median. In Table V, it can be observed that for all three sets (1st, 171st, 340th), the unbiased variance tends to be smaller for the ascending order strategy. The variance is larger for the 1st set, while for the 171st and 340th sets, the variance values are almost equal.

From Tables VI and VII, for the 1st set, there is little difference between the mean and median values for any strategy. On the other hand, for the 171st and 340th sets, the mean values are slightly higher than the medians, indicating a rightward skewness in the distribution.

From Table IV, the overall trend for the 5 parameter sets around the median indicates an intermediate nature compared to the top 5 and bottom 5. For the 169th and 172nd sets, although the maximum acceptance count $m$ is small, the difference between $p_{low}$ and $p_{high}$ is also small. On the other

TABLE VIII: QUANTILE OF $q$ DISTRIBUTION

| $q$ | Min | $Q_1$ | $Q_2$ | $Q_3$ | Max |
|------|------|-------|-------|-------|------|
| 0.15 | 1 | 58 | 136 | 220 | 340 |
| 0.10 | 23 | 112.5 | 189.5 | 282.5 | 339 |
| 0.05 | 185 | 250 | 275 | 305 | 336 |
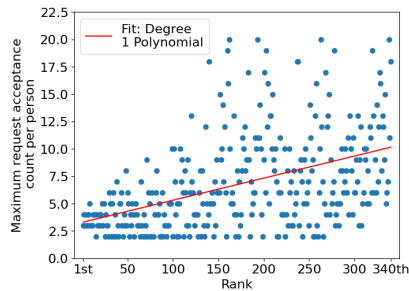


Fig. 3: Plots of $m$ distribution.

hand, for the 171st and 173rd sets, the difference between $p_{low}$ and $p_{high}$ is large, but the maximum acceptance count $m$ is also large. For the 170th set, the value of the maximum acceptance count $m$ is larger than that of the 169th and 172nd sets, and the difference between $p_{low}$ and $p_{high}$ is smaller compared to the 171st and 173rd sets.

Figure 1 illustrates the relationship between the daily insufficient number of employees and the number of requests for the top 5 and bottom 5 parameter sets with the largest difference in the insufficient number of employees. Notably, sets with a significant difference in the insufficient number of employees show almost twice the difference between descending and ascending request strategies. Conversely, sets with minimal difference in the insufficient number of employees, such as the 336th set, exhibit little difference in the number of requests.

Figure 2 depicts the trend in the difference in the daily insufficient number of employees among strategies for all validated parameter sets in descending request strategy. The decrease in the difference is significant up to around the 50th set. Parameter sets with a difference exceeding 0.4 constitute only around 2.0% of the total (1st to 7th), and those exceeding
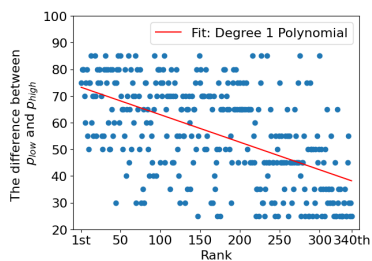


Fig. 4: Plots of distribution of the difference between $p_{low}$ and $p_{high}$.

TABLE IX: QUANTILE OF $n_{high}$ DISTRIBUTION

| $n_{high}$ | Min | $Q_1$ | $Q_2$ | $Q_3$ | Max |
|------------|------|-------|-------|-------|------|
| 15 | 1 | 23 | 47 | 88.75 | 229 |
| 10 | 4 | 60.5 | 110 | 185 | 324 |
| 5 | 56 | 174 | 237 | 293 | 340 |

TABLE X: PARAMETER SETS WITH THE LARGEST REQUEST FREQUENCY

| Rank | 1st | 2nd | 3rd | 4th | 5th |
|------|------|------|------|------|------|
| $X$ | 50 | 50 | 50 | 50 | 50 |
| $d$ | 28 | 28 | 28 | 28 | 28 |
| $q$ | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 |
| $m$ | 3 | 2 | 4 | 3 | 4 |
| $p_{low}$ | 0.05 | 0.05 | 0.10 | 0.10 | 0.05 |
| $p_{high}$ | 0.90 | 0.90 | 0.90 | 0.90 | 0.90 |
| $n_{low}$ | 35 | 35 | 40 | 40 | 35 |
| $n_{high}$ | 15 | 15 | 10 | 10 | 15 |
| $\eta$ | 18.098 | 18.003 | 17.949 | 17.918 | 17.763 |

0.3 account for approximately 8.8% (1st to 30th). Conversely, sets with a difference below 0.1 make up 43.5% of the total (193rd to 340th).

Table VIII displays quartiles, minimum, and maximum values of the absenteeism probability $q$ rank for each parameter set in descending request strategy of the difference $\delta$ in the insufficient number of employees between strategies. The table indicates that the difference in insufficient employees tends to increase with higher absenteeism probability, signifying more absences.

In Figure 3, the plot of the maximum acceptance count $m$ for each parameter set, arranged in descending request strategy of the insufficient number of employees difference between strategies, is presented. The regression line suggests that as the maximum acceptance count $m$ decreases, the difference in the insufficient number of employees between strategies tends to increase.

Additionally, Figure 4 shows the plot of the difference between acceptance probabilities $p_{low}$ and $p_{high}$ for each parameter set, ordered by the difference in the insufficient number of employees between strategies. The regression line illustrates that a larger difference between $p_{low}$ and $p_{high}$ corresponds to a greater difference in the insufficient number of employees between strategies.

Table IX presents quartiles, minimum, and maximum values of the rank of the number of employees $n_{high}$ with high acceptance levels, in descending request strategy of the difference in the insufficient number of employees between strategies. The table highlights that a higher number of employees with high acceptance levels contributes to a larger difference in the insufficient number of employees between strategies.

*3) Number of requests by manager:* Table X displays the top 5 parameter sets with the largest average number of daily requests in 300 simulations, denoted by $\eta$. Characteristics of sets with higher daily request numbers include a high absenteeism probability $q$, indicating a larger number of absences $F$, and a relatively small maximum acceptance count $m$. $p_{low}$ is small, $p_{high}$ is large, and $n_{low}$ is 40 for the 3rd and 4th sets and 35 for the others.

