

Computer Vision and Deep Learning-Enabled UAVs: Proposed Use Cases for Visually Impaired People in a Smart City

Moustafa M. Nasralla¹, Ikram U. Rehman², Drishty Sobnath³ and Sara Paiva⁴

¹ Dep. of Communications and Networks Engineering, Prince Sultan University, Saudi Arabia

² School of Computing and Engineering, University of West London, United Kingdom

³ Research, Innovation and Enterprise, Solent University, Southampton, United Kingdom

⁴ ARC4DigiT, Instituto Politécnico de Viana do Castelo, Portugal

Abstract. Technological research and innovation have advanced at a rapid pace in recent years, and one group hoping to benefit from this, is visually impaired people (VIP). Technology may enable them to find new ways of travelling around smart cities, thus improving their quality of life (QoL). Currently, there are approximately 110 million VIP worldwide, and continuous research is crucial to find innovative solutions to their mobility problems. Recent advances such as the increase in Unmanned Aerial Vehicles (UAVs), smartphones and wearable devices, together with an ever-growing uptake of deep learning, computer vision, the Internet of Things (IoT), and virtual and augmented reality (VR)/(AR), have provided VIP with the hope of having an improved QoL. In particular, indoor and outdoor spaces could be improved with the use of such technologies to make them suitable for VIP. This paper examines use cases both indoors and outdoors and provides recommendations of how deep learning and computer vision-enabled UAVs could be employed in smart cities to improve the QoL for VIP in the coming years.

Keywords: Deep Learning, Computer Vision, UAVs, Drone, Visually Impaired People (VIP), Smart City.

1 Introduction

According to the authors in [1], the World Health Organization (WHO) categorizes visual function under four headings: normal and moderate vision, vision impairment and blindness. Globally, it is estimated that there are 38 million blind people and 110 million people with visual disabilities. The aims of research in this area, according to [2], have centered on finding solutions to enhance the quality of life (QoL) of those afflicted, particularly in terms of safe mobility in urban areas, and enabling them to undertake daily tasks safely. In recent years there have been substantial technological developments that could potentially improve the QoL of the visually impaired. This is particularly evident in the advancement of navigation within smart cities, which has greatly encouraged visually impaired people (VIP).

The authors in [3] identify the expression ‘smart city’ as a term originating in the 1990s that encompasses the evolution of city development, technological advancement, exponential modernisation and globalisation. Poor eyesight should not preclude people from taking an active part in everyday life. The current pace of life necessitates tasks involving an increasingly complex input of information. In particular, education and employment require access to a wide range of detailed information.

Visually impaired people in smart cities, according to [4], benefit from the wide-ranging support provided by a profusion of innovations: cloud computing, embedded systems, remote sensors, wireless networks and robotics, sophisticated smartphones, UAVs, the Internet of Things (IoT), augmented and virtual reality (AR/VR) and artificial intelligence (AI). UAVs, sometimes referred as drones, are a class of aircrafts that can fly without the onboard presence of pilots, which could also provide potential solutions for the mobility of VIP.

It is the nascent time for smart cities, like those now developing in New York, Beijing, Singapore and Dubai. Smart cities of the future will expand and develop to incorporate new technology and ideas that will be

tested and subsequently incorporated into their infrastructure. This bodes well for VIP, improving their QoL and if the current investment is sustained, this will lead towards full incorporation into the experience.

The European Innovation Partnership on Small Cities and Communities (EIP-SCC) is an initiative partly funded by a one-billion-euro investment from the European Commission. The aim of the partnership, according to the European Commission in [5], is to unite citizens and industry in more than 300 cities. The visually impaired will be one group that gains advantage from this project.

