JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

CVPR2023

GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

Zhengxi Hu[1,2,3] · Yuxue Yang[1] · Xiaolin Zhai[1,2,3]
Dingye Yang[1,2,3] · Bohan Zhou[1] · , Jingtai Liu[1,2,3]*

[1]IRAIS, College of Artificial Intelligence, Nankai University
[2]tjKLIR, Nankai University [3]TBI center, Nankai University

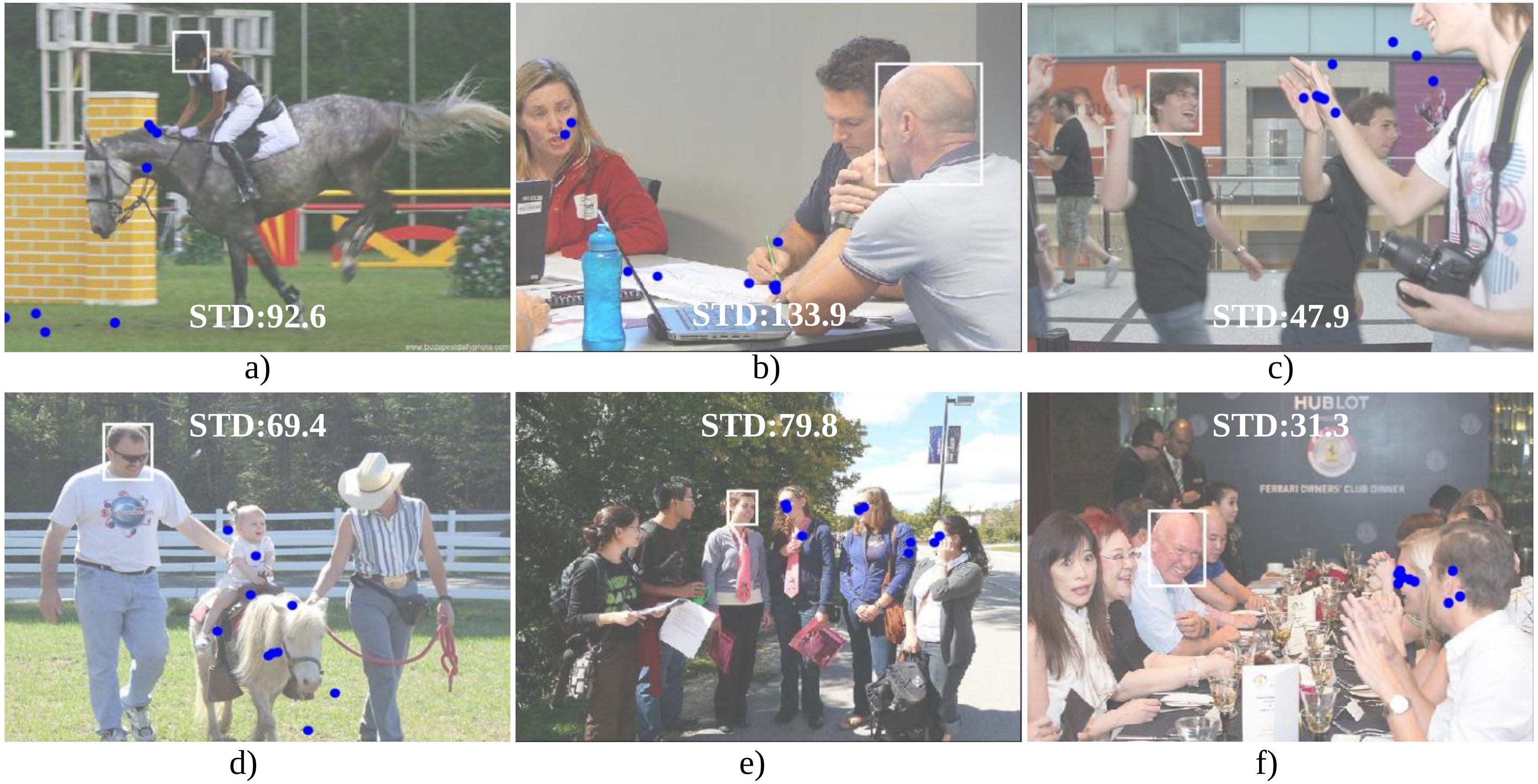
IRAIS

南开大学机器人与信息自动化研究所
Institute of Robotics & Automatic Information System
天津市智能机器人技术重点实验室
Tianjin Key Laboratory of Intelligent Robotics

Paper Tag

WED-AM-065

Preview → Key Problem



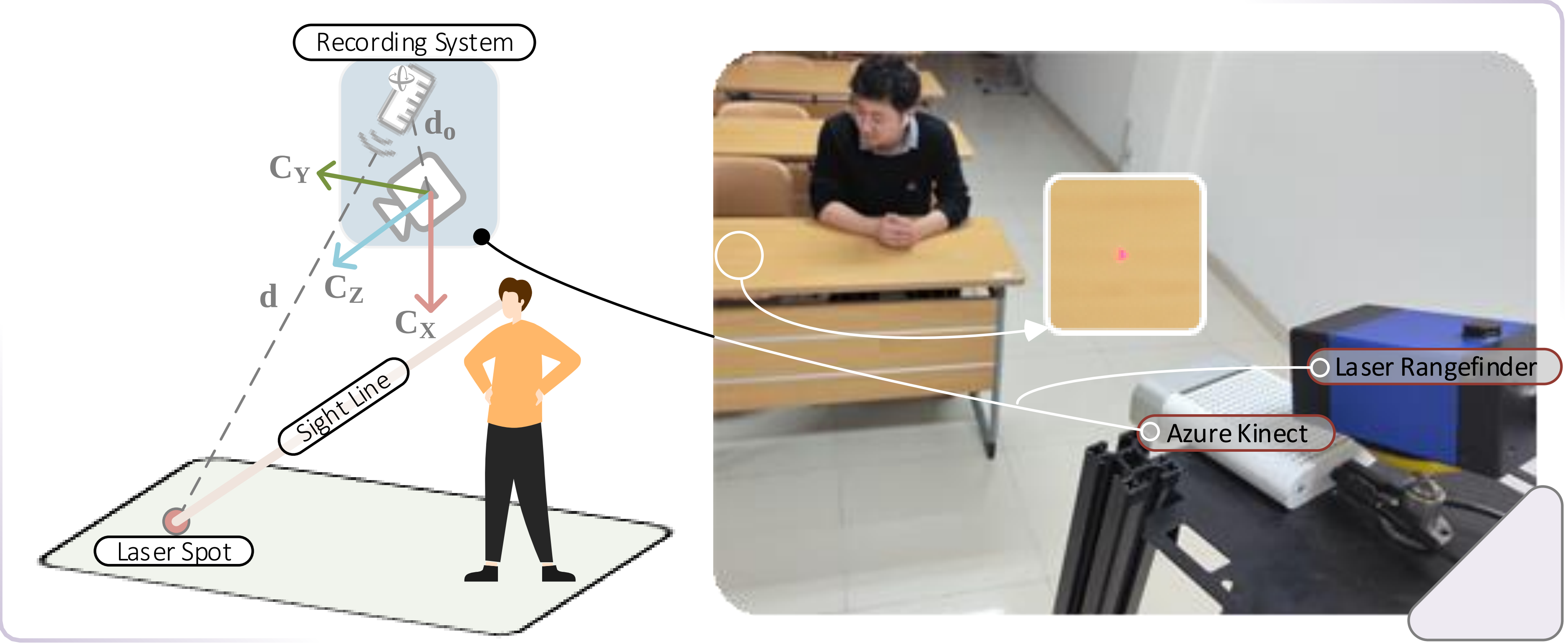
Some samples from the GazeFollow dataset

- How to establish a dataset with reliable gaze annotations?
- How to locate the gaze target accurately when collecting gaze behavior?

