Oculaone: Exploring Drones-based Assistive Technologies for the Visually Impaired

Suman Raj, Swapnil Padhi and Yogesh Simmhan
sumanraj@iisc.ac.in,swapnil18800@gmail.com,simmhan@iisc.ac.in
Indian Institute of Science (IISc)
Bengaluru, Karnataka, India

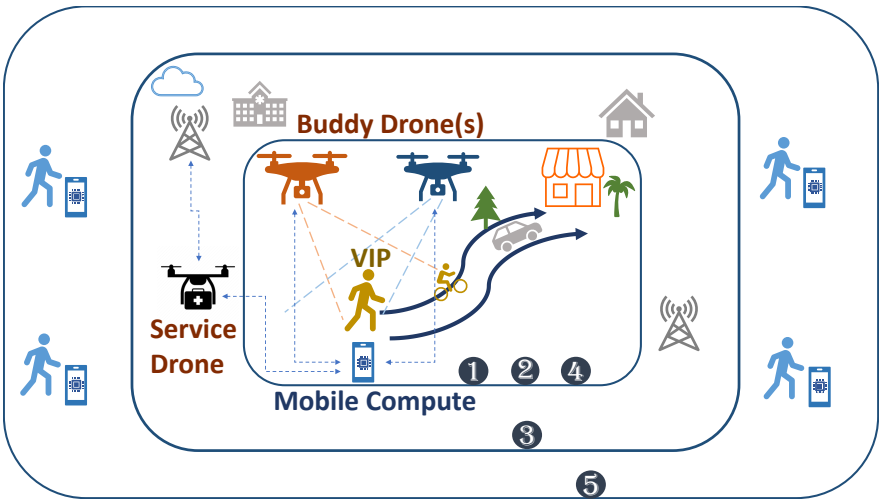


Figure 1: Proposed Design of Oculaone

ABSTRACT

Fleets of Unmanned Aerial Vehicles (UAVs), also called drones, are becoming commonplace to support diverse applications in logistics and urban safety. These rely on advances in computer vision and Artificial Intelligence (AI) for autonomous navigation and control, facilitated by onboard sensors and accelerated edge computing devices attached to their base stations. This article examines how drones can be used for social good to enhance the lifestyle of Visually-Impaired People (VIP) through navigational assistance and situation awareness, enabled through visual analytics over drone video streams. We propose *Oculaone*, a platform for coordinating and managing a heterogeneous UAV fleet that offers drones as buddies to guide users. It uses onboard sensors and edge accelerators, and two-way communication using gestures and audio prompts, to enhance the safety and autonomy of the VIP. We validate our early prototype using Tello nano-drones and Jetson Nano edge accelerators, and present preliminary results for several safety scenarios.

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585863

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; **Collaborative and social computing**; • **Computing methodologies** → **Distributed computing methodologies**; *Cooperation and coordination*; *Vision for robotics.*

KEYWORDS

Assistive technologies, Visually Impaired, Unmanned Aerial Vehicles, Drones, Computer Vision, Artificial Intelligence, Edge computing, Cloud computing, Distributed coordination

ACM Reference Format:
Suman Raj, Swapnil Padhi and Yogesh Simmhan. 2023. Oculaone: Exploring Drones-based Assistive Technologies for the Visually Impaired . In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3544549.3585863

1 INTRODUCTION

Context. Over 200 million people suffer from moderate or severe visual impairment worldwide, of whom 38 million are blind [36]. Vision impairment severely impacts the quality of life among adults resulting in poorer workforce participation and greater rates of depression. Among older adults, blindness can lead to social isolation, difficulty in mobility, and a higher risk of falls [34]. Assistive technologies can empower Visually Impaired People (VIP), e.g., enabling them to safely take a morning walk or buy groceries from a local store, thus promoting autonomy and social acceptance.

Opportunity. Advances in video and LIDAR sensing, computer vision and Deep Neural Network (DNN) models, capable edge accelerators, and control systems have led to self-driving cars plying on city streets and highways. However, for developing countries where city traffic is intractable, the promise of such autonomous vehicles in Unmanned Aerial Vehicles (UAVs), or drones, is much more feasible. UAVs are increasingly becoming a flexible mobility and observation platform. Progressive regulations by the US FAA [12], Europe's EASA [1], and India's DGCA [27] are bringing together industry and researchers to examine disruptive use cases. Drones are also more affordable and come in a variety of form factors, starting from nano drone quad-copters that weigh under 250 g, are about 200 mm in size, and yet have onboard cameras and sensors. Fleets of drones are being put to use in large cities for logistics and delivery [29], urban safety [33], infrastructure monitoring [18], and even to help the visually challenged [5].

Edge accelerators such as Nvidia Jetson are enabling rapid inferencing of DNN models and computer vision algorithms through low-end GPU modules integrated with ARM-based processors. They are also compact and power efficient. E.g., the Jetson Nano [16] used in our experiments is the size and weight of a pack of cards, consumes < 10W of power, and can fit into a purse or a backpack. Despite their small size, they can connect to the drone through communication modules and achieve real-time inferencing over video feeds. This enables a portable solution for closed-loop decision-making to operate one or more drones autonomously, driven by DNN models and control algorithms running on the edge.

Proposal. We propose a drone-based assistive platform, *Ocularone*, to enable VIPs to live an active yet safe outdoor lifestyle. We use one or more "buddy drones" with onboard cameras coupled with edge accelerators as mobile valets to help with outdoor safety and navigation in the vicinity of the VIP and to take commands using gestures. This is coupled with "service drones" that can be summoned to perform more complex tasks like fetching first-aid kits. Global path planning and coordination among the drones, along with more complex visual analytics, is managed from the cloud. Apps can be deployed to add personalized functionality.

