Monocular Depth Estimation for Human-Robot Locomotion

Qilong Cheng[1-3] and Brokoslaw Laschowski[2-4]

[1]Electrical and Computer Engineering, University of Toronto; [2]Robotics Institute, University of Toronto; [3]KITE Research Institute, Toronto Rehabilitation Institute; [4]Mechanical and Industrial Engineering, University of Toronto

kite UHN

UNIVERSITY OF TORONTO

Introduction

Depth perception can improve control and navigation of robotic leg exoskeletons and prostheses in complex real-world walking environments, ensuring safe and effective assistance. However, traditional methods have used large and expensive 3D sensors such as stereo cameras. As an alternative, we developed a novel monocular depth estimation algorithm, which uses deep learning to reconstruct and understand 3D walking environments.

Objectives

Our objective was to develop a novel algorithm for monocular depth estimation of human-robot walking environments using deep learning. Our preliminary application focused on extracting stair height and distance parameters.

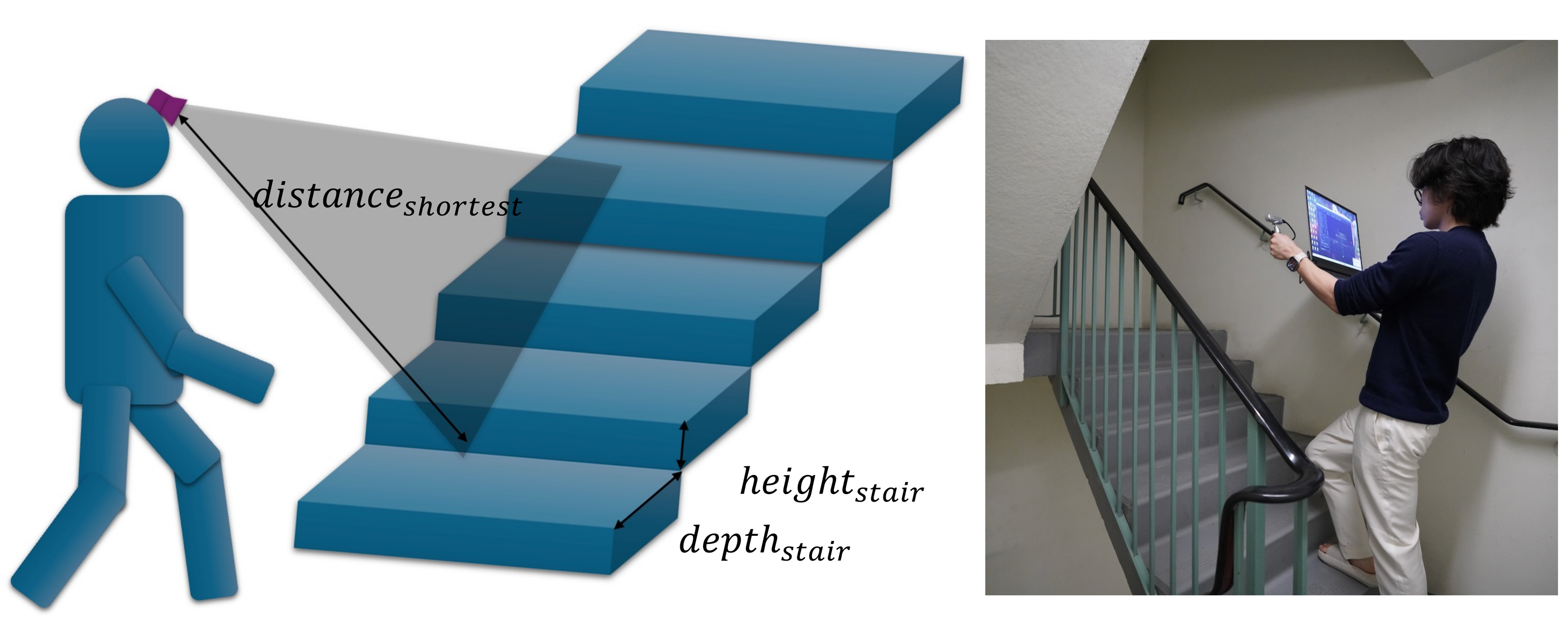


Figure 1: Experimental design for sensing 3D walking environments.

Methods

We used the ZoeDepth deep neural network for our baseline, finetuned on both the NYU-V2 video dataset and our custom-built dataset. Our custom video dataset was collected using an Intel RealSense D415 depth camera, focusing on stair environments for human-robot locomotion.