Zhengxi Hu et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

Preview → Main Contribution

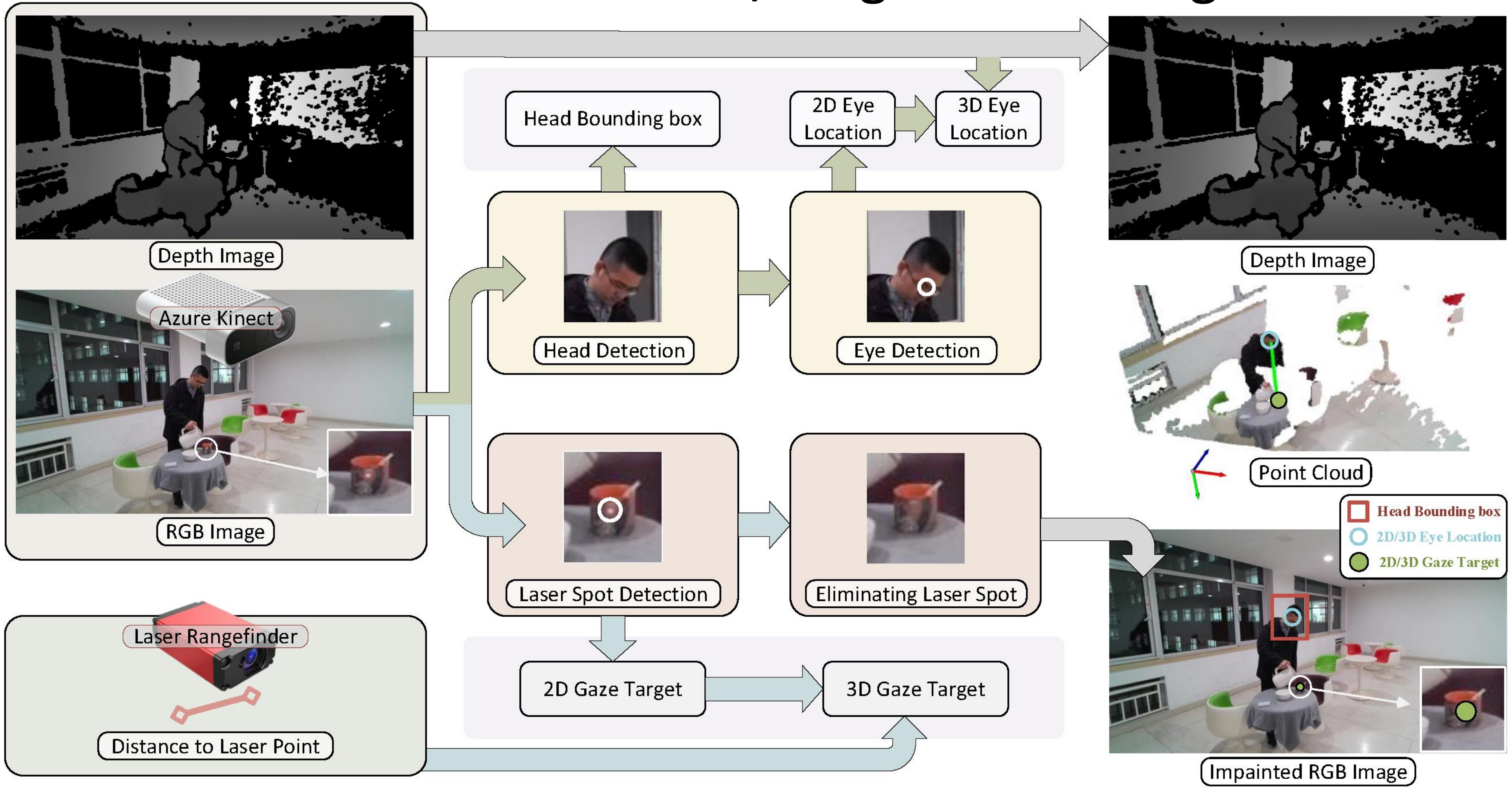
- We develop a system to guide and localize gaze target while recording gaze behavior



Zhengxi Hu et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

Preview → Main Contribution

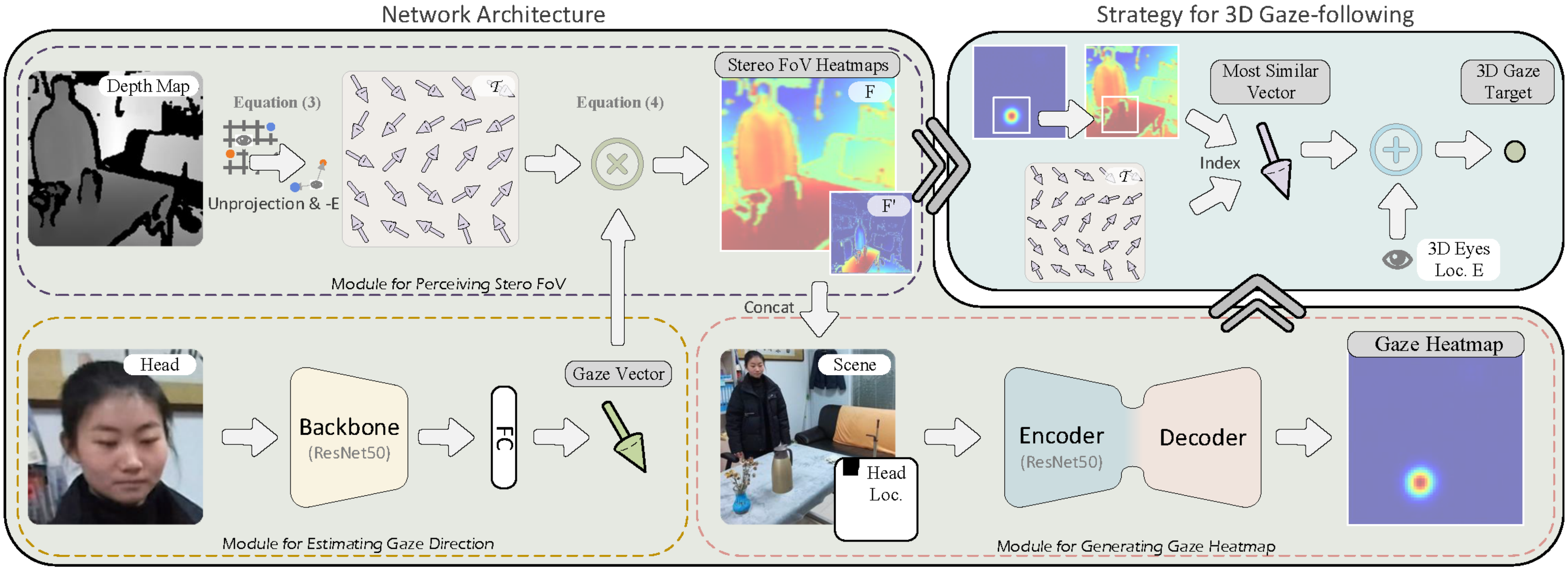
- Release a new GFIE dataset with reliable annotations for 2D/3D gaze-following



Zhengxi Hu et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

Preview → Main Contribution

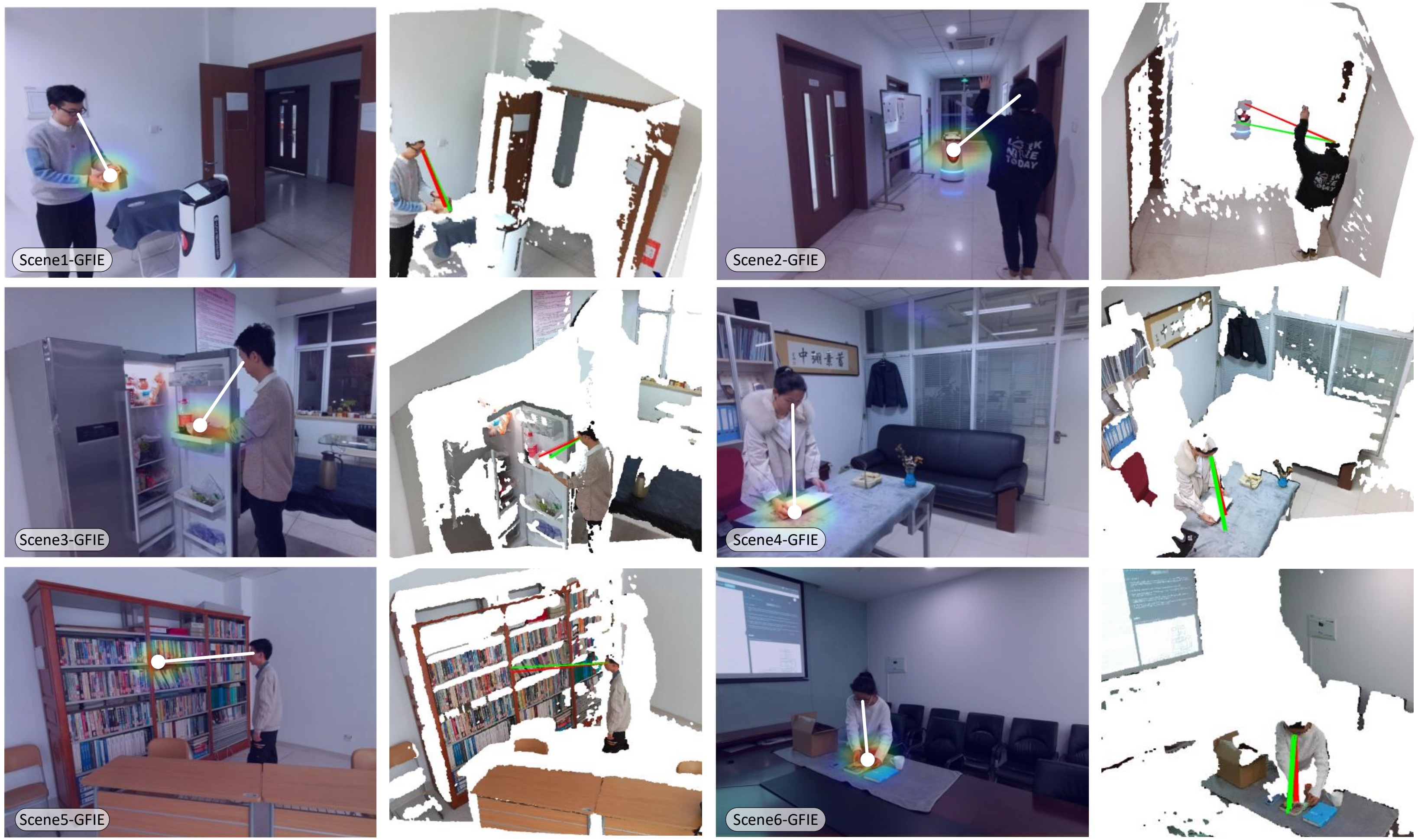
- We introduce a stereo field of view (FoV) in the proposed baseline method for improving gaze-following



Zhengxi Hu et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

Preview → Performance

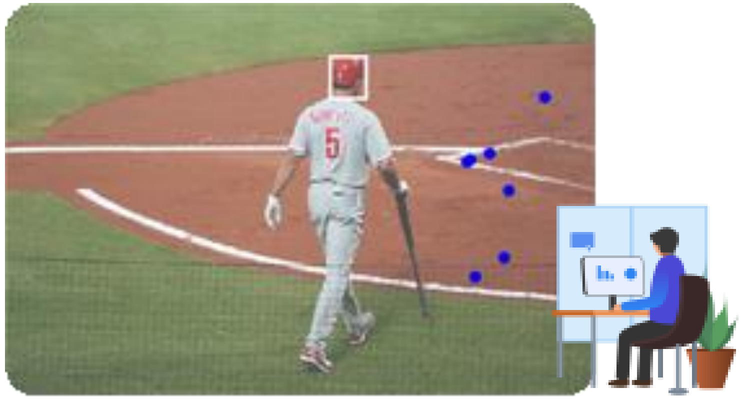
- Test examples on the GFIE dataset



Zhengxi Hu et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

Motivation

- The way of collecting gaze data in the existing dataset



a) Manual annotation

b)Automatic annotation with eye-tracking device

- Weakness
  - ✓ Most datasets are manually annotated, but the subjectivity of annotators may cause annotations to deviate from the actual gaze target. In addition, labor-intensive is another drawback.
  - ✓ The eye-tracking device can capture annotations automatically but alter subjects’ appearance in the dataset, which brings the gap with the gaze-related behavior in the natural environment.