Gaps and Challenges. Among contemporary technologies for navigation assistance of VIPs, smart canes and smart wearables offer sensor and video-based guidance but suffer from a restricted range and Field of View (FoV) that limit the detection of hazards. Recently, drone-based solutions have also been explored in managed environments where the drone is manually navigated to track a Bluetooth bracelet or the VIP follows the rotor sound of the drone. These have limited autonomous behavior and do not work in an urban environment. Our proposed solution for outdoor environments use drones to provide a comprehensive view of the VIP's surroundings, operate independently and interact with the VIP to guide them through a collision-free path using real-time visual analytics.

Achieving this, however, poses several challenges, some of which we address in this paper:

- Existing off-the-shelf DNN models to analyze the drone video feeds need tuning and post-processing before they can be used within Ocularone. In this paper, we retrained object detection models on our curated dataset to accurately detect a hazard vest as a unique identification of the VIP. Further, we designed and implemented a classifier model as a post-processing step to identify gestures made by the VIP.
- We need the drone to autonomously follow the VIP while adapting to the speed of the VIP. This requires us to tune the control parameters of the buddy drones to achieve a smooth flight with minimum jerks.

Contributions. In this article, we describe the high-level design goals and highlight the technical challenges of the proposed *Ocularone* platform (§ 3). We then describe an early prototype that addressed a subset of these challenges (§ 4). Our implementation uses customized object detection and body pose estimation DNN models to follow a VIP in an outdoor environment and detect any safety hazard or gestures. We conduct preliminary experiments in a controlled campus environment using a (proxy) VIP to evaluate this prototype (§ 5). These demonstrate the ability of the drone to autonomously follow the VIP, react to hand gestures, and detect and take action if they fall.

These early results are promising and motivate further exploration of the *Ocularone* architecture. Future work (§ 6) will address navigation and obstacle avoidance for the VIP, robustness in realistic environments, and coordination among multiple drones.

2 RELATED WORK

Guiding the visually impaired in familiar and new places helps enhance the VIP's confidence and efficiency when traveling [41]. Bandukda, et al. [9] highlight the experience of VIPs during mobility and orientation training, where they use non-visual clues for sensing the surroundings and navigation. However, the non-visual cues can be misleading in multiple scenarios as they are not tightly coupled with the locations. Nagraj, et al. [30] investigate the navigational habits of the blind in developing countries like India where the VIPs prefer interacting with other people for assistance when navigating in public spaces. Our proposed Ocularone design using "buddy drones" supports this factual experience. It mimics human assistance by sensing and sharing details of the visual surroundings and offers personalized cues for independent navigation. This can reduce their reliance on human support, which may not always be forthcoming or may lower their self-confidence.

Besides traditional means of assistance using white canes and guide dogs, there has been an emergence of sensory signals in human-computer interaction, including wearable devices, which use vibrotactile sensors for indoor navigation [19]. Drishti [38] is a wearable device that uses audio cues to provide an optimal route in indoor and outdoor environments. However, it weighs around 4 kgs, and can curtail the mobility of VIPs. iMerciv's BuzzClip [20] complements the cane and uses ultrasound to detect obstacles and notifies the users through discrete vibrations. But, ultrasound sensors are susceptible to noise, which makes them impractical in busy outdoor environments with multiple static and dynamic obstacles. Similarly, sensor-based walking assists [23] such as WeWalk [40] and GesturePod [35] include gesture-based communication but have a limited range in outdoor environments that restricts their ability to recognize safety hazards [26]. In addition to alerts on proximate obstacles, it is also important to offer guidance on upcoming path changes and hazards in real time. Drones are useful

in this regard. They have more flexibility in the sensors (including cameras) and actuators they carry, and offer a 360° view coupled with mobility for more holistic situation awareness.

With the increasing capability of machine learning, camera-based systems have emerged for visual perception to assist VIPs. An AI-powered backpack has been proposed [11], which analyzes the environment using computer vision over camera feeds and warns users about threats using a voice assistant. Smart guiding glasses [8] use depth cameras along with ultrasonic sensors to alert users using vibrations and voice. Blavigator [15] presents a wearable system built on top of SmartVision [17] to detect and avoid obstacles. A major limitation of wearables is their limited field of view. They are dependent on the direction in which the VIP faces and may be inadequate to detect peripheral hazards. Drones being highly mobile, can offer a variety of perspectives independently, and multiple buddy drones can coordinate to cover all blind spots.

A number of recent apps like Lookout [2] from Google India, Seeing AI [14], VStroll [22] by Microsoft, and Nearby Explorer [4] provide visual information service to the VIPs by describing their surroundings. These engage and promote walking by VIPs, but do not provide vision-based active navigational assistance like we propose. Liu et al. [24] has identified features essential to the VIPs as part of a user retention study of Microsoft Soundscape [21], an audio-based navigation. These help us understand and develop specific features which will help VIPs use our framework.

Using drones for navigation assistance of VIPs is gaining traction. DroneNavigator [5] explores the use of drones for guiding the visually impaired by maintaining the drone at a distance of $\approx$ 1–2 m in front of the participant, as measured using the Bluetooth signal strength between the drone and a bracelet that the VIP carries. However, the drone itself was manually being flown by an experimenter, and was not autonomous. In a subsequent work [6], they use a quadcopter connected using a physical leash to provide a tactile feedback to the VIP. Here again, the navigation instructions were triggered manually. We instead propose for the drones to autonomously navigate themselves and also guide the VIP using vision-based techniques. Bluetooth-based accessories can be added to our design in future for tactile or audio feedback.

