AI-POWERED OBSTACLE DETECTION FOR SAFER HUMAN-MACHINE COLLABORATION

Maros KRUPÁŠ, Mykyta KOT, Erik KAJÁTI, Iveta ZOLOTOVÁ
Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic, Tel. +421 55 602 2569
E-mail: maros.krupas@tuke.sk

ABSTRACT

This article deals with ensuring and increasing the safety of mobile robotic systems in human-machine collaboration. The goal of the research was to design and implement an artificial intelligence application that recognizes obstacles, including humans, and increases safety. The resulting mobile Android application uses a MiDaS model to generate a depth map of the environment from the drone's camera to approximate the distance from all obstacles to avoid the drone's collision. Besides, this work introduced us to DJI Mobile SDK and neural network optimizations for their use on smartphones.

Keywords: *human-machine collaboration, safety, monocular depth estimation, obstacle detection, mobile robots, MiDaS*

1. INTRODUCTION

Human-Machine Collaboration (HMC), which focuses on solutions where human workers and machines share their skills and tasks to improve productivity, still poses many challenges in industry applications, especially in becoming human-centric [1]. Safety is one of the primary challenges that must be tackled to ensure workers' physical and mental health [2].

Despite safety being one of the most researched areas in HMC, it still poses challenges with new systems being implemented, particularly with ground and aerial mobile robots, which are associated with greater risks mainly due to their mobility, which is usually their main advantage [3].

There are many ways and methods of ensuring safety in HMC. Vision-based safety systems have recently gained popularity due to their affordable price and ease of customization. In industry, vision-based safety systems are used for various tasks, usually calculating the distance between points and obstacles, collision avoidance, human intention recognition, visualizations and monitoring of safety zones [4]. One of the methods that has gained popularity recently is monocular depth estimation.

Depth estimation is a fundamental task in computer vision, crucial for various applications, including augmented reality, target tracking, and autonomous driving [5]. Monocular depth estimation involves deriving the distance of objects in the image from a single image, essentially creating a depth map. Monocular depth estimation uses only one camera to obtain an image or video sequence; thus, it doesn't require additional equipment but requires methods to regress depth in 3D space. Recently, researchers have proposed deep learning methods to handle traditional approaches [5].

This article presents our vision-based Artificial Intelligence (AI) safety system solution. The system uses monocular depth estimation to enhance safety and is controlled via a mobile application that operates the drone and its camera.

2. RELATED WORK

Safety concerns have been addressed since the introduction of industrial robots, where the initial task was to create physical separation from humans. With the introduction of mobile robots and cobots working close to humans, new challenges have emerged, particularly in collision avoidance [6]. However, collision avoidance alone is insufficient to ensure the psychological comfort of workers [7]. According to [2], safety protection mechanisms should be proactive, including environments that can sense and predict human behaviour to ensure safe human-machine interactions, especially in human-centric manufacturing. It was shown that the development and validation of guidelines for safety in human-machine collaborative assembly systems are also done using digital twins [8].

Based on the reviewed literature [9], many safety applications use AI methods paired with RGB-D cameras, such as deep learning, for object detection, collision avoidance, calculating distances between humans and robots, or motion detection to predict human actions. In one study [10], authors proposed using Visual Question Answering in HMC applications to improve effectiveness and safety, combining computer vision and natural language processing for a multimodal AI approach. For example, authors in [11] created a method to detect dangerous situations for humans in industrial settings with computer vision. In [12], the authors presented deep learning and fuzzy logic-based safety systems for collaborative ground mobile robots for scene understanding using instance segmentation in a warehouse environment.

Several monocular depth estimation applications have been proposed and used for mobile robot tasks. While safety is usually not the main topic of the articles, it is clear that the research and work can enhance safety since depth maps can be used to select collision-free waypoints to steer robots in a safe path [13]. For example, authors in [14] proposed autonomous navigation of mobile robots for object detection, and authors in [15] proposed a method for monocular visual odometry.

Various deep learning methods have been proposed in the literature, such as GAN [16] (Generative Adversarial Network) or CNN [12] (Convolutional Neural Network). MiDaS deep neural network model for monocular depth estimation, which is also used in this work, was also used in other works [17, 18] for autonomous navigation and mapping of indoor environments using a drone equipped with a monocular camera.

3. SOLUTION DESIGN

This paper aimed to design and implement an AI application to increase safety using DJI drones. For this, Google Pixel 7 Pro and DJI Mobile SDK version 5.8 were chosen for implementation. To test our solution, we used the DJI Mini 3 drone, which is suitable for its low price, small size, and low weight. Mini 3 uses several sensors to ensure safety, including a camera, visual positioning system, inertial measurement unit, barometer and positional sensors such as GPS, Galileo and BeiDou, which couldn't be used due to testing in indoor laboratory conditions only. As seen in the Fig. 1, the final architecture for this application also included DJI RC-N1 Remote Controller to control the drone movement and secure connection between the drone and mobile application.