Motivation

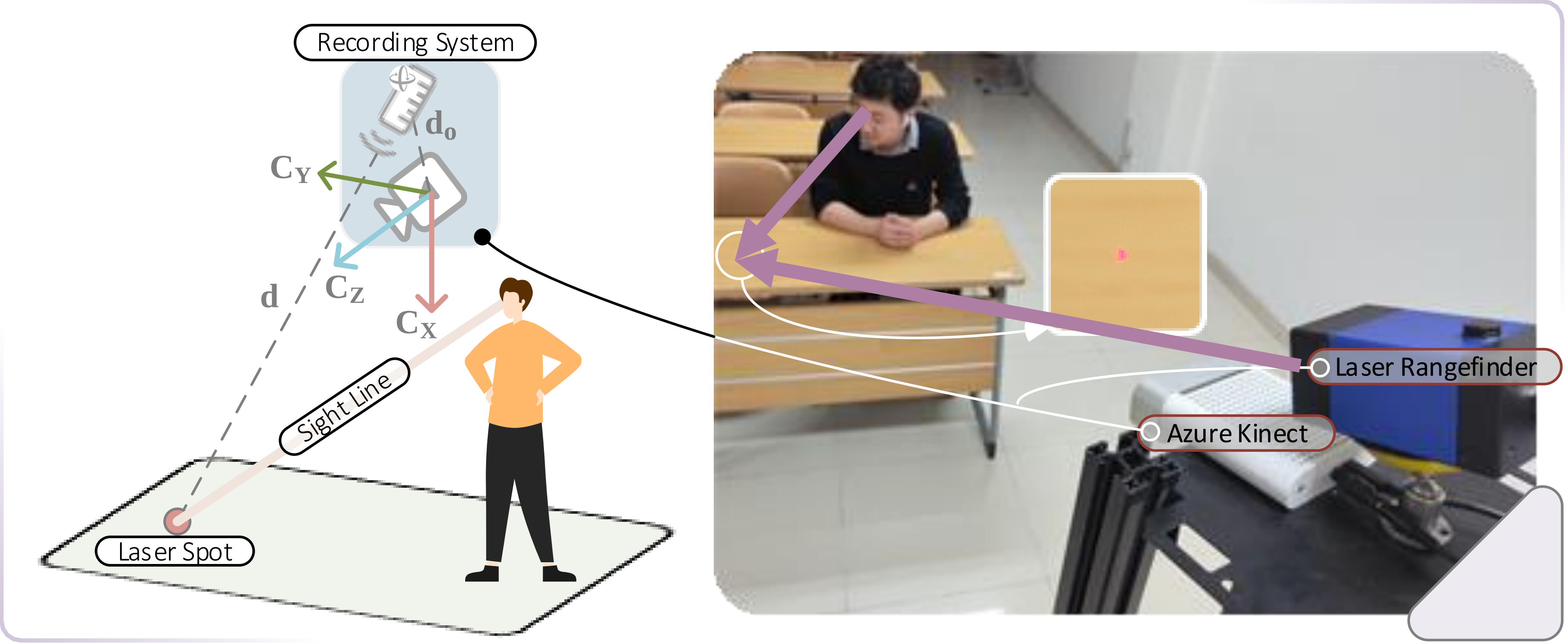
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

- Our main contributions:

✓ We develop a system consisting of a laser rangefinder and RGB-D camera to guide and localize gaze target

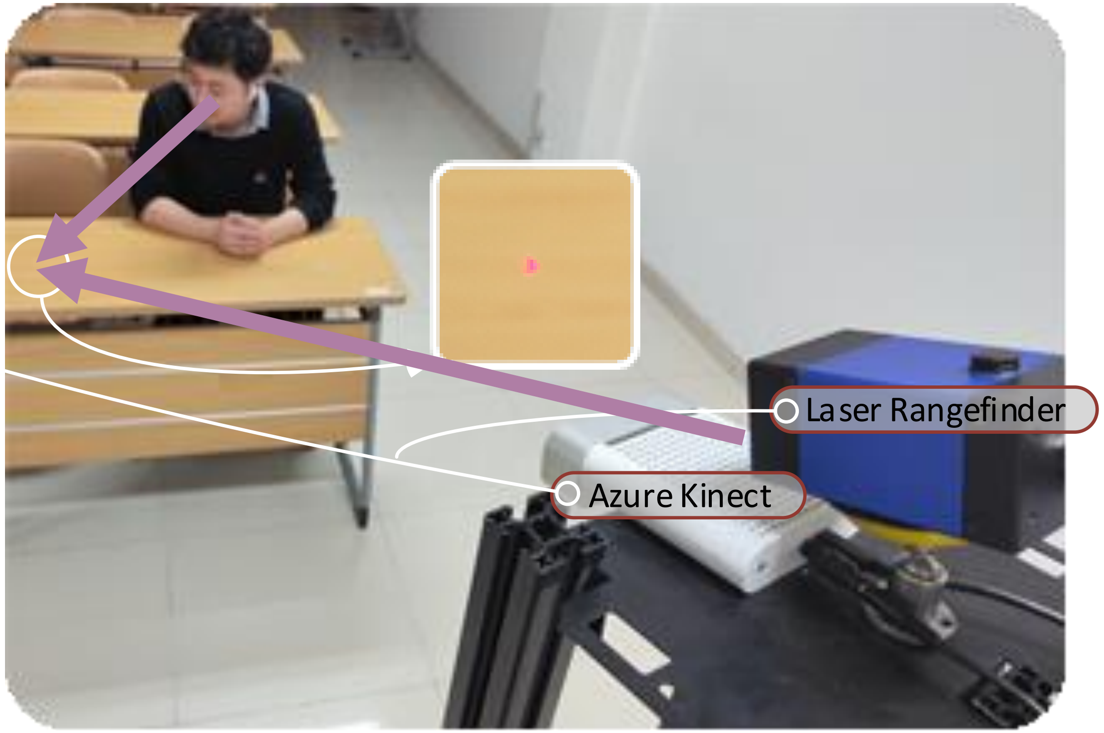
✓ We release a new GFIE dataset for 2D/3D gaze-following that contains reliable annotations and diverse human activities in indoor environments.

✓ We introduce a stereo field of view (FoV) in the proposed baseline method for improving gaze-following.



Zhengxi Hu et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

Workflow for GFIE Dataset Generation



- System setup:

✓ Azure Kinect is set to capture RGB images and depth images with a resolution of 1920 × 1080

✓ Laser Rangefinder are used to measure distance while emitting laser light.