The authors in [6] observe that technology is advancing to the extent that with IoT-enabled wireless development, the increased usage of UAVs and the employment of big data analysis and AI, image recognition and the achievement of certain activities are within the realms of possibility. Innovative recommendations include the employment of UAVs, utilising their inbuilt navigation systems to replace the established guide dog and assist VIP with their mobility, guiding them to their objective (Figure 1). Drones or UAVs have the added benefits of image analysis and recognition. A built-in camera is able to identify likely hazards, and via either audio signals, messages or by vibratory sensors, can warn VIP of potential threats in their path. Furthermore, the provision of a collision detection system in intelligent drones would alert a VIP of any threat of contact with people or objects [7].



Figure 1. VIP receiving drone-provided navigation guidance [7].

This paper makes recommendations referring to the development of technologies to meet these requirements. The aim is to provide VIP with the timely enhancement of movement and activity in a smart city environment.

The rest of the paper is organised as follows. Section 2 provides background of deep learning and computer vision application. Section 3 provides related work. In Section 4, we present deep learning and computer vision-enabled UAV use cases to assist VIP. Finally, Section 5 concludes the paper.

2 Deep learning and computer vision applications

Lately in some fields, state-of-the-art machine learning techniques have been surpassed by deep learning methods. This is evident in a number of areas, particularly in computer vision. Computational models of multiple processing layers, utilising deep learning, are able to learn and reproduce data with multiple layers of abstraction, emulating the perception and comprehension of multimodal information as processed by the brain, thereby assimilating complex structures of large-scale data [8].

Deep learning has enabled major advancements in solving a number of computer vision problems: face and action recognition, object detection, motion tracking and human pose estimation, and semantic segmentation [8]. Convolutional neural networks (CNNs), Deep Boltzmann Machines (DBMs), deep belief networks (DBNs) and stacked denoising autoencoders (SDAs) are some of the most prominent deep learning skills, described as follows:

1. Convolutional neural networks

The structure of the visual system led to the development of convolutional neural networks (CNNs), made up of convolutional, pooling and fully connected main types of layers. This highly effective system is used in diverse computer applications, where multiple layers are used in an end-to-end format [9], and it provides one of the most effective forms of deep learning applications.

2. Deep belief networks and Deep Boltzmann Machines

First proposed by Hinton et al. (1986), the Restricted Boltzmann Machine (RBM) is a Generative Stochastic Network (GSN) framework. Both the DBM and the DBN are part of the 'Boltzmann family', deep learning models that use the RBM as a learning module [10]. Based on the RBM, the machine is constrained, in that both the visible and hidden units need to form a part of the bipartite graph. This enables efficient training algorithms, especially gradient-based contrastive divergence algorithms [8].

3. Stacked denoising autoencoders

RBM is a major component of DBNs; SDAs are built around the autoencoder as the principal component. On this basis, it is necessary to describe the basic features of the autoencoder and its denoising model, prior to explaining the deep learning architecture of the SDA. As a particular form of the artificial neural network, the autoencoder is employed to learn efficient encodings. Rather than present given inputs to the network in order to pre-define a specific target value, the autoencoder is programmed to create its own inputs, whereby the output and input vectors have the same ambit [11].

Deep learning will help those with disabilities, particularly VIP. The crucial benefit is to have real-time support. To achieve this, drones or UAVs need to have systems installed for maximum deep learning benefits. Outside real-time support will not necessarily fulfil the need for access to be unrestricted by time or location. The IoT and similar advances in technology would provide backend information and data. However, the drone is the front-end provider, delivering essential and timely information to VIP during their daily activities. Recently, research, academic and industrial entities have more widely adopted AI and deep learning-enabled intelligent computer vision systems, cooperating to improve the QoL of VIP. The emphasis has been on developing more advanced navigation and hazard avoidance systems, utilising AI-based solutions. The

requirements have encompassed such diverse subjects as navigation tools to traverse indoor and outdoor areas, computer vision, AR/VR tools to enhance human vision, road crossing assistance, the location and identification of bus stops and drone guidance. This has been made possible by advancements in image and object capturing systems, and AI and machine learning development techniques such as deep learning, fuzzy inference systems (FISs) and neural networks.

