

Egocentric Future Localization

Hyun Soo Park, Jyh-Jing Hwang, Yedong Niu, and Jianbo Shi



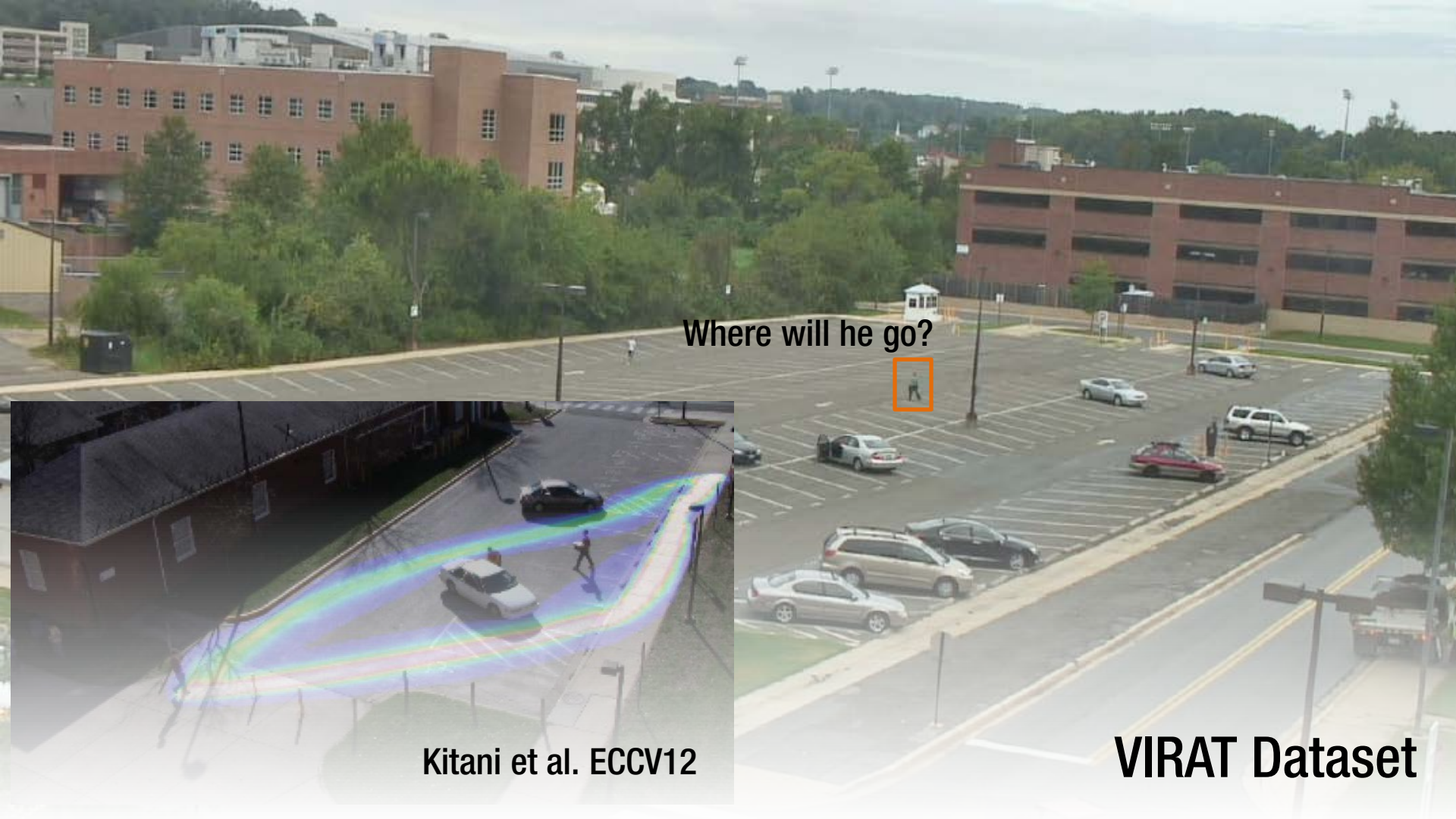


VIRAT Dataset

Where will he go?