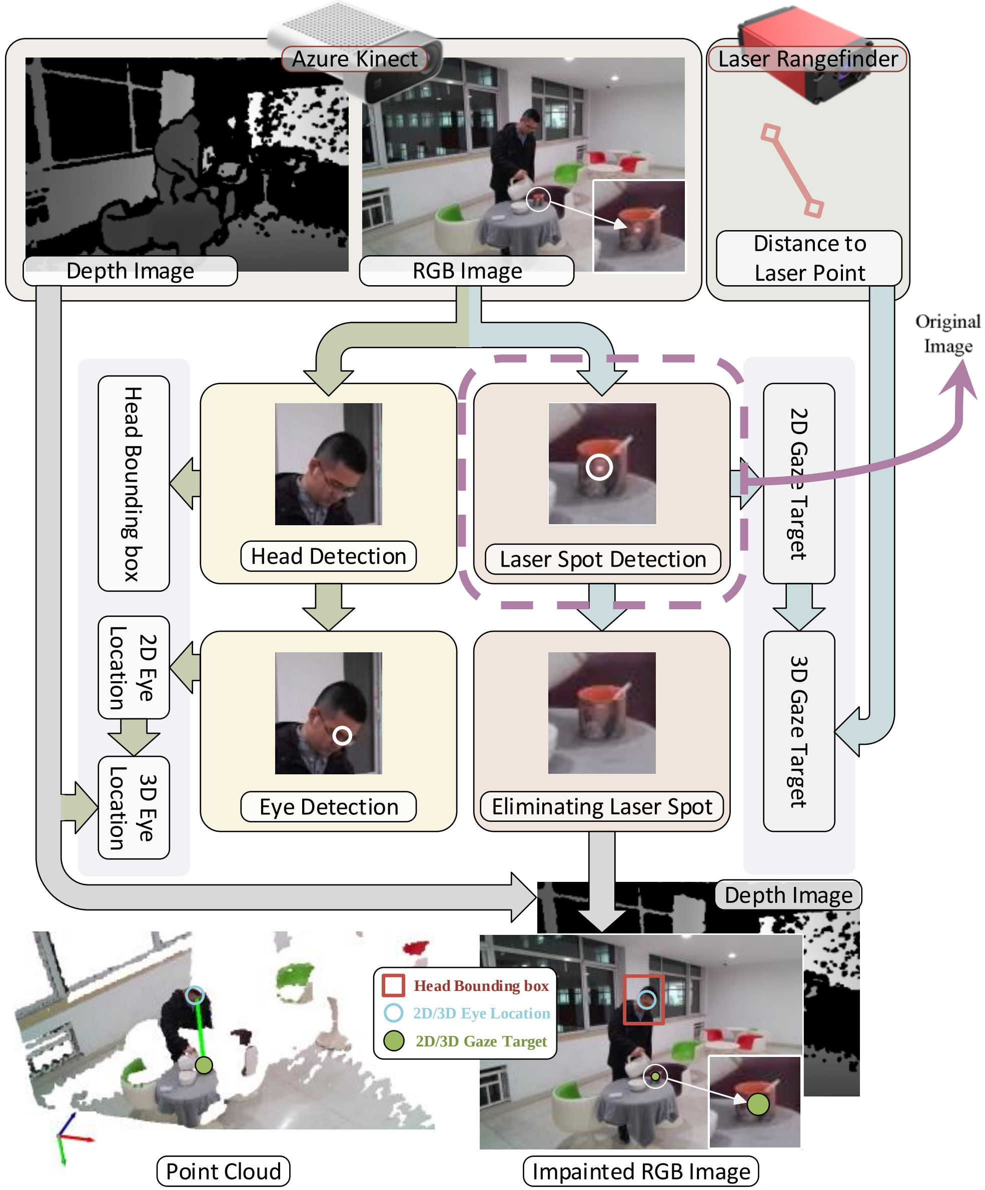
- How to record the gaze behavior:

✓ We operate the laser rangefinder to guide the subject’s attention target through the laser spot.

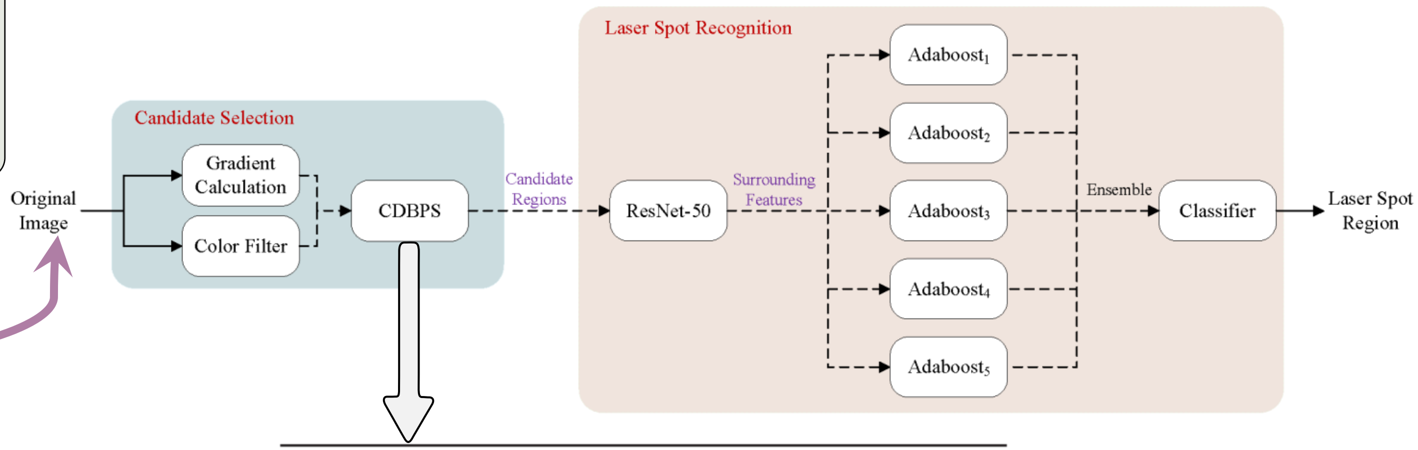
✓ The subject is always staring at the laser spot while performing in front of the camera.

Workflow for GFIE Dataset Generation

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA



- Laser Spot Detection:



Algorithm 1 CDBPS

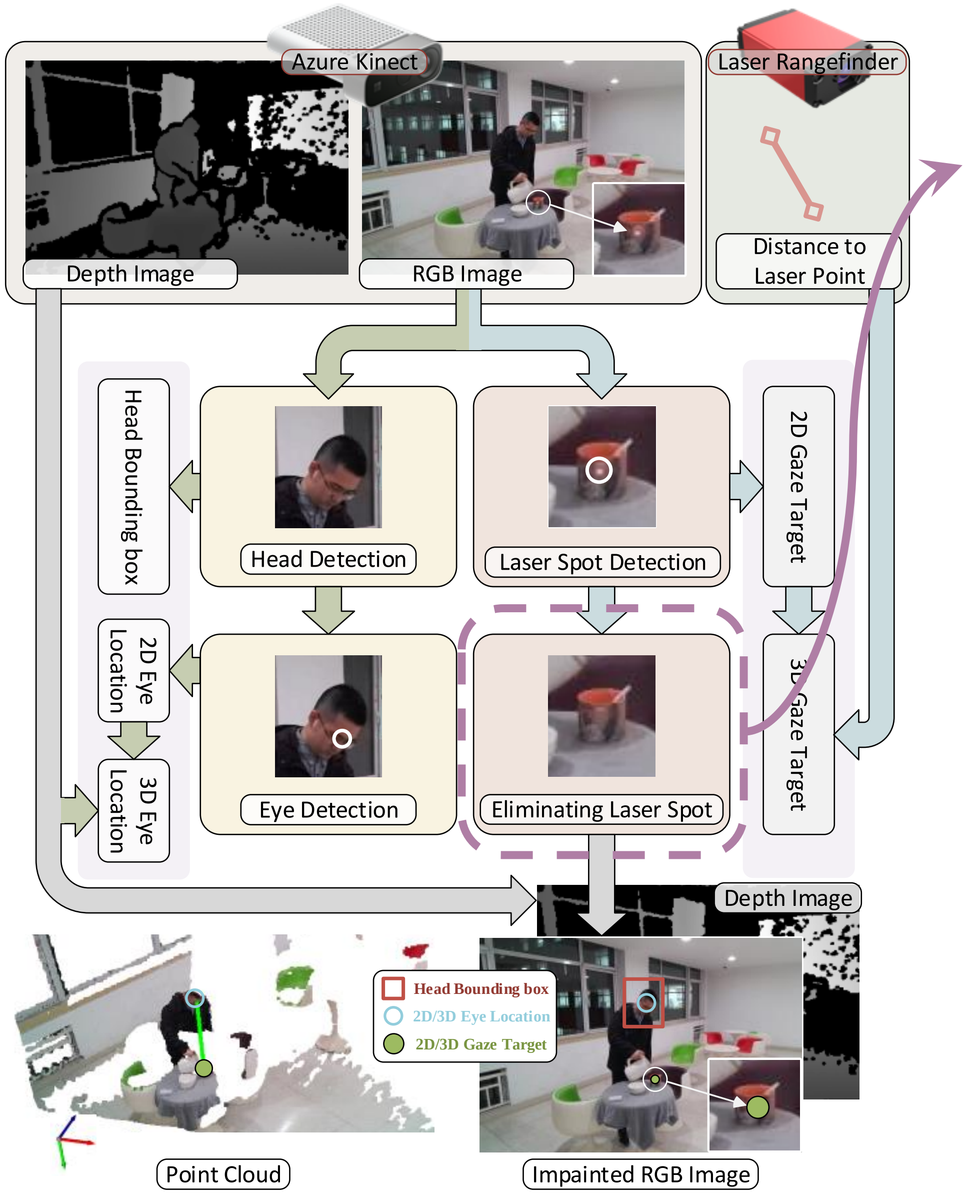
Input: Boundary point sets S, Threshold η, Minimum radius $R_{\min}$, Maximum radius $R_{\max}$

Output: Boundary point sets of candidate regions S_{target}

```
for set $s$ in $S$ do
    $c, r \leftarrow$ find the minimum enclosing circle of set $s$
    if $R_{\min} \leq r \leq R_{\max}$ then
        for point $p_i$ in $s$ do
            $D_i \leftarrow \|p_i - c\|$
        end for
        $v \leftarrow$ compute the variance of $D$
        if $v \leq \eta$ then
            add $s$ in $S_{\text{target}}$
        end if
    end if
end for
```

Zhengxi Hu et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

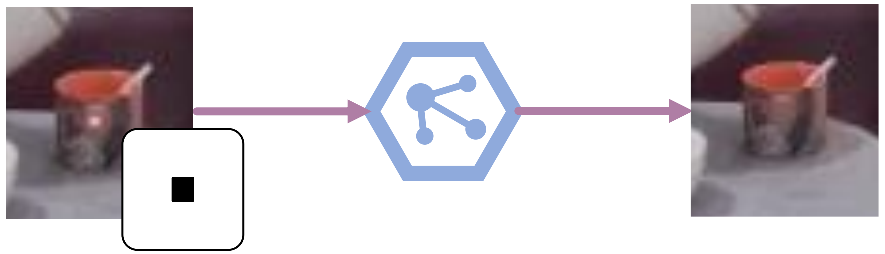
Workflow for GFIE Dataset Generation



- Eliminating Laser Spot:

✓ We adopt the generator network proposed by Ulyanov et al. [1] to inpaint the regions of laser spots in images.