*4) Difference in number of requests among strategies:* Table XI shows the top 5 parameter sets (1st-5th), the 5 sets around the median (169th-173rd) with the largest differences in request numbers between strategies, and the 5 parameter sets (336th-340th) with the smallest difference, denoted as $\epsilon$. Among the top 5 sets, all have the smallest $p_{low}$ value

TABLE XI: PARAMETER SET WITH THE MOST SIGNIFICANT DIFFERENCE IN THE NUMBER OF REQUEST BETWEEN STRATEGIES, THE LEAST SIGNIFICANT DIFFERENCE, AND AN INTERMEDIATE LEVEL OF DIFFERENCE

| Rank | 1st | 2nd | 3rd | 4th | 5th | 169th | 170th | 171st | 172nd | 173rd | 336th | 337th | 338th | 339th | 340th |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X$ | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 |
| $d$ | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 |
| $q$ | 0.15 | 0.15 | 0.10 | 0.10 | 0.10 | 0.10 | 0.15 | 0.05 | 0.15 | 0.10 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 |
| $m$ | 6 | 5 | 6 | 4 | 5 | 4 | 10 | 2 | 3 | 7 | 2 | 2 | 2 | 2 | 2 |
| $p_{low}$ | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.10 | 0.20 | 0.05 | 0.15 | 0.15 | 0.25 | 0.25 | 0.20 | 0.25 | 0.25 |
| $p_{high}$ | 0.90 | 0.90 | 0.90 | 0.70 | 0.90 | 0.90 | 0.50 | 0.50 | 0.70 | 0.70 | 0.90 | 0.70 | 0.90 | 0.90 | 0.50 |
| $n_{low}$ | 35 | 35 | 40 | 35 | 40 | 45 | 40 | 45 | 40 | 45 | 40 | 45 | 45 | 45 | 45 |
| $n_{high}$ | 15 | 15 | 10 | 15 | 10 | 5 | 10 | 5 | 10 | 5 | 10 | 5 | 5 | 5 | 5 |
| $\epsilon$ | 11.917 | 10.872 | 10.415 | 10.220 | 9.878 | 3.461 | 3.447 | 3.421 | 3.387 | 3.384 | 0.635 | 0.594 | 0.583 | 0.373 | 0.340 |

TABLE XII: VARIANCE OF NUMBER OF REQUESTS(TABLE XI)

|  | random | des | asc |
|---|---|---|---|
| 1st | 4.671 | 4.577 | 2.750 |
| 171st | 3.062 | 3.275 | 3.030 |
| 340th | 1.123 | 1.096 | 1.021 |

TABLE XIII: AVERAGE OF NUMBER OF REQUESTS(TABLE XI)

|  | random | des | asc |
|---|---|---|---|
| 1st | 16.590 | 10.713 | 22.410 |
| 171st | 9.424 | 8.428 | 10.816 |
| 340th | 11.498 | 11.313 | 11.624 |

TABLE XIV: MEDIAN OF NUMBER OF REQUESTS(TABLE XI)

|  | random | des | asc |
|---|---|---|---|
| 1st | 16.714 | 10.732 | 22.428 |
| 171st | 9.464 | 8.535 | 10.714 |
| 340th | 11.482 | 11.321 | 11.678 |

(0.05) except for the 4th set. Except for the 4th set, they also exhibit the largest $p_{high}$ values (0.90). The maximum acceptance count $m$ falls within intermediate values for these sets. Conversely, the bottom 5 sets, all with an employee absenteeism probability $q$ of 0.15, show a small difference between $p_{low}$ and $p_{high}$ values. The $n_{high}$ is 5 for all except the 336th set, and the maximum acceptance count $m$ is 2 in all five sets. The 5 sets near the median exhibit intermediate characteristics. For sets 169th, 171st, and 172nd, $m$ is relatively small, but the difference between $p_{low}$ and $p_{high}$ is large. For sets 171st and 173rd, the difference between $p_{low}$ and $p_{high}$ is significant, but $m$ is large. For set 170th, both $m$ and $p_{low}$ are large.

For the 1st, 171st, 340th paremeter sets in Table XI, the distribution of daily average requests for random, descending, and ascending request strategies is presented in Tables XII, XIII, and XIV, respectively. Table XII shows that, similar to daily insufficient employee numbers, the ascending request strategy has the smallest unbiased variance for all three sets (1st, 171st, and 340th). In random and descending request strategies, the 1st set has the largest variance, followed by 171st, while 340th has the smallest variance. Regarding mean values and medians (Tables XIII and XIV), the 1st set shows little difference, but for sets 171st and 340th, mean values are slightly higher than medians.

The relationship between daily insufficient personnel and the number of requests for the top 5 parameter sets with the largest difference and the bottom 5 parameter sets is shown in Figure 5. Notably, for parameter sets with a significant difference in the number of requests, the daily number of requests between descending and ascending request strategies differs approximately twofold. Conversely, for parameter sets with minimal differences in request counts, the occurrence of the daily insufficient number of employees is also limited.

Figure 6 illustrates the transition of the difference in the number of requests between strategies, sorted in descending order for all validated parameter sets. Similar to the trend observed in Figure 2, there is a significant decrease in the range of 1st to around 50th in Figure 6. For instance, parameter sets with a difference in the number of requests between strategies exceeding 10 represent approximately 1.1% of the total (1st to 4th), and even when limited to a difference of 8 or more, it accounts for around 5.8% of the total (1st to 20th). Conversely, parameter sets with a difference of 4 or less constitute 58.5% of the total (142nd to 340th).

