

Research Methods and Professional Development (ACFA7004) Research Proposal

Gesture-Driven Control of Mobile Robotics

University of Wales Trinity St David

Erhan Kocal 2521441

Lecturer : Seena Joseph

Project Supervisors : Simon Thomas & Mark Broadhead

ABSTRACT: The use of quadrupedal robots in search and rescue (SAR) operations provides a revolutionary way to navigate high-altitude, unstable terrain that is frequently too dangerous for human rescuers or conventional wheeled vehicles. These self-governing devices are being used more for initial exploration, medical supply transportation and victim location in hazardous environment or any other natural disaster. However, these robots' effectiveness is now dependent on manual control interfaces, either tablets or specialized joysticks, which create serious operational difficulties in alpine environments. Rescuers must regularly remove protective insulation when using a touch-screen or joystick in freezing conditions, which causes physical discomfort and decreased efficiency. Additionally, a rescuer's attention is diverted from the mountain's present environmental risks, including shifting snowpacks or thin edges, by the demand for "eyes-down" supervision of a controller. The most promising approach for high-risk field robotics appears to be gesture-based control, which enables "eyes-up, hands-free" contact, according to recent developments in computer vision (CV) and human-robot interaction (HRI). The use of gesture recognition in alpine search and rescue is still in its early stages, despite its advanced status in consumer electronics.

This research is aiming to develop real-time hand gesture control interface for Boston Dynamics Spot, specifically optimized for mountain rescue operations. The lack of an edge-deployed, multi-modal (Thermal-RGB) gesture framework capable of precisely interpreting high-level commands from a rescuer wearing bulky PPE in high-glare, alpine settings gap leads to this research.

The "Glove-Aware" gesture dataset will be created especially to get beyond the hand deformation and visual noise that are typical in alpine settings. Using Temporal Shift Modules (TSM), a lightweight MobileNetV3 model optimized for local operation on a Jetson Orin Nano edge processor will be trained to identify dynamic rescue requests. Finally, a low-latency target is <200ms and "eyes-up" control loop that operates independently of cellular or cloud connectivity is ensured by the direct integration of these classified gestures with the Boston Dynamics SDK.

This research addresses the gap between experimental gesture recognition and rough alpine reality by providing a glove-compatible, thermal-resilient AI framework that allows rescuers to maintain full situational awareness while maintaining physical protection.

INDEX TERMS: human-robot interaction (HRI), search and rescue (SAR), computer vision (CV), gesture recognition, spot hand gesture.