Zayer, et al. [3] adapt drones to guide blind runners, where the runners follow the drone using its rotor sound. This is not reliable in a busy urban environment where the rotor sound can be lost. Further, Nasralla, et al. [32] study the prospects of using deep learning and computer vision-enabled UAVs for assisting VIPs in a smart city. This motivates us to propose and explore DNN-based visual navigation of VIPs using drones. However, using drones in proximity with humans can be challenging. Recent studies [7] have explored the use of service robots to guide VIPs in a building. The study suggests that the robot should respect the VIP's autonomy and independence by allowing them to have the ultimate control of the interactions. We take this study into consideration and incorporate features for an intuitive VIP-drone interface.

Our proposed Ocularone solution uses drones with diverse roles, together with computer vision and deep learning capabilities, to coordinate among themselves. The 360° perspective offered by drones overcome the limitations of other technologies. This can assist VIPs with navigational tasks, situation awareness and safety alerts, and thus improve their independence. At the same time, drone-based

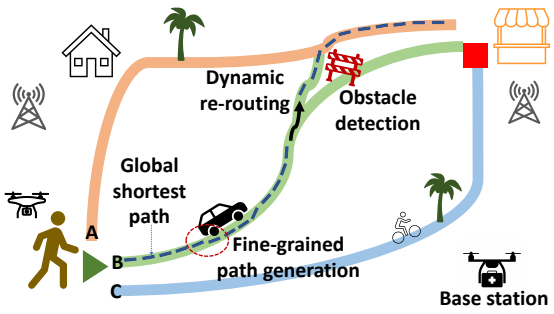


Figure 2: Path Planning

solutions also have limitations such as being sensitive to wind and rain, potentially hitting obstacles and failing, or the vision based algorithms being inaccurate in challenging lighting conditions. So such a solution can complement more basic technologies such as smart canes, smart glasses or Bluetooth tethering.

3 DESIGN GOALS AND REQUIREMENTS

We wish to support an active and safe lifestyle for a visually challenged person within an urban environment using the Ocularone platform, as shown in Fig. 1. The platform uses a fleet of heterogeneous drones consisting of *buddy drones* and *service drones* to support one or more VIPs. A buddy drone is assigned to provide navigational assistance and situation awareness to a single VIP, while service drones equipped with more advanced sensors/accessories are shared across VIPs to perform specific tasks on-demand.

Specifically, the platform should support these **design goals**.

(1) *Plan the path from the current location of the VIP to a destination they wish to go to.* This requires path planning from a starting point (Fig. 2, green triangle) to a destination (red square) at a coarse level, e.g., using a digital city map, similar to GPS-based guidance. This is complemented by planning at fine scales in the immediate vicinity of the VIP as they walk along that path. This first requires orienting the drone with respect to the VIP, following them as they move, prompting them to stop or take turns at relevant locations, and dynamic re-routing of the path, say, in case of road closures. E.g., in Fig. 2, once the closed road along the Path B (green) is detected, the mobile device requests our cloud service for an alternate route from the current location and, switches to Path A (orange) for the rest of the walk. These need to be supported using a mix of GPS and computer vision-based guidance.

(2) *Obstacle avoidance for the VIP and the drone.* This includes detecting static obstacles in the proximity and direction of walk of the VIP, such as potholes or trees. We must also identify dynamic obstacles, such as incoming traffic, or safety situations, such as an unmasked crowd during COVID. Based on these, we need to give cues to help the VIP navigate around such proximate hazards. The drone should also maintain its own safety by avoiding collisions with obstacles such as street lights or tree branches, while keeping the VIP within its FoV.

(3) *Detect and alert if the VIP has a health or safety incident.* The drone must detect if the VIP encounters a safety incident or suffers from a medical emergency, e.g., being injured or having a fall, and respond by taking specific actions, such as notifying emergency contacts with sufficient context, informing first responders, and/or requesting first aid through a service drone. This may require a mix of visual analytics complemented by smartphone or wearable based sensing using accelerometers.
(4) *Provide an intuitive interface with the VIP for bi-directional communication.* A natural, discrete and robust means of communication, from the Ocularone system to the VIP to provide cues and alerts, and from the VIP to the system to give predefined commands and respond to choices, is required. This may include audio cues using a bluetooth headset that communicates using smartphones and wearables for the former, and audio commands and hand gestures for the latter.
(5) *Scale to a city-wide scale with 100s of VIPs and shared drones operating collaboratively.* This allows buddy and service drones to be shared among users as part of a larger platform, and have vision-based help with visual sensing, analytics or service. It helps the system scale with fewer drones and reduced idle time, enhancing the resilience and affordability.

Through these goals, Ocularone can support several real-world **use cases**. In an urban environment, the drone can help plan and guide the VIP along a pre-set path for a walk while navigating them around local hazards. In a managed environment like a university campus, the VIP may take a stroll without a pre-defined destination, and the drone may follow the person while describing points of interest and looking for threats. If the VIP runs in a marathon, they may request the drone for water or an energy bar through a gesture, and this may request a remote drone with a grasping arm to fetch the requested supplies from the nearest pick-up location to the VIP's current location. Lastly, if the VIP is injured, the drone may alert their emergency contact along with a video recording of the VIP's current condition and also summon a first-aid service drone to assist. While our solution is aimed at VIPs, the system once developed can be adapted to other human and social-centric applications such as shepherding kids at school and assisting runners during a marathon.