Table XV displays quartiles, minimum, and maximum values of absenteeism probability $q$ ranks for each parameter set, sorted by the difference in manager-requested counts between strategies. The trend suggests that, overall, lower absenteeism probabilities are associated with larger differences in request numbers, though exceptions exist. Next, Figure 7 presents the plot of maximum acceptance count $m$ ranks for each parameter set, ordered by the difference in manager-requested counts. The red line indicates a fifth-degree approximation curve. Notably, both very large and very small $m$ values result in smaller differences in request numbers, with the most significant difference occurring at intermediate levels (around 5 to 6). Furthermore, Figure 8 illustrates the plot of the difference between acceptance probabilities $p_{low}$ and $p_{high}$ for each parameter set, ordered by the difference in manager-requested counts. The red line represents the regression line, showing that a larger difference between $p_{low}$ and $p_{high}$ corresponds to a larger difference in request numbers between strategies. Table XVI presents the quartiles, minimum, and maximum values of the ranks of the number of employees with high acceptance levels $n_{high}$ for each parameter set, ordered by the magnitude of the difference in the number of requests made by managers between strategies. It can be observed that a larger number of employees with high acceptance probabilities corresponds to a larger difference in the number of requests between strategies.

*5) Substitute attendance count of employees and the number of received requests:* Tables XVII, XVIII, and XIX show the

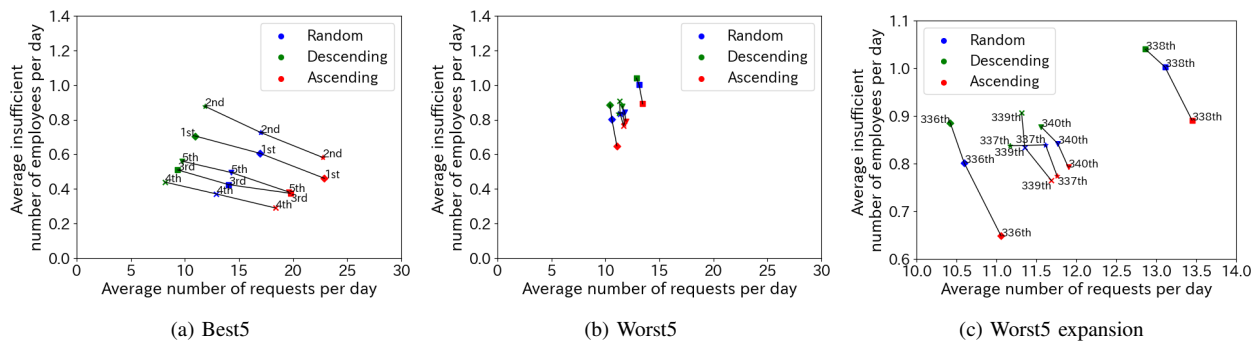(a) Best5      (b) Worst5      (c) Worst5 expansion

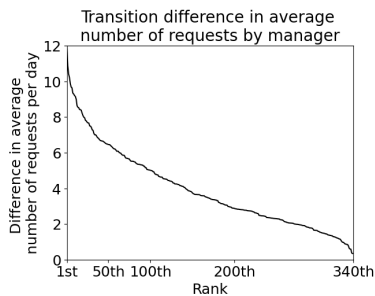Fig. 5: Relationship between insufficient number of employees and number of requests per day (Table XI).



Fig. 6: Transition of difference in average number of requests by manager.

TABLE XV: QUANTILE OF $q$ DISTRIBUTION

| $q$ | Min | $Q_1$ | $Q_2$ | $Q_3$ | Max |
|------|-----|-------|-------|-------|-----|
| 0.15 | 1 | 108 | 199 | 275 | 340 |
| 0.10 | 3 | 90.75 | 160.0 | 222.75 | 334 |
| 0.05 | 31 | 53 | 79 | 111 | 171 |



Fig. 7: Plots of $m$ distribution.



Fig. 8: Plots of distribution of the difference between $p_{low}$ and $p_{high}$.

average number of employees based on substitute attendances over 300 trials for the 1st, 171st, and 340th parameter sets (Tables IV). In Table XVII, employees with 3 substitute attendances are the most numerous regardless of the request strategy (random, descending, or ascending). Conversely, Tables XVIII and XIX indicate that the descending request strategy yields the largest proportion of employees with more substitute attendances, while the ascending strategy shows the smallest proportion.

Tables XX, XXI, and XXII display the average number of requests received by each employee from managers (regardless of acceptance) over 300 trials for the 1st, 171st, and 340th parameter sets (Tables IV). Across all three sets, the ascending request strategy results in a larger proportion of employees receiving more requests. In contrast, the descending strategy leads to more employees receiving fewer requests, with the random strategy showing an intermediate trend.

Tables XXIII, XXIV, and XXV show average employee counts for substitute attendances in the 1st, 171st, and 340th parameter sets from Table XI over 300 trials. In the 1st set, with a maximum acceptance count of 6, substitute attendances range from 0 to 6 per employee, while in the 171st and 340th sets, this range is 0 to 2. Table XXIII highlights that the de-
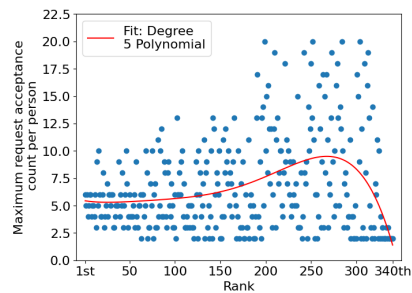
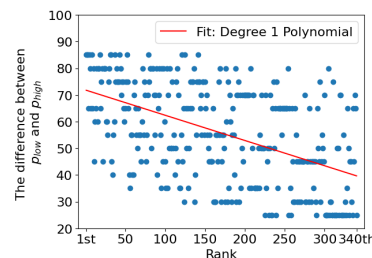scending request strategy leads to more substitute attendances for employees in the 1st set. However, the difference between strategies is less pronounced in Tables XXIV and XXV, likely due to smaller maximum acceptance counts in these cases.