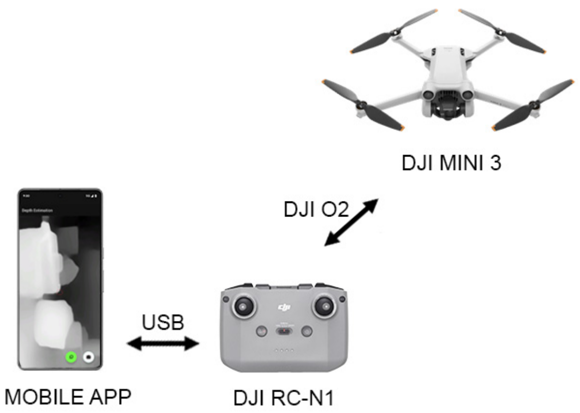

Fig. 1 Solution architecture

For the AI part, we used MiDaS (Multi-scale Differentiable Architecture for Depth Estimation) 2.1 v21 small 256, a machine learning CNN model developed by Intel Labs specifically designed for monocular depth estimation. MiDaS is particularly useful for our solution, as it allows drones to react to all obstacles. Our solution used the drone's camera to obtain image data. MiDaS model was then used on our mobile phone in the Android application, which was also responsible for obtaining video stream data from the drone's camera and all other drone functionalities.

4. IMPLEMENTATION

In this section, the final functions of the applications are described. The application user interface (UI) consists of the camera feed (see Fig. 2a) and buttons to control the drone's functionalities, such as turning on MiDaS monocular depth estimation, as seen in Fig. 2b. The model's output is depth data, represented by a 2D image array, representing the relative distance to all objects in the environment in the image. Values of the array range from 0 to 255, where black on the depth map represents 0 and the furthest objects from the drone, while white represents 255 and the nearest objects to the drone.

When the value limit for an obstacle is reached, a warning is shown on the screen. Furthermore, the image output from the monocular depth estimation is divided into 9 zones in a 3x3 grid to further specify the exact location of the detected obstacles. Based on the grid, if the obstacle is detected on the side of the drone, the drone is moved away to the opposite side, and when the obstacle is detected in the middle zone, the drone is stopped. When the drone stops in front of the obstacle, the user cannot control the drone with the controller for a few seconds.

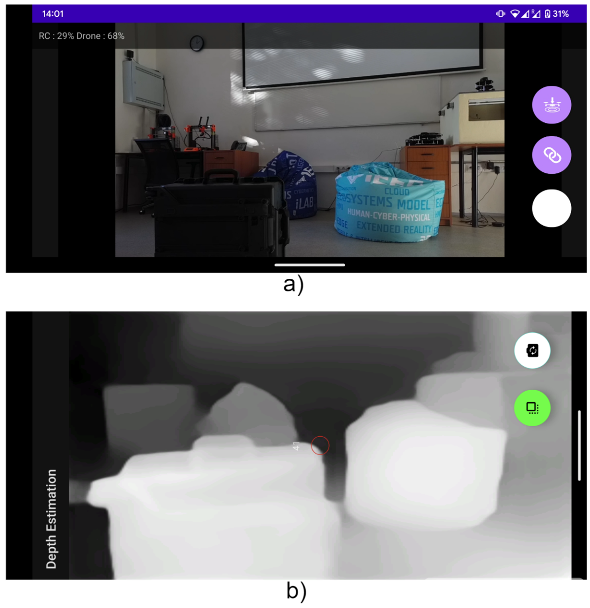

Fig. 2 Application UI: a) drone's camera video stream, b) MiDaS real-time depth estimation

5. RESULTS

After the application implementation, experiments evaluating our solution were done for static and moving obstacles. Since the depth data represent relative distance only, we measured all distances with a meter. In this case, three metrics were evaluated:

- detection distance (m) - the distance from which the drone detects an obstacle.
- speed (m/s) - speed of the drone at the moment when the obstacle is detected.
- obstacle distance (m) - the distance between the drone and the obstacle after the drone completely stops.

The average results for static objects, depending on the drone speed, are shown in Table 1. Based on our experiments, we can also conclude that the complexity of the patterns or images on these objects minimally affects the distance at which the drone detects them. This suggests that our detection system is robust to variations in object appearance and focuses on other factors, such as object size and contrast with the background, for successful detection.

