International Journal on Advances in Systems and Measurements

2023 vol. 16 nr. 1&2

The *International Journal on Advances in Systems and Measurements* is published by IARIA.
ISSN: 1942-261x
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Systems and Measurements
Volume 16, Number 1 & 2, 2023

International Journal on Advances in Systems and Measurements
Volume 16, Numbers 1 & 2, 2023

CONTENTS

pages: 53 - 65

Mapping of a Low-Textured Environment Using Visual Simultaneous Localization and Mapping to Use Augmented Reality Simulation for Testing Advanced Driver Assistance Systems in Future Automotive Vehicles

Michael Weber, Institute of Energy Efficient Mobility Hochschule Karlsruhe - University of Applied Sciences, HKA, Germany
Tobias Weiss, Institute of Energy Efficient Mobility Hochschule Karlsruhe - University of Applied Sciences, HKA, Germany
Franck Gechter, CIAD (UMR 7533) Univ. Bourgogne Franche-Comte, UTBM; LORIA-MOSEL (UMR 7503) Universite de Lorraine, France
Reiner Kriesten, Institute of Energy Efficient Mobility Hochschule Karlsruhe - University of Applied Sciences, HKA, Germany

pages: 66 - 79

Modelling Player Combat Behaviour for Dynamic Difficulty Scaling and Combat Perception Analysis in First Person Shooter Games

Paul Williamson, University of South Wales, United Kingdom
Christopher Tubb, University of South Wales, United Kingdom

pages: 80 - 90

Simulation of the Emergency Department Management during the Pandemic

Ramona Galeano, Universidad Autónoma de Barcelona, España
Dolores Rexachs, Universidad Autónoma de Barcelona, España
Alvaro Wong, Universidad Autónoma de Barcelona, España
Eva Bruballa, Escuela Universitaria Gimbernat, España
Cynthia Villalba, Universidad Nacional de Asunción, Paraguay
Diego Galeano, Universidad Nacional de Asunción, Paraguay
Emilio Luque, Universidad Autónoma de Barcelona, España

A Multi-modal AI Approach for Intuitively Instructable Autonomous Systems

Ferran Gebellí Guinjoan[1], Erwin Rademakers, Anil Kumar Chavali, Abdellatif Bey Temsamani
Flanders Make
Lommel, Belgium
[1] email: ferran.gebelli@flandersmake.be
Gorjan Radevski[3], Tinne Tuytelaars
KU Leuven, ESAT
Leuven, Belgium
[3] email: gorjan.radevski@esat.kuleuven.be

Matthias Hutsebaut-Buysse[2], Kevin Mets, Tom De Schepper, Steven Latré, Erik Mannens
University Of Antwerp - imec
Antwerpen, Belgium
[2] email: matthias.hutsebaut-buysse@uantwerpen.be
Hugo Van hamme[4]
KU Leuven, ESAT
Leuven, Belgium
[4] email: hugo.vanhamme@esat.kuleuven.be

***Abstract*— We present a multi-modal AI framework to intuitively instruct and control Automated Guided Vehicles. We define a general multi-modal AI architecture, which has a loose coupling between three different AI modules, including spoken language understanding, visual perception and Reinforcement Learning navigation. We use the same multi-modal architecture for two different use cases implemented in two different platforms: an off-road vehicle, which can pick objects, and an indoor forklift that performs automated warehouse inventory. We show how the proposed architecture can be used for a wide range of tasks and can be implemented in different hardware, demonstrating a high degree of modularity.**

Keywords - AI based autonomous systems; Multi-modal AI; Natural language processing; deep learning; neural networks; reinforcement learning

I. Introduction

Autonomous Guided Vehicles (AGVs), which are often also referred as Autonomous Mobile Robots (AMRs), are becoming more and more popular in industrial applications. In previous works [1] [2] we presented two particular use cases where multi-modal AI leverages AGV tasks. In this paper, we propose a multi-modal AI framework that allows to intuitively and easily (re-)configure an AGV to perform different and variable tasks. The proposed multi-modal software architecture has a loose coupling between the different modules, which allows to easily exchange the components and deploy them in different hardware units.

AGVs can pick up and deliver materials around a manufacturing facility or warehouse [3]. However, with the continuously increase of mass customization [4], a return on investment of production AGVs can only be obtained if these AGVs can easily perform large variability of tasks and / or deal with large variability of products.

Task scheduling has been done by a central entity for a fleet of AGVs following predefined configurations. But driven by flexibility, robustness and scalability requirements, the current trends in AGV systems are customization and decentralization [5]. In a decentralized architecture, an AGV broadcasts the information about its states in a local way and decides which actions to take [6].

Although new generations of AGVs are highly instrumented with different sensors, they are more suited for long-distance transportation of materials between multiple destinations, and tuned for repetitive and predictable tasks [7].

(Re-)configurating AGVs to perform multiple tasks in a non-predictable environment remains, however, a challenge today in industrial settings due to dynamically changing environments. Classic navigation pipelines typically need to construct a map by scanning the environment with sensors, such as lidars [8], while manually driving the AGV. Sometimes the usage of floor markings or fiducial landmarks (e.g., reflectors) are used as well. These approaches do not only require an updated map, but also require a different module to set destinations or missions with waypoints, meaning that a high set-up time for new or modified environments is needed. Because of the increasing variability in industry settings, it is common that the environment is modified after short periods of time. This exposes the need for an increased flexibility in the whole navigation approach.

Research on a voice controlled AGV remains in the level of performing basic operations (e.g., moving with constant speed) in a prescribed path [9].

In this work we show how a general multi-modal architecture can be applied on two different use cases, which run on two different platforms (Figure 1). On the one hand, we implement an application on an off-road vehicle where the main task is to pick certain objects. On the other hand, we deploy an automated inventory monitoring on a forklift. In both cases, an operator can intuitively instruct the AGV by speech interaction that can be done locally or remotely.

In Section II, the common multi-modal AI architecture is presented. Section III explains the off-road vehicle use case, while Section IV describes the forklift use case. Finally, Section V contains the conclusions.

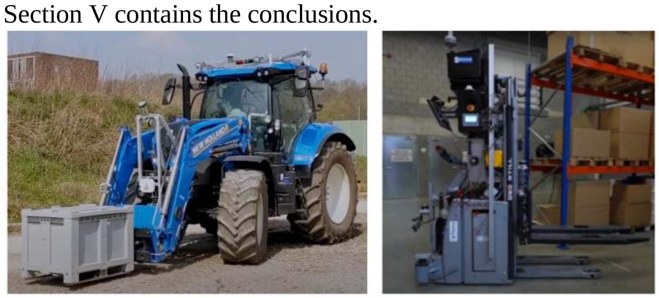


Figure 1. Platforms used for off-road vehicle picking objects (left) and forklift warehouse automated inventory (right) use cases.

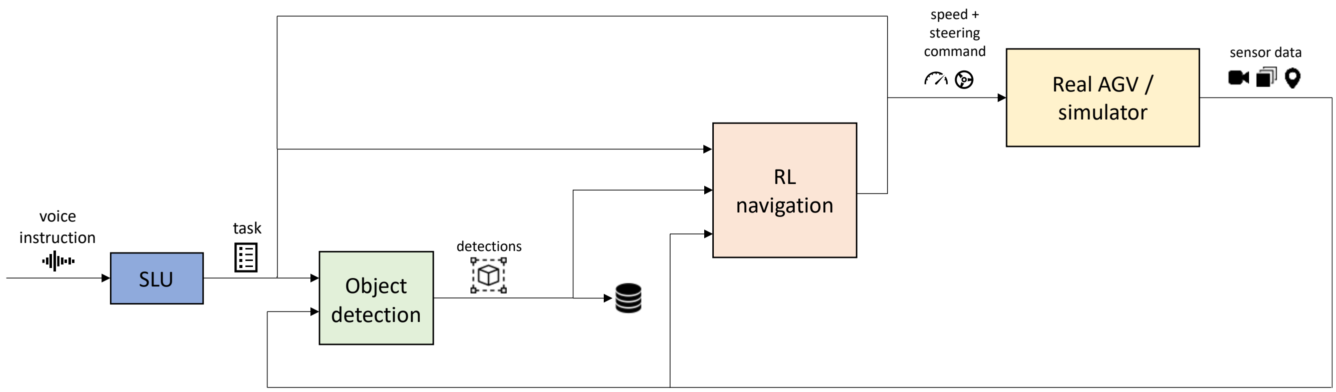


Figure 2. General architecture for a multi-modal AI autonomous platform.

II. Multi-Modal AI Architecture

The presented multi-modal AI architecture (Figure 2) is a general software architecture for the implementation of autonomous vehicles that based on AI can perform a particular set of tasks, instructed by speech. The architecture defines the different modules and interface, and can be implemented in different platforms with different hardware typologies. Even within the same implementation, different modules can run in different hardware units. Because the architecture exhibits a loose coupling, the modules can be easily exchanged for other models or algorithms, as far as they share the same interface. The proposed interface has human-understandable signals, which helps to improve the explainability of the system. The suggested architecture has a directed flow of information between the modules, which is represented by arrows in Figure 2. This defines and constrains the exchange of information between the different modules. The architecture has 4 main building modules: **(i)** Spoken Language Understanding (SLU), **(ii)** association between speech cues with sensor data for objects detection and localization, **(iii)** RL for navigation, which uses information from, speech, vision and sensor data and (**iv**) vehicle platform, which receives motion commands and sends processed sensor data.

(i) Spoken interaction offers fast and natural interaction with machines and AGVs, while operators keep their hands and eyes free for other tasks. The task of a SLU component is to map speech onto an interpretation of the meaning of a command, while taking the variability in the input signal into account: differences in voice, dialect, language, acoustic environment (noise, reverberation), hesitation, filled pauses and pure linguistic variation. Traditionally, SLU is approached as a cascade of Automatic Speech Recognition (ASR) mapping speech into text followed by Natural Language Understanding (NLU) mapping text onto meaning. This cascaded approach tends to propagate and inflate ASR errors and requires application-specific textual data, which is unnatural to acquire. Instead, this work uses End-to-End SLU (E2E SLU), where spoken instructions are directly mapped onto meaning without textual intermediate representations. The output of the speech module is a semantic definition of the task, which is then used by all the other modules. This module provides the unique interface where the user can provide inputs.

(ii) For agents to interact with the environment, they must process and understand visual input, i.e., extract the semantically relevant cues from the environment in order to execute the desired task. Should the input be provided from an RGB camera, a plethora of Deep Learning techniques could be leveraged to achieve visual understanding. Deep Learning techniques rely on Neural Networks, commonly (pre-)trained on large-scale general-purpose datasets, e.g., for visual recognition [10] such as object detection [11]. Since our goal is to interpret a language-based instruction, we need to locate the object(s) in the environment. To this end, we build on state-of-the-art object detection methods. Given an RGB input, the object detector's role is to locate (detect) the relevant objects. This serves as a backbone to perform multi-modal interaction by associating the representation of the language-based instruction with the representation of the spatial layout of the scene (2D location and categories of the detected objects). The RGB can be enhanced with depth information (RGBD camera or lidar) and vehicle localization for precise 3D location of the detections in world coordinates. The output of the vision module is used by the navigation module. However, for some tasks (e.g., automated inventory, finding/locating an object, getting attributes of an object, etc.) the output of the vision module is itself the principal result of the task, and is saved in a database, which the user can access.

(iii) Egocentric navigation is one of the core problems intelligent systems need to master. An agent needs this skill not only to execute the task at hand, but also to navigate, in order to collect experience that can be used to learn from. In the presented approach we have chosen for an end-to-end learning-based navigation approach. Such an approach is able to outperform Simultaneous Localization and Mapping (SLAM) based approaches [12], it does not suffer from propagation errors due to mapping errors, and excels in visually sparse environments [13]. In our architecture, we foresee several available RL agents, each one trained for a specific set of tasks. The navigation module receives the task directly form the speech module, and switches to the appropriate RL agent. Sensor data coming directly from the platform is used for dynamic obstacle avoidance and general exploration. Finally, the output of the vision module is used to direct the navigation to ensure that the exploration is done considering the relevant objects. As we need to train the RL agent in a simulation environment due to the large amount of required interactions, it is very important to couple the RL agent with a simulator that has the same interface as the real platform. Therefore, it is necessary to bridge the sim-2-real gap in two points: the acting gap and the observation gap. On the one hand, the acting gap refers to the interaction of the agent into the environment. For our architecture, this means making sure that the speed and steering commands have similar effects both in the real world and simulator. On the other hand, bridging the observation gap requires not only that the sensor data is similar in simulation and reality, but also that the simulator is able to produce similar object information as it would come from the real object detection.

(**iv**) The vehicle platform receives the control commands (set speed and steering wheel angle) from the navigation module. However, it can also be controlled directly by speed in case the speech task is directly affecting only navigation (e.g., “move slightly to the right slow”). The platform provides sensor data from the environment (camera, lidar and localization data) to the vision and navigation modules.

III. CASE STUDY – AUTONOMOUS OFF-HIGHWAY VEHICLE

In this case study, the vehicle is able to navigate towards a specific object, which is in the field of view, given a speech command [1]. The AGV used in this case study consists of the off-highway tractor developed at Flanders Make [14]. To perceive the environment we use cameras, lidars, a GNSS system and a microphone. The sensors data is then processed in separate computing platforms and stored on middleware (ROS), from where the Speech and Vision units send the information to the control block. This later is divided in two levels, (i) a High-level controller that controls the tractor via a state machine and (ii) a Low-level controller, built in a dSpace platform [15], that controls the trajectory such that velocity and heading can be followed. The output signals are sent to different actuators that consist of the brakes, throttle, steering and fork implement that are controlled via servo motors. Autonomous vehicle upgrades to deal with Multi-modal AI

An example of intuitive instructions given by an operator to the AGV to execute a task and their high level interpretations by the Multi-modal AI framework, described in this paper, is illustrated in Figure 3.

The instruction: ‘*Pick up the red pallet and put it on the truck*’, needs first to be communicated to the computer that runs the speech AI module (described in Section A). In the next level, a vision module, where real time 2d vision data is processed and fed to a pretrained NN, allows objects classification and their association to different attributes such as object’s type, color, etc. (as described in Section B). The AGV should then move towards the recognized object. This step is supported by the association made so far between speech and vision data as well as the navigation data. This later makes use of the cartesian coordinates of the AGV in the navigation space and the reinforcement learning module (as described in Section III.C) that allows to estimate the optimal trajectory between the AGV and the object of interest.

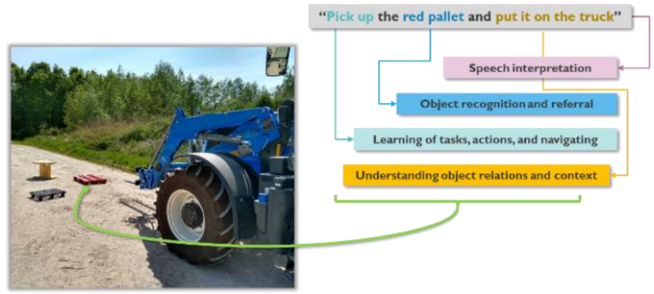


Figure 3. Example of speech-based instruction and multi-modal mapping.

In order to implement and demonstrate the Multi-modal AI framework, The AGV is updated by a newly installed system for interfacing through speech with a dedicated PC. This PC is also used for developing and testing the neural networks. It is equipped with a powerful Nvidia GPU and a new headset microphone for giving audio commands. The autonomous tractor internally uses ROS to communicate between the different sub-systems. Originally it was only used sparingly in the autonomous tractor, mainly to communicate lidar sensor data. After the system upgrade, also the control unit, the dedicated PC and the Nvidia Drive platform have a ROS interface. While the Nvidia Drive could technically runs the neural networks, for more convenience, during testing we installed the neural networks on the dedicated PC. Data from the cameras on the Nvidia Drive, LiDAR and navigation all come in as ROS messages while for speech a simple microphone is connected to the PC. The output of the multi-modal setup is the location of a specific object together with the task the tractor must complete. This information can be communicated through ROS to the navigation module.

A. *Spoken language understanding*

1) *Speech data generation*

To train the SLU model, training dataset with audio fragments is made. It is important that the recorded speech seems natural, as if the participants are really interacting with the AGV. To this end, we believe that a visual feedback to the participant would be very useful. Therefore, a simple automotive simulator called Webots [16] was used and a set of API calls were written in order to control the simulated tractor in the simulated environment (Figure 4).

Figure 4. Simulator that provides visual feedbacks to participants for speech recording.

The participants are given some high-level objectives and it is up to them to control the tractor with speech commands in order to fulfil these tasks. With the '*high-level*' objectives (in contrast with explicitly providing the primitive commands to the participants) we aim to improve the variability of commands that participant's would naturally choose to control the tractor. Every time the participants speak a relevant command, the experiment supervisor presses a button to invoke the correct API call. This way, we already have some automatically generated annotations linking the participant's speech command to the supervisor's API call invocation. We recorded the audio in Audacity in WAV format using a headset microphone and a separate standalone microphone. The commands were mainly basic control commands like turning a direction or driving speed. A total of 14 people who speak *Dutch* language (different dialects) were recorded with mixed female and male voices.

2) SLU model architecture & training

Classical *semantic frames* are used for representing the semantics of an utterance. A semantic frame is composed of *slots* (e.g., "direction") that take one of multiple *slot values* (e.g., "forward" or "backward"). This encoding represents the affordances of the AGV and corresponds to API calls with parameters filled in. The task of the SLU component is to map an utterance (spoken command) to a completed semantic frame. The SLU architecture follows the encoder-decoder structure first described in [17] and later refined in [18] to allow for encoder pretraining for ASR targets on generic *Dutch* data. The decoder is trained on the task-specific data.

The encoder encodes an utterance in a single high-dimensional embedding in two steps. The first step maps MEL-filterbank speech representations to letter probabilities using a transformer network [19] preceded by a down sampling CNN, trained maximal cross-entropy between predicted and ground truth transcriptions in a 37-letter vocabulary. The training data consist of 200 hours of Flemish speech with its textual transcription from the CGN corpus [20], fourfold augmented with noise (0-15 dB) and reverberation (sampled from [21]) to achieve acoustic robustness. The second step counts bigram occurrence frequencies of all letter pairs across the utterance and repeats the same while skipping one position in the bigram, resulting in a $2(37^2) = 2738$ dimensional utterance embedding.

The decoder maps the utterance embedding onto a multi-hot encoding of the slot values via non-negative matrix factorization (NMF) [22] as described in [17]. Other than in the pretraining stage, the training pairs here do not require textual transcription, but are pairs of speech with the completed semantic frame. Here, a neural network could be taken as well, but the chosen decoder has several advantages: (1) it requires few training data, (2) it retrains in a fraction of a second when user interaction data becomes available and (3) it establishes a bag-of-words model making the SLU system less sensitive to the rather free word order in *Dutch* (at least compared to *English*). Learning a stricter word order would require more task-specific training data exhibiting the word order variability.

The approach is evaluated on the Grabo corpus [23], which contains a total of 6000 commands to a robot spoken by ten *Flemish* speakers and one *English* speaker. The commands were recorded with the participants' own hardware in a quiet room at their homes. The semantics are described in eight different semantic frames describing driving, turning, grabbing, pointing, etc. using one (e.g., "*close* gripper") to three (e.g., "*quickly* drive *forward a little bit*") of ten slots (e.g., angle, direction, etc.), which can take between two and four different values. In total, 33 different meanings occur in the data. The accuracy is evaluated as the F1-score for slot values as a function of the number of task-specific training examples. The trained decoder is speaker-specific. The average accuracy over speakers is plotted in Figure 5 and shows that with the minimal of 33 training utterances, i.e., one example per meaning, an accuracy of over 98.5% is reached. The performance saturates around 180 task-specific utterances.

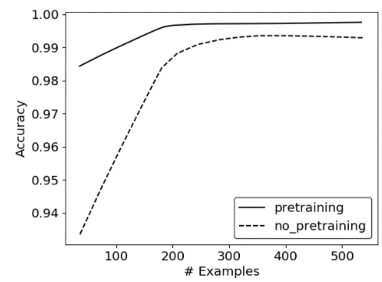


Figure 5. F1-score as a function of the task-specific training examples.

3) SLU model validation

For deployment we set up a docker container to run all the code. We developed a user interface to be able to easily visualize the results of the SLU model and provide training examples for training the decoder. In this interface, it is possible to record samples, open the microphone so the tractor can listen, give feedback to the model and retrain the model. After each command is given the confidence value of the prediction is estimated. Commands with sufficient confidence are forwarded to the tractor through ROS to the control PC.

The initial accuracy of the model depends a lot on the person giving the commands and their accent. But we were able to achieve high levels of accuracy of more than 90 percent in the noisy tractor environment using an active learning approach. In this approach, the operator can give feedback samples to retrain the model. In this experimental set-up, repeating an instruction in 5 instances proved to achieve high accuracy (90%). The retraining flow is quite time-efficient and takes less than a second to retrain.

B. Visual perception

1) Vision AI Objects detection and classification

a) Vision data generation

The dataset for training the vision model contains images with mostly objects that the AGV can pick up. This means mostly pallets and boxes of varied materials, shape and sizes containing materials like bobbins and wooden planks. This data was recorded on the Flanders Make local site, spread over two occasions: one on an early cloudy morning in spring

TABLE 1. QUANTITATIVE EVALUATION OF THE VISION AI TRAINED MODEL

	Vid. 1	Vid. 2	Vid. 3	Vid. 4	Vid. 5	Vid. 6	Vid. 7	Vid. 8	Vid. 9	Avg.
mAP	55.04	40.90	56.03	66.42	68.35	50.50	65.25	61.9	51.42	57.30

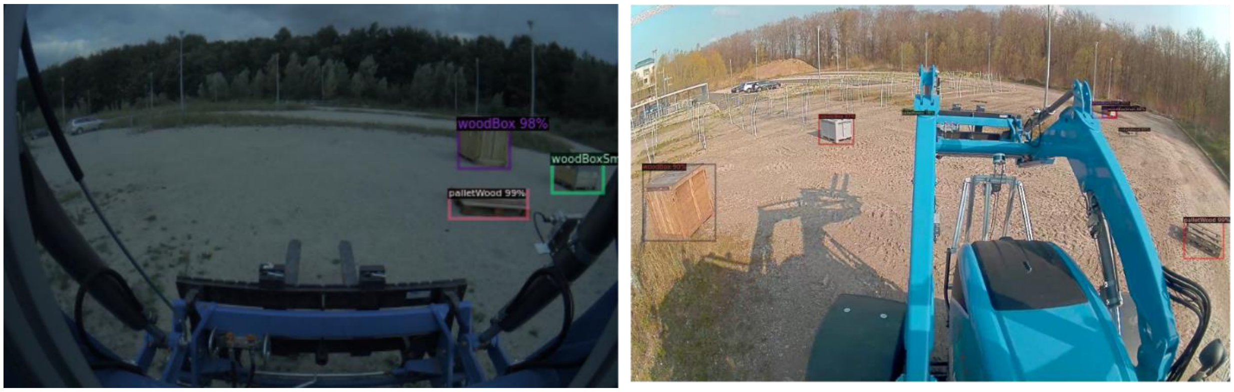


Figure 6. (left) all objects are correctly classified, (right) some objects are not detected.

and one just after noon in summer with sunny weather. Every image was recorded with a resolution of 960 x 608 pixels. The entire dataset contained 1100 images, derived from 9 videos. Each of these videos recorded one configuration of objects from many angles.

b) Vision NN architecture & training

The main building block of the vision pipeline is the object detector. It gets an RGB image I as input, where $I \in \mathbb{R}^{3 \cdot H \cdot W}$ and H and W are the image height and width respectively. The model we use is a state-of-the-arts two-stage object detector, where in the first stage, a region proposal network generates regions of interest for the image, and in the second stage, bounding boxes and object classes are predicted for each proposal, which exhibits an objectness score above a certain threshold. The region proposal network generates region proposals by sliding a spatial window over features map obtained from a Convolutional Neural Networks (CNN), i.e., a backbone. Additionally, the object detector includes a Feature Pyramid Network [24], a fully-convolutional module, which generates features maps at different levels, thus enabling the model to recognize objects at different scales. The object detector we use is a Faster R-CNN [25], with a ResNet101 backbone [26], pre-trained for general purpose object detection on COCO [11].

Even though less resource intensive FasterR-CNN backbones exist, such as MobileNets [27], given our computational budget, we find the FasterR-CNN variant we use to yield the best tradeoff between detection performance and speed (near real-time).

The model's outputs are object bounding boxes and classes with a confidence score for each. The confidence score for the predicted class is obtained as the Softmax probability of the highest scoring class.

We perform fine-tuning of the Faster R-CNN on images consisting of scenes from the environment, where the objects of interest are annotated with bounding boxes and classes. The images we use are video frames, extracted from 9 videos of the AGV navigating the environment while encountering the objects. Considering that the amount of data at our disposal is limited, we determine the optimal hyperparameters by training the object detector in a leave-one-out fashion, such that we train on a subset of 8 videos and perform evaluation on the remaining one. We iterate this process until we train a separate model on all unique subsets. The final model performance is averaged over each of the videos. We evaluate the model's performance using the standard COCO [11] mean average precision (mAP). The final model, i.e., the model used in the AGV, is trained on all 9 videos using the hyperparameters determined during the leave-one-out training/evaluation process.

We train the model for 5 epochs with a learning rate of 1e-4. We perform random horizontal flip data augmentation, enabling us to synthetically increase the dataset size and make the detector invariant to such transformations of the data. We sample a subset of 128 region proposals to estimate the regression and classification loss of the region proposal network.

We quantitatively evaluate each of the trained models on the videos, which were held-out during training. In TABLE **1** the lowest score is highlighted in red, while the highest scoring one is green. Overall, we observe that the performance is relatively high across all different videos (57.3 mAP). We further observe that the performance on Video 2 (Vid. 2), is significantly lower compared to the average performance. To inspect the reason for the lower performance, we qualitatively inspect the samples from Video 2 as discussed below.

We qualitatively evaluate the object detector's performance by visualizing the predictions on the held-out videos during training. In Figure 6 (left), we observe that the model correctly predicts all objects, which is in line with our expectations as the objects are fully visible and of a reasonable size. On the other hand, in Figure 6 (right) we observe several mis-detected objects of a frame from Video 2. We conclude that even though the model performs well, it struggles to recognize objects, which are (1) far from the camera (small size), and (2) occluded in the environment – both of which are active areas of object detection research.

2) Visual grounded SLU

To deal with the data sparsity, and to be able to ground (localize) the speech model output in the image, we perform discretization of the spatial layout (the bounding boxes and classes obtained as output from the object detector). To be specific, we perform mapping of both modalities to a canonical space, where we later measure the similarity between the output of the speech model and each of the detected objects in the image. To that end, we encode each detected object as a collection of one-out-of-k encodings of its label (box, pallet, etc.), material (wooden, plastic, etc.), size (regular or small), and location in the image. Note that the object category, material and size are jointly predicted by the object detector as object class. Lastly, we quantize the location of the object, i.e., we represent the object's location based on the object's horizontal and the lower vertical position. We showcase the grid over the image including the spatial references according to the x and y axes in Figure 7.

Finally, we represent each detected object as a vector of size 12, where we allocate 3, 2, 3, 3, 1 indices for the object's class, material, x-location, y-location and size respectively. When measuring the similarity between the speech model output (a vector of size 12 as well) and each encoded object detection, we explore different weighting strategies for each object attributes, which we discuss next.

Figure 7. Grid over image with object's spatial reference.

3) Adding Spatial relations

We evaluate different strategies for measuring the similarity between each (discretized) object detection and the speech model output. The output is a bounding box, which represents the grounding location of the instruction. We evaluate each grounding strategy on two variants of the dataset, namely (1) a descriptive variant, where the objects are commonly described based on their attributes, e.g., pick up the wooden box, and (2) a spatial variant, where the referred object is described based on its location in the frame, e.g., pick up the box furthest on the left. The grounding strategies we evaluate are:

1. Random matching (RM): A naïve baseline, where we ground the speech given instruction to a randomly selected bounding box. We establish a lower bound on the grounding performance with this baseline.
2. Basic matching (BM): We obtain the dot-product between the one-hot encodings of speech instruction and each object detection, representing the similarity.
3. Weighted matching (WM): We (re-)scale the contributions of the individual elements in the dot-product with pre-defined weights.
4. Confidence matching (CM): We represent the speech model with the confidence scores.
5. Weighted confidence matching (WCM): We use confidence scores for the speech model output and additionally weight the individual contributions using the pre-determined weights.

We perform evaluation using the standard grounding accuracy metric, where we score a hit if the predicted grounding bounding box has intersection over union (IoU)>0.5 with the ground truth box. For the random baseline, we perform inference 5 times and report the average performance. Through grid-search, for the weighted modules (WM, WCM) we use a weight of 0.1, 0.7, 0.2, and 0.05 for the spatial indicators, the object class, the object material and the object size respectively. We report the results in TABLE 2.

We observe consistent gains when we weigh (WM) or use the speech model confidence scores (CM) in the grounding, compared to the baseline basic matching (BM) method. Additionally, a combination of the weight and confidence matching (WCM) yields superior results across the different data (descriptive, spatial) and significantly outperforms the other methods. Lastly, even though the spatial data is more challenging than the descriptive data, the WCM module performs well, indicating that by re-weighting and adding confidence scores, we can ground spatial speech data reasonably well.

C. Reiforcement learning based navigation

In this section, the navigation part of the Multi-modal AI and its association with the speech-vision data is described. The currently developed proof of concept consists of a simulation environment with the hardware in the loop.

To make this simulator as close to real life as possible, a 3-D scan of the test environment by using an aerial scanning using a drone with photogrammetry capabilities that allows us to map images to a high fidelity 3-D twin of the area. This twin was then imported to the simulator for the purpose of reinforcement learning.

1) RL archiecture & training

The presented Reinforcement Learning (RL) approach makes use of the DD-PPO (Decentralized Distributed Proximal Policy Optimization) architecture [28] (Figure 8). The Reinforcement Learning (RL) approach is able to map high dimensional inputs to discrete actions. The DD-PPO model consists of a visual pipeline, for which in our case we use a ResNet18 [26].

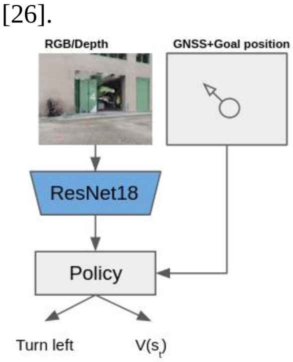


Figure 8. DD-PPO architecture overview.

TABLE 2. EVALUATION OF THE AI MODEL WITH SPATIAL RELATIONS

Method	**Dataset type**	
	Descriptive	**Spatial**
RM	25.91	17.14
BM	65.91	59.52
WM	70.45	57.94
CM	76.14	62.70
WCM	**79.55**	**65.87**

The resulting learned visual representation is concatenated together with a GNSS sensor. This output is then passed onto a recurring policy consisting of 2 Long short-term memory (LSTM) **[29]** layers. The final outputs of the model consists of a state value estimation, and an action distribution from which actions (move forward, turn left, turn right and stop) can be sampled. The stop-action should be executed by the agent when positioned less than 2 meters of the goal position. As inputs for the model we tested a single depth camera, a single RGB camera, or a combination of both RGB and depth. We use these sensors as they are cheap and widely available. The camera is positioned on the front of the AGV.

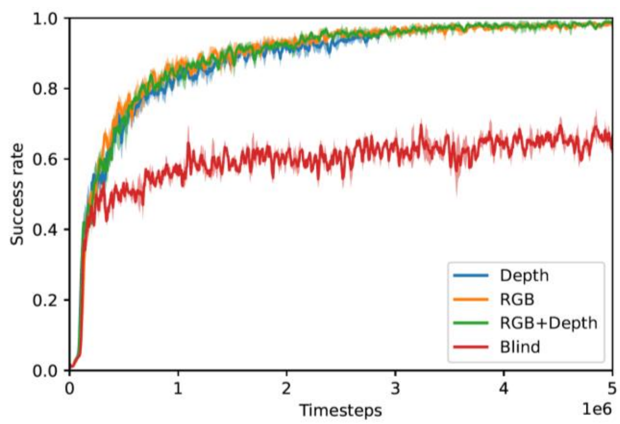


Figure 9. Training performance. The blind agent can perform basic navigation by relying on the GNSS sensor, however to further improve to near perfect results an additional RGB of depth sensor is required to detect and avoid collisions.

To train the agent we use the improvement in geodesic distance between the agent and the goal position as a dense reward signal. A slack penalty of -0.01 is subtracted on each step, and a termination bonus of 2.5 is awarded upon successfully utilizing the done action. We train the agent entirely in the Habitat simulator [12] where a photorealistic scan of the environment is used. This allows the agent to interact with the terrain in a safe way. While in this case we trained the agent to specifically work on a single environment, DD-PPO also allows generalization to unseen environments, given enough different training environments and training samples. Figure 9 shows the required number of interactions with the environment. These results indicate that in this setting the agent relies mostly on the GNSS sensor, as the blind agent performs reasonably (60% success rate after 5M training interactions). However, by adding either a depth or RGB sensor the agent achieves near perfect navigation capabilities on the training set after 5M interactions with the simulated environment.

2) RL validation

Realizing Reinforcement learning on a large autonomous platform brings in multiple challenges to the board. For safety concerns, the approach to validate the system was to use a Hardware-in-loop setup (Figure 9) along with the digital twin of the environment. The main input from the real world was the signal from the GNSS receiver (Septentrio AsteRx-U) on the AGV, which was then mapped to the digital twin coordinates system. The GNSS had a dual antenna setup, which could then provide the heading of the platform as well. Using a cloud-based service updates were provided in real time to the simulator/digital twin environment to position the simulated tractor same as the one in real world. The output from the simulator was the suggested trajectory to the goal pose.

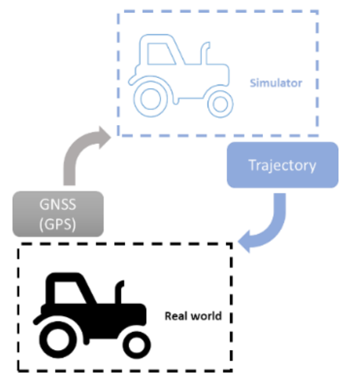


Figure 10. Hardware In Loop setup (overview).

To evaluate the navigation capabilities of the agent, we created a holdout dataset. This holdout dataset contains goal positions the agent did not see during training. TABLE 3 contains the results of 100 tested episodes. In TABLE 3, the success rate indicates the amount of episodes the agent could complete successfully. The Success weighted by Path Length (SPL) measurement also considers the length of the path taken.

TABLE 3. SUMMARY OF TESTED EPISODES

Sensors	**Success Rate**	**SPL**	**Avg. Collisions**
RGB	100%	0.9454	0.4355
Depth	100%	0.8882	0.1129
RGBD	100%	0.9272	0.5161
Blind	91.94%	0.7294	4.3548

Figure 11. Snapshot of the demonstrator of the AGV Multi-modal AI framework: (top left), the Speech model interface, (bottom left), the Vision model interface, (right), the Navigation digital twin interface, (bottom middle), the estimated trajectory between the AGV.

D. AGV Multi-modal AI Demonstration

To demonstrate the full methodology, we combined the methods respectively described in Sections A, B and C in one demonstrator implemented in the AGV. We added all the information in a new docker environment to be able to run on the dedicated PC in the AGV. There is a similar user interface compared to the SLU model where you can record your voice and use the NLP model to predict the voice commands. These commands consist of the description of the object and the task the AGV should do. Then the fusion model uses this information to link an object description with a detection from the vision model to predict the location of the describer object on the image. As a last step the lidar data is used to link the 2D location on the image to a 3D location of the object in the world coordinate space. This location can then be sent further as a goal to the control systems together with the described task from the NLP model. A significant improvement could be made in the parameters of the fusion model. There was a bias against using spatial information in the voice command. The material of the object is more difficult to extract on the image than its location, so using the location for finding the correct object is more reliable. Hence, we tuned some of the weights to have a bigger focus on this kind of information. Another small improvement could be made to the audio side. The person dedicated to controlling the AGV added some voice samples and gave feedback to the model through the user interface. This way the model was more confident in recognizing their accent and way of talking. With regards Navigation, although the approach is not fully implemented in the rea system, the approach can already be demonstrated by Hardware-In-the-Loop. In this setting an instance of the simulator is constantly synchronized with the AGV. This is done by using the GNSS position from the real-world AGV to set the position of the agent in the simulator. We can use the digital twin to generate trajectory paths. These generated trajectories can then be used in the real-world by the AGV. A snapshot from the full demonstrator is depicted in Figure 11.

IV. CASE STUDY – INDOOR AUTOMATED INVENTORY

The second case focuses on the task of automated inventory of unknown warehouse settings [2]. The goal is to explore with a good tradeoff between navigation time and inventory accuracy. An open experimental platform has been built on top of an AGV, automating a standard pallet forklift [30]. Localization is provided by a commercial system with reflector landmarks with known positions across the warehouse. Triangulation allows to get the AGV position with an accuracy of the order of few centimeters.

Two Ouster OS1 lidar with 64 vertical layers have been used. They have a vertical field of view of 45° and a maximum range of 120 m. They are placed in the front and the back of the AGV, and they are merged into a single point cloud that has a full 360-degree coverage. A camera (Zed mini) is used for inventory detection and is placed at the front of the forklift.

ROS is used as a middleware to provide communication between the different perception modules. Then, control commands are sent to a motion module via ethernet, which is responsible for executing the actions on the AGV. There is a safety system mainly based on safety scanners that stops the forklift in case of an expected imminent collision.

The dynamics of the forklift can be summarized in the kinematic bicycle model [31]. This model is used in the RL training bridge the sim2real gap in the actuation. In Figure 12 the vehicle model can be seen.

The kinematics for a forklift AGV are defined by the following equations [32]:

$$\dot{x} = V(t)\cos\theta(t)$$
$$\dot{y} = V(t)\sin\theta(t)$$
$$\dot{\theta} = \frac{V(t)\tan\delta(t)}{l - a\tan\delta(t)}$$

The following values apply for this work AGV: l = 1.5m, a = 0:15m The forward velocity is denoted as V and δ is the steering angle in radians.

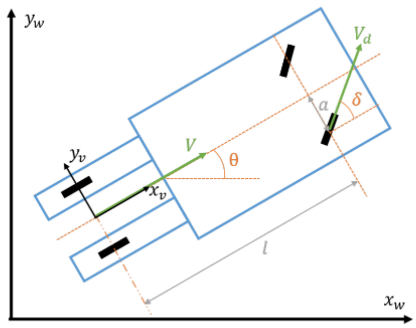


Figure 12. Bicycle kinematic model for AGV.

A. Spoken language understanding

We have used the same SLU module than the one from the previous use case and we have retrained it to work for a new set of tasks. For this use case we have trained the model in English, showing that the speech recognition can work well in different languages given pairs of audio signals and tasks.

If the operator wants to give a speech command, he/she can either press a button and then start talking, or enable the open microphone feature and say a pre-defined keyword to indicate that an instruction will be given. Three different kind of possible tasks have been selected for this use case. First, a command is available to start a new inventory session ("count"). Then, there are 3 options available: steer the AGV manually, trigger the RL autonomous exploration ("explore"), or further give speech instructions to control the movement of the AGV ("move"), such as "forward", "a little bit to the left", or "stop".

B. Visual Perception

1) Object Detection and tracking

The detector uses an RGB image as input and produces bounding boxes with associated confidence scores. We do not use depth sensors or lidar. The reason is that training models which use these sensors would require 3D annotations, generally not available in industrial datasets. An alternative is to label point clouds, which is prohibitive, and therefore we opt only for 2D object detection applied on RGB images.

We use the 3D lidar sensor, available in the navigation module, to obtain depth information which is pixel-by-pixel aligned with the RGB images. This approach provides better depth accuracy than depth cameras. The point cloud from the lidar is projected on the camera plane, with some inflation proportional to the depth value, leading to higher inflation for closer points. This provides a richer depth image, as illustrated in Figure 13. The projection of a point cloud into a camera plane only works well only if the two sensors are mounted close enough, which is the case for our platform.

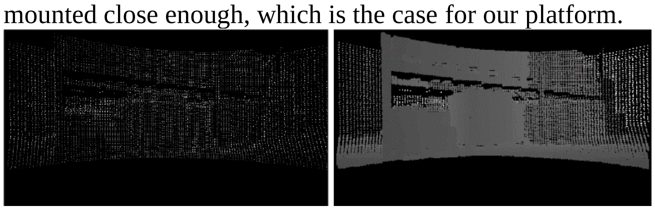


Figure 13. Depth image from the point cloud without inflation (left) and with inflation (right).

We choose the Yolov7 detector [33], as it is one of the latest open-source detectors with a better trade-off between accuracy and real time performance. Starting from a pre-trained version on COCO dataset [11], 4 videos recorded in the test warehouse have been annotated, making a total of around 1500 frames. The detector is trained to detect only one class, which is the cardboard box.

We select BYTETrack [34] as an object tracker, because it can be easily coupled with any other detector and yields to good accuracy in the MOT20 [35] benchmark. The main building block is a Kalman filter [36] with a constant speed model for the bounding box position and size of the detections. In most cases, trackers are employed in applications with a static camera and moving objects, while we use a moving camera with static objects. We have slightly modified the default version to be able to tune the covariance matrices Q and R of the Kalman filter in order to put a higher confidence on the detections (measurements) than in the model (constant speed motion). Especially when the camera is turning, the model will be less reliable, so we want to give higher importance to the new detections. Tracking provides unique IDs across frames, but does not solve the problem of tracking objects when they re-enter the camera FOV after some time. This will be addressed in the 3D map creator.

2) 3D map creator

The individual 2D detections, the generated depth image and the AGV location in the warehouse are inputs to the 3D map creator, which is responsible to merge new detections to the ones in the map. This way, it keeps an updated version of the counted items locations, which are represented as cuboids with an ID, confidence score, internal point cloud, center, width, height and depth. The 3D map also keeps track of the uncertain areas, which are represented in the same way but with a negative value for the ID. Algorithm 1 shows the pseudo-code of the map creator, including also the object detector and tracker.

For each new frame the algorithm iterates over the bounding boxes from the tracker. For each track, the corresponding depth pixel values are retrieved with a padding to discard pixels that may belong to the background. Then, depth values are converted back to a point cloud per detection. This point cloud goes through a filtering process that includes a Statistical Outlier Removal (SOR), a passthrough filter to remove far points and a SAmple Consensus (SAC) test: using the domain knowledge that boxes have flat surfaces and that they are never seen from above, we fit a plane and require it to be vertical in the world coordinate system.

At this point, we have for each detected object a point cloud, which generally contains points on the main surface of the box. There are two reasons to consider it uncertain:

- Uncertainty in the detector output: If the confidence score provided by the detector is below a certain threshold, then the corresponding object is marked as uncertain in detection.

Algorithm 1: Pseudo-code of the inventory monitoring

Input: sequence S with image I, lidar point cloud L and vehicle position P ; threshold for tracking T_t ; detection confidence threshold for counting T_d ; position confidence threshold for counting T_p

Output: goods map M (list of objects with ID, confidence score, point cloud and 3D cuboid)

Initialization: $M \leftarrow 0$
for I, L, P in S **do**
  $Dets$ = detector(I)
  $Tracks$ = tracker($Dets$, T_t) % Tracks contain an ID, confidence score and bounding box
  $Depth$ = project_pointcloud(L)
  for $Track$ in $Tracks$ **do**
    $Depth_{filtered}$ = filter_depth($Depth$, $Track$) % Depth with padding
    O_{track} = to_pointcloud_object($Depth_{filtered}$, T_d) % object with point cloud, ID (<0 for uncertain) and confidence fields
    $O_{filtered}$ = filter_poitcloud(O_{track}, T_p) % SOR, passthrough, SAC filters + ID becomes <0 if uncertain position
    O_{world} = to_world($O_{filtered}$, P) % transform from ego view
    $O_{current}$ = compute_cuboid(O_{world}) % add 3D box to object
    $Test$ = overlap_test($O_{current}$, M) % compare to all map objects
    if $Test$ **then**
      M = merge_to_map($O_{current}$, M) % discard new ID & merge
    else
      M = add_to_map($O_{current}$, M) % new detection added to map
    end
    M = voxel_grid_filter(M)
    M = delete_uncertain_areas(M)
  end
end
Return M

Figure 14. Algorithm for inventory monitoring

- Uncertainty in the object location: In case the SAC plane is too far away, has a low number of inliers, or is not seen frontally (the boxes are too much at the side of the image), then the corresponding object is marked as uncertain in position.

The point cloud is finally transformed using the vehicle location into world coordinates, and a 3D cuboid that encloses the point cloud is computed.

Then, all the detections are merged with the map. There are two possibilities:

- The ID of the current detection is already in the map. In that case, the default option is to merge it with the map's object with the same ID. However, it could be the case that the 2D tracker fails, so an overlapping volume comparison is done with all the other detections already in the map, and if there is enough overlapping, the current detection is merged with the map object with more overlapping volume.
- The current detection is not in the map. The same overlapping test is done as in the case above. If there is not enough overlapping, it is a new detection, and a new object is initialized in the map. Otherwise, the new detection is merged into the matched object in the map.

When a detection is merged to one in the map, the point clouds are concatenated and then reduced using a voxel grid filter. The confidence is updated to the maximum of the ones being merged, and the centroid and vertex locations are updated fitting a cuboid to the point cloud. Since only one surface per box is considered, the cuboid corners are extended so each dimension is bigger than a user defined minimum object size. The current vehicle position and relative viewpoint respect the detection are used to know the direction of the extension.

Uncertain detections are merged in a similar way as certain ones. Certain and uncertain detections are never merged between them. When an uncertain detection with a particular ID becomes certain, all the uncertain data is deleted. Moreover, whenever there is a certain detection being added or merged to the map, nearby uncertain detections are deleted. Finally, in case that the AGV gets close enough to an uncertain detection and it remains uncertain, the object is completely discarded, since after having a good viewpoint the certainty did not increase enough, so it is assumed to be a detection false positive.

C. Reinforcement learning based navigation

We address the sim-2-real gap in the sensing part by using lidars, which are more robust to sensor noise. While lidar-based simulations are often very compute-intensive, our approach allows fast simulations by rendering obstacles into top-down images containing the lidar data, without any need for ray casting. Rack locations are similarly added as a second image channel, and a third channel contains past vehicle positions. This 3-channel image in the ego view (see Figure 15) determines the only input of the RL agent. The same 3-channel image is created in the real setup:

- The obstacles channel comes from a projection of the 3D lidar point cloud to the plane parallel to the floor.
- The second channel contains the areas to direct the exploration, which come from the detection module. A 3D point cloud is projected as in the first channel.
- The third channel contains the past trajectory, which is obtained by concatenating the last positions given by the AGV positioning system.

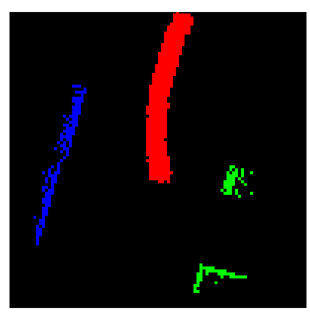


Figure 15. Input image to the RL agent. Blue are obstacles, green represents uncertain areas and red is the past trajectory

Simulations use a kinematic model of the AGV to bridge the sim-2-real gap in the acting part. The RL policy utilizes a discrete set of 15 actions, that map to specific steering angles and forward speeds. At a low speed (0.3 m/s) the vehicle can turn at 3 different angles (small, medium and large) to the left, and 3 to the right. The vehicle can also go straight. This makes a total of 7 actions, which are also available for backward moving. The 15th action allows to go forward straight at a higher speed (0.5 m/s). The simulation environments are randomly generated to create several rack configurations and generalize to any warehouse setting. We use Proximal Policy Optimization (PPO) [37] to train the agent.

D. AGV Multi-modal AI Demonstration

We have integrated all the algorithms in the forklift AGV platform and performed several online real-time experiments. Figure 16 shows the available inventory visualization in an experiment sequence. The locations of the racks are provided by the user and are only employed to improve the visualization, as they are not part of the algorithm. In the Figure 16 top image it is seen how several boxes in the middle rack have already been detected while in another rack there are uncertain detections. White points denote areas with low detection certainty, while grey points correspond to low certainty in location. Those areas direct the navigation to move closer, and once better viewpoints are obtained, they become certain detections that are added to the inventory, as seen in the middle image. Finally, in the bottom image it is seen how after performing a loop around the middle rack, the previous 2 racks are seen again, but only new objects are added to the inventory count. Detections that are assigned to an object already in the map are merged, and the object location is slightly adjusted accordingly if necessary.

TABLE 4 contains the results for object detection. We have used a test subset of 188 frames of around 30 seconds where the vehicle goes towards a rack and then performs a turn. The "*Detector alone"* row contains the results of the detector without any tracking or merging on the map. Then, the following rows represent the results for different ablations on the map creator, where the thresholds to track (T_t) and to count (T_d, T_p) are modified. *H* represents a version where the several thresholds for the position certainty are high, while *L* is for low values. We denote as $T_t=0$ the case where the 2D tracker is not used. The results include the precision and recall values, as well as the number of detected uncertain objects that are remaining in the map at the end of the sequence. A distinction is done between remaining uncertain objects that would become true and false positives if added to the count. Although accuracy values in the *"Detector alone"* are high, all versions with the 3D map creator have a higher precision and similar or higher recall. Depending on the thresholds to track the objects and to count them in the inventory, the trade-off between precision and recall changes. In our application a high precision would be desired, while we expect to improve the recall by the active navigation. Results show there is still room for improvement in the directed exploration, since there are several true positive uncertain detections that were not yet included in the map. Alternatively, counting and position thresholds could be further reduced to count those uncertain detections and increase the recall, but that would reduce precision. Results show how the usage of a 2D tracker ($T_t \neq 0$) helps to avoid false positives, as seen in the TABLE 4.

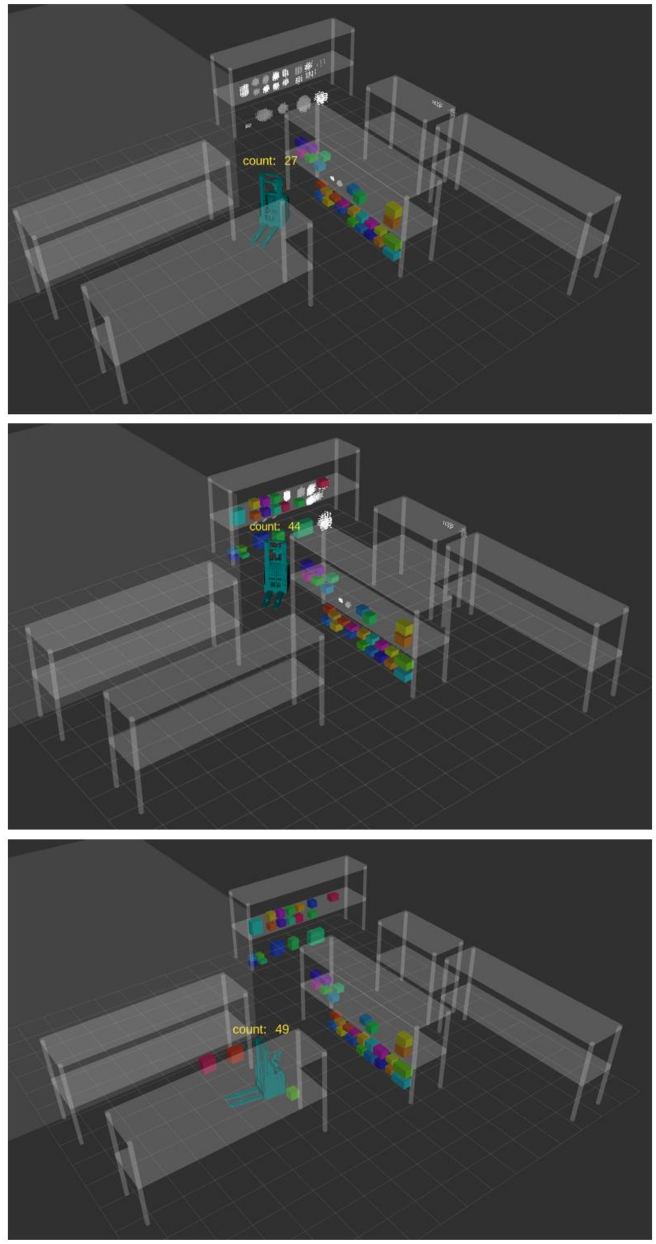


Figure 16. Sequence of the forklift around some racks in a warehouse.

TABLE 4. RESULTS OF THE OBJECT DETECTION

	Precision	Recall	Uncertain (T/F)
Detector alone	0.89	0.85	-
T_t=0.3, T_d=0.9, T_p=H	1	0.76	9/0
T_t=0.3, T_d=0.5, T_p=H	1	0.81	7/0
T_t=0, T_d=0.5, T_p=H	1	0.81	7/7
T_t=0.3, T_d=0.9, T_p=L	0.96	0.86	5/0
T_t=0.3, T_d=0.5, T_p=L	0.97	0.89	3/0
T_t=0, T_d=0.5, T_p=L	0.94	0.86	3/8

Results show how, by using spatial-temporal information of the same object while actively navigating to obtain better viewpoints, we can rely in a less accurate detector and achieve higher accuracy results on the high-level task of inventory count. This directly translates into a faster set up of the detector (less required labeled data, less time doing hyperparameter tuning, etc.), which is critical to reduce the implementation time of the solution in a new or modified warehouse. In this direction, the usage of an instance segmentation detector would have provided pixel level detections, which could be better matched to depth information leading to better position accuracy in the map. However, this would have increased the inference rate and the labeling effort. Our results show, how by post-processing the lidar data and registering to the inventory only detections with high position accuracy, a bounding box detector is enough instead of a more advanced pixel level instance segmentation detector.