Achieving these design goals requires us to solve several **technical challenges** in Ocularone.

- The buddy drone should uniquely identify and follow the VIP autonomously at a specific distance and height in order, and adjust its trajectory to maintain a comprehensive view of the VIP and their surroundings. This should be done based on visual sensing
- The framework should generate a global shortest path from the current location to the destination the VIP wishes to reach using a city's map information, possibly optimized for accessibility considerations. This can be complemented with local visual details collected while the VIP is taking the path and fed into a global model of the world to help with future plans.
- While the VIP is on a path, the buddy drone should detect obstacles ahead of the VIP based on computer vision algorithms, and identify a trajectory to help them avoid the obstacle based on visual sensing. At the same time, the drone itself should be able to plan a trajectory that avoids obstacles in its own path while still keeping the VIP in its FoV.
- The drone should be able to use video analytics to have full situation awareness with a 360° perspective. This should be used to detect threats in real-time and any health emergencies the VIP is undergoing, based on video analytics.
- The drone should be able to interpret audio and visual cues from the VIP to take commands under diverse urban conditions.
- Service drones should be able to operate autonomously using onboard sensors, local computing and cellular connectivity to the cloud to exchange control signals and transfer data for analysis.
- Multiple drones should be able to collaborate among each other. This includes two or more buddy drones assisting a single VIP by maximizing the FoV, service drones helping different VIPs as it carries out tasks, or mobile devices for a buddy drone helping with edge analytics for a service drone in its vicinity.
- The platform should leverage the capability of cloud along with the edge devices to perform complex visual analytics tasks in a timely manner to support diverse applications.

In this paper, we address a subset of these technical challenges and achieve some of the design goals, as described next, while the rest are left to future work.

4 PROTOTYPE

We have designed and implemented an early prototype of *Ocularone* to support some of the design requirements specified above. Here, we describe our approach for a buddy drone to uniquely identify and shadow the VIP (§ 4.1), have situation awareness to detect a safety incident (§ 4.2) and respond to commands based on visual gestures (§ 4.3).

The VIP carries a mobile compute device – a *Jetson Nano* in our prototype – which is connected through a wireless link to a buddy drone assigned to them. Each buddy drone, a *DJI Tello quad-copter* in our implementation, has an onboard camera and other sensors such as an altimeter, GPS, etc. The buddy drones are connected to the mobile device over a wireless link for streaming live videos from the camera for analysis and receiving automated navigation control signals. The mobile device performs onboard analytics over the videos to drive these controls, and optionally offloads some analysis to cloud resources accessed through a cellular link.

Our platform predominantly relies on video-based analytics over drone feeds for perception and decision making. Performing such video inferencing in real-time is essential, but computer vision algorithms and DNN inferencing over videos are computationally demanding. Also, each analytic may have a different priority depending on the application consuming it at that time, e.g., for drone navigation, detecting safety hazards, or just informing the VIP of nearby points of interest. We need to intelligently use the limited computing power of the GPU/TPU accelerated edge device available instantly, along with larger cloud resources available remotely but with higher latency. Further, cellular connectivity to the cloud may be variable, and hence time-sensitive analysis and decisions

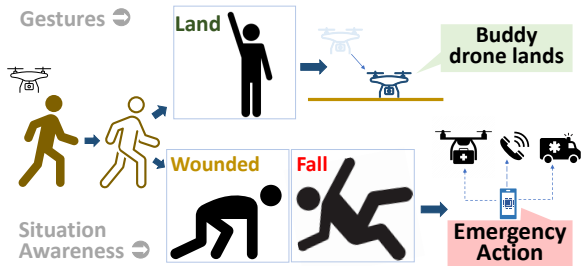


Figure 3: Situation Awareness

must happen on the edge device. We leverage our prior work on such deadline and priority-driven video analytics across edge and cloud resources for this [37].

Next, we discuss design approaches to achieve the functional capabilities that are motivated by the platform design goals.

4.1 Shadowing the VIP

A key requirement for the buddy drone is to identify and follow the VIP autonomously. We leverage computer vision algorithms over the drone's video feed to locate and localize the VIP who is nearby.

To make this task robust in our prototype, the VIP wears an *identification marker*, such as a distinctive hazard vest in our experiments. The drone initially performs a 360° pirouette to locate the VIP in its video feed. We then determine the spatial orientation of the person relative to the drone, i.e., to decide if they are facing toward or away from the drone. For the prototype, we have collected over 600 hazard vest images using the Tello drone to retrain *YOLO v4 (tiny) [10]*, a popular and light-weight DNN model for object detection to accurately detect a VIP wearing a hazard vest. We use these detections to navigate the drone to face the rear of the VIP based on their spatial orientation.

Next, we *track and follow* the VIP when they are walking. This is done by ensuring that they remain within the FoV of the drone with the relevant orientation and proximity required for the current conditions, e.g., the drone facing the rear of the VIP and with a view in front of the path they are walking along. We use the same retrained YOLO v4 DNN to detect the movement of the VIP within a video frame and assess their spatial position relative to the drone. We use the area of the bounding box generated around the hazard vest in the frame to estimate the distance of the VIP from the drone – smaller the bounding box, farther is the drone from the VIP and vice versa. We set a desired bounding box size which corresponds to a target setpoint distance between the drone and VIP.