Tables XXVI, XXVII, and XXVIII present the average number of requests received by each employee from managers (regardless of acceptance) during 300 trials for the 1st, 171st, and 340th parameter sets (Tables XI). In all three parameter sets, it is evident that, when the ascending request strategy is employed, the proportion of employees receiving a higher number of requests from managers is larger.

### E. Consideration

Table III highlights that insufficient employee numbers are more common with high absenteeism probability ($q$) and a low maximum acceptance count ($m$). This is likely due to the challenge of securing replacement workers for frequently absent employees when flexibility is limited. Furthermore, a significant gap between acceptance probabilities ($p_{low}$ and

TABLE XVI: QUANTILE OF $n_{high}$ DISTRIBUTION

| $n_{high}$ | Min | $Q_1$ | $Q_2$ | $Q_3$ | Max |
|---|---|---|---|---|---|
| 15 | 1 | 25.25 | 67.5 | 137.25 | 308 |
| 10 | 3 | 61 | 123 | 181.5 | 336 |
| 5 | 31 | 156 | 227 | 285 | 340 |

TABLE XVII: NUMBER OF EMPLOYEES BY FREQUENCY OF SUBSTITUTE ATTENDANCE DURING PERIOD(TABLE IV, 1ST)

| times | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| random | 10.47 | 13.03 | 8.83 | 17.66 |
| des | 12.76 | 13.41 | 7.31 | 16.50 |
| asc | 7.51 | 11.98 | 12.27 | 18.23 |

$p_{high}$), coupled with a substantial number of employees with high acceptance probabilities ($n_{high}$), leads to noticeable shortages in employee availability. This suggests that in scenarios with frequent absences, restricted employee availability for substitute work (due to workplace rules or other factors), a relatively high acceptance rate (around 30Conversely, observing opposite trends for each parameter may help mitigate the occurrence of a shortage of employees. In workplaces with these characteristics, Figure 1 highlights a significant employee shortage difference between descending and ascending request strategies. Increasing requests in the ascending strategy notably reduces employee shortages. Prioritizing shortage reduction favors the ascending strategy, while opting for the descending strategy is advisable for reducing managerial burden. However, Tables XX through XXII show that the ascending request strategy often results in more requests received by employees. In situations where excessive requests may impact employee motivation negatively, adopting this strategy may not be ideal. Conversely, Tables XVII through XIX suggest that the descending request strategy tends to lead to more substitute work occurrences for employees, potentially indicating better motivation retention.

In workplaces with infrequent absences and around 10However, Tables XVII, XVIII, XIX, XXIII, XXIV, and XXV indicate that the descending request strategy slightly increases the proportion of employees with more substitute work occurrences. This may be desirable, reflecting sustained employee motivation despite frequent substitute work. In cases with no significant difference in request counts between strategies (e.g., 337th to 340th sets), there's little distinction in the shortage of employees and the manager's request count. Therefore,

TABLE XVIII: NUMBER OF EMPLOYEES BY FREQUENCY OF SUBSTITUTE ATTENDANCE DURING PERIOD(TABLE IV, 171ST)

| times | 0-2 | 3-4 | 5-6 | 7-8 |
|---|---|---|---|---|
| random | 35.87 | 10.37 | 3.12 | 0.62 |
| des | 38.03 | 5.04 | 3.99 | 2.93 |
| asc | 34.45 | 13.66 | 1.79 | 0.08 |

TABLE XIX: NUMBER OF EMPLOYEES BY FREQUENCY OF SUBSTITUTE ATTENDANCE DURING PERIOD(TABLE IV, 340TH)

| times | 0-5 | 6-9 | 10-12 | 13-18 |
|---|---|---|---|---|
| random | 49.35 | 0.64 | 0.0 | 0.0 |
| des | 47.83 | 2.09 | 0.08 | 0.0 |
| asc | 49.49 | 0.50 | 0.0 | 0.0 |

TABLE XX: NUMBER OF EMPLOYEES BY FREQUENCY OF REQUESTS RECEIVED DURING PERIOD(TABLE IV, 1ST)

| times | 0-5 | 6-10 | 11-15 | 16-20 | 21-25 | 26-30 |
|---|---|---|---|---|---|---|
| random | 19.87 | 18.88 | 9.60 | 1.53 | 0.09 | 0.003 |
| des | 23.85 | 19.49 | 6.02 | 0.59 | 0.03 | 0.00 |
| asc | 17.69 | 12.12 | 13.52 | 5.77 | 0.84 | 0.04 |

TABLE XXI: NUMBER OF EMPLOYEES BY FREQUENCY OF REQUESTS RECEIVED DURING PERIOD(TABLE IV, 171ST)

| times | 0-4 | 5-9 | 10-14 | 15-19 | 20-24 | 25-27 |
|---|---|---|---|---|---|---|
| random | 9.81 | 25.01 | 13.01 | 2.04 | 0.10 | 0.00 |
| des | 17.97 | 22.78 | 7.12 | 1.99 | 0.12 | 0.00 |
| asc | 10.78 | 16.24 | 16.27 | 5.93 | 0.74 | 0.01 |

TABLE XXII: NUMBER OF EMPLOYEES BY FREQUENCY OF REQUESTS RECEIVED DURING PERIOD(TABLE IV, 340TH)

| times | 0-5 | 6-11 | 12-17 | 18-23 | 24-29 | 30-33 |
|---|---|---|---|---|---|---|
| random | 14.62 | 27.83 | 7.20 | 0.33 | 0.003 | 0.00 |
| des | 18.37 | 25.51 | 4.90 | 1.06 | 0.13 | 0.006 |
| asc | 13.03 | 26.38 | 9.16 | 0.52 | 0.00 | 0.00 |

TABLE XXIII: NUMBER OF EMPLOYEES BY FREQUENCY OF SUBSTITUTE ATTENDANCE DURING PERIOD(TABLE XI, 1ST)

| times | 0 | 1-2 | 3-4 | 5-6 |
|---|---|---|---|---|
| random | 22.23 | 14.40 | 4.51 | 8.84 |
| des | 27.62 | 9.05 | 3.15 | 10.16 |
| asc | 17.14 | 19.35 | 6.57 | 6.93 |