V. Conclusion

In this work, we developed and demonstrated a multi-modal AI framework that allows to intuitively instruct production AGVs to perform multiple tasks. The interface with operators is allowed by speech interaction that is decoded through an AI NLP model to translate speech commands to interpretable instructions by all the components of the AI Framework. Associations with vision and navigation data are done to be able to perform a wide range of tasks. We show how the loose coupling between the modalities creates a an architecture which is general enough to be applied in a wide set of tasks for two different use cases, which run on very different hardware platforms. Moreover, the loose coupling of the modules provides a clear interface between the modalities (e.g., task, object detections, motion commands) which is interpretable by humans, thus leveraging the explainability. The main outputs of the system are the control commands that enable the vehicle navigation, and relevant task information (e.g., location of objects) which is provided to the user. The demonstrators remain, however, a research proof of concept (to demonstrate the approach) and require different improvements before effective industrial usage. This includes, amongst others, training with larger datasets (speech, vision, navigation) and evaluation in an extended number of scenarios. Moreover, bridging the sim-2-real gap for the RL navigation is still a challenge in terms of achieving the necessary robustness for industrial applications.

Acknowledgment

This research is done in the framework of Flanders AI Research Program (https://www.flandersairesearch.be/en) that is financed by EWI (Economie Wetenschap & Innovatie), and Flanders Make (https://www.flandersmake.be/en), the strategic research Centre for the Manufacturing Industry who owns the AGV infrastructure. The authors would like to thank everybody who contributed with any inputs to make this publication.

References

[1] A. B. Temsamani et al., "A multimodal AI approach for intuitively instructable autonomous systems : a case study of an autonomous off-highway vehicle," *The Eighteenth International Conference on Autonomic and Autonomous Systems,* pp. 31-39, 2022.

[2] F. Gebelli Guinjoan et al., "A Multi-modal AI Approach For AGVs: A Case Study On Warehouse Automated Inventory," *The Nineteenth International Conference on Autonomic and Autonomous Systems,* pp. 25-33, 2023.

[3] D. Li, B. Ouyang, D. Wu and Y. Wang, "Artificial intelligence empowered multi-AGVs in manufacturing systems," in *ArXiv abs/1909.03373,* 2019.

[4] L. Radder and L. Louw, "Mass customization and mass production," *The TQM magazine,* vol. 11, pp. 35-40, 1999.

[5] M. De Ryck, M. Versteyhe and F. Debrouwere, "Automated guided vehicle systems, state-of-the-art control algorithms and techniques," *Journal of Manufacturing Systems,* vol. 54, pp. 152-173, 2020.

[6] D. Herrero-Perez and H. Martinez-Barbera, "Decentralized coordination of automated guided vehicles," *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems,* vol. 3, pp. 1195-1198, 2008.

[7] M. Mousavi, H. J. Yap, S. N. Musa, F. Tahriri and S. Z. Md Dawal, "Multi-objective AGV scheduling in an FMS using a hybrid of genetic algorithm and particle swarm optimization," *PloS one,* vol. 12, p. 12(3): e0169817, 2017.

[8] C. Stachniss, J. J. Leonard and S. Thrun, "Simultaneous localization and mapping," *Springer Handbook of Robotic,* no. Springer, pp. 1153-1176, 2016.

[9] S. HT and C. Arjun, "Design of Voice Controlled Automated Guided Vehicle," *International Journal of Science Technology & Engineering,* vol. 3, pp. 90-93, 2017.

[10] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM,* vol. 6, pp. 84-90, 2017.

[11] T.-Y. Lin et al., "Microsoft COCO: Commeon objects in Context," *13th European Conference in Computer Vision,* pp. 740-755, 2014.

[12] M. Savva et al., "Habitat: A platform for embodied ai research," *Proceedings of the IEEE/CVF international conference on computer vision,* pp. 9339-9347, 2019.

[13] D. Mishkin, A. Dosovitskiy and V. Koltun, "Benchmarking classic and learned navigation in complex 3d environments," in *arXiv preprint arXiv:1901.10915,* 2019.

[14] Flanders Make, "Automated off-highway vehicle test platform," [Online]. Available: https://www.flandersmake.be/en/testing-validation/product-validation/automated-off-highway-vehicle-test-platform. [Accessed 10 2 2023].

[15] dSpace, "Real-time testing system (dSpace)," [Online]. Available: https://www.dspace.com/en/pub/home.cfm. [Accessed 10 2 2023].

[16] Cyberbotics, "Webots - Open source robot simulator," [Online]. Available: https://cyberbotics.com/. [Accessed 2 10 2023].

[17] B. Ons, J. F. Gemmeke and H. Van hamme, "Fast vocabulary acquisition in an NMF-based self-learning vocal user interface," *Computer Speech & Language,* vol. 28, pp. 997-1017, 2014.

[18] P. Wang and H. Van hamme, "Pre-training for low resource speech-to-intent applications," in *arXiv preprint arXiv:2103.16674,* 2021.

[19] A. Vaswani et al., "Attention is all you need.," *Advances in neural information processing systems,* vol. 30, 2017.

[20] N. Oostdijk, "The Spoken Dutch Corpus. Overview and First Evaluation," in *Proceedings of LREC*, 2000.

[21] RWTH Aachen, "Aachen Impulse Response Database," [Online]. Available: https://www.iks.rwth-aachen.de/en/research/tools-downloads/databases/aachen-impulse-response-database/. [Accessed 10 2 2023].

[22] D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances in neural information processing systems,* vol. 13, 2000.

[23] KU Leuven, "ALADIN: Adaptation and Learning for Assistive Domestic Vocal Interfaces," [Online]. Available: https://www.esat.kuleuven.be/psi/spraak/downloads/. [Accessed 10 2 2023].

[24] T.-Y. Lin et al., "Feature Pyramid Networks for Object Detection," *Proceedings of the IEEE conference on computer vision and pattern recognition,* pp. 2117-2125, 2017.

[25] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in neural information processing systems 28,* 2015.

[26] K. He, X. Zhang and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[27] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861,* 2017.

[28] E. Wijmans et al., "Dd-ppo: Learning near-perfect pointgoal navigators from 2.5 billion frames," *arXiv preprint arXiv:1911.00357,* 2019.

[29] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural computation,* vol. 9, pp. 1735-1780, 1997.

[30] A. Bartic, "Autonomous vehicles can perform an increasing array of tasks all by themselves," Flanders Make, 28 April 2020. [Online]. Available: https://www.flandersmake.be/en/blog/autonomous-vehicles-can-perform-increasing-array-tasks-all-themselves. [Accessed 1 February 2023].

[31] P. Polack, F. Altche, B. d'Andrea-Novel and A. de La Fortelle, "The kinematic bicycle model: A consistent model for planning feasible trajectories for autonomous vehicles?," *IEEE intelligent vehicles symposium (IV),* pp. 812-818, 2017.

[32] K. Jung, J. Kim, J. Kim, E. Jung and K. Sungshin, "Positioning accuracy improvement of laser navigation using UKF and FIS," *Robotics and Autonomous Systems,* vol. 62, pp. 1241-1247, 2014.

[33] C.-Y. Wang, A. Bochkovskiy and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *arXiv preprint arXiv:2207.02696*, 2022.

[34] Y. Zhang et al., "ByteTrack: Multi-object Tracking by Associating Every Detection Box," *European Conference on Computer Vision,* pp. 1-20, 2022.

[35] P. Dendorfer et al., "Mot20: A benchmark for multi object tracking in crowded scenes," in *arXiv preprint arXiv:2003.09003*, 2020.

[36] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Fluids Eng,* vol. 82, pp. 35-45, 1960.

[37] J. Schulman, F. Wolski, P. Dhariwal, A. Radford and O. Klimov, "Proximal Policy Optimization Algorithms," in *arXiv:1707.06347*, 2017.

Project Management for Online Course Quality Management

Brayden Milam, Julie R. Newell, Stephen Bartlett, Tamara Powell, Kaylee Polk, Elly Sloman, Aiden Reichner, Lesley Gabel
Norman J. Radow College of Humanities and Social Sciences, Office of Digital Education
Kennesaw State University, Kennesaw, Georgia, USA
Email: bmilam3@kennesaw.edu, jnewell2@kennesaw.edu, sbartlet@kennesaw.edu, tpowel25@kennesaw.edu, kpolk7@students.kennesaw.edu, esloman@students.kennesaw.edu, areichne@students.kennesaw.edu, lgabel@kennesaw.edu

Abstract— **Post-pandemic, a requirement that online courses be certified for accessibility and meet federal guidelines for instructional equivalency created a backlog of courses that needed to be processed quickly and efficiently. This situation was particularly acute in the Radow College of Humanities and Social Sciences, the largest College at Kennesaw State University, and its 11 component schools and departments. Without the availability of increased staff or funding, efficient workflow management systems needed to be established quickly and using software already available at the institution. Team members developed the "Bucket System" based on multiple components of Microsoft 365 to map and track workflows across multiple criteria and multiple reviewers. This paper builds on a previous presentation by exploring existing project management methodologies and expanding on how the team built their framework, software system, and supplementary tools to meet needs specific to higher education institutions.**

Keywords-course review; online course quality; project management methodology (PMM); monitoring and evaluation (M&E) framework; Microsoft 365.

I. Introduction

This project, first presented at eLmL 2022: The Fourteenth International Conference of Mobile, Hybrid, and On-line Learning [1], originally detailed a year-long, large-scale course quality review endeavor conducted by the Office of Digital Education (ODE) in the Radow College of Humanities and Social Sciences (RCHSS) at Kennesaw State University (KSU), a public university in metro-Atlanta, Georgia. In summer 2022, shortly after the conference we presented our findings at, college and university leadership decided to reassign the course review process to the university-level equivalent of the ODE. While no further data was collected, this paper explores the existing project management methodologies consulted during our initial conversations and expands on how our team built the framework, software system, and supplementary tools we used to support the process we developed.

In 2018, KSU President Pamela Whitten chose to discontinue the requirement of Quality Matters (QM) Certification for all online classes. Colleges within the university could still require quality control, but it was left to college leadership to decide if and how to implement this control. RCHSS continued to offer training and certification options for those departments requesting support. Training and certification were also built into compensation contracts offered by the college. Despite these continued options, the college decided to allow the requirement for instructor and course certification to lapse.

Instead of following QM, which cost money and received opposition from faculty resistant to the latest version update, digital education experts in what was then the Distance Learning Center partnered with faculty to develop a set of standards that fit the university's needs and aligned with best practices. This process ultimately resulted in the KSU Course Quality Checklist (KSU CQC) and was approved through various shared governance bodies. While some colleges chose to remain with QM (after memberships were provided through the University System) and others developed their own rubrics, the KSU CQC was adopted across multiple colleges at the university and has become the standard quality measure in RCHSS in instances when course review is required.

When the pandemic hit in March 2020, the focus across KSU shifted from offering high quality courses to simply *offering* courses. Suddenly, everyone needed to be online. Professional development shifted with the needs of faculty and training facilitators as they struggled alike to adapt to the relatively new modality of synchronous online education. The nearly four-month period of online-only education at the start of the pandemic allowed misconceptions about online teaching and learning to spread as digital learning experts scrambled to teach basics to faculty with little-to-no online experience. Then, in Fall 2020, as the return to campus began, faculty were expected to teach in various rotational and hybrid formats with many having had little-to-no training in these new modalities. This trend continued into Spring 2021, with a complete return to campus planned for Fall 2021. By this point, the combined effects of eliminating required instructor and course certification for online offerings and unsound pedagogical approaches created by pandemic pandemonium were beginning to manifest in student complaints about the quality of online education. Studies [2][3][4] show "(1) stress and negative emotions increased, (2) positive emotions, enjoyment with class, satisfaction with class decreased, (3) engagement decreased, and (4) extroverts and those who prefer on campus classes were impacted the most by the conversion to online [2]."

The authors first present a description of the context and magnitude of the challenge to be addressed. The paper then details the project management system developed to facilitate the workflow of the overall process and the ancillaries intended to help faculty master the skills necessary to have all courses certified by the university's deadlines.

II. Overview and Rationale

After a change in leadership in July 2021, KSU's Office of the President charged the Division of Curriculum and Academic Innovation (CAI) with ensuring that all digital course content met recently updated federal guidelines and the University System of Georgia Board of Regents policy on accessibility standards and sustained instructor interaction. Each college, in turn, was asked to submit a review procedure for all online and hybrid courses. While many colleges already possessed the requested review procedure, the lack of required review since 2018 meant RCHSS was faced with creating and implementing a new review process in a limited window of time.

A. The RCHSS Proposal

The RCHSS ODE solicited input from faculty and college administrators and created three plans for reviewing asynchronous online, synchronous or hybrid online, and *template* courses (defined as courses created in entirety by faculty designers but taught by other course facilitators). These processes were then vetted through the appropriate college faculty governance channels. The CAI charged each college with developing and implementing their plans by the start of Spring 2022, which gave the ODE team roughly five and a half months to solicit feedback, adapt the KSU CQC for hybrid and template courses, update faculty certification trainings, and develop all tools (including communication and project management tools) needed to implement the initial two-year review cycle. No new resources were provided at the university or college level to support this initiative.

B. A Larger College, a Larger Problem

The five-month implementation timeline and two-year review timeline were reasonable for most other colleges at KSU. However, RCHSS is, by far, the largest college on campus, with 425 faculty and 8,500 students enrolled in majors within the College. Additionally, RCHSS is responsible for eight of the fourteen general education standards required for institutional accreditation. As shown in Figure 1, of those eight standards, the college offers 50 different courses that can be completed to fulfill the requirements, compared to 22 courses offered by a mix of the Coles College of Business, the College of Science and Mathematics (CSM), and the College of the Arts (COTA) across the remaining six standards. In Fall 2021, RCHSS offered 35% of undergraduate course sections and 41% of graduate course sections, totaling 750 sections, in the online modality (Figure 1).

Of the eleven departments and schools in RCHSS, two have completely rejected the idea of template courses outside of use for emergency hires, and two regularly use template courses. The remaining seven departments reserve template courses for adjunct and emergency use only. This practice contrasts with other colleges at the university. For the most part, other colleges only utilize template courses or are so limited in size as to require one or two course builds to meet class section needs. Because RCHSS is focused on disciplines in the humanities and social sciences, RCHSS has dozens of sections of one course, all with unique designs, to allow faculty to teach to their strengths and provide their expertise to students. For example, many faculty teach American literature online, and yet each professor has a different specialization; one professor might be an expert on African American literature in the 19th century and build his/her/their course around those authors, while another might focus on Gothic authors in the American South, which could produce a thematically similar course with content drastically different from his/her/their colleague. When each faculty member creates his/her/their course according to best practices and capitalizes on his/her/their area of expertise, students get the best instruction and experience. This, of course, contrasts to programs taught in other colleges that are built around state/national/international accreditation standards, which often utilize a specific curriculum across all course sections. Additionally, bespoke course designs allow for continued variety in perspective and approach, avoiding some of the pitfalls of intellectual and cultural homogenization.

The ODE team had little time to prepare for this project and had no formal structures or tools in place to accommodate a project this large. The structures and tools previously used to review courses on a training basis needed to be adapted to accommodate a higher volume of course designs reviewed in a shorter amount of time.

It takes an average of three to five hours to review each course, contingent on the course content and approach. Based upon pre-pandemic numbers, the team had anticipated reviewing a total of approximately 750 courses over a two-year period. In spring 2022, we expected 135 courses to be submitted for review. 192 courses have been submitted. Projected forward, we are now anticipating reviewing around 1000 courses over the next two years using a team of two instructional designers, three faculty, and four student assistants.

The most comparable project underta”en b’ the team was a review of 115 partial courses across a seven-month period, which was tracked in an Excel spreadsheet. At the time, our team consisted of one instructional designer, two faculty, and two senior student assistants, and all were able to assist in the review process.

III. Project Management Software

A. Possible Paths Forward

Knowing the scope of the project, we began to look at how we would manage *our* end of the course review process:

1. Conducting the initial course review,
2. Sharing the feedback written during each review with the faculty designer so they could make changes,

3. Assisting the faculty designer with any technology issues they ran into during their changes,
4. Re-reviewing the course after appropriate changes had been completed,
5. Submitting the course to the institutional database managed by CAI.

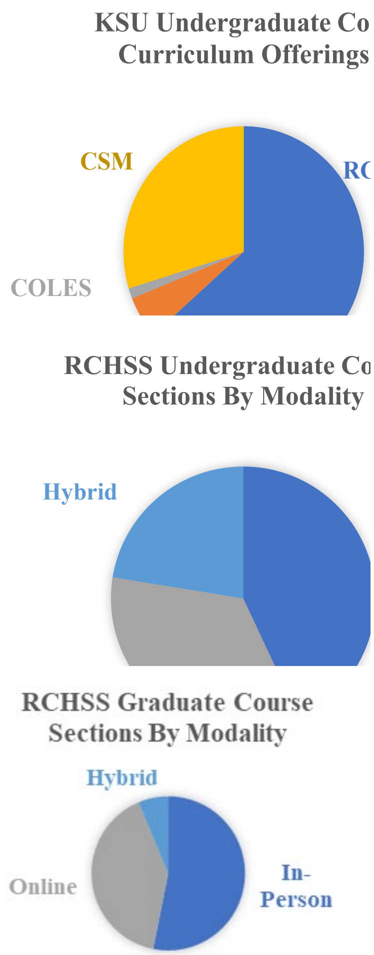


Figure 1. Three charts showing the division of courses across KSU and within the college.

We had three factors to consider when developing our workflow [5]. First, we needed to abide by the processes put forth by the departments in our college. Because our college is so large, the RCHSS Digital Education Council requested that each department be allowed to submit their own process to us. The outcomes from this request were varied: seven departments decided to defer to the ODE and have us carry out all reviews; one department decided to do the same but added additional faculty support by way of a mentoring committee to assist faculty designers with updating their courses to meet the reviews put forth by the ODE team; and three departments decided to conduct all parts of the review, with minimal support from the ODE.

Second, our undergraduate student assistants are only able to conduct the accessibility portion of the review, not to make decisions concerning content and pedagogy. One section of the KSU CQC is dedicated to digital accessibility. The review process for this section is very clinical, so it can be completed by students without years of instructional design experience.

Third, the sheer number of courses was overwhelming to a team with only five people qualified to evaluate entire courses. Department schedulers indicated a slow return to pre-pandemic scheduling practices, so we estimated that we would need to review 135 courses across the Spring 2022 semester. We were wary of more "traditional" project management methodologies (PMM) that required adherence to strict procedures, given the ever-changing nature of our institution and the natural progression of shared governance proceedings that might affect RCHSS's course review policy. At the same time, we knew that the CAI was planning for this process to be a multi-year undertaking, scaffolded by course level (e.g., 1000-level courses were reviewed in Spring 2022, 2000- and 3000-level courses would be reviewed in Fall 2022, 4000- and 5000-level courses would be reviewed in Spring 2023, and so on through upper-level undergraduate and then graduate courses). This would require us to reevaluate at the end of each semester with the little downtime that classes were not in session, 9-month faculty were not on contract, but that the University was still open.

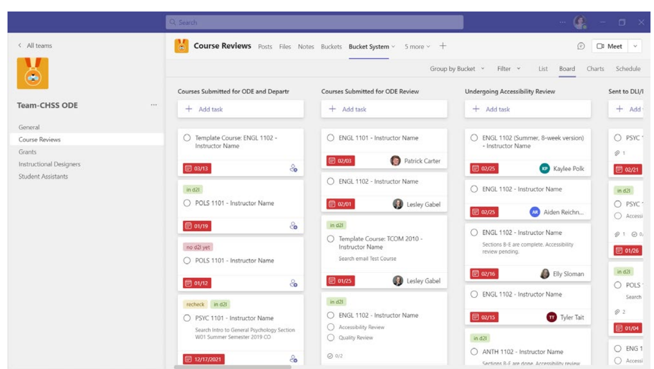


Figure 2. Screenshot of the "Bucket System" dashboard.

As previously stated, our most comparable undertaking was tracked in a spreadsheet and was conducted when our team was half the current size. Initial research about implementing PMM revealed some of our own hesitations: fear about unforeseen changes in policies that our office would have to coordinate; implementing a structured process in a naturally flexible and collaborative office; and an unfamiliarity with how to manage every document and piece of data we needed to track [6]. Further, reviews [7] indicated that they were "more adequate to be implemented in larger and more complex projects… [while] agile methodologies… are more suitable… for some smaller and less complex projects We looked more closely at agile methodology and found better matches for our needs with the SCRUM approach, which emphasizes teamwork and accountability and meshes flexibility with repetitive process [8]. The only problem was that length of our "sprints" would follow a semesterly pattern—nearly triple the length recommended by experts [8]. Essentially, we would be able to utilize most aspects of a SCRUM framework, but we would need to expand our timeline and build in some added flexibility.

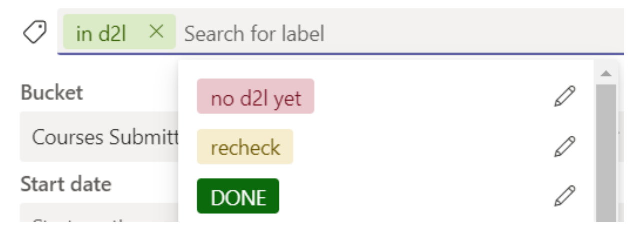


Figure 3. A screenshot of the tagging system the ODE utilizes within the bucket system.

We decided that the reevaluation/sprint process would instead follow a monitoring and evaluation (M&E) framework [9], which is typically implemented to track large-scale resources and goods for government agencies and non-governmental organizations (NGOs). In our case, neither the university nor the college could commit any additional resources, and the ODE had very little resources of our own. Instead of having the typical outputs, outcomes, and impacts of resource development, our outputs would be course reviews, outcomes would focus on bringing courses up to meet the course quality standards set by the federal government and the University, and the impact would be approved courses (ideally leading to improved retention and graduation rates, though these would not be tracked by the ODE).

After reading and watching project management software (PMS) reviews and testimonies from other higher education teams [10][11], the team met and compiled a list of the features we would need.

- Access for at least 10 people
- Reporting functions
- Ability to connect files to a task
- Ability to assign tasks
- Ability to create subtasks
- A visual marker for where the course was in the review workflow.

Our team also decided that we preferred a more visual interface, as opposed to a simple task list. A comprehensive literature shows [12] that "creative discovery processes are almost never structured and require lots of interaction with the data." Because this process was new and we knew that we would need a comprehensive understanding of our workflow to both increase efficiency and advise College and University leadership, we wanted to ensure that we had a holistic view at the ready.

Most importantly, all of this needed to be located in a no-cost or extremely low-cost tool. PMS like Slack [13], Trello [14], and ClickUp [15] had been utilized by the team before but required monthly subscriptions for large projects. Open-source PMS like OpenProject [16] and Focalboard [17] all had the key features the ODE team needed but also had annual hosting costs, usually upwards of $150USD. Products offered by ServiceNow [18], BMC [19], and IBM [20] were much more robust than needed for this one project, so they were not considered.

B. "The Bucket System"

After developing our list of wants and needs, we developed an original system we have christened as the "Bucket System." Our system was devised to be hosted in existing PMS and as an easy way to physically move items along in a workflow, while also providing more flexibility than an Excel spreadsheet. The team devised a plan to build something similar to other tools by utilizing various Microsoft applications, as the University had recently moved to Microsoft 365 [21]. This project management tool would be modeled after existing tools [22] but would be customized to our specific needs.

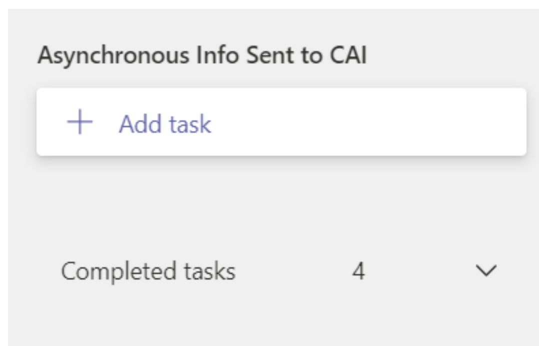


Figure 4. Screenshot showing how tasks become invisible once marked "complete."

Microsoft Planner had the tools we would need at a task-management level (file connection, task assignment, reporting functions, and subtasks) and had several layout options. It was also integrated with Microsoft Teams—which would create a localized "dashboard"—and Power Automate, Microsoft's backend automation and workflow application, which would be crucial in populating the system. Microsoft Lists had similar functions—and the bonus of custom metadata fields—but lacked the visual layout we wanted. Ultimately, the goal was to produce a Kanban-like system of workflow visualizations, as we unfortunately did not have the luxury of limiting work in progress (WIP) that was required in a true Kanban system [23]. Microsoft Planner utilizes columns called "buckets" and projects called "cards" (Figure 2). Each bucket represents a step in the workflow, and each card within a bucket represents an individual course. Additionally, we could follow a loose interpretation of SCRUM methodology with built in data visualization, which allowed us to catch up on what had been completed, by whom, and what obstacles people were facing.

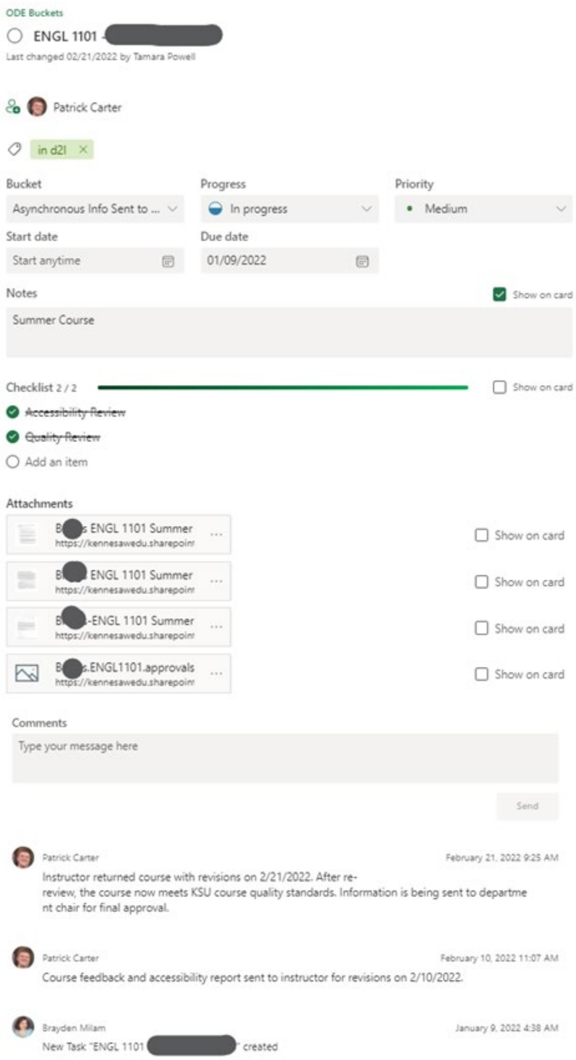


Figure 5. A course review card showing the light green "in D2L tag," a completed checklist, attached documents, and comment trail.

Within the cards, we initially utilized the tag feature to indicate whether ODE directors had access to the course in the LMS (and, thus, had the ability to add reviewers) and whether the course was undergoing a "recheck" or re-review. We ultimately added a tag to mark course reviews as finished because we initially lost courses that had been dismissed as "complete" (Figures 3 and 4).

We also use the comment section to create a documentation trail regarding when communication had been sent to faculty. The team found that using the "start date" feature did little to help prioritize courses, as Planner does not offer a "start date" option under the sorting tool. We instead pivoted to using the "due date" feature, which had the bonus of posting the date on the card. This all creates an easily accessible history of what courses have been reviewed, where they are in the process, and who has completed what (Figure 5).

After building the process in planner, the decision was made to integrate the structure with Microsoft Forms, which would allow faculty to submit "review requests." We reviewed the Certified Course Build SmartSheet developed by the CAI, which was designed to track courses certified according to quality standards and included key questions on our Form (Figure 6). Originally, we created one Form for all of RCHSS. As we tested the process, we discovered we needed individual Forms for all eleven departments and schools. This decision was made to minimize confusion, as we initially offered too many review-customization options on a single form, and to streamline the automation process. The automation process is as follows: a faculty member fills out a form, Power Automate pulls information from the Form to generate a Planner card in the correct bucket and then sends a personalized confirmation email to the faculty member (Figure 7).

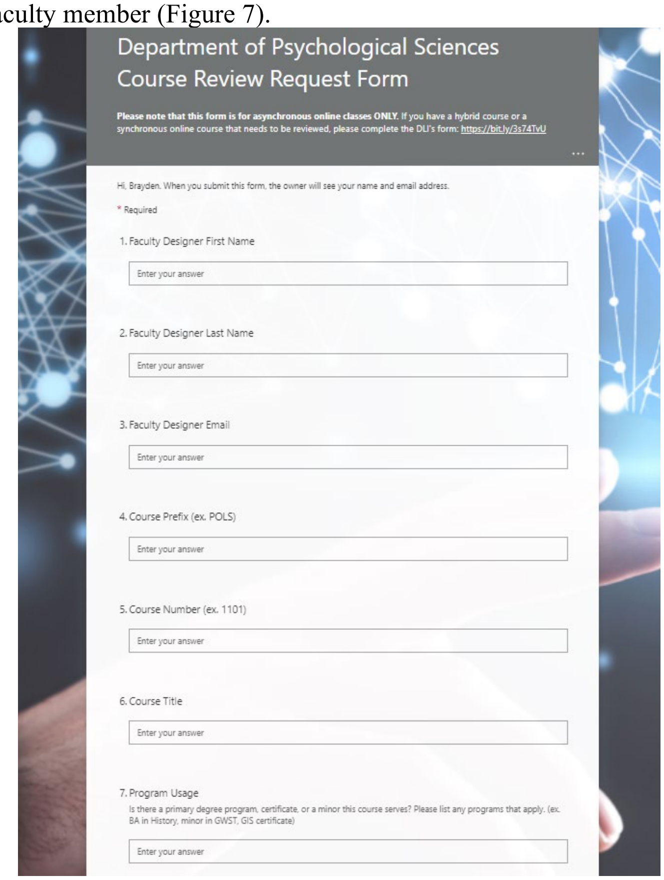


Figure 6. The first page of the Review Request Form for the Department of Psychological Science.

After the faculty member has added the ODE directors to their course shell, the formal review process begins. At the beginning of each week, the ODE Associate Director assigns courses to ODE reviewers (instructional designers and faculty) and student assistants depending on the courses available for review. Instructional designers and faculty prioritize courses taught by departments with their own review process, while student assistants complete accessibility reviews for courses that are reviewed by the ODE in full. The instructional designers and faculty then complete sections A and B of the review for the courses assigned to them, which require greater familiarity with digital accessibility standards and state and federal mandates, respectively. They then pass the course along to the department liaison/representative so the department can finish the rest of the review. Alternately, students complete section A of the review, which only requires familiarity with standards of digital accessibility, for the courses assigned to them. After the accessibility review has been completed, students pass their reviews on to the ODE instructional designers and faculty, who complete the rest of the review.

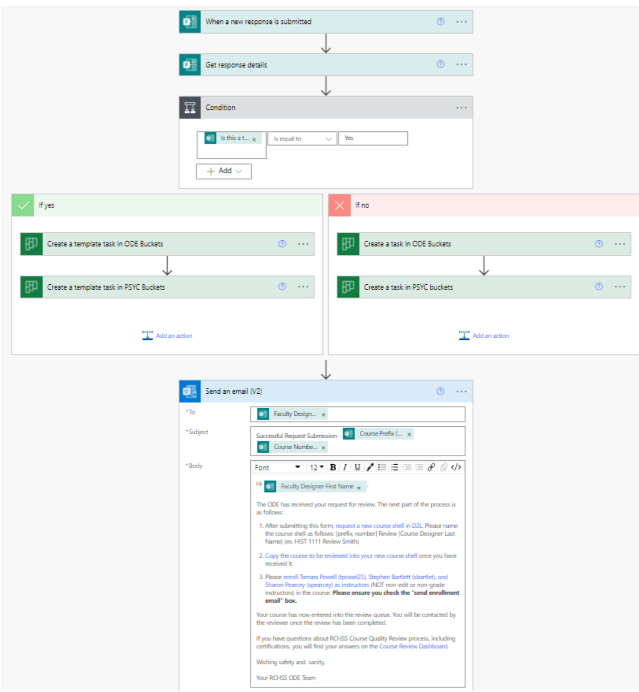


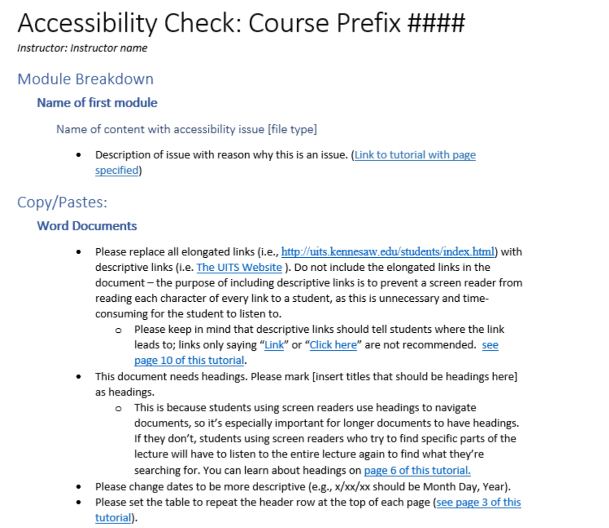
Figure 7. The automated process behind the Psychological Science Request Form, built with Power Automate.

Following the completion of the initial review, the review documents are sent to the faculty designer. The faculty designer then works to make the appropriate changes and consults with ODE team members (or their department liaisons) if they need assistance. After the appropriate changes have been made, the course undergoes a secondary review. If the course passes this review, the review documents will then be sent to the department chair for approval. If the course does not pass the review, the revision process begins again. This revision process may be completed up to two times before the course is removed from the queue. If it is removed from the queue, the chair must request reactivation and the faculty designer must resubmit the course to be reviewed without priority.

Once a course is approved by the department chair, the ODE Director submits the course to the CAI's Certified Course Design SmartSheet and notifies the program coordinator, who updates the public-facing spreadsheet maintained by the ODE. As of Spring 2022, this process is set to repeat on a five-year cycle. The initial review cycle is based on course level, so all existing courses will have completed the review process by December 2023 and been scheduled for their next re-review.

C. Supplemental Tools for the Team

The most refined and evaluated portion of our review process involved managing individual workloads. To contend with the sheer number of courses, we developed clinical language guides for our accessibility reviewers and fillable documents for our content equivalency reviewers.



Accessibility Check: Course Prefix ####

Instructor: Instructor name

Module Breakdown

Name of first module

Name of content with accessibility issue [file type]

- Description of issue with reason why this is an issue. (Link to tutorial with page specified)

Copy/Pastes:

Word Documents

- Please replace all elongated links (i.e., http://uits.kennesaw.edu/students/index.html) with descriptive links (i.e. The UITS Website). Do not include the elongated links in the document – the purpose of including descriptive links is to prevent a screen reader from reading each character of every link to a student, as this is unnecessary and time-consuming for the student to listen to.
    - Please keep in mind that descriptive links should tell students where the link leads to; links only saying "Link" or "Click here" are not recommended. see page 10 of this tutorial.
- This document needs headings. Please mark [insert titles that should be headings here] as headings.
    - This is because students using screen readers use headings to navigate documents, so it's especially important for longer documents to have headings. If they don't, students using screen readers who try to find specific parts of the lecture will have to listen to the entire lecture again to find what they're searching for. You can learn about headings on page 6 of this tutorial.
- Please change dates to be more descriptive (e.g., x/xx/xx should be Month Day, Year).
- Please set the table to repeat the header row at the top of each page (see page 3 of this tutorial).

Figure 8. An Accessibility Check Template, with verbiage available to copy and paste into communications with faculty members.

The language guides were developed first, as the first step of our end of the review process. Our student assistants were trained in reviewing digital materials for accessibility (e.g., ensuring videos had intelligible captions, documents contained heading structures, color contrast was appropriate, etc.). Ms. Milam, the staff expert on accessibility and a former ODE student assistant, worked with Sam Lee to develop a template in January 2021. By the time the reviews were well underway, she and Kaylee Polk worked to adapt the language in the template into a guide (Figure 8), so that students could copy and paste language to help faculty address the issues at hand. Each software supported by KSU (all Microsoft Office products, D2L, Adobe Acrobat, and Kaltura MediaSpace) had a section with professional, courteous language instructing the faculty designer of the issue and providing resources on how to solve it. The language was then pasted into the final review document with information about where the issue was located (the weekly or unit folder—called "modules"—along with the document name).

After developing the language guides, the content equivalency reviewers began developing fillable documents (Figure 9) for their own use. Three of these documents were created: one for asynchronous online courses, one for hybrid and synchronous online courses, and one for on-site/face-to-face courses. Each document utilized MS Word's checkbox, content controls, and grouped text features to create a mix of "locked" content and "fillable" content. This was to ensure that faculty received an accessible, standardized document with direct yet courteous language to assist them through the review process.

Section C: Pedagogy, Structure, Navigation, Course Objectives, and Module Objectives

☐ 21. The course includes measurable course objectives (course goals/learning outcomes) at the appropriate level of Bloom's Taxonomy.
☐ 22. Modules include measurable objectives that align with course objectives.
☐ 23. The course includes digital course content, assignments, and assessments that align with course and module objectives.
☐ 24. The course aligns with the stated description in the current KSU Undergraduate Catalog or Graduate Catalog course, as appropriate, including its learning objectives, module objectives, and competencies.
☐ 25. The course modality – asynchronous online (95%, 100%) – is clearly stated in the course materials. (95% asynchronous is defined as students required to come to campus once during the semester, usually to take a final exam).
☐ 26. Materials are organized in a way that creates an obvious path for learners by organizing content into sections based on weekly, chronological modules, not based on type of content.
☐ 27. The course schedule is included and makes clear expectations for each class session (including dates and, in the case of hybrid courses, modality).
☐ 28. The course has a clear and consistent structure and navigation, which is clearly stated and explained to students via navigational videos or guides within the "Start Here" module and within modules and/or checklists/task lists within modules.
☐ 29. Course content is labeled clearly from a learner's perspective.

Comments for Section C:

30. Click or tap here to enter text.

Figure 9. The KSU Course Quality Checklist converted into a fillable document in MS Word, created with Content Controls.

IV. Tools for Shareholder Involvement

Our most important shareholders in this process were faculty. To initiate a review, faculty must submit their courses for review, request new course shells for each course review needed, and add the ODE directors to the new course shells. We wanted to streamline the process as much as possible to make it as simple as possible for faculty, so as not to induce confusion or increase their workloads. The simple form and automated email accomplished most of these goals. We kept the forms as simple as possible, included tutorials on how to request, copy, and add the ODE directors to the course shells in Brightspace D2L, our learning management system; eliminated an additional step of communication by requesting faculty utilize the enrollment notification option in D2L; and added a layer of familiarity with personalization tools [24]. Ultimately, we have received positive, anecdotal feedback from faculty about how easy the process is.

A. The Course Review Dashboard

The ODE decided to anticipate the needs of faculty and create resources to aid them in the course review process. The primary resource was the "Course Review Dashboard," which functions as a one-stop-shop for all

After consulting with the RCHSS Digital Education Council, the ODE found that the new emphasis on accessibility requirements made syllabi templates more desired by faculty. Again, lockability remained the greatest difficulty, as faculty would frequently mistakenly render ADA compliant syllabi inaccessible by altering certain document features. After many hours of research, the team was able to use Microsoft Word to develop a "lockable" template, much like the fillable document mentioned earlier, that met accreditation and accessibility standards but remained customizable enough to allow for academic freedom. The team also worked with the First-Year Composition (FYC) program to develop two syllabi specifically for use in ENGL 1101: Composition I and ENGL 1102: Composition II, two courses all students are required to complete prior to graduation. These templates were created with a variety of Developer Tools, including content controls and "lockable" groups, and general accessibility tools, like Styles and list/table structures.

V. Conclusion

With varying requirements across the eleven schools and colleges that comprise the Radow College, the unexpectedly high number of course submissions, and the need to validate review practices and feedback information across reviewers, much fine-tuning of the Bucket System is still in progress. The overall system, however, is proving to be robust and flexible while providing consistent tracking and information capture. things related to course reviews, including tutorials, policies and checklist criteria, review request forms, and contacts. This project was designed to be far-reaching and creative. We knew the website needed to be easy to navigate and user friendly, but we also knew how expansive the topics and resources needed to be. We ultimately built eight pages hosted on the site and built an additional site specifically for course reviewers within the departments. We also completely redesigned our tutorial library, which hosted 115 simple technology and software tutorials, for a cleaner interface with a more visual navigation structure (Figure 9).

A. Syllabus Templates

With a great deal of the accessibility and quality checklist requirements focused on course syllabi, a significant need for an accessible syllabus template arose. The need to include an ever-increasing number of policies and links to student success tools had placed syllabi templates at the forefront of academic concerns of the university for well over a decade. The updated federal guidelines and University System of Georgia Board of Regents policy only increased the urgency in addressing this need. The greatest difficulty was in "locking" the template so that required items could not be altered or deleted while keeping the remainder of the document open for editing by the instructor. An early attempt at an electronic solution to the syllabus template issue ended in a costly failed attempt to partner with a professional technical company. Further attempts at a lockable form were abandoned.

Because the Bucket System is built on components of Microsoft 365 and accessed through familiar interfaces, the time required to learn to use the system effectively is far shorter than would be the case using other software tools to manage the same project. Bringing new reviewers or content specialists online is also simplified, with no need to obtain additional software licenses or install additional software. It also bridges seamlessly across units within the College and across the University as needed.

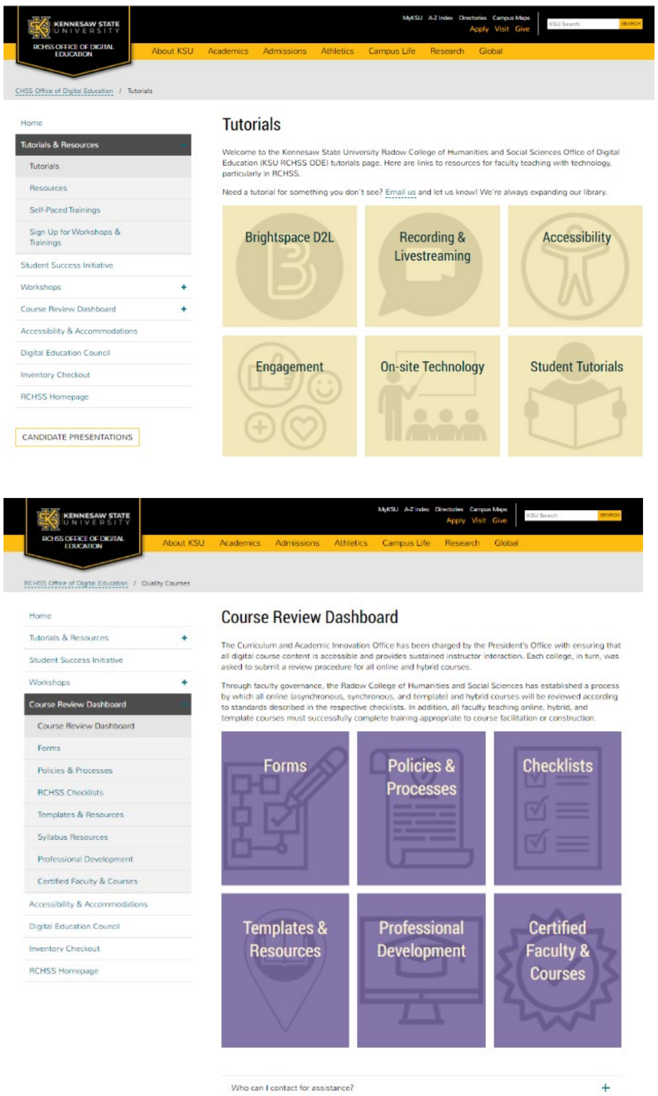


Figure 10. Two screenshots showing the Tutorial and Course Review Dashboard websites.

The Bucket System has been adapted across multiple course review efforts and departmental processes. Each department with their own review process has a system customized to their needs, including one built for the university-level equivalent of ODE. For example, the faculty member designated as the Hybrid Specialist in the ODE is piloting an adaptation of the Bucket System to track and manage a grant-funded Open Educational Resource (OER) project. Currently, this version of the Bucket System functions more as a task board but will be adapted at the conclusion of the OER's pilot semester (Spring 2023) into a ticketing system for instructors to provide feedback and request support from the original grant team on any issues that may present. Using the Bucket System in this way will allow for data collection, as the team plans to build in survey software and track open feedback and support tickets. Additionally, the Bucket System is being used as an event management tool by RCHSS's Office of the Dean. This version of the Bucket System more closely mirrors the original version. Each event hosted by the College has a dashboard, is used for knowledge management, task completion, budget tracking, contact tracking, and collects data that feeds into a more comprehensive data visualization page. The Office of the Dean plans to duplicate the system at the end of the fiscal year to use for future years.

REFERENCES

[1] B. Milam, T. Powell, J. R. Newell, S. Bartlett, L. Gabel, K. Polk, E. Sloman, A. Reichner, "Using the Tools at Hand: Creative Online Course Quality Management," In *The Fifteenth International Conference on Mobile, Hybrid, and On-Line Learning* (June 2022): 42-47.

[2] A. Whiting, W. Ritz, and J. S. Hain, "Exploring the Effects of Students from Converting On-Campus Classes to Online due to the COVID-19 Pandemic," *Journal for Advancement of Marketing Education* 29, no. 1 (2021): 13-24.

[3] T. H. Reisenwitz and J. G. Fowler, "Transitioning from Face-To-Face to Online Classes During a Pandemic: Factors That May Affect Student Satisfaction of the Administration and Instructors," *Marketing Education Review* 31, no. 3 (2021): 199-208.

[4] G. Bulman and R. W. Fairlie, "The Impact of COVID-19 on Community College Enrollment and Student Success: Evidence from California Administrative Data," *National Bureau of Economic Research Working Paper Series* (2022).

[5] Project Management Institute, *A Guide to the Project Management Body of Knowledge (PMBOK(R) Guide–Sixth Edition*, Project Management Institute, 2017.

[6] E. Ozmen, *Project management methodology (PMM): how can PMM serve organisations today?*, PMI® Global Congress 2013—EMEA, Istanbul, Turkey. Newtown Square, PA: Project Management Institute, 2013.

[7] P. Jovanović and I. Berić, "Analysis of the Available Project Management Methodologies," *Management: Journal of Sustainable Business and Management Solutions in Emerging Economies* 23, no. 3 (2018): 3-13.

[8] M. Sliger, "Agile project management with Scrum," PMI® Global Congress 2011—Dallas, TX (2011).

[9] *Handbook on Planning, Monitoring and Evaluating for Development Results*, United Nations Development Programme, 2009.

[10] J. D. Frame, *Managing Projects in Organizations: How to Make the Best Use of Time, Techniques, and People* (3rd Edition), Jossey-Bass, 2003.

[11] R. Michalak and M. D. T. Rysavy, "Managing Remote Projects Effectively with an Action Dashboard," *Journal of Library Administration*, 60(7), pp. 800–811, 2020.

[12] V. González and A. Kobsa, "Benefits of information visualization systems for administrative data analysts," In

Proceedings on Seventh International Conference on Information Visualization IV (2003): 331-336.

[13] Slack, https://slack.com. Date accessed: 01 June 2023.

[14] Trello, https://trello.com/en-US. Date accessed: 01 June 2023.

[15] ClickUp, https://clickup.com. Date accessed: 01 June 2023.

[16] OpenProject, https://www.openprojec.org. Date accessed: 01 June 2023.

[17] Focalboard, https://www.focalboard.com. Date accessed: 01 June 2023.

[18] ServiceNow, https://www.servicenow.com/products/project-portfolio-management.html. Date accessed: 01 June 2023.

[19] BMC, https://www3.bmcgroup.com/solutions/virtual-data-room/difference/project-management/. Date accessed: 01 June 2023.

[20] IBM, https://www-50.ibm.com/partnerworld/gsd/solutiondetails.do?&solutions=50480. Date accessed: 01 June 2023.

[21] Microsoft 365, https://www.microsoft.com/en-us/microsoft-365. Date accessed: 01 June 2023.

[22] S. L. Catto, & E. A Maccari, "Innovation Projects Management: A Systematic Literature Review." *Brazilian Journal of Management / Revista de Administração Da UFSM,* 14 (4), pp. 848–863, 2021.

[23] N. Damij and T. Damij, "An Approach to Optimizing Kanban Board Workflow and Shortening the Project Management Plan," *IEEE Transactions on Engineering Management*, forthcoming, doi: 10.1109/TEM.2021.3120984, 2021.

[24] R. V. Waters & S. A. Ahmed, "Beyond the Spreadsheets: Quality Project Management," *Performance Improvement*, *59* (10), pp. 16–29, 2020.

Acoustic Emission Sensing of Materials and Structures at Mechanical Strengths

Irinela Chilibon
Optospintronics Department
National Institute of Research and Development for Optoelectronics, INOE-2000
Magurele, Romania
e-mail: qilib@yahoo.com

***Abstract*—Acoustic Emission (AE) sensing is used in the field for the testing of metal, composite and structures, as non-destructive techniques. The AE technique allows determining the size of the cracks, damages, fractures and failures into materials. Possible causes of the internal-structure changes are crack initiation and growth, crack opening and closure, dislocation movement, twinning, and phase transformation in monolithic materials and fibre breakage and fibre-matrix debonding in composites. AE technology involves the use of ultrasonic transducers to listen for the sounds of failure occurring in materials and structures. Monitoring the stability of construction structures, by AE techniques get involve detecting the onset and evolution of stress-induced cracks and preventing the structural failures.**

Keywords-Acoustic Emission (AE); Acoustic Emission technique (AET); Non-Destructive Testing (NDT); Structural Health Monitoring (SHM); Polyvinylidene fluoride (PVDF).

I. INTRODUCTION

Acoustic Emission (AE) techniques have attracted attention to the diagnostic applications, material testing and study of deformation, fracture and corrosion, because they give an immediate indication of the response and behavior of materials under stresses, intimately connected with strength, damage, fracture and failure. AE technology involves the use of ultrasonic sensors (20 kHz - 1 MHz) to listen to the sounds of failure occurring in materials and structures [1], [2].

The roughest localization method is guessing the source origin using the "first hit" technique. The sensor which detects an AE first defines a radius or a half sphere, respectively, in which the signal originated. For instance, this can be done for some cases in combination with other techniques or knowledge to "localize" the source of failures.

Fibre breakage, matrix cracking, and delaminating are three mechanisms that can produce AE signals when stress is applied to the material or structure. Most of the sources of AEs are damage-related to the detection of these emissions.

The continuous research evolution in this field may be useful for a targeted diagnosis of the corrosion-induced damage severity and the recognition of corrosion sources through the AE online inspection and monitoring [3]. Also, integrated with additional information, such as metallography, AE technique can provide a valid tool for identifying specific features related to crack initiation and propagation mechanisms.

AE technology uses ultrasonic transducers in the frequency range (20 kHz - 1 MHz) to detect sounds emitted by defects that occur in materials and structures, which are subjected to mechanical pressure or temperature variations.

Determining the degree of degradation of mechanical properties and the residual life of metal structures under complex dynamic deformation demands has various applications related to bending, static and dynamic tensile loads and defect initiation processes [4].

In case of metal structures, fatigue failure occurs due to cyclic stress from operating conditions. The main mechanisms of failure occur from mechanical fatigue or thermal fatigue, such as: mechanical fatigue failure is due to cyclic stresses and thermal fatigue failure is due to cyclic temperature changes. The tipping point for failure is when the material fails at loads lower than the yield strength of the material. The acoustic emission as a monitoring tool has capabilities to detect fatigue crack initiation and propagation in mooring chains [5].

Structural Health Monitoring (SHM) allows the early detection of potential damages resulting from the natural deterioration of structural materials and the optimization of decisions over maintenance, repair, and reconstruction of the bridge asset [6].

The structure of this paper is organized as follow: Section II describes the Acoustic Emission sensors. Section III describes the Acoustic emission monitoring methods. The conclusions and acknowledgment close the article.

II. ACOUSTIC EMISSION SENSORS

The future market for electronic devices will focus on miniaturized flexible electronic devices with low power consumption. The development of piezoelectric films with excellent piezoelectric responses and low coercive voltages would therefore be advantageous [7]. Acoustic emission sensors are usual piezoelectric receiver transducer, having as active elements discs made by piezoceramic lead titanate zirconate (PZT), lead titanate (PT), barium titanate (BP) [8], PVDF materials, copolymers, and composites. Piezoelectric materials are among the most important new materials in this century because of their excellent performances.

Polyvinylidene fluoride (PVDF), also known as polyvinylidene difluoride and PVF2, is part of the fluoropolymers family, a group of specialized, versatile polymeric materials with distinct properties that result from the strong bond between their carbon atoms and fluorine atoms and the fluorine shielding of the carbon backbone. PVDF is a polymer with pyroelectric and piezoelectric properties and is used in the manufacturing of diverse high-purity, high-strength, and high-chemical-resistance products for applications in electrical, electronic, biomedical, construction, etc.; PVDF has a similar structure to poly (tetrafluoro-ethylene) PTFE, except that the hydrogen atoms are only replaced by fluorine on every alternate carbon [9].

Piezoceramic-based AE sensors measure the acoustic emissions and are sensitive to the flexural wave motion (vertical motion to the surface). The acoustic emissions due to impact hammering, fatal failure, large disturbances, and glass fracture had a high-amplitude vertical component of wave motion, detected by AE sensors. Attempts were made to capture the in-plane component of the wave motion using fiber-Bragg grating sensors. Coupled piezoelectric film strain sensors, monolithic piezoceramic patches were used to measure the acoustic waves [10]. The schematic of the cross-section of a typical commercial AE sensor is shown in Figure 2. It has several components inside a steel housing. It has a backing plate, PZT material, electrodes, and damping material inside the housing. The top electrode of the PZT material is connected to the center conductor of the connector and the bottom electrode is grounded to the housing. A bonding agent is used to connect the AE sensor to the host structure [10]. Piezoelectric elements inside the sensor convert this pressure into current, which then converted to a voltage signal**.**

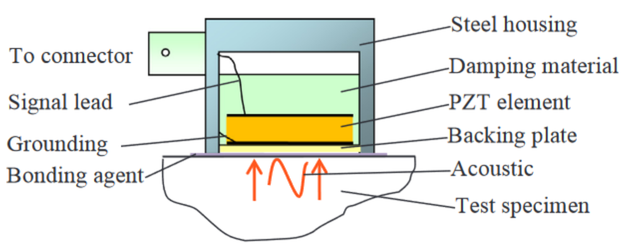


Figure 1. Cross-section of a typical commercial AE sensor which measures out-of-plane wave motion [10].