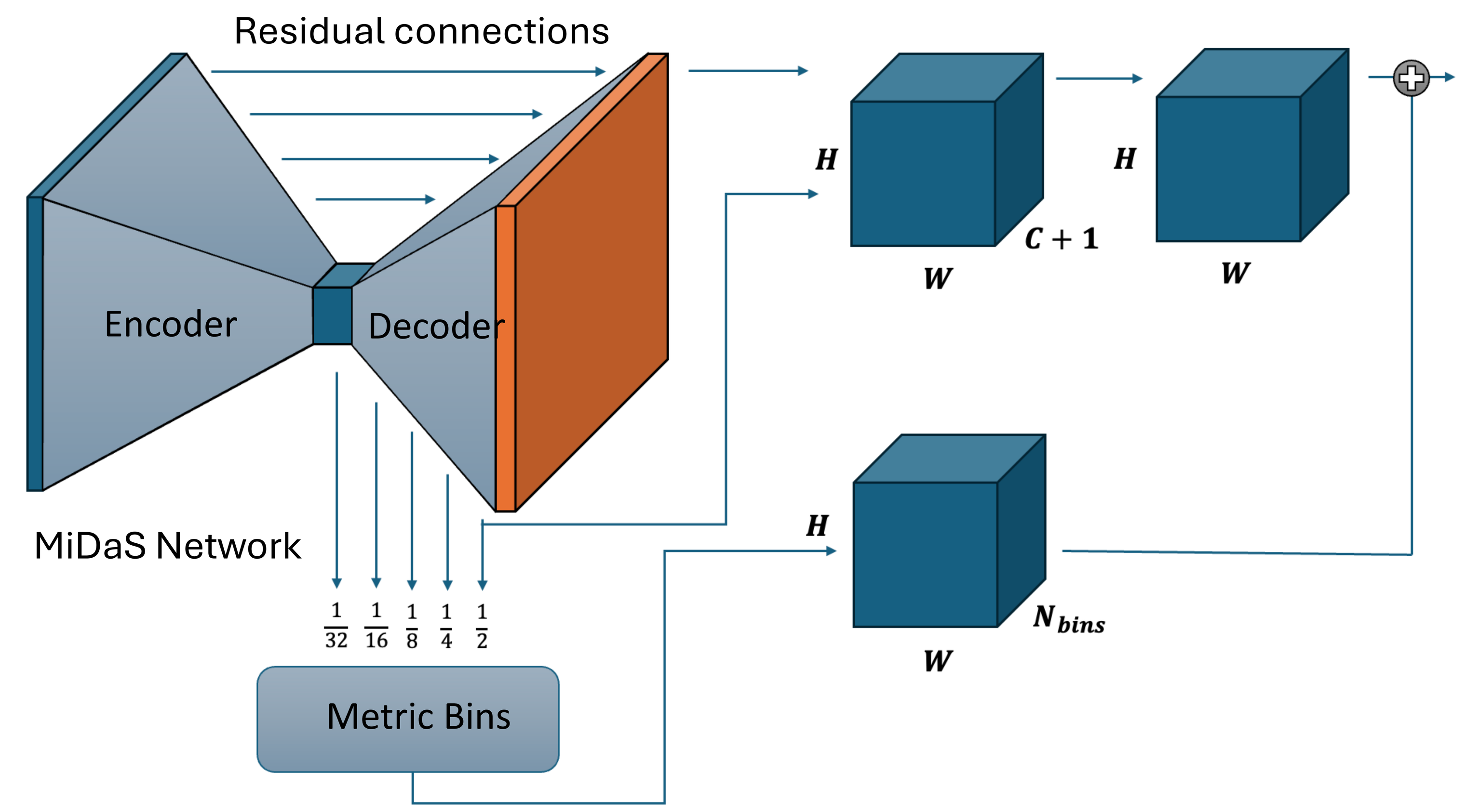


Figure 2: System overview of ZoeDepth deep neural network architecture.

Once we calculate the metric depth estimate of the images, we obtain the 3D point cloud using the intrinsic matrix for the Intel depth camera.

