Case Study
Assistive Technology

A Vision System for the Visually Impaired

A prize-winning implementation of Artificial Intelligence and Machine Learning results in a smart, reliable system for helping the visually impaired move through their environment.

An example of how technology can truly benefit human beings.

Challenge

The difficulties people with visual impairments experience in navigating through the world—especially in public spaces—are both obvious and, many times, profound. Their lives and opportunities are often seriously impacted by the need to depend on others for assistance in simply getting to work, going shopping, meeting friends, and on and on. What sighted people take for granted often becomes an enormous obstacle that the visually impaired must deal with on a daily basis.

Now, for what is practically speaking the first time, technology is on the verge of providing workable solutions to help overcome this challenge. Progress in AI and Machine Learning, coupled with advances in hardware performance by industry leaders such as Intel Corporation, mean that researchers now have more powerful, flexible, and sophisticated tools to work with in crafting solutions for the visually impaired.

One of these researcheers, Jagadish K. Mahendran, recently won the grand prize in the OpenCV Spatial AI 2020 Competition, the world's largest spatial AI competition. Sponsored by Intel and Microsoft, the contest challenged AI engineers to build a smart camera assembly that leverages the power of neural inference in real time. In taking home the prize, Mahendran's effort stood out from hundreds of projects entered by engineers from around the world.

Jagadish K. Mahendran

Computer Vision/Artificial Intelligence Engineer

Motivation from Real Life

For Jagadish, the spark for the project came from the issues faced by a visually impaired friend. As he explained, "Back in 2013, as I started my Master's in Artificial Intelligence, the idea of developing a visual assistance system first occurred to me. The infeasibility of using smart sensors then, combined with deep learning techniques and edge AI not being mainstream in computer vision, made it difficult to make any progress on this project.

"When I met my visually impaired friend Breean Cox, I was struck by the irony that while I have been teaching robots to see, there are many people who cannot see and need help. This motivated me to build the visual assistance system. The timing of the OpenCV Spatial AI competition could not have been better, it was the perfect channel for me to build this system and bring this idea to life."

The OpenCV team and Intel have a long history of fruitful collaboration. The OpenCV project is a library of programming functions supporting real-time computer vision. Launched in 1999 as an Intel Research initiative aimed at advancing CPU-intensive applications, it has enjoyed over two decades of success.

Besides promoting research in AI by providing a common infrastructure for developers to build on, it serves as a source for the latest research in vision-related technology. In 2012, a newly formed non-profit organization, OpenCV.org, took over responsibility for the initiative. In addition to supporting annual competitions, it maintains a developer and user site at https://opencv.org. Intel remains a Platinum level sponsor; to learn more, visit https://opencv.org/intel/.

Defining the Problem

When Jagadish began developing his project, visual assistance systems for navigation ranged from GPS-based, voice-assisted smartphone apps to camera-enabled smart walking stick solutions. No system available was designed to capture the visual scene accurately—to give the visually impaired a picture of the environment around them to the depth needed to facilitate independent navigation.

Jagadish and his team interviewed a number of people with visual impairments to understand the daily challenges facing them. The difficulty of the project quickly became evident, as there were a number of quite different obstacles that would have to be dealt with. Some were stationary, such as crosswalks and traffic lights, while others were moving, such as vehicles and other pedestrians. An obstacle could come into play from above, such as a hanging flower basket, or from below, such as the elevational change of stepping off a curb. It didn't take long to come up with a lengthy list of obstacles, which the team ranked by priority.

In approaching the problem, the team developed and tested a range of technologies, from image processing and point cloud processing to advanced AI/ML techniques. Their goal was to develop a cost-effective solution with high accuracy and inference rate. They also looked at supportive technologies such as GPS and SMS services to handle supplemental features.

The Solution

The team developed an AI solution that can be readily used by visually impaired people for navigation and scene perception in the real world.

The physical setup of the system uses a small backpack containing a host computing unit (which could be a Raspberry Pi, Chromebook, or laptop). A vest jacket conceals a camera, and a fanny pack is also used to hold a power unit.

Two Luxonis OAK-D sensors powered by the Open Source DepthAI Repositories are affixed to the chest area of the vest, then connected to the computing unit in the backpack. Three tiny holes in the vest provide viewports for the OAK-D unit, which is attached to the inside of the vest with Velcro tape.