3 Related work

There is the prospect of major development for the mobility of VIP in a smart city environment. Deep-learning-based UAVs and the advancement of computer vision are evolving and being enhanced constantly to improve hazard detection.

Currently, information and communication applications (ICTs) are being employed for the improvement of QoL for VIP. The recently completed Ambient Assisted Living (AAL) project was an initiative encompassing a number of EU projects with this aim. There are three stages in drone image recognition, namely (1) object detection and capture, (2) segmentation, and (3) object recognition. Deep neural networks (DNNs) are used increasingly for object recognition, replacing previous machine learning algorithms such as support vector machines (SVMs) and neural networks. The DNN system works on a deep-learning form of machine learning technique and is proving to be adaptable in incorporating environmental changes and identifying minor changes in an extracted object feature [12]. Current and predicted development results identify image recognition-based drones as paramount in improving the QoL of VIP in a smart city environment. However, there is a need for a large dataset for DNN training, which is one of the challenges in implementing such systems. This places a resultant high demand on time and computer power resources, requirements that can seriously curtail drone battery life.

The methods used for image-based object detection have proved efficient, particularly in an agricultural setting, where techniques employing classical shape and colour-based ground vehicle detection have proved to be effective [13]. The development of CNNs [8] has enabled the resolving of object detection issues, based on the deep learning method. In the ImageNet competition, algorithms like Single Shot Detector (SSD) and You Only Look Once (YOLO) [14] performed well in the detection phase, while AlexNet and GoogLeNet displayed excellent performance and speed results. Nevertheless, the algorithms lacked the computing power requirement needed for an embedded system.

As an alternative, a drone-based prototype navigation system was presented, incorporating an en-route auditory information and warning system for VIP travelling to their objective [7]. It was recommended that VIP wear a device such as a bracelet to facilitate the passage of voice directions from the VIP to the drone. Five walking trials were conducted, using a blind VIP as a case study. Various movements were tested such as making full 180-degree turns, walking in a straight line and navigating through a crowded room. The drone gave a clearly audible signal to warn both the VIP and people nearby of its presence. The trials proved successful, which pleased the VIP, although neither deep learning nor object recognition was tested, as the aim of the trial was to evaluate the navigation capabilities of the system. In addition, during the trial, the VIP used his white cane to locate nearby obstacles.

Studies investigating methods to improve the QoL of VIP ([15] – [20]) recorded a range of aids used to assist in avoiding hazards; these included microcontrollers, webcams, audio feedback mechanisms and

ultrasonic sensors. The limitations of mobility and costs were taken into consideration, particularly as the equipment tended to be in the form of cumbersome prototypes with neither deep learning nor AI in their tactile vision systems.

4 Use cases

A set of use cases was presented to highlight the requirement for the employment of drones incorporating AI and deep learning, to enhance QoL. The aim was to establish safe and effective movement indoor and outdoor using VIP as subjects. Every case presented evidence of the advantages of intelligent drone usage.

4.1 Indoor scenario

It is possible to configure drones to operate indoors, employing a deep-learning form of machine-learning mechanism. For safe and effective navigation, a number of features need to be provisioned:

1. The drones need to be programmed to navigate the indoor paths available.
2. Real-time images of the environment to be used by the VIP should be installed into the drone built-in camera.
3. Accurate information about en-route hazards should be available from image and object recognition mechanisms.
4. Drones should be fitted with collision detection IoT-based sensors to protect against hitting people or structures.

Navigation is not the only service intelligent drones are able to provide for VIP. Reference images of misplaced or lost items can be uploaded via the computer vision and image analysis mechanisms. Once the drone has identified and located the item and established its coordinates it can guide the user to its position.

Sensors linked to the guide drones with their sensor-based surveillance systems could monitor VIP health states, particularly in emergency situations such as falling or sudden illness. Sensors, either worn by VIP or sited in smart homes, would be able to ascertain whether the subject had suffered a fall or was generally experiencing movement difficulties. Microsoft Kinect, an example of a drone-mounted motion sensor, is able to identify certain tasks being carried out by VIP. In a study, the authors in [21] observe that the system is also able to identify faces and objects, which further assists in reducing the possibility of indoor accidents. Google Home and Amazon Echo Dot are drone-installed voice activation programmes allowing access to a range of services such as news and weather, music and audiobooks.