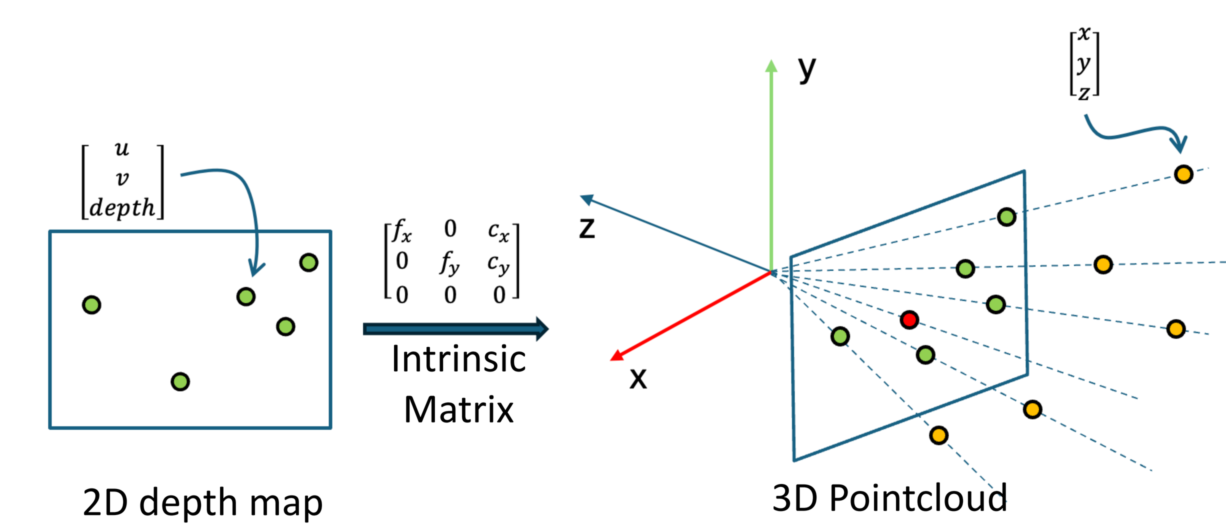


Figure 3: Conversion of the 2D metric depth images to 3D point clouds.

We explored different methods for 3D point cloud processing to extract the stair planes from the environment. Iterative RANSAC was used to extract the staircase planes and calculate the stair height and depth, and the distance to the nearest stair.

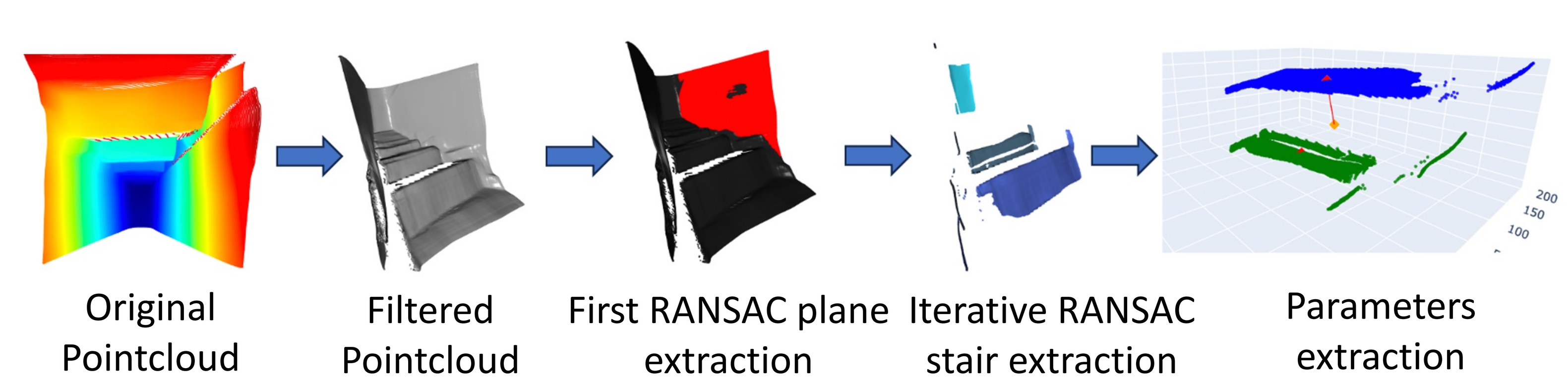


Figure 4: Our pipeline for the 3D point cloud staircase plane extraction.

Result

Our optimized monocular depth model was able to reconstruct the depth of real-world stair environments with high accuracy, despite the ambiguous lighting and surface conditions.