We develop a *Proportional Derivative (PD) Control loop algorithm* [31] that uses these detections across video frames to alter the drone's direction of motion along its degrees of freedom: up, down, forward, backward, clockwise or anti-clockwise rotation (yaw), and the speed along that direction. These controls attempt to maintain the VIP in the center of the camera's FoV and within the required setpoint distance from them as they are moving. This DNN model runs on the edge device given the real-time nature of decisions and controls required.

In future, we need to evaluate and improve the accuracy of the DNN models under different lighting and background conditions, and have robust means to re-acquire the VIP if they are lost from the FoV. We also need to consider different markers such as Hiro [13] or QR code on the front/back of the vest to reliably determine their orientation (or to distinguish between two proximate VIPs). Use of alternate accessories like a cap or even an Bluetooth tether based on their smart phone to complement the spatial accuracy is also worth examining. Lastly, having reliable tracking of the VIP and control of the drone under adverserial conditions, such as varying speeds of walk/run or taking many sharp turns, needs to be supported.

4.2 Situation Awareness

Here, the goal is to examine the ambient environment around the VIP based on the drone video feed and implicitly determine if the situation requires specific actions to be taken. When navigating the VIP, they may require assistance but it may not be directly communicated to the drone, e.g., when they sits down due to exhaustion, get hurt or take a fall. Here, the VIP may be in an pose that does not match a standard standing, walking or sitting position, such as a bent-down or kneeling position or in a fallen position on the ground (Fig. 3, lower part).

To detect such conditions, our prototype uses a *pose classification model* to determine if the current situation warrants some action. We use Google's MediaPipe body-pose DNN model [25] coupled with a custom Support Vector Machine (SVM) we train. The pose estimation model returns 33 body landmarks for each person detected in the frame, e.g., right shoulder, left elbow, nose, etc. Each landmark is a tuple (x, y, z) that gives the $3D$ offset in meters of that landmark from the waist. We can use the relative position of these landmarks to determine the orientation and pose of the person. But this offers a large permutation-space of landmark positions for the same logical pose, such as a person standing or lying, and explicitly enumerating this becomes challenging and error-prone. E.g., there may be multiple kneeling positions or half upright positions with different landmark values that should be classified as kneeling.

Instead, we train an *SVM classifier* that takes the landmark vector for a person and returns the logical pose we are interested in. Towards this, we collect up to 5 minutes of video for individuals (authors of this paper) who perform each pose we wish to classify and extract representative frames from them. We are interested in five pose classes in this evaluation: *upright*, *kneel* and *fall* for situation awareness, and *single hand-up pose* and for *hand-pointing pose* gesture detection. We collect between 300–500 sample images for each class, which are used as training data for the SVM classifier. Our trained classifier achieves an accuracy of 97% for the kneel and fall safety situations.

Detecting false positives is a challenge, e.g., if we mix up a brief wave to someone with a hand pointing gesture of interest. This is reduced by checking if the current pose is held for a certain threshold duration before identifying it as an explicit command. In our prototype, we set this threshold to seeing a pose in at least 5 of the last 7 poses to detect it as an event. This translates to $\approx 3s$ for which the pose is held.

Once a situation of interest has been determined, we perform the action associated with that event. E.g., if the VIP is detected as being hurt, the drone may capture their current state, say, as a video clip from performing a 360° rotation around them, by taking a close-up

photo, or by getting close and recording an audio message from them , and send this through the mobile device to their emergency contacts and/or first responders.

In future, we plan to extend the set of such situations we can auto-detect and their corresponding actions, e.g., existing crowd-counting DNN models can be used to detect if the VIP is approaching a crowded area and alert them, even to the extent of detecting if the crowd is masked or not during a flu or pandemic season. Future actions may also include requesting a service drone to reach the VIP's current location with a first aid kit [28]. We may also use the visual sensing with other mediums such as the accelerometer on their smart phone or some audible gasp detected using a Bluetooth headset to confirm a fall. We also propose to extend the accuracy of the detections by limiting the pose detection only for the VIP (rather than all entities in a frame) by limiting the pose estimation to the bounding box where the VIP is identified by the hazard vest detection model and tracking it in the subsequent frames.

4.3 Gesture-based Intuitive Interface

We use a vision-driven gesture approach for the VIP to intuitively instruct the Ocularone platform through the buddy drone to perform specific actions. Here, specific hand gestures by the VIP are mapped to specific programmable tasks to be performed. This can complement audio-based instructions provided through a Bluetooth headset in case of a noisy neighborhood, a location requiring silence, or if the VIP does not wish to reveal what they are instructing.

The buddy drone captures the gesture from the VIP and using a similar body-pose estimation model like the one above, it determines the gesture and takes the corresponding action. We have trained the pose classifier for *two* additional pose classes: *single hand-up pose* which when detected will have the drone hover in-place, and *hand-pointing pose* to request the drone to land. E.g., to take a break when walking, the VIP can raise their hand to ask the drone to pause and hover in-place while keeping the VIP in its FoV; or if the VIP completes their walk, they can show a gesture to land and park the drone, or release it to be returned to the base station for charging. Here again, we have a threshold duration for which the gesture is held before it is treated as a command.

In future, false positives for gestures can be further mitigated by the drone asking for a positive confirmation from the VIP of the detected gesture by using an audio cue. We can also configure the gestures to conform to existing community or accessibility standards for such actions that may exist.