2024

At a glance: the year in numbers

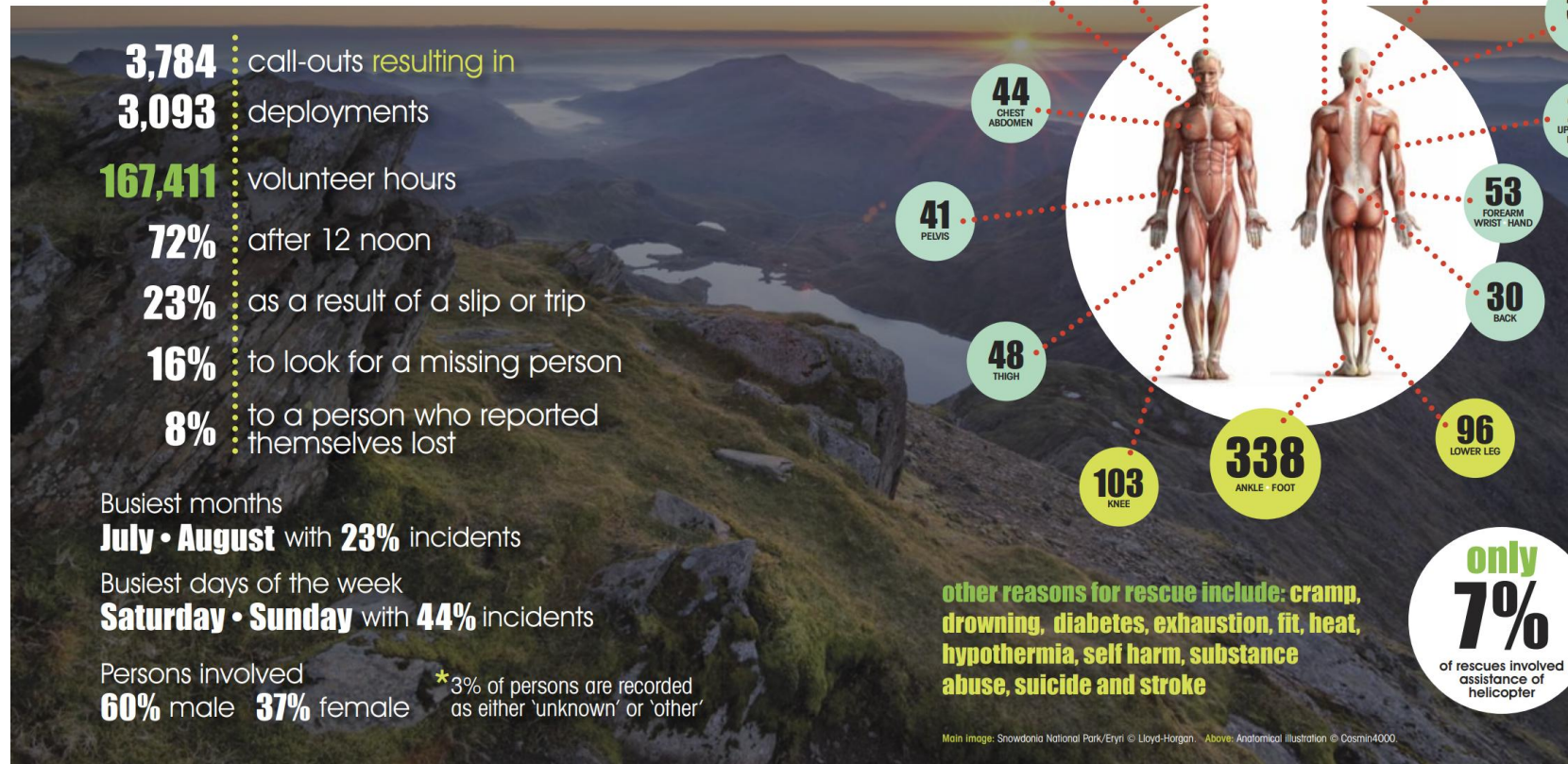


Fig 1. Mountain Rescue England and Wales Incidents in 2024 [4].

1. INTRODUCTION: Human-Robot Interaction (HRI) is a multidisciplinary field in modern science and technology that focuses on human-robot interactions. The employment of robots has evolved significantly over time in parallel with the enormous advancements in science and technology. In the past, the primary requirement for human-robot interaction was mass automation. Nowadays, robots are more useful in our daily lives in addition to their industrial uses. Drone cameras, for instance, transport packages and serve food in restaurants. As a result, one of the most challenging research areas in the last ten years has been gesture-based robot interaction. Gesture recognition is a popular technique for robot navigation since a variety of sensing devices are easily accessible. For robot navigation, pointing or directing gestures are thought to be more reliable and effective than voice recognition, which is dependent on accent and voice frequency [1].

1.1. Project Impact: Reduction in Operator Cognitive Load: Rescuers can concentrate on environmental threats instead of a control screen by turning on "eyes-up" navigation. Studies show that in high-stress SAR situations, intuitive HRI significantly boosts mission success rates [2].

1.2. Project Limitations:

1.2.1. Environmental Navigation: Robots have to navigate across congested, debris-filled, and constantly shifting surroundings.

1.2.2. Victim Interaction: Mechanisms that guarantee safety, prevent additional harm, and can adjust to the various physical circumstances of victims are necessary for any physical contact with casualties.

1.2.3. Operational Safety: To protect both the victims and the machinery, strict safety regulations must be upheld, requiring sophisticated control techniques and exact operation [3].

1.3. Perception Sensors: To improve awareness and understanding perception sensors collect data from the surroundings. SAR robots use these sensors to gather information about their environment. Decision-making processes are supported by the data gathered by perception sensors, which enables systems to react appropriately to shifting circumstances or situations.

1.3.1. Light Detection and Ranging (Lidar): Provides a practical method of providing reliable perception in a range of environments and weather conditions, including potential underwater applications, by using infrared (IR) lasers as a light source. This technology, which is well known for its excellent resolution and low cost, faces challenges because to its susceptibility to lens contamination and reduced accuracy in fog and heavy rain.

1.3.2. Radio Detection and Ranging (Radar): Despite having a lesser resolution, it is crucial for SAR applications because of its adaptability in a variety of settings and wide detection range, which make it necessary for efficient remote sensing. Radar shows durability and less sensitivity to external elements like interference or terrain changes in difficult situations. Radar's importance in supplying crucial information for SAR applications is highlighted by the trade-off between resolution and detection range [5].

REFERENCES

- [1] J. A. Mahmud, B. C. Das, J. Shin, K. M. Hasib, R. Sadik, and M. F. Mridha, "3D Gesture Recognition and Adaptation for Human–Robot Interaction," IEEE Access, vol. 10, pp. 116485–116503, 2022, DOI: 10.1109/ACCESS.2022.3218679.
- [2] S. Zimmermann, R. Poranne and S. Coros, "Go Fetch! – Dynamic Grasps using Boston Dynamics Spot with External Robotic Arm," 2021 IEEE International Conference on Robotics and Automation (ICRA), 2021, pp. 1170-1176. DOI: [10.1109/ICRA48506.2021.9561835](https://doi.org/10.1109/ICRA48506.2021.9561835)
- [3] R. P. Saputra et al., "Rescue Robots for Casualty Extraction: A Comprehensive Review," IEEE Access, vol. 13, pp. 2451–2483, 2025, DOI: 10.1109/ACCESS.2025.3637567.
- [4] https://www.mountain.rescue.org.uk/annual_review/annual-review-2025/
- [5] K. Tong, Y. Hu, B. Dikic, S. Solmaz, F. Fraundorfer, and D. Watzenig, "Robots Saving Lives: A Literature Review About Search and Rescue (SAR) in Harsh Environments," in 2024 IEEE Intelligent Vehicles Symposium (IV), Jeju Island, Korea, 2024, pp. 953-960, DOI: 10.1109/IV55156.2024.10588685.