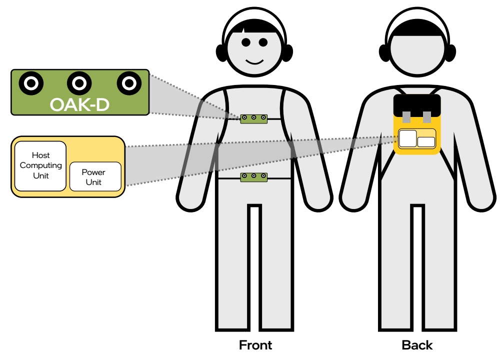
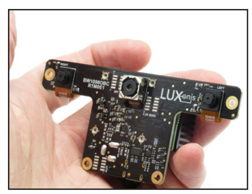


Figure 1. System design

The OAK-D unit is a versatile and powerful AI device capable of simultaneously running advanced neural networks while also providing depth of field from two stereo cameras, as well as color information from a single 4k camera. Containing an on-chip edge AI processor compatible with OpenVINO, it is powered by a pocket-size battery pack capable of providing approximately 8 hours of use.



In addition, a USB-enabled GPS is mounted on top of the backpack and connects to the host computing unit, and a wireless, Bluetooth-enabled earphone enables the user to interact with the system.

Figure 2. Luxonis OAK-D sensor

The host computing unit the team decided on was an older model Lenovo Yoga laptop powered by an Intel® Core™ i7 processor. It had 8GB RAM.

Attention was paid to hiding the sensors to make them less noticeable to other people. The final configuration is convenient to wear and is not noticeable as an assistive device.

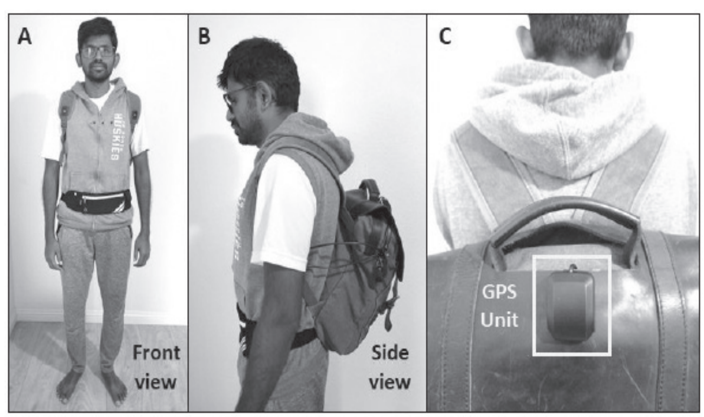


Figure 3. A: Front. B: Side. C: GPS unit placement

User Interface

The user interacts with the system via voice queries and commands, and the system responds with verbal information. Creating a well-designed user interface is critical to providing an effective and comfortable user experience. It's important not to overwhelm the user with an avalanche of information. To reduce the annoyance of continual, trivial messages, the team elected to provide most updates only on request—with the exception of critical updates for safety (avoiding collisions, awareness of elevation changes, etc.).

The system is completely controlled using voice interactions enabled by the earbuds. Pre-set commands are used to trigger system actions, such as:

- Starting the system ("Start")
- Learning about detected objects in the user's path ("Describe")
- Saving landmarks for future reference ("Save location")
- Finding saved locations ("Locate")
- Sending GPS information via SMS ("Send location")

Interaction using these commands is always subject to being overridden by the critical safety updates mentioned earlier.

What the User Experiences

Visually impaired users can take advantage of the system for both indoor and outdoor navigation, as well as for gaining an understanding of their local environment. Simple to put on, it's inconspicuous, and the user can take advantage of the system to walk freely on public streets without attracting undesired attention.

As the user moves through the locale, the system will audibly convey information about common challenges such as traffic signs or hanging obstacles, as well as moving obstacles such as other pedestrians. It will warn of upcoming crosswalks, as well as elevation changes such as curbs and entryways. No other assistance, such as a cane or guide dog, is required.

Here are some of the things the system can help enable the visually impaired user to do:

- Walk more safely on a city sidewalk.
- Better avoid obstacles, such as trash barrels, low-hanging branches, and other pedestrians.
- Be better informed of the presence of traffic and street signs.
- Be more aware of, stop at, and cross crosswalks.
- Be warned of elevation changes, such as a step up to or down from a curb.
- Instruct the system to perform actions, such as describing the current scene.

For example, if the user is walking down the sidewalk and is approaching a trash bin, the system can issue a verbal warning of "left," "right," or "center," indicating the relative position of the receptacle. When the user is approaching a corner, the system will describe what's ahead by saying "Stop sign" or "Enabling crosswalk," or both. Similarly, if the user is approaching an area where bushes or branches overhang the sidewalk, the system will issue a notice such as "Top, front," warning of something in the way.