4.2 Outdoor Scenario

The prime concern for the safety of VIP and the elderly is mobility related, to allow the negotiation of both familiar and unfamiliar hazardous areas. Drones, according to the authors in [4], can assist in navigating VIP safely in smart city open areas, by providing:

1. Signal-based alerts, audible and vibrotactile, to provide pedestrians with accurate positioning information.
2. Shopping support using a mobile product recognition system.
3. Adapted mobile-assisted city apps for VIP use.

The mobility of VIP can be identified and aided by the employment of deep learning and computer vision. The systems can be developed further, recording known VIP routes, particularly in smart cities where the network infrastructure is developed, and drones can be used effectively, and when combined with voice instruction functions can effectively direct people to their nominated destination. An added benefit is in tourist areas, where the systems are able to describe the local setting through which the person is passing, thereby expanding user cognisance of their location.

Guidance for university and college VIP students greatly enhances their experience. Drones assist them in moving unaided from lecture to lecture around the campus, increasing their safety and reducing possible anxiety, thereby improving their QoL in the learning environment. The computer-based algorithms help provide information and direction to places of interest such as libraries, lecture halls and cafes. They can be programmed to provide real-time information on possible problem areas, such as busy thoroughfares and traffic crossings, all of which can be accessed and questioned audibly by the VIP. The aim is to establish a personalised VIP/system relationship.

The authors in [22] observe that transportation is a major issue for VIP at every level; local, regional, national and international. Within smart cities, drones are able to provide detailed assistance, locating bus stops, and audibly giving route details together with departure and arrival times, and recognising the right bus for the user. This greatly benefits VIP, enabling them to travel on public transport without having to ask for assistance.

Supermarket shopping is a task where drones can greatly assist VIP. Those UAVs equipped with systems such as OS X developed by Apple, which has VoiceOver features, use deep learning-based image recognition technologies, and are able to identify products and make comparisons. Audible instructions and information programmes need to be user friendly in order to allow VIP to operate mobile phone apps easily.

5 Conclusion

Indoors, VIP need to be able to move without colliding with obstacles or objects and also be able to locate specific items. Outdoors, the system should assist them in safe passage, identifying the best routes and crossing points, in addition to giving audible transport information, bus stops, timings and bus identification.

Smart cities have the functionality to improve the QoL of VIP. Their development, together with technological advancements, lead to the goal of allowing the visually impaired to take an active part in social events and provides them with a greater degree of independence.

The exponential development of technological advances promises greater freedom for VIP, not only in safer and freer movement but also the benefit of a better QoL. Smart cities are modelled to allow the straightforward integration and management of ICT-based technology. Computer vision and deep learning functions, together with sensing and monitoring capabilities, are integrated to exploit current and future evolution.

The aim of this paper is to identify the development of computer vision and deep learning-enabled drones, and the need for obstacle and collision detection systems combined with image recognition features to be integrated into a single unit. The driving force has been to enhance the safety and QoL of VIP in smart cities, identifying use cases for both indoor and outdoor movement. There is an ongoing need to develop the capabilities and efficiency of deep learning and computer vision-enabled drones, particularly in smart cities.