[1] Dmitry Ulyanov et al. Deep image prior. ICCV 2018



- Annotation

2D Gaze Target

Head bounding box

2D/3D eye location

(g_u, g_v)

$\mathcal{K} = (f_u, f_v, c_u, c_v)$

intrinsics of the RGB camera

$$\begin{cases} \frac{(g_u - c_u)\, g_z}{f_u} = g_x \\ \frac{(g_v - c_v) g_z}{f_v} = g_y \\ \sqrt{g_x^2 + g_y^2 + g_z^2} = d - d_o \end{cases}$$

measured distance

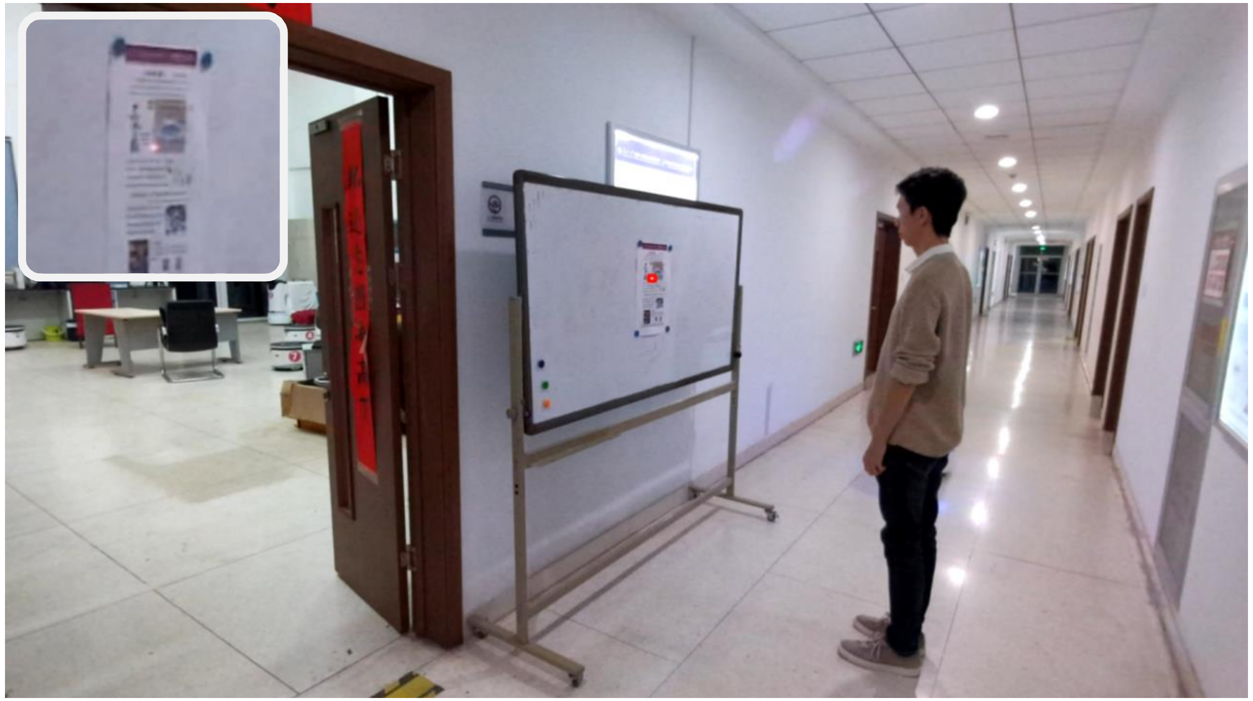
3D Gaze Target

Solve

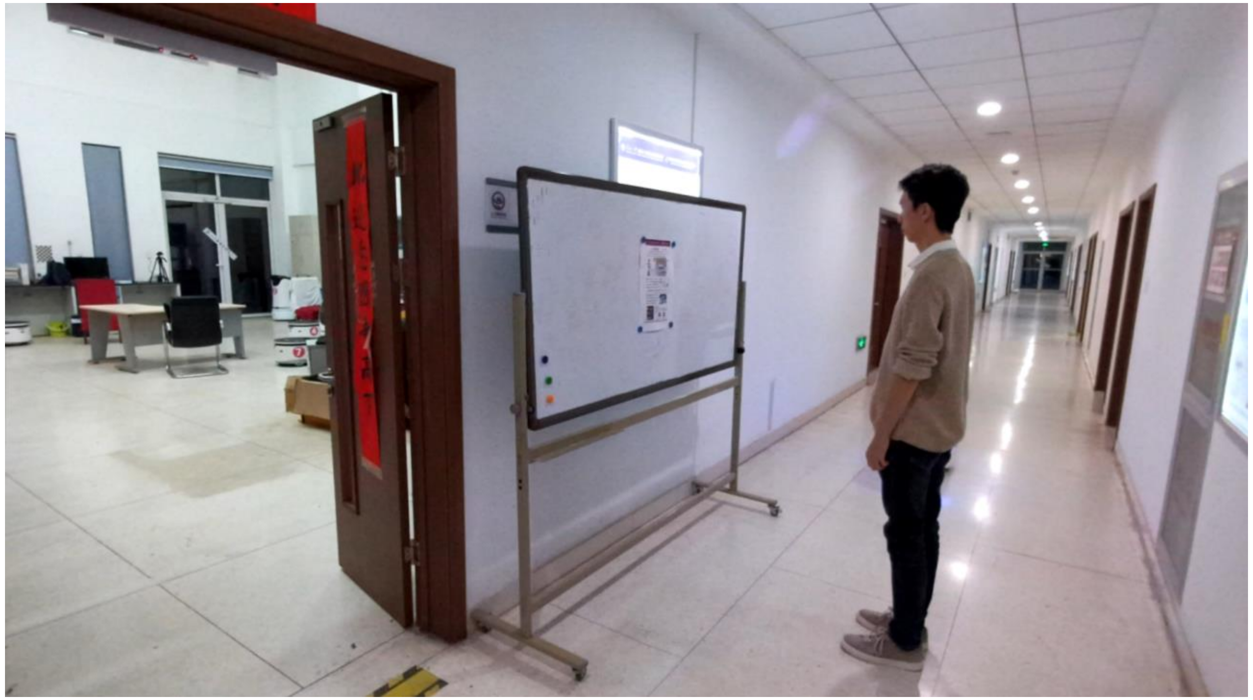
Zhengxi Hu et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