5 VALIDATION

We validate our prototype implementation of Ocularone using a managed experimental setup within our campus environment. Specifically, we design scenarios to verify the feasibility of *buddy drone shadowing the VIP*, *situation awareness*, and the *gesture-based interface*. The experimental setup and results are described below.

5.1 Setup

Ryze Tello drone [39] powered by DJI is used as the buddy drone for a VIP. Tello is a lightweight quad-copter with a nano form factor that weights just $80g$. It has an onboard 720p HD monocular camera that generates feeds at 30 frames per second (FPS), infrared sensors, an internal downward-facing camera to keep the Tello stable and an altimeter to determine the height from the ground. We use an *Nvidia Jetson Xavier Nano* developer kit [16] as our accelerated edge mobile device. It has a 128-core Maxwell GPU, quad-core ARM A57 CPU @ 1.43 GHz, 4GB RAM shared by CPU and GPU, and uses an SD-card for storage. It measures just 70×45 mm in size, is $280g$ heavy, and consumes $< 10W$ of power. It can be powered using a USB battery pack. The Tello uses WiFi 802.11n protocol at 2.4 GHz to interface with the Jetson Nano with a WiFi dongle. This has a range of $100m$ with a clear line of sight. This link is used to send live video data from the drone camera and to receive control commands from the edge.

The Nano runs *two DNN inferencing models*: YOLOv4-tiny for hazard vest detection model, and the Google MediaPipe pose estimation model for situation awareness and gestures. These have an average inferencing latency of $90ms$ and $170ms$ per frame, respectively. We run YOLO on the video frames sampled at 6 FPS whereas pose estimation is inferenced at 2 FPS. These frame rates are chosen to balance the computational budget of performing the inferencing on the edge ($90ms \times 6 + 170ms \times 2 = 880ms$ of computation load per second of video), and the responsiveness required for the task performed – with shadowing the VIP using the hazard vest detection being more time sensitive compared to pose detection that is tolerant to latency. The latency to decide the next action based on the DNN inferencing and send the relevant commands from the Nano to the drone takes a negligible ≈ 1 ms. This allows for real-time tracking of the VIP, and their pose estimation for the situation awareness and gesture recognition.

The scenarios described below are conducted in a controlled outdoor environment at our university campus. There is no active vehicle or foot traffic during the experiments. One of the co-authors served as a proxy subject for the VIP. The proxy is not visually challenged and performs his tasks in a measured but natural manner. Our system is configured to have the Tello drone follow behind the VIP at a constant distance of $\approx$ $2m$ while maintaining a constant height of $\approx$ $130cm$ from the ground. In our experiments, the Nano was placed in the vicinity of the drone and the VIP (but not on their person). This does not impact the functional accuracy of the experiments.

5.2 Results

Fig. 4a shows a third-person view captured using an external camera of the buddy drone autonomously following the person wearing the hazard vest. Further, we evaluate five scenarios for gesture recognition and situation awareness described below. For each, we show in Figs. 4b–4f a representative video frame from the drone camera overlaid with the body landmarks returned by the the body pose estimation model. We report the classification of different poses by the model in the top left of the frame and the corresponding commands being sent to the drone on the top right corner. The supplementary video accompanying this article has more details.

In Fig. 4b, the body pose model and our trained SVM classifier are able to detect from the drone video that the VIP has raised one hand above their shoulder level, and this is mapped to the `HOVER` command. This requests the drone to hover at its current position and stop following the VIP, until the same gesture is shown again.

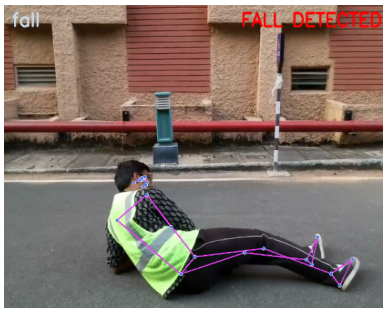
(a) Tello drone (circled) following a VIP

(b) VIP gesturing for drone to hover *(Drone camera view)*

(c) VIP gesturing for drone to land *(Drone camera view)*

(d) VIP in safe upright position *(Drone camera view)*

(e) VIP kneeling upon a simulated injury *(Drone cam view)*



(f) VIP after a simulated fall *(Drone camera view)*

Figure 4: Validation of Ocularone prototype using a proxy VIP for several scenarios in a controlled campus environment

In Fig. 4c, the body pose and SVM models detect that the VIP has raised their hand perpendicular to the body and enacts the `LAND DRONE` command, which instructs the drone to land.

Finally, we report evidence for real-time situation awareness. These poses are critical as they may imply a health and safety condition requiring immediate attention. In Fig. 4d, the body pose and SVM models detect the default pose of the VIP in an upright position and treats this as a `SAFE` situation, which does not require the drone to take any alternate action. In Fig. 4e, the VIP sits in a kneeling position to simulate an injury. The models classify this position accurately as kneel. Similarly, in Fig. 4f, the VIP simulates a fall which is also correctly detected by the models. In both these case, the drone maps the pose of the VIP to the situation of `FALL DETECTED`, which can be used to trigger some emergency command.

6 DISCUSSION AND CONCLUSION

In this paper, we have proposed the Ocularone distributed analytics platform to leverage a fleet of drones with onboard cameras for guiding the visually impaired in an urban environment. We have also described a light-weight prototype implementation of the platform with initial validation in a controlled space. The vision features provided by the drones add an extra dimension to the sensory surroundings of the VIP. Existing technologies either curtail the autonomy of the users by adding weight or provide a limited FoV to the VIP, thus restricting their mobility in an unaided manner. Further, our platform can be used in collaboration with other smart devices with a higher battery capacity, lower form factor, etc. This socio-technical collaboration can help the VIPs feel more connected to their surroundings and help them experience life from a different perspective.