VIRAT Dataset

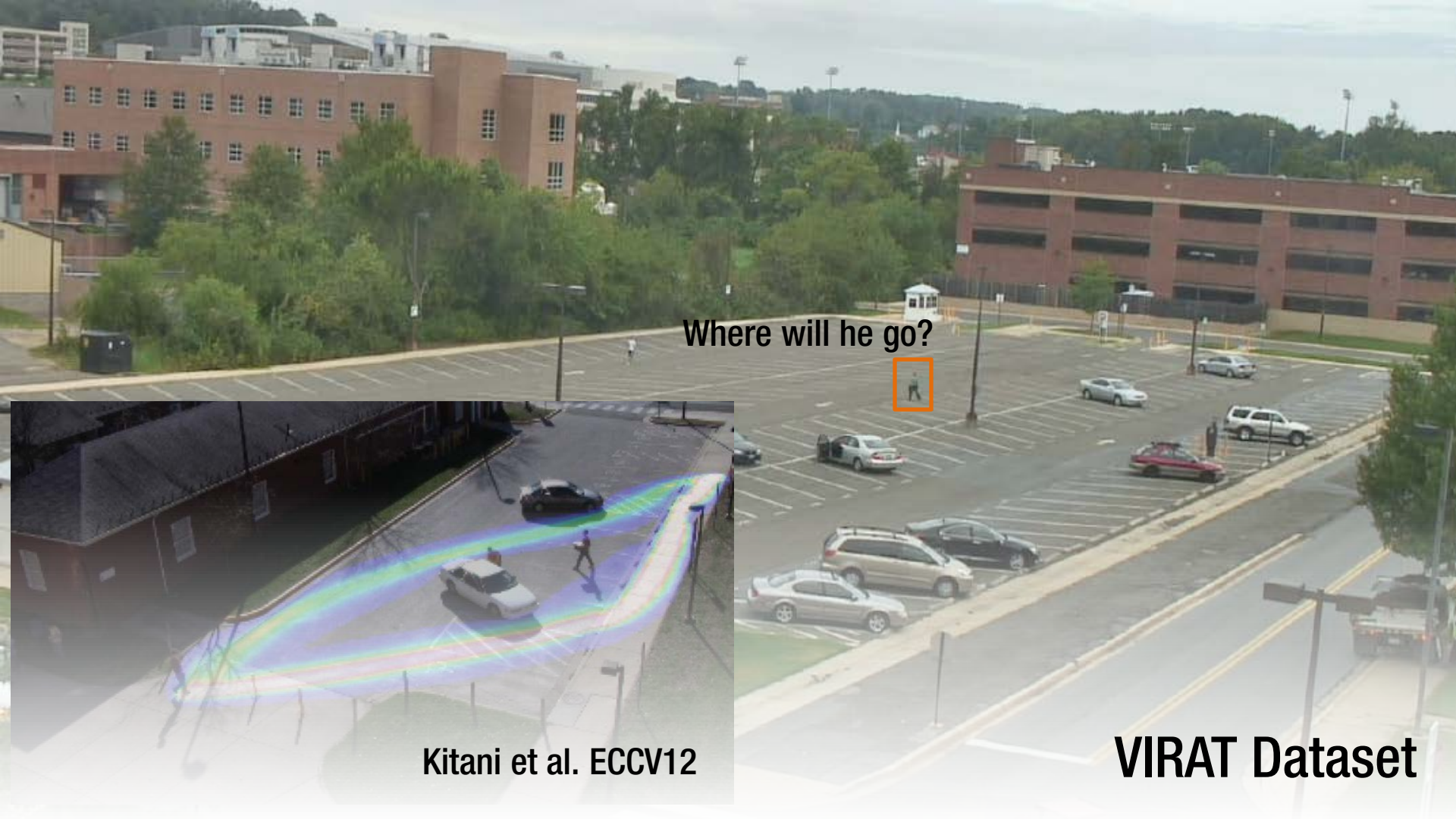


Where will he go?



Kitani et al. ECCV12

VIRAT Dataset



Where will he go?



Kitani et al. ECCV12

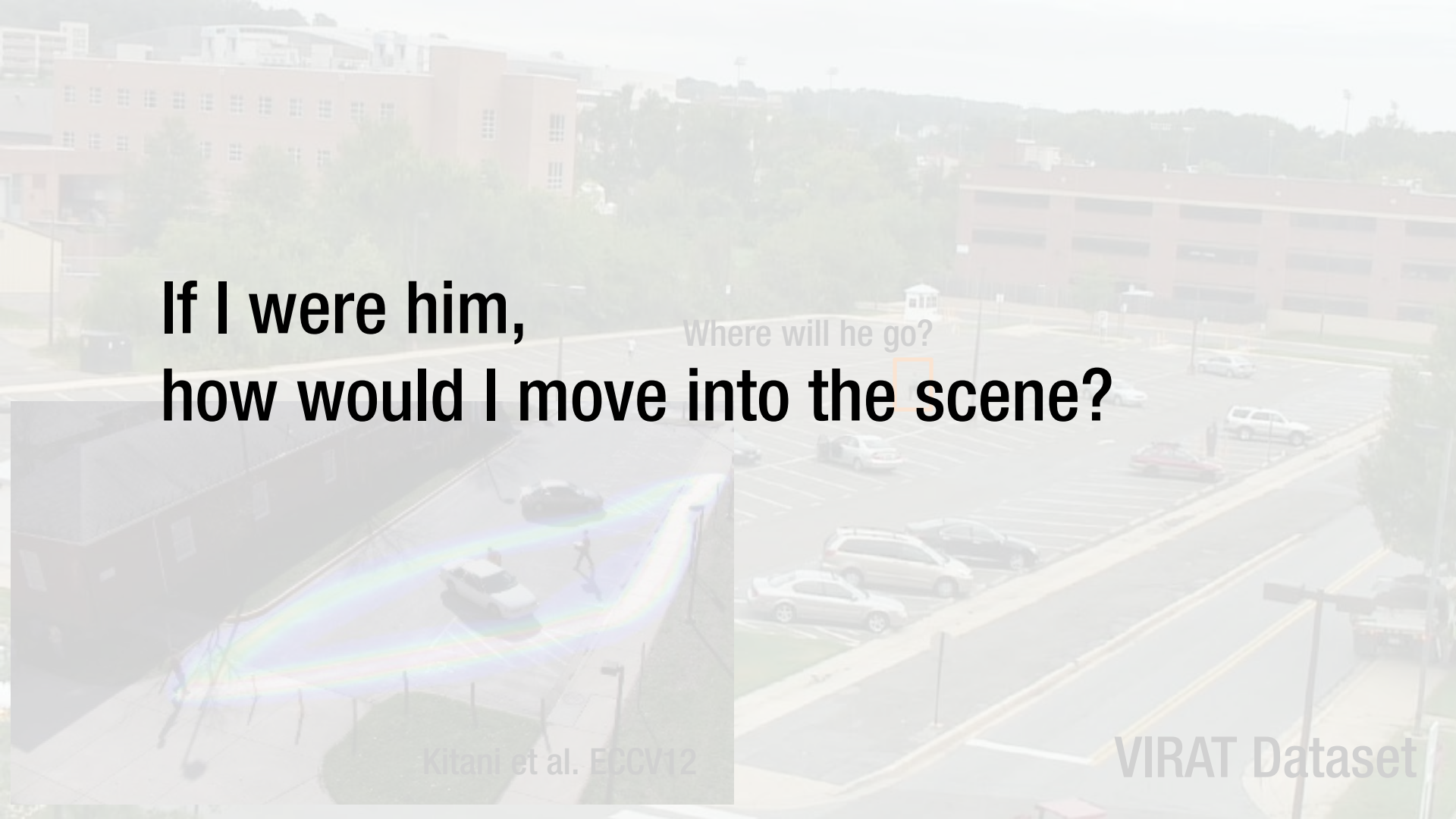
VIRAT Dataset

**If I were him,
how would I move into the scene?**

Where will he go?

Kitani et al. ECCV12

VIRAT Dataset



What is he experiencing visually?