Workflow for GFIE Dataset Generation

- Demo of Laser Spot Detection

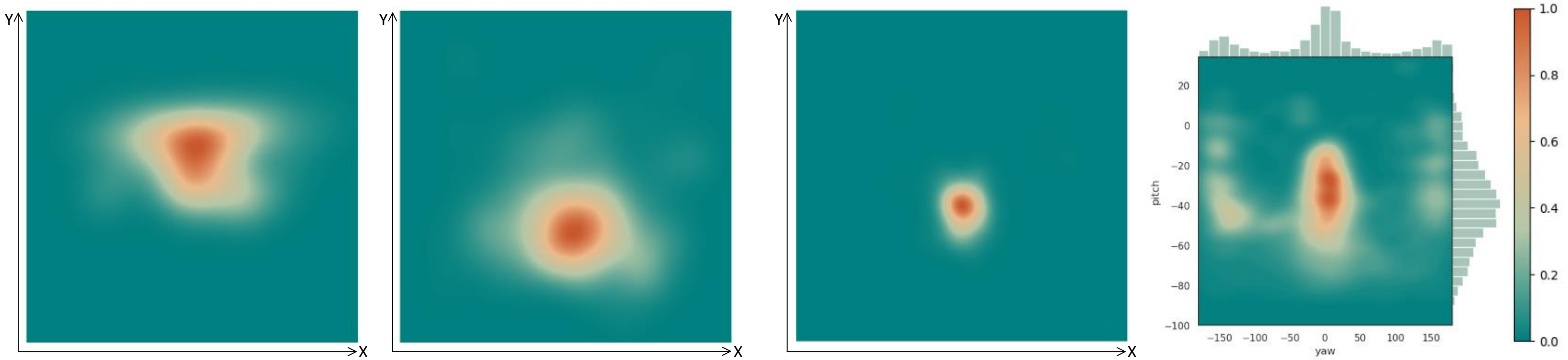


- Demo of Eliminating Laser Spot



GFIE Dataset Statics

- Annotation distribution on 2D plane and Gaze angle density



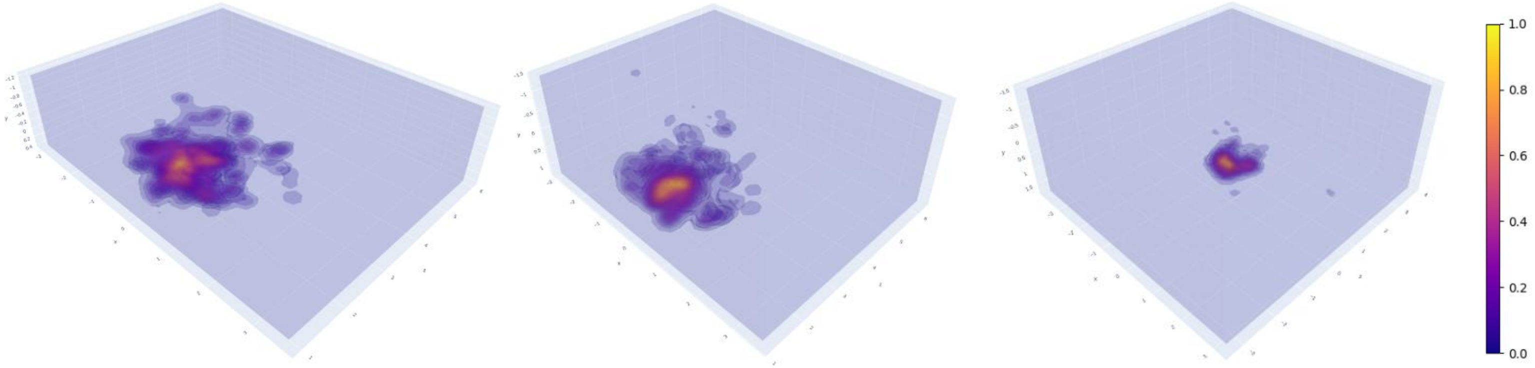
a) 2D Head Location Density

b) 2D Gaze Target Loaction Density

c) 2D Gaze Target Wrt. Head Location Density

d) Gaze Angle Density

- Annotation distribution in 3D space



e) 3D Head Location Density

f) 3D Gaze Target Loaction Density

g) 3D Gaze Target Wrt. Head Location Density

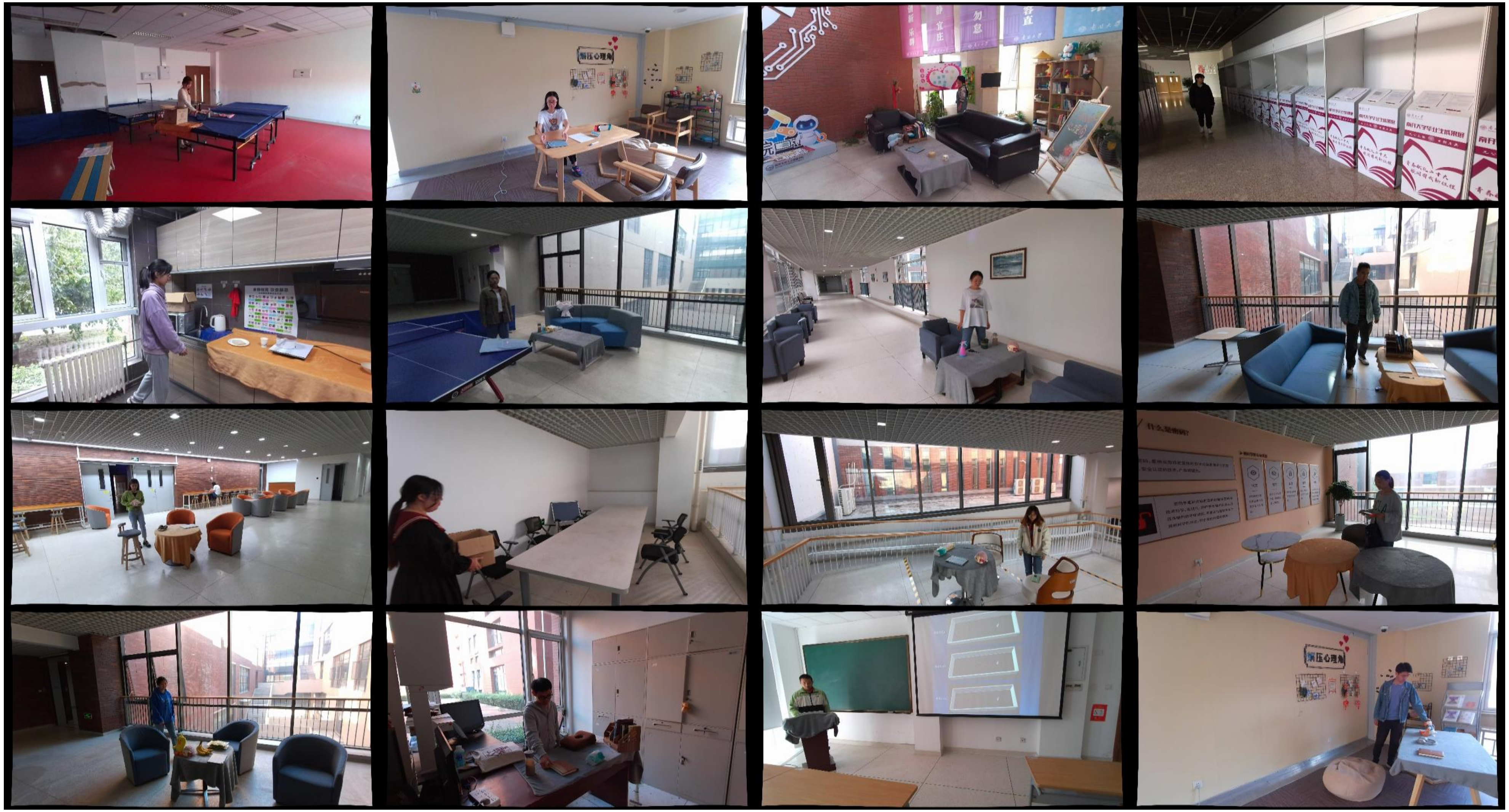
- Quantitative statistics

Modality	RGB/Depth
Frames	71,799
Subjects	61 (27 male and 34 female)
Dataset splits	Train set : 59,217
	Validation set: 6,281
	Test set: 6,281

Zhengxi Hu et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

GFIE Dataset Statics

- An overview of the GFIE dataset (part of all scenes)

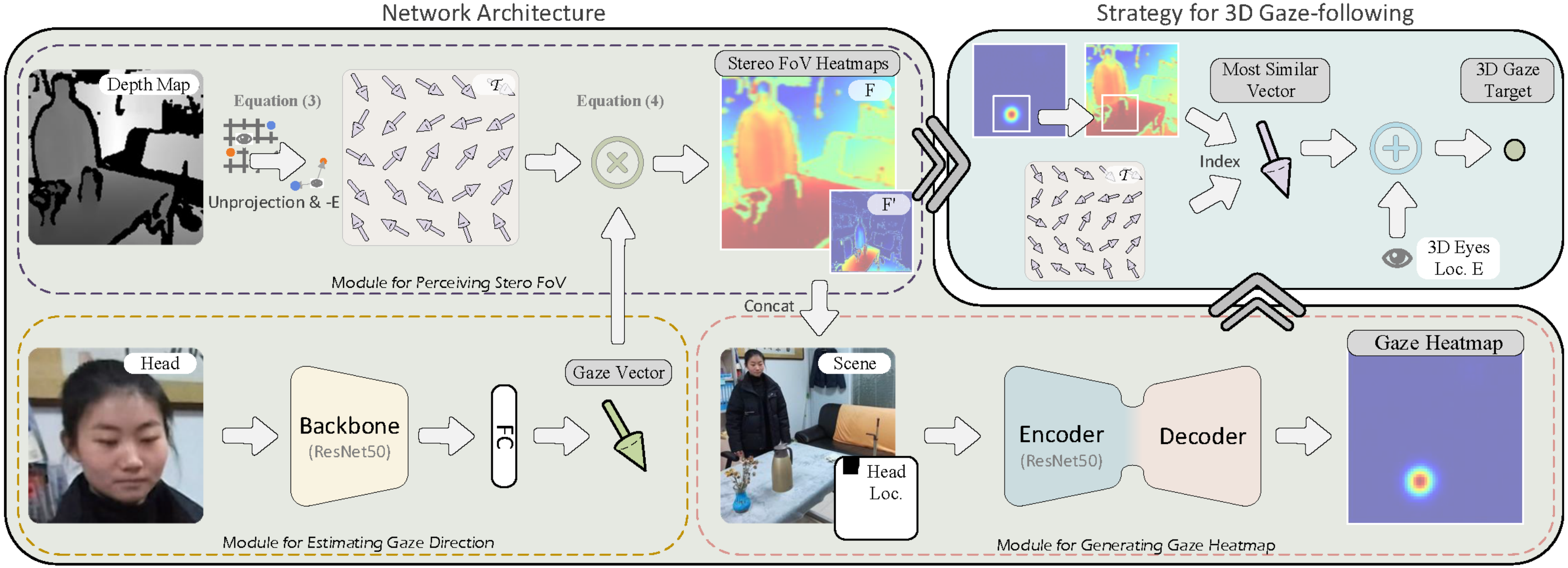


Zhengxi Hu et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

GFIE Baseline

- Three key components:
  - ✓ Module for estimating gaze direction
  - ✓ Module for perceiving stereo FoV (field of view)
  - ✓ Module for generating a gaze heatmap