Figure 2. Image of an AE sensor

Figure 1 presents a cross-section of a typical commercial AE typical commercial AE sensor which measures out-of-plane wave motion [10], and Figure 2 is the image of a usual AE sensor.

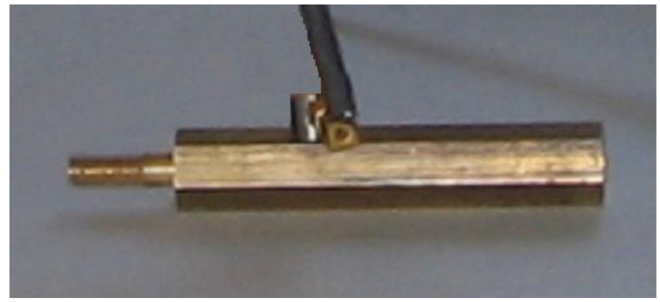


Figure 3. AE sensor placed on a metal bar

As an application, a piezoceramic AE sensor was fixed on the surface of a metal bar by means of a silicone Vaseline-type coupling material (Figure 3), which ensures the maximum coupled transmission coefficient of the acoustic elastic waves at the piezoelectric element of the sensor.

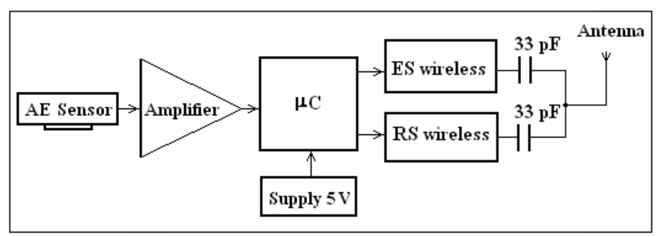


Figure 4. Block diagram of a module with acoustic emission sensor (SEA) [11].

More sophisticated devices were constructed as intelligent modules with AE sensor (SEA), used to pick up the acoustic emission signals. For example, an intelligent module with AE sensor (MSEA) is composed by: Acoustic emission sensor (SEA), Wireless transmitter (ES), Wireless receiver (RS), Stabilized voltage source with battery, Amplifier, PIC 18F452 microcontroller (µC) (Figure 4) [11].

III. ACOUSTIC EMISSION MONITORING METHODS

A. *Non-Destructive Evaluation Techniques*

For historical buildings, Non-Destructive Evaluation (NDE) techniques are used for several purposes: (1) detecting hidden structural elements, such as floor structures, arches and piers; (2) determining masonry characteristics, mapping the heterogeneity of the materials used in the walls (e.g., use of different bricks during the life of a building); (3) evaluating the extent of the mechanical damage in cracked structures; (4) detecting voids and flaws; (5) determining moisture content and rising by capillary action; (6) detecting surface decay phenomena; and (7) evaluating the mechanical and physical properties of mortar and brick, or stone [12]. The choice of a technique for controlling and monitoring reinforced concrete or masonry structures is strictly correlated with the kind of structure to be analyzed and the data to be extracted.

A general approach for AE data analysis, which will be followed here is stepwise: (a) evaluating AE activity, e.g., the rate or cumulative number of selected AE hits or located events and noting their correlation with time or applied load; (b) evaluating AE intensity, e.g., the burst signal peak amplitude, burst signal energy, or continuous signal parameters and their behavior with load; (c) AE source location, if more than one AE sensor has been used, e.g., spatial or spatial-temporal clustering of AE event sources; and finally (d) looking for indications of different damage mechanisms, e.g., from AE intensity or waveform analysis [13].

Delamination behaviour of composites is a standard reference for all those researching laminated composites and using them in such diverse applications as microelectronics, aerospace, marine, automotive and civil engineering. In AE small amounts of elastic energy are released within a structure by a mechanical mechanism. Such energy release may arise from a variety of mechanisms, such as crack tip advance, plastic deformation, or other mechanical behaviour like friction and rubbing. This energy radiates from its point of release, known as the source, in all directions, propagating as an elastic wave [14]. The nature of this technique means that a source mechanism must be active (i.e., damage must be growing) in order for it to be detected, making it ideal for in-service structural health monitoring (SHM) and non-destructive testing (NDT).

AE monitoring appears to be a promising technique that can be used for bridge inspection to quantify the condition of steel-reinforced concrete, where corrosion is occurring, and where repair is needed.

Material study is another field of acoustic emission application. Particularly, acoustic emission is used for studies of:

- Environmental cracking including stress corrosion cracking, hydrogen embrittlement.
- Fatigue and creep crack growth.
- Material properties including material ductility or embrittlement, inclusions content.
- Plastic deformation development.
- Phase transformation, and many other.

B. *Acoustic Emission Testing Methods*

Acoustic emission is a very versatile, non-invasive way to gather information about a material or structure. Acoustic Emission Testing (AET) can be applied to inspect and monitor pipelines, pressure vessels, storage tanks, bridges, aircraft, and bucket trucks, and a variety of composite and ceramic components. It is also used in process control applications such as monitoring welding processes. Acoustic Emission Testing is a non-destructive testing method that "listens" for transient elastic-waves generated due to a rapid release of strain energy caused by a structural alteration in a solid material.

Piping inspection is another common application, and Acoustic Emission is used efficiently and fast for detection of cracks, corrosion damage and leaks. There are multiple advantages of the method in case of piping inspection.

For example, in case of buried or insulated pipelines (Figure 5), there is no need to open the entire surface of the pipe but just a small opening for installation of sensors, while a distance between sensors can be from few meters to 100 meters. Acoustic emission testing is applied also for inspection of high pressure and temperature piping systems during their normal operation.

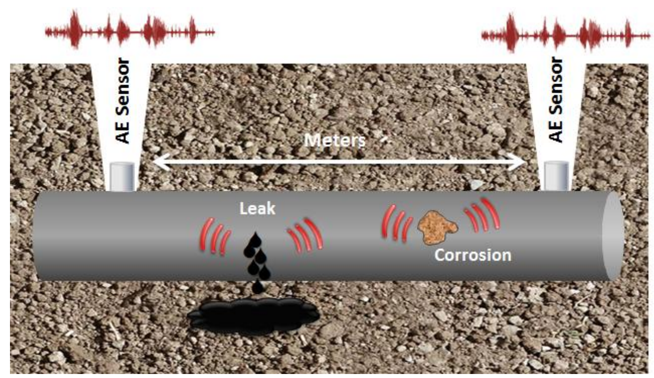


Figure 5. AE sources related to corrosion development and a leak in an underground pipeline.

As example, inspections of concrete and reinforced concrete bridges are applications where acoustic emission is used for detection of cracks, other concrete flaws, rebar corrosion, failure of cables and other. The method allows an overall inspection of a structure and long-term condition monitoring when it is necessary providing important information for bridge maintenance.

An elastic wave is a combination of longitudinal, transverse, and reflected waves, with a broadband frequency range from kHz to MHz. The AE is a phenomenon in which transient elastic waves are generated by rapid release waves, and a monitoring system requires a source and crack propagation or a tendon failure (Figure 6) [6].

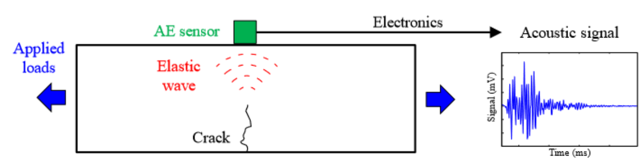


Figure 6. Working principle of an AE monitoring system [6].

The evaluation by AE monitoring of the complex mechanisms acting during stress corrosion crack (SCC) is still far from a clear, well accepted interpretation [3]. SCC is one of the most critical corrosion types and can also cause premature failures of structural components and should not be neglected in damage risk managements.

Monitoring a structure by means of the AE technique makes it possible to detect the onset and evolution of stress-induced cracks. Crack opening, in fact, is accompanied by the emission of elastic waves that propagate within the bulk of the material (Figure 7) [3].

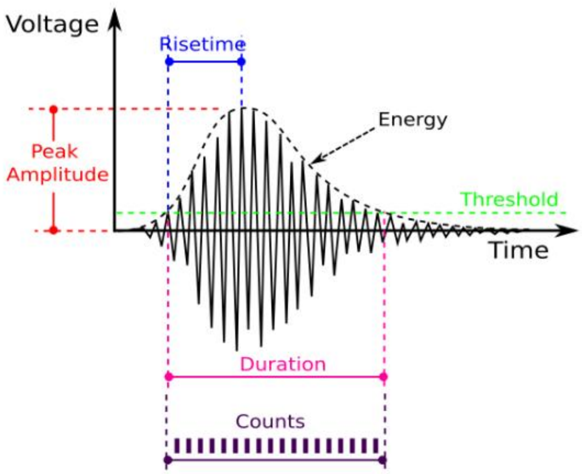


Figure 7. Schematic of an Acoustic Emission event and related parameters [3].

These waves can be captured and recorded by transducers applied to the surface of the structural elements [12]. The signal identified by the transducer (Figure 8) is preamplified and transformed into electric voltage; it is then filtered to eliminate unwanted frequencies, such as the vibrations caused by the mechanical instrumentation, which are generally lower than 100 kHz. In the Ring-Down Counting method, the signal is analyzed by a threshold measuring unit, which counts the oscillations exceeding a certain voltage value.

Non-destructive techniques were not accepted for long time for the testing bridges and other components of the infrastructures, because of inability of AE technique to determine the size of cracks [15].

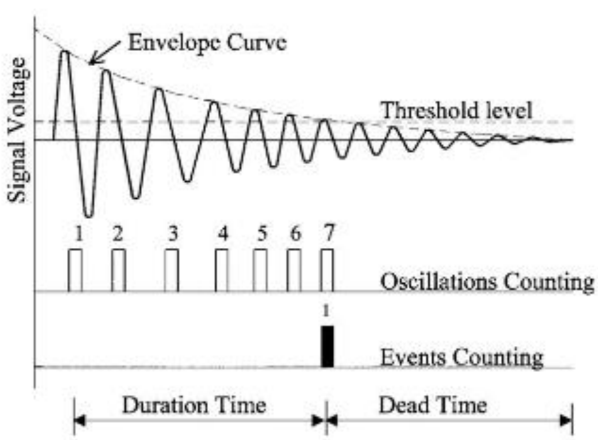


Figure 8. Counting methods in AE technique [12].

For example, one of the major issues in offshore equipment design is preventing the accumulation of fatigue damage over a long period of time.

A full-scale fatigue test rig and the monitoring setup were arranged to perform the AE measurements [5]. The chain failure most likely occurs at the point of the intrados (the lower or inner curve of an arch) (KT point) and crown positions, due to higher localized stresses in these areas (Figure 9). Four sensors were used on links and a water-based couplant was used to facilitate the transmission of the sound signal between the transducer and the link's surface. Also, each sensor was equipped with an integrated 34 dB pre-amplification.

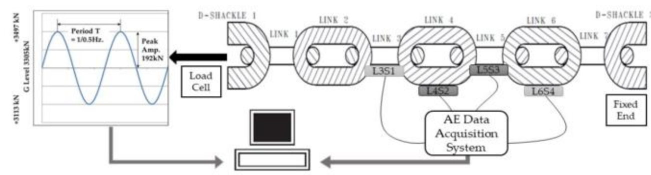


Figure 9. Test rig illustration: the chain is fixed at one end (right) and the loading (strain) is applied at the other end (left) [5].

After careful evaluation of the different AE signal features and all possible correlations, it appears that the frequency content of the AE signals is the most promising parameter. An increase of the average frequency is observed with the growth of the crack in the chain [5].

AE sensor manufacturers typically provide a sensitivity curve based on face-to-face calibration. This calibration procedure has been treated as proprietary information and described only inadequately. Calibration curves are usually in reference to the reference level of 1 V/μbar, but this reference remains undefined [16].

AE technique (AET) has found applications in monitoring the health of aerospace structures because sensors can be attached in easily accessed areas that are remotely located from damage prone sites. AET has been used in laboratory structural tests, as well as in flight test applications.

By signal graphic representation analysis in time and frequency can be determined the attenuation coefficient of the pulse into material. Figure 10 shows an aluminum pipe structure subjected to the mechanical stretching after its rupture [17].

Figure 10. Broken aluminium pipe at maximum mechanical stretch [17].

Figure 11 presents prevailed AE signals by the sensor in the breaking moment of aluminium pipe at the maximum stretch. Therefore, one can detect these emissions and predict the moment of material failure, and causing damage to the overall structure.

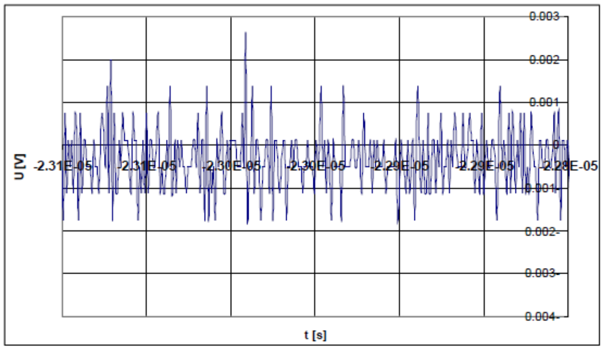


Figure 11. AE time spectrum for aluminium pipe (ϕ20x150 mm, and thickness 1 mm) subjected to the mechanical stretching [17].

The use of the phenomenon of acoustic emission during deformations of tension and bending makes it possible to predict the onset of critically dangerous states of loss of working capacity of metal structures.

The Acoustic Emission Technique (AET) is one of the non-destructive methods usually implemented to investigate the damage onset, damage evolution, and damage location in civil engineering structure [18].

Investigating the performance of acoustic emission (AE) technique on material deterioration of masonry is subjected to uniaxial monotonic and cyclic tensile loadings [19]. The basic acoustic emission (AE) counting, considering the cumulative or averaged number of AE hits or AE energy, proves the capacity of AET for damage assessment in masonry. Auto Sensor Test (AST) is carried out preliminarily to ensure no variation of sensor sensitivity.

In the case of localization AE events at a minimum error, an experimental set-up with four AE sensors, formed into a three-dimensional space is placed at the four corners of specimen, near the expected initial crack, each sensor placed about 2 cm from the stone edge (Figure 12) [19] .

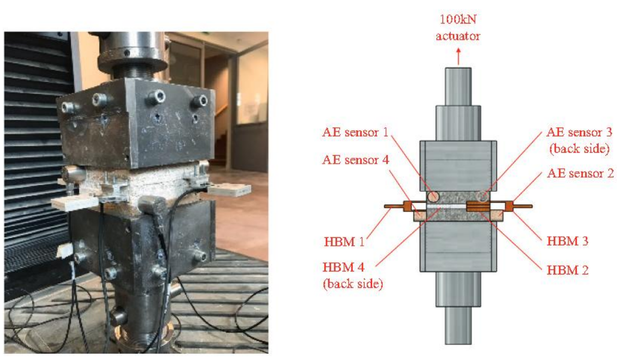


Figure 12. Experimental setup and instrumentations [19].

Experimental tests reveal that the signal durations represent a dominant parameter which can distinguish the signal from micro crack to macro crack, based on the correlation between the evolution of AE energy and fracture energy. Also, the scale effect should be considered, according to [20], as the mechanical analysis of masonry specimen under tensile loading is concentrated in the mesoscale, while the emission is received in a micro scale, more fundamental research is needed to uncover this relation at microscale.

Average Frequency (AF) is defined as the ratio between the number of counts and duration of an AE waveform. It basically determines the number of threshold crossings per unit time of an AE waveform [21].

$$\text{Average Frequency (AF) [in kHz]} = \text{Counts/Duration} \qquad (1)$$

where "Counts" indicates the number of times the signal amplitude exceeds the fixed threshold over the entire duration of the AE waveform, and *RA value* of an AE waveform is defined as the ratio between the rise time and amplitude:

$$\text{RA [in } \mu s/V] = \text{Rise time/Amplitude} \qquad (2)$$

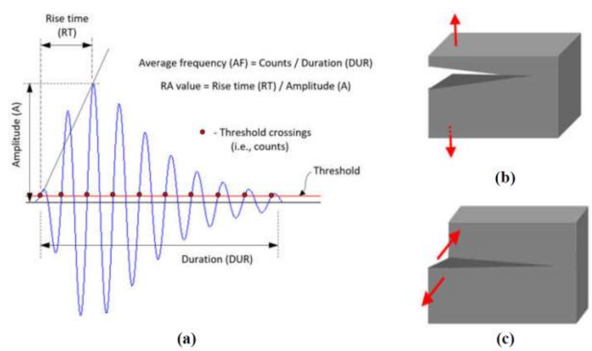


Figure 13. Schematic representation of (**a**) an AE signal and some important AE parameters, (**b**) tensile crack, and (**c**) shear crack [21].

AE waveforms generated due to tensile cracks (Figure 13) have shorter rise time, tensile-type cracks usually generate AE signals with lower *RA values* and higher AF, and in the case of shear-type cracks, the AE waveforms are longer, *RA values* are relatively higher, and AF is lower.

Various features, such as, amplitude, rise time, counts, duration, energy, number of hits, etc. [21] of the acoustic emission signals can be analyse and monitored the progressive damage of the samples under the four-point bending test.

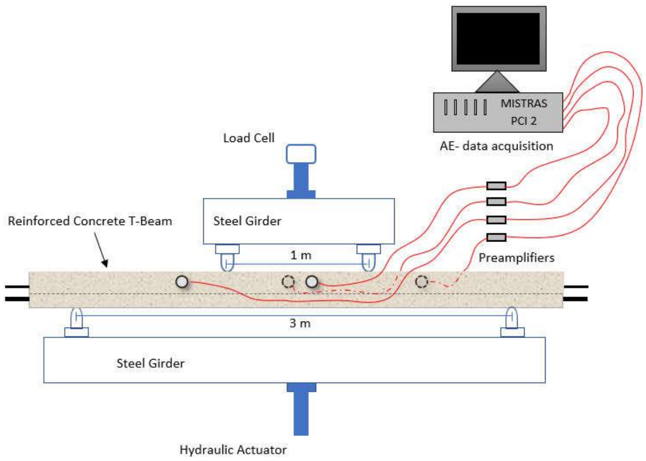


Figure 14. Reinforced concrete T-beam subjected to four-point bending and acoustic monitoring [21].

Much more, AE activity evolves as function of the applied load, where amplitudes of the involved AE hits become higher in the case of a major crack, and the cumulative number of hits can be considered as a global parameter, which can be sensitive to the mechanical test conditions for instance the load rate. The evolution of the cumulative hits per unit time is constant for each mechanical test and the evolution of the cumulative hits per unit time is proportional to the loading rate, where the highest value is for the loading rate corresponding to 4 mm/s, which creates the shortest nonlinear zone.

In order to classify the AE data, machine learning methods type supervised and unsupervised are employed.

Guo et al. present experimental quasi-static splitting tensile tests corresponding to a strain rate of $\sim 2\times10{-}6\ s^{-1}$ using a Shimadzu universal testing machine. Each cylindrical specimen is placed on the lower loading platen with its axis parallel to the platen, and the upper loading platen was set to move downwards at a speed of 0.2 mm/min to compress the specimen (Figure 15) [22].

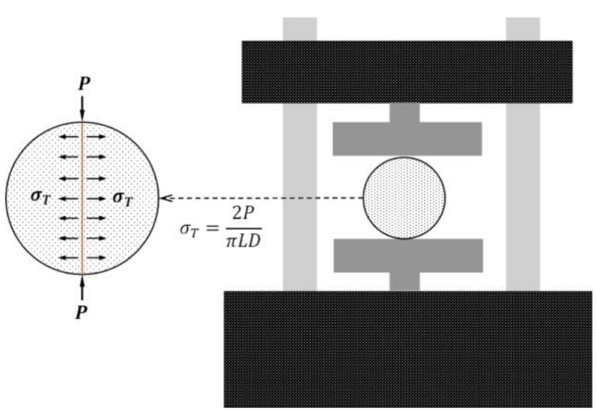


Figure 15. Experimental arrangement for quasi-static splitting tensile tests [22].

It is useful to verify, for the length scale studied, whether a homogeneous-linear-elastic analysis yields reasonable estimates of stress and deformation. Using elastic analysis one can determine material tensile stresses ($\mathcal{E}$) derived at 0^0, 45^0 and 90^0 [22], and the theoretical values are compared with those measured by the strain gauges (Figure 16).

Concrete is heterogeneous at the meso-scale, comprising a mortar matrix with randomly distributed coarse aggregate particles, and is able to display a small degree of plasticity.

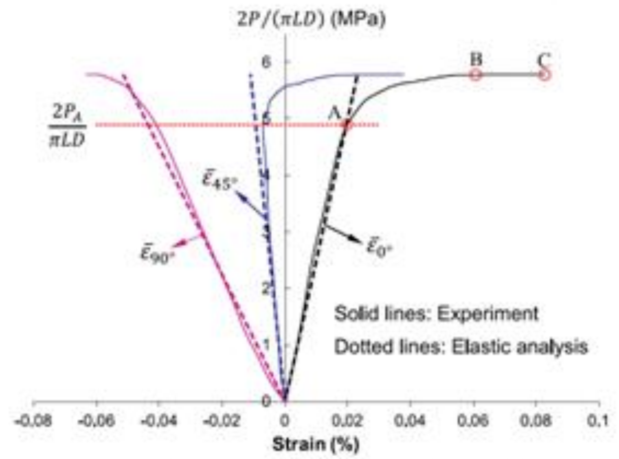


Figure 16. Variation of tensile stress at specimen centre ($2P/(\pi LD)$) with average strain in three directions (0°, 45° and 90°), for a splitting test on a C80 specimen – comparison of experimental results with theoretical predictions based on elastic analysis [22].

A monitoring system can analyse parameters like count, hit, event, rise time, duration, peak amplitude, energy, RMS (root mean square), voltage, frequency spectrum, arrival – time difference, etc. Considering AE measurement in concrete, one can use the standard ISO 16836 / 2019[23]. A measuring system with a digital-signal processor includes the following main elements: (S) - AE sensor, (Pa) - pre-amplifier, (Ma) - main amplifier and (Bf) - band-pass filter.

The AE localization algorithms are based on the distance travelled by the emitted waves. Their value can be achieved by multiplying arrival time to the propagation speed in the specific medium. The speed of the P-wave in a structure is closely related to the travel path.

Localization of AE sources in different structures can be done by using one or more AE sensors and representation in 1D linear, 2D planar and 3D localization technique (Figure 17) [24].

Single-sensor approaches are based on modal acoustic emission (MAE), which requires identification of the arrival time of extensional and flexural wave modes.

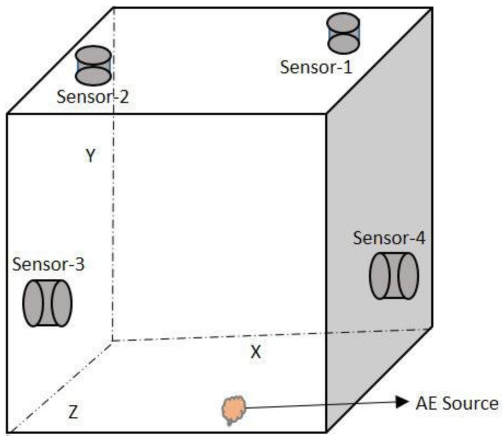


Figure 17. Localization of AE sources in 3-D structures [24].

Since the velocities for both modes are different, it is possible to calculate the distance traveled by the propagation of the wave modes if the respective velocities are known. The implementation of MAE has several associated problems, such as wave reflections and the separation of multiple AE events [24].

The attenuation of the extension wave is measured to appreciate its applicability in the long-range localization of sources. Time reversal technique and the artificial neural network are used in the literature for the AE localization problem. The convolutional neural network is more stable in allowing more information rich inputs in the frequency domain.

Triangulation method for AE source localization relies on the identification of precise arrival times and the knowledge of an appropriate propagation velocity.

The experimental works can determine practical information concerning the AE response of different materials and structures at mechanical strength and predict their behaviour and maximum strengths.

As novelty, monitoring the stability of important construction structures, by AE technique makes possible to detect the onset and evolution of stress-induced cracks and prevent the structural failures. Also, AETs investigate the damage onset, damage evolution, and damage location in civil engineering structure, material deterioration of masonry.

IV. Conclusion

The acoustic emission techniques are well applied for the identification, characterization, and localization of deformations in civil engineering structures. Numerous localization techniques, i. e. Structural Health Monitoring, Modal Acoustic Emission, Neural Networks, Beamforming, and Triangulation methods with or without prior knowledge of wave velocity can detect into material structure the fatigue cracks, delaminations, corrosions, etc.

Monitoring structures by AE techniques could detect the onset and evolution of stress-induced cracks into materials (metals, non-metals, composites) and the crack opening is accompanied by the acoustic emission of elastic waves that propagate within the bulk of the materials and structures. These waves (AE) can be captured and recorded by transducers, type AE sensors, applied to the surface of the structural elements and provide the material behavior at mechanical stresses, fatigue and vibrations.

Crack initiation into material is determined by the appearance of the AE signal at low stretch stress levels. After the crack initiated, the AE signals around the zero stress could be the cause by crack-face grinding when the cracks are closed. By experimental works one can determine practical information concerning the strength of different type of materials, their mechanical limits at different stretch values. More, it can obtain practical information about AE of monitories complex construction structures, such as: bridges, containers fulfilled with liquids, etc., in order to prevent their possible breaking due to hostile environmental (shocks, long time vibrations, bending, temperature differences, etc.).

The internal stress redistribution of the materials caused by the changes in the internal structure can be predicted by the monitoring the acoustic emissions of the stress waves generated by the structures.

Possible causes of the internal-structure changes are crack initiation and growth, crack opening and closure, dislocation movement, twinning, and phase transformation in monolithic materials and fiber breakage. A number of challenges related to accurate localization and classification of sources still remain, which must be addressed in order to exploit the full potential of the AE technique.

Acknowledgment

This work was supported by the Romanian Ministry of Research, Innovation and Digitalization (MRID) for:1) Core Program PN 18N/2019 and PN 23 05 No. 11N /2023, and 2) Program I – Development of the National R & D System, Subprogram 1.2 Institutional Performance-Projects for Excellence Financing in RDI, contr.18PFE/2021.

References

[1] I. Chilibon, "Acoustic Emission Sensing of Materials and Structures", The Thirteenth International Conference on Sensor Device Technologies and Applications (SENSORDEVICES 2022) IARIA, Oct. 2022, pp. 32-37, ISSN: 2308-3514, ISBN: 978-1-68558-006-3.

[2] D. H. Kim, W. K. Lee, and S. W. Kim, "Analysis of Acoustic Emission Signal for the Detection of Defective Manufactures in Press Process," World Academy of Science, Engineering and Technology, vol. 53, pp. 1301-1305, 2009.

[3] L. Calabrese and E. Proverbio, "A Review on the Applications of Acoustic Emission Technique in the Study of Stress Corrosion Cracking," Corros. Mater. Degrad. vol. 2, pp. 1–30, 2021, https://doi.org/10.3390/cmd2010001.

[4] P. Louda, A. Sharko, and D. Stepanchikov, "An Acoustic Emission Method for Assessing the Degree of Degradation of Mechanical Properties and Residual Life of Metal Structures under

Complex Dynamic Deformation Stresses," Materials, vol. 14, pp. 2090, 2021, https://doi.org/10.3390/ma14092090.

[5] Á. Angulo et al. "Acoustic Emission Monitoring of Fatigue Crack Growth in Mooring Chains," Appl. Sci., vol. 9, pp. 2187, 2019; doi:10.3390/app9112187.

[6] D. Tonelli, M. Luchetta, F. Rossi, and P. Migliorino, "Structural Health Monitoring Based on Acoustic Emissions: Validation on a Prestressed Concrete Bridge Tested to Failure," Sensors, vol. 20, pp. 7272, 2020, doi:10.3390/s20247272.

[7] J-X. Chen, J-W Li, C-C. Cheng, and C-W. Chiu, "Piezoelectric Property Enhancement of PZT/Poly(vinylidenefluoride-***co***-trifluoroethylene) Hybrid Films for Flexible Piezoelectric Energy Harvesters," ACS Omega, vol. 7, pp. 793−803, 2022.

[8] R. Lay, G. S. Deijs, and J. Malmstrom, "The intrinsic piezoelectric properties of materials –a review with a focus on biological materials," RSC Adv., vol. 11, pp. 30657, 2021.

[9] J. E. Marshall et al. "On the Solubility and Stability of Polyvinylidene Fluoride," Polymers, vol. 13, Issue 9, pp. 1354, 2021, https://doi.org/10.3390/polym13091354.

[10] Y. Bhuiyan, B. Lin, and V. Giurgiutiu, "Characterization of piezoelectric wafer active sensor for acoustic emission sensing," Ultrasonics, vol. 92, pp. 35-49, 2019.

[11] I. Chilibon, M. Mogildea, and G. Mogildea, "Wireless acoustic emission sensor device with microcontroller," Edited by: R. Walczak, J. Dziuban, Conference: 26th European Conference on Solid-State Transducers (Eurosensors), September 2012.

[12] A. Carpinteri, S. Invernizzi, and G. Lacidogna, "Historical brick-masonry subjected to double flat-jack test: Acoustic emissions and scale effects on cracking density," Construction and Building materials, vol. 23, Issue 8, pp. 2813-2820, August 2009.

[13] J. Bohse, "Acoustic emission characteristics of micro-failure processes in polymer blends and composites," Composites Science and Technology, vol. 60(8), pp. 1213–1226, 2000, doi:10.1016/S0266-3538(00)00060-9.

[14] K. M. Holford, M. J. Eaton, J. J. Hensman, R. Pullin, S. L. Evans, N. Dervilis, and K. Worden,"A new methodology for automating acoustic emission detection of metallic fatigue fractures in highly demanding aerospace environments: An overview," Progress in Aerospace Science, vol. 90, pp. 1-11, April 2017, https://doi.org/10.1016/j.paerosci.2016.11.003.

[15] S. Uppal, D. Yoshino, and H.I. Dunegang, "Using Acoustic Emission to Monitor Fatigue cracks on the Bridge at FAST," Technology Digest, February 2002.

[16] K. Ono, "Calibration Methods of Acoustic Emission Sensors," Materials, vol. 9(7), pp. 508, 2016, https://doi.org/10.3390/ma9070508.

[17] I. Chilibon, "Metallic Structures Behaviour under Mechanical Stretches" 17th International Congress on Sound and Vibration (ICSV17), pp. 1-5, July 2010.

[18] A. Boniface, J. Saliba, Z.M. Sbartaï, N. Ranaivomanana, and J.-P Balayssac, "Evaluation of the acoustic emission 3D localisation accuracy for the mechanical damage monitoring in concrete," Eng. Fract. Mech., vol. 223, pp. 106742, 2020, https://doi.org/10.1016/j.engfracmech.2019.106742.

[19] S. Peng et al., "Mechanical damage evaluation of masonry under tensile loading by acoustic emission technique," Construction and Building Materials, vol. 258, pp. 120336, 2020

[20] A. Carpinteri, G. Lacidogna, M. Corrado, and E. Di Battista, "Cracking and crackling in concrete-like materials: a dynamic energy balance," Eng. Fract. Mech., vol. 155, pp. 130–144, Apr. 2016, https://doi.org/10.1016/j.engfracmech.2016.01.013.

[21] D. D. Mandal, M. Bentahar, A. E. Mahi, A. Brouste, R. E. Guerjouma, S. Montresor, and F.-B. Cartiaux, "Acoustic Emission Monitoring of Progressive Damage of Reinforced Concrete T-Beams under Four-Point Bending," Materials, vol. 15, 3486, pp. 1-25, 2022, https://doi.org/10.3390/ma15103486.

[22] Y. B. Guo, G. F. Gao, L. Jing, and V. P. W. Shim, "Quasi-static and dynamic splitting of high-strength concretes—Tensile stress–strain response and effects of strain rat," Int. J. Impact Eng. vol. 125, pp. 188–211, 2019.

[23] International Standard ISO 16836: 2019, Non-destructive testing – acoustic emission inspectrion measurement method for acoustic emission signals in concrete

[24] F. Hassan *et al.* "State-of-the-Art Review on AE Source Localization Techniques," IEEE Access, vol. 9, pp. 101246-101266, 2021.

Using Visible Light Communication to Guide Mobile Users Inside Large Buildings

Manuela Vieira, Manuel Augusto Vieira, Paula Louro, Alessandro Fantoni
ADETC/ISEL/IPL,
R. Conselheiro Emídio Navarro, 1959-007
Lisboa, Portugal
CTS-UNINOVA
Quinta da Torre, Monte da Caparica, 2829-516,
Caparica, Portugal

e-mail: mv@isel.ipl.pt, mv@isel.pt, plouro@deetc.isel.pt, afantoni@deetc.isel.ipl.pt

Pedro Vieira
ADETC/ISEL/IPL,
R. Conselheiro Emídio Navarro, 1959-007
Lisboa, Portugal
Instituto das Telecomunicações
Instituto Superior Técnico, 1049-001,
Lisboa, Portugal
e-mail: pvieira@isel.pt

Abstract— **The main goal of this paper is a Visible Light Communication (VLC) based guidance system to be used by mobile users inside large buildings. This system is composed of several transmitters (ceiling luminaries), which transmit map information and path guidance messages. Mobile devices, with VLC support, decode the information. A mesh cellular hybrid structure is proposed. The luminaires, via VLC, deliver their geographic position and specific information to the users, making them available for whatever use they request. The communication protocol, coding/decoding techniques, and error control are examined. Bidirectional communication is implemented and the best route to navigate through venue calculated. We propose several guidance services and multi-person cooperative localization. By analyzing the results, it became clear that the system not only provides self-location, but also the capability to determine the direction of travel and to interact with information received in order to optimize the route towards a static or dynamic destination.**

Keywords- Visible Light Communication; Assisted indoor navigation; Bidirectional Communication; Optical sensors; Transmitter/Receiver; Edge-Fog architecture.

I. INTRODUCTION

This paper is an extended and polished version of the paper presented in SENSORDEVICES 2022 conference [1]. The main goal is to specify the system conceptual design and define a set of use cases for a VLC based guidance system to be used by mobile users inside large buildings. The most obvious method of using guidance signs is through billboards located in high traffic areas. Handheld devices allow customers to stay informed, gather information and communicate with others without being tied to a physical location.

With the rapid increase in wireless mobile devices, the continuous increase of wireless data traffic has brought challenges to the continuous reduction of radio frequency (RF) spectrum, which has also driven the demand for alternative technologies [2][3]. In order to solve the contradiction between the explosive growth of data and the consumption of spectrum resources, VLC has become the development direction of the next generation communication network with its huge spectrum resources, high security, low cost, and so on [4][5].

With the increasing shortage of radio frequency spectrum and the development of Light-Emitting Diodes (LEDs), VLC has attracted extensive attention. Compared to conventional wireless communications, VLC has higher rates, lower power consumption, and less electromagnetic interferences. VLC is a data transmission technology that can easily be employed in indoor environments since it can use the existing LED lighting infrastructure with simple modifications [6][7]. The use of white polychromatic LEDs offers the possibility of Wavelength Division Multiplexing (WDM), which enhances the transmission data rate. A WDM receiver based on tandem a-SiC:H/a-Si:H pin/pin light controlled filter can be used [8][9] to decode the received information. Here, when different visible signals are encoded in the same optical transmission path, the device multiplexes the different optical channels, performs different filtering processes (amplification, switching, and wavelength conversion) and finally decodes the encoded signals recovering the transmitted information.

Visible light can be used as an Identifier (ID) system and can be employed for identifying the building itself. The main idea is to divide the service area into spatial beams originating from the different ID light sources and identify each beam with a unique timed sequence of light signals. The signboards, based on arrays of LEDs, positioned in strategic directions [10], are modulated acting as down- and up-link channels in the bidirectional communication. For the consumer services, the applications are enormous. Positioning, navigation, security and even mission critical services are possible use cases that should be implemented.

In this paper, a VLC based guidance system to be used by mobile users inside large buildings is proposed. After the Introduction, in Section II, a model for the system is proposed and the communication system described. In Section III and Section IV, the main experimental results are presented, downlink and uplink transmission is implemented and the best route to navigate calculated. In Section V, the conclusions are drawn.

II. SYSTEM MODEL

The system model of the proposed system will be presented in this section.

A. *Communication system*

The system model is composed by two modules: the transmitter and the receiver. The block diagram is presented in Figure 1. Both communication modules are software defined, where modulation/demodulation can be programed.

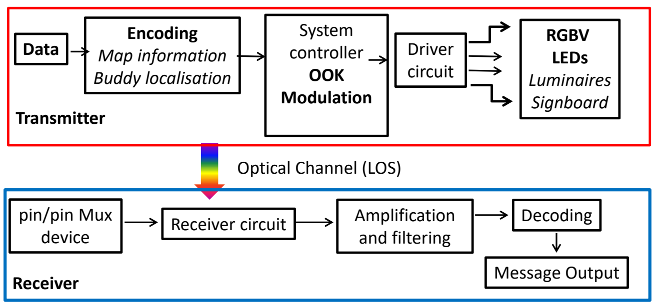


Figure 1. Block diagram. System model of the proposed control scheme applied to OOK modulation.

Data from the sender is converted into an intermediate data representation, byte format, and converted into light signals emitted by the transmitter module. The data bit stream is input to a modulator where an ON–OFF Keying (OOK) modulation is utilized. On the transmission side, a modulation and conversion from digital to analog data is done. The driver circuit will keep an average value (DC power level) for illumination, combining it with the analog data intended for communication. The visible light emitted by the LEDs passes through the transmission medium and is then received by the MUX device.

To realize both the communication and the building illumination, white light tetra-chromatic sources (WLEDs) are used providing a different data channel for each chip. The transmitter and receiver relative positions are displayed in Figure 2a. Each luminaire is composed of four polichromatic WLEDs framed at the corners of a square. At each node, only one chip is modulated for data transmission (see Figure 2b), the Red (R: 626 nm¸ 25 μW/cm^2), the Green (G: 530 nm, 46 μW/cm^2), the Blue (B: 470 nm, 60 μW/cm^2) or the Violet (V, 400 nm, 150 μW/cm^2). Data is encoded, modulated and converted into light signals emitted by the transmitters. Modulation and digital-to-analog conversion of the information bits is done using signal processing techniques. An OOK modulation scheme was used to code the information. This way digital data is represented by the presence or absence of a carrier wave.

The signal is propagating through the optical channel, and a VLC receiver, at the reception end of the communication link, is responsible to extract the data from the modulated light beam. It transforms the light signal into an electrical signal that is subsequently decoded to extract the transmitted information. The obtained voltage is then processed, by using signal conditioning techniques (adaptive bandpass filtering and amplification, triggering and demultiplexing), until the data signal is reconstructed at the data processing unit (digital conversion, decoding and decision) [11] [12]. At last, the message will be output to the users.

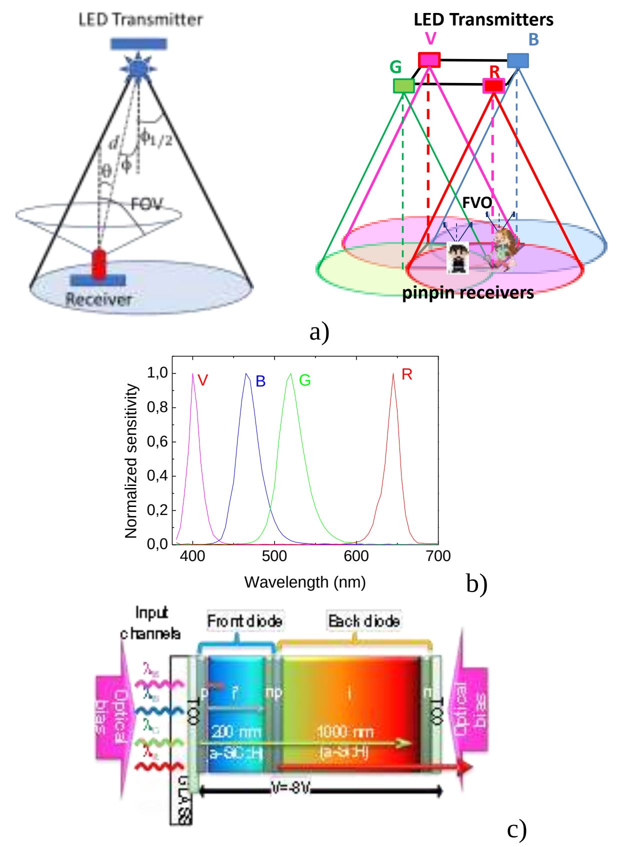


Figure 2. a)3D relative positions of the transmitters and receivers. b) Spectra of the input channels. c)Configuration and operation of the pin/pin Mux device

On the receiving side, a MUX photodetector acts as an active filter for the visible region of the light spectrum. The integrated filter consists of a p-i'(a-SiC:H)-n/p-i(a-Si:H)-n heterostructure with low conductivity doped layers [7] as displayed in Figure 2c. Independent tuning of each channel is performed by steady state violet optical bias (λ_{bias}= 2300 μW/cm^2) superimposed from the front side of the device and the generated photocurrent measured at -8V. The generated photocurrent is processed using a transimpedance circuit obtaining a proportional voltage. Since the photodetector response is insensitive to the frequency, phase, or polarization of the carriers, this kind of receiver is useful for intensity-modulated signals. After receiving the signal, it is in turn filtered, amplified, and converted back to digital format for demodulation. The system controller consists of a set of programmable modules.

In this system model, there are a few assumptions that should be noted: The channel state information is available both at the receiver and the transmitter; compared with the direct light, the reflected light is much weaker in the indoor VLC systems; only the Line OF Sight (LOS) path is considered and the multipath influence is not considered in the proposed indoor VLC system.

The received channel can be expresssed as:

$$y=\mu hx+n \tag{1}$$

where *y* represents the received signal, *x* the transmitted signal, μ is the photoelectric conversion factor which can be normalized as μ = 1, *h* is the channel gain and n is the additive white Gaussian noise of which the mean is 0.

The LEDs are modeled as Lambertian sources where the luminance is distributed uniformly in all directions, whereas the luminous intensity is different in all directions. The luminous intensity for a Lambertian source is given by Equation (2) [13]:

$$I(\emptyset)=I_N \cos \emptyset^m \,;\ m=\frac{ln(2)}{ln\left(\cos \emptyset_{1/2}\right)} \tag{2}$$

I_N is the maximum luminous intensity in the axial direction, ϕ is the angle of irradiance and *m* is the order derived from a Lambertian pattern. For the proposed system, the commercial white LEDs were designed for illumination purposes, exhibiting a wide half intensity angle ($\phi_{1/2}$) of 60º. Thus, the Lambertian order *m* is 1. Friis' transmission equation is frequently used to calculate the maximum range by which a wireless link can operate. The coverage map is obtained by calculating the link budget from the Friis transmission equation [14].

The Friis transmission equation relates the received power (P_R) to the transmitted power (P_E), path loss distance (L_R), and gains from the emitter (G_E) and receiver (G_R) in a free-space communication link.

$$P_{R\ [dBm]}=P_{E\ [dBm]}+G_{E\ [dB]}+G_{R\ [dB]}-L_{R\ [dB]} \tag{3}$$

Taking into account Figure 2a, the path loss distance and the emitter gain will be given by:

$$L_{R\ [dB]}=22+20\ln\frac{d}{\lambda} \tag{4}$$

$$G_{E\ [dB]}=\frac{(m+1)A}{2\pi d_{E-R}^2} I(\emptyset)\cos(\theta) \tag{5}$$

With *A* the area of the photodetector and d_{E_R} the distance between each transmitter and every point on the receiver plane. Due to their filtering properties of the receptors the gains are strongly dependent on the wavelength of the pulsed LEDs. Gains (G_R) of 5, 4, 1.7 and 0.8 were used, respectively, for the R, G, B and V LEDs. I_N of 730 mcd, 650 mcd, 800 mcd and 900 mcd were considered. The coverage map, for a square unit cell, is shown in Figure 5.

B. *Building model and Architecture*

Lighting in large environments is designed to illuminate the entire space in a uniform way. The proposed scenario is a multi-level building. Ceiling plans for the LED array layout, in floor level is shown in Figure 3. A square lattice topology was considered for each level.

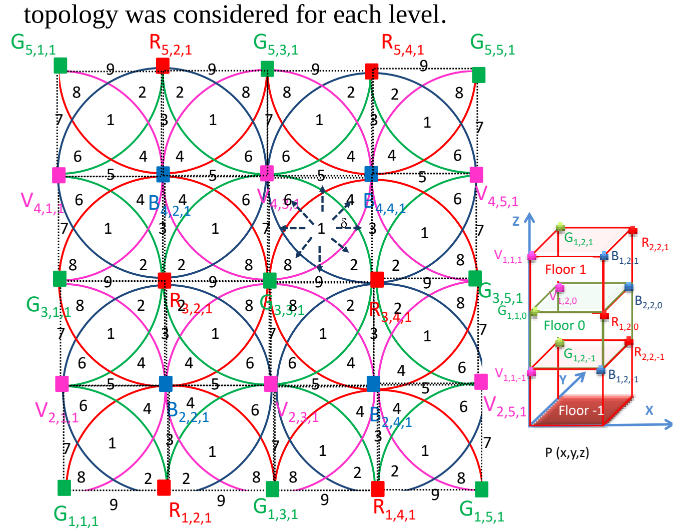


Figure 3. Clusters of cells in square topology. Illustration of the optical scenario. (RGBV =modulated LEDs spots).

In fog /edge computing, computing, storage, networking, and data management services are provided on nodes within close proximity to IoT devices, bridging the gap between the cloud and end devices.

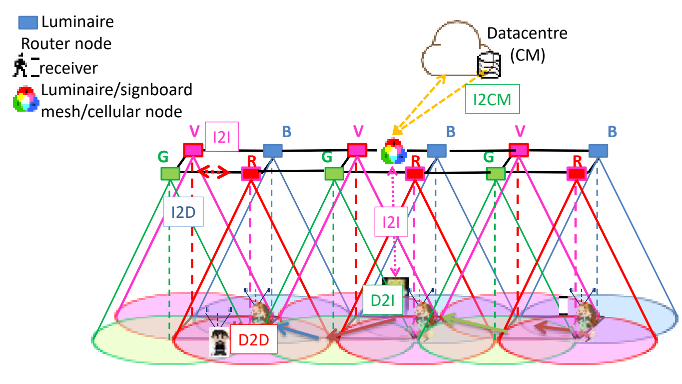


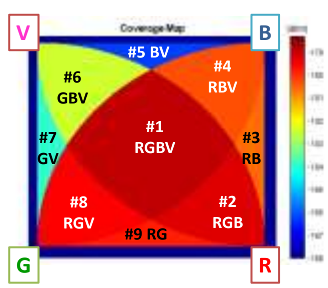
Figure 4. Mesh and cellular hybrid architecture.

In Figure 4, the proposed architecture is illustrated. Under this architecture, the short-range mesh network purpose is twofold: enable edge computing and device-to-cloud communication, by ensuring a secure communication from a luminaire controller to the edge computer or datacenter (I2CM), through a neighbor luminaire/signboard controller with an active cellular connection; and enable peer-to-peer communication (I2I), to exchange information.

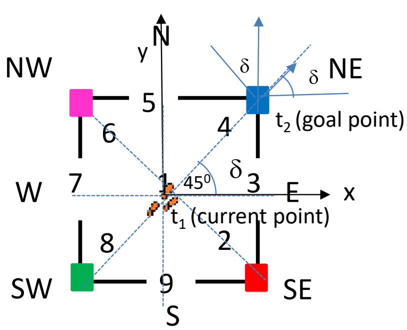
A user navigates from outdoor to indoor. It sends a request message (D2I) to find the right track and, in the available time, he adds customized points of interest (guidance services). The requested information (I2D) is sent by the emitters at the ceiling to its receiver.

In this architecture, the polychromatic WLEDs are placed on the ceiling in a square lattice topology (see Figure 3), but only one, chip is modulated (R, G, B, V). The principle is that each WLED transmits a VLC signal with a unique identifier. The optical receiver uses this information and a position algorithm, based on the received joint transmission, calculates the track of the user.

To receive the I2D information from several transmitters, the receiver must be located at the overlap of the circles that set the transmission range (radial) of each transmitter.



a)



b)

Steering angle code	1	2	3	4	5	6	7	8	9
Square topology (δ)	-	(SE)°	(E)	(NE)	(N)	(NW)	(W)	(SW)	(S)

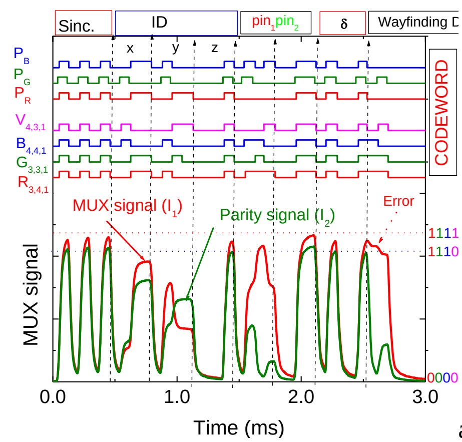
Figure 5. Illustration of the coverage map in the unit cell: a) Footprint regions (#1-#9). b) Steering angle codes (2-9).

Taking into account (1)-(5), the coverage map for a square unit cell is displayed in Figure 5. All the values were converted to decibel (dB). The nine possible overlaps (#1-#9), defined as fingerprint regions, as well as receiver orientations (2-9 steering angles; δ) are also pointed out for the unit square cell, in Figure 5. The input of the aided navigation system is the coded signal sent by the transmitters to an identified user (I2D), and includes its position in the network *P(x, y, z)*, inside the unit cell and the steering angle, δ, that guides the user across his path at a given time, *t*. The device receives multiple signals, finds the centroid of the received coordinates, and stores it as the reference point position. Nine reference points, for each unit cell, are identified giving a fine-grained resolution in the localization of the mobile device across each cell.

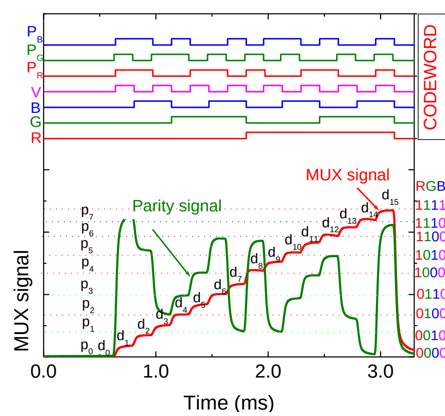
The indoor route throughout the building (track; *q(x, y, z,* δ, *t)*) is presented to the user by a responding message (I2D) transmitted by the ceiling luminaires that work also either as router or mesh/cellular nodes.

Two-way communication (D2I-I2D) between users and the infrastructure is carried out through a neighbor luminaire/signboard controller with an active cellular connection (I2CM). With this request/response concept, the generated landmark-based instructions help the user to unambiguously identify the correct decision point where a change of direction (pose) is needed, as well as offer information for the user to confirm that he/she is on the right way.

C. *Communication protocol, coding/decoding techniques and error control*



a)



b)

Figure 6. Code and parity MUX/DEMUX signals. On the top the transmitted channels [R G B V : P_R P_G P_B] are shown. a) Calibrated cell. b) Error control assigned to a request from user "7261" at $C_{4,3,1}$; #1 N.

Using the photocurrent signal measured by the photodetector, it is necessary to decode the received information. A calibration curve is previously defined to establish this assignment [15]. As displayed in Figure 6b, calibration curves make use of 16 distinct photocurrent thresholds which correspond to a bit sequence that allows all the sixteen combinations of the four RGBV input channels (2^4). If the calibrated levels (d_0-d_{15}) are compared to the different four-digit binary codes assigned to each level, then the decoding is obvious, and the message may be read [14]. Due to the proximity of successive levels (see Figure 6a) occasional errors occur in the decoded information. A parity check is performed after the word has been read [16]. The parity bits are the SUM bits of the three-bit additions of violet pulsed signal with two additional RGB bits and defined as:

$$P_R = V x\ R \text{ x } B;\ P_G = V x\ R \text{ x } G;\ \ P_B = V \text{ x } G \text{ x } B \quad (6)$$

In Figure 6b, the MUX signal that arises from the transmission of the four calibrated RGBV wavelength channels and the MUX signal that results from the generation of the synchronized parity MUX are displayed. On the top the seven bit word [R,G,B,V, P_R, P_G, P_B] of the transmitted inputs guides the eyes. The colours red, green, blue and violet were assigned respectively to P_R, P_G , P_B and P_V. For simplicity the received data ($d_{0\text{-}15}$ levels) is marked in the correspondent MUX slots as well as the parity levels marked as horizontal lines. On the top the decoded 7-bit coded word is exhibited. In the right side 4-bit binary codes assigned to the eight parity sublevels are inserted**.**

In Figure 6a we illustrate how error control is achieved using check parity bits. A request from user "7261" is shown at $C_{4,3,1}$; #1 N, along with the matching parity signal. To automate the process of recovering the original transmitted data, an algorithm was developed. The transmitted data is decoded by comparing the code MUX signal with the parity MUX levels. The decoding algorithm is based on a proximity search [17].

For each time slot, the data are translated into a vector in multidimensional space, which is determined by the signal currents I_1 and I_2, where I_1 is the d level and I_2 is the p level for the 4-bit codeword (RGBV). The corresponding parity levels, [P_R, P_G P_B] in the respective time slot are also obtained and are assumed to be correct. The result is then compared with all vectors resulting from the calibration sequence (Figure 6a) where each code level, d (0-31) is assigned the corresponding parity level, p (0-15). Euclidean metrics are used to calculate distances.

The tests were done with a variety of random sequences, and we were able to recover the original colour bits. Figure 7 illustrates the encoding/decoding process with and without check parity error. In Figure. 7a, the encoded optical signals (codewords) and the experimental received signals are depicted. After encoding, Figure 7b shows how information can be recovered with and without error control. The encoded signals transmitted by the LEDs are determined through the interpolation of the signals received by the photodiode, (MUX and Parity, Figure 6a), with the calibration curves (Figure 6b).