Table 1 Results for static objects

speed (m/s)	detection distance (m)	obstacle distance (m)
0,2	0,5	0,4
0,4	0,4	0,2
0,5	0,3	0,15
0,6	0,28	0,10

In the next phase, we examined the system's ability to stop before colliding with moving objects. The average results for different speeds of the drone before obstacles suddenly appeared in the drone's camera view are shown in Table 2. Based on the results, we can conclude that the maximum safe speed of the drone to avoid collisions is 0,6 m/s for static objects and 0,35 m/s for moving objects. It was shown that as the drone's speed increases, the detection distance and obstacle distance decrease. Comparing the results between detecting the static and moving obstacles highlights the limitations of the reaction time of our application.

Table 2 Results for suddenly appearing objects

speed (m/s)	detection distance (m)	obstacle distance (m)
0,2	0,2	0,14
0,3	0,2	0,10
0,35	0,2	0,05

Additionally, to evaluate reaction time, different metrics regarding application processing time are evaluated on three different phones with various calculating powers to evaluate the relation between the power of the phone and the effectivity of our application. The first phone was our original Google Pixel 7 Pro, the second phone was the Realme X7, and the third was the Samsung S23 Ultra, presenting the phone with the best power. All phones used GPU to calculate the depth map. These metrics were:

- Conversion from raw to bitmap (ms) - Time of the conversion from raw video source from the drone's camera to the bitmap format suited for the neural network.
- MiDaS processing (ms) - Time it took for the MiDaS model to calculate the depth map for the input bitmap image.
- Decision making (ms) - Time for application to detect obstacle based on the threshold value.

In Fig. 3, we can see that conversion from raw to bitmap took the least amount of time from the three metrics. There was also an additional delay not included in the figure in communication between a drone and a mobile phone, which was 120ms. This delay represented the time it took for the phone to get the new image from the drone.

Thus, mobile phones' processing capability directly influences the efficiency and responsiveness of the obstacle detection system, which is critical for ensuring the safety and performance of the drone in real-time applications.

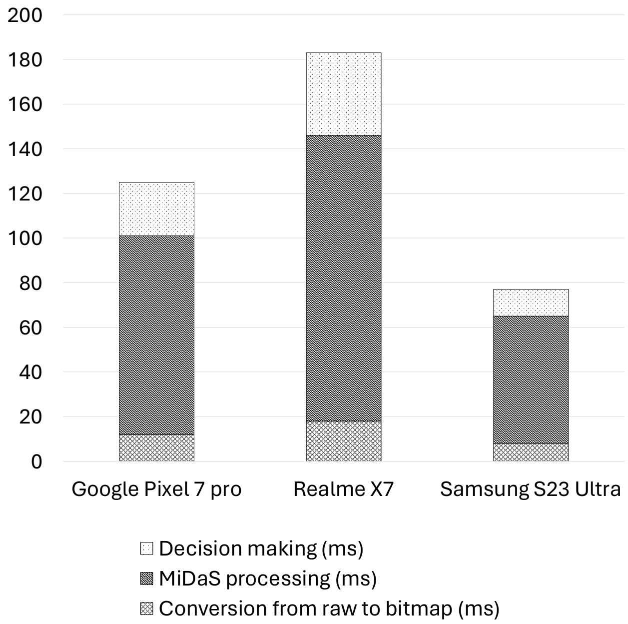

Fig. 3 Processing times of testing mobile phones in ms

6. CONCLUSION

Our experimental findings highlight the reliability of our application in detecting obstacles and preventing collisions with humans or objects. The results also indicate our application's responsiveness to sudden obstacles, although this ability is limited to scenarios where the drone's speed is lower. This indicates that the application's reaction time may not be fast enough to handle sudden obstacles at higher speeds.

Overall, our results indicate that future work should focus on optimizing other aspects of the system, such as response time and the ability to estimate distances accurately, rather than fine-tuning the ability to distinguish between different patterns and images. Furthermore, future work can also focus on distinguishing between different objects and humans. Digital twin-enhanced validation of our solution could also be implemented

A significant strength of our solution is its potential to enhance safety in various environments significantly. As drones are increasingly used in complex environments such as urban areas, industrial sites, and delivery services, it is essential for drones to accurately detect and avoid obstacles to prevent accidents and ensuring the safety of both workers and machinery is important. However, the system faces challenges in more dynamic settings that must be addressed to enhance overall safety. Integrating advanced AI techniques for motion detection and behaviour prediction further contributes to creating safer, more responsive systems. By continuing to advance these technologies, we can develop solutions that meet current safety standards and set new benchmarks for reliability and effectiveness.

ACKNOWLEDGEMENT

This publication is the result of the research project funded by KEGA grant IVI: Internet of Vehicles in the context of Industry 4.0 (KEGA 068TUKE-4/2024).