References

- [1] C. Teixeira, A. S. Toledo, A. da S. Amorim, S. T. Kofuji, and V. Rogério dos Santos, "Visual Impairment and Smart Cities: Perspectives on Mobility.," *JOJ Ophthalmol.*, vol. 3, 2017.
- [2] Ali Jasim Ramadhan, "Wearable Smart System for Visually Impaired People," *Sensors* , vol. 13, no. 3, p. 834, 2018.
- [3] D. V. Gibson, G. Kozmetsky, and R. W. Smilor, *The Technopolis phenomenon : smart cities, fast systems, global networks*. Rowman & Littlefield Publishers, 1992.
- [4] Knud Erik Skouby, Anri Kivimäki, Lotta Haukipuro, Per Lynggaard, and Iwona Windekilde, "Smart Cities and the Ageing Population". In 32nd Meeting of WWRF, Marrakech, Morocco. Retrieved from: <https://pdfs.semanticscholar.org/d7a5/84f867996dbdf78a34697523c537dae218bc.pdf>
- [5] European Commission, "General Assembly of the European Innovation Partnership on Smart Cities and Communities (EIP-SCC) | European Commission," 2018.
- [6] N. Y. Philip and I. U. Rehman, "Towards 5G health for medical video streaming over small cells," in *IFMBE Proceedings*, 2016.
- [7] M. Avila, M. Funk, N. Henze, "Dronenavigator: Using drones for navigating visually impaired persons,.". In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (pp. 327-328), 2015.
- [8] A. Voulodimos, N. Doulamis, A. Doulamis and E. Protopapadakis, "Deep learning for computer vision: A brief review." *Computational intelligence and neuroscience*, 2018.
- [9] M. Liang, and X. Hu. "Recurrent convolutional neural network for object recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3367-3375. 2015.
- [10] G. E. Hinton and T. J. Sejnowski. "Learning and relearning in Boltzmann machines." *Parallel distributed processing: Explorations in the microstructure of cognition*, pp: 282-317, 1986.
- [11] T. Baltrušaitis, C. Ahuja, and L. P. Morency. "Multimodal machine learning: A survey and taxonomy." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, no. 2, pp: 423-443, 2018.
- [12] A. Carrio, C. Sampedro, A. Rodriguez-Ramos, and P. Campoy. "A review of deep learning methods and applications for unmanned aerial vehicles." *Journal of Sensors*, 2017.
- [13] C. Hung, Z. Xu, and S. Sukkarieh. "Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a UAV." *Remote Sensing* 6, no. 12, pp: 12037-12054, 2014.
- [14] S. Jung, S. Hwang, H. Shin, and D. H. Shim. "Perception, guidance, and navigation for indoor autonomous drone racing using deep learning." *IEEE Robotics and Automation Letters* 3, no. 3, pp: 2539-2544, 2018.
- [15] S. Cardin, D. Thalmann, and F. Vexo, "A wearable system for mobility improvement of visually impaired people," *Vis. Comput.*, 2007.
- [16] L. D. Dunai, I. L. Lengua, I. Tortajada, and F. B. Simon, "Obstacle detectors for visually impaired people," in *2014 International Conference on Optimization of Electrical and Electronic Equipment, OPTIM*, 2014.
- [17] P. Gharani and H. A. Karimi, "Context-aware obstacle detection for navigation by visually impaired," *Image Vis. Comput.*, 2017.
- [18] A. Jonnalagedda et al., "Enhancing the Safety of Visually Impaired Travelers in and around Transit Stations," 2014.
- [19] C. L. Lee, C. Y. Chen, P. C. Sung, and S. Y. Lu, "Assessment of a simple obstacle detection device for the visually impaired," *Appl. Ergon.*, 2014.

- [20] M. Poggi and S. Mattoccia, "A wearable mobility aid for the visually impaired based on embedded 3D vision and deep learning," in Proceedings - IEEE Symposium on Computers and Communications, 2016.
- [21] M. Rahman, B. Poo, A. Amin, and H. Yan, "Support System Using Microsoft Kinect and Mobile Phone for Daily Activity of Visually Impaired," in An Embedded SIP-VoIP Service in Enhanced Ethernet Passive Optical Network , 2015, pp. 425–440.
- [22] M. Patricio, L. Haidee, L. Ciro, R. Telma, and F. Felipe, "Analysis and Proposed Improvements in the Support for the Visually Impaired in the Use of Public Transportation," in SMART 2015 : The Fourth International Conference on Smart Systems, Devices and Technologies, 2015, pp. 978-1-61208-414–5.