TABLE XXIV: NUMBER OF EMPLOYEES BY FREQUENCY OF SUBSTITUTE ATTENDANCE DURING PERIOD(TABLE IV, 171ST)

| times | 0 | 1 | 2 |
|---|---|---|---|
| random | 34.68 | 10.48 | 4.82 |
| des | 35.77 | 9.13 | 5.09 |
| asc | 33.81 | 11.73 | 4.44 |

TABLE XXV: NUMBER OF EMPLOYEES BY FREQUENCY OF SUBSTITUTE ATTENDANCE DURING PERIOD(TABLE IV, 340TH)

| times | 0 | 1 | 2 |
|---|---|---|---|
| random | 6.03 | 10.92 | 33.05 |
| des | 6.15 | 11.07 | 32.77 |
| asc | 5.70 | 10.56 | 33.73 |

TABLE XXVI: NUMBER OF EMPLOYEES BY FREQUENCY OF REQUESTS RECEIVED DURING PERIOD(TABLE IV, 1ST)

| times | 0-4 | 5-9 | 10-14 | 15-19 | 20-24 | 25-28 |
|---|---|---|---|---|---|---|
| random | 5.93 | 21.52 | 16.97 | 4.76 | 0.76 | 0.03 |
| des | 16.96 | 26.07 | 6.73 | 0.22 | 0.006 | 0.0 |
| asc | 5.72 | 12.95 | 11.62 | 11.56 | 6.38 | 1.47 |

TABLE XXVII: NUMBER OF EMPLOYEES BY FREQUENCY OF REQUESTS RECEIVED DURING PERIOD(TABLE IV, 171ST)

| times | 0-3 | 4-6 | 7-9 | 10-12 | 13-15 | 16-20 |
|---|---|---|---|---|---|---|
| random | 14.37 | 20.17 | 11.67 | 3.28 | 0.45 | 0.04 |
| des | 17.25 | 21.19 | 9.41 | 1.95 | 0.18 | 0.00 |
| asc | 10.94 | 18.10 | 14.08 | 5.60 | 1.10 | 0.15 |

TABLE XXVIII: NUMBER OF EMPLOYEES BY FREQUENCY OF REQUESTS RECEIVED DURING PERIOD(TABLE IV, 340TH)

| times | 0-3 | 4-7 | 8-11 | 12-15 | 16-19 | 20-26 |
|---|---|---|---|---|---|---|
| random | 9.89 | 23.48 | 12.48 | 3.51 | 0.58 | 0.04 |
| des | 10.33 | 23.68 | 12.17 | 3.30 | 0.38 | 0.11 |
| asc | 10.89 | 21.88 | 12.31 | 4.16 | 0.70 | 0.04 |

choosing the ascending request strategy can address imbalances in substitute work occurrences among employees, while the descending strategy is preferable for reducing the number of requests received by employees. Table X highlights that higher manager-to-employee request numbers coincide with elevated absenteeism probability ($q$), necessitating more substitute attendance. Even in workplaces with low absenteeism probability, notable differences in request numbers persist. Figure 5 indicates significant contrasts between descending and ascending request strategies, with the former recommended for reducing managerial burden and the latter for addressing insufficient employees. However, the effectiveness of the ascending strategy may be limited in certain scenarios (e.g., cases 1 to 5 in Figure 5). Further analysis is needed to assess the correlation between request differences and insufficient employee numbers. Additionally, caution is advised regarding the potential impact on employee motivation, particularly with the descending request strategy favoring individuals with more substitute attendances. In such cases, the random request strategy may offer a preferable alternative, aligning closely with trends observed in both requests and insufficient numbers compared to the descending strategy. Conversely, when the maximum acceptance count ($m$) is either too small or too large, and there is minimal difference between acceptance probabilities ($p_{low}$ and $p_{high}$), with few employees having high acceptance probabilities ($n_{high}$), there is little to no notable difference in the number of requests from managers across strategies. In such cases, Figure 5 suggests adopting the ascending request strategy, which prioritizes reducing the insufficient number of employees. However, it is important to consider that the ascending request strategy may lead to an increase in the number of requests accepted by employees. Thus, in situations where the rise in requests may not significantly impact employee motivation, opting for the ascending request strategy to mitigate the insufficient number of employees is advisable.

## V. Conclusion

This paper presents a simulation model for substitute attendance requests, exploring three request order strategies: random, descending, and ascending. Our findings indicate that in workplaces with high absenteeism probability and specific characteristics, the ascending strategy effectively reduces employee shortages. Conversely, the descending strategy is effective for reducing managerial burden in workplaces with the opposite trend. In scenarios influenced by various factors, the descending or random strategies prove effective for reducing managerial burden, while the ascending strategy is suitable for mitigating employee shortages in workplaces with opposing trends.

Future research should focus on developing substitute shift request strategies based on easily observable managerial information, including past substitute shift counts and potential future availability, alongside employee acceptance probabilities.

## References

[1] A. Ikegami and A. Niwa, "A subproblem-centric model and approach to the nurse scheduling problem," vol. 97, pp. 517–541, 2003.

[2] K. Nonobe and T. Ibaraki, "A tabu search approach for the constraint satisfaction problem as a general problem solver," *Eur. J. of Oper. Res.*, vol. 106, pp. 599–623, 1998.

[3] U. Aickelin and K. Dowsland, "An indirect genetic algorithm for nurse scheduling problem," *J. of Oper. Res. Society*, vol. 31, pp. 761–778, 2003.

[4] M. Kitada and K. Morizawa, "Heuristic method in dynamic nurse scheduling following a sudden absence of nurses," vol. 65, no. 1, pp. 29–38, 2014.

[5] A. Clark, P. Moule, A. Topping, and M. Serpell, "Rescheduling nursing shifts: Scoping the challenge and examining the potential of mathematical model based tools," *J. of Nursing Management*, 2013.