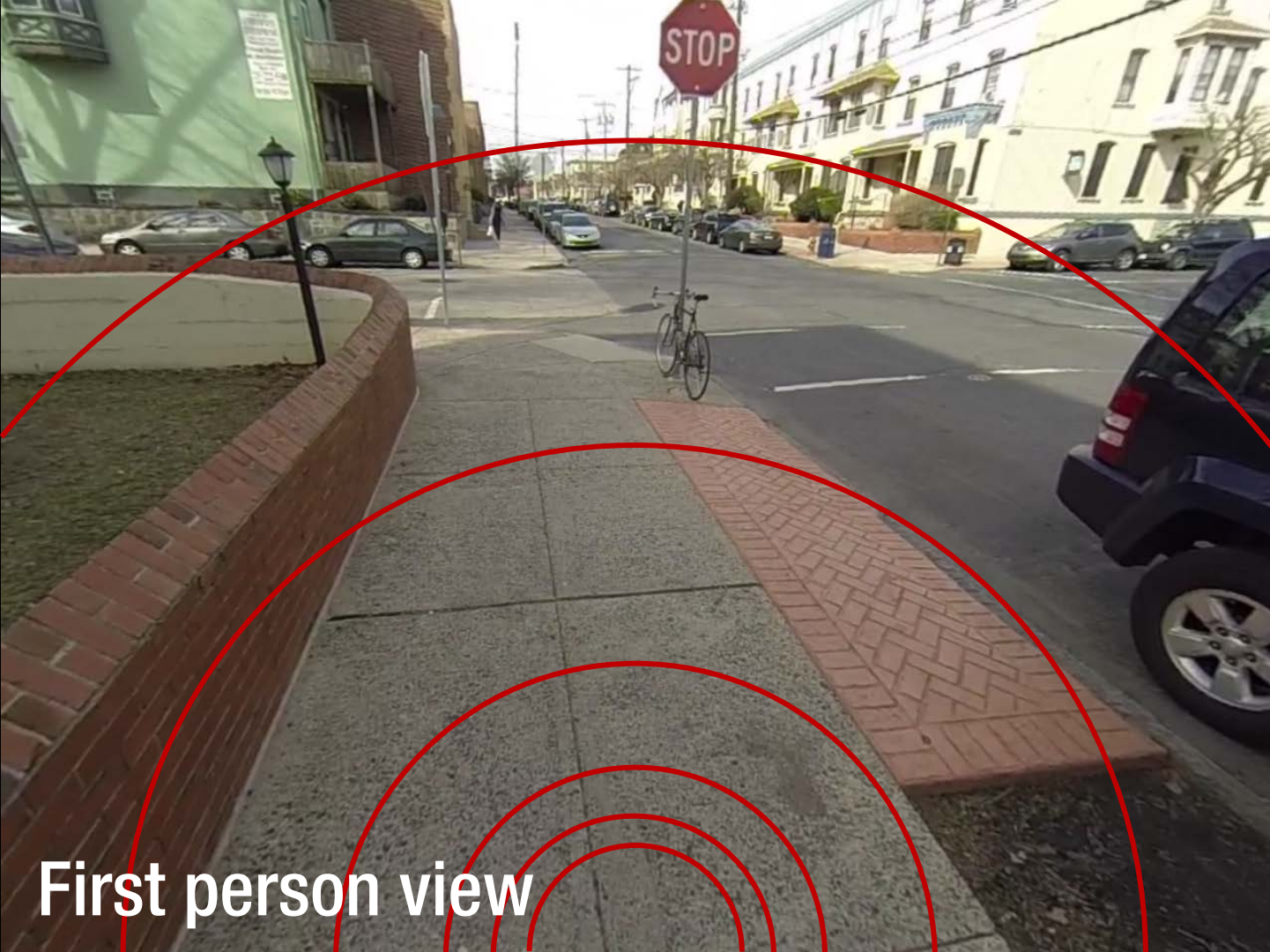
First person view



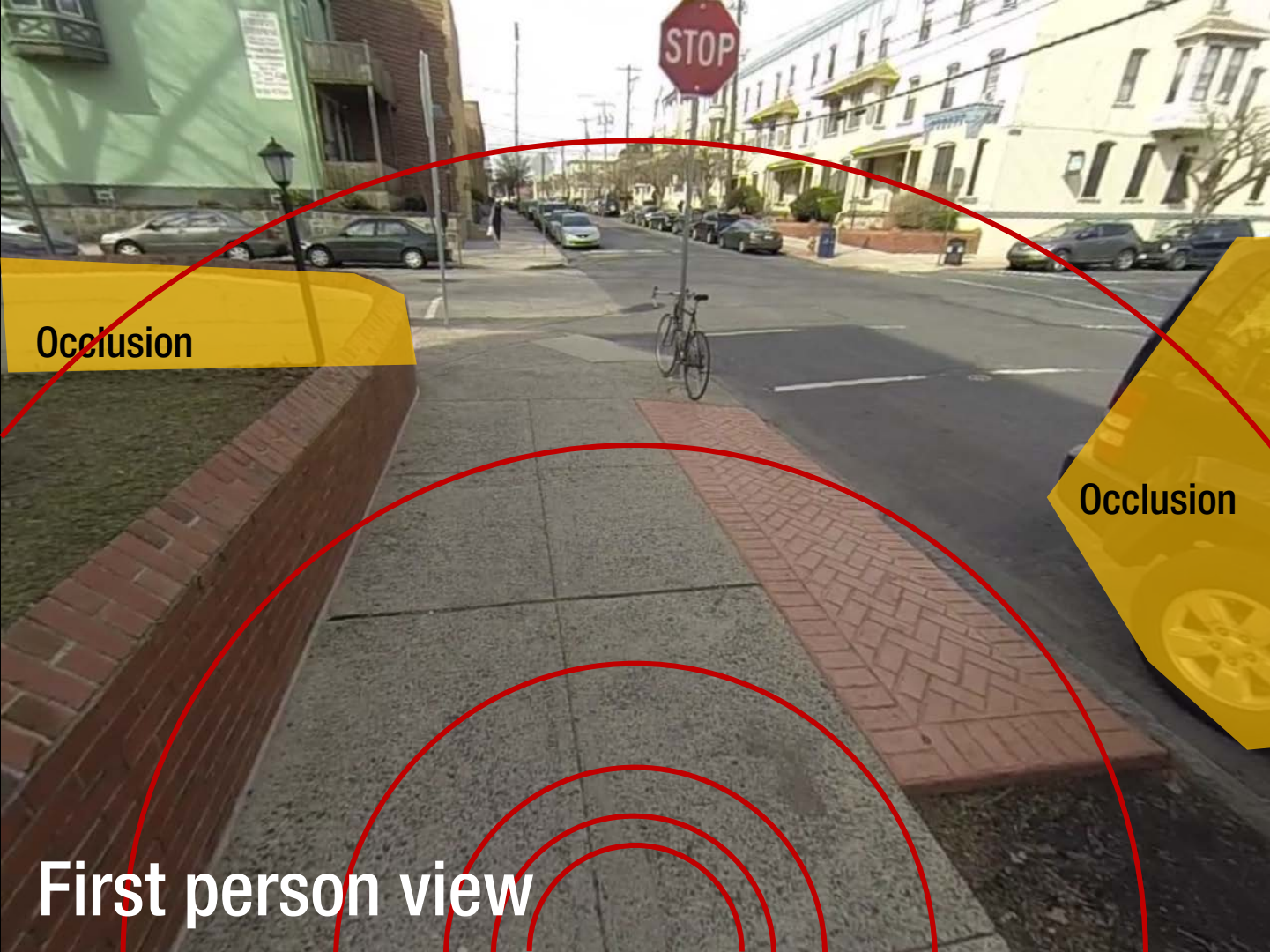
First person view



RGBD



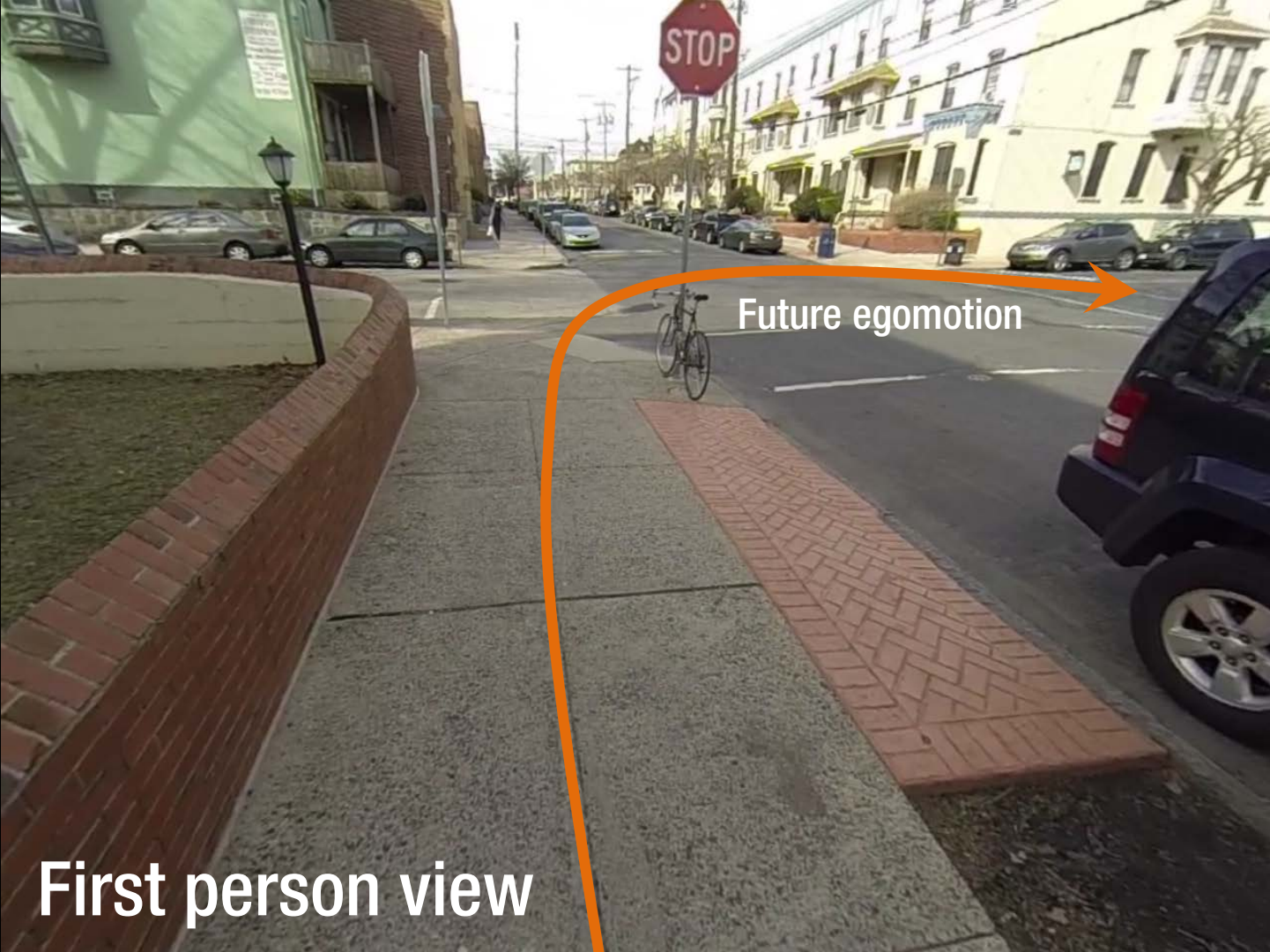