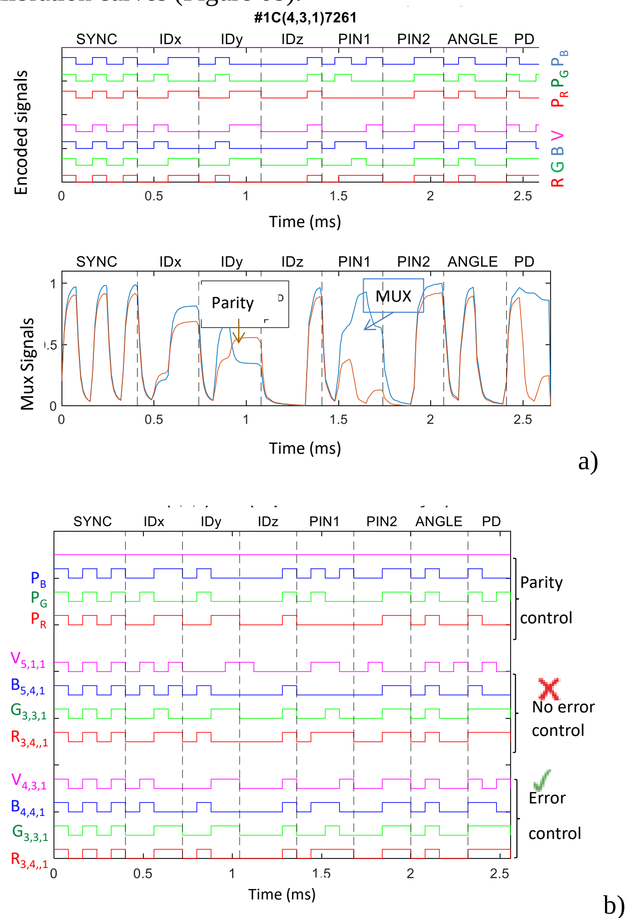


Figure 7. Encoding/decoding process with and without check parity error. a) Transmitted code signals [R G B V : P_R P_G P_B] and received MUX and Parity signals. b) Decoded information with and without error control assigned to a request from user "7261" at cell $C_{4,3,1}$ #1 NE.

Results show that without check parity bits, decoding was difficult primarily when levels were close together (dotted arrow). Based on the results for the analysed cases, the BER is high (4.6% without error correction) whereas it is negligible with error correction.

III. COOPERATIVE SELF-LOCALIZATION AND GUIDANCE SERVICES

Concepts of self-localization and route control will be present in this section.

A. *Self-localization*

Self-localization is a fundamental issue since the person must be able to estimate its position and orientation (pose) within a map of the environment it is navigating. We consider the path to be a geometric representation of a plan to move from a start pose to a goal pose (Figure 5). Let us consider a person navigating in a 2D environment. Its non-omnidirectional configuration is defined by position (x, y) and orientation angle δ, with respect to the coordinate axes. $q(t) = [x(t), y(t), \delta(t)]$ denote its pose at time t, in a global reference frame. In cooperative positioning systems, persons are divided into two groups, the stationary persons and the moving persons. Let us consider that $q_i(t,t')$ represents the pose of person i at time t' relative to the pose of the same person at time t and $q_{ij}(t)$ denotes the pose of person j relative to the pose of person i at time t. $q_i(t,t')$ is null for people standing still and non-zero if they move. These three types of information $q_i(t)$, $q_i(t,t')$and $q_{ij}(t)$ compose the basic elements of a pose graph for multi-person cooperative localization.

We consider that the risk of catching a disease exists if $q_{ij}(t)$ is less than 2 m. The system will alert the users to stay away from those regions and to plan the better route to the desired wayfinding services. To estimate each person track the pure pursuit approach [18][19] is used. The principle takes into account the curvature required for the mobile receiver to steer from its current position (t_1) to its intended position (t_2). By specifying a look-ahead distance, it defines the radius of an imaginary circle. Finally, a control algorithm chooses a steering angle in relation to this circle. This then allows to iteratively construct the intermediate arcs between itself and its goal position as it moved, thus, obtaining the required trajectory for it to reach its objective position. To avoid the risk of transmission, in the same frame of time and in known crowded regions, $q_{ij}(t)$ is estimated and the steering angle readjusted [20].

B. Geotracking, Navigation and Route Control

VLC geotracking is the process of identification or estimation of the geographic position of a device through visible light. It involves the generation of a set of geographic coordinates but its usefulness is enhanced by the use of these coordinates to determine a meaningful location, to help the user navigation through an unfamiliar building or, to guide him to an intended destination or planned meeting.

In Figure 8, the MUX received signal and the decoding information that allows the VLC geotracking and navigation in successive instants (t_0, t_1, t_2) from user "7261" guiding him along his track is exemplified. In the right side, the match between the MUX signals and the 4-binary codes are pointed out (horizontal dotted lines) to facilitate the decoding process. On the top, the decoded channels packets are shown [R, G, B, V]. The visualized cells, paths and the reference points (footprints) are also shown as inserts.

After decoding the MUX signals, and taking into account the frame structure, the ID of the receiver in the network (footprint and 3D coordinates), wayfinding services required (pin_1, pin_2) orientation (δ) and wayfinding data are revealed.

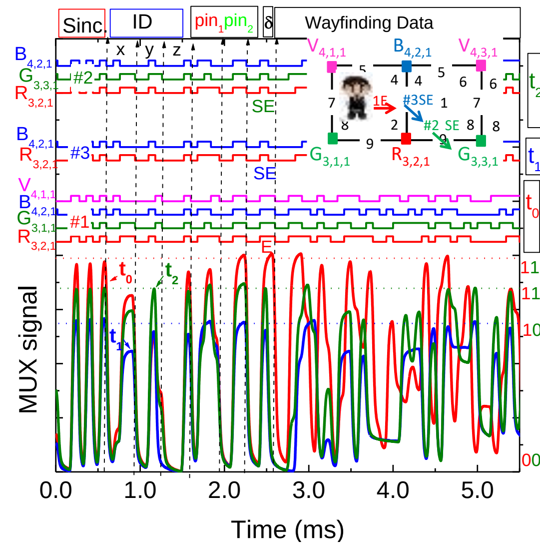


Figure 8. Fine-grained indoor localization and navigation in successive instants. On the top the transmitted channels packets are decoded [R, G, B, V]

Data shows that as a receiver moves between generated point regions, the received information pattern changes. Between two consecutive data sets, there is always a navigation data bit transition (channels are missing or added). At t_0, a maximum in the synchronism amplitude was detected and corresponds to the binary word [1111], meaning that the user has received the overlap transmission from all the four channels, footprint #1, while at t_1 the green and violet channels were missing, indicating footprint #3, and at t_3 the green channel from an adjacent cell was added, footprint #2. The transmitter's node address comes directly from the ID bits. At t_0 the network location of the received signals are $R_{3,2,1}$, $G_{3,1,1}$, $B_{4,2,1}$ and $V_{4,1,1}$, at t_1 the user receives the signal only from the $R_{3,2,1}$, $B_{4,2,1}$ nodes and at t_2 he was moved to the next cell since the node $G_{3,1,1}$was added at the receiver. The next 12 bits identify the user code (pin_1, "7621"), the meeting code, (pin_2, 3) and the steering angles (δ) required for the mobile receiver to steer from its current position (#1) to its intended position (#9). Finally, the last block is reserved for the transmission of the wayfinding message (Payload data). Hence, the mobile user "7261" begins his route into position #1(t_0) and wants to be directed to his goal position, in the next cell (# 9). During the route the navigator is guided to E (code 3) and, at t_1, steers to SE (code 2), cross footprint #2 (t_3) and arrives to #9.

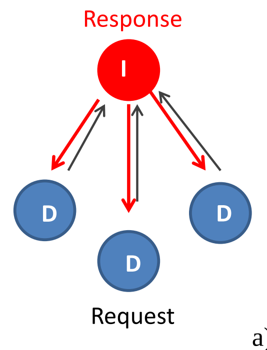
The proposed VLC dynamic geolocation system gives every location a unique perceptual identity, so that the navigator can associate their immediate surroundings (unit cell) with a location in the larger space (building network). This is one of the main principles of wayfinding. The ceiling lamps (landmarks) are spread over all the building and can act as edge/fog nodes in the network (see Figure 4), providing well-structured paths that maintain a navigator's orientation with respect to both the next landmark along the path and the distance to the eventual destination. Also, the VLC dynamic system enables cooperative and oppositional geolocation. In some cases, it is in the user's interest to be accurately located, so that they can be offered information relevant to their location and orientation (pin_1, pin_2 and δ blocks). Since geolocation software can get the information of user location, companies using geomarketing may provide web content or products that are famous or useful in that specific location. In other cases, users prefer not to disclose their location for privacy, in this case these last three blocks are set to zero and the user only receives its own location. This architecture also allows that advertisements and content on a website that uses geolocation software may be tailored to provide the information that a certain user wants.

IV. MULTI-PERSON COOPERATIVE LOCALIZATION AND GUIDANCE SERVICES

A VLC system consists of two entities, infrastructure and device (I and D), which establish bidirectional communication through cellular edge nodes. Bi-directional communication between VLC emitters and receivers is available at a VLC ready handheld device, through the control manager interconnected with a signboard receiver located at each unit cells (#1). Via the control manager, a handheld device with VLC connectivity communicates bidirectionally with a signboard receiver in each unit cell (#1). Each user (D2I) uplinks to the local controller a "request" message with the pose, $q_i(t)$, (x, y, z, δ), user code (pin_1) and also adds its needs (code meeting and wayfinding data). For route coordination the CM, using the information of the network's VLC location capability, downlinks a personalized "response" message to each client at the requested pose with his wayfinding needs (I2D).

These communication channels constitute the uplink (D2I) and downlink channels (I2D) as exemplified in Figure 9a. Each user (D2I) sends to the local controller a "request" message with the pose, $q_i(t)$, (x,y,z, δ), user code (pin_1) and also adds its needs (code meeting and wayfinding data). For route coordination the CM, using the information of the network's VLC location capability, sends a personalized "response" message to each client at the requested pose with his wayfinding needs.

In Figure 9b, the MUX synchronized signals received by two users that have requested guidance services, at different times, are displayed. We have assumed that a user located at $C_{2,3,-1}$, arrived first (t_1), auto-identified as ("7261") and informed the controller of his intention to find a friend for a previously scheduled meeting (code 3). A buddy list is then generated and will include all the users who have the same meeting code. User "3009" arrives later (t_3), sends the alert notification ($C_{4,4,1}$; t_3) to be triggered when his friend is in his floor vicinity, level 1, identifies himself ("3009") and uses the same code (code 3), to track the best way to his meeting.



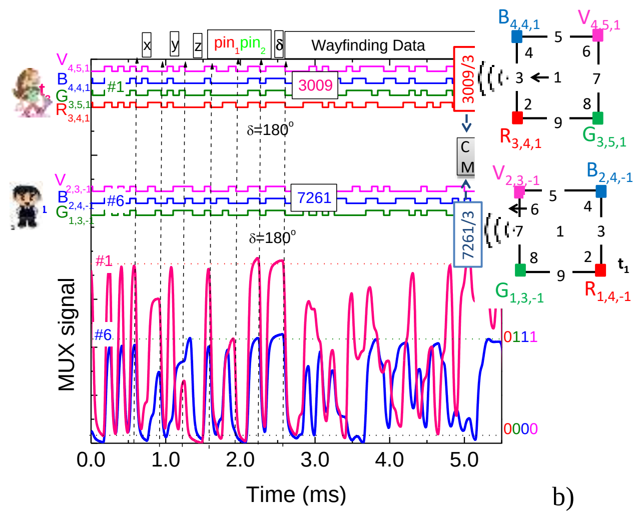


Figure 9. a) Communication channels established between the infrastructure and the vehicles for geolocation and navigation. b) MUX/DEMUX signals assigned requests from two users ("3009" and "7261") at different poses (C4,4,,1; #1W and C2,3,-1; #6 W) and in successive instants (t1 and t3).

Upon receiving this request (t_3), the buddy finder service uses the location information from both devices to determine the proximity of their owners (q_{ij} (t)) and provides the best route to the meeting, avoiding crowded areas.

An example of the MUX signals assigned to a request/response received by user "3009" during his path to reach user "7261" is displayed in Figure 10. In the top of the figure, the decoded information is shown and the simulated scenario is inserted to guide the eyes.

The "request" message includes, beyond synchronism, the identification of the user ("3009"), its address and orientation ($C_{4,4,1}$, #1W) and the help requested (Wayfinding Data). Since a meet-up between users is expected, its code was inserted before the right track request. In the "response", the block CM identifies the sender [0000] and the next blocks the cell address ($C_{4,4,1}$), the user (3009) for which the message is intended and finally the requested information: meeting code 3, orientation NE (code 4) and wayfinding instructions.

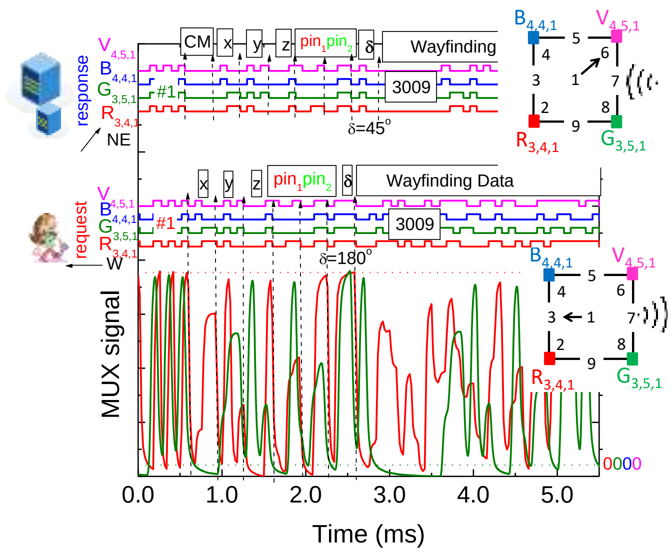


Figure 10. Request from user “3009” and response from the CM to him. On the top the transmitted channels packets are decoded [Xi,j,k].

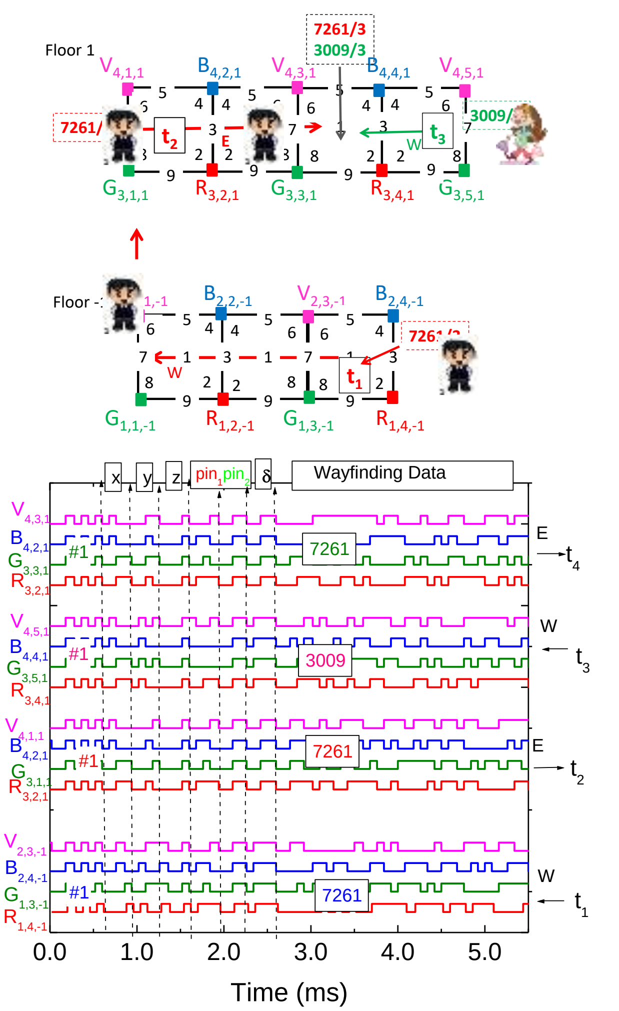


Figure 11. Decoded messages from the two users as they travel to a pre-scheduled meeting.

In response to the estimated relative pose position, $q_{i,j}(t)$, between the users with the same meeting code, the CM sends a new alert that takes into account the occupancy of the service areas along the paths, $q_i(x,y,z, \delta, t)$, which optimizes the path without crowding the users. This allows the CM to recalculate, in real time, the best route for the users, $q_i(t,t')$, that request wayfinding services avoiding crowded regions.

In Figure 11, the decoded messages from the two users as they travel to the pre-scheduled meeting is displayed.

Decoded data shows that user “7621” starts (t_1) his journey on floor -1, $C_{2,3,-1}$; #1W, goes up to floor 1 in $C_{2,1,-1}$ and at t_2 he arrives at $C_{4,1,1}$ heading for E. During his journey, user “3009” from $C_{4,4,1}$ #1 asks the CM (t_3) to forward him to the scheduled meeting and follows course to W. At t_4 both friends join in $C_{4,3,1}$.

The pedestrian movement along the path can be thought as a queue, where the pedestrians arrive at a path, wait if the path is congested and then move once the congestion reduces. In Figure 12, a graphical representation of the simultaneous localization and mapping problem using connectivity as a function of node density, mobility and transmission range is illustrated.

.

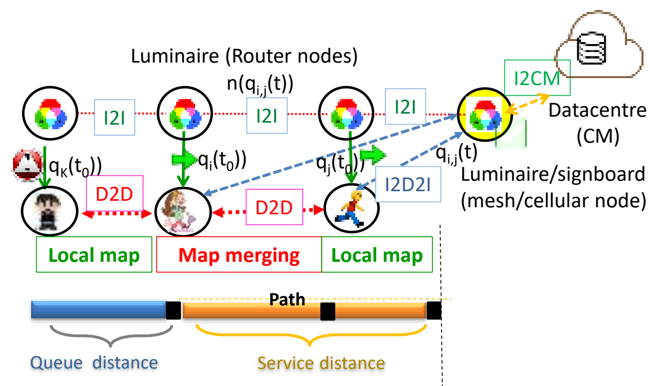


Figure 12. Graphical representation of the simultaneous localization and mapping problem using connectivity as a function of node density, mobility and transmission range.

The following parameters are therefore needed to model the queuing system: The initial arrival time (t_0) and the path, defined as the time when the pedestrian leaves the previous path and the actual movement along the path, $q_i(t, t')$. Here, the service time is calculated using walking speed and distance of the path. The number of service units or resources is determined by the capacity of the pathway, $n(q_i(x,y,z, \delta, t))$ and walking speed which depends on the number of request services, and on the direction of movement along the pathway $q_i(x,y,z, \delta, t)$. The pedestrians are served as soon as the request message is appended by the CM (response message) as displayed in Figure 12.

If the number of pedestrians exceeds the path capacity, a backlog is automatically formed until the starting node. The hybrid controller integrates the number of requests and individual positions received during the same time interval. Once the individual positions are known, $q_i(t)$, the relative positions are calculated, $q_{ij}(t)$. If the relative position is less than a threshold distance, a crowded region locally exists, and an alert message is sent for the users. This alert allows the CM to recalculate, in real time, the best route for the users, $q_i(t,t')$, that request wayfinding services avoiding crowded regions.

V. Conclusions

A VLC based guidance system to be used by mobile users inside large buildings was proposed and characterized. According to global results, the location of a mobile receiver is found in conjunction with data transmission. VLC's dynamic LED-aided guidance system is designed to give users accurate route guidance and enable navigation and geotracking. The multi-person cooperative localization system detects crowded regions and alerts the user to reschedule meetups, as well as provides guidance information. With those alerts, the CM can recalculate, in real time, the best route for users requesting wayfinding services, avoiding crowded areas. Further research activities are still necessary to optimize the coverage, namely the effects of synchronization, shadowing and ambient light. Also, the LED design and positioning has to be improved in the future, in order to optimize the communication performance while meeting the illumination constraints.

Acknowledgements

This work was sponsored by FCT – Fundação para a Ciência e a Tecnologia, within the Research Unit CTS – Center of Technology and Systems, reference UIDB/00066/2020 and IPL/2022/POSEIDON_ISEL.

References

[1] M. Vieira, M. A. Vieira, P. Louro, A. Fantoni, and P. Vieira, “VLC Based Guidance System to Be Used by Mobile Users Inside Large Buildings,” SENSORDEVICES 2022: The Thirteenth International Conference on Sensor Device Technologies and Application Copyright (c) IARIA, 2022. ISSN: 2308-3514, ISBN: 978-1-68558-006-3 pp. 9-14.

[2] F. Zafar, D. Karunatilaka, and R. Parthiban, “Dimming Schemes for Visible Light Communication: The State of Research, “ IEEE Wireless Commun., 22, pp. 29–35, 2015.

[3] L. U. Khan, “Visible Light Communication: Applications, Architecture, Standardization and Research Challenges,” Digit. Commun. Netw. 3, pp. 78–88, 2017.

[4] N. U. Hassan, A. Naeem, M. A. Pasha, T. Jadoon, and C. Yuen, “Indoor positioning using visible led lights: A survey,”ACM Comput. Surv., vol. 48, pp. 1–32, 2015.

[5] E. Ozgur, E. Dinc, and O. B. Akan, “Communicate to illuminate: State-of-the-art and research challenges for visible light communications,” Physical Communication 17, pp. 72–85, 2015.

[6] D. Tsonev et al., “A 3-Gb/s single-LED OFDM-based wireless VLC link using a Gallium Nitride μLED,” IEEE Photon. Technol. Lett. 26 (7), pp. 637–640, 2014.

[7] D. O’Brien et al., “Indoor visible light communications: challenges and prospects,” Proc. SPIE 7091, 709106, 2008.

[8] M. Vieira, P. Louro, M. Fernandes, M. A. Vieira, A. Fantoni and J. Costa “Three Transducers Embedded into One Single SiC Photodetector: LSP Direct Image Sensor, Optical Amplifier and Demux Device,” Advances in Photodiodes InTech, Chap.19, pp. 403-425, 2011.

[9] M. A. Vieira, P. Louro, M. Vieira, A. Fantoni, and A. Steiger-Garção, “Light-activated amplification in Si-C tandem devices: A capacitive active filter model,” IEEE Sensor Journal, 12, NO. 6, pp. 1755-1762, 2012.

[10] S. B. Park et al., “Information broadcasting system based on visible light signboard,” presented at Wireless and Optical Communication 2007, Montreal, Canada, 2007.

[11] M. Vieira, M. A. Vieira., P. Louro., P. Vieira, A. Fantoni, “Light-emitting diodes aided indoor localization using visible light communication technology,” Opt. Eng. 57(8), 087105, 2018.

[12] M. A. Vieira, M. Vieira, P. Louro, and P. Vieira, "Bi-directional communication between infrastructures and vehicles through visible light," Proc. SPIE 11207, Fourth International Conference on Applications of Optics and Photonics, 112070C (3 October 2019); doi: 10.1117/12.2526500. 2019.

[13] Y. Zhu, W. Liang, J. Zhang, and Y. Zhang, “Space-Collaborative Constellation Designs for MIMO Indoor Visible Light Communications,” IEEE Photonics Technology Letters, vol. 27, no. 15, pp. 1667–1670, 2015.

[14] H. T. Friis, “A note on a simple transmission formula,” Proc. IRE34, pp. 254–256, 1946.

[15] M. Vieira, M. A. Vieira, P. Louro, A. Fantoni, P. Vieira, "Dynamic VLC navigation system in Crowded Buildings," International Journal On Advances in Software, vol. 14, no. 3&4, pp. 141-150, 2021.

[16] M. A. Vieira, M. Vieira, V. Silva, P. Louro, and J. Costa, “Optical signal processing for data error detection and correction using a-SiCH technology,” Phys. Status Solidi C vol. 12, no. 12, pp. 1393–1400, 2015.

[17] M. A. Vieira, M. Vieira, P. Louro, V. Silva, J. Costa, and A. Fantoni, “SiC Multilayer Structures as Light Controlled Photonic Active Filters,” Plasmonics, vol. 8, no. 1, pp. 63-70, 2013.

[18] R. Rajamani, “Dynamic Vehicle and control,” Mechanical Engineering Series, ISBN 978-1-4614-1433-9 . 2012.

[19] J. Ackermann, J. Guldner, W. Sienel, R. Steinhauser, and V. Utkin, “Linear and Nonlinear Controller Design for Robust Automatic Steering,” IEEE Transactions on Control Systems Technology, vol. 3, no. 1, pp. 132 – 143, Mar. 1995, DOI: 10.1109/87.370719.

[20] M. Vieira, M. A. Vieira, P. Louro, A. Fantoni, and P. Vieira, "Geolocation and communication in unfamiliar indoor environments through visible light," Proc. SPIE 11706, Light-Emitting Devices, Materials, and Applications XXV, 117060P (5 March 2021).

Environmental Map Building and Moving Object Tracking Using Helmet-Mounted LiDAR and IMU for Micromobility

Ibuki Yoshida, Akihiko Yoshida
Graduate School of Science and Engineering
Doshisha University
Kyotanabe, Japan
e-mail: {ctwh0151, ctwj0112}@mail4.doshisha.ac.jp

Masafumi Hashimoto, Kazuhiko Takahashi
Faculty of Science and Engineering
Doshisha University
Kyotanabe, Japan
e-mail: {mhashimo, katakaha}@mail.doshisha.ac.jp

Abstract— **This paper presents a method of environmental map building and moving object tracking using Light Detection And Ranging (LiDAR) and Inertial Measurement Unit (IMU) mounted on a smart helmet worn by a micromobility rider. This presented method can be used for active safety for micromobility, such as bicycles, e-bikes, and electric scooters. Distortion in scan data from LiDAR is corrected by estimating the helmet pose (three-dimensional (3D) position and attitude angle) based on information obtained from normal distributions transform scan matching and IMU. The corrected LiDAR scan data are classified into scan data on road surfaces, road boundaries, stationary objects, and moving objects in the environments. Moving object scan data are used for moving object tracking. Stationary object and road boundary scan data are used to represent 3D stationary objects and road obstacles, such as curbs, gutters, and steps. Road surface scan data from LiDAR are in conjunction with IMU acceleration, and small road unevenness, such as potholes and humps, is detected to reduce the falling risk of micromobility. Furthermore, road surface conditions are identified by integrating IMU acceleration data, and a road surface condition map is constructed to provide safety and comfortability for micromobility riders in environments. Experiments conducted on a road on our university campus demonstrate the effectiveness of the proposed method.**

Keywords-helmet LiDAR/IMU; micromobility; environmental map building; moving object tracking; road boundary detection; road unevenness detection; road surface condition estimation.

I. INTRODUCTION

This paper is an extended and improved version of an earlier paper presented at the IARIA Conference on Sensor Device Technologies and Applications (SENSORDEVICES 2022) [1] in Lisbon.

Numerous studies on active safety and autonomous driving in the field of Intelligent Transportation Systems (ITS) have been conducted [2]. In the field of last-mile automation, there has also been flourishing study on delivery robots [3]. The environmental map building [4, 5] and tracking of moving objects, such as cars, cyclists and pedestrians [6, 7], are important issues for autonomous driving and the active safety of vehicles and mobility robots. Many related studies have been conducted using cameras, radars, and Light Detection And Ranging (LiDAR) [8, 9]. In this paper, we focus on environmental map building and moving object tracking with vehicle-mounted LiDAR.

To reduce carbon emissions and resolve congestion, the use of single-seater micromobility, such as bicycles, e-bikes, electric scooters, and standing-type personal mobilities, has been attracting attention as a means of short-distance transportation in urban cities [10]. The coronavirus disease 2019 has made people highly resistant to using conventional means of urban transportation, such as crowded trains and buses. Therefore, micromobilities have become more prevalent as a means of reducing the risk of infection in future endemic societies.

Micromobilities can be driven on a road, a bicycle lane, or a section of sidewalk. In addition, micromobilities are prone to tipping over due to unevenness in road surfaces, and there is also the issue that pedestrians in the vicinity who are “doing something while walking” will not notice the approach of micromobilities.

Although the frequency of traffic accidents involving micromobility has increased recently, study on active safety for micromobility is far behind. As a result, this paper explores environmental sensing for micromobility, such as environmental map building and moving object recognition for active safety for micromobility.

In the case of micromobility systems, it is difficult to mount many sensors on the vehicle body, as is the case with cars, because of size and theft concerns. Thus, it is desirable to mount small and easily detachable sensors on the handlebar of a micromobility or the helmet worn by the micromobility rider. This paper presents a Helmet-Mounted LiDAR (HML)-based environmental map building and the tracking of moving objects, such as cars, two-wheelers, and pedestrians, in dynamic and Global Navigation Satellite System (GNSS)-denied environments. Moving object tracking and map building related to stationary 3D objects and road obstacles, such as curbs, gutters, and steps, are essential for preventing collision accidents of micromobilities.

The detection of small road unevenness, such as potholes and humps, as well as the estimation of road surface conditions, is also required to prevent the falling risk of micromobility and improve rider comfort. However, it is difficult to detect small road unevenness and estimate road surface conditions using only LiDAR scan data. In this paper, these are performed using acceleration information from a helmet-mounted Inertial Measurement Unit (IMU).

The rest of this paper is organized as follows. Section II presents an overview of related work, and Section III describes the experimental system. Section IV overviews the experimental map building and moving object tracking. Section V explains the estimation method of the helmet self-pose and the distortion-correction method of LiDAR scan data, and Section VI presents the method for classifying LiDAR scan data into scan data on road surfaces, road boundaries, stationary objects, and moving objects. Section VII presents the method for estimating the road boundaries and tracking moving objects, and Section VIII presents the method for detecting small road unevenness and estimating road surface conditions using IMU acceleration. Section IX presents experimental results to verify the proposed method, followed by the conclusions and future works in Section X.

II. RELATED WORK

Many studies on environmental mapping and moving object tracking have been conducted [4–7]. We previously presented an environmental map-building method using car and motorcycle-mounted LiDAR based on Normal Distributions Transform (NDT)-Graph Simultaneous Localization And Mapping (SLAM) [11, 12] to build a three-dimensional (3D) point cloud map in community road environments. We also presented a method of moving object tracking using car and motorcycle-mounted LiDAR [13, 14].

Recently, we proposed a method of building a 3D point-cloud map in sidewalk and roadway environments using LiDAR attached to the rider's helmet (HML) of a micromobility [15]. This study is an extension of our previous works [14, 15] on environmental map building using HML and moving object tracking by motorcycle-mounted LiDAR, and these methods are integrated into our HML system.

Several studies have been conducted on surrounding environmental sensing using HML. Indoor SLAM, where people on helmets equipped with two-dimensional (2D) or one-dimensional (1D) LiDAR walk around in building and factory environments, has been reported [16–18]. Niforatos et al. [19] presented a method of skier detection using 1D LiDAR attached to ski helmets to reduce the risk of accidents on ski slopes.

Apart from map building and moving object tracking, many studies on road surface condition estimation have been proposed using car and bicycle-mounted sensors, such as LiDAR, accelerometer, and smartphone sensors [20–24]. However, the estimation of road surface conditions using a helmet-mounted IMU remains a challenge.

To the best of our knowledge, no studies have been conducted on environmental sensing including environmental map building and moving object tracking in sidewalks and roadways using 3D LiDAR and an IMU attached to the rider helmet of a micromobility. Although there have been several studies on helmets with sensors (smart helmets) in the ITS field [25], their use is limited to alcohol detection in motorcycle riders and collision-accident detection, as well as confirming rider safety after accidents.

III. EXPERIMENTAL SYSTEM

Figure 1 shows an overview of the smart helmet. The upper part of the helmet is equipped with a mechanical 64-layer LiDAR (Ouster, OS0-64) and an IMU (Xsens, MTi-300).

The HML has a maximum range of 55 m, a horizontal field of view of 360° with a resolution of 0.35°, and a vertical field of view of 90° with a resolution of 1.4°. LiDAR can obtain 1024 measurements (distance, direction, and reflected light intensity) every 1.56 ms (every 5.6° in the horizontal direction). Therefore, approximately 66,000 scan data points are acquired in one rotation (360° observation) period (100 ms).

The attitude angle (roll and pitch angles), angular velocity (roll, pitch, and yaw angular velocities), and three-axis acceleration of the helmet are obtained from the IMU every 10 ms. The measurement error for the attitude angle is less than ± 0.3°, that of the angular velocity is less than ± 0.2°/s, and that of the acceleration is less than ± 5 mG.

The weight of the mechanical LiDAR is 0.5 kg, and the smart helmet is heavier and larger than usual helmets. Therefore, the LiDAR reduces the usability and practicability of the smart helmet. Moreover, it affects the performance of the helmet in the event of a crash. However, modern LiDAR technology [26] has been developing smaller, more lightweight, and lower power consumption solid-state LiDARs than mechanical LiDARs. The use of solid-state LiDARs will significantly improve the usability and practicability of smart helmets.

IV. OVERVIEW OF ENVIRONMENTAL MAP BUILDING AND MOVING OBJECT TRACKING

Figure 2 shows the sequence of environmental map building and moving object tracking. LiDAR scan data captured in the helmet coordinate system attached to the HML are mapped onto the world coordinate system using the self-pose (3D position and attitude angle) information of the helmet. For this, an accurate self-pose of the helmet is required. NDT scan matching [27] is employed to estimate the self-pose in GNSS-denied environments.

Because LiDAR scans lasers in the omnidirection, all scan data within one scan cannot be obtained at a single location when the micromobility is moving or swinging, or when the rider's body is swinging. Therefore, if all scan data within one scan is transformed using the pose information of the helmet simultaneously, distortion arises in the LiDAR scan data mapped in the world coordinate system. Because distortion leads to inaccurate results in map building and moving object tracking, distortion correction of the LiDAR scan data is required. The distortion-correction method [15] is briefly described in Section V.

Figure 1. Overview of the experimental smart helmet.

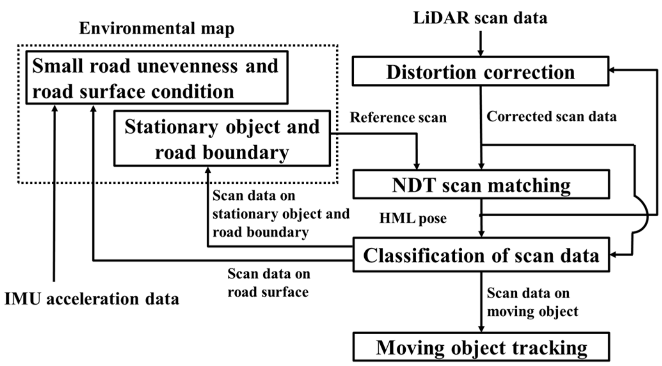


Figure 2. Sequence of environmental map building and moving object tracking.

The distortion-corrected LiDAR scan data are classified into scan data on road surfaces, road boundaries, stationary objects, and moving objects. Scan data on road surfaces (referred to as road surface scan data) are used to recognize areas where micromobility can travel. Scan data on stationary objects in the environment (stationary object scan data) are used to build a stationary object map. Scan data on moving objects (moving object scan data) are used for moving object tracking. Scan data relating to road boundaries (road boundary scan data) are used to detect road obstacles, such as fallen objects on the road, and boundaries of the road, e.g., grooves and curbs. The classification method of road surface, road boundary, stationary object, and moving object scan data from entire LiDAR scan data is described in Section VI. The methods for accurately estimating road boundaries from road boundary scan data and for accurately tracking moving objects from moving object scan data are explained in Section VII.

Although the position determination of small unevenness on road surfaces, such as potholes, cracks, and manhole covers, is necessary for preventing fall accidents from micromobilities, it is difficult to detect them with only LiDAR. Therefore, such small road unevenness is detected using acceleration data from IMU. Furthermore, the road surface condition is estimated using IMU acceleration data, and the related map is built to predict the safety of micromobility and the comfort of the rider. These methods are described in Section VIII.

Maps for stationary objects, road boundaries, and road unevenness, as well as moving object tracking, are built recursively after each LiDAR scan data and IMU acceleration data are obtained, whereas maps for road surface conditions are built in batch after a micromobility run is complete.

V. Self-Pose of Helmet and Distortion Correction of LiDAR Scan Data

This section briefly explains the self-pose (position and attitude angle) estimation of the helmet using NDT scan matching and Extended Kalman Filter (EKF)-based distortion correction of LiDAR scan data.

A. Estimation of Helmet Self-Pose

For the i-th measurement point ($i = 1, 2, \ldots n$) in the LiDAR scan data, the position in the helmet coordinate system is denoted by $\boldsymbol{p}_{Hi} = (x_{Hi}, y_{Hi}, z_{Hi})^T$, and that in the world coordinate system by $\boldsymbol{p}_i = (x_i, y_i, z_i)^T$. The following relationship is then represented by the homogeneous transformation:

$$\begin{pmatrix} \boldsymbol{p}_i \\ 1 \end{pmatrix} = \boldsymbol{T}(\boldsymbol{X}) \begin{pmatrix} \boldsymbol{p}_{Hi} \\ 1 \end{pmatrix} \tag{1}$$

where $\boldsymbol{X} = (x, y, z, \phi, \theta, \psi)^T$. $(x, y, z)^T$ and $(\phi, \theta, \psi)^T$ are the 3D position and attitude angle (roll, pitch, and yaw angles), respectively, of the helmet in the world coordinate system. $\boldsymbol{T}(\boldsymbol{X})$ is the following homogeneous transformation matrix:

$$\boldsymbol{T}(\boldsymbol{X}) = \begin{pmatrix} \cos\theta\cos\psi & \sin\phi\sin\theta\cos\psi - \cos\phi\sin\psi & \cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi & x \\ \cos\theta\sin\psi & \sin\phi\sin\theta\sin\psi + \cos\phi\cos\psi & \cos\phi\sin\theta\sin\psi - \sin\phi\cos\psi & y \\ -\sin\theta & \sin\phi\cos\theta & \cos\phi\cos\theta & z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

A voxel map with a cell size of 0.2 m per side is defined in the world coordinate system. In NDT scan matching, a normal distributions transformation is performed on the scan data obtained up to the previous time (referred to as reference scan data) in each cell of the voxel map, and the mean and covariance of the scan data in each cell are calculated.

The current self-pose $\boldsymbol{X}$ of the helmet is calculated by matching the scan data obtained at the current time (referred to as current scan data) with the reference scan data. The current scan data are mapped onto the world coordinate system by performing a coordinate transformation according to (1) using the pose $\boldsymbol{X}$. These data are then merged into the reference scan data. By repeating this process every LiDAR scan period, a LiDAR-based map is built.

B. Distortion Correction of LiDAR Scan Data

The helmet pose is determined every 100 ms (LiDAR scan period) by NDT scan matching. Scan data are acquired every 1.56 ms during one rotation of LiDAR. During LiDAR scanning, all scan data within one scan cannot be obtained at a single location when the micromobility is moving or swinging, or when the rider's body is swinging. If all scan data within one scan are transformed using the pose information of the helmet simultaneously, distortion appears in the mapping of the LiDAR scan data onto the world coordinate system. Therefore, the distortion in the scan data is corrected by estimating the helmet's pose using the EKF every 1.56 ms, i.e., every LiDAR scan data are obtained.

Figure 3 shows the sequence of distortion correction. The LiDAR scan period (100 ms) is denoted as τ, the IMU observation period (10 ms) as $\Delta\tau_{IMU}$, and the scan data observation period (1.56 ms) as $\Delta\tau$. Here, the method for correcting the distortion in scan data obtained between time $(t-1)\tau$ and $t\tau$ is described [15].

Let us suppose that at the time $(t-1)\tau$, the pose of the helmet is calculated by NDT scan matching and estimated using the EKF. IMU data are obtained 10 times per LiDAR scan ($\tau = 10\Delta\tau_{IMU}$). Using the IMU data obtained every $\Delta\tau_{IMU}$, the EKF estimates the pose $\hat{X}(t-1, k)$ at the time $(t-1)\tau + k\Delta\tau_{IMU}$, where k = 0–10. Because the observation period $\Delta\tau_{IMU}$ of the IMU is 10 ms, and the scan-data observation period $\Delta\tau$ is

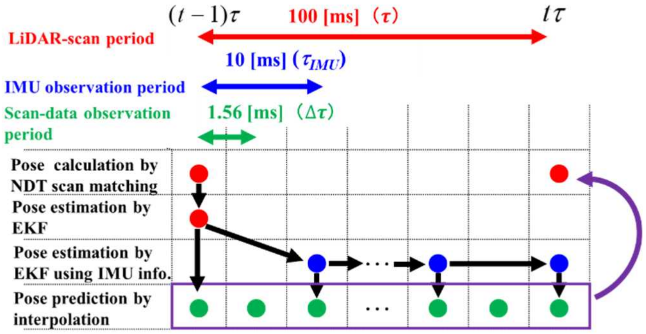


Figure 3. Sequence of distortion correction of LiDAR scan data.

1.56 ms, the LiDAR scan data are obtained six times within the IMU observation period ($\Delta\tau_{IMU}=6\Delta\tau$).

From the estimates, $\hat{\boldsymbol{X}}(t-1,k)$ at the time $(t-1)\tau+k\Delta\tau_{IMU}$ and $\hat{\boldsymbol{X}}(t-1,k+1)$ at $(t-1)\tau+(k+1)\Delta\tau_{IMU}$, the interpolation algorithm predicts the helmet's pose $\hat{\boldsymbol{X}}(t-1,k,j)$ at $(t-1)\tau+k\Delta\tau_{IMU}+j\Delta\tau$ (where j = 1–5), at which the scan data are acquired.

The scan data $\boldsymbol{p}_{Hi}(t-1,k,j)$ (where i = 1, 2, …, n) obtained at $(t-1)\tau+k\Delta\tau_{IMU}+j\Delta\tau$ in the helmet coordinate system are transformed into $\boldsymbol{p}_i(t-1,k,j)$ in the world coordinate system using (1) as follows:

$$\begin{pmatrix} \boldsymbol{p}_i(t-1,k,j) \\ 1 \end{pmatrix} = \boldsymbol{T}(\hat{\boldsymbol{X}}(t-1,k,j)) \begin{pmatrix} \boldsymbol{p}_{Hi}(t-1,k,j) \\ 1 \end{pmatrix} \quad (2)$$

Using the pose estimate $\hat{\boldsymbol{X}}(t-1,10)$ at $t\tau$ (= $(t-1)\tau$ + $10\Delta\tau_{IMU}$), the scan data $\boldsymbol{p}_i(t-1,k,j)$ obtained by (2) is again transformed into the scan data $\boldsymbol{p}^*_{Hi}(t)$ in the helmet coordinate system at $t\tau$ by

$$\begin{pmatrix} \boldsymbol{p}^*_{Hi}(t) \\ 1 \end{pmatrix} = T(\hat{\boldsymbol{X}}(t-1,10))^{-1} \begin{pmatrix} \boldsymbol{p}_i(t-1,k,j) \\ 1 \end{pmatrix} \quad (3)$$

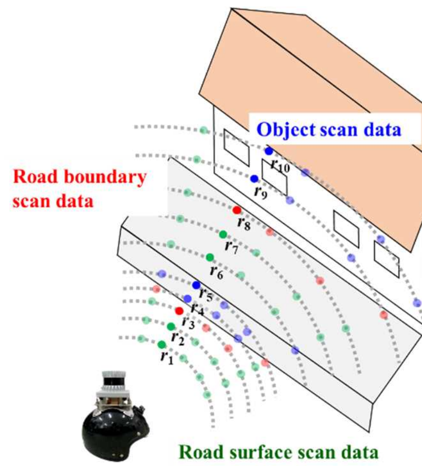
The $\boldsymbol{p}^*_{Hi}(t)$ denotes the scan data corrected for distortion at $t\tau$. Using the corrected scan data, environmental map building and moving object tracking are performed. In the EKF for correcting scan data distortion, a constant velocity model is used as the helmet motion (see Appendix).

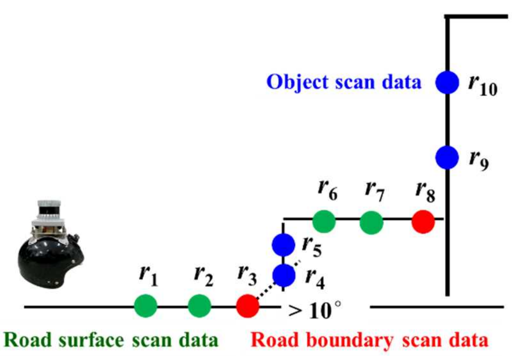
VI. Classification of LiDAR Scan Data

The distortion-corrected LiDAR scan data are classified into scan data on road surfaces, road boundaries, and 3D objects using a rule-based method. Following that, the 3D object scan data are classified into stationary and moving object scan data using map subtraction (MS)-based classification and occupancy grid methods.

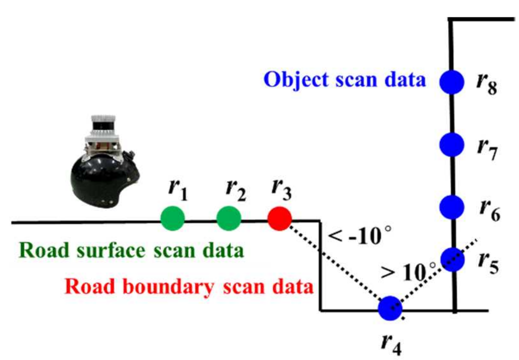
A. Classification of Road Surface, Road Boundary, and Object Scan Data

As shown in Figure 4 (a), 64 scan data points obtained for each 0.35° (vertical resolution) at a horizontal angle of the LiDAR laser beam, are arranged in order of decreasing elevation angle $r_1, r_2, r_3, \ldots$. First, we obtain the angle α of the line connecting points r_1 and r_2 relative to the xy plane in the world coordinate system. If $|\alpha| \le 10°$, r_2 is considered road surface scan data. This process is performed sequentially in other measurements r_3, …, r_n. When $\alpha > 10°$, the scan data r_4, r_5, r_9, and r_{10} in Figure 4 (a) are considered convex object data. Then, the scan data r_3 and r_8 are considered the road boundary scan data. On the other hand, if $\alpha <$-10°, as r_4 in Figure 4 (b), the scan data are considered concave object data, and r_3 is considered road boundary scan data. This process is performed on all LiDAR scan data, and only the object scan data are used for subsequent processes.





(a) Convex obstacle



(b) Concave obstacle

Figure 4. Classification of LiDAR scan data. The green, red, and blue circles indicate scan data on road surfaces, road boundaries, and objects, respectively.

Because the boundary between the road surface and a 3D object can be obtained for fallen objects on the road, the method described above can recognize obstacles on the road.

The angle threshold to classify the scan data is set to 10°. If it is small, road slopes will be mis-detected as 3D objects. Generally, the slope of steep roads for vehicles is approximately 6°. Thus, a threshold value of 10°, which is larger than 6°, is set.

B. Classification of Stationary and Moving Object Scan Data

For map building, moving object scan data have to be eliminated, and stationary object scan data have to be extracted from the entire LiDAR scan data. Moving object tracking conversely requires the removal of stationary object scan data and the extraction of moving object scan data from the entire LiDAR scan data. For this, accurate classification of stationary and moving object scan data is required. Although, in our previous work [15], the classification was performed using the occupancy grid method, LiDAR noises and outliers frequently result in misclassification.

To mitigate the misclassification, we compare the current object scan data with stationary object scan data in a map built up to the previous time. We call this approach MS-based classification or dynamic background subtraction-based method [14]. Figure 5 shows the MS-based classification method. In this method, we subtract the stationary object scan data in the map from the current object scan data to remove as much stationary scan data as possible from the object scan data.

The scan data extracted using the MS-based method are mapped onto a grid map. The cell on the grid map is a square with a side length of 0.3 m. A cell with scan data is called an occupied cell. For the moving object scan data, the time to occupy the same cell is short (less than 0.8 s in this paper), whereas for the stationary object scan data, the time is long (not less than 0.8 s). Therefore, by using the occupancy grid method based on the cell occupancy time [28], cells occupied by moving object scan data (or stationary object scan data) can be detected as moving cells (or stationary cells).

Because an object can occupy multiple cells, adjacent occupied cells are clustered. Then, clustered moving cells (or stationary cells) are obtained as a moving cell group (or stationary cell group). The scan data contained in the moving cell group are finally determined as the moving object scan data. The stationary object scan data are extracted by subtracting the moving object scan data from the current object scan data.

The LiDAR field of view also moves along with the micromobility movement. Although an object that has recently entered the LiDAR field of view is stationary, it is misclassified as a moving object because the cell occupancy time is short. To address this problem, new-observation cells are defined on the grid map, which corresponds to the new LiDAR field of view. The time of cells entering the LiDAR field of view (T_{NC}) and the cell occupancy time (T_{OC}) are measured, and the occupancy time rate (β) is calculated by $\beta = T_{OC} / T_{NC}$. Cells with β of 10% or more are determined to be new-observation cells and then considered moving cells. This can minimize the false classification of stationary objects that recently entered the LiDAR field of view as moving objects.

The scan data in the map are sparser in the areas in front of the micromobility and the occlusion areas. Therefore, the stationary object scan data likewise exist in a sparse state when the map is subtracted from the current object scan data. If the scan data, sparsely extracted using the MS-based method, are mapped onto a grid map, they may be erroneously determined as moving cells.

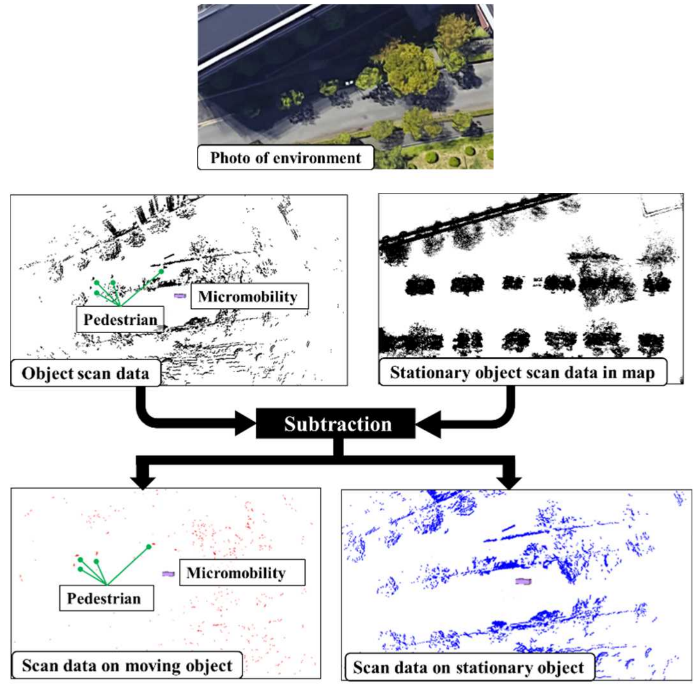


Figure 5. Sequence of MS-based classification (top view).

To overcome this issue, the scan data removed using the MS-based method are also mapped onto the grid map as stationary cells. As a result, sparse stationary object scan data that tend to be moving cells and stationary object scan data removed by the MS-based method are both mapped onto the grid map. The neighboring cells, which these stationary object scan data occupy, are clustered, and the cell group is then determined to be a stationary cell group. Accordingly, sparse stationary object scan data are correctly determined as stationary object scan data using the occupancy grid method.

In our preliminary experiments, LiDAR could correctly detect objects located within a range of approximately 50 m from the helmet. A grid map is therefore set up in ±35 m squares from the helmet; the distance from the helmet to the vertex of the square is approximately 50 m. Considering the resolution (0.35°) of the horizontal viewing angle of LiDAR, the cell size of the grid map is set to 0.3 m so that at least one measurement of LiDAR could be occupied in a cell 50 m away from the helmet. Assuming that the width and length of a pedestrian are 0.4 m and the walking speed is greater than 1 m/s, the time that the LiDAR measurements related to the pedestrian occupying a cell is less than 0.8 s. Therefore, the threshold of the occupation time to determine stationary and moving cells is set to 0.8 s.

VII. Road Boundary Estimation and Moving Object Tracking

In this section, the methods for estimating road boundaries and tracking moving objects are described using road boundary and moving object scan data, respectively.

A. Road Boundary Estimation

Although large boundaries between motor roads and sidewalks, such as grooves and curbs, can be detected every LiDAR scan period (100 ms), misdetection due to sensor errors and disturbances frequently occurs. Therefore, detection accuracy for road boundaries is improved using the following method. A grid map is created with the helmet position in the *xy*-plane in the world coordinate system as its center. Here, the cell on the grid map is a square with a side length of 0.5 m. Road boundary scan data obtained every LiDAR scan period (100 ms) are mapped onto the cell (referred to as road boundary cell) of the grid map.

The occupancy time (R_{OC}) is measured when road boundary scan data occupy each road boundary cell, and the unoccupied time (R_{UOC}) is also measured when they are unoccupied. If the total time (R_{OC} + R_{UOC}) is greater than 2 s and the ratio of occupation time to the total time, $R_{OC}/(R_{OC} + R_{UOC})$ is greater than 70 %, the scan data in the road boundary cell are considered road boundary scan data.

Because the width of grooves is generally 0.3–0.5 m, the size of the boundary cell is set to 0.5 m.

B. Moving Object Tracking [29]

In moving object tracking, the shape of the moving object is represented by a cuboid with a width W, a length L, and a height H, as depicted in Figure 6. The width W_{meas} and length L_{meas} of the moving object are extracted from the moving object scan data. With these values, W and L of the moving object are estimated by

$$\begin{cases} W(t) = W(t-1) + G(W_{meas} - W(t-1)) \\ L(t) = L(t-1) + G(L_{meas} - L(t-1)) \end{cases} \tag{4}$$

where G represents the gain.

The height estimate H of the moving object is obtained from the height measurements of the moving object scan data.

The Kalman filter is used to estimate the position and velocity of the moving object in the world coordinate system based on the centroid position of the rectangle estimated from (4). In crowded environments, the rule-based data association method is used to accurately match multiple moving objects with multiple moving scan data.