[6] B. Maenhout and M. Vanhoucke, "An evolutionary approach for the nurse rerostering problem," *Computers Oper. Res.*, vol. 38, pp. 1400–1411, 2011.

[7] M. Moz and M. V. Pato, "Solving the problem of rerostering nurse schedules with hard constraints: New multicommodity flow model," *Annals of Operations Research*, vol. 128, pp. 179–197, 2004.

[8] K. Hatamoto, S. Yokoyama, T. Yamashita, and H. Kawamura, "Development of efficient request method using messaging app for substitute fulfillment," vol. 60, no. 10, pp. 1757–1768, 2019.

[9] IBM, "IBM ILOG CPLEX 20.1.0 User's Manual for CPLEX," 2021. [Online]. Available: https://www.ibm.com/docs/en/icos/20.1.0?topic=cplex-users-manual

# Enhancing Light Field Video Compression Efficiency via View Selection and Synthesis Techniques

Tala Bazzaza, Maissan Bazazeh, Morteza Rashidinia, Weige Qian, Hamid Reza Tohidypour, and Panos Nasiopoulos

Dept. of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada

e-mail: {tbazzaza, mbazazeh, mr1997, wgqian, htohidyp, panos}@ece.ubc.ca

*Abstract*— **Light Field (LF) imaging represents a transformative technology evolving to replicate human-like visual data and emulate our visual environment. Departing from traditional single-viewpoint cameras, light field cameras capture scenes from multiple perspectives, preserving realistic parallax and capturing the direction of light rays. Despite its potential, the substantial increase in data capture underscores the need for efficient compression techniques. This paper investigates the possibility of enhancing LF data compression by strategically omitting certain views during transmission, and then recreating them at the receiver's end through a specialized synthesis method tailored for light fields. The performance evaluation reveals a delicate balance between bandwidth efficiency and image reconstruction quality in the LF compression and transmission. We believe that a more efficient view synthesis approach that capitalizes on all directional light field views, holds the promise of enhancing LF compression performance.**

*Keywords— Light field video compression, prediction structures, view synthesis.*

## I. INTRODUCTION

Light Field (LF) imaging, also referred to as plenoptic imaging, stands as a transformative technology in constant evolution, aiming to replicate human-like visual data and faithfully emulate our visual environment [1]. Diverging from traditional cameras, which capture scenes from a single viewpoint, light field cameras capture light from multiple perspectives, preserving realistic vertical and horizontal parallax. This not only records light intensity but also captures the direction of light rays. The richness of data obtained facilitates post-scene adjustments such as depth of field, focal point, and resolution. Moreover, the inclusion of depth and distance data enhances capabilities in segmentation and object detection. Light field technology finds diverse applications in cinematography, augmented/virtual reality, and medical contexts.

Despite its potential, the significant increase in data capture underscores the crucial need for efficient compression techniques. Traditional compression standards are inadequate for handling the unique characteristics of light field data. Therefore, the development of an effective encoding method is pivotal for managing this vast amount of data, enabling the technology to thrive and unlocking new market opportunities.

State-of-the-art LF compression methods focus on organizing keyframes (I and P frames) and leveraging horizontal and vertical similarities within the hierarchical bi-directional (B) frames. Khoury et al. [2] innovatively positioned the I-frame at the center and expanded the structure by placing the P-frames at the furthest cells horizontally and vertically, achieving a 38% bitrate reduction compared to LF-MVC. Mehajabin's et al. [3] approach utilizes a Structural Similarity Index Measure (SSIM) based selection strategy, determining correlations among views being predicted and their references, and accordingly choosing different types of frames. This method, an extension of Multiview-HEVC, demonstrates a 17% improvement in compression over [2], establishing it as the current state-of-the-art for LF video compression.

In this paper, we explore the potential for achieving greater compression efficiency with light field data by selectively omitting certain views during transmission and subsequently synthesizing them at the receiver end using a synthesis method tailored for LF [4].

The rest of this paper is organized as follows. In Section II, we describe our proposed approach. Section III presents the
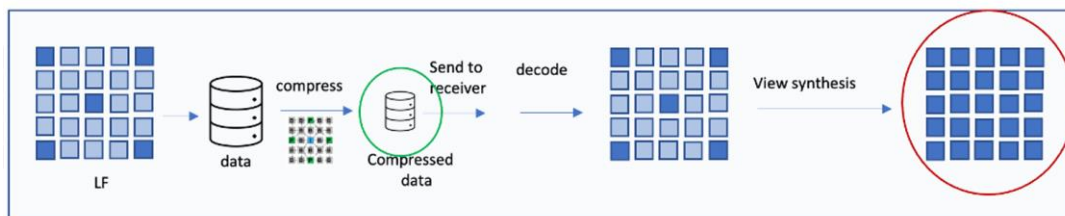


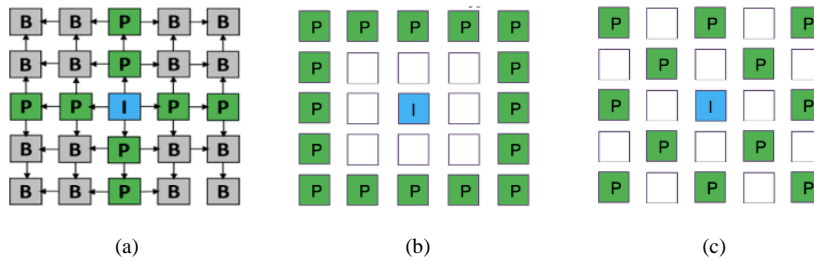Figure 1. Overview of our proposed workflow.

Figure 2.  LF video structures, (a) all original views, (b) peripheral, (c) raster skip.

performance evaluation of our method and discusses the results. Finally, Section IV concludes our paper.

## II. OUR PROPOSED METHOD

Our objective is to investigate the potential for achieving greater compression efficiency with light field data by omitting certain views during transmission and then synthesizing them at the receiver end using a synthesis method tailored for LF. To this end, our first task is to compress all the views of the original video sequence as well as two other view structures, which have some of the views omitted. Then, we employ a view synthesis approach designed for LF to reconstruct the missing views at the receiver end. Figure 1 shows an overview of the proposed workflow. The following subsections describe in detail our end-to-end approach.