First person view



Occlusion

Occlusion

First person view

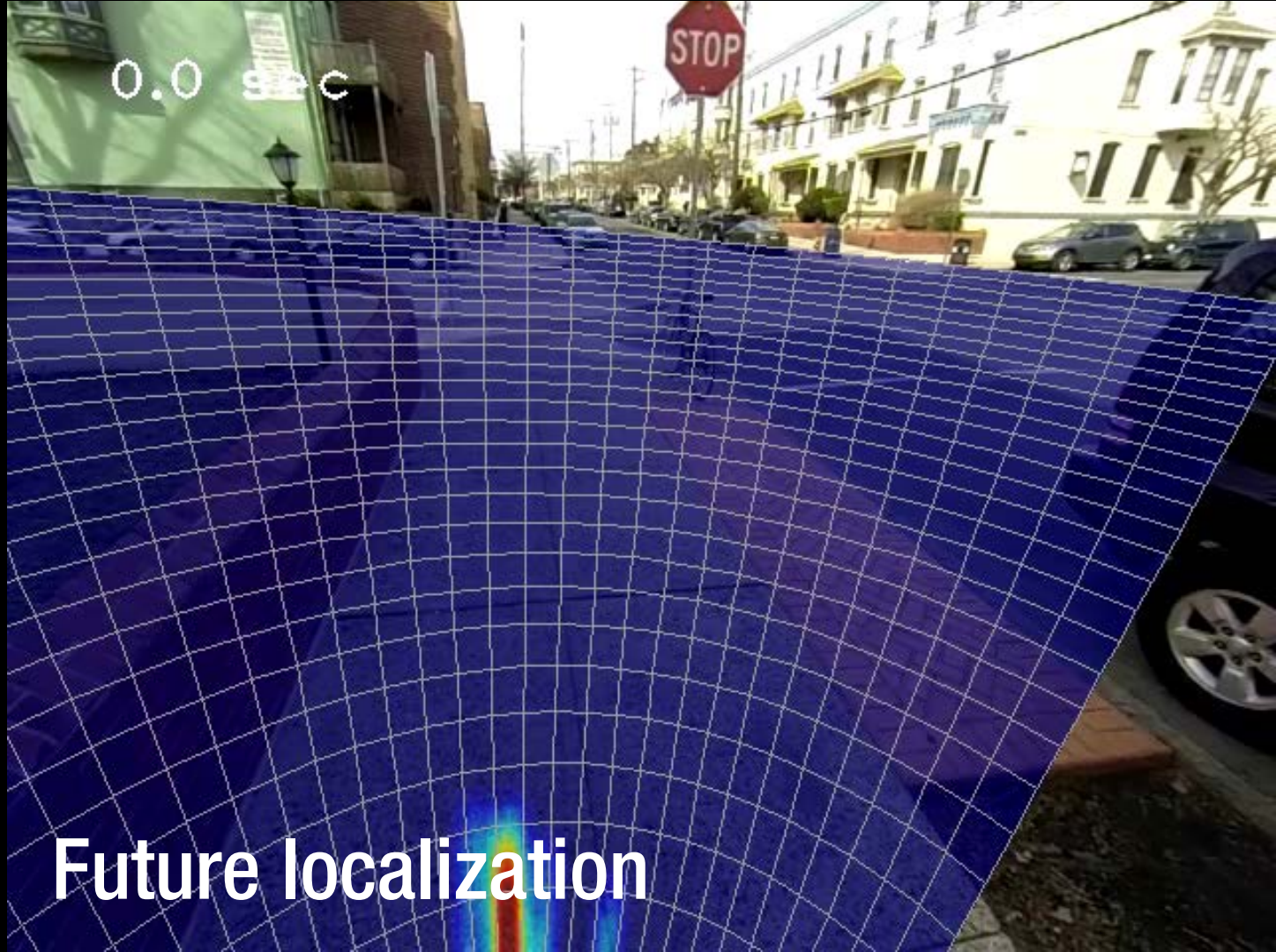


Future egomotion

First person view

