11

Comparing depth estimation bias using large-scale human data

3D world captured by humans and AI

Abstract

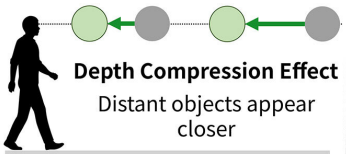
Humans can naturally estimate 3D structures from 2D images, and recent advances in artificial intelligence (AI) have enabled physical devices to develop similar abilities. Our research investigates whether these systems rely on the same visual cues as humans in depth estimation. To this end, we collected large-scale human-annotated data for indoor and outdoor images and compared them with predictions from various AI models. Our results show that many AI models exhibited estimation biases similar to humans (e.g., perceiving distant objects as closer than they physically are). Additionally, we identify an accuracy-similarity trade-off: highly accurate AI models often behave less like humans. By precisely modeling human-like error patterns, our work contributes to the development of AI models that better align with human perception. This may support safer and more intuitive applications, such as remote robot operation, where visual misunderstandings can lead to accidents.

Background

Humans can estimate 3D structure using various cues from the information projected onto their retinas.

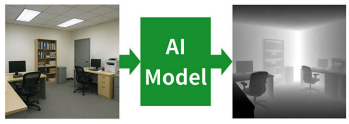
Human Depth Perception

Various estimation errors



AI Depth Perception

Estimation from a 2D image

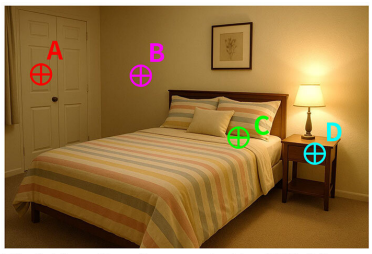


Q. How similar are the error patterns of human and AI depth perception?

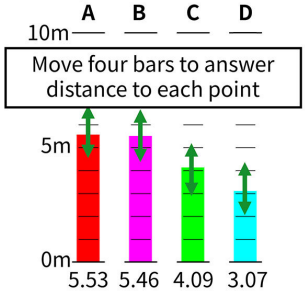
Why important? Human-centered information presentation by precisely predicting human error patterns in 3D perception

Human-annotated Data Collection

We collected a large-scale human-annotated dataset in depth perception through online experiments.



*Mimicking stimuli generated by GPT-4.0



■Procedure

- Participants answered perceived distances in meters.
- We collected eight or sixteen responses per image.

■Stimuli Indoor Images / NYU Depth V2 dataset
Outdoor Images / KITTI dataset

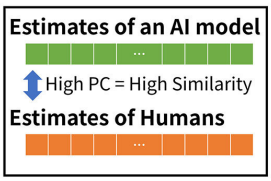
Error Comparison of Humans and AIs

■AI Models: 69 pre-trained AI models

- We employed and trained 69 pre-trained AI models for physical depth estimation based on prior work.
- The models vary in training strategies, datasets, and architectures.

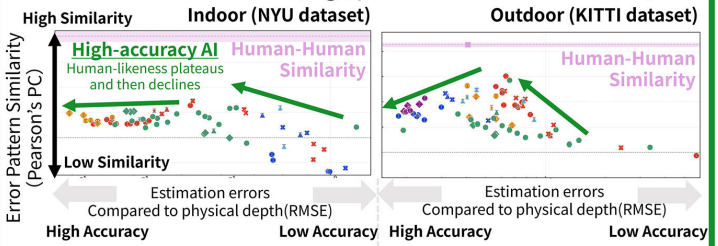
■Metrics: Partial correlation (PC)

- We removed the effect of physical depth values and compared the error patterns between humans and AI models.



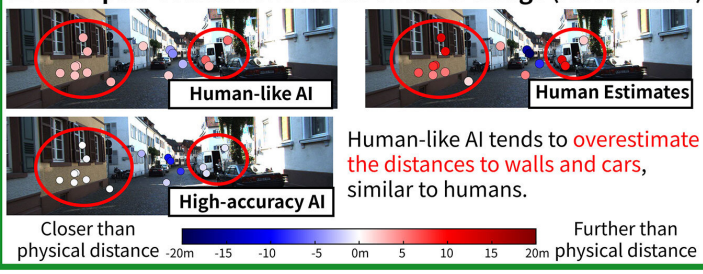
Model Accuracy and "Human-likeness"

- Many AI models exhibited estimation biases similar to those of humans. (e.g., depth compression effect)
- We found the trade-off between accuracy and human-likeness: highly accurate AI models exhibit tendencies different from humans. (e.g., positions of walls and cars)



*Each point represents an AI model, with estimation accuracy on the x-axis and similarity to human error patterns on the y-axis. (Marker color: Training data, Marker shape: Training strategy)

■ Examples estimations for an outdoor image (KITTI dataset)



Human-like AI tends to overestimate the distances to walls and cars, similar to humans.

References

[1] Y. Kubota, T. Fukiage, "Human-like monocular depth biases in deep neural networks," *PLOS Computational Biology*, Vol. 21, No. 8, e1013020, 2025.
[2] Y. Kubota, T. Fukiage, "Accuracy does not guarantee human-likeness in monocular depth estimators ," *arXiv*, 2512.08163, 2025.
[3] Y. Kubota, T. Fukiage, "Benchmarking human and DNN biases in monocular depth estimation," under review, 2026.

Contact

Yuki Kubota, Sensory Representation Research Group, Human Information Science Laboratory