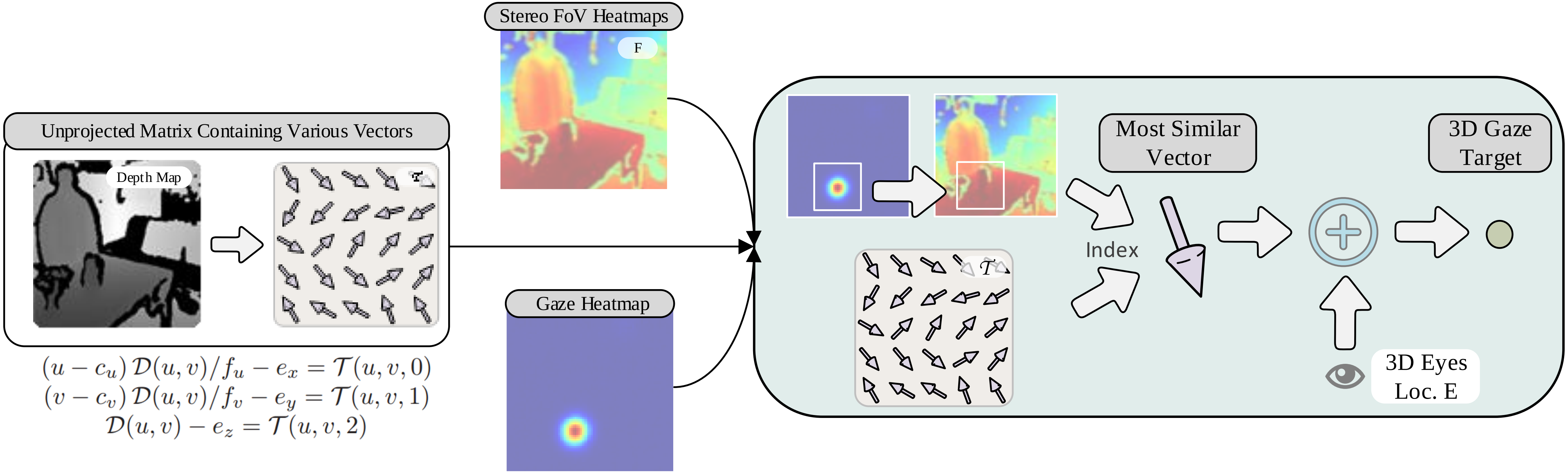
FoV is defined as the extend to which a person can observe in 3D space.



Zhengxi Hu et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

GFIE Baseline

- Strategy for 3D Gaze-following
  - ✓ Based on the 2D gaze target and the stereo FoV heatmap, the reliable 3D gaze vector is selected from the matrix containing candidate vectors, and then the 3D gaze target can be obtained.

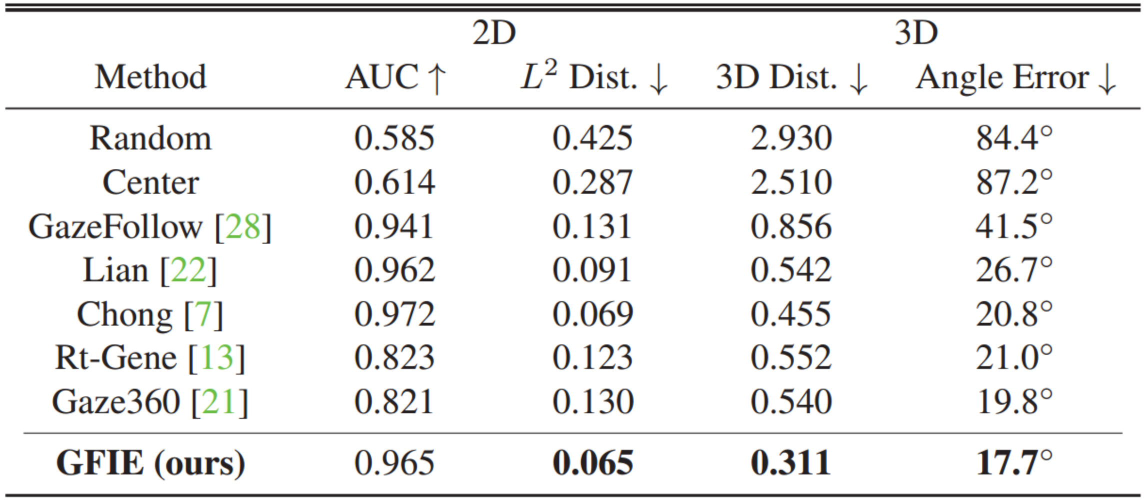


$$(u - c_u)\,\mathcal{D}(u,v)/f_u - e_x = \mathcal{T}(u,v,0)$$
$$(v - c_v)\,\mathcal{D}(u,v)/f_v - e_y = \mathcal{T}(u,v,1)$$
$$\mathcal{D}(u,v) - e_z = \mathcal{T}(u,v,2)$$

Experiment

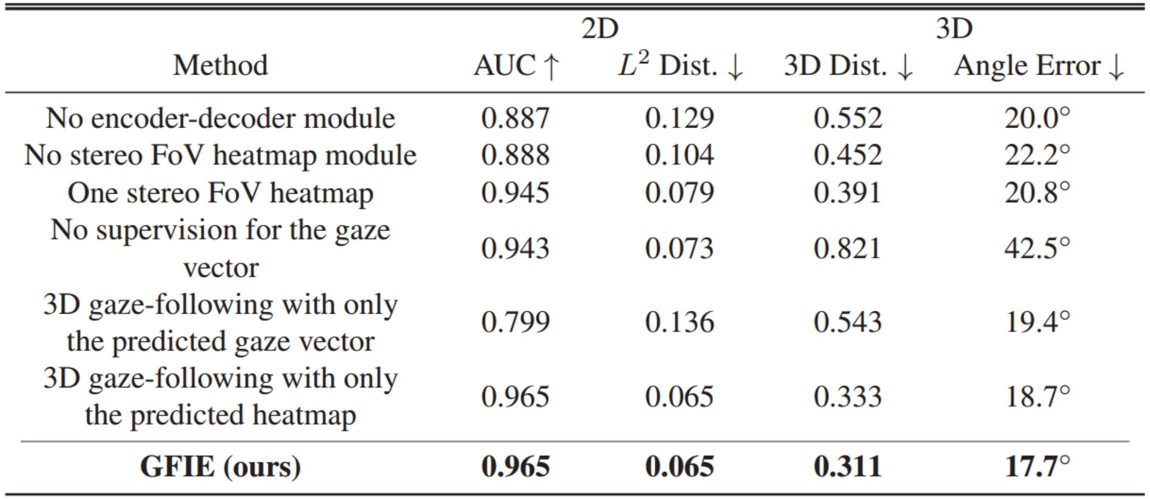
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

- Performance comparison on the GFIE dataset



	2D		3D	
Method	AUC ↑	L^2 Dist. ↓	3D Dist. ↓	Angle Error ↓
Random	0.585	0.425	2.930	84.4°
Center	0.614	0.287	2.510	87.2°
GazeFollow [28]	0.941	0.131	0.856	41.5°
Lian [22]	0.962	0.091	0.542	26.7°
Chong [7]	0.972	0.069	0.455	20.8°
Rt-Gene [13]	0.823	0.123	0.552	21.0°
Gaze360 [21]	0.821	0.130	0.540	19.8°
GFIE (ours)	0.965	**0.065**	**0.311**	**17.7°**

- Quantitative results of ablation study on the GFIE dataset



	2D		3D	
Method	AUC ↑	L^2 Dist. ↓	3D Dist. ↓	Angle Error ↓
No encoder-decoder module	0.887	0.129	0.552	20.0°
No stereo FoV heatmap module	0.888	0.104	0.452	22.2°
One stereo FoV heatmap	0.945	0.079	0.391	20.8°
No supervision for the gaze vector	0.943	0.073	0.821	42.5°
3D gaze-following with only the predicted gaze vector	0.799	0.136	0.543	19.4°
3D gaze-following with only the predicted heatmap	0.965	0.065	0.333	18.7°
GFIE (ours)	**0.965**	**0.065**	**0.311**	**17.7°**

- 2D evaluation metrics:

✓ AUC: The area under curve proposed by [17] is introduced to use the predicted heatmap as the confidence to draw the ROC curve.

✓ L^2 Dist.: The Euclidean distance between the predicted gaze point and the ground truth, we assume the size of the image is 1 × 1.

- 3D evaluation metrics:

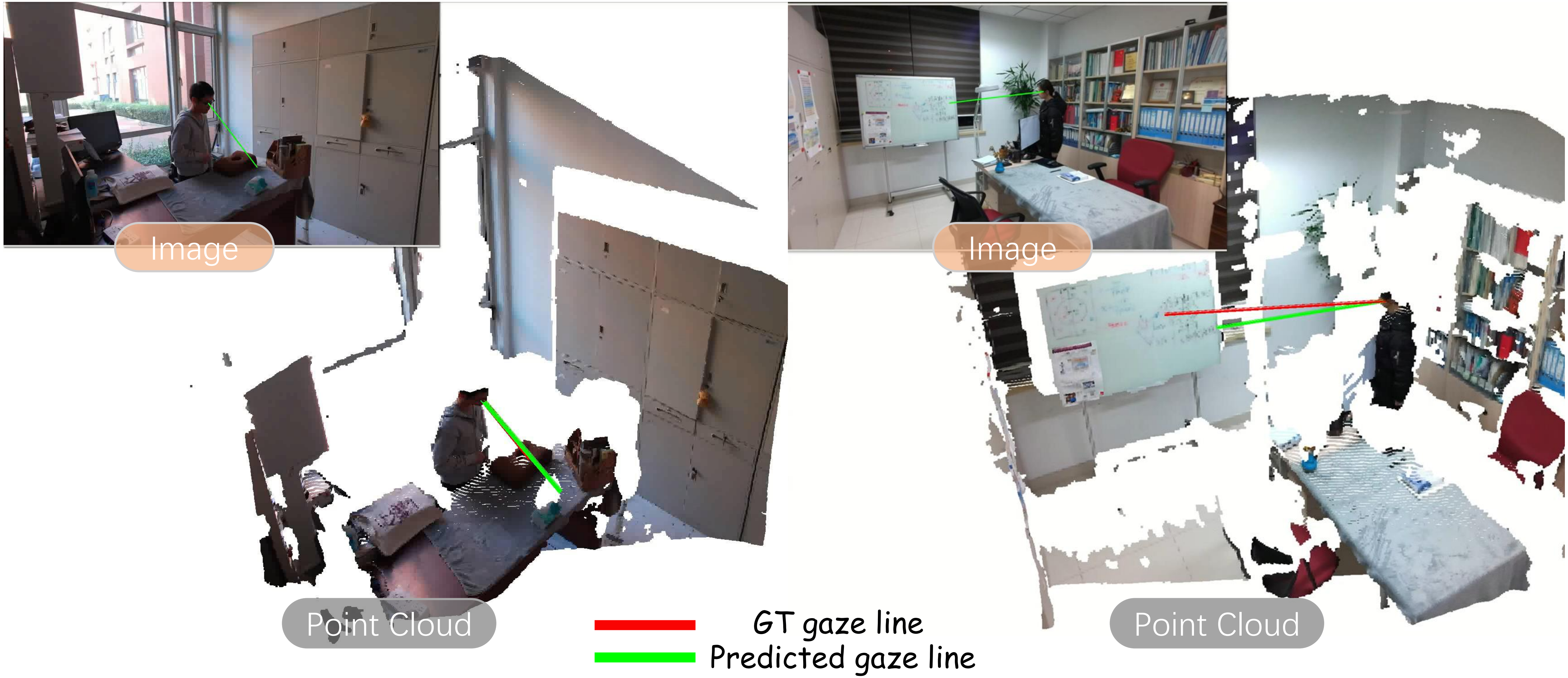
✓ 3D Dist.: Similar to L2 Dist., but for 3D scenes, its unit is m.