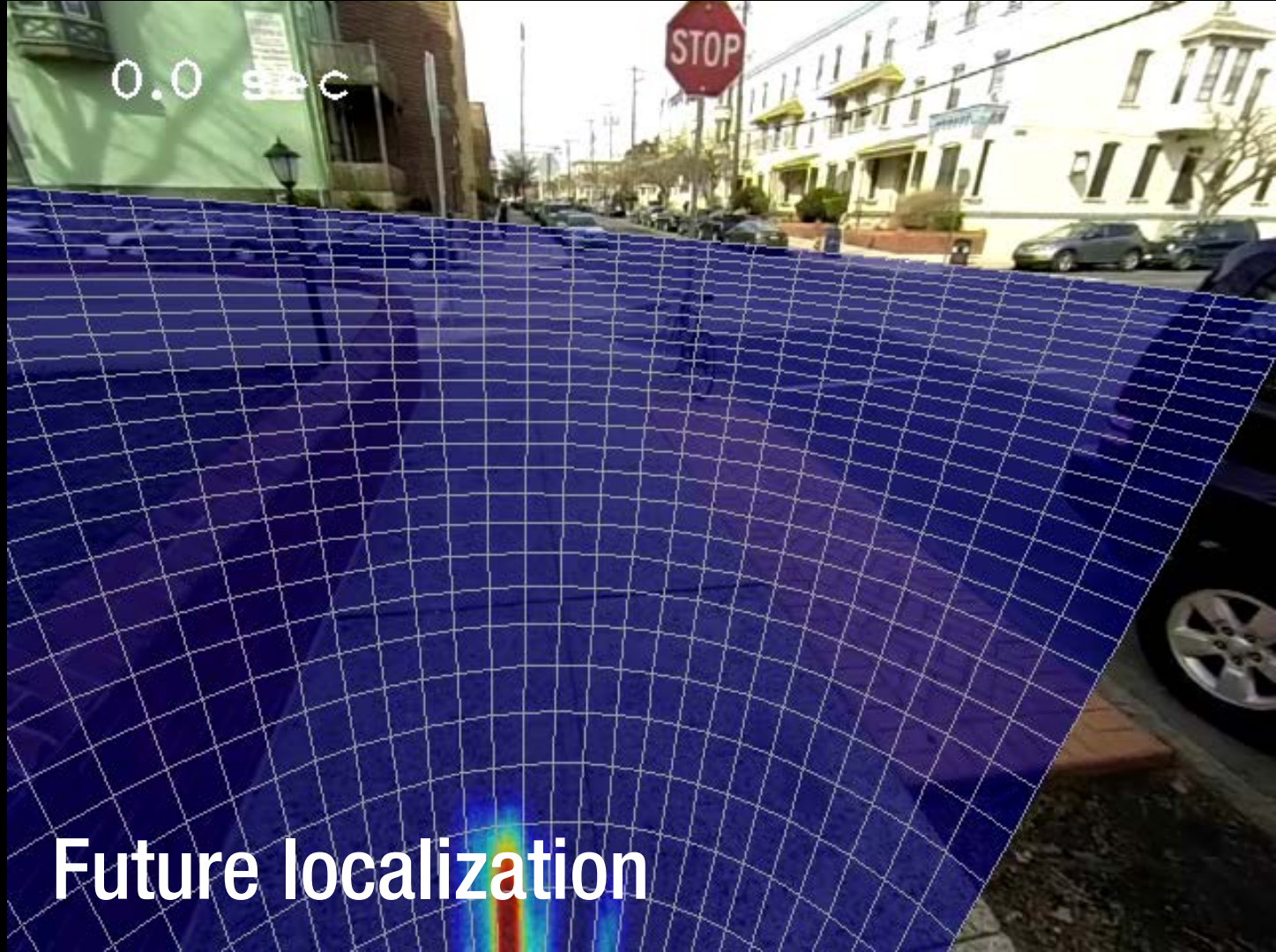
0.0 sec

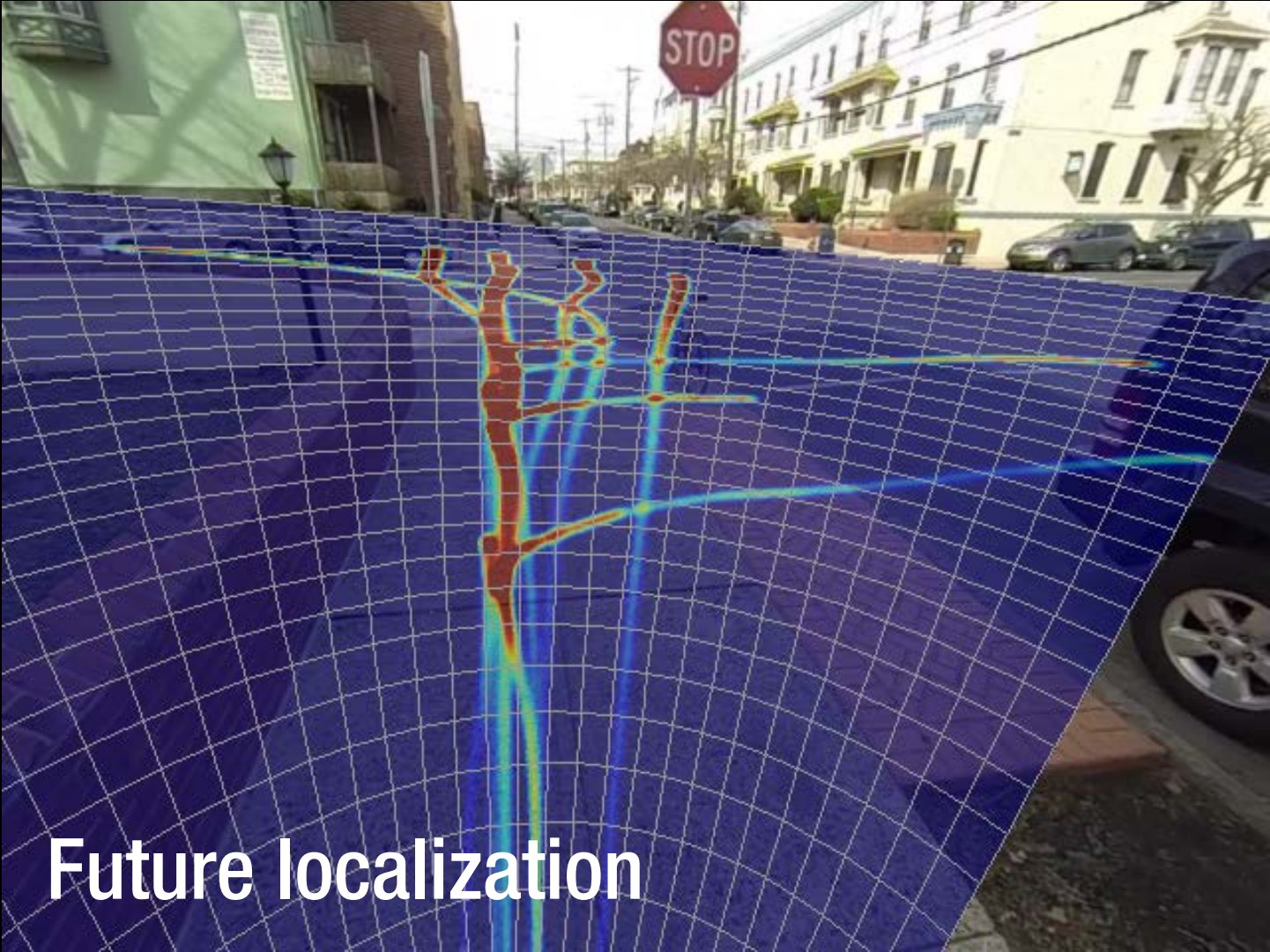
Future localization



0.0 sec

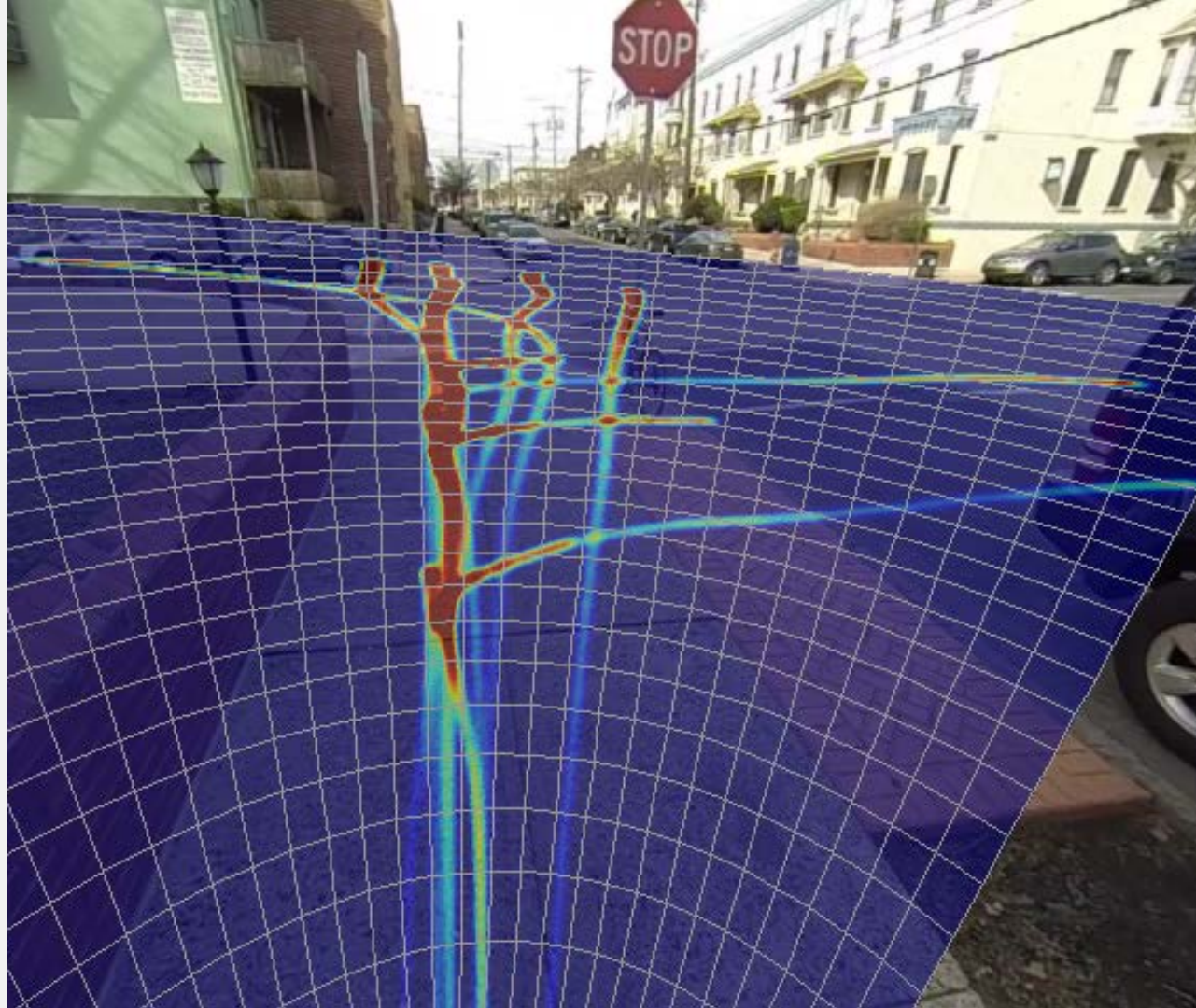