At any time, the user can query the system for information using simple voice commands. For example, if the user says "Describe," the system might respond with "Yellow pavement, 10 o'clock...person, 11 o'clock...traffic light, 2 o'clock...car, 3 o'clock." If the user is waiting to cross a street and a car passes, the system can advise, "car, 12 o'clock", informing the user that a vehicle is dead ahead.

The user can also instruct the system to save specific locations. The command "Save location, Ice Cream store" will inform the system to store this spot for future visits.

As the user continues and the environment changes, the system can issue updates about what it is "seeing." The user can select to be updated continuously or only in case of critical warnings (where potential physical harm is possible).

The system was developed and tested in an urban area of Monrovia, California.

Solution Details

Perceiving the physical world in the immediate vicinity of the unit is the overarching objective of the project. There are three main types of perception the system must handle.

- *Primitive perception* mainly deals with object presence—what's around me? It is highly focused on detecting obstacles in order to update the user and avoid collisions. Due to the importance of this task, primitive perception concentrates on detection of objects as opposed to their identification—hence the name "primitive." It must be able to detect obstacles of various types and shapes, and at various heights, and do so with great reliability.
- *Advanced perception* moves beyond primitive perception, carrying out more complex tasks than detection, such as image classification and semantic segmentation, to provide the user with a more detailed description of the environment. Outdoor navigation for pedestrians is in many ways similar to that of self-driving cars—though those require not only robust, reliable sensors, but also huge datasets and training on complex deep learning models with high accuracies and fast inference rates. While Jagadish and the team took advantage of a number of techniques and models developed for self-driving cars,

they faced significant down-scaling challenges due to limitations on computing power, cost, and form factor.

- The third type of perception involves a *localizer*. Its purpose is to identify the user's position in the real world.

Specific Perceptual Tasks

Object detection

The team used DepthAI's SSD-MobileNet object detection pre-trained model on the PASCAL VOC dataset to detect many different classes of objects, freeing them to focus on others that are important to visually impaired pedestrians, such as traffic signs, lights, cones, hydrants, yellow-painted pavement, crosswalk buttons, public trash cans, etc.

Semantic segmentation / Crosswalk detection

Similar to what is used in self-driving cars, semantic segmentation is optimized for roads, sidewalks, road markings, and background, making it useful for detecting crosswalks at intersections.

Elevation change detection

Though non-visually impaired people don't often notice, navigating through a typical public scene involves a number of changes of elevation. Streets have curbs, buildings have stairs. Failure to spot them can result in serious injury. This project trained on thousands of images to perceive color and depth images of up-curb and down-curb movement.

Figure 4. Curb dataset for elevation

Datasets

The team incorporated a large collection of existing geographical datasets, including Google Open Image, LISA traffic, Cityscapes, and more. In addition, the team collected and labeled thousands of custom images taken on the streets of Monrovia. Because all these datasets combined would be extremely large, the team winnowed the data down, selecting images necessary for navigation, such as traffic lights, street signs, etc., and converting them to PASCAL VOC format for AI training purposes.

Models

Before deciding on models that would be accurate, fast and cost effective, the team tested a variety of possibilities. To enable the level of sophisticated understanding of the scenes needed for assistive navigation, a number of important tasks had to be accomplished, including image classification, object detection, and semantic segmentation of models. The Intel® Distribution of OpenVINO™ Toolkit played a major role in the development of lighter models with higher accuracy, and OpenVINO zoo models saved dataset collection and model training time. OpenVINO optimizations provide a huge boost in inference speed versus other technologies (up to approximately 13x). Custom models were trained using GPUs, then converted to OpenVINO format for inference. A few pre-trained TensorFlow Lite models were also used for training.

Conclusion

As an example of how technology can truly benefit human beings, the system Jagadish's team invented is hard to top. There's little doubt that this system, and the even more advanced versions that will follow, will have a profound and liberating effect on many visually impaired people.

In that spirit, the project is non-commercial and will be open sourced as the contributors to the project increase, including code, models, and datasets. In addition, the complete project will be published as a research paper in the near future.

Learn More

Intel projects related to visually impaired:

https://software.intel.com/content/www/us/en/develop/articles/success-story-using-ai-to-help-visually-impaired-people-identify-cash.html

https://devmesh.intel.com/projects/intelligent-walking-aid-for-the-visually-impaired

https://devmesh.intel.com/projects/a-visual-aid-glasses-for-visually-impaired-or-blind-people

https://devmesh.intel.com/projects/cash-recognition-for-visually-impaired

Open CV

https://store.opencv.ai/products/oak-d

Luxonis

https://luxonis.com/

https://docs.luxonis.com/en/latest/#luxonis-github-repositories

Mira

https://miravisinfo.wixsite.com/mira