For the Kalman-filter-based tracking of moving object, it is assumed that the object moves at an approximately constant velocity. The motion model of the object is then given by

$$\boldsymbol{x}(t) = \begin{pmatrix} 1 & \tau & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \tau \\ 0 & 0 & 0 & 1 \end{pmatrix} \boldsymbol{x}(t-1) + \begin{pmatrix} \tau^2/2 & 0 \\ \tau & 0 \\ 0 & \tau^2/2 \\ 0 & \tau \end{pmatrix} \boldsymbol{\Delta x}(t-1) \tag{5}$$

where $\boldsymbol{x} = (x, \dot{x}, y, \dot{y})^T$. $(x, y)^T$ and $(\dot{x}, \dot{y})^T$ are the position and velocity, respectively. $\Delta\boldsymbol{x} = (\Delta\ddot{x}, \Delta\ddot{y})^T$ is an unknown acceleration (plant noise).

If the object moves with various different motions, such as moving at a constant speed, going or stopping suddenly, or turning suddenly, the use of multi-model-based tracking, such as an interacting-multiple-model estimator [30], will improve the tracking performance.

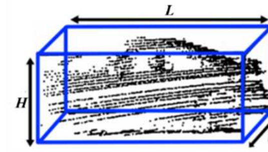


Figure 6. Cuboid around the tracked object (car).

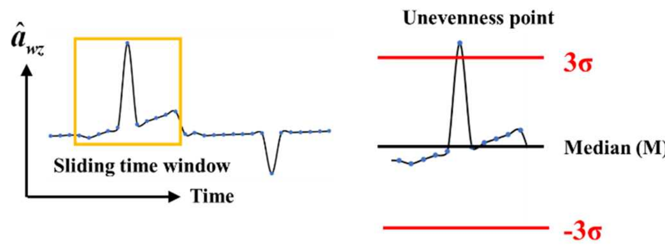


Figure 7. Threshold for road unevenness detection.

VIII. Detection of Road Unevenness and Estimation of Road Surface Condition

In this section, the methods for detecting small road unevenness and estimating road surface conditions with acceleration data from IMU are described. The related information can be used to prevent falling risks of micromobility and to improve rider comfort.

A. Detection of Small Road Unevenness

Because it is difficult to detect small unevenness on roads with LiDAR, such as dents, cracks, and manhole covers, they are detected using IMU data. The acceleration obtained from the IMU in the helmet coordinate system is denoted by $\boldsymbol{a}_H = (a_{Hx}, a_{Hy}, a_{Hz})^T$. Because the acceleration data include the gravitational acceleration G, the acceleration $\hat{a}_H$, where the gravitational acceleration is removed, is given by

$$\hat{\boldsymbol{a}}_H = \boldsymbol{a}_H - G \begin{pmatrix} -\sin\theta \\ \sin\phi\cos\theta \\ \cos\phi\cos\theta \end{pmatrix} \tag{6}$$

where ϕ and θ represent the roll and pitch angles, respectively, of the helmet obtained from the IMU in the world coordinate system.

Then, the acceleration $\hat{\boldsymbol{a}}_W = (\hat{a}_{Wx}, \hat{a}_{Wy}, \hat{a}_{Wz})^T$ in the world coordinate system is given by

$$\hat{\boldsymbol{a}}_W = \begin{pmatrix} \cos\theta & \sin\phi\sin\theta & \cos\phi\sin\theta \\ 0 & \cos\phi & -\sin\phi \\ -\sin\theta & \sin\phi\cos\theta & \cos\phi\cos\theta \end{pmatrix} \hat{\boldsymbol{a}}_H \tag{7}$$

When $|\hat{a}_{wz}|$ exceeds a threshold, it is assumed that the micromobility encounters small road unevenness. The threshold of unevenness detection is determined using a Hampel filter [31]. As shown in Figure 7, the acceleration data

$\hat{a}_{wz}$ are separated by a sliding-time window (window size of 5 s in this paper), and median (M) and standard deviation (σ) are calculated for the acceleration data in this window. From the calculation, the points $\hat{a}_{wz} \le M - 3\sigma$ or $\hat{a}_{wz} \ge M + 3\sigma$ are detected as road unevenness.

The locations of road unevenness in the world coordinate systems are determined based on the self-pose of the helmet.

B. Estimation of Road Surface Condition

Micromobilities travel on various roads, including non-paved roads, such as gravel roads, and paved roads that are not suitably maintained. Because the road surface condition is closely related to the safety of the micromobility, as well as rider comfort, it is also estimated from acceleration data to build a related map.

The International Roughness Index (IRI) [32] is used as an evaluation index for road surface conditions. To obtain IRI values, the vertical displacement caused by road unevenness, is calculated by integrating the acceleration data $\hat{a}_{wz}$. Because of the low-frequency noise of acceleration data, double integration of acceleration data with respect to time significantly leads to drift errors in the estimation of the vertical displacement.

To accurately obtain the vertical displacement, the acceleration data are integrated in the frequency domain, as follows. First, the acceleration data are passed through a Hanning window to remove sidelobes, and Fourier transform is performed. Following this, low-frequency noise is removed using a high-pass filter, and double integration is performed. The vertical displacement is then obtained by inverse Fourier transform.

The vertical displacement calculated from the acceleration data, which is obtained every IMU observation period (10 ms), is denoted by d_i. Then, the IRI value, J, for every S (10 m in this paper) of the traveled distance of the helmet is given by

$$J = \frac{1}{S} \sum_{i=1}^{n} |d_i| \tag{8}$$

where S is obtained from the self-pose of the helmet.

Note that the units of d_i, S, and J are mm, m, and mm/m, respectively. The higher the IRI value J, the rougher the road surface condition.

Because the IMU is mounted on the helmet, the head motion of the rider may affect the estimation of the vertical displacement. To reduce the effect of the rider's head motion as well as the low-frequency noise of acceleration data, the cut-off frequency of the high-pass filter is set to 10 Hz.

IX. Fundamental Experiments

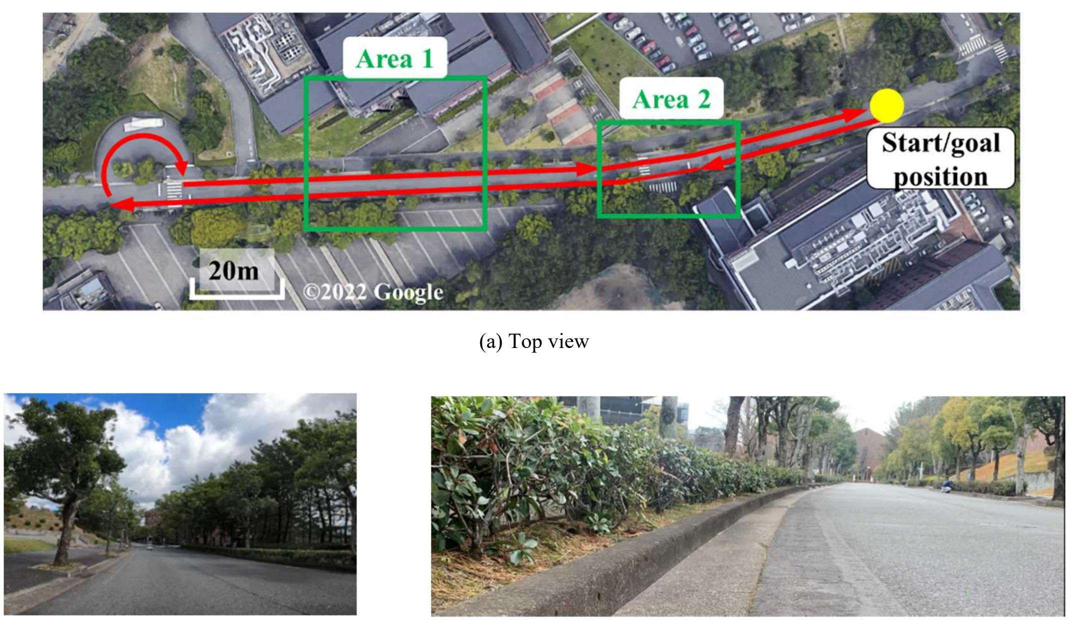
In this section, the performance of our method is evaluated through experiments in a road environment on our university campus.

A. Results of Map Building and Moving Object Tracking

A micromobility was moved on our university campus road, as shown in Figure 8, and environmental map building and moving object tracking with LiDAR scan data were performed. The traveled distance of micromobility is approximately 500 m, and the maximum speed is approximately 30 km/h. Figure 9 shows the attitude angle and angular velocity of the helmet during movement, which are observed by the IMU.

The experiments are conducted in the following two cases:

- Case 1: Map building and moving object tracking with the distortion correction of LiDAR scan data and the MS-based classification method (the proposed method)
- Case 2: Map building and moving object tracking without using either method.

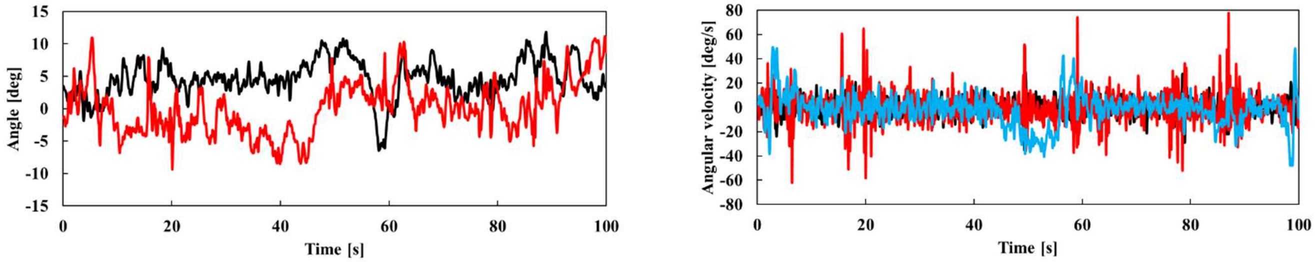


(a) Top view

(b) Side view

(c) Curbstone. The longitudinal offset from road is 150 mm.

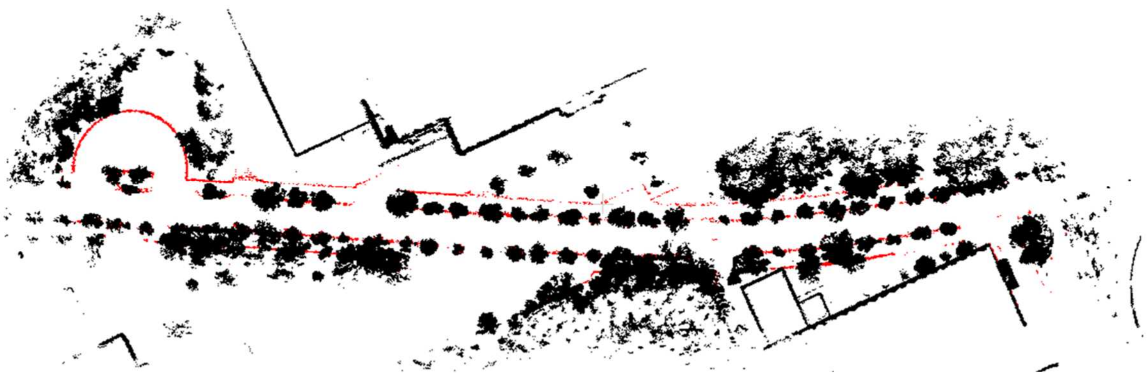
Figure 8. Photo of experimental environment.



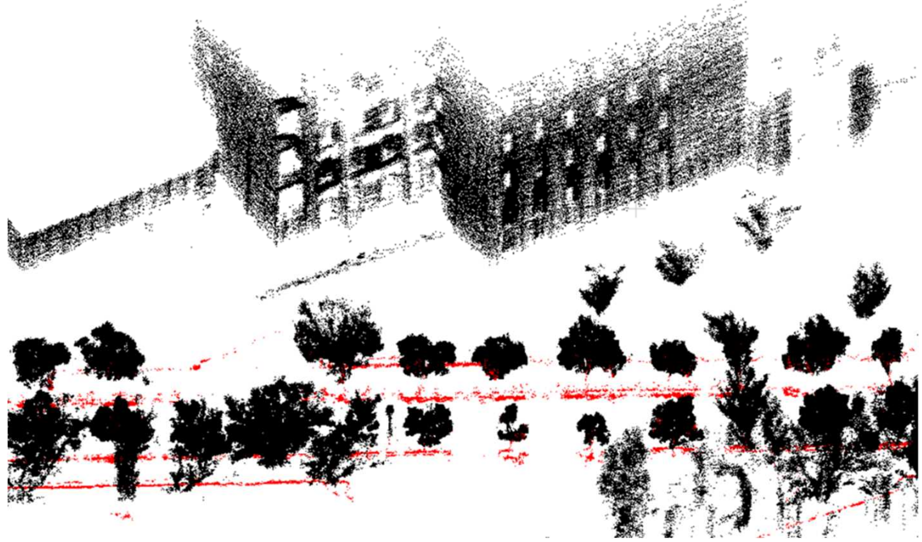
(a) Roll (black) and Pitch(red) angles

(b) Roll (black), pitch (red) and yaw (blue) angular velocities

Figure 9. Attitude angle and angular velocity of helmet.



(a) Overall map (top view)



(b) Enlarged map of area 1 (bird's-eye view)

Figure 10. Map-building result. The black dots indicate stationary object scan data, and the red dots indicate estimated road boundaries.

In case 2, although the distortion of LiDAR scan data is not corrected, the self-pose of the helmet is estimated using the EKF.

Figure 10 shows the result of map building obtained using the proposed method (case 1). This figure shows that the proposed method can build an environmental map.

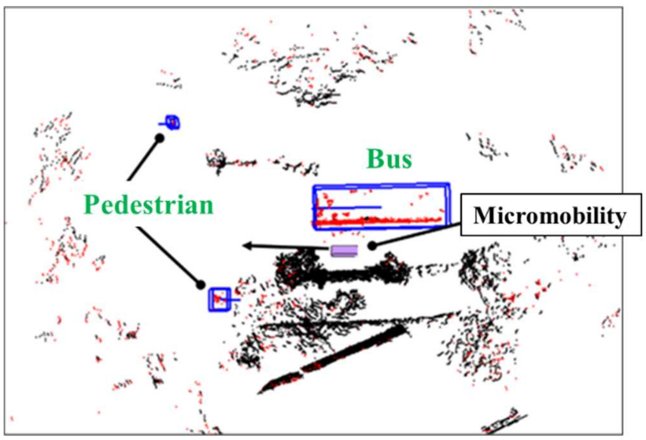
In our SLAM-based-map building method, the accuracy of map building is equivalent to that of the self-pose estimate of a helmet. Therefore, to evaluate the accuracy of map building, the error of position estimate of the helmet at the goal position is measured using a GNSS/LiDAR positioning system set at the goal position. Table I shows the result, where the micromobility is moved thrice on the road shown in Figure 8. According to the table, the proposed method (case 1) provides better map-building accuracy than case 2.

Figure 11 presents the result of moving object tracking in area 2 shown in Figure 8. In (b), the blue rectangle indicates the assessed size of the moving object, and the blue stick indicates

TABLE I. ERROR IN POSITION ESTIMATE OF HELMET AT GOAL POSITION.

	Experiment 1	Experiment 2	Experiment 3
Case 1	0.38 m	1.98 m	0.68 m
Case2	5.91 m	15.10 m	6.03 m

(a) Photo of area 2



(b) Estimated position and size of moving objects in area 2

Figure 11. Result of moving object tracking (top view).

the moving direction of the moving object obtained from the velocity estimate. The black (or red) dots indicate the scan data removed (or extracted) from the LiDAR scan data using the MS-based method.

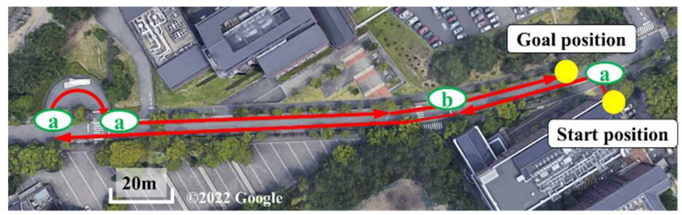
In experiments 1–3, there are 111 moving objects (106 pedestrians and five cars). As a result, in case 1 (proposed method), 109 objects (104 pedestrians and five cars) can be successfully tracked, and two pedestrians cannot be tracked. Conversely, in case 2, 103 objects (98 pedestrians and five cars) can be successfully tracked, and eight pedestrians cannot be tracked. From these results, our proposed method achieves better accuracy in moving object tracking.

Pedestrians who are not being tracked walk close to trees. In case 2, moving cells related to pedestrians are merged with adjacent stationary cells related to trees using the occupancy grid method, and pedestrians are falsely detected as stationary objects. On the other hand, because, in case 1, the stationary cells related to trees are removed using the MS-based classification method, pedestrians are detected correctly. Consequently, untracking of pedestrians occurs more often in case 2 than in case 1.

B. Results of Detection of Road Unevenness and Estimation of Road Surface Condition

A micromobility was moved on an asphalt-paved road on our university campus, as depicted in Figure 12. Four unevenness (three curbs and a crack) on the road shown in Figure 12 are detected, and road roughness condition is estimated. It is expected that the accuracy of detecting road unevenness and estimating road surface conditions is affected by rider weight, vehicle speed, vehicle dynamics, and so on. Therefore, the experiments are conducted under conditions of different rider weights (55, 65, and 76 kg) and micromobility velocities (10 and 15 km/h).

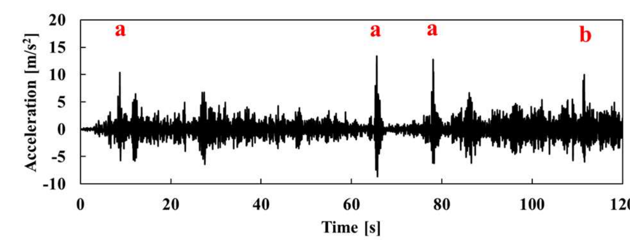
Figure 13 shows the vertical acceleration $\hat{a}_{wz}$ calculated from IMU data obtained while driving of the micromobility. Table II shows the success rate of detecting road unevenness. From the results, when the rider weight is light, the vertical acceleration to road unevenness becomes small, and the success rate of detection decreases.



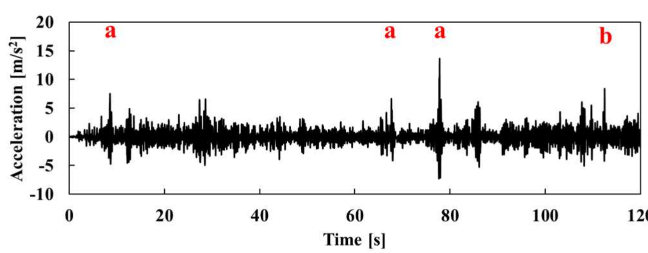
a) Curb (+-15mm)

b) Crack (-6mm)

Figure 12. Moved path of micromobility, and curb (a) and crack (b) on road. Numerals indicate the longitudinal offsets of the road unevenness.



(a) Rider weight of 75 kg, and micromobility velocity of 15 km/h



(b) Rider weight of 55 kg, and micromobility velocity of 15 km/h

Figure 13. Vertical acceleration. The "a" and "b" mean the times when the micromobility encounters on curb and crack, respectively, shown in Figure 12.

TABLE II. SUCCESS RATE OF ROAD UNEVENNESS DETECTION

		Road Unevenness	
		a: curb	***b: crack***
Rider Weight	***55 kg***	44.4 % (4/9)	66.7 % (2/3)
	65 kg	83.3 % (10/12)	100 % (4/4)
	75 kg	100 % (9/9)	66.7 % (2/3)
Total		76.7 % (23/30)	80.0 % (8/10)

*The “c” and “d” in c/d indicate the number of detected unevenness and the number of times when the micromobility encounters the unevenness, respectively.

Next, as shown in Figure 14, the path traveled by the micromobility is divided into 79 sections of 10 m each, and the IRI values are obtained, where the road unevenness (curb and crack) shown in Figure 12 are in sections 4, 44, 50, and 75. Figure 15 (a) shows the results. Because the IRI values for the rider with a weight of 55 kg are small and lead to inaccuracy, the mean of the IRI values for riders with weights of 65 and 75 kg is calculated. The mean is shown in Figure 15 (b). Figure 16 shows a map of the road surface condition, which is drawn based on the mean of the IRI values in Figure 15 (b).

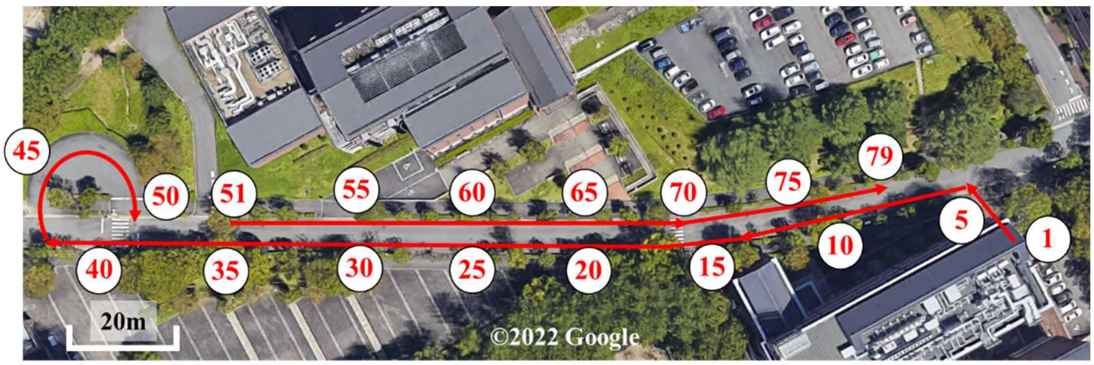
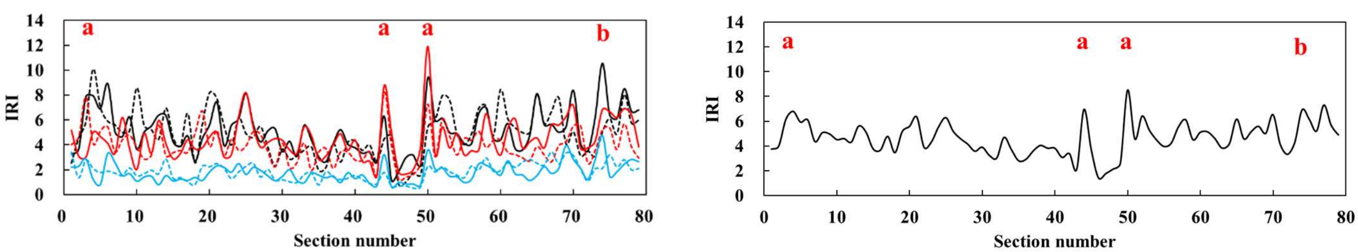


Figure 14. Section for IRI value estimation.



(a) IRI value by different rider weights and micromobility velocities

(b) Mean of IRI value

Figure 15. Estimation result of road surface conditions. In (a), black solid line (weight of 75 kg and velocity of 15 km/h), black dashed line (75 kg and 10 km/h), red solid line (65 kg and 15 km/h), red dashed line (65 kg and 10 km/h), light blue solid line (55 kg and 15 km/h), and light blue dashed line (55 kg and 10 km/h).

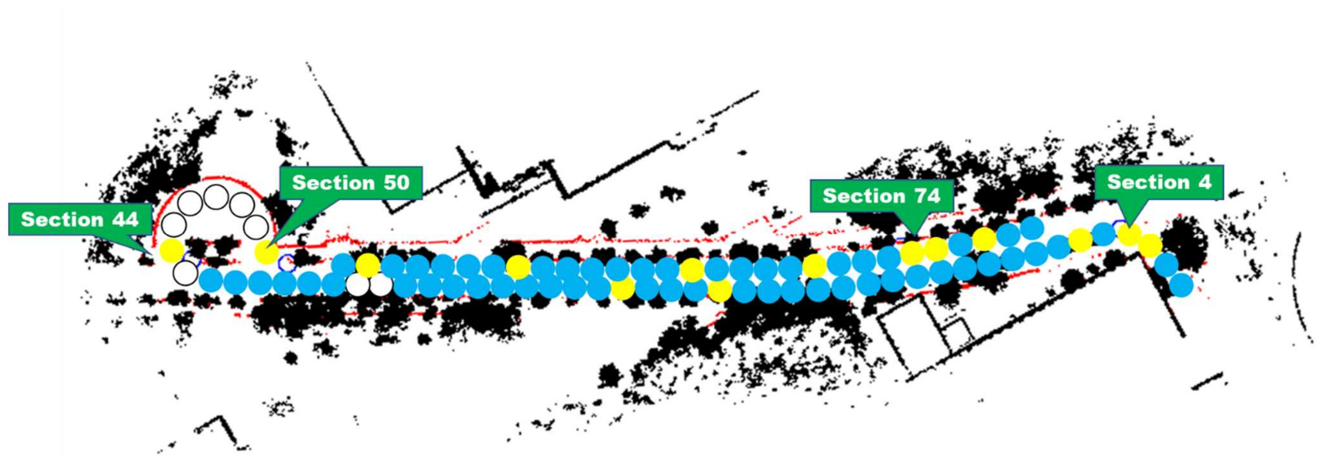


Figure 16. Map of road surface conditions. Different colored circles indicate IRI levels (white:0–3 mm/m, light blue: 3–6, and yellow:6–9).

(a) Section 25 (6.28 mm/m)

(b) Section 52 (6.40 mm/m)

(c) Section 70 (6.52 mm/m)

(d) Section 77 (7.29 mm/m)

(e) Section 13 (5.59 mm/m)

(f) Section 49 (2.57 mm/m)

(g) Section 61 (5.13 mm/m)

(h) Section 66 (4.56 mm/m)

Figure 17. Photo of road surfaces in eight sections. Numerals in brackets indicate the estimated IRI values. The IRI level of (a)–(d) is high (yellow circle in Figure 16), and that of (e)–(h) is low (white and light blue circles in Figure 16).

Because most of the sections on which the micromobility travels are flat asphalt-paved roads, the IRI values are small in most sections. Figure 17 shows the photos of road surfaces and the estimated IRI values in eight of 79 sections. The IRI values in sections 4, 44, 50, and 74, where there is road unevenness (curb and crack), as shown in Figure 12, are 6.78, 6.94, 8.49, and 6.90, respectively.

Because we currently do not have any professional instruments to accurately measure IRI values on roads, such as a car-mounted laser pavement scanner, the true IRI value is unknown, and the performance of our road surface condition estimation cannot be evaluated quantitatively.

The root mean square (RMS) values of the IRI values shown in Figure 15 (a) are 5.48 mm/m (rider weight of 75 kg and micromobility velocity of 15 km/h), 5.30 mm/m (75 kg and 10 km/h), 4.77 mm/m (65 kg and 15 km/h), 4.13 mm/m (65 kg and 10 km/h), 1.93 mm/m (55 kg and 15 km/h), and 1.98 mm/m (55 kg and 10 km/h). These RMS values and Table II show that the accuracy of the road surface condition map built using acceleration data is affected by rider weight and micromobility velocity. Eliminating those effects and improving map accuracy could be performed using a machine learning-based method, where the results of road unevenness detection and road surface condition estimation are collected from many micromobilities traveling on the same road. A related study in this regard is one of our future works.

The building of maps related to stationary objects, road boundaries, and road unevenness, as well as tracking of moving objects, should be performed in real time after each LiDAR scan data and IMU acceleration data are obtained. In our experiments, LiDAR scan data and IMU data are recorded, and map building and moving object tracking are performed offline using a computer. The specifications of the computer are as follows: Windows 10 Pro OS, Intel(R) Core (TM) i7-1065G7 @1.30GHz CPU, 16 GB RAM, and C++ software language. The point cloud library [33] is used for NDT scan matching.

The RMS values of the processing times are as follows:

- Distortion correction and NDT scan matching: 4048 ms
- Classification of LiDAR scan data: 2153 ms
- Map update of stationary objects and road boundaries: 39 ms
- Moving object tracking: 247 ms
- Detection of road unevenness: 3 ms
- Total: 6317 ms

Although a long computational time is currently required for map building and moving object tracking, the computational time can be reduced by optimizing the program code and using a graphical processing unit for real-time operations.

X. Conclusion and Future Work

This paper presented a method of experimental map building and moving object tracking using LiDAR and IMU attached to a helmet worn by a rider of micromobility. To accurately perform LiDAR-based environmental mapping and moving object tracking, the distortion of scanning LiDAR data was corrected using the self-pose information by NDT scan matching and IMU information using the EKF.

The distortion-corrected LiDAR scan data were classified into different data types to build a map composed of scan data relating to stationary objects and road boundaries and to track moving objects. Furthermore, road unevenness was detected, and road surface conditions were estimated using acceleration information from the IMU, and the related map was built to reduce the falling risk of micromobility and to provide comfort of the rider. The performance of the presented method was examined through experiments in a road environment on our university campus.

In future works, experiments in various sidewalk and roadway environments will be conducted to thoroughly evaluate the proposed method. The accuracy of the environmental map will be improved by collecting and processing map information obtained from many micromobilities. In addition, the realization of environment map building and moving object tracking using small and lightweight solid-state LiDAR instead of the mechanical LiDAR used in this paper is an important future direction.

APPENDIX: MOTION MODEL OF HELMET

As shown in Figure A, the translational velocity of the helmet in the helmet coordinate system (O_H-$x_H y_H z_H$) is denoted by (V_x, V_y, V_z), and the angular velocity (roll, pitch, and yaw angular velocities) by ($\dot{\phi}_H, \dot{\theta}_H, \dot{\psi}_H$).

The following motion model of the helmet can be obtained assuming that the helmet moves at nearly constant translational and rotational velocities:

$$
\begin{pmatrix} x_{(t+1)} \\ y_{(t+1)} \\ z_{(t+1)} \\ \phi_{(t+1)} \\ \theta_{(t+1)} \\ \psi_{(t+1)} \\ V_{x(t+1)} \\ V_{y(t+1)} \\ V_{z(t+1)} \\ \dot{\phi}_{H(t+1)} \\ \dot{\theta}_{H(t+1)} \\ \dot{\psi}_{H(t+1)} \end{pmatrix}
=
\begin{pmatrix}
x_{(t)} + a_{1(t)}\cos\theta_{(t)}\cos\psi_{(t)} \\
\quad + a_{2(t)}\left\{\sin\phi_{(t)}\sin\theta_{(t)}\cos\psi_{(t)} - \cos\phi_{(t)}\sin\psi_{(t)}\right\} \\
\quad + a_{3(t)}\left\{\cos\phi_{(t)}\sin\theta_{(t)}\cos\psi_{(t)} + \sin\phi_{(t)}\sin\psi_{(t)}\right\} \\
y_{(t)} + a_{1(t)}\cos\theta_{(t)}\sin\psi_{(t)} \\
\quad + a_{2(t)}\left\{\sin\phi_{(t)}\sin\theta_{(t)}\sin\psi_{(t)} + \cos\phi_{(t)}\cos\psi_{(t)}\right\} \\
\quad + a_{3(t)}\left\{\cos\phi_{(t)}\sin\theta_{(t)}\sin\psi_{(t)} - \sin\phi_{(t)}\cos\psi_{(t)}\right\} \\
z_{(t)} - a_{1(t)}\sin\theta_{(t)} + a_{2(t)}\sin\phi_{(t)}\cos\theta_{(t)} \\
\quad + a_{3(t)}\cos\phi_{(t)}\cos\theta_{(t)} \\
\phi_{(t)} + a_{4(t)} + \left\{a_{5(t)}\sin\phi_{(t)} + a_{6(t)}\cos\phi_{(t)}\right\}\tan\theta_{(t)} \\
\theta_{(t)} + \left\{a_{5(t)}\cos\phi_{(t)} - a_{6(t)}\sin\phi_{(t)}\right\} \\
\psi_{(t)} + \left\{a_{5(t)}\sin\phi_{(t)} + a_{6(t)}\cos\phi_{(t)}\right\}\dfrac{1}{\cos\theta_{(t)}} \\
V_{x(t)} + \tau w_{\dot{V}x} \\
V_{y(t)} + \tau w_{\dot{V}y} \\
V_{z(t)} + \tau w_{\dot{V}z} \\
\dot{\phi}_{H(t)} + \tau w_{\ddot{\phi}_H} \\
\dot{\theta}_{H(t)} + \tau w_{\ddot{\theta}_H} \\
\dot{\psi}_{H(t)} + \tau w_{\ddot{\psi}_H}
\end{pmatrix}
\qquad \text{(A)}
$$

where (x, y, z) and (ϕ, θ, ψ) denote the position and attitude angle (roll, pitch, and yaw angles), respectively, of the helmet in the world coordinate system (O_w-$x_w y_w z_w$). $(w_{\dot{V}x}, w_{\dot{V}y}, w_{\dot{V}z}, w_{\ddot{\phi}_H}, w_{\ddot{\theta}_H}, w_{\ddot{\psi}_H})$ denotes the acceleration disturbance. τ denotes the sampling period of LiDAR scan data and IMU data. $a_1 = V_x\tau + \tau^2 w_{\dot{V}_x}/2$, $a_2 = V_y\tau + \tau^2 w_{\dot{V}_y}/2$, $a_3 = V_z\tau + \tau^2 w_{\dot{V}_z}/2$, $a_4 = \dot{\phi}_H\tau + \tau^2 w_{\ddot{\phi}_H}/2$, $a_5 = \dot{\theta}_H\tau + \tau^2 w_{\ddot{\theta}_H}/2$, and $a_6 = \dot{\psi}_H\tau + \tau^2 w_{\ddot{\psi}_H}/2$.

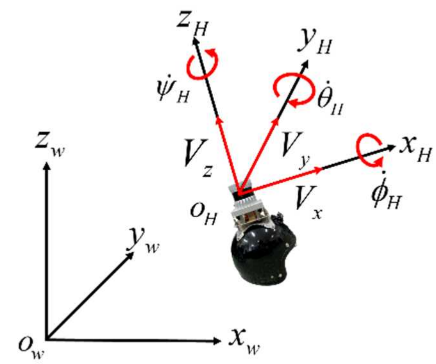


Figure A. Notation for helmet motion.

ACKNOWLEDGMENT

This study was partially supported by the KAKENHI Grant #23K03781, the Japan Society for the Promotion of Science (JSPS).

REFERENCES

[1] I. Yoshida, A. Yoshida, M. Hashimoto, and K. Takahashi, "Simultaneous Localization, Mapping and Moving-Object Tracking Using Helmet-Mounted LiDAR for Micromobility," Proc. the 13th Int. Conf. on Sensor Device Technologies and Applications, pp. 25–31, 2022.

[2] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A Survey of Autonomous Driving: Common Practices and Emerging Technologies," IEEE Access, vol. 8, pp. 58443–58469, 2020.

[3] V. Balaska et al., "A Viewpoint on the Challenges and Solutions for Driverless Last-Mile Delivery," Machines 2022, 10, 1059, 2022.

[4] B. Huang, J. Zhao, and J. Liu, "A Survey of Simultaneous Localization and Mapping," eprint arXiv:1909.05214, 2019.

[5] S. Kuutti et al., "A Survey of the State-of-the-Art Localization Techniques and Their Potentials for Autonomous Vehicle Applications," IEEE Internet of Things Journal, vol.5, pp. 829–846, 2018.

[6] A. Mukhtar, L. Xia, and TB. Tang, "Vehicle Detection Techniques for Collision Avoidance Systems: A Review," IEEE Trans. Intelligent Transportation Systems, vol. 16, pp. 2318–2338, 2015.

[7] Á. Llamazares, E. J. Molinos, and M. Ocaña, "Detection and Tracking of Moving Obstacles (DATMO): A Review," Robotica, vol. 38, pp. 761–774, 2020.

[8] E. Marti, J. Perez, MA. Miguel, and F. Garcia, "A Review of Sensor Technologies for Perception in Automated Driving," IEEE Intelligent Transportation Systems Magazine, pp. 94–108, 2019.

[9] Q. Chena, Y. Xiea, S. Guob, J. Baic, and Q. ShudaKey, "Sensing System of Environmental Perception Technologies for Driverless Vehicle: A Review of State of the Art and Challenges," Sensors and Actuators A: Physical, vol. 319, 112566, 2021.

[10] B. Sengul and H. Mostofi, "Impact of E-Micromobility on the Sustainability of Urban Transportation-A Systematic Review," Applied Science. 2021, 11(13), 5851, 2021.

[11] S. Tanaka, C. Koshiro, M. Yamaji, M. Hashimoto, and K. Takahashi, "Point Cloud Mapping and Merging in GNSS-Denied and Dynamic Environments Using Only Onboard Scanning LiDAR," Int. J. Advances in Systems and Measurements, vol. 13, pp. 275–288, 2020.

[12] K. Matsuo, A. Yoshida, M. Hashimoto, and K. Takahashi, "NDT Based Mapping Using Scanning Lidar Mounted on Motorcycle,"

Proc. the Fifth Int. Conf. on Advances in Sensors, Actuators, Metering and Sensing, pp. 69–75, 2020.

[13] S. Sato, M. Hashimoto, M. Takita, K. Takagi, and T. Ogawa, "Multilayer Lidar-Based Pedestrian Tracking in Urban Environments," Proc. IEEE Intelligent Vehicles Symp., pp. 849–854, 2010.

[14] S. Muro, I. Yoshida, M. Hashimoto, and K. Takahashi, "Moving-Object Tracking by Scanning LiDAR Mounted on Motorcycle Based on Dynamic Background Subtraction," Artificial Life and Robotics, vol. 26, issue 4, pp. 412–422, 2021.

[15] I. Yoshida, A. Yoshida, M. Hashimoto, and K. Takahashi, "Map Building Using Helmet-Mounted LiDAR for Micro-Mobility," Artificial Life and Robotics, vol. 28, issue 2, pp. 471–482, 2023.

[16] Y. Cai, S. Hackett, G., Ben, F. Alber, and S. Mel, "Heads-Up Lidar Imaging with Sensor Fusion," Electronic Imaging, The Engineering Reality of Virtual Reality 2020, pp. 338-1–338-7, 2020.

[17] B. Cinaz and H. Kenn, "Head SLAM- Simultaneous Localization and Mapping with Head-Mounted Inertial and Laser Range Sensors," Proc. 12th IEEE Int. Symp. on Wearable Computers, 2008.

[18] H. Sadruddin, A. Mahmoud, and M. M. Atia, "Enhancing Body-Mounted LiDAR SLAM using an IMU-based Pedestrian Dead Reckoning (PDR) Model," Proc. 2020 IEEE 63rd Int. Midwest Symp. on Circuits and Systems, 2020.

[19] E. Niforatos, I. Elhart, A. Fedosov, and M. Langheinrich, " s-Helmet: A Ski Helmet for Augmenting Peripheral Perception," Proc. the 7th Augmented Human Int. Conf.,2016.

[20] K. Zang, J. Shen, H. Huang, M. Wan, and J. Shi, "Assessing and Mapping of Road Surface Roughness based on GPS and Accelerometer Sensors on Bicycle-mounted Smartphones," Sensors, 2018, 18, 914, 2018.

[21] S. Cafiso, A. D. Graziano, V. Marchetta, and G. Pappalardo, "Urban Road Pavements Monitoring and Assessment Using Bike and E-scooter as Probe Vehicles," Case Studies in Construction Materials 16, 2022.

[22] K. R. Opara, K. Brzezinski, M. Bukowicki, and K. Kaczmarek-Majer, "Road Roughness Estimation Through Smartphone-Measured Acceleration," IEEE Intelligent Transportation Systems Magazine, pp. 209–220, 2022.

[23] W. Titov and T. Schlegel, "Monitoring Road Surface Conditions for Bicycles – Using Mobile Device Sensor Data from Crowd Sourcing," HCI in Mobility, Transport, and Automotive Systems, pp. 340–356, 2019.

[24] V. Douangphachanh and H. Oneyama, "Formulation of a Simple Model to Estimate Road Surface Roughness Condition from Android Smartphone Sensors," Proc. 2014 IEEE Ninth Int. Conf. on Intelligent Sensors, Sensor Networks and Information, 2014.

[25] A. Pangestu, M. N. Mohammed, S. Al-Zubaidi, S. H. K. Bahrain, and A. Jaenul, "An Internet of Things Toward a Novel Smart Helmet for Motorcycle: Review," AIP Conf. Proceedings 2320, 050026, 2021.

[26] T. Raj, F. H. Hashim, A. B. Huddin, M. F. Ibrahim, and A. Hussain, "A Survey on LiDAR Scanning Mechanisms," Electronics, vol. 9, 2020.

[27] P. Biber and W. Strasser, "The Normal Distributions Transform: A New Approach to Laser Scan Matching," Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 2743–2748, 2003.

[28] M. Hashimoto, S. Ogata, F. Oba, and T. Murayama, "A Laser Based Multi-Target Tracking for Mobile Robot," Intelligent Autonomous Systems 9, pp. 135–144, 2006.

[29] Y. Tamura, R. Murabayashi, M. Hashimoto, and K. Takahashi, "Hierarchical Cooperative Tracking of Vehicles and People Using Laser Scanners Mounted on Multiple Mobile Robots," Int. J. Advances in Intelligent Systems, vol. 10, no. 1 & 2, pp. 90–101, 2017.

[30] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting Multiple Model Methods in Target Tracking: A Survey," IEEE Trans. Aerospace and Electronic Systems, vol.34, pp.103–123, 1998.

[31] R. Wicklin, "The Hampel Identifier: Robust Outlier Detection in a Time Series," https://blogs.sas.com/content/iml/2021/06/01/hampel-filter-robust-outliers.html, 10 September, 2022.

[32] O. G. Dela Cruz, C. A. Mendoza, and K. D. Lopez, "International Roughness Index as Road Performance Indicator: A Literature Review," IOP Conf. Series: Earth and Environmental Science, vol. 822, 2021.

[33] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," Proc. 2011 IEEE Int. Conf. on Robotics and Automation, 2011.

Mapping of a Low-Textured Environment Using Visual Simultaneous Localization and Mapping to Use Augmented Reality Simulation for Testing Advanced Driver Assistance Systems in Future Automotive Vehicles

Michael Weber
Institute of Energy Efficient Mobility
Hochschule Karlsruhe -
University of Applied Sciences, HKA
Karlsruhe, Germany
email: michael.weber@h-ka.de

Tobias Weiss
Institute of Energy Efficient Mobility
Hochschule Karlsruhe -
University of Applied Sciences, HKA
Karlsruhe, Germany
email: tobias.weiss@h-ka.de

Franck Gechter
CIAD (UMR 7533)
Univ. Bourgogne Franche-Comte, UTBM
Belfort, France

LORIA-MOSEL (UMR 7503)
Université de Lorraine
Nancy, France
email: franck.gechter@utbm.fr

Reiner Kriesten
Institute of Energy Efficient Mobility
Hochschule Karlsruhe -
University of Applied Sciences, HKA
Karlsruhe, Germany
email: reiner.kriesten@h-ka.de

***Abstract*—Taking advantage of Advanced Driver Assistance Systems (ADAS) testing in simulation and reality, this paper presents a new approach to using Augmented Reality (AR) to test ADAS. Our procedure creates a link between simulation and reality besides existing methods like Vehicle in the Loop (ViL) and should enable a faster development process for ADAS tests, which will become increasingly complex. High computer power is needed for complex automotive environmental conditions, such as high vehicle speed and fewer orientation points on a test track compared to AR applications inside a building. A three-dimensional model with accurate information about the urban test site is generated based on the combination of Image Segmentation (IS), Artificial Intelligence (AI) for object recognition, and Visual Simultaneous Localization and Mapping (vSLAM). The use of AI and IS aims to significantly improve performance, such as robustness, calculation speed, and accuracy for AR applications in complex automobiles. Another focus of this work is to make the relocalisation stable even in low-texture environments.**

***Index Terms*—Artificial Intelligence; Augmented Reality; Advanced Driver Assistance Systems; Visual Simultaneous Localization and Mapping; Oriented FAST and BRIEF**

I. Introduction

Advanced Driver Assistance Systems (ADAS), such as the active lane departure warning (LDW)-system and traffic sign recognition support the driver, offer comfort, and take responsibility for increasing road safety. These complex systems endure an extensive testing phase resulting in optimization potential regarding quality, reproducibility, and costs. ADAS of the future will support ever-larger proportions of driving situations in increasingly complex scenarios. Due to the increasing complexity of vehicle communication and the rising demands on these systems in terms of reliability to function safely even in a complex environment and to support the driver and increase safety, the test scenarios for ADAS are constantly further developed and adapted to higher requirements. European New Car Assessment Programme (Euro NCAP) has therefore introduced a series of new safety tests for ADAS into its program and created a road map until the year 2025 [2] [3].

Current testing methods for ADAS can be divided into simulation and reality. The core concept behind simulation is to replicate the vehicle behaviour as realistically as possible in virtual test drives. The goal of this approach is to leverage its benefits, such as reproducibility, flexibility, and cost reduction, by testing and evaluating specifications and solutions in the early stages of development. Suitable simulation methods allow for the efficient design, development, and implementation of vehicles and vehicle components. However, simulation cannot yet completely replace real-world tests. Physical conditions, such as weather, road surface, and other variables, play a crucial role in evaluating ADAS road tests and cannot be fully replicated in a virtual environment [4] [5].

However, the test and evaluation effort correlate with the complexity of an ADAS. The more complex the system, the greater the testing effort. The robustness, functional safety, and reliability of the ADAS must be proven in increasingly dynamic, complex, and chaotic traffic situations. This also includes the interaction with different road users, each with their natural movements, such as, e.g., the interaction of road users with each other. MAURER and STILLER say *"If testing and assessment methods cannot keep pace with this functional growth, they will become the bottleneck of the introduction of Advanced Driver Assistance Systems to the market."* [2] Therefore, new and efficient test methods are required to pave the way for future ADAS [6]. A new approach *Vehicle in the Loop (ViL)*, which is already being used in the industry today, combines the advantages of simulation and reality. The approach in this paper is a new method besides existing ViL-solutions.

This paper describes the usage of Augmented Reality (AR) for testing camera-based ADAS. Section II gives an introduction to the field of AR. Based on this, we describe the different single modules of Simultaneous Localization and Mapping (SLAM) to map the environment in Section III. Section IV gives a short Introduction to ADAS. Section V describes the current methods used for testing ADASs. In Section VI we describe a possible usage of AR to test ADAS and its challenges in detail. The paper concludes with Section VII, where the research, results, and lessons learned are discussed.

II. Augmented Reality

According to Azuma's proposal, AR can be defined as a combination of three fundamental characteristics: the combination of real and virtual worlds and the precise three-dimensional registration of real and virtual objects, both in a real-time interactive environment [7]. The basic principle of AR is mainly known from the mobile game Pokémon Go [8]. Within this game, users can interact with digital creatures through their smartphones. These creatures are placed virtually in the user's environment. One such AR application is shown in Figure 1. Figure 1 shows also a self-created AR-App for demonstrating a possible scenery with traffic signs and a pedestrian. The three parts of the algorithms behind AR are image analysis, 3D modeling, and augmentation [1].

Image analysis serves to identify points or areas of interest within the given image. Feature detection, such as corner detection is often used for this step [9]. A three-dimensional model of the environment is created using the results of the image analysis. The types of algorithms used for this step vary

Fig. 1. Pokémon Go-App on the left side of the figure [8] and a self-created Augmented Reality application showing a possible scenery on the right side of the figure [1].

depending on the type of AR application. SLAM or Structure-from-Motion (SfM) algorithms are often used for AR in unknown locations [9]. The augmentation is based on the results of 3D modelling. The scene model is typically provided as a positional description of a plane or coordinate system that represents the real world [9]. With this information, a virtual object can be placed on the plane or in the coordinate system with appropriate characteristics, such as size and orientation. After object placement, the virtual content is combined with the real image [1] [9].

There are different versions of applications for AR. These applications are very diverse in their fields, from the use of AR in psychology [7] to use in hospital operating rooms [9] to mobile games [9] to military applications [9]. What all these apps have in common is that human reality is expanded. With humans as users of AR, there are implications for the application. One is that, in most cases, the human user is forgiving of not accurately placed virtual objects, if the error lies within a small margin. In addition, the speed of human movement, and therefore the distance covered in a given time, is limited. Because of these limitations, localization, mapping, object placement, and runtime requirements are not as strict and demanding as in the automotive environment given in this paper [1].

In order to use AR, the entire system must first be able to orient itself in the environment in order to finally augment virtual objects on reality. SLAM is a possible approach for this and is described in more detail in the following section.

III. Simultaneous Localization and Mapping

SLAM, or Simultaneous Localization and Mapping, is a method for determining the 3D structure of an unknown environment and the movement of sensors within it. Originally developed for autonomous robot control, the use of SLAM has since expanded to various applications, such as online 3D modeling using computer vision, AR visualization, and self-driving cars [10]. Early SLAM algorithms utilized multiple sensor types, such as laser range sensors, rotary encoders, inertial sensors, GPS, and cameras. However, more recent developments focus on using only cameras due to the simplicity of the sensor setup and greater technical challenges. SLAM systems that still use visual information as input are known as visual SLAM, or vSLAM.

A. Visual Simultaneous Localization and Mapping

SLAM techniques that rely solely on visual information are called visual Simultaneous Localization and Mapping (vSLAM). These algorithms are widely used in computer vision, robotics, and AR, and are particularly useful for camera pose estimation in AR systems [10]. Because AR systems often require real-time processing on light portable devices, various low-computational vSLAM algorithms have been developed. These algorithms have applications beyond AR systems and are also valuable for unmanned autonomous vehicles in robotics [11]. Most vSLAM approaches are comprised of five technical modules: three basic modules and two additional modules. These modules will be briefly described.

B. Basic Modules of Visual Simultaneous Localization and Mapping

The basic modules of vSLAM are *Initialization*, *Tracking*, and *Mapping* and are shortly presented in the following:

- *Initialization* is a crucial step in vSLAM that involves defining a coordinate system for estimating camera position and reconstructing a 3D environment. During initialization, a global coordinate system must be established, and a part of the environment is reconstructed as an initial map in this system [12].
- *Tracking* is a component of vSLAM that follows the reconstructed map in an image to continuously estimate the camera position relative to the map. That is achieved by determining distinctive matches between the captured image and the created map using feature matching or feature tracking [12].
- *Mapping* involves expanding the map by determining the 3D structure of an environment when the camera encounters previously unmapped regions. It involves understanding and calculating the unknown parts of the environment [12].

C. Additional Modules of Visual Simultaneous Localization and Mapping

The basic modules of vSLAM are supplemented by the additional modules *Relocalization* and *Global Map Optimization*:

- *Relocalization* refers to determining the current camera position in a reconstructed map when tracking has failed, which can occur due to fast camera movements. That allows the system to recompute the camera's location [12].
- *Global Map Optimization*, which includes *Loop Closing*, is a method to refine the map by eliminating cumulative estimation errors that accumulate with camera movement. The map is optimized by considering the consistency of all map information. If previously recorded map elements are recognized, loops are closed, and the estimation error is corrected from the beginning to the present. Loop Closing is used to acquire reference information by comparing the current image to previously acquired images. Relocalization is used to recover the camera position, and Pose Graph Optimization is employed to suppress cumulative error by optimizing the camera positions. Bundle Adjustment (BA) is also used to minimize the map reprojection error by optimizing both the map and camera positions. In large environments, this optimization method is used to effectively minimize estimation errors, while in small environments, BA can be performed without loop closure as the cumulative error is small [12].

Different sensors, and combinations of sensors, can be used in vSLAM. These sensors will be described in the following section.

D. Sensors for Visual Simultaneous Localization and Mapping

We discuss the primary sensors for visual SLAM in this section.

1) Monocular Camera: Monocular cameras consist primarily of an image sensor and a lens. After removing lens distortion, the camera can be modeled as a pinhole camera [13], allowing a 3D point in the camera's coordinate reference system to be projected into 2D pixel coordinates. Most industrial cameras are global shutter cameras, which capture the entire image at once, while consumer cameras tend to be rolling shutter cameras that capture pixel rows at different times, which can affect accuracy in visual SLAM if not adequately modeled. Due to the nature of monocular cameras, they cannot accurately determine the actual scale of the world and therefore monocular SLAM can only estimate the map and camera trajectory up to scale. To resolve this issue, additional information, such as an Inertial Measurement Unit (IMU) or known distances in the map, is necessary to scale the solution correctly [14].

2) Stereo Camera: Stereo cameras consist of two cameras that are attached rigidly to each other. Ideally, they have synchronized for simultaneously capturing the images. The depth of an object can be estimated from a single stereo frame by finding correspondences between pixels in the left and right cameras. To achieve this, both cameras must be calibrated internally and the rotation and translation between both cameras must be calibrated using several stereo frames with a calibration pattern. The distance between the cameras, known as the baseline, along with the focal length and image resolution, determine the range of depth where depth estimation is accurate. The projection function for a rectified stereo camera maps a 3D point in the camera coordinate system to another 3D point [14].

3) RGB-D Camera: Red Green Blue-Depth (RGB-D) cameras consist of a color camera and a depth sensor that uses structured light or time-of-flight technology. By calibrating the intrinsic parameters of the camera and the extrinsic parameters between the color camera and depth sensor, the depth measurements can be transformed into a depth map with a

1:1 relationship with the color image, eliminating the need for stereo matching as with a stereo camera. However, their use is limited to indoor environments due to the nature of the depth sensor [14].

4) Inertial Measurement Unit: IMUs detect an object's motion by combining a gyroscope that measures angular velocity and an accelerometer that measures linear acceleration. They provide information about the object's self-motion, complementing what is seen by vision. IMUs can be used to determine the motion between camera frames or calculate the scale of monocular SLAM, and they can also estimate gravity, allowing for the calculation of absolute pitch and roll. They typically measure acceleration and angular velocity hundreds of times per second. To effectively use IMUs with vision, they should be synchronized and calibrated to match the reference frame of the camera [15].

E. Solving Visual Simultaneous Localization and Mapping

Exploiting information from a stream of images captured by a vision sensor is crucial for visual SLAM. There are two main approaches to this problem: feature-based and direct methods. When applied to automotive vehicles, vSLAM faces challenges such as fast scene changes and low environmental texturing. Several vSLAM algorithms are available and listed in [16], where they are compared in terms of accuracy and robustness, among other factors. In the following chapter, a specific vSLAM approach relevant to our research work is described.