### A. Compression

In the first phase of our approach, we consider compressing three different view arrangements, namely the original video sequence that includes all 25 views and two other arrangements shown in Figure 2, where some of the views have been intentionally omitted. We name these two arrangements (structures) peripheral (Figure 2b) and raster skip (Figure 2c).

In our implementation, we choose to use the state-of-the-art LF compression method proposed by Mehajabin et al., also known as Universal Pseudocode Structure (UPS) [3]. This compression method uses a SSIM based selection strategy to determine the correlation among the views being predicted and their references and choose accordingly the different type of frames. Based on that, it utilizes a hierarchical B-frame prediction structure, which leverages both horizontal and vertical correlation to encode the different views/frames (see Figure 2a). An important advantage of this approach, beyond its exceptional compression efficiency [3], is that it is scalable to different view arrangement as well as view numbers and has both encoding and decoding parallelisms. This is very important for our approach, as we can consider any arrangement of horizontal and vertical views based on which ones we want to transmit.

### B. View Synthesis

The missing LF views which were dropped at the transmitting end, are synthesized using the Wafa's et al.'s method [4], a state-of-the-art view synthesis approach that is based on a GAN learning-based model that is trained using the spatial and angular information of light field video content.
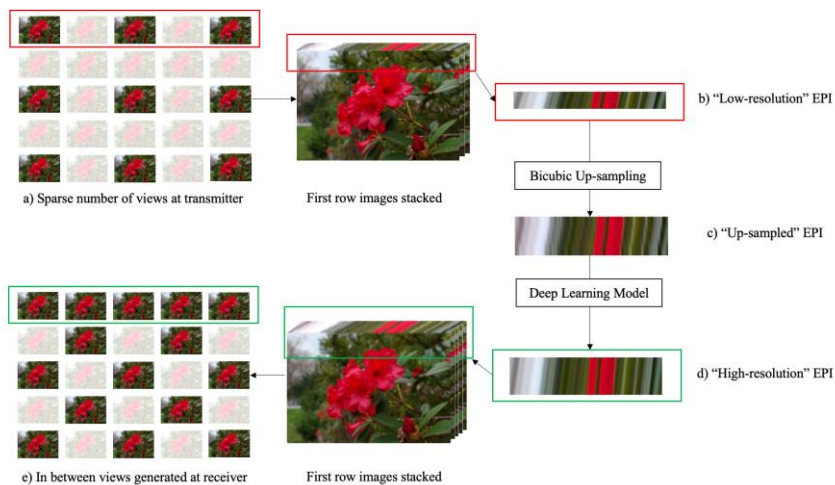


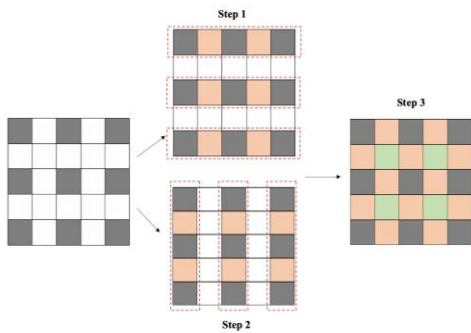Figure 3.  Framework of the LF view sythesis approach used in our implementation.

Figure 4. The three steps for synthesizing in between views.

This view synthesis approach is based on a deep learning model that is trained using the Epipolar Plane Images (EPI) of light field content. The overall framework is shown in Figure 3. The transparent views are the ones dropped at the transmitting end, leaving only the other views to be transmitted or stored. The EPI of rows and columns are fed to a modified Deep Recursive Residual Network (DRRN), which produces the synthesized views at the receiver end.

As seen in Figure 3 (a), some of the views are omitted at the transmitting end. The frames of the first row are then stacked to obtain the "low-resolution" EPI (Figure 3 (b)). Next, the "up-sampled" EPI is obtained by using bicubic interpolation, as seen in Figure 3 (c), to obtain an EPI of the same size as the original number of views. The upscaled EPI is used by the trained model to attain the "full size" EPI, shown in Figure 3 (d). This EPI is finally used to generate the in between views at the receiver end (Figure 3 (e)).

In general, the reconstruction process of synthesizing the in between views for the 4D light field, involves the three steps that are shown in Figure 4. In the first step, a horizontal 2D EPI is generated for each row of the inputs, as explained in Figure 3. This step is repeated for every row of views in the scene. In the second step, the vertical columns of views are used to generate the up-sampled vertical 2D EPIs and then these are used to reconstruct the intermediate views in the LF columns. In the third step, the synthesized views from either the horizontal or the vertical EPIs are utilized to generate the remaining views, represented by green boxes in Figure 4 (step 3).

## III. PERFORMANCE EVALUATION AND DISCUSSION

We evaluated our approach on publicly available microlens based light field videos captured by the Raytrix LF camera [5]. The video sequences are 30fps with 2K resolution and duration of 10 seconds. We examined various compression quality levels by setting QP values to 25, 30, 35 and 40.

Figure 5 provides a comparative analysis of bitrate versus average Peak Signal-to-Noise Ratio (PSNR) for light field video compression structures Raster skip, Peripheral, and the Universal Prediction Structure applied to all the original views, which is the standard for comparison in this context. Note that here we only consider the bitrate and PSNR of the views transmitted, meaning that the raster skip and peripheral
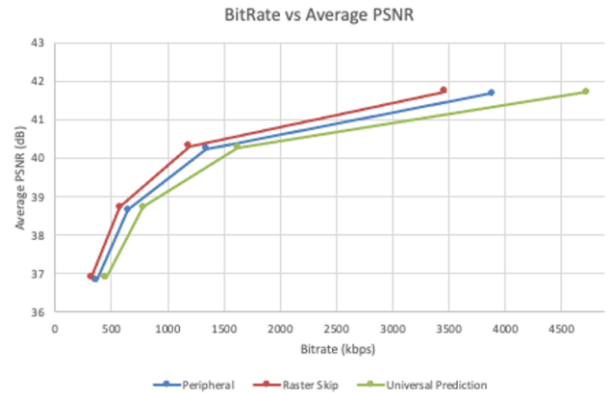


Figure 5. Performance comparison of bitrate vs. average PSNR of the three view structures, all views (UPS), raster skip and peripheral, only for the transmitted views.

have fewer views. We observe that although all the methods seem to achieve similar compression results, there are still distinct differences in their performance.