REFERENCES

[1] NAHAVANDI, S.: *Industry 5.0-a human-centric solution*, Sustainability **11**, No. 16 (2019) 4371

[2] LU, Y., et. al.: *Outlook on human-centric manufacturing towards Industry 5.0*, Journal of Manufacturing Systems **62**, No. (2022) 612-627

[3] KRUPAS, M. – CHAND, S., et. al.: *Human-Centric UAV-UGV Collaboration*, 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE) , No. (2023) 1-6

[4] HALME, R. J., et. al.: *Review of vision-based safety systems for human-robot collaboration*, 51st CIRP Conference on Manufacturing Systems **72**, No. (2018) 111-116

[5] MING, Y., et. al.: *Deep learning for monocular depth estimation: A review*, Neurocomputing **438**, No. (2021) 14-33

[6] SIMOES, A. C., et. al.: *Designing human-robot Collaboration (HRC) workspaces in industrial settings: A systematic literature review*, Journal of Manufacturing Systems **62**, No. (2022) 28-44

[7] GOMEZ, S. R., et. al.: *Working Together: A Review on Safe Human-Robot Collaboration in Industrial Environments*, Ieee Access **5**, No. (2017) 26754-26773

[8] GUALTIERI, L., et. al.: *Development and validation of guidelines for safety in human-robot collaborative assembly systems*, Computers & Industrial Engineering **163**, No. (2022)

[9] KRUPAS, M. – LIU, CH., et. al.: *Towards Human-Centric Digital Twin for Human-Machine Collaboration: A Review on Enabling Technologies and Methods*, Sensors **24**, No. 7 (2024) 1-20

[10] WANG, T., et. al.: *Digital twin improved via visual question answering for vision-language interactive mode in human–machine collaboration*, Journal of Manufacturing Systems **58**, No. (2021) 261-269

[11] ZUBAL, M., et. al.: *IoT gateway and industrial safety with computer vision*, 14th International Symposium on Applied Machine Intelligence and Informatics , No. (2016)

[12] HATA, A., et. al.: *AI-based Safety Analysis for Collaborative Mobile Robots*, 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA) , No. (2019)

[13] DONG, S., et. al.: *Towards Real-Time Monocular Depth Estimation for Robotics: A Survey*, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, **23**, No. 10 (2022) 16940-16961

[14] LIU, Y., et. al.: *Ground-Aware Monocular 3D Object Detection for Autonomous Driving*, IEEE ROBOTICS AND AUTOMATION LETTERS **6**, No. 2 (2021) 919-926

[15] BAN, X., et. al.: *Monocular Visual Odometry Based on Depth and Optical Flow Using Deep Learning*, IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT **70**, No. (2021)

[16] ALEOTII, F., et. al.: *Generative Adversarial Networks for Unsupervised Monocular Depth Prediction*, Computer Vision – ECCV 2018 Workshops , No. (2019) 337-354

[17] HENSEL, S., et. al.: *Monocular Depth Estimation for Autonomous UAV Navigation Based on Deep Learning*, 2023 XXXII International Scientific Conference Electronics (ET) , No. (2023)

[18] KUMAR, R., et. al.: *Autonomous Drone Navigation using Monocular Camera and Light Weight Embedded System*, 2023 International Conference for Advancement in Technology (ICONAT) , No. (2023)

Received June 18, 2024, accepted July 23, 2024

BIOGRAPHIES

Maroš Krupáš is a PhD student at the Department of Cybernetics and Artificial Intelligence, Technical University of Kosice. He received his Bachelor's degree in 2020 and Master's degree in 2022. His Bachelor's thesis deals with the cooperation of ground mobile robots in map exploration. His master thesis deals with the cooperation of ground and aerial robots in mapping and navigation. His current research deals with digital twin and HMC in Industry 5.0.

Mykyta Kot is a bachelor student at the Department of Cybernetics and Artificial Intelligence, Technical University of Kosice. His Bachelor's thesis deals with the safety of humans collaborating with flying robots.

Erik Kajáti leads the Centre of Applied Cybernetics at the Department of Cybernetics and Artificial Intelligence (DCAI), FEEI, TUKE. His research delves into Industry 4.0/5.0, Human Cyber-Physical Systems (H-CPS), the Internet of Things (IoT), Edge-Enabled Computing, and the energy sector, particularly resilient and sustainable energy. His primary research focus is on the Edge-Enabled Approach for Intelligent Human-System Interoperability. Additionally, he co-founded the startup CHECkuP. Erik is also an active IEEE and Slovak Society for Artificial Intelligence member.

Iveta Zolotová is a professor and a leader of the research group Intelligent Cybernetic Systems at the DCAI FEEI TUKE. Her expertise are: intelligent systems, human-system symbiosis and interaction, digital twin, Industry

4.0/5.0, industrial informatics, applied artificial intelligence, edge AI, industrial IoT, smart health and transportation, computer vision, robotics. She leads many research and innovation projects, and the labs: Think lab: Vision, IoT and Robotics; Intelligent Health Lab, and Smart Industry/Industry 4.0 lab. She is a member of IFIP and a senior member of IEEE.