1) Direct and Feature-Based Simultaneous Localization and Mapping: Feature-based methods analyze images to identify and extract unique, recognizable *keypoints*. These keypoints can be detected repeatedly and consistently in images of the same scene, even under varying viewpoints and lighting conditions. Each keypoint is then assigned a descriptor, a numerical representation used to match keypoints across images by comparing the descriptors. The keypoint and its descriptor make up a *feature*. Once features have been extracted, the original image is no longer necessary as the features are used for all subsequent processing. That is beneficial as features are easier to match and manipulate for solving geometry problems in visual SLAM, such as triangulation, epipolar geometry, Perspective-n-Point (PnP) problem, and transformations between reference systems. The direct approach in SLAM involves utilizing sensor measurements, such as pixel intensity in an image directly. It can be categorized into dense (using all pixels) [17], semi-dense (using pixels with high gradients) [18], or sparse (using only a few pixels) [19] methods. These direct methods are considered more accurate and reliable when there is minimal texture or blur in the image, as they do not depend on keypoint detectors. The objective of direct SLAM is to determine the depth of each pixel in selected cameras by minimizing photometric error through optimization [14].

2) Oriented FAST and Rotated BRIEF (ORB)-SLAM: The ORB-SLAM algorithm was first presented in 2015 and is the current state of the art as it has higher accuracy than comparable SLAM algorithms [20]. ORB-SLAM represents a complete SLAM system for monocular, stereo, and RGB-D cameras. The system operates in real-time and achieves remarkable results in terms of accuracy and robustness in a variety of different environments. ORB-SLAM is used for indoor sequences, drones, and cars driving through a city. The ORB-SLAM consists of three main parallel threads: *Tracking*, *Local Mapping*, and *Loop Closing*. It is possible to create a fourth thread to execute the *BA* after a closed loop. This algorithm is a feature-based approach that represents the detected points in a three-dimensional *Point Cloud* [16]. ORB-SLAM seems to be the best algorithm for our approach [1]. To use ORB-SLAM for testing ADAS using AR, the following chapter first provides an introduction to the topic of ADAS.

IV. Advanced Driver Assistance Systems

ADAS enhances the driving experience by offering assistance to the driver when operating a vehicle. Depending on the specific system, it can enhance comfort, improve safety, decrease energy usage, or streamline traffic flow. The technology uses sensors to monitor driving conditions, employs powerful computers to process the collected data, and provides feedback to the driver through visual, auditory, or touch signals. In some cases, ADAS can even take control of the vehicle by accelerating, braking, signaling, or steering, potentially leading to fully autonomous driving. For ADAS to be effective, it requires rapid data processing with near real-time response and a highly dependable system [21] [22]. The proliferation of environmental sensors, including radio detection and ranging (radar), cameras, ultrasonic sensors, and Light Detection and Ranging (LiDAR)-sensors, makes it possible to use ADAS and related autonomous driving functions in modern vehicles. However, each sensor has its limitations and cannot provide all of the necessary information about the vehicle's surroundings to guarantee safety. Only through the fusion of data from multiple sensors a complete environment model can be created, which is crucial for the dependability and safety of driver assistance systems and autonomous driving [23]. Both simulation and real-world testing are required for thoroughly evaluating individual ADAS functions [1]. The following section outlines the different testing options for ADAS.

V. Testing of Advanced Driver Assistance Systems

Simulative test procedures during the development process and test procedures, in reality, are used to evaluate the functionality of individual ADAS sensors and their joined interaction in ADAS-relevant scenarios. While in the early concept stage, all components of the road test are still virtual and characterized by test procedures such as:

- Model in the Loop (MiL)
- Software in the Loop (SiL) or
- Hardware in the Loop (HiL)

Through the various stages of development, a gradual exchange of virtual for the equivalent physical world test components enrolls. By the end, reality replaces the simulated elements completely [22]. Table I shows an overview of different stages for testing ADAS in simulation and in reality. In the following, an overview of the respective possibilities and the advantages and disadvantages in simulation and in reality is given.

A. Testing Advanced Driver Assistance Systems in Simulation

The objective of the virtual road test is to recreate the experience of an actual road test as closely as possible in a virtual environment. The goal is to take advantage of simulation, such as reproducibility, versatility, and ease of use, and to evaluate and test vehicle specifications and solutions early in the development process. By using appropriate simulation methods, vehicle and component design, development, and implementation can be made more efficient. These methods reduce the time it takes for real-world prototypes to become available. Optimizing simulation techniques based on actual driving tests and actual test results requires balancing modeling, parameterization, and simulation effort with the efficiency gained. The approach mainly employs methods from embedded mechatronic system development, including SiL, MiL, and HiL methods [4].

B. Testing Advanced Driver Assistance Systems in Reality

Validation of vehicle dynamic control systems, despite their complexity and wide range of variations, can be expensive to test in actual driving tests. However, this approach is not feasible for driver assistance systems with environmental perception due to the high-level system complexity, the complexity of test cases, and the extent of testing required. Even if the tests are performed identically, it is impossible to ensure that the same conditions are being tested, due to the numerous and unknown factors influencing the results. That makes the reproducibility of results uncertain. Function-relevant features may require the involvement of multiple road users and can also be affected by complex interactions of various conditions, such as glare from the sun and reflections on a wet road at a certain angle. Current ADAS access information about the environment, often gathered by multiple sensors with different functions and processed in an environment representation [5] [24]. As suggested in literature Euro NCAP is a European standard for actual ADAS driving tests, which focuses on the vehicle's behavior and response in safety-critical scenarios. Dangerous situations are simulated using dummy vehicles, pedestrians, and cyclists to evaluate the effectiveness of ADAS systems. To improve road safety, the requirements for ADAS are being increased, and as a result, Euro NCAP test procedures will become increasingly complex in the future. The roadmap for 2025 includes the inclusion of additional road users, such as scooters, motorcycles, and wild animals [26]. Figure 2 displays a test configuration for pedestrian detection. In the test, a pedestrian dummy (representing a child) is crossing the road behind a parked car, and the test vehicle must detect the pedestrian and apply the brakes to prevent personal injury or damage to property [25].

It is rarely possible to conduct tests with prototypes in real road traffic and test persons due to legal and safety restrictions. The validation of safety-critical functions such as Automatic Emergency Braking (AEB) for pedestrians cannot be replicated adequately, even within the framework of NCAP test scenarios on a test track. As the driving situation becomes more complex for the assistance system under test, it becomes increasingly challenging to realistically and reliably assess the interaction between system behavior and driver experience and behavior. ViL bridges the gap between driving simulation and real-world testing. By combining virtual visual representation with the experienced haptics, kinaesthetics, and acoustics from actual vehicle movement, ViL provides a new augmented reality-based approach to efficiently and safely develop and evaluate ADASs [5]. This approach is described in the following chapter.

C. Combining Virtual and Real Testing

To enable the testing of camera-based assistance systems in real environments earlier in the development phase and thus

TABLE I
OVERVIEW OF DIFFERENT STAGES FOR TESTING ADVANCED DRIVER ASSISTANCE SYSTEMS IN SIMULATION (VIRTUAL) AND IN REALITY. V: VIRTUAL, R: REALITY [22].

	MiL	SiL	HiL	Chassis Dynamometer	ViL	Road Test
Functional Code	V	R	R	R	R	R
ECU	V	V	R	R	R	R
System	V	V	R	R	R	R
Vehicle	V	V	V	R	R	R
Driver	V	V	V	V/R	V/R	R
Driving Dynamics	V	V	V	V	R	R
Perceptibility	V	V	V	V	R	R
Lane	V	V	V	V	R	R
Environment	V	V	V	V	V	R

Fig. 2. Test setup for pedestrian detection, where a pedestrian (child as a dummy) crosses the road behind a parked car in a scenario for testing Automatic Emergency Braking (AEB) based on Euro NCAP regulations [25].

increase the quality of the systems, the use of AR as a link between virtual and real testing lends itself to this. Using AR to test camera-based systems combines the advantages of a virtual environment and these of the real world: Reproducible, complex scenes with realistic environmental conditions. AR thus makes it possible to dispense with test dummies or second vehicles including drivers even in the initial phases of testing. This reduces the testing costs and increases the safety of the test engineers. The combination of different test situations is also possible: The display of several vehicles, lane markings, and road signs allow the simultaneous testing of all camera-based driver assistance systems. The unlimited variety of test scenarios allows a significant increase in the depth of testing at an early stage of development. This increases the quality of the testing and the overall system. In 2010, a Swedish team led by Jonas Nilsson presented a software framework at a conference that used AR to evaluate a pedestrian detection system. The framework was able to augment the images from the vehicle camera to include a walking pedestrian. The resulting detection system results were comparable to test results obtained with real obstacles. As summarized in this paper, deeper investigations are needed further advance an AR test system [10]. A different method than ours for combining virtual and reality testing is discussed in the subsection that follows.

D. Vehicle in the Loop

ViL stands for a newer method to meaningfully complement and improve the development of the V-Model for driver assistance systems. It addresses the need for many driver assistance functions for an elaborate driving test and a high demand for functional safety. This group of driver assistance functions will increasingly gain in importance and scope. One main reason is the ever-growing number of vehicle derivatives in which driver assistance-functions are offered, and the accompanying ever-higher level of automation and networking. The ViL method allows the operation of the test vehicle in a virtual environment. The coupling between the vehicle and the virtual environment can be done in two ways. One way is to replace the physical sensor system with an interface. At this interface, the simulation environment feeds in simulated sensor signals which correspond to the sensor response from a physical environment. The second way is to retain and artificially stimulate real sensor technology, as is feasible, for example, with ultrasonic sensor technology exposed to artificially generated response signals via ultrasonic transducers [27]. In both variants, the physical test vehicle reacts to features and events in the virtual environment. Critical driving manoeuvres towards obstacles or objects on a collision course can thus be tested safely and reproducibly. The interface created can also be used to generate the sensor signals in a way that would occur due to a changed position in a vehicle derivative or due to different tolerances. Thus, this method enables testing corresponding derivatives or tolerances with a test vehicle.

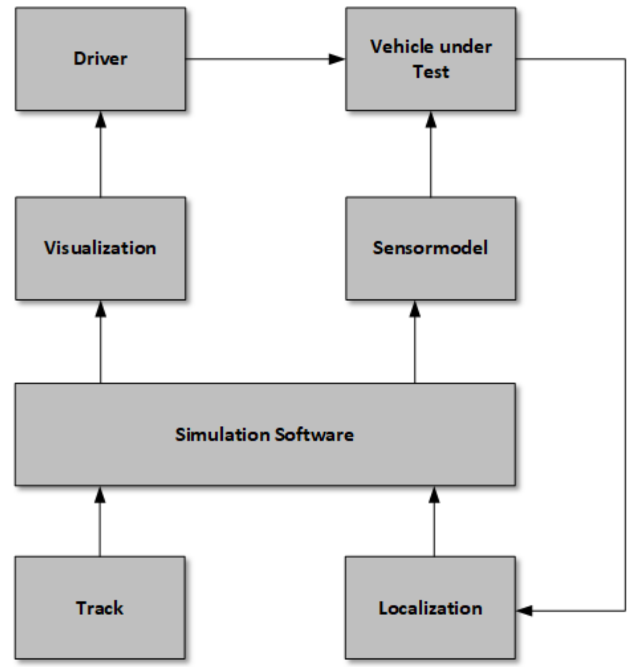


Fig. 3. ViL-architecture as well as the flow of information [22].

In addition to the considerably safer test operation, this allows effective testing and application of driver assistance-functions. That results in the considerable economic potential for driving tests in driver assistance. The use of virtual integration in conjunction with the ViL method allows efficient application of the customer function. The efficiency and reproducibility of the test cases required for this can thus increase significantly. Figure 3 illustrates the general operating principle of the ViL by the architecture as well as the flow of information [22]. The following section describes another approach besides existing ViL-Solutions using AR.

VI. Augmented Reality Simulation in Advanced Driver Assistance Systems

With a focus on the camera-based ADAS sensors, the area around the test field is recorded, as shown in Figure 4. The path between the sensor fusion module and the Electronic Control Unit for Advanced Driver Assistance Systems (ADAS-ECU), which causes the vehicle to intervene, for example by braking, has to be disrupted and a new path has to be found through the Augmented Reality Electronic Control Unit (AR-ECU). Within the AR-ECU, the captured environmental data is augmented with virtual objects, such as traffic signs or lane markings. The aim here, is a realistic and consistent behavior of the ADAS-ECU as in real object detection. For the final augmentation of the virtual objects on the real image of the sensor, a detailed 3D environment of the test environment must first be created. This section will describe the single steps as well as the results so far to use AR for testing ADAS.

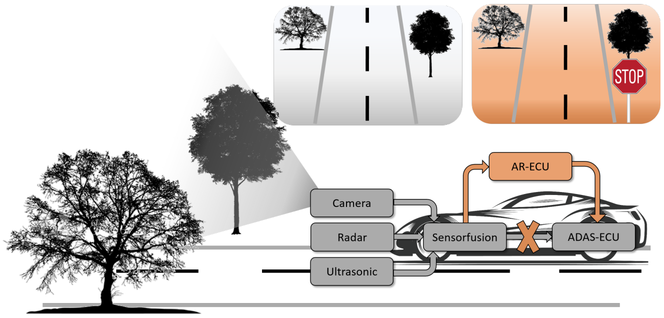


Fig. 4. Our approach for using augmented reality in advanced driver assistance systems.

A. Technical Steps to use Augmented Reality in Advanced Driver Assistance Systems

Figure 5 shows the technical steps of our approach to using AR in ADAS. The *Sensor Input* as a camera stream is transferred in greyscale to the element function block *ORB-SLAM3*. The keypoints are *tracked* and transferred to relevant *featurepoints*. These featurepoints are filtered by the Image Segmentation (IS) in the function *DS-SLAM*. The relevant featurepoints are merged into a *Point Cloud* in *Mapping*. *Loop Closing* and *Bundle Adjustment* are further elements of *ORB-SLAM3* to enable corrections of the Point Cloud and increase accuracy. The next function block is the *Point Cloud Preparation*. The first step is a *Preparation* of the Point Cloud to identify relevant points for *Plane Detection*. Based on these results, a plane is inserted on which virtual objects like e.g., a cyclist or traffic signs from the *Virtual Objects Library* can be placed. These objects can be static or animated. After the successful placement of the virtual elements, *rendering* is carried out in the function block *AR Viewer*, whereby the virtual objects are augmented with the real image stream. In the final step the *Output is Prepared* and adapted to the target medium as *Augmented Images*. The following subsections describe the single steps in a more detailed way.

B. Mapping of the Environment

The environment mapping is done using different stereo cameras such as the *ZED2i* equipped with a polar filter and a baseline of 120 mm, and the *Intel RealSense D455* with a baseline of 95 mm. The vSLAM approach ORB-SLAM3 is applied to these cameras. The results presented in the images were obtained using the *Intel Realsense D455*. The detected features are recorded in a three-dimensional Point Cloud. The test drives were performed on appropriate NCAP test areas, as shown in Figure 6. The impact of the low texture of the environment must be taken into consideration. To overcome the repetition of scene images, pylons are placed along the test track at intervals of 20 metres, alternating in number on either side of the track. That ensures that feature matching and proper orientation in the Point Cloud occur. The camera is mounted at the rear-view mirror height on the top of the windshield, as is typical for cars. Figure 6 displays the recorded test track scene. The rectangles represent the feature points detected

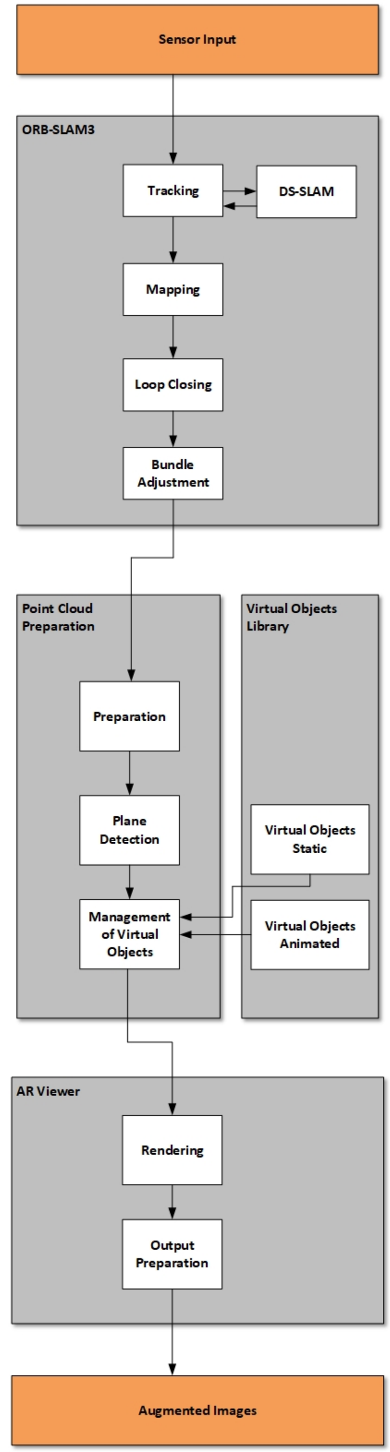


Fig. 5. Technical steps to use augmented reality in advanced driver assistance systems.

by the ORB-SLAM algorithm. The generated Point Cloud can be viewed in Figure 7 using the *RViz* tool. The Point Cloud depicts a straight road with pylons, road markings, and other objects such as trees on the left side or a hill on the right. Stereo-based cameras allow for accurate mapping of the environment with the correct scale. Approaches that

Fig. 6. Test area on the top of the figure with pylons and test vehicle as well as detected feature points using ORB-SLAM (rectangles) on the bottom.

utilize a combination of mono-based cameras and IMUs do not result in usable outcomes due to the lack of depth and scale information. As shown in Figure 6, feature points are detected in the sky as well as on the hood of the ego vehicle, which hinder the performance of our vSLAM algorithm. To address this issue, we decided to employ Image Segmentation (IS) to detect these false feature points.

C. Image Segmentation

We have selected *Deeplabv3+* developed by Google as the model for IS and semantic segmentation after evaluating various options. The pixellib library, a python tool for training and testing deep neural networks (NN), was utilized in conjunction with Deeplabv3+. This choice was based on Deeplabv3+ delivering the best results for our ORB-SLAM implementation, as it outperformed most recent NN models for semantic segmentation. For a comprehensive overview of different models and their performance, we refer to [28]. The Xception-65 model, with pre-trained weights from the ADE20k-Dataset, was utilized as the backbone. The pre-trained model provided 120 labels for indoor and outdoor environments with a color mapping for each label. The model

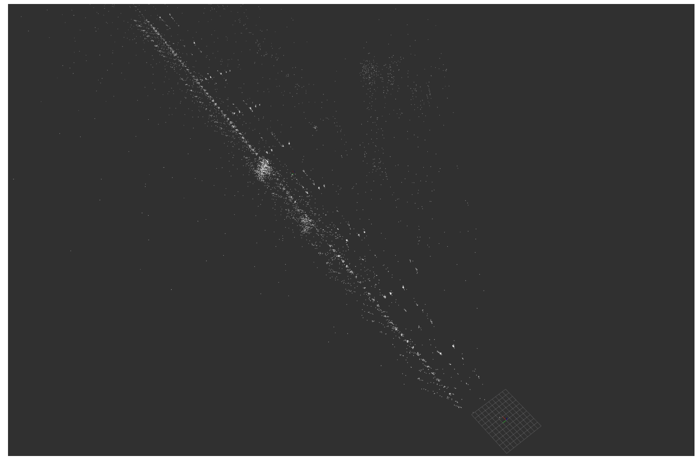


Fig. 7. Created three-dimensional Point Cloud based on detected feature points using ORB-SLAM.

performed so effectively on our test images and videos that additional custom training was not required, although it could be easily implemented for further development. As depicted in Figure 8, relevant feature points were successfully detected in the testing ground, while false feature points in the sky or on the car hood were disregarded. The use of this IS led to an improvement in the robustness of relocalization.

D. Map Preprocessing

The subsequent step is to fit a plane suitable for the Point Cloud. This plane ensures that virtual objects can be realistically inserted with the correct height and position in the scene image. As an initial attempt, we applied the RAndom SAmple Consensus (RANSAC)-algorithm to determine a plane on the road surface. Figure 9 illustrates that using the

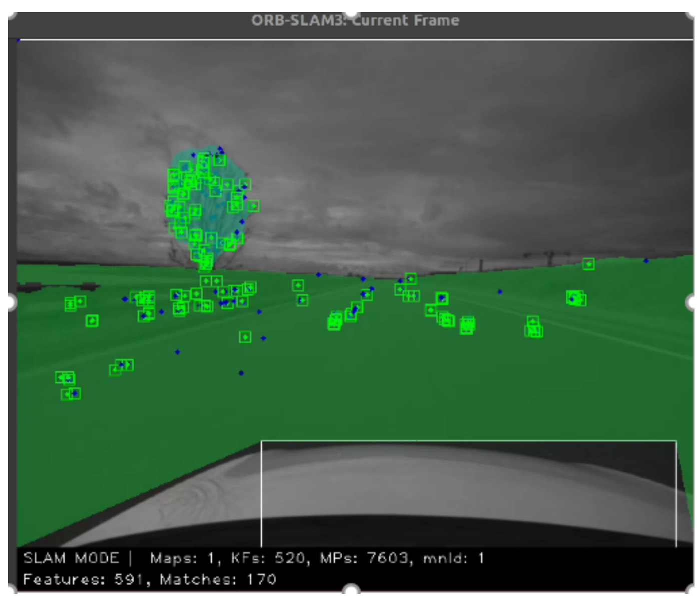


Fig. 8. Results of our image segmentation to detect feature points in relevant regions (green area).

entire Point Cloud results in an inclined plane due to the abundance of feature points on the hill and the scarcity of points on the street. To address this issue, we extract only the relevant feature points from the Point Cloud. That is achieved by using a cylindrical area around the camera trajectory to select the feature points on the street and eliminate irrelevant feature points outside the test track. We can define the cylinder by specifying 2 points along the axis, and the radius of the cylinder:

$A(x1, y1, z1)$
$B(x2, y2, z2)$
and radius = R

Consider the line coordinates with the direction

$$e = r_B - r_A \tag{1}$$

and moment

$$m = r_A \times r_B \tag{2}$$

These two vectors represent the infinite line between A and B. A point P with position rP lies in the cylinder between A and B and radius R if:

1. Distance of P to line AB is equal or less than R:

$$d = \frac{||m + e \times r_P||}{||e||} <= R \tag{3}$$

2. Closest point Q on line to P is:

$$r_Q = r_P + \frac{e \times (m + e \times r_P)}{||e||^2} \tag{4}$$

3. The barycentric coordinates of $Q(w_A w_B)$ such that r_Q = $w_A r_A$ + $w_B r_B$ are:

$$w_A = \frac{||r_Q \times r_B||}{||m||} \tag{5}$$

$$w_B = \frac{||r_Q \times r_A||}{||m||} \tag{6}$$

4. Check that point Q lies between A and B by making sure the barycentric coordinates are between 0 and 1:

inside = $(w_A >= 0)$ and $(w_A >= 1)$ and $(w_B >= 0)$ and $(w_B >= 1)$

After defining which feature points are included by our cylinder and which radius is to be used, we applied RANSAC again. This cylindrical approach is depicted in Figure 10.

Virtual NCAP-relevant objects such as traffic signs, pedestrians, and road signs are created using the Blender software. Figure 11 demonstrates a created cyclist for our approach. The use of Blender software enables us to design both static and dynamic objects with intricate details. The virtual objects can be placed on the plane as shown in Figure 12, with no restrictions on their placement or orientation. After the virtual objects have been placed, the next step is to render them into the camera stream.

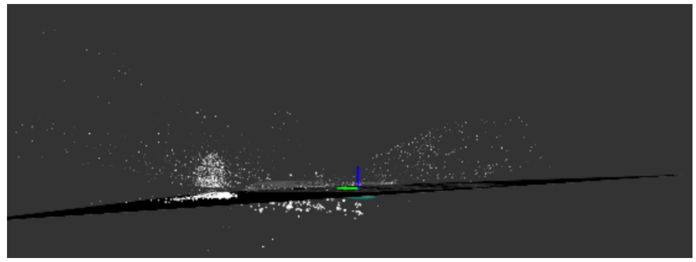


Fig. 9. Incorrectly placed inclined plane using the random sample consensus algorithm.

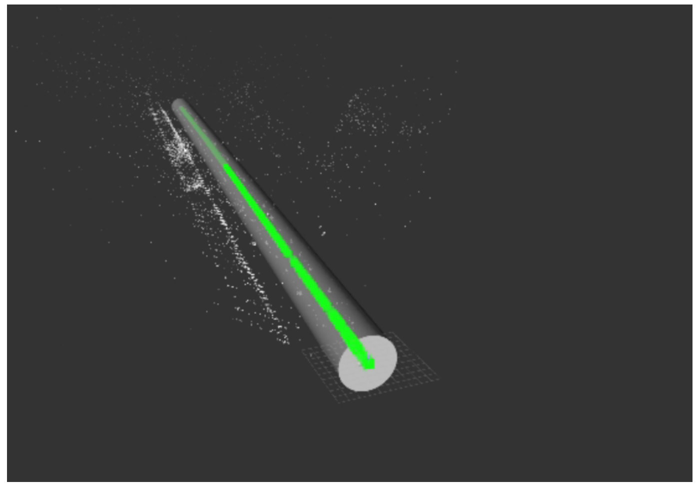


Fig. 10. Approach with a cylinder to select the relevant feature points to place a correct plane on the street.

E. Augmentation-Process

Open Graphics Library (OpenGL) is utilized to integrate virtual objects from Blender into a scene. This tool enables the rendering process. As depicted in Figure 13, the virtual cyclist has been added to the scene without material properties such as color or shadow. The grid displayed on the floor demonstrates that the cyclist is positioned on a recognized plane. The final step in our AR-ADAS pipeline involves

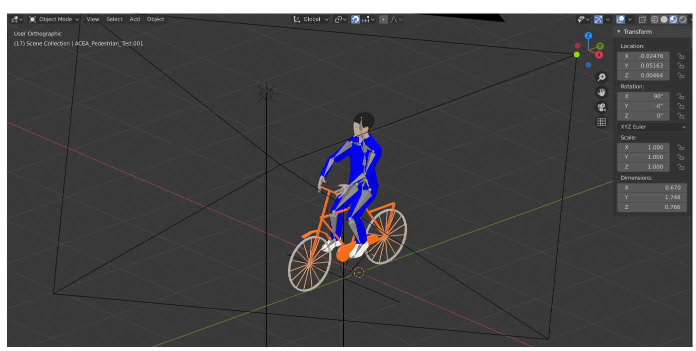


Fig. 11. NCAP-testobject cyclist using software *Blender*.

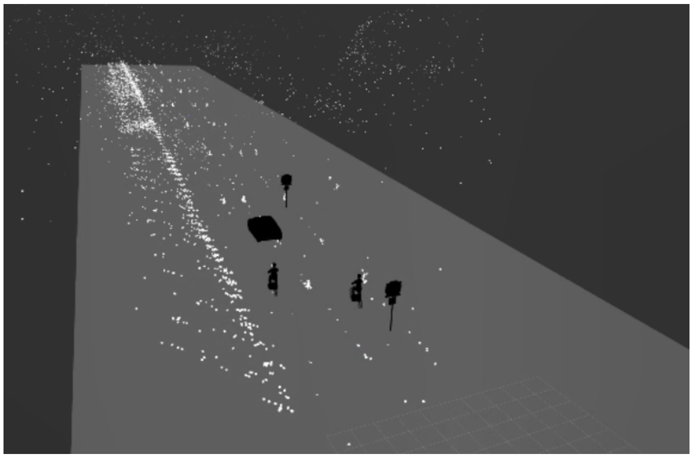


Fig. 12. Object placement on the surface in the Point Cloud.

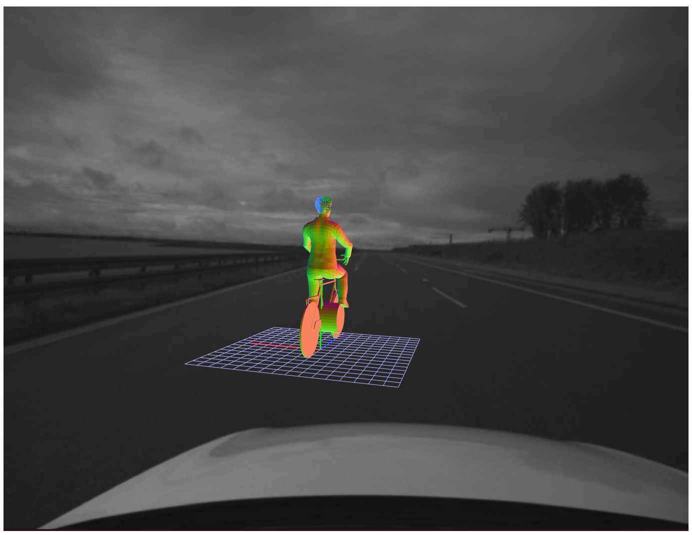


Fig. 13. Augmented cyclist in NCAP-scene.

driving the physical vehicle through the test scene once more. This time, ORB-SLAM is not utilized in mapping-mode, but rather in localization mode, which results in a faster vehicle speed due to reduced computational requirements. The real camera images are now augmented with virtual objects.

F. Advantages of our Approach

Our approach aims to merge the benefits of testing in simulation, such as reproducibility, flexibility, and efficiency, with the complexities of the real-world vehicle and environmental conditions. This approach seeks to bridge the gap between testing methods and enable the testing of more intricate scenarios, thus enhancing road safety.

Industry-based ViL-approaches demonstrate the need for such a method. Unlike conventional ViL-methods, which simulate a virtual environment to stimulate vehicle sensors while maintaining the vehicle's realistic dynamics, our approach represents a step towards more realistic vehicle testing. By utilizing AR, our approach allows us to use the real environment with all its characteristics, eliminating any simplifications made in simulation. Figure 14 contrasts existing ViL-methods, represented at the top, with our approach, depicted at the bottom.

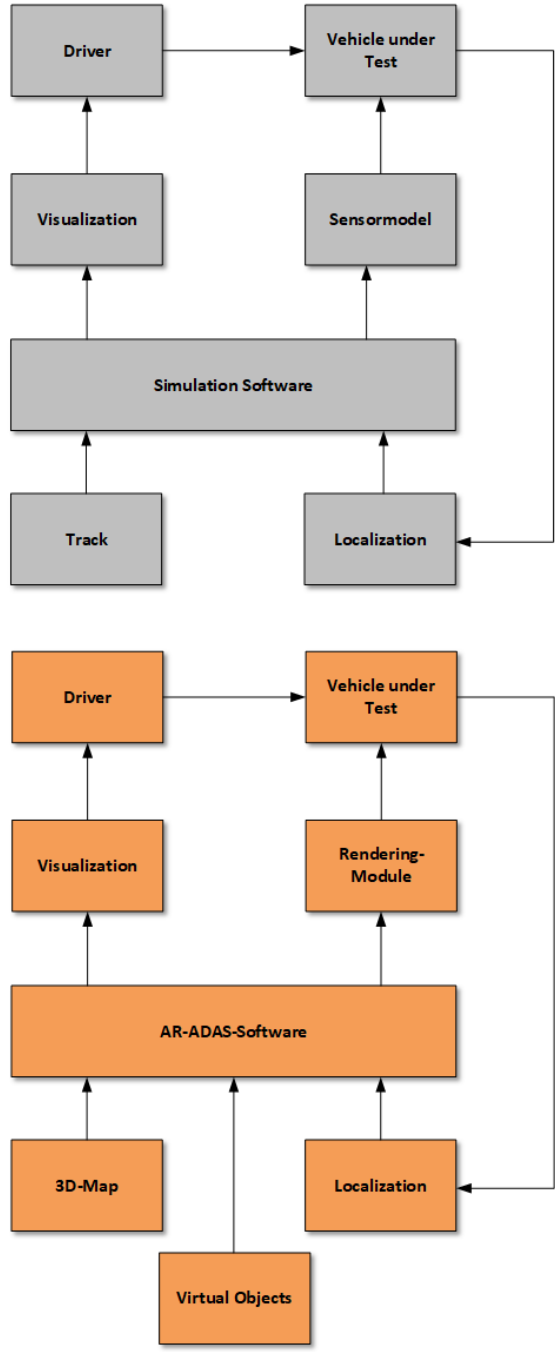


Fig. 14. Existing ViL-approach on top [22] and our approach of using augmented reality in advanced driver assistance systems on the bottom.

The overlay of lanes allows for independent testing of a lane departure warning system, regardless of the testing ground. Scenarios such as the presence of temporary lane markings or missing sections can be tested in the same area. Variations in lane width or international differences in lane markings can be represented. The camera image can also be augmented with superimposed vehicles ahead to test congestion assistance systems. That eliminates the need for second vehicles and drivers in the initial testing phase, reducing costs and increasing safety for test engineers. Test cases with traffic signs, pedestrians, and cyclists can be situational and quickly added, and a combination of different test scenarios is also possible. The limitless variety of test scenarios allows for a significant increase in the depth of testing during the early stages of development, leading to improved testing quality and overall system. With the increasing number of driver assistance systems and the move towards autonomous driving, the application area of the software can be expanded as needed.

G. Challenges for our Approach

Further advancements and optimizations regarding accuracy, robustness, and runtime can be seen in developments based on the ORB-SLAM approach, such as ORB2-SLAM [20] and ORB3-SLAM [16]. While ORB-SLAM demonstrates impressive performance in well-structured sequences, it can encounter errors in poorly structured sequences, such as those found in Euro NCAP test scenarios, or when feature points temporarily disappear due to factors like motion blur [29]. Along with accuracy, the runtime of the algorithm is also a crucial factor. Today, camera systems operate at a frame rate of 30 to 60 frames per second (fps), and the maximum overall runtime for processing a single frame can be found in Table II.

TABLE II
SEVERAL FRAMERATES AND THE ACCORDING MAXIMUM RUNTIME

Framerate	Maximum Runtime
10 fps	$\frac{1}{10}\,s = 0.1000\,s$
30 fps	$\frac{1}{30}\,s = 0.0333\,s$
40 fps	$\frac{1}{40}\,s = 0.0250\,s$
45 fps	$\frac{1}{45}\,s = 0.0222\,s$
50 fps	$\frac{1}{50}\,s = 0.0200\,s$
60 fps	$\frac{1}{60}\,s = 0.0167\,s$

For a successful evaluation of ADAS-test scenarios, the AR system must be able to orient itself in the environment very accurately [20]. One cause is the missing feedback about the impact intensity of test dummies when crashing them. For this reason, it is necessary to know the exact position of the car on the test track to calculate the intensity of the impact based on the braking distance. When using Euro NCAP test scenarios, velocities up to

$$130\,\frac{km}{h} \,\widehat{=}\, 36.111\,\frac{m}{s} \tag{7}$$

are tested [26]. The AR algorithm must have a faster runtime compared to the speed of the camera system. The distance d the vehicle covers within a frame at any given velocity and framerate can be calculated by:

$$d = \frac{v_{Vehicle}\,[\frac{m}{s}]}{Framerate\,[\frac{frames}{s}]} \tag{8}$$

At a speed of 130 $\frac{km}{h}$ and a camera framerate of 30 fps, the vehicle travels

$$d = \frac{36.111\,[\frac{m}{s}]}{30\,[\frac{frames}{s}]} = 1.204\,\frac{m}{frame}. \tag{9}$$

Accordingly, for a framerate of 60 fps at the same speed, a distance of

$$d = \frac{36.111\,[\frac{m}{s}]}{60\,[\frac{frames}{s}]} = 0.602\,\frac{m}{frame} \tag{10}$$

is covered. A deceleration of one frame means a deviation of the test results of 0.602 to 1.204 metres. Based on the high speed of the car and the camera, and the high need for precision in object placement, it is clear that the requirements for this application of AR are far more strict than for the usual application for human users. Another task we have to deal with is the low texture of the environment and the uniformity that comes with it, so that *Image Matching* can occur. The use of multiple objects such as pylons to enrich the texture of the environment or approaches to *Optical Flow* can support this.

H. Visualization for the Testdriver

The main task of the Human-Machine-Interface (HMI) is to make the AR perceptible to the test driver in real time that the ADAS functions of the test vehicle, as well as the human interaction, can be evaluated. The acceptance of the HMI as an interface for the experience plays an important role. This depends for the most part on the quality of the display, interaction, and haptics [30]. For our approach, the selection of a suitable HMI concept focuses on visualization and interaction. To display AR visibly, the use of a suitable HMI or a corresponding display is necessary. Possible screen approaches are classified into feature classes based on their properties. Displays that use a medium-direct view through to the real environment in 3D belong to the class of See-Through (ST)-displays. Monitor-Based (MB)-displays only allow an indirect view of the real environment. Live or stored videos (2D) are used for this technology. Indirect displays (3D objects: video ST), which visualize AR in 3D using video, also belong to the group of ST displays. The three-dimensional concept is crucial here. The processing of the 2D-camera data of the real environment used, through 3D-scene modeling, makes it possible in the first place to integrate the virtual objects in the correct perspective (2D). Video-based ST displays (video ST) are used if the recording and playback of this same AR on an indirect display take place almost simultaneously. Optical ST-displays (3D-Objects:optical ST) are used when the reproduction of the virtual objects in combination with the direct view of the real environment is correctly integrated. The visualization of AR according to Azuma limits the AR-capable displays to those that can display virtual three-dimensional objects correctly oriented in perspective [7]. For the identification of suitable HMI approaches for testing camera-based ADAS, only these ST displays fulfill the necessary criteria. Figure 15 shows a summary of the different categories for visualization.

HMI approaches in which stationary displays are mechanically fixed to the vehicle for the duration of the test belong to the Head-Up-Display (HUD) group. Head-Up-Displays used in automotive vehicles to show the driver the actual speed or using the display of a smartphone or tablet belong to this category. Those in which the display is attached to the head like when using Virtual Reality (VR)-glasses or AR glasses belong to the Head-Mounted-Display (HMD) group [31]. In both HMD and HUD, HMI approaches of optical and video-based ST displays are identified. In the further progress of the approach to use AR as a visualization for the driver, different evaluations must be carried out [1].

I. Further Thoughts about Using Augmented Reality Simulation for Testing Advanced Driver Assistance Systems in Future Automotive Vehicles

In the first step, our approach will be transferred to camera-based sensors. As already highlighted in the previous chapters, only a few ADAS functions, such as traffic sign recognition or LDW, only access the camera. To evaluate further tests and achieve the equal behavior of the ADAS-ECU (cf. Figure 4) in reality as in using AR, the integration of further sensor technology such as radar is needed. It should also be mentioned that Euro NCAP test scenarios according to the current state only take place under ideal conditions (sun position noon - no or only a few shadows and reflections, no other road users, no rain, etc.) [26]. Using our approach is intended to further increase the complexity and realism of Euro NCAP test scenarios. Another aspect is the visualization of the Augmentation for the driver. Here, one considerable aspect is the acceptance of the user by AR. Further investigations into a visualization for the user are being pursued as part of this project.

VII. Conclusion

In this paper, we have proposed an approach using AR in automotive vehicles. The use of AR in ADAS is intended to combine the advantages of simulative test procedures, such as reproducibility and cost savings, with the advantages of test procedures in reality (complexity of the entire vehicle and the environment). We modeled the problem of creating an urban environment to use AR for testing in high-speed ADAS. Our approach is based on combining vSLAM-algorithms with Artificial Intelligence (AI) to use Object Detection. That should help generate a better overall performance concerning computing speed and accuracy. Creating a virtual three-dimensional environment with a superior understanding of the individual objects should, in a further step, make it possible to augment other sensors such as the car's radar and LiDAR with objects in addition to the camera data. That should once again increase the overall performance of the entire system. In addition to providing a link between virtual and real test procedures, this approach intends to increase the complexity of potential test procedures, accelerate the development speed of ADAS functions, and improve safety for future mobility solutions.

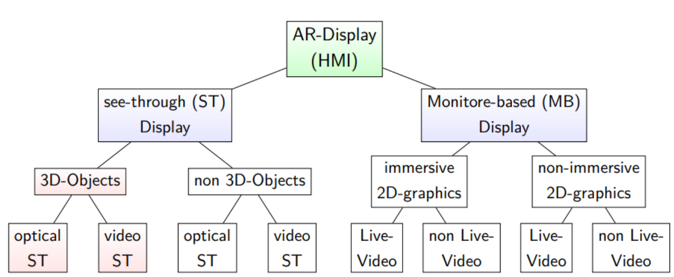


Fig. 15. Augmented reality display classes inspired by Milgram [7].

References

[1] M. Weber, T. Weiß, F. Gechter, R. Kriesten, "Augmented Reality Simulation for Testing Advanced Driver Assistance Systems in Future Automotive Vehicles," in: SIMUL 2022, The Fourteenth International Conference on Advances in System Simulation, Lisbon, Portugal, October 2022.

[2] K. Bengler et al., "Three decades of driver assistance systems: Review and future perspectives," in: IEEE Intelligent Transportation Systems Magazine vol. 6, no.4, pp. 6–22, Winter 2014.

[3] F. Schuldt, F. Saust, B. Lichte, M. Maurer and S. Scholz, "Efficient systematic test generation for driver assistance systems in virtual environments - Effiziente systematische Testgenerierung für Fahrerassistenzsysteme in virtuellen Umgebungen," 2013, [Online], Available from: https://publikationsserver.tu-braunschweig.de/servlets/MCRFileNodeServlet/dbbs-derivate-00031187/AAET-Schuldt-Saust-Lichte-Maurer-Scholz.pdf, accessed 2023.02.09.

[4] B.-J. Kim and S.-B. Lee, "A study on the evaluation method of autonomous emergency vehicle braking for pedestrians test using monocular cameras," Applied Sciences 10, no. 13: 4683, July 2020, doi: 10.3390/sapp10134683.

[5] C. Miquet et al., "New test method for reproducible real-time tests of ADAS ECUs: "Vehicle-in-the-loop" connects real-world vehicles with the virtual world," in: 5th International Munich Chassis Symposium 2014, pp. 575-589, July 2014.

[6] J.E. Stellet et al., "Testing of Advanced Driver Assistance towards automated driving: A survey and taxonomy on existing approaches and open questions," in: 2015 IEEE 18th International Conference on Intelligent Transportation Systems, pp. 1455–1462, September 2015.

[7] R.T. Azuma, "A survey of augmented reality," in: Teleoperators and Virtual Environments, pp. 355–385, August 1997.

[8] Pokémon GO, "Developer Niantic is working on a game for tourists - Pokémon GO: Entwickler Niantic arbeitet an einem Spiel für Touristen," [Online], available from: https://mein-mmo.de/pokemon-go-entwickler-app-touristen/, accessed 2023.02.09.

[9] A. State, G. Hirota, D. Chen, W. Garrett and M. Livingston, "Superior augmented reality registration by integrating landmark tracking and magnetic tracking," in: SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, pp. 429-438, August 1996.

[10] R. Chatila and J-P. Laumond, "Position referencing and consistent world modeling for mobile robots," in: IEEE International Conference on Proceedings Robotics and Automation, pp. 138-145, 1985.

[11] J. Engel, J. Sturm and D. Cremers, "Camera-based navigation of a low-cost quadrocopter," in: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2815-2821, 2012.

[12] T. Taketomi, H. Uchiyama and S. Ikeda, "Visual SLAM algorithms: a survey from 2010 to 2016," in: IPSJ Transactions on Computer Vision and Applications, doi: 10.1186/s41074-017-0027-2, 2017.

[13] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision," in: Cambridge University Press, second edition, ISBN: 0521540518, 2004.

[14] R. Mur Artal, "Real-Time Accurate Visual SLAM with Place Recognition," Ph.D.-thesis, Universidad de Zaragoza, Prensas de la Universidad, Zaragoza, Spain, 2017.
[15] P. Furgale, J. Rehder and R. Siegwart, " Unified temporal and spatial calibration for multi-sensor systems," in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1280–1286, Tokyo, Japan, 2013.
[16] C. Campos et al., "ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM," in: IEEE Transactions on Robotics, pp. 1-17, 2021.
[17] R. A. Newcombe, S. J. Lovegrove and A. J. Davison, "DTAM: Dense tracking and mapping in real-time," in: IEEE International Conference on Computer Vision (ICCV), pp. 2320–2327, Barcelona, Spain, 2011.
[18] J. Engel, T. Schoeps and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," in: European Conference on Computer Vision (ECCV), pages 834–849. Zurich, Switzerland, 2014.
[19] J. Engel, V. Koltun and D. Cremers, "Direct sparse odometry," in: arXiv:1607.02565, July 2016.
[20] R. Mur-Artal, J. Montiel and J. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," in: IEEE Transactions on Robotics, pp. 1147-1163, 2015.
[21] M. Nagai, "Research into ADAS with autonomous driving intelligence for future innovation," in: 5th International Munich Chassis Symposium 2014, pp. 779–793, January 2014.
[22] H. Winner, S. Hakuli, F. Lotz and C. Singer, "Manual driver assistance systems - basics, components and systems for active safety and comfort - Handbuch Fahrerassistenzsysteme - Grundlagen, Komponenten und Systeme fuer Aktive Sicherheit und Komfort," in: Springer Vieweg, Wiesbaden, March 2015, [Online], Available from: https://link.springer.com/content/pdf/10.1007/978-3-658-05734-3.pdf, accessed 2023.02.09.
[23] M. Darms, "A basic system architecture for sensor data fusion of environmental sensors for driver assistance systems - Eine basis-systemarchitektur zur Sensordatenfusion von Umfeldsensoren fuer Fahrerassistenzsysteme," Ph.D.-thesis, Technische Universität Darmstadt, 2007, [Online], available from: https://tuprints.ulb.tu-darmstadt.de/914/ accessed 2023.02.09.
[24] P. Seiniger and A. Weitzel, "Testing procedures for consumer protection and legislation - Testverfahren fuer Verbraucherschutz und Gesetzgebung," in: Manual driver assistance systems - Basics, components and systems for active safety and comfort - Handbuch Fahrerassistenzsysteme - Grundlagen, Komponenten und Systeme fuer Aktive Sicherheit und Komfort, pp. 167–182, Springer Vieweg, Wiesbaden, March 2015. [Online]. Available from: https://link.springer.com/content/pdf/10.1007/978-3-658-05734-3.pdf, accessed 2023.02.09.
[25] Euro NCAP, "AEB Pedestrian," [Online], available from: https://www.euroncap.com/en/vehicle-safety/the-ratings-explained/vulnerable-road-user-vru-protection/aeb-pedestrian/, accessed 2023.02.09.
[26] R. Fredriksson, M.G. Lenné, S. van Montfort and C. Grover, "European NCAP program developments to address driver distraction, drowsiness and sudden sickness," November 2021, [Online], available from: https://www.frontiersin.org/articles/10.3389/fnrgo.2021.786674/full, accessed 2023.02.09.
[27] M. Sieber et al., "Validation of driving behavior in the vehicle in the loop: Steering responses in critical situations," in: 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), 2013.
[28] C. Kamann and C.Rother, "Benchmarking the Robustness of Semantic Segmentation Models," in: IEEE Conference on Computer Vision and Pattern Recognition CVPR, 2020.
[29] Yu et al., "DS-SLAM: A Semantic Visual SLAM towards Dynamic Environments," in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018), 2018.
[30] J. Brade and A. Koegel, "Presence in virtual reality - Key to acceptance and transferability?!," in: 5. Fachkonferenz zu VR/AR-Technologien in Anwendung und Forschung, VAR² 2019, pp. 59-71, December 2019.
[31] R. Doerner, "Fundamentals and methods of virtual and augmented reality - Grundlagen und Methoden der Virtuellen und Augmentierten Realitaet," in: Virtual and Augmented Reality (VR/AR), Springer Viweg, pp. 1-143, 2019.

Modelling Player Combat Behaviour for Dynamic Difficulty Scaling and Combat Perception Analysis in First Person Shooter Games

Paul Williamson
School of Computing and Mathematics
University of South Wales
CF37 1DL, Pontypridd
e-mail: paul.williamson@southwales.ac.uk

Christopher Tubb
School of Computing and Mathematics
University of South Wales
CF37 1DL, Pontypridd
e-mail: christopher.tubb@southwales.ac.uk

***Abstract*—Non-Player Characters are a core aspect of a modern video game. They fulfil a wide range of roles and features in most genres. How they are perceived can have an impact on the overall enjoyment of the game, and when the NPCs are poorly modelled, the experience can be negative. This paper explores modelling human player behaviour to develop a combat model for NPCs, which will provide a dynamic solution for skill scaling. Game difficulty is a subjective notion and when games have a predefined classification, being able to satisfy all players is not realistic. This paper investigates if the combat model can therefore be used to scale the difficulty of the NPC in real-time, by dynamically adjusting a skill attribute, which is used by several key combat behaviours. As combat perception is critical to maintaining an immersive and entertaining experience, this paper also explores the degree of combat awareness exhibited by players.**

Keywords—NPC; Player Modelling; Difficulty; Combat Behaviour; Gameplay; FPS; Perception; Combat Awareness.

I. Introduction

This paper explores a player driven approach for dynamically adjusting the difficulty of Non-Player Characters (NPCs) in real-time [1]. This is achieved by modelling player combat gameplay. The model is then applied directly to NPCs. The combat efficiency is controlled by dynamically adjusting the variable parameters in real-time for the purpose of difficulty scaling. The reasoning behind modelling human-player gameplay is to help define the generalised upper and lower skill level of average players. These bounds are then, in determining how skilfully an NPC should behave, based on data rather than developer interpretation.

This paper conducts three experiments. The first experiment uses predetermined combat scenarios that records the data of human subjects, which is used to model generalised combat behaviours. The second experiment applies the combat model to NPCs, then evaluates the performance of the model when human subjects conduct a deathmatch scenario with the skill of the NPC increasing with each round. While the third experiment is also a deathmatch scenario, however, some behaviours have been altered to determine if subjects notice this irregularity.

The results show that generalising human combat data provides a suitable base for controlling the combat effectiveness of NPCs. It demonstrates that by using a variable to denote the skill level of the NPC, the difficulty can be dynamically adjusted to reflect a desired outcome.

The perception test had mixed results, with subjects generally displaying good combat awareness but lacking awareness of their surroundings. It also raised the question about the authenticity of subject opinions, because in several cases the subject made statements about the NPC that were not factual.

In Section II, this paper provides a background for NPCs in first-person shooter (FPS) games, the role of difficulty, and perception. Section III discusses the motivation behind this paper and what impact this research could have on the gaming industry. In Section IV, a literature review is undertaken to discuss relative research, which is consistent with the NPCs in the FPS genre. Section V details the three experiments, the first is used to capture and model human subject combat behaviour, while the second experiment evaluates the model by having human subjects compete against NPCs with the combat model in a deathmatch scenario. The third experiment assesses the combat awareness of subjects by implementing common flaws in the combat model and then having the subject compete in a deathmatch. In Section VI, the results of the experiments are analysed and displayed. Firstly, a numerical comparison between the human subject data and NPCs is discussed. An evaluation of the feedback from human subjects, regarding the perception of the combat model. An assessment of the model when determining if the combat model provided a suitable solution for dynamic skill scaling. Lastly, the data from the combat awareness experiment is analysed to determine the impact of the flaws had on the combat experience. Section VII concludes the paper and discusses the impact of the combat model. Finally, Section VIII discusses the potential future work that can be undertaken to build upon the work presented in this paper.

II. Background

A. Non-Player Character

A non-player character (NPC) is a character in a video game that is not controlled by a player. In a first-person shooter (FPS) game, NPCs are often used to add depth to the game's story and provide opponents for the player to fight against. There are several types of NPCs that can be found in FPS games. Some NPCs are simply background characters that exist to add atmosphere to the game world. These NPCs may not interact with the player in any meaningful way and are simply there to create a sense of realism. Other NPCs are more interactive and may offer the player quests or other tasks to complete [2].

Some NPCs are friendly to the player and will offer assistance or information, while others are hostile and will attack the player on sight. There are also NPCs that are neutral and will only attack the player if provoked. The behaviour of NPCs is often determined by the player's actions and choices within the game.

NPCs can also play a role in the game's story by advancing the plot or providing important information to the player. In many FPS games, the player will encounter NPCs who are key characters in the story and will have to interact with them in order to progress through the game.

The appearance and behaviour of NPCs in FPS games can vary widely. Some NPCs may be highly detailed and realistic, while others may be more stylized or cartoonish. The AI (artificial intelligence) that controls NPCs can also vary in complexity, with some NPCs behaving in a very lifelike manner while others are more robotic in their actions.

Overall, NPCs play a vital role in FPS games by providing opponents for the player to fight against, advancing the game's story, and adding depth and realism to the game world.

Research has shown that poorly designed and/or developed NPCs can be detrimental to the overall enjoyment of a game, and that thorough modelling should be considered when created NPCs [3][4].

B. *Difficulty in Games*

Difficulty is a key factor that can greatly impact a player's enjoyment of a video game. For some players, a high level of challenge is an essential part of the enjoyment of a game, as it allows them to feel a sense of accomplishment upon completing a difficult task. These players may seek out games that are known for their high level of difficulty and enjoy the sense of satisfaction that comes from overcoming that challenge.

On the other hand, some players may find a high level of difficulty to be frustrating and may not enjoy a game that is too difficult. These players may prefer games that offer a more casual or easy-going experience, allowing them to relax and enjoy the game without feeling overwhelmed or frustrated.

The ideal level of difficulty for a game can vary widely from player to player, and many games offer options for adjusting the difficulty level to better suit the player's preferences. Some games may allow the player to choose from several different difficulty levels at the start of the game, while others may have the difficulty based on the type of experience the player is interested in playing. For instance, the difficulty could be 'Story' or 'Challenge'.

In general, a well-designed game will offer a range of difficulty levels that allow players to choose the level of challenge that best fits their preferences. This can help to ensure that players of all skill levels can enjoy the game and find it to be a rewarding experience.

Ultimately, the enjoyment of a game is subjective and will depend on the individual player's preferences and skill level. Some players may find a high level of difficulty to be a key part of the enjoyment of a game, while others may prefer a more relaxed and easy-going experience. By offering a range of difficulty levels, developers can help to ensure that a wider range of players can enjoy their games.