The raster skip strategy outperforms the others, including the peripheral method, which is slightly more aligned with compressing all the available views (UPS). The superiority of raster skip can be rationalized by the fact that it omits the most views, therefore reducing the amount of data that needs to be compressed. It also means that Mehajabin's compression approach is very efficient, performing extremely well for the case that views are dropped, not really being affected from the fact that the remaining views are farther apart than the case that all views are available for compression (UPS).

Regarding view synthesis, we observe that when we consider a 3x3 window within the raster skip and peripheral structures, both share the same four corners as reference points (see Figure 6). It is worth noting that the raster skip structure has fewer views that need to be synthesized compared to the peripheral, which includes two adjacent views.

For a visual evaluation of the view synthesis approach, Figure 7a shows the original view and 7b represents the synthesized view for the raster skip structure. As it can be seen, the synthesized view looks almost identical to the original, careful observation of the background text and numbers indicates some visual artifacts such as reduction in text boldness.

Figure 8 depicts the performance comparison of bitrate vs. average PSNR of the three view structures, all views (UPS), raster skip and peripheral, including the transmitted and synthesized views for the raster skip and peripheral structures.
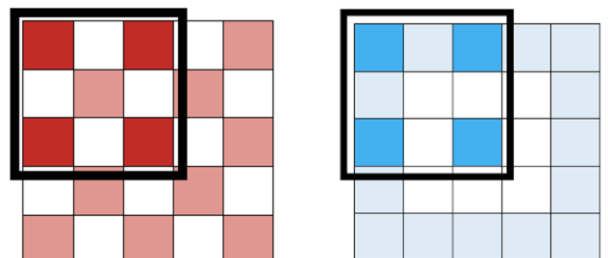


Figure 6. Window for the raster skip and peripheral structures.

(a)



(b)

Figure 7. (a) Original, (b) Synthesized (raster skip).

We observe that although at low bitrates the raster skip performs better, at medium and high bitrates compressing and transmitting all the views shows better efficiency. We could conclude that despite the advancements shown in our synthesized outputs, there remains a gap when compared to the results obtained from configurations utilizing all available views. Our synthesized version, while impressive, understandably falls short of this benchmark due to the fact that it does not take simultaneously take advantage of horizontal, diagonal and vertical views. Our future work focuses on improving the existing view synthesis approach to address this shortcoming and we feel very optimistic that this improvement will yield the desirable bandwidth savings.

## IV. CONCLUSIONS

In this paper, we explore the potential for improving compression efficiency of light field video content by selectively omitting certain views during transmission and subsequently synthesizing them at the receiver end. The main reason for our quest is that although there have been notable advancements in compression algorithms for light field video content, there is still a need for additional enhancements in compression. These improvements are essential to fulfill the bandwidth requirements for the practical application of light field technology.
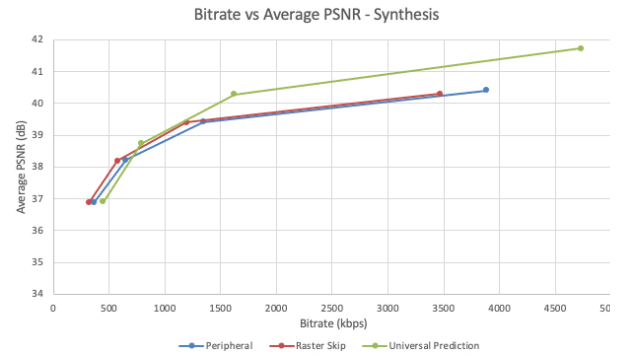


Figure 8. Performance comparison of bitrate vs. average PSNR of the three view structures, all views (UPS), raster skip and peripheral, including the transmitted and synthesized views for the raster skip and peripheral structures.

Our findings highlight the delicate balance between bandwidth efficiency and reconstruction quality in light field compression and transmission. While our current attempts did not yield the desired outcomes, we believe that a more efficient view synthesis approach, capitalizing on all directional light field views, holds the promise of enhancing compression performance. This will be the primary focus of our future work in this domain.

## REFERENCES

[1] I. Ihrke, J. Restrepo, and L. Mignard-Debise, Principles of Light Field Imaging: Briefly revisiting 25 years of research. In IEEE Signal Processing Magazine, vol. 33, no. 5, pp. 59 – 69, September 2016.

[2] J. Khoury, N. Mehajabin, M. T. Pourazad, P. Nasiopoulos, and V. C. M. Leung, An efficient three-dimensional prediction structure for coding light field video content using the MV-HEVC standard. In International Journal of Multimedia Intelligence and Security, vol. 4, no. 1, pp. 47-64, 2022.

[3] N. Mehajabin, M. T. Pourazad and P. Nasiopoulos, An Efficient Pseudo-Sequence-Based Light Field Video Coding Utilizing View Similarities for Prediction Structure. In IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 4, pp. 2356-2370, April 2022, doi: 10.1109/TCSVT.2021.3092282.

[4] A. Wafa and P. Nasiopoulos, Light Field GAN-based View Synthesis using full 4D information. ACM SIGGRAPH European Conference on Visual Media Production (CVMP), London, UK, December 2022.

[5] L. Guillo, X. Jiang, G. Lafruit, and C. Guillemot, Light field video dataset captured by a R8 Raytrix camera (with disparity maps), International Organisation for Standardisation iso/iec jtc1/sc29/wg1 & wg11, phd diss., 2018.