

Researchers help robots navigate crowded spaces with new visual perception method

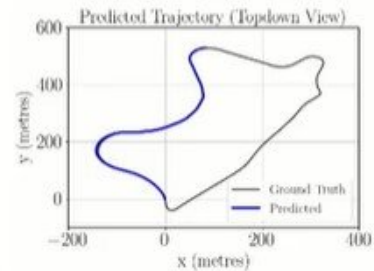
November 10 2022, by Safa Jinje



Taking a stream of raw images as input, our tightly coupled networks produce:



Dense (per-pixel) depth predictions



Visual egomotion predictions

Credit: University of Toronto

A team of researchers at the University of Toronto has found a way to enhance the visual perception of robotic systems by coupling two different types of neural networks.

The innovation could help [autonomous vehicles](#) navigate busy streets or enable medical robots to work effectively in crowded hospital hallways.

"What tends to happen in our field is that when systems don't perform as expected, the designers make the networks bigger—they add more parameters," says Jonathan Kelly, an assistant professor at the University of Toronto Institute for Aerospace Studies in the Faculty of Applied Science & Engineering.

"What we've done instead is to carefully study how the pieces should fit together. Specifically, we investigated how two pieces of the motion estimation problem—accurate perception of depth and motion—can be joined together in a robust way."

Researchers in Kelly's Space and Terrestrial Autonomous Robotic Systems lab aim to build reliable systems that can help humans accomplish a variety of tasks. For example, they've designed an electric wheelchair that can automate some common tasks such as navigating through doorways.

More recently, they've focused on techniques that will help robots move out of the carefully controlled environments in which they are commonly used today and into the less predictable world humans are accustomed to navigating.

"Ultimately, we are looking to develop situational awareness for highly dynamic environments where people operate, whether it's a crowded hospital hallway, a busy public square or a city street full of traffic and pedestrians," says Kelly.

One challenging problem that robots must solve in all of these spaces is known to the robotics community as "structure from motion." This is the process by which robots stitch together a set of images taken from a moving camera to build a 3D model of the environment they are in. The process is analogous to the way humans use their eyes to perceive the world around them.

In today's robotic systems, structure from motion is typically achieved in two steps, each of which uses different information from a set of monocular images. One is [depth perception](#), which tells the [robot](#) how far away the objects in its field of vision are. The other, known as egomotion, describes the 3D movement of the robot in relation to its environment.

"Any robot navigating within a space needs to know how far static and dynamic objects are in relation to itself, as well as how its motion changes a scene," says Kelly. "For example, when a train moves along a track, a passenger looking out a window can observe that objects at a distance appear to move slowly, while objects nearby zoom past."

The challenge is that in many current systems, depth estimation is separated from motion estimation—there is no explicit sharing of information between the two neural networks. Joining depth and motion estimation together ensures that each is consistent with the other.

"There are constraints on depth that are defined by motion, and there are constraints on motion that are defined by depth," says Kelly. "If the system doesn't couple these two neural network components, then the end result is an inaccurate estimate of where everything is in the world and where the robot is in relation."

In a recent study, two of Kelly's students—Brandon Wagstaff, a Ph.D. candidate, and former Ph.D. student Valentin Peretroukhin—investigated and improved on existing structure from motion methods.

Their new system makes the egomotion prediction a function of depth, increasing the system's overall accuracy and reliability. They recently presented their work at the International Conference on Intelligent Robots and Systems (IROS) in Kyoto, Japan.

"Compared with existing learning-based methods, our new system was able to reduce the motion estimation error by approximately 50%," says Wagstaff.

"This improvement in motion estimation accuracy was demonstrated not only on data similar to that used to train the network, but also on significantly different forms of data, indicating that the proposed method was able to generalize across many different environments."

Maintaining accuracy when operating within novel environments is challenging for neural networks. The team has since expanded their research beyond visual motion estimation to include inertial sensing – an extra sensor that is akin to the vestibular system in the human ear.

"We are now working on robotic applications that can mimic a human's eyes and inner ears, which provides information about balance, motion and acceleration," says Kelly.

"This will enable even more accurate [motion](#) estimation to handle situations like dramatic scene changes—such as an [environment](#) suddenly getting darker when a car enters a tunnel, or a camera failing when it looks directly into the sun."

The potential applications for such new approaches are diverse, from improving the handling of self-driving vehicles to enabling aerial drones to fly safely through crowded environments to deliver goods or carry out environmental monitoring.

"We are not building machines that are left in cages," says Kelly. "We want to design robust robots that can move safely around people and environments."

Provided by University of Toronto

Citation: Researchers help robots navigate crowded spaces with new visual perception method (2022, November 10) retrieved 23 April 2024 from <https://techxplore.com/news/2022-11-robots-crowded-spaces-visual-perception.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.