Our design, however, has certain limitations. Using drones for the visually impaired on a day-to-day basis has not been deployed earlier, thus making it challenging for them to be accepted as safety devices. Passers-by may be curious and attempt to interfere physically with the drone's operations. The framework is also restricted by technical challenges. Small consumer drones have limited battery capacity, currently ranging in 10s of minutes. Lightweight drones may fail in adverse environmental conditions like heavy rains, wind gusts, etc. Due to the mobility of VIP, there may be network variation between the edge device and the cloud that increases the latency for off-loading costly analytics. We also need to investigate the trade-offs between the speed of the VIP and the accuracy of our system in processing real-time results. Inaccuracies in the computer vision models due to low lighting, etc., may cause a drone to operate incorrectly and potentially collide with the VIP, nearby people, or proximate objects.

Besides attempting to address some of these limitations, we plan to further extend the prototype implementation to achieve the other design goals we have proposed. As ongoing work, we are examining path-planning for real-time collision avoidance for the VIP and for the buddy drone. We also propose to make our system robust in case the VIP goes out of FoV of the drone momentarily. We will also explore algorithms for the coordination of multiple drones deployed as a part of our platform, which can be challenging. We need to further validate our work using real VIP subjects (after

human subject committee approvals) to understand the practical, technical and social gaps, and adapt it to suit the end-user. Overall, we believe using multiple drones with edge computing, computer vision, and AI can significantly enhance the independence and quality of life for visually impaired persons.

ACKNOWLEDGMENTS

The authors thank Dibyajyoti Nayak, Ruchi Bhoot, Shreeparna Dey and Tuhin Khare from IISc for their assistance. The first author was supported by the Prime Minister's Research Fellowship (PMRF) from the Government of India.

REFERENCES

[1] European Union Aviation Safety Agency. 2022. *Easy Access Rules for Unmanned Aircraft Systems (Regulations (EU) 2019/947 and 2019/945).* Technical Report.

[2] Abrar Al-Heeti. 2020. *Google expands Lookout app for people who are blind or vision-impaired.* Technical Report.

[3] Majed Al Zayer, Sam Tregillus, Jiwan Bhandari, Dave Feil-Seifer, and Eelke Folmer. 2016. Exploring the Use of a Drone to Guide Blind Runners. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS).* https://doi.org/10.1145/2982142.2982204

[4] Inc. American Printing House for the Blind. 2020. *Nearby Explorer (Full) and Nearby Explorer Online for Android User Guide.* Technical Report.

[5] Mauro Avila, Markus Funk, and Niels Henze. 2015. DroneNavigator: Using Drones for Navigating Visually Impaired Persons. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS).* https://doi.org/10.1145/2700648.2811362

[6] Mauro Avila Soto, Markus Funk, Matthias Hoppe, Robin Boldt, Katrin Wolf, and Niels Henze. 2017. DroneNavigator: Using Leashed and Free-Floating Quadcopters to Navigate Visually Impaired Travelers. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS).* https://doi.org/10.1145/3132525.3132556

[7] Shiri Azenkot, Catherine Feng, and Maya Cakmak. 2016. Enabling building service robots to guide blind people a participatory design approach. In *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* https://doi.org/10.1109/HRI.2016.7451727

[8] Jinqiang Bai, Shiguo Lian, Zhaoxiang Liu, Kai Wang, and Dijun Liu. 2017. Smart guiding glasses for visually impaired people in indoor environment. *IEEE Transactions on Consumer Electronics* (2017). https://doi.org/10.1109/TCE.2017.014980

[9] Maryam Bandukda, Catherine Holloway, Aneesha Singh, Giulia Barbareschi, and Nadia Berthouze. 2021. Opportunities for Supporting Self-Efficacy Through Orientation & Mobility Training Technologies for Blind and Partially Sighted People. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS).* https://doi.org/10.1145/3441852.3471224

[10] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. https://doi.org/10.48550/ARXIV.2004.10934

[11] Dalvin Brown. 2021. *Researchers design an AI-powered backpack for the visually impaired.* Technical Report.

[12] Federal Aviation Administration Code of Federal Regulations. 2022. *General Operating and Flight Rules (14 CFR Part 91).* Technical Report.

[13] Ashley Colley, Dennis Wolf, Klaus Kammerer, Enrico Rukzio, and Jonna Häkkilä. 2020. Exploring the Performance of Graphically Designed AR Markers. In *Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia (MUM).* https://doi.org/10.1145/3428361.3432076

[14] Microsoft Corporation. 2017. Seeing AI.

[15] Paulo Costa, Hugo Fernandes, João Barroso, Hugo Paredes, and Leontios J. Hadjileontiadis. 2016. Obstacle detection and avoidance module for the blind. In *2016 World Automation Congress (WAC).* https://doi.org/10.1109/WAC.2016.7582990

[16] NVIDIA Developer. 2019. Jetson Nano Developer Kit.