In FPS games, difficulty is often represented in how adept NPCs are during combat encounters with the player. This presents a problem when having static difficulties because some players will have a skill level that is situated between two of the difficulty categories. Furthermore, when NPCs are artificially boosted, whether by increasing weapon damage or unnatural awareness, the player could feel a sense of unfairness.

C. *Definition of Combat Perception*

In an FPS game, combat perception refers to a player's ability to effectively and efficiently gather and process information about their surroundings, enemies, and allies during combat situations. This includes things like identifying enemy positions, anticipating enemy movements, spotting opportunities for cover and flanking, and using terrain and other environmental features to their advantage.

Good combat perception is crucial, as it allows players to make quick and accurate decisions in the heat of battle, which can be the difference between success and failure. There are many factors that can influence a player's combat perception, including their skill level, game experience, and the design of the game itself. Some players may have naturally strong combat perception, while others may need to practice and improve their skills over time.

As combat perception refers to the player's ability to gather and interpret information about the game environment, opponents, and their own character to make strategic decisions and take actions in real-time. Good combat perception involves being aware of one's surroundings, anticipating enemy movements, and reacting quickly and effectively to changing circumstances.

- Situational awareness: This refers to the player's ability to understand what is happening around them in the game world. This includes things like knowing where enemies are located, what weapons they are using, and what their current objectives are. It also involves being aware of the layout of the environment, including cover, choke points, and potential flanking routes. Good situational awareness allows the player to make informed decisions about how to approach combat situations.
- Target acquisition: This refers to the player's ability to identify and engage enemy targets effectively. This involves aiming and shooting accurately, as well as choosing the appropriate weapons and tactics for different ranges and situations. Players with good target acquisition can quickly and effectively eliminate enemy threats.
- Movement: In many FPS games, the player's character can move around the game world in a variety of ways, including walking, running, crouching, and jumping. Good combat perception involves being able to use movement effectively to avoid enemy fire, take cover, or position oneself for an attack. Players may also need to consider factors such as the speed and agility of their character, the layout of the environment, and the positioning of enemies when deciding how to move.
- Decision-making: Good combat perception also involves being able to make quick and effective decisions in the heat of battle. This may involve choosing the right

weapons and tactics for a given situation, deciding whether to attack or retreat, or choosing which objectives to prioritize. Players with good combat perception can adapt to changing circumstances and make informed, strategic decisions.

Therefore, when modelling NPCs from key human-based combat behaviours, it is pivotal to identify and model generalised combat patterns. This should help control the degree of skill of NPCs and help maintain a realistic combat experience where NPCs operate in a similar manner as human players.

III. MOTIVATION

A goal of this paper is to explore the purpose of different difficulty levels in video games and to consider how these levels might be classified. The difficulty of a game can affect how it is perceived by players, and it is important that the chosen difficulty level is perceived appropriately in order to avoid negative experiences. This is a subjective topic, as different players may have different preferences when it comes to the level of challenge they prefer in a game.

An FPS game was chosen as the subject of study because it is a well-established and popular genre in the gaming industry and can provide useful insights into how players perceive and respond to increasing difficulty. FPS games often rely on player skills such as quick reaction time, accuracy, and decision-making in combat situations, which makes them a suitable choice for evaluating how players perceive changes in difficulty. By focusing on a single genre, the scope of the research can be kept narrow and more specific conclusions can be drawn from the data. It is common practice in the video game industry to adjust certain gameplay elements such as reaction time, damage dealt, and precision when changing the difficulty level in FPS games.

The idea behind real-time skill scaling is that a single, fixed difficulty setting may not be suitable for all players. By adjusting NPC abilities based on the player's performance, the game can offer a more personalized and player-centric approach to difficulty. Maintaining immersion in combat is important because NPCs and human players are often perceived differently in FPS games, even when they are fulfilling similar roles. Real-time skill scaling could allow developers to balance uneven teams with NPCs without breaking immersion for the player, by making NPCs indistinguishable from human players. Combat perception is also a key characteristic of distinguishing which players are human or NPC controlled. The motivation around determining the generalised awareness of players, is to provide a better understanding which where a combat model for NPCs can be improved, with the focus of maintaining immersion.

A study undertaken by Williamson and Tubb [5] indicated that gameplay could be categorised into fundamental behaviours and patterns that could be modelled. They suggested that the effects of having poorly modelled and designed NPCs, such as high accuracy in combat, can result in a negative experience and that their opponent could be mistaken as cheating.

IV. RELATED RESEARCH

A. Non-Player Characters in Combat Roles

When discussing NPCs in a combat role, the focus will be primarily on the FPS genre. While NPCs have been present since the inception of the FPS genre, there has been some progress in the scope of their capabilities. Orkin [6] showed that NPCs could display seemingly complex characteristics, such as tactics, by using Goal Orientated Action Planner (GOAP), which enabled individual actions to be undertaken based on current circumstance. This technique sparked a host of research based around maximising the effectiveness of using action planners. Influenced by GOAP, Pezzato et al. [7] researched a technique that uses active interference and Behavioural Trees (BT) to develop decision-based plans, which are adaptable and created in real-time for robotics.

However, Agis et al. [8] suggest that because GOAP are individual by design, NPCs are not actively cooperating with allies and just giving the illusion of working as a team. They proposed using a new form of Event Driven Behavioural Tree (EDBT), where three extra nodes are added for communication purposes between same team NPCs. These nodes facilitate actions in a sub-tree, so the sender can request the receiver/s to perform certain actions.

Neufeld et al. [9] propose that behavioural trees are a good approach to NPC decision-making, when combined with a Hierarchical Task Network (HTN). The HTN was able to instruct multiple NPCs what to do but leave the low-level execution of tasks to the BTs. When comparing this hybrid approach against a pure HTN solution. The results showed the hybrid approach to be more flexible and fail less often.

B. Human Imitation

The idea to measure and quantify the skill of human players is not a new idea and in essence this research is akin to work undertaken to develop NPCs to appear more believable. While believability can have many interpretations depending on the context of the topic, this paper is focused on NPCs with a generalised skill level. Camilleri et al. [10] identify believability as:

> *Player believability is a highly subjective notion commonly viewed as the ability of a game playing character to convince observers that it is being controlled by a human player*

This suggests that for an NPC to be deemed as a believable character, it needs to exhibit human-like behaviours. So, observers are convinced the character is human controlled. In the context of the FPS genre, these behaviours will range from combat efficiency to pathfinding characteristics and decisions.

Polceanu et al. [11] presented two solutions for imitating human player behaviour. The first is a human behaviour mirror technique, while the second uses interactive genetic algorithms. This is supported by Mora et al. [12], they combined the techniques to create a hybridisation model. When experimented in a Turin Test scenario, the results showed that a high level of “humanness” was achieved.

The importance of imitating human players extends beyond difficulty scaling. Wehbe et al. [13] argue that the difference between playing with other human participants and

NPC is profound because it intersects the artificial world of the game with real life. They also imply that quasi-feudalistic tendencies of human players should be encouraged in NPCs where possible.

C. Difficulty Control in FPS Games

When discussing difficulty control, it is important to define the terminology in the context of a video game. Smeddinck et al. [14] state:

> *Game difficulty choices that are presented in menus with typical labels such as "easy, medium, hard" can be found even in very early and simple games. The "classic way to present difficulty choices" has arguably evolved largely as a matter of technical circumstance.*

This suggests that difficulty has been part of video games for a long time and has needed to evolve out of necessity.

Hendrix et al. [15] showed a positive result when developing a six-point system, which gathers data and applies an algorithm to determine the capabilities of the player through engagement.

Blom et al. [16] have taken a different approach to identifying the currently perceived difficulty experienced by the player. They developed a system that monitors facial expressions to determine the difficulty the subject is experiencing, achieving 72% accuracy. The model was able to adapt to the individual's performance level via their facial expression and establish which tasks or challenges presented the most difficulty.

The research shows that while difficulty may be an easily defined concept, problems occur when putting the theory into practice. For the remainder of this paper, difficulty will be defined as:

> *The degree of challenge presented to an individual when undertaking game related tasks or mechanics.*

These tasks can range from game objectives, such as solving puzzles or combat interactions that relies on eliminating opponents.

The idea of a personalised approach to NPC development is not a novel idea. Research by Bakkes [17] discusses the notion of using player models to generate an individualistic experience. Their research shows that a few viable techniques have been examined [18]-[20]. These techniques mainly use NPC training methods or manipulate external constraints, such as environmental, to achieve the desired outcome. However, the purpose of this paper is to use generalised player data to control the combat efficiency of NPCs, which can be adapted in real-time.

D. Impact of Perception

There is a growing body of research relating to combat perception and/or behaviours in FPS games. Moreira et al. [21] argue that there are negative experiences when playing against an opponent that is much weaker or stronger. For weaker players, they experience frustration because they are likely to lose, and the stronger players have a lower sense of challenge which can manifest into boredom. They proposed three systems to help bridge the gap between the varying skill levels:

- Better aim assist: They used a system in which implements a bullet magnetism effect, which is where bullets are slightly attracted to the target.
- Activation and duration: This system dynamically controls when aim assist should be active.
- Variable manipulation: This system varies some of the game variables, such as weapon damage.

While this approach yielded extremely positive results, and the overall enjoyment of their subjects increased, there is an issue of fairness which could manifest into something negative if a player thinks their opponent is be artificially boosted because it could appear as though their opponent is cheating. This perception is essential to this research because when a player 'feels' they are at a disadvantage through no fault of their own, it can affect the entertainment value [22]-[24].

Choi et al. [25] used a different approach in which they developed a cognitive architecture to control how NPCs act throughout an FPS game. They created a system called 'ICARUS' which exhibited reactive and goal orientated behaviours, that were designed to imitate human-like gameplay. This system is based on psychological theories of human intelligence, it differentiates between short-term and long-term memories, and stores information in symbolic list structures. It operates in cycles of recognition and action, and its learning process is incremental and integrated with performance. The ICARUS architecture records the actions and observations of the NPC during exploration and tries to explain how these actions contributed to achieving the goal. It does this by linking known skills and concepts, and by creating generic templates for basic action models to explain any unexpected changes. ICARUS converts these explanations into new skills that can be used to achieve similar goals in the future, including strategies, route knowledge, and ways to overcome obstacles.

J. Asensio et al. [26] conducted research on the idea of imitating human behaviour by implementing a two-tier system. The system used symbols to represent and understand the world and artificial neural networks (ANN) to replicate the human nervous system's ability to adapt and recognize patterns. When this system was tested using the Turin Test for Bots, the results indicated a positive trend towards the NPC's perceived "humanness". This research shows that players' perceptions of NPCs are important and experimenting via a Turin Test, it removes knowledge that the subject is playing against an NPC. Therefore, identifying the degree of generalized combat perception exhibited by players is crucial in determining how much imitation is necessary to pass as a human player.

V. Experiments

A set of experiments were done to study the combat effectiveness of human subjects, including a perception test to measure their level of combat awareness. First, a combat modelling experiment was required, where subjects undertake a series of combat scenarios and their data recorded. The data is used to develop a combat model and then the NPC with the combat model undertakes the same experiment. Next, subjects are tasked with competing against NPCs with the combat model, with increasing difficulty as the stages progress. Lastly, subjects fight against NPCs with the combat model in a deathmatch scenario, however, key gameplay behaviours have been altered which will help evaluate the awareness of the subjects.

The purpose of these experiments is to find patterns in the combat gameplay, which can be modelled and applied to NPCs. After modelling, the experiments evaluate how well the model works in a deathmatch scenario, which is a popular game mode in FPS games. The objective for the perception test is to see where the focus of the modelling should be, because if subjects notice that the model has unnatural elements to the combat, it will mean the attention to detail of the model will need to be very high.

A. *Experimantal Environment*

The experiments in this paper were created using Unity3D and were uploaded to an online server. Participants had to download and run the experiments on their own computers. The first experiment took about 15 minutes to finish, and data was saved and uploaded to an online database after each stage. The second and third experiment took 10 minutes each, and also had an accompanying questionnaire that participants completed through a web browser. Data from this experiment was also saved online after each stage.

In both experiments, the subjects were not identifiable by name and no personal information was collected. No specific skills or knowledge were required to participate. The data collected during the experiments did not contain any identifying information that could be linked to a specific individual. As the subjects were anonymous participants and it was conducted online, a certain level of control was reduced. To help protect and verify the integrity of the data, the actions of the subjects was also saved, this enabled the researchers to watch all of the entries and look for any irregularities. It should be noted that as a consequence of anonymity, it was possible for participants to run the experiments multiple of times, there was no data to suggest this happened. For the combat behaviour experiment, a total of 30 subjects fully completed the experiment, and for the dynamic difficulty experiment, a total of 21 participants.

B. *Combat Behaviour Experimental Protocol*

The experiment involved five scenarios in which the actions and decisions of the subjects were recorded as they completed each task. Each scenario had three levels of difficulty, with the targets appearing at various distances from the subjects. The subjects were not able to move, except to turn, and the targets randomly appeared within their field of view. These scenarios were chosen because they represent typical combat situations in first-person shooter games. It was determined that it would be best to only model the basic scenarios and evaluate the effectiveness of the skill scaling system rather than including more complex scenarios that may not be applicable in all modern games. The specific scenarios were derived by analysing how combat emerges and the various conditions, and through an examination of literature focused on combat [6][27][28]. The importance of these scenarios can be seen in research undertaken by Conroy [27], they devised a model where NPCs calculated the degree of threat based on their current situation and part of the threat analysis is whether to engage in combat based on state of the opposition. They showed that distance to opponents has an impact on combat behaviour and was reflected in the results. While Rayner [28] conducted research focused on cursor-based mechanics, which partly discusses the implications of modelling player behaviour when the crosshair is in the centre of the screen. They also suggest that skill is therefore based on how fast a player can control the targeting across a 2D plane, whilst also being accurate so the target remains in the centre of the screen.

Figure 1 shows an example of a target, all targets spawn facing the subject and are eliminated from a single shot.

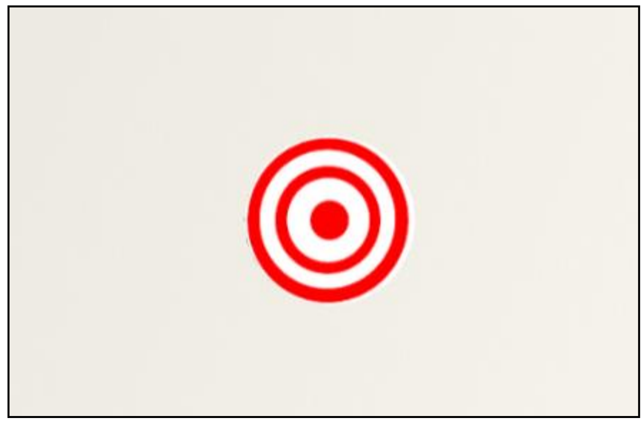


Figure 1. Example of Target

While it is understood that in most FPS games targets will take more damage if hit in a particular spot, however, this experiment does not employ such mechanics as the objective is to observe the combat behaviour as well as combat efficiency.

The experiment cycles through all weapons and stances until the subject has completed all scenarios using both weapons and both stances. The weapons and stances are as follows:

- Pistol: Off hip
- Pistol: Aiming down sights
- Assault Rifle: Off hip
- Assault Rifle: Aiming down sights

Figure 2 shows an example of both pistol and assault rifle from the perspective of shooting off the hip. In this stance each weapon has a crosshair in the middle of the screen and when shooting, the shots will be towards the centre of the crosshair. When the weapon is continuously shot, the size of the crosshair grows, making the weapon less accurate, until they stop firing and the crosshair recedes back to normal size. The purpose of this is see how accurate subjects are when tasked with this constraint, because it is possible to increase accuracy by controlling the about of burst firing.

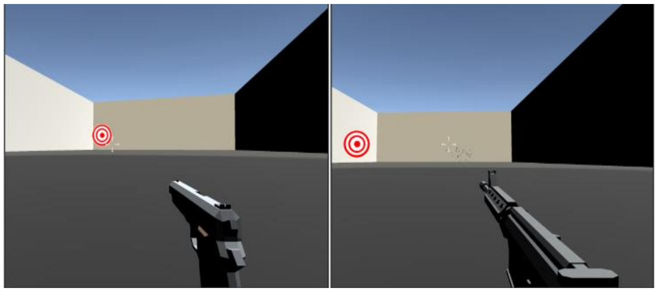


Figure 2. Off the Hip

Figure 3 shows an example of both assault rifle and pistol when aiming down the sights. When aiming down the sights there are no crosshairs, the subject will need to use pip on the sights. Both guns have kickback, with the pistol having a large kickback, and the assault rifle small. The reason this experiment is required is because it will compare how efficient subjects are when attempting the same problem with different means.

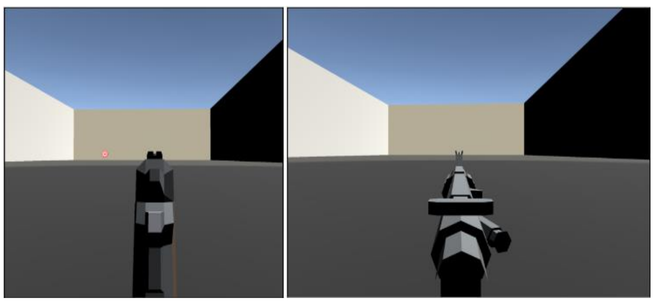


Figure 3. Aiming Down Sights

The purpose of the experiments is to understand how the use of different weapons and the way they are held can influence combat behaviour and effectiveness. This information will be used to create realistic models of weapon behaviour and to assess the impact of these behaviours on game difficulty. All targets in the experiments are identical and have a fixed amount of health that is depleted when they are hit. When the health of a target reaches zero, it is destroyed.

a) Normal Scenario

In this scenario, the objective is to eliminate the target before it self-destructs. Targets spawn one at a time and the subject has 2 seconds to eliminate before it disappears, a new target will not spawn until the current target is destroyed or self-destructs, there is also a small delay between target being destroyed and a new one spawning. The purpose of this scenario is to measure the player's accuracy and reaction time at different distances from the target (near, medium, far) to understand the average reaction time in a normal scenario. A total of five targets spawn at each distance, is within view of the subject's field of view, and are facing the subject. Figure 4 provides an example of a target spawning at a medium distance.

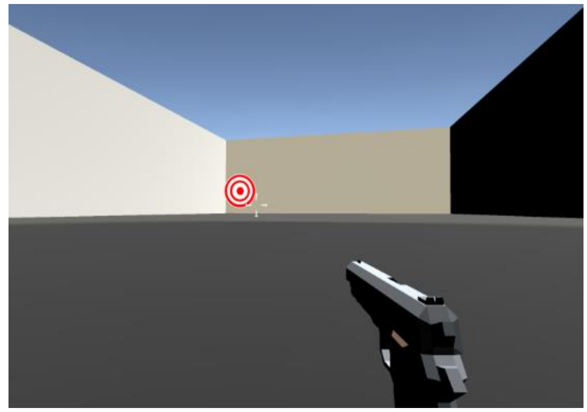


Figure 4. Normal Scenario Screenshot

This scenario represents a common encounter in an FPS game because it highlights how subjects react when they first observe their target.

b) Increased Spawn Rate

This scenario was conducted to see if the presence of additional targets while a subject is already in the process of eliminating another target, affects the subject's reaction time and combat efficiency. Specifically, the experiment aimed to investigate whether external influences can negatively impact combat efficiency and if so, to what extent. There are three stages which correspond to near, medium, and far distance, each stage will have a total of five targets. Similar to the single target scenario, the target self-destruct after 2 seconds, however, the spawn rate of targets is increased as the stages advance. The first stage (near) has a spawner time of 1.8 seconds, second stage (medium) 1.6 seconds, and last stage (far) 1.4 second, this provides a situation where a subject is likely to have multiple targets on the screen at a given time. While more aggressive spawn times could have been used, it was decided that the scenario was to observe how subjects react when a new target appears and if it affects their combat performance and decision-making, rather than overwhelming subjects.

Figure 5 shows an instance of the scenario when the targets are at a far distance.

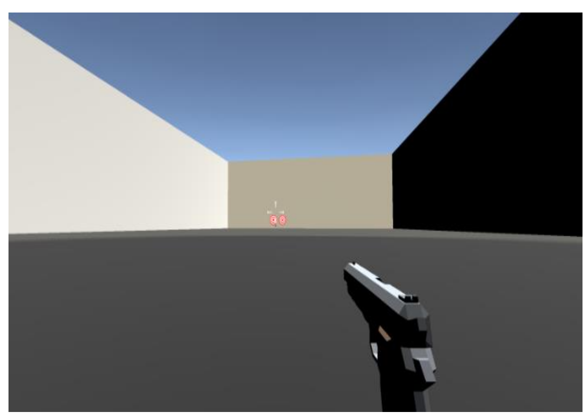


Figure 5. Increased Spawn Rate Screenshot

The importance of this scenario is to observe the combat behaviour when a new target appears, because in most FPS there will be times when combat is underway, and a new threat appears. It is important to determine if subjects' efficiency decreases because their focus is momentarily altered.

c) Grouped Targets

In this scenario, multiple targets may appear at the same time, as is common in modern games. The scenario also involves a three-stage approach, with an increasing number of targets spawning simultaneously as the stages progress. In the near stage, there are three targets, in the medium stage there are five targets, and in the far stage there are seven targets. Each target remains visible for five seconds before disappearing. This scenario is designed to study the pattern of behaviour exhibited when multiple opponents suddenly appear, rather than appearing after the combat is already underway.

This scenario differs from the increased spawn rate because the targets are all present and in view from the start of a round. The increase spawn rate scenario analyses how subjects adapt to a new target appearing while they are currently engaging a different target, whereas the grouped targets scenario addresses how subjects formulate attack strategies on multiple targets and the impact it has on combat efficiency.

Figure 6 displays a screenshot of the grouped targets at a medium distance.

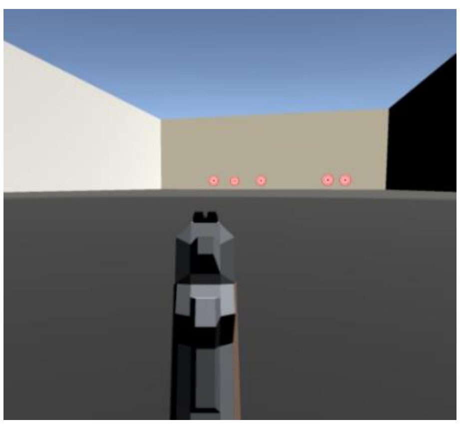


Figure 6. Grouped Targets Screenshot

This scenario is the often experienced when a player enters a new and/or open area, multiple enemies can be in view and then player needs to quickly form a targeting strategy.

d) Varying Size

In this scenario, the normal scenario stage is repeated but with target size decreasing as the stages progress. The purpose of this is to determine if precision has a significant impact on combat efficiency. The near stage will have a normal target size, medium stage the size will be 0.9x and the far stage a size of 0.8x. This scenario will help factor precision into the modelling, if the results suggest that there is a combat performance drop due to the target size. Figure 7 shows a screenshot of a target during the last stage when the targets spawn far away from the subject.

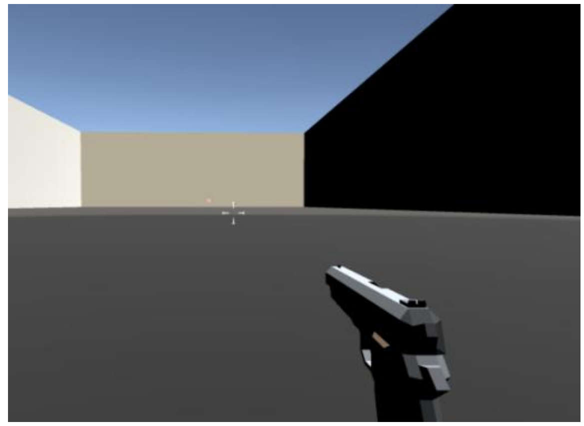


Figure 7. Smaller Target Screenshot

This scenario represents conditions when a player spots a target that is far away, or when smaller and specific parts of the target provide bonus damage when attacked. For instance, it is common in FPS games when attacking a humanoid enemy in the head, often referred as a 'headshot', it yields a lot higher damage than attacking not critical parts such as arms or legs.

e) Moving Target

In this scenario, targets appear every 2.5 seconds and disappear after 5 seconds. The targets move from left to right at a constant speed, reversing direction when they reach the edge of a predefined boundary. It follows the same set up as the normal target scenario, regarding near, medium, and far stages. A total of five targets spawn per stage, and when moving targets rotate to face the subject, this was to keep the targets area size the same regardless of the offset angle from the location of the subject. A new target spawns every 2 seconds, therefore, if the targets are not eliminated quickly enough, they may overlap on the screen. The purpose of this scenario is to see if combat efficiency is affected by the presence of moving targets.

C. Dynamic Difficulty Experimental Protocol

This experiment involves a deathmatch with the goal of achieving a set number of eliminations before the opposition does. In this case, the requirement is set at 5 eliminations. The experiment has three stages, and as the stages progress, the base skill level of the non-player characters (NPCs) is increased. The stages and their corresponding base skill levels are as follows:

- Stage One: NPC Base Skill is 3
- Stage Two: NPC Base Skill is 6
- Stage Three: NPC Base Skill is 9

Along with the subject there are two NPC opponents, as it is a deathmatch, all players are hostile, this means NPCs will attack each other as well as the subject. One NPC uses a pistol and the other uses an assault rifle. Both weapons have a sub model that affects attack distance and has its own attributes, such as clip size and bullet damage.

Subjects will be equipped with three different weapons, which can be selected by pressing 1,2 or 3 on the keyboard. All weapons have separate ammunition, weapon damage and bullet spread. Table 1 highlights the key attributes for each weapon.

Table 1. Weapon Details

Weapon	Description
Pistol	The pistol has high recoil and kickback, slow fire-rate, but deals high damage
Assault Rifle	This weapon has high fire-rate, medium damage, and moderate recoil
Shotgun	The shotgun fires 12 pellets per shot, each of which deal low damage, high recoil, and high pellet spread over range

There are two types of collectables scattered throughout the map. There are three medic-packs and three ammunition packs, the location for each collectable is the same across all instances and available to both subject and NPC. The collectable is collected by moving over it, after which it will be destroyed and then respawns 5 seconds later at the same location. The medic pack provides up to 50% health, depending how much missing health the character has, and the ammunition resupplies up to one clip worth for the currently equipped weapon. Figure 8 displays an example of a medic and ammunition pack.

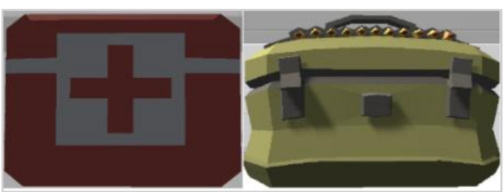


Figure 8. Health and Ammunition

The purpose of this experiment is to assess the effectiveness of the combat model in terms of adjusting the combat capabilities of NPCs based on the base skill level. This skill scaling is important because in a real-world setting, there is a range of player skill levels, and the model needs to be able to adapt to these different levels. If NPCs can increase or decrease their combat efficiency in real-time, it would allow for more challenging NPCs and provide a way to match the skill level of the player during a game.

a) Combat Model

The model is based on data and results from an experiment called the "Combat Behaviour Experiment." The goal of the model is to make the NPCs behave in a way that resembles human combat behaviour. The NPCs' combat efficiency is linked to a skill variable that can range from 1 to 10, with 10 representing very high efficiency and 1 representing very poor efficiency. In this paper, the NPC's skill level is set to 5, which represents an average level of efficiency based on the data from the Combat Behaviour Experiment. The concept of skill in this context refers to how well the NPC performs in combat compared to the average performance level observed in the experiment.

When creating the combat model, only standardised development techniques were considered during the modelling, which were suitable with the data captured from the combat behaviour experiment. As the experiment uses Unity3D, it was decided that scripting and object orientated programming would be the best solution, as the model will be using coroutines and multiple scripts accessing the model data at a given time.

It was important to capture some of the highs and lows that players often experience during a game. This is when players have a run of successful eliminations called ‘hot streak’, or a run of losses called ‘cold streak’. These streaks are directly linked to the combat efficiency of a player, therefore, the model tracks how well the NPC is currently performing and adjusts the active skill level to incorporate cold and hot streaks. However, it was important that there was a limit to how much influence a streak can have on the NPC, this prevents NPCs drifting to far from the initial base skill level. There are three main factors that affect the active skill level (Table 2).

Table 2. Active Skill Modifiers

Modifier	Description
Weapon	Like players, NPCs have a ‘preferred’ weapon/s. When they have a favourite weapon equipped, it increases the active skill by up to +1.
Death	Resets the active hot streak, if the NPC is not in a hot streak, they enter a cold steak adding -0.5 to the active skill.
Eliminations	Resets the active cold streak, if the NPC is not in a cold streak, they enter a hot steak adding +0.5 to the active skill.

The active skill is accumulative, with an upper limit of +2.5 and a lower limit of -2.5. Therefore, if an NPC has a base skill level of 5, the active skill level can be a maximum of 7.5 and a minimum of 2.5.

As NPCs could not change weapons during this experiment, it was decided to provide them with +1 to the weapon skill modifier.

The activate skill is determined by the base skill, weapon modifier and current streak. This allows for some flexibility in how the NPC performs, rather than a static or manufactured combat efficiency. The active skill directly features in the equations for reaction time and accuracy.

Equation (1) shows how active skill was used to control the aiming speed, considering the location and size of the target in the field of view of the game camera.

$$\text{Reaction Time} = \frac{Ts + \left((S * M) * Dx\right)}{Dz} \tag{1}$$

Where, *Ts* is the target size, *S* the active skill attribute, *M* is a generic modifier, *Dx* the 2D distance from the centre of the screen to the target and *Dz* the Euclidean distance from the NPC to the target. This represents a comprehensive approach to targeting because it considers not only the size of the target (*Ts*) but also the distance (*Dz*). This was required because the further away the target was in game, the smaller it would appear on the screen. This underscores the importance of precision when determining the initial reaction targeting speed.

The Equations (2) and (3) are for accuracy, it was decided to separate the two shooting stances as looking down sights should provide better control than shooting off the hip.

$$\text{Aiming} = \text{Vp} + \text{Vf} * (\text{Mx} + (\text{My} * \text{S})) \qquad (2)$$

Vp is the vector position and *Vf* the vector forward direction of the NPC, *Mx* and *My* are generic modifiers and *S* the active skill of the NPC.

$$\text{Off Hip} = \frac{\text{Vp} + \text{Vf} * (\text{Mx} * \text{S})}{Sx * Ws} \qquad (3)$$

Like the aiming equation, *Vp* is the vector position and *Vf* the vector forward direction of the NPC, *Mx* a generic modifier and *S* the active skill of the NPC. However, *Sx* is the locale scale of the crosshair and *Ws* is the equipped weapon bullet spread. As previously stated, the bullet spread is focused on the crosshair, and the longer the weapon is continuously fired, the bigger the crosshair grows, making it more inaccurate.

D. Combat Perception Experiment

The perception experiment was developed in the same way as the dynamic difficulty experiment. It consisted of a deathmatch in which the subject will be against two NPCs in an all vs all scenarios. As the intention is to determine if subjects can identify inconsistencies in the combat, both NPCs have a static skill level of 5. This represents an averaged skilled opponent, and it was important to maintain an even level of combat efficiency so that perception was consistent across all subjects.

The objective of the experiment was to achieve 8 eliminations before any of the NPCs. There were three combat perception focused areas:

- Omniscient Player Awareness: When the NPC is ambushed and attacked from behind, the NPC will be given the location of the attacker. This will enable them to have omniscient awareness of its target in combat only.
- Aggressive Combat Personality: During combat the NPCs will make a beeline towards the subject, attacking as it moves. The NPC will not break off combat until its target is eliminated or it loses sight of the target for a period.
- Non-Smooth Pathing when Roaming: During roaming the path smoothing has been removed. This is where when the NPC requests a path, no extra smoothing technique has been used. This makes NPCs movement somewhat jagged/jerky at times, when scrutinised from an observer.

These perceptions were specifically designed as they target different areas of interaction with players. The omniscient player awareness will determine if subjects can identify that they have no advantage when trying to ambush the NPC. This is essential because it will show that players have an understanding about opponent reaction times, and when they know their opponent has not seen them, the reaction behaviour should different than when they know they have been spotted.

The aggressive combat personality will determine if subjects can identify and form opinions of opponents attacking behaviours. If subjects identify the aggressive attacking, it will mean that to fully encapsulate believability, NPCs will need to have a range of combat personalities in which how the engage in combat will impact if they are believable.

The non-smooth pathing when roaming will test the observational ability of subjects when not in combat. The main purpose is to see how much attention subjects give to their opponents, or if the main concern is focused on maximising their combat advantage. If subjects can notice the NPCs sometimes jagged pathing, it will mean that all areas of the NPCs behaviours will be under scrutiny by players. This has a severe impact on the modelling because it means that a consistent level of detail will be required across all gameplay.

VI. Results

According to the results of the study, it is effective to model human behaviour in order to create NPCs with dynamic skills in a video game. It is also possible to generalize the efficiency of combat in this way. Additionally, the method of skill scaling, which adjusts the combat efficiency of NPCs in real-time gameplay, was found to be very effective. However, when evaluating the level of awareness possessed by players, specifically when in combat situations, the results was mixed. Furthermore, some subjects had negative opinions regarding the combat, which were based on a false awareness that NPCs had perfect accuracy.

a) Combat Model Analysis Comparison

The analysis of the combat model shows that it was generally well received, but there are some areas of the combat that need more attention. The purpose of this analysis is to determine two key objectives about the combat model:

- Statistically: How does did the model perform from numerical perspective, when compared to human subjects.
- Perceptually: What was the overall perception of the combat behaviour of NPCs and how human-like did they appear.

The results suggest subjects were fairly accurate in their perception of NPC combat efficiency because the combat model was over tuned which increase NPC accuracy.

The graph in Figure 9 illustrates the average shooting accuracy for both NPCs at each stage, as well as the overall shooting accuracy for all subjects when playing against NPCs with a base skill of 3 (stage 1). The data suggests that while the scaling was effective, the NPCs may have been too difficult in general.

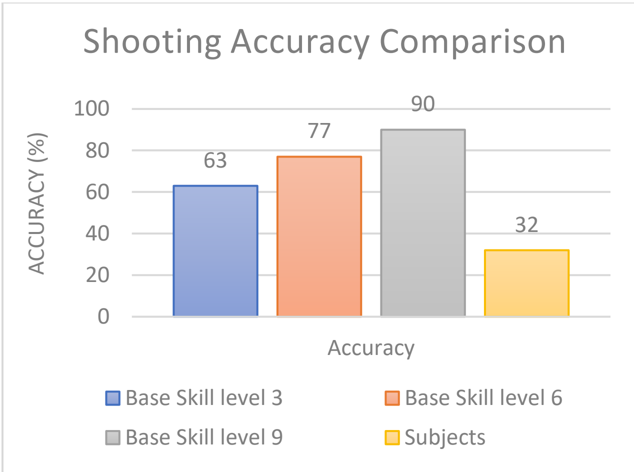


Figure 9. Shooting Accuracy Comparison.

Figure 10 displays the feedback from subjects when asked which part of the NPCs combat behaviour was the weakest. This was a multichoice question, therefore, subjects were able to select more than one aspect of combat. The accuracy relates to the overall shooting accuracy of NPCs, the targeting refers to how much the NPCs lock on to subjects, player awareness is focused on how much omniscient awareness did the NPC appear to have, and reaction speed in how quickly NPCs could aim and shoot when spotting a subject.

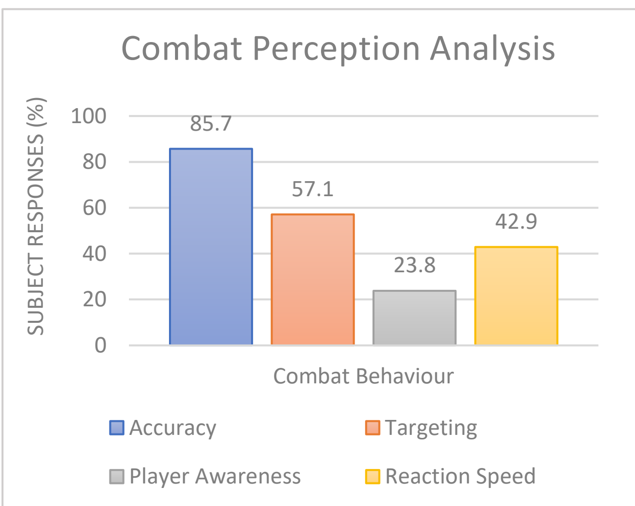


Figure 10. Combat Perception Analysis.

The results highlight that accuracy was significantly the most pressing issue, with more than 85% of the subjects stating it was not human-like. This is followed by targeting with more than half of the subjects suggesting NPCs had unnatural targeting in combat. The reaction time of the subjects was higher than expected. This indicates that the subjects had difficulty competing with the NPCs. One potential solution to this issue is to lower the modifier in the reaction time equation. However, more testing will be needed to determine the optimal adjustments to the model.

When asking subjects about how believable the NPCs appeared, in relation to human players, the results were mixed. Figure 11 shows that none of the subjects thought the NPCs were human-like, but 50% thought they were somewhat human-like.

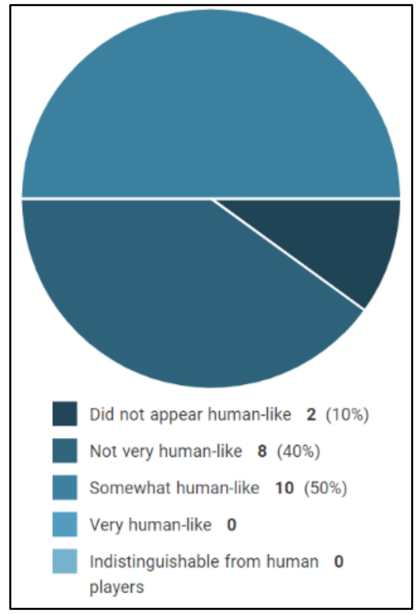


Figure 11. Combat Feedback.

While the feedback did not state that the model had human-like combat behaviour, it did show that potentially, the model could be tweaked and revised to be more accurate.

b) Gameplay Influence

When looking at the feedback, it became clear that other gameplay elements were influencing people's perceptions of the combat's realism. This meant that issues with the NPC's navigation were causing subjects to have negative views on the NPC's combat abilities. This is highlighted when analysing the specific responses from subjects when asked about non-combat elements. One of the responses, shown in Figure 12, emphasises that other gameplay aspects like navigation can affect how believable the combat appears.

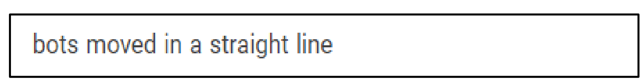


Figure 12. Specific Combat Subject Feedback.

This shows that different gameplay elements are interconnected. When one aspect is not modelled well, it can affect the perceptions of other elements and cause undesirable results. This means that when creating models for certain behaviours, it's important to consider how they may impact other behaviours in certain situations.

c) Skill Scaling

The skill scaling seemed to work well based on the feedback, when asking the subjects if they noticed the difficulty increasing as they progressed through the stages (Figure 13). Of the respondents, 57.1% noticed an increase in difficulty, while 42.9% did not notice if the difficulty changed as the stages progressed.

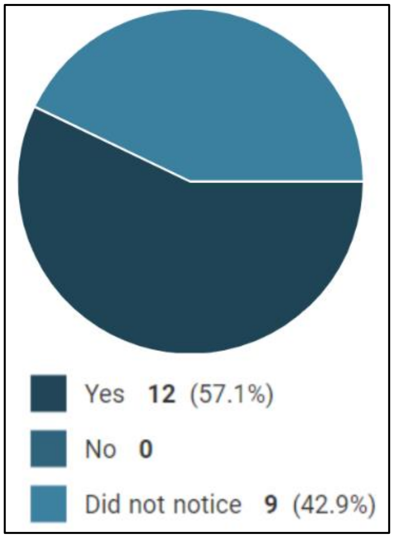


Figure 13. Difficulty Identification Feedback.

This suggests that the combat model could be a viable option for creating NPCs with adaptable real-time difficulty. It should be noted that this is the perception of difficulty, and it could be suggested it is common for a player to experience higher difficulty when getting familiar with a game. However, when analysing the combat efficiency of NPCs, the data shows they did become better, and this could have been noticed by the subjects. This would support the idea that players have some understanding of the combat abilities of their opponents. As a result, this would emphasize that a one-size-fits-all approach to difficulty will not work for all players.

When looking at the pistol combat performance for the three different difficulties, a pattern emerges where the individual combat behaviours improved as the base skill level increased (Figure 14).

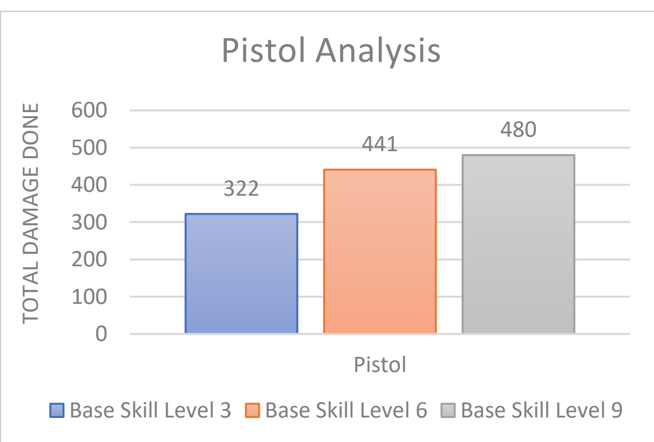


Figure 14. Pistol Damage Done Skill Comparison.

This indicates that as the base skill level increased, the average amount of damage inflicted by the NPC also increased. This suggests that the NPC's combat efficiency improved with a higher base skill level. This is supported by the analysis of the average initial reaction time for the assault rifle across all three stages (Figure 15).

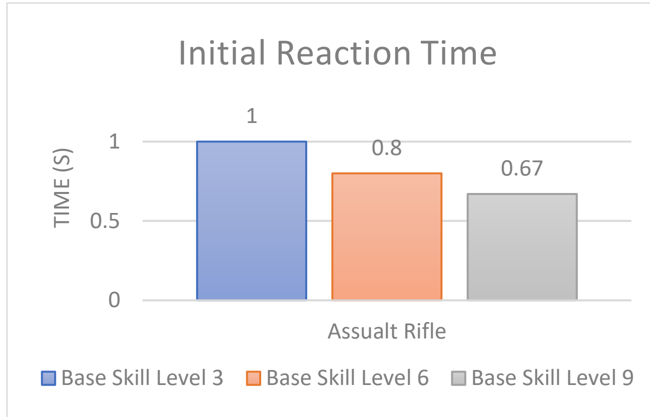


Figure 15. NPC Skill Reaction Time Initial Comparison.

The results suggest that there is a small relationship between reaction time and active skill level. While there is some variation, this could be caused by changes in active skill level due to the triggering of hot and cold streaks.

Figure 16 looks at the first and third stages only and shows the average number of eliminations for each character. On average, the subjects did not perform very well and the NPC with the assault rifle was particularly effective, especially in the first stage.

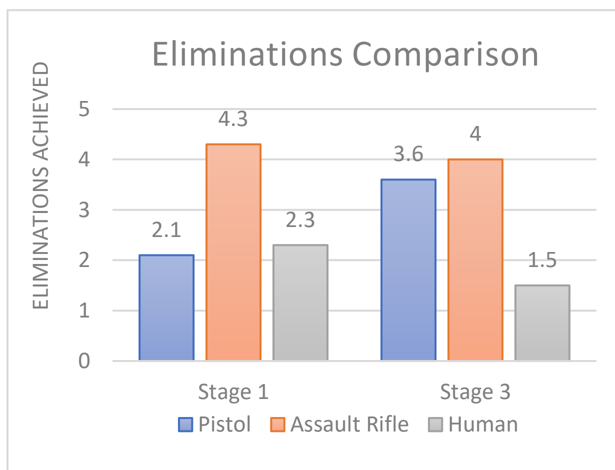


Figure 16. Eliminations Comparison.

This data supports the idea that the NPC's accuracy was too high and will be less effective after balancing. However, when looking at the NPC with the pistol, the data showed that it improved as its skill level increased. This had a noticeable impact because both the subjects and the assault rifle NPC's elimination count decreased. While further experimentation and refinement of the model is necessary, this is a promising result.

This model and set of algorithms are innovative because they could allow the NPC's combat efficiency to be directly affected by the skill level of the player. More research is needed to accurately track the skill of the player in real time so that the NPC can adjust its own skill level to create a personalized difficulty experience that is tailored to the player's individual skill level.

d) Perception

The results indicate that subjects had a good sense of awareness. When specifically analysing the omniscient player awareness test, most of the subjects stated not only did they notice the unnatural ability to locate the subject when they are ambushing, but it has a detrimental effect on how the NPC was perceived. Figure 17 shows a selection of responses from subjects when they were asked which aspect of the NPCs combat was the least believable when compared to a human.

the aim was pretty cracked
Locating players from long range.
As said before, they target you very quickly no matter where you are related to where they're facing.
the accuracy
Accuracy
The AI always had better awareness than me and out shot me
laser like accuracy

Figure 17. Combat Perception Feedback

This is further supported when highlighting the subject responses when asked a multichoice question about key combat behaviours from the combat perception experiment only (Figure 18).

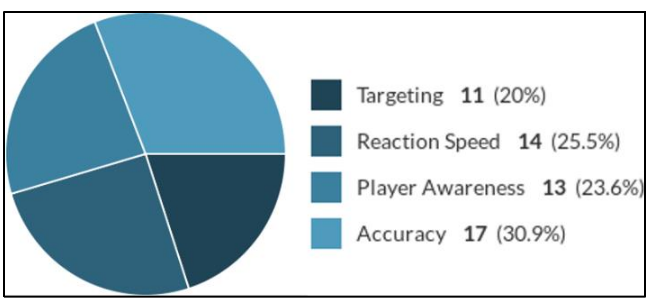


Figure 18. Combat Multichoice Feedback

This negative reaction shows that when developers try to take shortcuts on NPC combat modelling, the results can severely impact the perception of the NPC. This is further supported when discussing the aggressive combat personality because a selection of subjects commented on how they noticed NPCs would move in such a way, where it appeared, the NPC was always committed to a destroy or be destroyed mentality (Figure 19).

the fact it did not run to get ammo when it ran out of mach9ine gun bullets and switched in an inferior weapon
Returning fire at long range instead of taking cover.
They could've been hit many times and they would still attempt to kill you when they have very little chance to win in the engagement.

Figure 19. Specific Feedback

This feedback highlights that subjects' perceptions may not always be correct or a fair reflection of the NPCs. The comment about NPCs not acquiring ammunition is factually incorrect, as the data shows NPCs consistently looked for ammunition when they were below a threshold. When scrutinising the feedback further, it became apparent that subjects' opinions about some of the NPC's characteristic was wrong. The main complaint about the NPC combat was that they were overly accurate and never missed, however, the data shows that while the accuracy was higher than an average subject, NPCs did not have perfect accuracy when shooting. This adds complexity to the modelling because if the subjects' perceptions do not match the numerical data, it could mean the modelling will need to account for incorrect perceptions.

It is interesting that subject perceptions were significantly higher in combat situations, than when they were observing the NPCs from a distance. Figure 20 shows the feedback from subjects when asked to comment on the navigation performance of the NPCs, and which part of the pathing finding was the least believable.

They always seemed to move in a straight line and never tried to dodge or strafe
Straight line walking
they walked directly towards me and often stood still while shooting at me
Seemed to move in dead straight lines
move towards me in combat
I guess they moved in straight lines
Looked 'off'
bots moved towards me attacking, its not realistic

Figure 20. Navigation Feedback

There is no commenting on the NPCs having a jagged pathing when roaming, but significant feedback on how the NPCs moved in combat. This suggests that perceptions are not even across all aspects of gameplay, instead the subjects were more likely to perceive something as 'off' when directly interacting with the NPCs.

VII. Conclusion

The objective of this paper was to develop a dynamically skill scaling combat modelled based off data acquired from human subjects.

This paper has shown that the combat model provided a good solution for modelling human combat behaviours, however, more research is required to improve the modifier weight in relation to the active skill of the NPC. There is evidence to suggest that modelling combat alone would not be enough to create dynamic NPCs, several subjects referred to the pathfinding/navigation when asked to provide feedback on the combat behaviours. Therefore, it could be argued that gameplay behaviours are uniquely intertwined and when one gameplay behaviour is poorly modelled, it can influence other behaviours.

The skill scaling showed a positive result, the data highlighted that there is a somewhat linear change in reaction time and accuracy, as the skill of the NPC was increased. This was noted by 57.1% of the subjects, with zero subjects stating that they thought the NPCs did not increase in difficulty. However, the combat efficiency of the NPC was deemed to be too high, even on the easiest difficulty stage, this was supported when analysing the number of eliminations achieved and the NPC with the assault rifle significantly outperformed the subjects.

The perception feedback indicated that subjects had specific awareness about their opponent when engaging in combat. The data suggests that subjects knew there was something unnatural when NPCs were able to know where they were when they were ambushed. This proved that subjects were able to assess the actions of an opponent during real time, and when something was 'off' it impacted how they viewed the opposition. However, there was data to conclude that not all the opinions were correct, and the incorrect assumptions influenced perceptions. This presents an interesting problem because when modelling combat for NPCs, should they be modelled on human gameplay, or on how players 'think' the NPCs should behave. There is the possibility that the problem could be with the experimental protocol, and that by telling the subject their opposition is an NPC, it subconsciously influences their perception. Therefore, an experiment should be undertaken where the subject does not know whether they are competing against NPCs or human.

VIII. Future Work

As the results indicated the combat behaviour was negatively impacted by the other gameplay behaviours, research is required to the extent of this influence. This means navigation and decision-making will need to be modelled then the experiment run again, and results compared.

Identifying real-time player skill level could further enhance the effectiveness of dynamic skill scaling, research in this area could help determine if dynamically adjusting NPC skill based on the current player perform provides a more engaging experience.

References

[1] P. Williamson and C. Tubb, "Modelling Player Combat Behaviour for Dynamic Difficulty Scaling in First Person Shooter Games," The Fourteenth International Conference on Advances in System Simulation. pp. 22-28, Lisbon, Portugal, October 16-20, 2022.

[2] H. Warpefelt, "The Non-Player Character: Exploring the believability of NPC presentation and behavior," (Doctoral dissertation, Department of Computer and Systems Sciences, Stockholm University), 2016.

[3] P. Hingston, "A turing test for computer game bots," IEEE Transactions on Computational Intelligence and AI in Games, pp.169-186, 2009.

[4] P. Lankoski and S. Björk, "Gameplay Design Patterns for Believable Non-Player Characters," *DiGRA Conference*, pp. 416-423, 2007.

[5] P. Williamson and C. Tubb, "Modelling Player Combat Behaviour for NPC Imitation and Combat Awareness Analysis," *ECMS*, pp. 205-212, 2021.

[6] J. Orkin, "Three states and a plan: the AI of FEAR," Game developers conference pp. 04, 2006.

[7] C. Pezzato, C. Hernandez, S. Bonhof, and M. Wisse, "Active Inference and Behavior Trees for Reactive Action Planning and Execution in Robotics," arXiv preprint arXiv:2011.09756, 2020.

[8] R. A. Agis, S. Gottifredi, and A. J. García, "An event-driven behavior trees extension to facilitate non-player multi-agent coordination in video games," Expert Systems with Applications, pp. 155, 2020.

[9] X. Neufeld, S. Mostaghim, and S. Brand, "A hybrid approach to planning and execution in dynamic environments through hierarchical task networks and behavior trees," In Proceedings of the Fourteenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment. AAAI Press, Article 29, pp. 201–207, 2018.

[10] E. Camilleri, G. N. Yannakakis, and A. Dingli, "Platformer level design for player believability," Conference on Computational Intelligence and Games (CIG) pp. 1-8. IEEE, 2016.

[11] M. Polceanu, A. M. Mora, J. L. Jimenez, C. Buche, and A.J. Fernandez-Leiva, "The believability gene in virtual bots," The Twenty-Ninth International Flairs Conference, pp. 346-349, 2016.

[12] A. M. Mora, A. Gutiérrez-Rodrıguez, and A. J. Fernández-Leiva, "Optimising Humanness: Designing the Best Human-Like Bot for Unreal Tournament 2004," In International Work-Conference on Artificial Neural Networks, pp. 681-693. Springer, Cham, 2017.

[13] R. R. Wehbe, E. Lank, and L. E. Nacke, "Left the 4 dead: Perception of humans versus non-player character teammates in cooperative gameplay," In Proceedings of the 2017 Conference on Designing Interactive Systems pp. 403-415, 2017.

[14] J. D. Smeddinck et al., "How to present game difficulty choices? Exploring the impact on player experience," CHI Conference on Human Factors in Computing Systems pp. 5595-5607, 2016.

[15] M. Hendrix, T. Bellamy-Wood, S. McKay, V. Bloom. and I. Dunwell, "Implementing adaptive game difficulty balancing in serious games," IEEE Transactions on Games, 11(4), pp. 320-327, 2018.

[16] P. M. Blom, S. Bakkes, and P. Spronck, "Modeling and adjusting in-game difficulty based on facial expression analysis," Entertainment Computing, 31, p. 100307, 2019.

[17] S. Bakkes, C. T. Tan, and Y. Pisan, "Personalised gaming: a motivation and overview of literature," In Proceedings of the 8th Australasian Conference on Interactive Entertainment: Playing the System, pp. 1-10, 2012.

[18] P. H. M. Spronck, I. G. Sprinkhuizen-Kuyper, and E. O. Postma, "Difficulty scaling of game AI," In A. E. Rhalibi and D. Van Welden, editors, Proceedings of the 5th International Conference on Intelligent Games and Simulation

(GAMEON'2004), pp. 33–37. EUROSIS-ETI, Ghent University, Ghent, Belgium, 2004.

[19] P. Demasi and A. J. de. O. Cruz, "Online coevolution for action games," International Journal of Intelligent Games and Simulation, 2(3), pp.80–88, 2002.

[20] G. N. Yannakakis and J. Hallam, "Towards optimizing entertainment in computer games," Applied Artificial Intelligence, 21(10), pp. 933–971, 2007.