Future localization





Future localization

Why challenging?



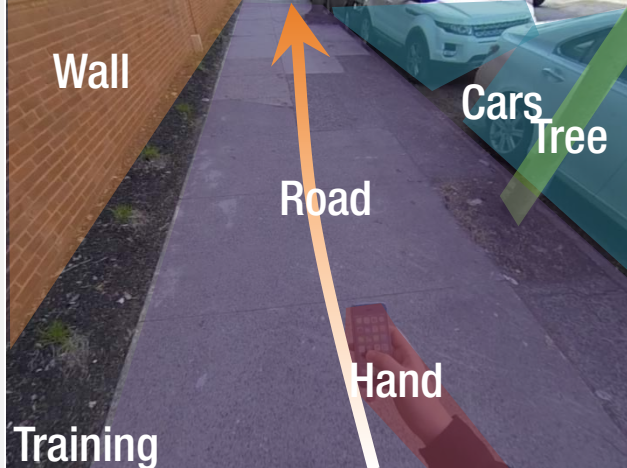
Why challenging?



Why challenging?

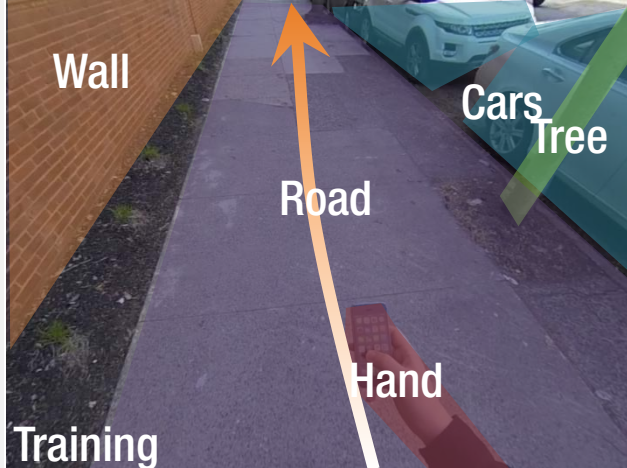


Why challenging?



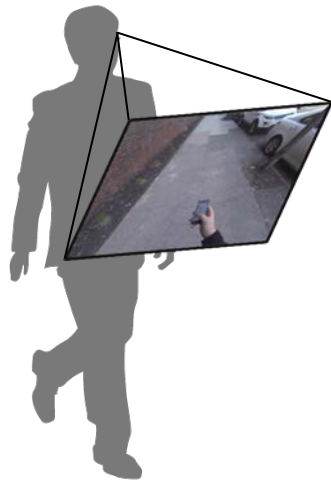
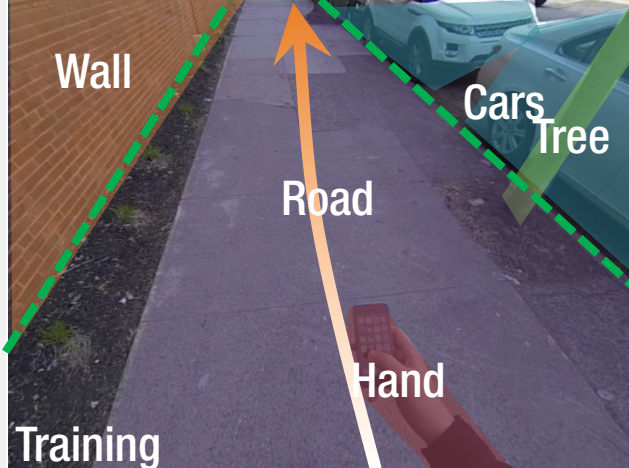
Why challenging?