[17] Hugo Fernandes, Nuno Conceição, Hugo Paredes, António Pereira, Pedro Araújo, and João Barroso. 2012. Providing accessibility to blind people using GIS. *Universal Access in the Information Society* (2012). https://doi.org/10.1007/s10209-011-0255-7

[18] Youngjib Ham, Kevin K Han, Jacob J Lin, and Mani Golparvar-Fard. 2016. Visual monitoring of civil infrastructure systems via camera-equipped Unmanned Aerial Vehicles (UAVs): a review of related works. *Visualization in Engineering* (2016). https://doi.org/10.1186/s40327-015-0029-z

[19] Stephan Huber, Anastasia Alieva, and Aaron Lutz. 2022. Vibrotactile Navigation for Visually Impaired People. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS).* https://doi.org/10.1145/3517428.3550387

[20] iMerciv Inc. 2014. iMerciv: A Wearable Mobility Tool for the Blind.

[21] Enable Group in Microsoft Research. 2018. *Microsoft Soundscape: A map delivered in 3D sound.* Technical Report.

[22] Gesu India, Mohit Jain, Pallav Karya, Nirmalendu Diwakar, and Manohar Swaminathan. 2021. VStroll: An Audio-Based Virtual Exploration to Encourage Walking among People with Vision Impairments. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS).* https://doi.org/10.1145/3441852.3471206

[23] Md. Milon Islam, Muhammad Sheikh Sadi, Kamal Z. Zamli, and Md. Manjur Ahmed. 2019. Developing Walking Assistants for Visually Impaired People: A Review. *IEEE Sensors Journal* (April 2019). https://doi.org/10.1109/JSEN.2018.2890423

[24] Tiffany Liu, Javier Hernandez, Mar Gonzalez-Franco, Antonella Maselli, Melanie Kneisel, Adam Glass, Jarnail Chudge, and Amos Miller. 2022. Characterizing and Predicting Engagement of Blind and Low-Vision People with an Audio-Based Navigation App. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA).* https://doi.org/10.1145/3491101.3519862

[25] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. 2019. MediaPipe: A Framework for Building Perception Pipelines. (2019). https://doi.org/10.48550/ARXIV.1906.08172

[26] Rezylle Milallos, Vinita Tibdewal, Yiwen Wang, Andre Ogueh Udegbe, and Tae Oh. 2021. "Would the Smart Cane Benefit Me?": Perceptions of the Visually Impaired towards Smart Canes. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS).* https://doi.org/10.1145/3441852.3476524

[27] India Ministy of Civil Aviation. 2023. *Accessibility Standards and Guidelines for Civil Aviation.* Technical Report.

[28] Alec Momont. 2014. *Ambulance Drone at TU Delft.* Technical Report. TUDelft.

[29] Mohammad Moshref-Javadi and Matthias Winkenbach. 2021. Applications and Research avenues for drone-based models in logistics: A classification and review. *Expert Systems with Applications* (2021). https://doi.org/10.1016/j.eswa.2021.114854

[30] Anirudh Nagraj, Ravi Kuber, Foad Hamidi, and Raghavendra SG Prasad. 2021. Investigating the Navigational Habits of People Who Are Blind in India. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS).* https://doi.org/10.1145/3441852.3471203

[31] IJ Nagrath. 2006. *Control systems engineering.* New Age International.

[32] Moustafa M. Nasralla, Ikram U. Rehman, Drishty Sobnath, and Sara Paiva. 2019. Computer Vision and Deep Learning-Enabled UAVs: Proposed Use Cases for Visually Impaired People in a Smart City. In *Computer Analysis of Images and Patterns.* Springer International Publishing. https://doi.org/10.1007/978-3-030-29930-9_9

[33] Norzailawati Mohd Noor, Alias Abdullah, and Mazlan Hashim. 2018. Remote sensing UAV/drones and its applications for urban areas: a review. *IOP Conference Series: Earth and Environmental Science* (2018). https://doi.org/10.1088/1755-1315/169/1/012003

[34] World Health Organization. 2022. *Blindness and visual impairment fact sheets.* Technical Report.

[35] Shishir G. Patil, Don Kurian Dennis, Chirag Pabbaraju, Nadeem Shaheer, Harsha Vardhan Simhadri, Vivek Seshadri, Manik Varma, and Prateek Jain. 2019. GesturePod: Enabling On-Device Gesture-Based Interaction for White Cane Users. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST).* https://doi.org/10.1145/3332165.3347881

[36] Louis Pizzarello, Adenike Abiose, Timothy Ffytche, Rainaldo Duerksen, R. Thulasiraj, Hugh Taylor, Hannah Faal, Gullapali Rao, Ivo Kocur, and Serge Resnikoff.

2004. VISION 2020: The Right to Sight: A Global Initiative to Eliminate Avoidable Blindness. *Archives of Ophthalmology* (2004). https://doi.org/10.1001/archopht.122.4.615

[37] Suman Raj, Harshil Gupta, and Yogesh Simmhan. 2023. Scheduling DNN Inferencing on Edge and Cloud for Personalized UAV Fleets. In *IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID)*. (To Appear).

[38] L. Ran, S. Helal, and S. Moore. 2004. Drishti: an integrated indoor/outdoor blind navigation system and service. In *Second IEEE Annual Conference on Pervasive Computing and Communications (PERCOM)*. https://doi.org/10.1109/PERCOM.2004.1276842

[39] Ryze Tech. 2018. RYZE Tello: Powered by DJI.

[40] WeWALK LIMITED UK. 2020. WeWalk: Enhancing the mobility of visually impaired people.

[41] Michele A. Williams, Caroline Galbraith, Shaun K. Kane, and Amy Hurst. 2014. "just Let the Cane Hit It": How the Blind and Sighted See Navigation Differently. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*. https://doi.org/10.1145/2661334.2661380