✓ Angle Error: The angular difference between predicted gaze direction and ground truth, in degrees.

Metrics

Experiment

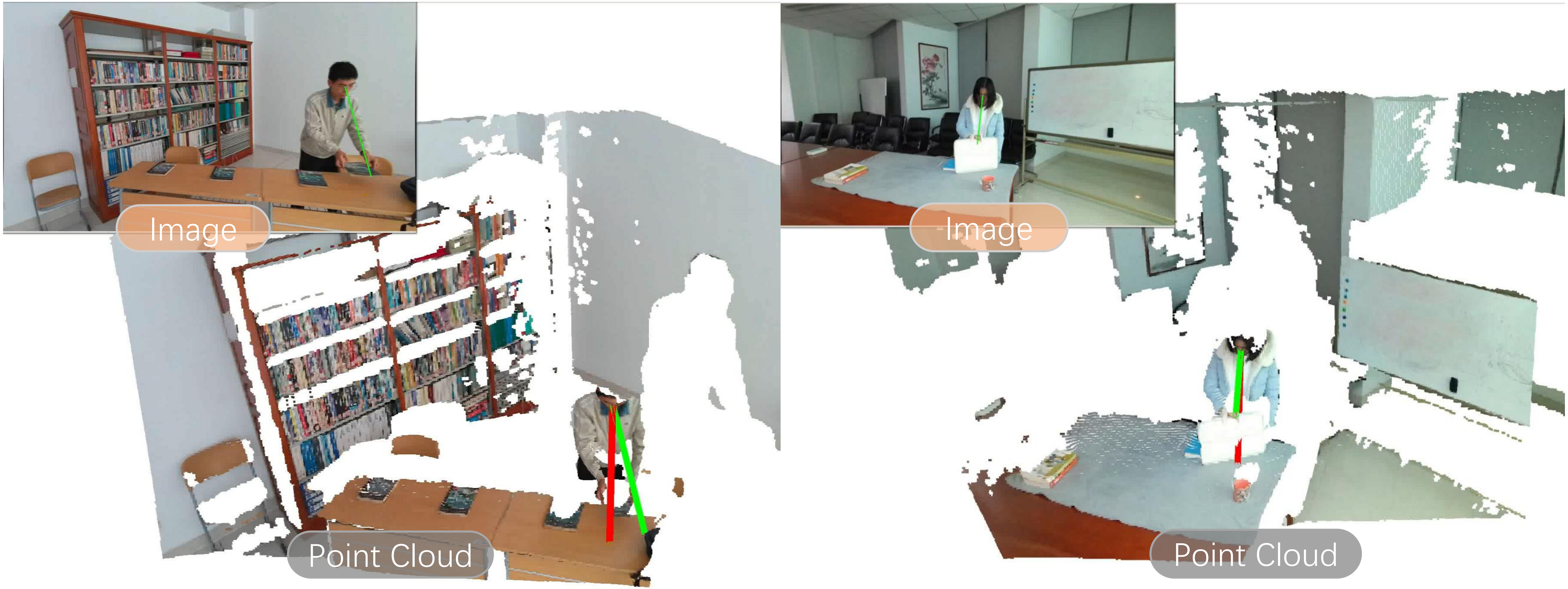
- Performance of our proposed baseline on the GFIE dataset



Zhengxi Hu et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

Experiment

- Performance of our proposed baseline on the GFIE dataset



🟥 GT gaze line

🟩 Predicted gaze line

Zhengxi Hu et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

Experiment

- Cross-dataset evaluation on CAD-120 dataset

✓ Quantitative evaluation on CAD-120 dataset

Method	2D AUC ↑	 L^2 Dist. ↓	3D 3D Dist. ↓	 Angle Error ↓
Random	0.469	0.758	1.910	70.3°
Center	0.456	0.706	1.280	75.9°
GazeFollow [5]	0.862	0.196	1.030	44.1°
Lian [4]	0.871	0.180	0.813	34.8°
Chong [1]	0.891	0.152	0.812	31.9°
Rt-Gene [2]	0.463	0.492	0.483	26.5°
Gaze360 [3]	0.463	0.474	0.427	20.6°
GFIE (ours)	**0.921**	**0.114**	**0.365**	**19.8°**

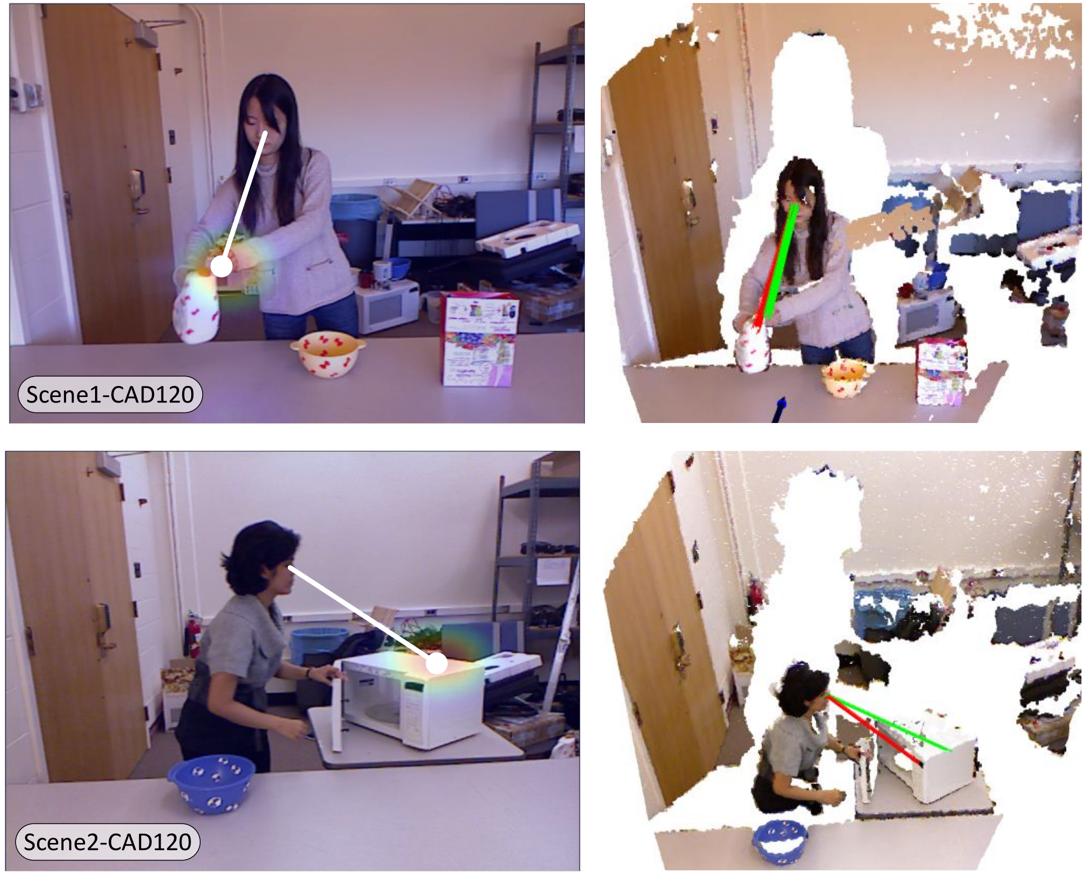
CAD-120 Dataset

✓ The CAD-120 dataset is built for human activity

✓ We selected 1737 frames and asked 3 annotators to annotate the 3D gaze targets manually in the software *CloudCompare*.

✓ The evaluation process performs testing without training

✓ Test examples on CAD-120 dataset



CVPR2023

GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments

Thank you

南开大学机器人与信息自动化研究所
Institute of Robotics & Automatic Information System
天津市智能机器人技术重点实验室
Tianjin Key Laboratory of Intelligent Robotics

IRAIS

Paper Tag

WED-AM-065