[21] R. Vicencio-Moreira, R.L. Mandryk, and C. Gutwin, "Now you can compete with anyone: Balancing players of different skill levels in a first-person shooter game," In Proceedings of the 33rd annual ACM conference on human factors in computing systems, pp. 2255-2264, 2015.

[22] S. Hladky and V. Bulitko, "December. An evaluation of models for predicting opponent positions in first-person shooter video games," IEEE Symposium on Computational Intelligence and Games, pp. 39-46, 2008.

[23] J.J. De Simone, T. Verbruggen, L.H. Kuo, and B. Mutlu, "Is cheating a human function?" The roles of presence, state hostility, and enjoyment in an unfair video game. Computers in Human Behavior, pp.2351-2358, 2012.

[24] S. Pontiroli, "The cake is a lie! Uncovering the secret world of malware-like cheats in video games," VB2019, pp.2-4, 2019.

[25] D. Choi, T. Konik, N. Nejati, C. Park, and P. Langley, "A believable agent for first-person shooter games," In Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, Vol. 3, No. 1, pp. 71-73, 2007.

[26] J.M.L. Asensio et al., "Artificial intelligence approaches for the generation and assessment of believable human-like behaviour in virtual characters," Expert Systems with Applications, 41(16), pp.7281-7290, 2014.

[27] D. Conroy, P. Wyeth, and D. Johnson, "Understanding player threat responses in FPS games," In Proceedings of the 9th Australasian Conference on Interative Entertainment: Matters of Life and Death, pp. 1-3, 2013.

[28] C. Rayner, "Player Modelling for Cursor-Driven Games," [Unpublished], 2007.

Simulation of the Emergency Department Management during the Pandemic

[1,2,4]Ramona Galeano, [1]Dolores Rexachs, [1]Alvaro Wong, [5]Eva Bruballa,
[2]Cynthia Villalba, [3]Diego Galeano, [1]Emilio Luque

[1]*Computer Architecture and Operating Systems Department*
Universitat Autónoma de Barcelona (UAB)
Barcelona, Spain
e-mail: RamonaElizabeth.Galeano@autonoma.cat, Dolores.Rexachs@uab.cat
alvaro.wong@uab.es, eva.bruballa@eug.es, emilio.luque@uab.es
[2]*Facultad Politécnica*
[3]*Facultad de Ingeniería*
[4]*Facultad de Ciencias Médicas*
Universidad Nacional de Asunción
San Lorenzo, Paraguay
e-mail: cvillalba@pol.una.py, dgaleano@ing.una.py
[5]*Escuela Universitaria de Informática*
Escuela Universitaria Gimbernat

***Abstract*—The COVID-19 pandemic has caused disasters worldwide, overwhelming public health systems and generating actions such as movement restrictions and containment orders. It has strained public health systems and exposed the healthcare needs and gaps for marginalized and vulnerable populations. Modeling and simulation can help make a physical or logical representation of a system to generate data and determine decisions or predict a given system or problem. In this paper, we present an adaptation of our previous work on a simulation of an agent-based model to simulate an emergency department for a different hospital in a pandemic situation; we compare some of the results we obtained from the simulations with reality to help the management of the emergency department.**

***Index Terms*—simulation; agent-based model; COVID-19; management**

I. Introduction

Pandemics occur when a new virus emerges for which the necessary natural defenses are not possessed, so it spreads rapidly, sometimes with disastrous results. It is a disease outbreak that spans several countries and continents, affects many people, crosses many borders, exceeds the expected cases, and persists over time. Pandemics are often caused by viruses, such as COVID-19, which can be easily spread from person to person.

The COVID-19 pandemic caused a severe economic, social, and health crisis never seen before. They cause unemployment, leading to an exponential increase in poverty, and their repercussions on developing unprecedented famines.

It constitutes a disruptive situation that generates high individual and collective stress levels. For many people, it implies a tragic situation due to the losses they must face: loss of loved ones, health, housing, property, or employment.

Many countries suffered many consequences from the crisis, with overloaded health systems, unemployment, business failures, and other problems. This is particularly true in the poorest areas, where the pandemic has exposed deep-seated inequalities.

The COVID-19 pandemic generated a shock wave that affected the entire world economy and triggered the biggest crisis in over a century. This led to a drastic increase in inequality within and between countries.

The COVID-19 pandemic affected different aspects of life and caused many collateral problems, such as biological damage and mental health problems. It has also generated unique ethical dilemmas stemming from the various demands on society generated by many of the decisions made and the existing health system.

Health professionals must make decisions about allocating scarce resources that can eventually cause moral distress and affect their mental health and patients.

Another edge has been the restrictions on freedom of movement, which has forced the closure of entire economies to flatten the epidemics' curve. Even the recommendation of good hygiene practices, proper hand washing, and physical distancing, which seem easy to comply with, may be challenging in the poorest societies.

Finally, some will question the ethics behind the search for effective treatments and the development of vaccines; research carried out at a time of uncertainty and anguish.

Most countries have been affected, including developed and developing countries. COVID-19 has caused death and exposed the severe limitations of the countries' health systems.

Emergency departments must be prepared to handle crises and disasters such as the pandemic. During this time, they required a quick solution to give a sustained response to the problem that came ahead.

Simulation can be used effectively to create or improve a system or process. It is beneficial to build hypothetical scenarios; it is used in many areas and also for emergency departments.

The main methods used for simulation in emergency departments are agent-based simulation [1], [2], discrete events [3], and system dynamics [4]. The advantage of simulation is facilitating the automatic search for scenarios that can provide the best solutions given a set of constraints and future states. This automation of the search for improvements to an emergency department can significantly help managers who need answers to problems.

We have previously developed an agent-based model for emergency departments; some of our works are: creating a simulator using NetLogo [5], performance optimization of these emergency systems [6], micro-level behavior of the emergency department for the prediction of characteristics at the macro level [7], contact transmission of MRSA between hospitalized patients [8] among others. An agent-based model is an approach to model systems where individual agents interact. It offers ways to more easily model interactions of individuals and also how these interactions affect other agents in the system [9].

Our work presented here builds upon the work of Liu et al. [10] that developed a simulator for planning and management of the emergency department, which has the advantage of having been verified and validated in several cycles or iterations, taking into account a wide variety of data and configurations, and with the participation of emergency department staff at the Hospital of Sabadell (Spain).

This work adapted an agent-based model that allows modeling the management of an emergency department during a pandemic such as COVID-19. The emergency department incorporated doctors and nurses from other specialties to support the emergency area, manage more resources (adding beds or boxes to increase the number of people who can be treated), and the severity levels of the patients to a pandemic situation (COVID-19), where the patients who attended were the most serious.

To adapt the incorporation of more personnel to the emergency department, our simulator was configured, which has personal resources as a configurable parameter and also considers two levels of junior and senior experience: juniors (with limited knowledge or low experience) and seniors (with experience); we use the two levels that the simulator already has to map the two types of classification on them doctors used during the pandemic. This is very important because during the COVID-19 pandemic, in addition to the emergency intensivists (seniors), health personnel from different specialties (juniors versus intensivists) were incorporated into the emergency department.

For the experimental part, we have prepared a set of synthetic input data for the simulation, arrival distribution of patients with COVID and NON-COVID with different levels of severity, and general distribution by the age of the simulated patients; with these input data, we have analyzed the occupation time of the doctors and nurses, the waiting time of the patients, the service time according to the level of severity. Then we compared it with reality, obtaining very promising results.

Our simulator describes the behavior of the emergency department during the COVID-19 pandemic and can help hospital personnel as a decision support system for emergency department management.

The paper is organized as follows. In section II, we explain related works. In section III, the adaptation of the emergency department model. Section IV describes the simulation. Section V presents the experiments and discussion, and section VI presents the conclusions and future work.

II. Related works

The simulation topics most frequently found in the literature in the COVID-19 simulation area are studies for contact tracing with COVID-19, transmission models of healthy patients with an infectious spread in health systems [11], patient flow improvement [12], how simulation modeling can help reduce the impact of COVID-19 [13], among others.

Some simulation methods found that were used in the area of COVID-19 simulations are discrete events [12] [14] [15], artificial intelligence [16] and agent-based simulation [17]. Some countries used the simulation to predict scenarios, including the behavior of the Delta variant to know the number of deaths, infected, and vaccinated infected; others to see the evolution of COVID-19; others to establish the infected, quarantined, recovered, and dead, using the Susceptible Exposed Infected Asymptomatic Quarantined Recovered (SEIAQR) model [18]. Other work presents an approach in health care in which combined techniques of discrete event simulation, simulation-based multi-objective optimization, and data mining are used [19].

There is an evolution in the literature on applying agent-based models to emergency department operations. The agent-based model (ABM) has grown tremendously over the last 15 years and, more recently, in hospitals and healthcare. One of the main applications of ABM in hospital environments is that it examines the flow of patients in emergency departments [20], [21].

The studies about the emergency department deal with the practice of protocols or objects for medical procedures [22]. Another job in the emergency department involves managing the resources in intensive care, intensive care beds, and their devices [14]. Length of stay and patient waiting times, optimizing resources [19].

Previous works in the emergency department area are: Create a simulator for the emergency department with the participation of the Sabadell Hospital emergency staft [5]. Active agents, passive agents, and the environment are identified, and an initial simulation is created using NetLogo [5]. Another task is to optimize the emergency departments' performance [6]. Extensive search optimization is used to find the optimal configuration of emergency department staff, a multi-dimensional, multi-objective problem [6]. An index is proposed to minimize the patients' length of stay in the emergency department. The results obtained using alternative Monte Carlo and Pipeline schemes are promising [6]. This work presents a layer-based application framework for discovering knowledge of an emergency department system by simulating micro-level behaviors of its components [7]. This work proposes using a simulation tool, the MRSA Simulator, to design and conduct virtual clinical trials to study contact transmission of MRSA among hospitalized patients [8].

The difference between our job and the others is that this model of COVID-19 in the emergency department simulator will allow the emergency department managers to analyze and evaluate potential solutions for the beds and the management of human resources such as nurses and doctors. And also has been considered at two levels with experience and without experience (juniors and seniors); when reinforcing, the health personnel treating COVID have incorporated emergency intensivists (seniors) and health personnel from different specialties (juniors). And to evaluate the effectiveness of different combinations of scenarios. Many countries have been experiencing extreme stress with patients unable to access therapy beds, dying in emergency department corridors while waiting for beds to be released, and a lack of experienced doctors.

III. Adaptation of the Emergency Department Model

This section presents a model for the emergency department during the COVID-19 pandemic. The general objective of this research is to propose a model that allows to expand the simulators' functionality to adapt it to changes in the emergency department operation when exceptional situations occur, such as a health alert, or restrictions, and when extraordinary temporary measures are adopted. As is the case of pandemics spread by air, in such a way that it helps in the planning and management of the service. First, our simulator was adapted to another hospital, verifying its operation, and then it was adapted to the pandemic.

The first step of the work consists in a description of a conceptual model of the systems' operation, from which the computational model that will allow the system to be simulated is elaborated. It is planned to use the simulation environment and a high-level platform.

A. Active agents

The active agents are the individuals who act dynamically; they are all human actors in the emergency department. They are:

Patients: The essential individuals in the system.

Admissions staff: The person, the patient, goes to request an appointment, update their data, and request the opening or search of their medical record.

The triage nurse: Responsible of the assignment of the acuity level to patients.

Doctor: They interact with patients to diagnose and treat them.

Nurse: They provide and supervise treatment to the patient and take and send tests to laboratories.

Laboratory Staff: These are the persons who perform the tests and analyze the patient if necessary.

B. State variables

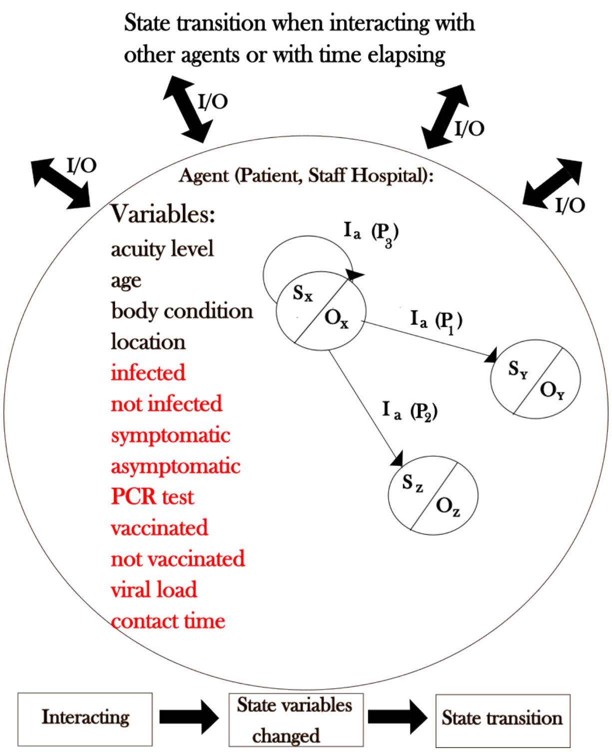


Fig. 1. State transition when interacting with other agents or with time elapsing

The agents move from one place to another by interacting with other agents. During this time, each agent changes its state due to the interactions. A state machine perfectly represents this behavior, so they have chosen a state machine to model all agents. Specifically, the agents are characterized by a probabilistic Moore machine. An initial set of state variables defined through the round of physician interviews is based on the minimum amount of information needed to model each patient and staff. An initial set of state variables is shown in Figure 1 for the agent patient and the hospital staff (admissions staff, doctor, triage nurse, laboratory staff, and nurse) [23].

Variables must be incorporated into each agent; the variables included by the agents' patient and hospital staff (admissions staff, doctor, triage nurse, laboratory staff, nurse) are acuity level, age, body condition, and location. And the new variables are infected or not, symptomatic or not, vaccinated, viral load, contact time, and PCR test.

The agents are divided by their state variables and their behaviors. The values of the state variables of an agent at a given time, t, define the agents' situation at that time, t. The behavior of each agent depends on the category to which it belongs and is determined based on the rules previously assigned to each. To represent the different states of the agents during the attention process are used finite state machines.

The agents' state transition from one state to another will then be determined by (a) the current state and (b) the input value it receives due to the interaction with another agent, always considering that this value will be granted based on a previously defined probability.

C. Output

Some of the outputs are the length of stay (LOS), the length of waiting (LOWT) for each stage (e.g., waiting time for service request: wtsr, time of admission: at, waiting time in admission: wta, waiting time in nursing: wtn, time nursing care: twnc, waiting time in doctors' treatment: wtd, doctor treatment care time: twmd and others), destination, acuity level, infected, symptomatic, PCR test, vaccinated, viral load and the occupations of hospital staff (admissions staff, doctor, triage nurse, laboratory staff, and nurse).

IV. Simulation

An initial simulation is created to verify the proposed model designed, using the NetLogo [24] agent-based simulation environment, a high-level platform especially suited for modeling complex systems that develop over time. NetLogo [24] allows visualizations of actions and agent interactions, an essential aspect considering that a primary use of the tool is gathering feedback from the emergency department.

The emergency department is divided into different zones where other agents can act, maintaining interactions that can also be different. The input to our model is a group of patients arriving in the emergency department. After the patients' arrival and the admission staff completes the registration, based on the seriousness of their situation in the triage, the patients are categorized, considering their acuity level. There are five different values, level 1 is for the most critical condition, and level 5 is not urgent [25]. The original simulator has the following areas, as shown in Figure 2: admissions area, triage area, diagnosis-treatment area, waiting rooms, etc. After triage, patients with diagnosed acuity levels 1, 2, and 3 are treated separately and assigned to Area A, and patients 4 and 5 are treated in Area B. The adapted simulator has the same areas but with different resources, for example, more boxes, doctors, and nurses, as shown in Figure 3 has the same areas but with different resources and added variables.

$$\begin{aligned} n_{scenarios} = & n_{admissions} * n_{timeadmissions} * n_{triagenursing} * \\ & n_{timetriagenursing} * n_{nursing} * n_{timenursing} * \\ & n_{doctor} * n_{timedoctor} \end{aligned} \tag{1}$$

The scenario adopted for this initial stage is to simulate the patients who move through the emergency department. The areas and types of active agents represented in this simulation are patients, admission staff, triage nurses, doctors, auxiliary staff, laboratory tests, internal tests, external tests, ambulance, and carebox. Each combination of values represents a different scenario simulation. Wide varieties of values make up the parameter space. The parameters can generate many different scenarios (1).

There is many parameter value combinations, large enough that there is no possibility of casting each one manually. For this reason, parametric simulations are required, in which the simulator launches a set of simulations with different combinations of parameter values.

In general, the time to compute a time interval of a simulation based on agents is the product of the time it takes to simulate the actions of an agent within the world of simulation in this step. In the model described, agents in the simulation are the hospital staff and patients. The simulator will be conducted by time. Time is divided into discrete, identical intervals and periods at each time step of the agents' operating system. Each time step is divided into two phases. Assuming that the simulator this at time t, the phases are: First, each agent processes the inputs of the last phase (It-1) and, according to that input and the state, as it was during the last step (St -1) and changes to its new state St. Second, each agent emits its output to its current state, Ot. This output uses receivers to switch to the next state. In time each agent changes state. It may change to the same state it was previously, but there is a change nonetheless. The metrics that are to be used for each state input It and output Ot are: waiting time to request service: twrs, time for register a required service: trs, time admission: ta, waiting time in admission: twa, waiting time in nursing: twn, time nursing care: te, waiting time in doctor: Twmc, health care time: tm.

$$Pi = f(LOS, age, level) \tag{2}$$

$$\sum_{i=1}^{n} P_i = 100 \tag{3}$$

$$P' = f'(TOT, age, level) \tag{4}$$

$$\sum_{i=1}^{n} P'_i = 100 \tag{5}$$

To generalize the process of all patients, probability distribution during simulation will decide the following status. The distribution model of the probability was based on the statistical data from the emergency department. Figure 4 indicates the general process during the patient stay in the emergency department; P1(%), P2(%), P3(%), and Pn (%) represent the probability of the following state transition separately, equation (2), (3), (4), (5) show the formulae [23]. All of the probabilities follow some probability distributions. The distributions' probability density function is decided by several key parameters based on the statistical analysis of the doctors' decision and the patients' behavior; a tuning process estimates the value of these parameters from actual historical

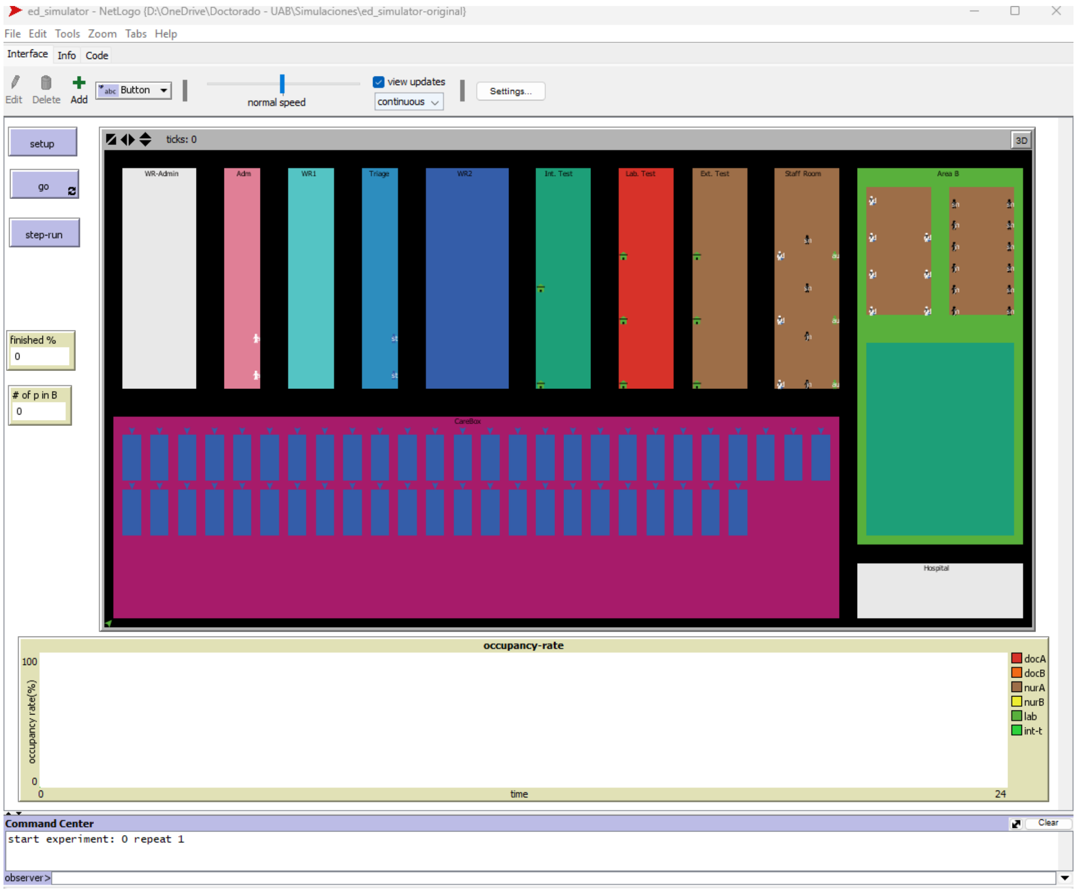


Fig. 2. Visualization of the original simulator in NetLogo before the adaptation

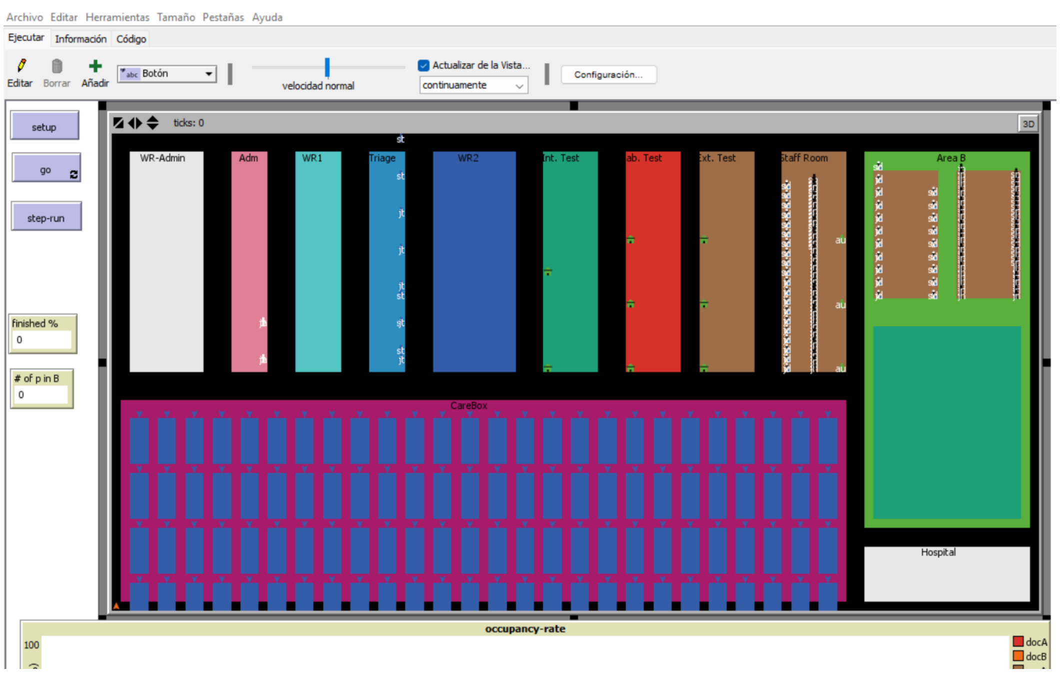


Fig. 3. Visualization of the simulator in NetLogo after the adaptation

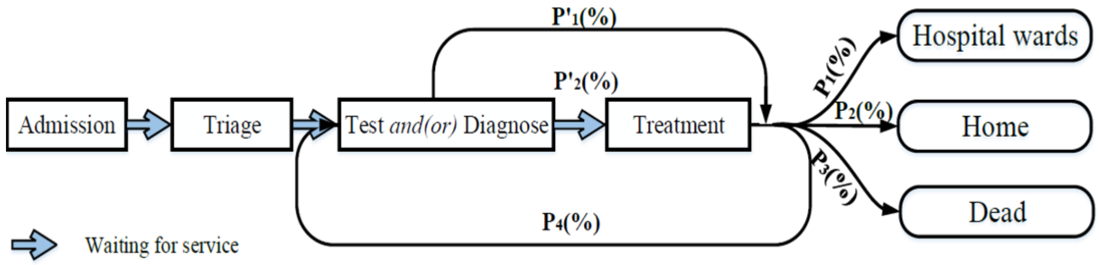


Fig. 4. Main processes in the emergency department

data of the specified emergency department. The uniform forms of the density functions are:

LOS is the patients' length of stay, and age is the patients' age, which also influences the probability of status transition. The level is the acuity level of the patient, and TOT is the type of test service or diagnosis by a doctor. The functions f and f' are the probability density function. These functions will be implemented by analyzing historical data in the tuning process. Therefore, combined with (2) - (5), every patient will show different behavior during the execution of the model because of the probability distribution and their differences in body condition. But the statistical property of agents will reflect their typical behavior.

In the case of active agents for medical staff, two different levels of experience are considered junior and senior. The less experienced user will need more time to complete the process than the most experienced. The simulator user can easily define the number of each type of personnel and their level of experience using the configuration console. The less experienced will use more time to carry out their work because they do not know; they could be a resident doctor who has just finished. The more experienced will take less time. They already know the process and treatment because they have much experience and years of service. To make a preliminary demonstration of how a simulation can be reproduced using only a few parameters, a simplified set of patient attributes and patient flow that is less complicated have been defined. The time of the doctors' attention changes according to each patient and its severity level.

V. Experiments and Discussion

A. Case Study: COVID-19 at the IPS Ingavi of Paraguay

The IPS Ingavi is a modern high-complexity hospital in Paraguay, offering medical care and emergency department to more than 2,000 insured persons per day, with approximately 1,500,000 insured persons. It is one of the reference hospitals for caring for patients with COVID-19 in the country.

From March 2020 to September 2021, IPS Ingavi Hospital treated approximately 15,000 COVID-19 patients, of whom 1,500 died despite medical efforts. In the most critical period of the disease, up to 10 daily deaths were recorded, and there were between 200 and 300 deaths per month. Table II shows the human resources configuration of the IPS hospital after hiring more personnel due to the high demand of patients due to COVID-19, and Table I shows the values of the human resources configuration of the IPS hospital before hiring more doctors and nurses.

TABLE I
Case 1: Low number of Human resources

Human resources configuration before the pandemic		
Label	***Interpretation***	***Number***
JA	Junior Admission staff	3
SA	Senior Admission staff	3
JTN	Junior Triage Nurse	5
STN	Senior Triage Nurse	5
JNA	**Junior Nurse area A**	**5**
SNA	**Senior Nurse area A**	**5**
JNB	**Junior Nurse area B**	**5**
SNB	**Junior Nurse area B**	**5**
JLE	Junior Outside Laboratory	3
SLI	Senior Internal Laboratory	3
JDA	**Junior Doctor area A**	**5**
SDA	**Senior Doctor area A**	**5**
JDB	**Junior Doctor area B**	**5**
SDB	**Senior Doctor area B**	**5**

[a]Human resources configuration before the pandemic

TABLE II
Case 2: High number of human resources

Human resources during the pandemic		
Label	***Interpretation***	***Number***
JA	Junior Admission staff	3
SA	Senior Admission staff	3
JTN	Junior Triage Nurse	5
STN	Senior Triage Nurse	5
JNA	**Junior Nurse area A**	**35**
SNA	**Senior Nurse area A**	**20**
JNB	**Junior Nurse area B**	**35**
SNB	**Senior Nurse area B**	**20**
JLE	Junior Outside Laboratory	3
SLI	Senior Internal Laboratory	3
JDA	**Junior Doctor area A**	**10**
SDA	**Junior Nurse area A**	**10**
JDB	**Junior Nurse area B**	**10**
SDB	**Senior Doctor area B**	**10**

[a]Human resources during the pandemic

The simulator developed with NetLogo [24] stores information about everything that happens during their execution and allows the creation of reports that can be exported and processed with statistics. We did an initial simulation with the data and got to analyze the simulators' behavior against the variables that influence the emergency department; several

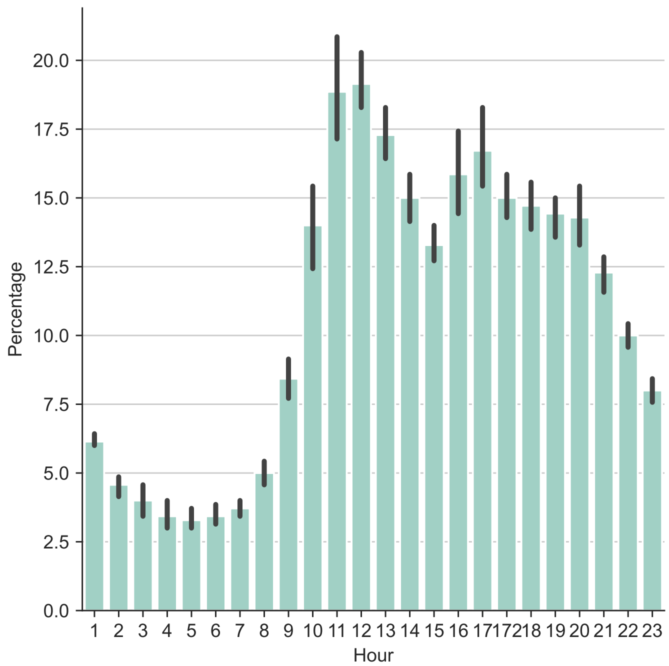


Fig. 5. Arrival of patients per hour at the hospital

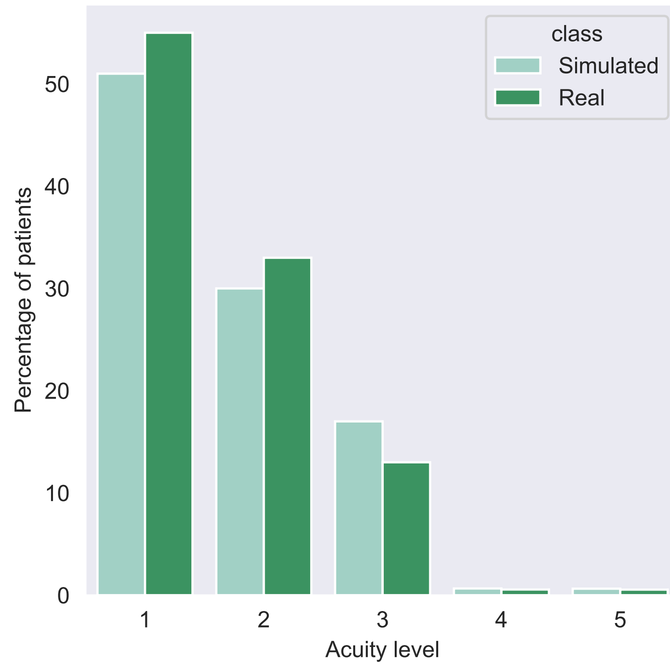


Fig. 7. Acuity level of patients between actual and simulated data

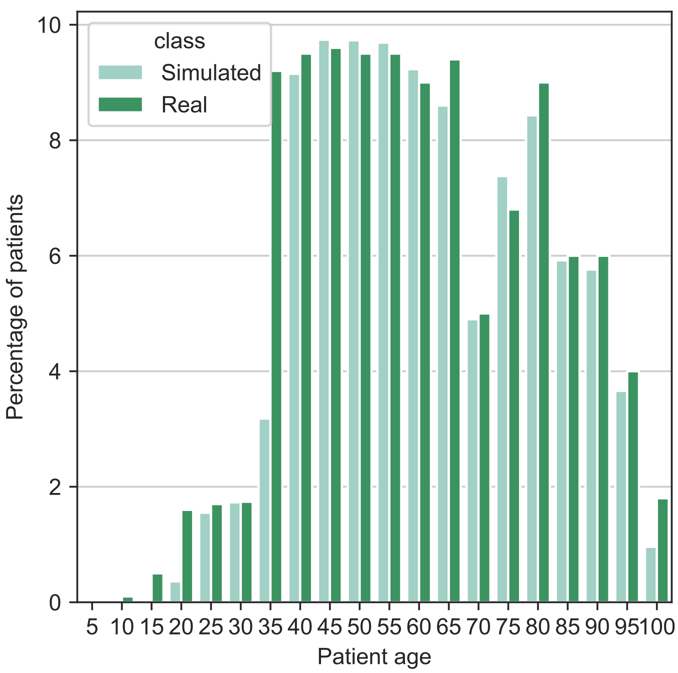


Fig. 6. Comparison of patient age between actual and simulated data

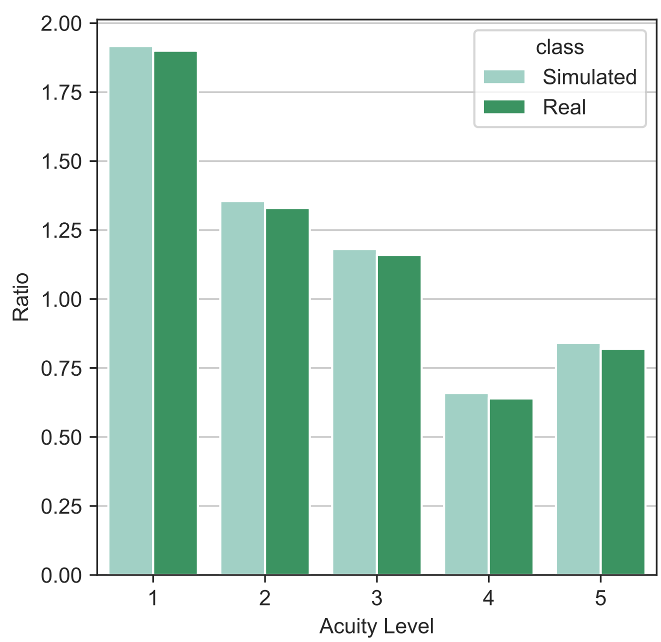


Fig. 8. Doctors' service time for each level of patients

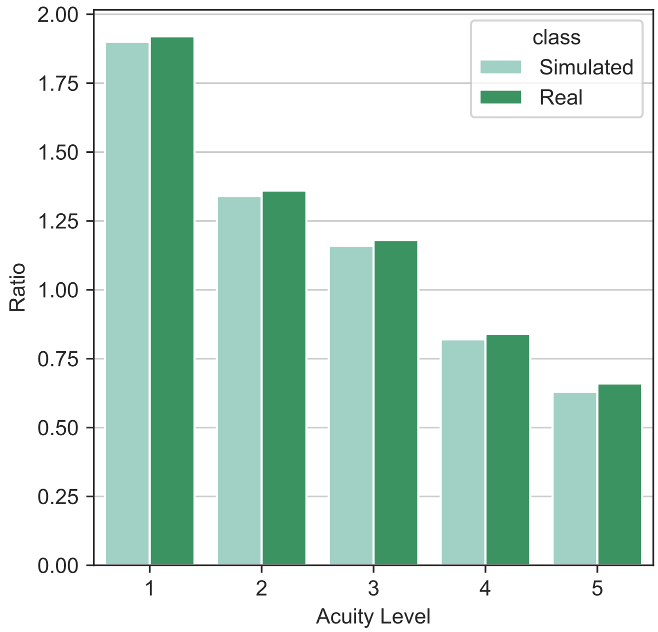


Fig. 9. Nurses' service time for each level of patients

simulations have been carried out with different values to observe what results we could get.

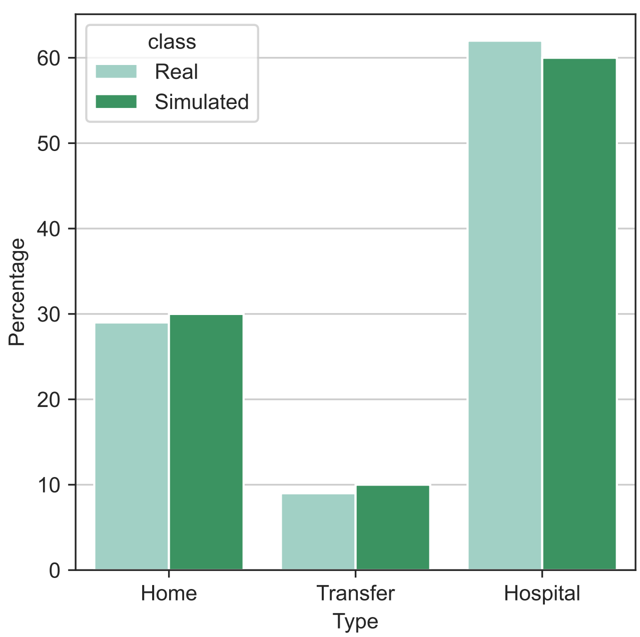


Fig. 10. Destination patient

Patient arrival is the emergency department simulators' input, directly influencing the systems' behavior. A precision model to reflect patient arrival is necessary to simulate and predict the behavior of an emergency department; the patient arrival model includes patients; Figure 5 shows the patient arrival rates distribution due to hours of the day in the hospital. This figure shows the arrival of patients at the hospital according to the time and day of the week, and it can be seen that the range of approximately 6 to 21 hours is the range where patients go to the hospital the most. Figure 6 shows the age range of the actual patients compared to the simulated ones, where you can see the age range of the patients who attended the hospital the most, which shows a rise from twenty-five to approximately ninety-five years.

TABLE III
UTILIZATION RESULTS ANALYSIS OF THE HUMAN RESOURCES (SIMULATION RESULTS).

Utilization results analysis of the human resources (percentage)				
Case	***Human resources***	***Average***	***Maximum***	***Confidence int***
Case 1	Doctor area A	70	100	(68, 70)
Case 2	Doctor area A	37	89	(37, 38)

[a]Utilization results analysis of the human resources

TABLE IV
MAXIMUM, THE AVERAGE SAMPLE SIZE TO ASSESS PATIENT STAY (LOS, LOWT, RELATIVE ERROR: 10%, CONFIDENCE: 95%).

Maximum, average sample size to assess patient stay (hours)				
Case	***Type***	***Average***	***Maximum***	***Confidence int***
Case 1	LOS	284	508	(282, 286)
Case 1	Lowt	282	433	(280, 284)
Case 2	LOS	256	497	(254, 258)
Case 2	Lowt	254	435	(252, 256)

[a]Maximum, average sample size to assess patient stay

The patient arrival model includes patients with severity L (acuity level). To do quantitative verification and validation, we built a patient arrival model according to the actual data from our cooperative hospital. As seen in Figure 7, comparing the simulated datas' severity levels to the actual data shows that levels 1, 2, and 3 are the levels of the patients who attended the hospital the most. The patients with severity levels 1, 2, and 3 are the most serious, and levels 4 and 5 are the mildest.

One of the simulator results is that the distribution of L (acuity level) among arrival patients was obtained through statistical analysis of the actual data and the overall patient age distribution.

The results from the patients' perspective are the LOS and the Lowt; the results from the management point of view are the occupations of the doctors, nurses, admission, and triage.

Patients with severity levels 1, 2, and 3 are the most serious and therefore take longer to be treated due to the seriousness of their situation. Figure 8 and Figure 9 show the comparison between the actual and simulated values of the ratio attention time of the doctors and nurses according to the severity level of the patients, where it is observed that the most extended attention times are between levels 1,2 and 3.

One of the outputs of the simulator is the destination of the patient after entering the emergency department; in Figure 10, you can see the comparison of the actual and simulated values of the percentage of patients who went home after the

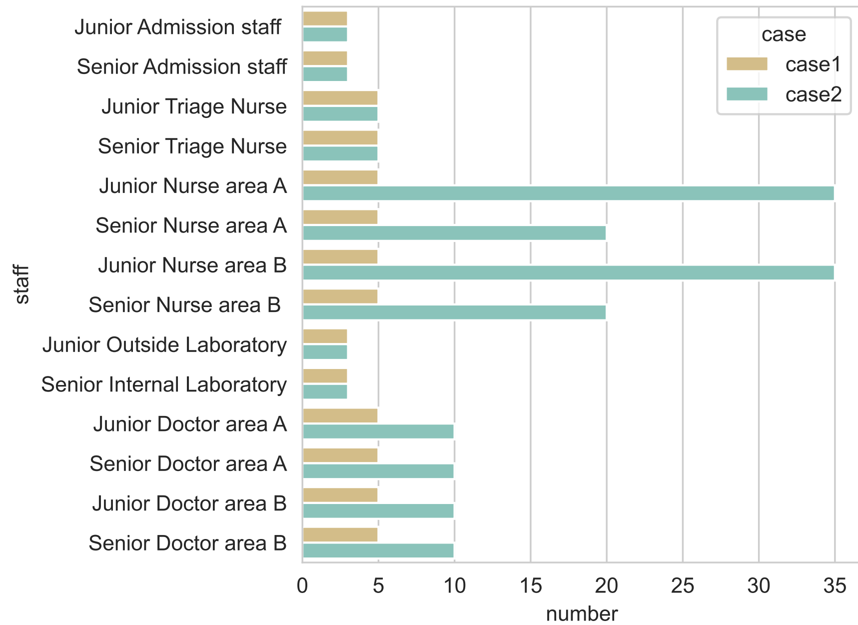


Fig. 11. Value of the human resources configuration case 1 and 2

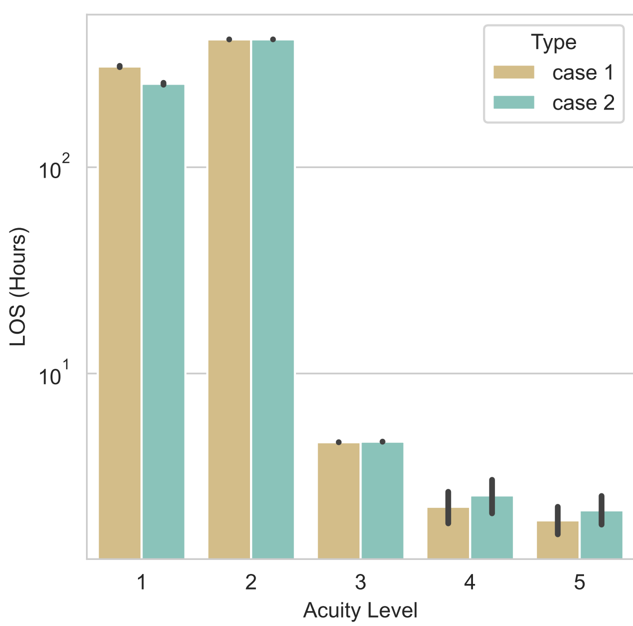


Fig. 12. Length of stay according to acuity level (LOS).

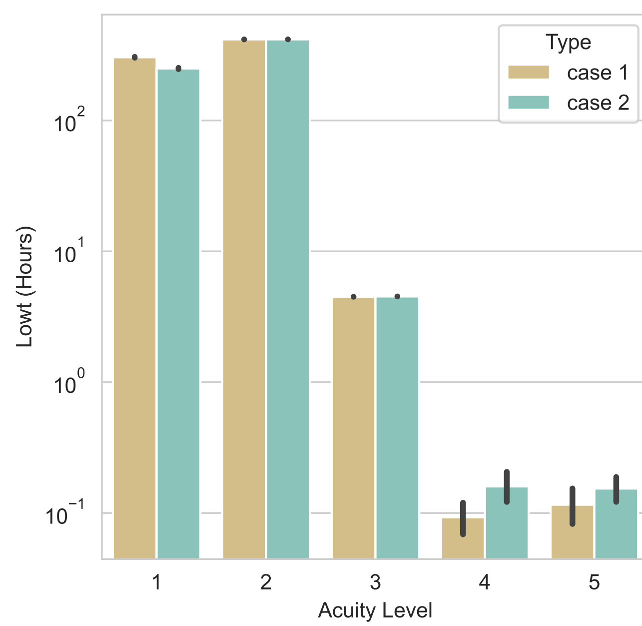


Fig. 13. Length of patient waiting time according to acuity level (Lowt).

emergency consultation, that is the case in which the patients were not seriously ill, the rate of patients who remained in the hospital to continue with their treatments and the percentage of those transferred.

Figure 11 shows the human resources configuration for cases 1 and 2 of the simulation; case 1 has a low number of doctors and nurses before hiring more hospital staff; case 2 has a high number of nurses and doctors after hiring more staff. Regarding the length of stay (LOS) and (Lowt) with the simulations carried out with cases 1 and 2, it can be observed that patients with severity levels 1 and 2 are those who have a longer waiting time to be treated in the hospital; this is seen in Figure 12 and Figure 13. This is because severity levels 1 and 2 are the most serious currently in the hospital; his critical condition requires further attention from hospital staff.

Regarding the percentage of occupation of doctors and staff,

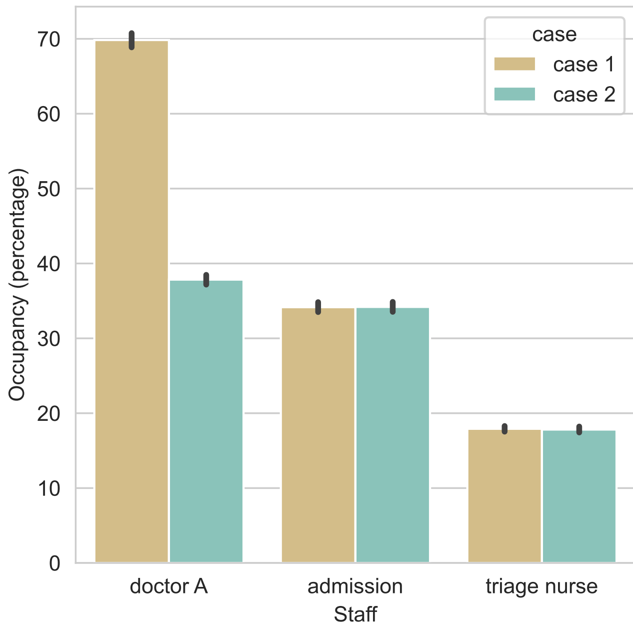


Fig. 14. Staff occupation time

the calculation is essential to determine when the system is about to be saturated, and it is necessary to hire more personnel; Also, when doctors are very saturated, their performance of patient care may decrease, and not care for patients correctly, in case they are already fatigued.

Figure 14 shows the occupation range of doctors, triage nurses and admission staff. It can be seen that the occupation range of the doctors for case 1 goes to approximately seventy percent; as the patients arrive, the occupation of the health personnel increases, and with an occupation range of about forty percent on average for case 2 by hiring more extra doctors and nurses due to the number of patients that are arriving due to the pandemic.

Table III shows the maximum value, the average (percentage), and the 95 percent confidence intervals of the doctors' area A for cases 1 and 2. When more health personnel are hired, the values of maximum and average occupations decrease. And when more health personnel are not engaged the occupations' average and maximum value increase.

Table IV shows the maximum value, the average, and the confidence intervals at 95 percent of the value of the LOS and Lowt for cases 1 and 2. It can be distinguished that when there is more health personnel hired, the values of the mean, maximum, and confidence interval of the LOS and Lowt (hours) waiting times of the patients are lower, and when there are fewer personnel, values of the means, maximum and confidence interval of the LOS and Lowt waiting times of the patients are higher.

VI. Conclusion

As a result of our research, we present an adaptation of an agent-based model for emergency department management. We handle different scenarios to adapt the simulator to pandemic situations, for example, using combinations such as hospital human resources and patient increases.

We used a set of synthetic input data for the input of data to the simulator for the arrival of COVID and NON-COVID patients with different ages, levels of severity, and the distribution of the arrival of patients by hours of the day, with these input data we have analyzed the occupation time of the doctors, the length of waiting time of the patients, the service time according to the level of severity and then we compare with reality. One of the advantages of our work is that this "COVID-19 in the emergency department" model/simulator allows emergency department managers to analyze and evaluate possible solutions to analyze the number of hospital staff, boxes/beds.

Our future work is to add the airborne spread of COVID-19 in the emergency department to validate and add more detail to the agent-based simulator to make it as consistent and close as possible in a pandemic situation and build different scenarios for decision-making. It will allow us to build virtual scenarios to understand the transmission phenomenon of COVID-19 and the potential impact of implementing different policies on viral spread rates.

Acknowledgment

This research has been supported by the Agencia Estatal de Investigacion (AEI), Spain and the Fondo Europeo de Desarrollo Regional (FEDER) UE, under contract PID2020-112496GB-I00 and partially funded by the Fundacion Escuelas Universitarias Gimbernat (EUG).

References

[1] R. E. Galeano Galeano, D. I. Rexachs del Rosario, A. Wong Gonzalez, E. Brugada, C. E. Villalba Cardozo, D. A. Galeano Galeano, and E. Luque Fadón, "An Agent-Based Model for the Management of the Emergency Department During the COVID-19 Pandemic," *The Fourteenth International Conference on Advances in System Simulation (SIMUL)*, pp. 54–62, Oct 2022.

[2] J. M. Adleberg, C. L. Catlett, R. E. Rothman, K. Lobner, and Y.-H. Hsieh, "Novel applications of agent-based modeling in emergency medicine research–a systematic literature review," *The American journal of emergency medicine*, vol. 35, no. 12, p. 1971, 2017.

[3] X. Zhang, "Application of discrete event simulation in health care: a systematic review," *BMC health services research*, vol. 18, no. 1, pp. 1–11, 2018.

[4] M. R. Davahli, W. Karwowski, and R. Taiar, "A system dynamics simulation applied to healthcare: A systematic review," *International Journal of Environmental Research and Public Health*, vol. 17, no. 16, p. 5741, 2020.

[5] M. Taboada, E. Cabrera, M. L. Iglesias, F. Epelde, and E. Luque, "An agent-based decision support system for hospitals emergency departments," *Procedia Computer Science*, vol. 4, pp. 1870–1879, 2011.

[6] E. Cabrera, E. Luque, M. Taboada, F. Epelde, and M. L. Iglesias, "Abms optimization for emergency departments," in *Proceedings of the 2012 winter simulation conference (WSC)*, pp. 1–12, IEEE, 2012.

[7] Z. Liu, D. Rexachs, E. Luque, F. Epelde, and E. Cabrera, "Simulating the micro-level behavior of emergency department for macro-level features prediction," in *2015 Winter Simulation Conference (WSC)*, pp. 171–182, IEEE, 2015.

[8] C. Jaramillo, D. Rexachs, F. Epelde, and E. Luque, "Virtual clinical trials: A tool for the study of transmission of nosocomial infections," *Procedia Computer Science*, vol. 108, pp. 109–118, 2017.

[9] C. Macal and M. North, "Introductory tutorial: Agent-based modeling and simulation," in *Proceedings of the winter simulation conference 2014*, pp. 6–20, IEEE, 2014.

[10] Z. Liu, E. Cabrera, M. Taboada, F. Epelde, D. Rexachs, and E. Luque, "Quantitative evaluation of decision effects in the management of emergency department problems," *Procedia Computer Science*, vol. 51, pp. 433–442, 2015.

[11] R. Moss et al., "Coronavirus disease model to inform transmission-reducing measures and health system preparedness, australia," *Emerging infectious diseases*, vol. 26, no. 12, p. 2844, 2020.

[12] M. Zeinalnezhad, A. G. Chofreh, F. A. Goni, J. J. Klemeš, and E. Sari, "Simulation and improvement of patients' workflow in heart clinics during covid-19 pandemic using timed coloured petri nets," *International journal of environmental research and public health*, vol. 17, no. 22, p. 8577, 2020.

[13] C. Currie et al., "How simulation modelling can help reduce the impact of covid-19," *Journal of Simulation*, vol. 14, no. 2, pp. 83–97, 2020.

[14] J. Le Lay, V. Augusto, X. Xie, E. Alfonso-Lizarazo, B. Bongue, T. Celarier, R. Gonthier, and M. Masmoudi, "Impact of covid-19 epidemics on bed requirements in a healthcare center using data-driven discrete-event simulation," in *2020 Winter Simulation Conference (WSC)*, pp. 771–781, IEEE, 2020.

[15] G. Fava, T. Giovannelli, M. Messedaglia, and M. Roma, "Effect of different patient peak arrivals on an emergency department via discrete event simulation: a case study," *Simulation*, p. 00375497211038756, 2021.

[16] V. Ahuja and L. V. Nair, "Artificial intelligence and technology in covid era: A narrative review," *Journal of Anaesthesiology, Clinical Pharmacology*, vol. 37, no. 1, p. 28, 2021.

[17] M. Laskowski, R. D. McLeod, M. R. Friesen, B. W. Podaima, and A. S. Alfa, "Models of emergency departments for reducing patient waiting times," *PloS one*, vol. 4, no. 7, p. e6127, 2009.

[18] Z. Chen, Z. Shu, X. Huang, K. Peng, and J. Pan, "Modelling analysis of COVID-19 transmission and the state of emergency in Japan," *International Journal of Environmental Research and Public Health*, vol. 18, no. 13, p. 6858, 2021.

[19] A. G. Uriarte, E. R. Zúñiga, M. U. Moris, and A. H. Ng, "How can decision makers be supported in the improvement of an emergency department? a simulation, optimization and data mining approach," *Operations Research for Health Care*, vol. 15, pp. 102–122, 2017.

[20] S. S. Jones and R. S. Evans, "An agent based simulation tool for scheduling emergency department physicians," in *AMIA Annual Symposium Proceedings*, vol. 2008, p. 338, American Medical Informatics Association, 2008.

[21] M. R. Friesen and R. D. McLeod, "A survey of agent-based modeling of hospital environments," *IEEE Access*, vol. 2, pp. 227–233, 2014.

[22] S. Diebel and E. Boissonneault, "A pan-canadian narrative review on the protocols for covid-19 and canadian emergency departments," *International Journal of Medical Students*, vol. 9, no. 2, pp. 157–161, 2021.

[23] Z. Liu, E. Cabrera, D. Rexachs, and E. Luque, "A generalized agent-based model to simulate emergency departments," in *Sixth Int Conf Adv Syst Simul*, pp. 65–70, 2014.

[24] U. Wilensky and W. Rand, *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. Mit Press, 2015.

[25] E. Bruballa, A. Wong, D. Rexachs, E. Luque, and F. Epelde, "Evaluation of response capacity to patient attention demand in an emergency department," *Int. J. Adv. Syst. Meas*, vol. 10, no. 1-2, pp. 11–22, 2017.