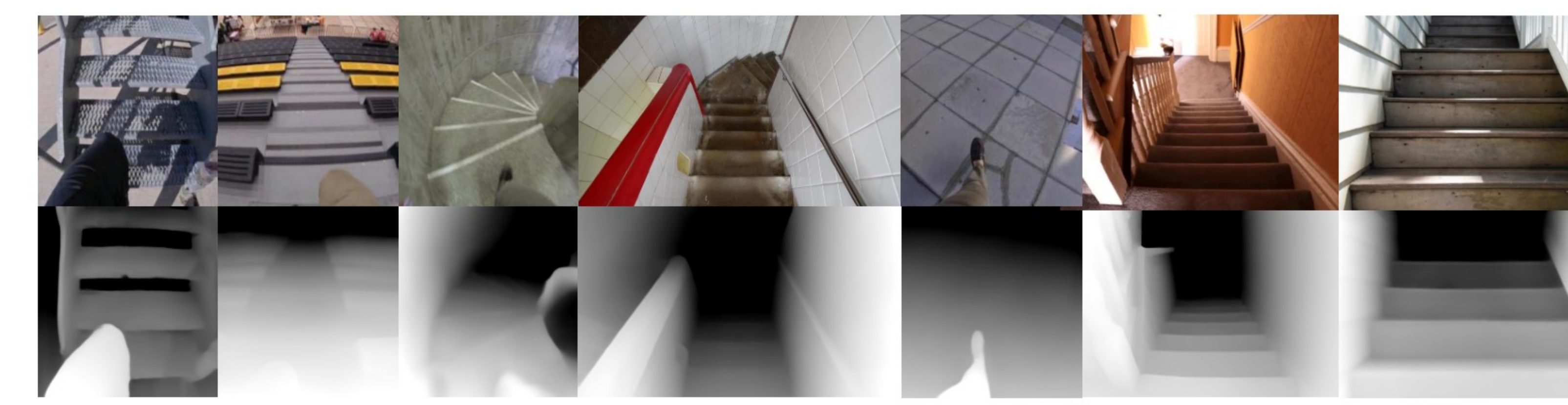
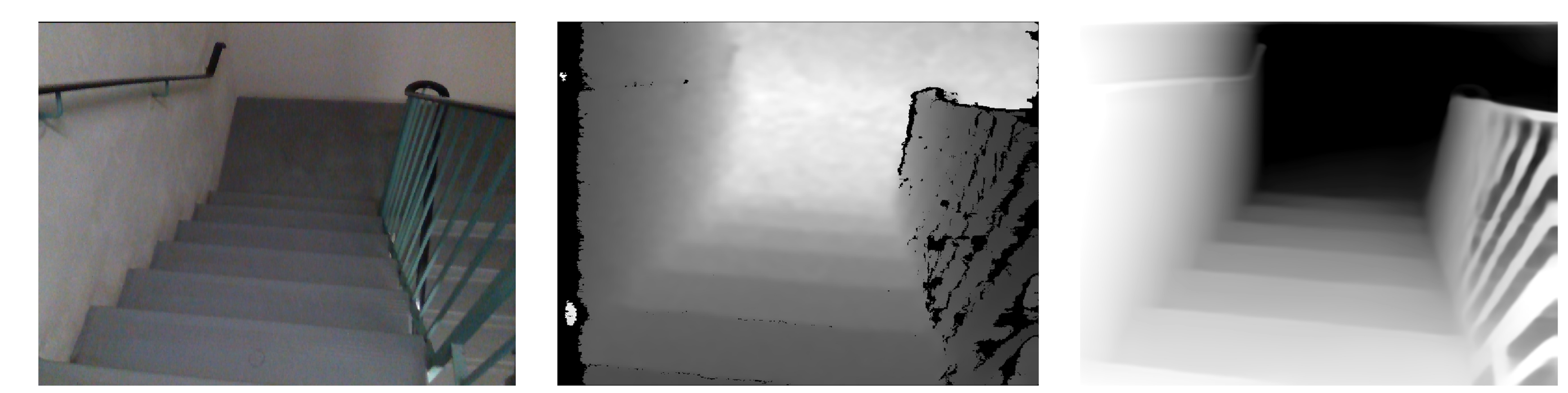


Figure 5: Examples of our depth estimation results. Note that these images are examples of ones that the previous state-of-the-art failed to classify, while we successfully predicted the depth.



Original RGB image of stair environment | Measurement with stereo camera | Estimation with our new monocular depth model

Methods	Stair Depth	Stair Height
Ground truth using April tags	261 mm	199 mm
Measurements using stereo camera	270 mm	190 mm
Estimates using our monocular depth estimation model	289 mm	204 mm

Table 1: Comparison between the difference depth perception methods.

Discussion

In this study, we developed a novel monocular depth estimation model for human-robot walking, which achieved accuracies on par with the state-of-the-art for 3D sensing (i.e., stereo camera). Future research will focus on improving the accuracy for different environments and optimizing the computational efficiency for real-time embedded computing for robot control.

Contact: brokoslaw.laschowski@utoronto.ca

Robotics Institute @KITE_UHN