1. Geometric inconsistency



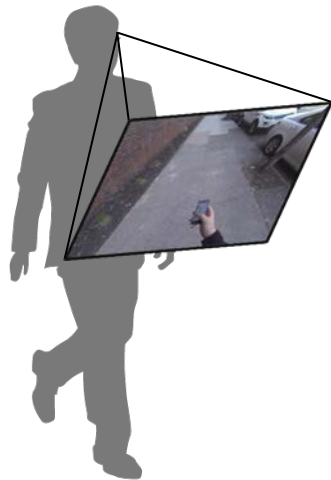
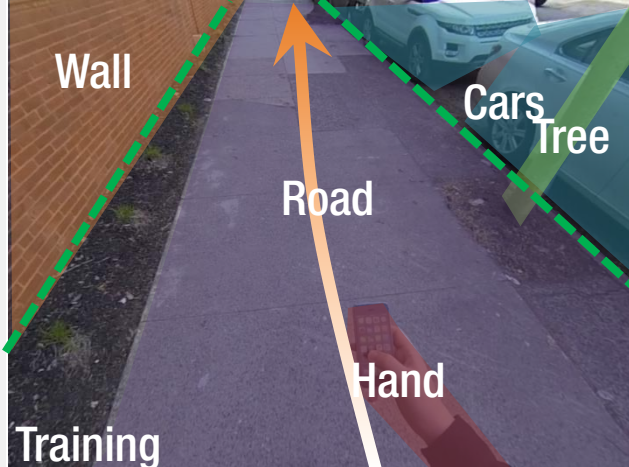
Why challenging?

1. Geometric inconsistency



Why challenging?

1. Geometric inconsistency
2. Semantic inconsistency

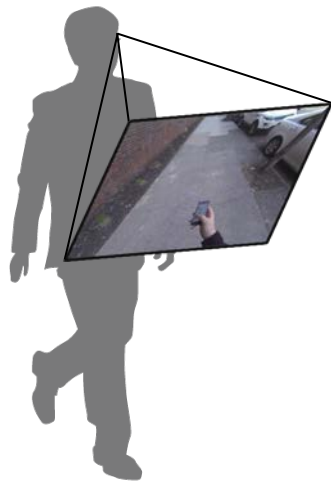
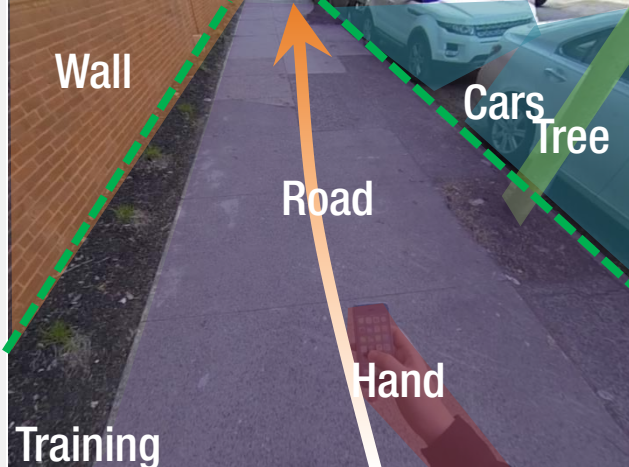


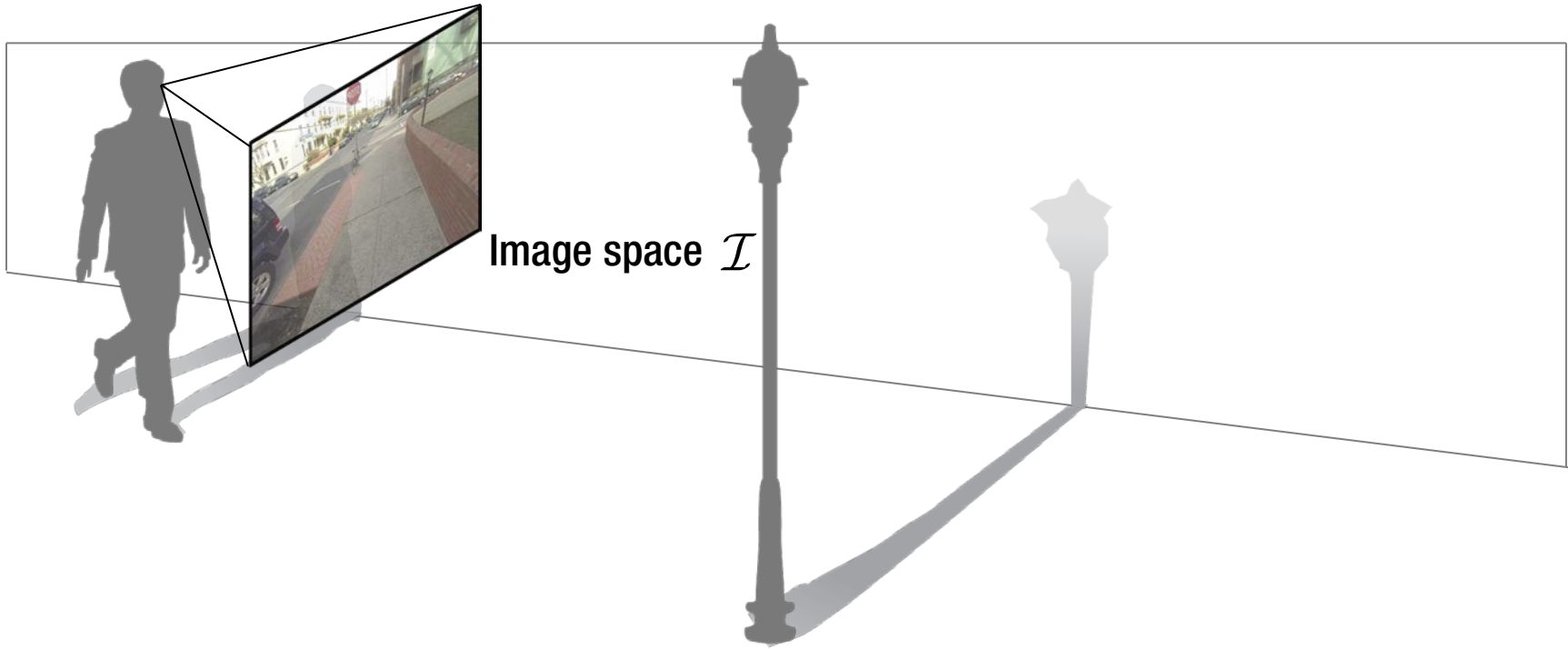
Looking down

Looking forward

Why challenging?

1. Geometric inconsistency
→ EgoRetinal representation
2. Semantic inconsistency
→ Preference learning





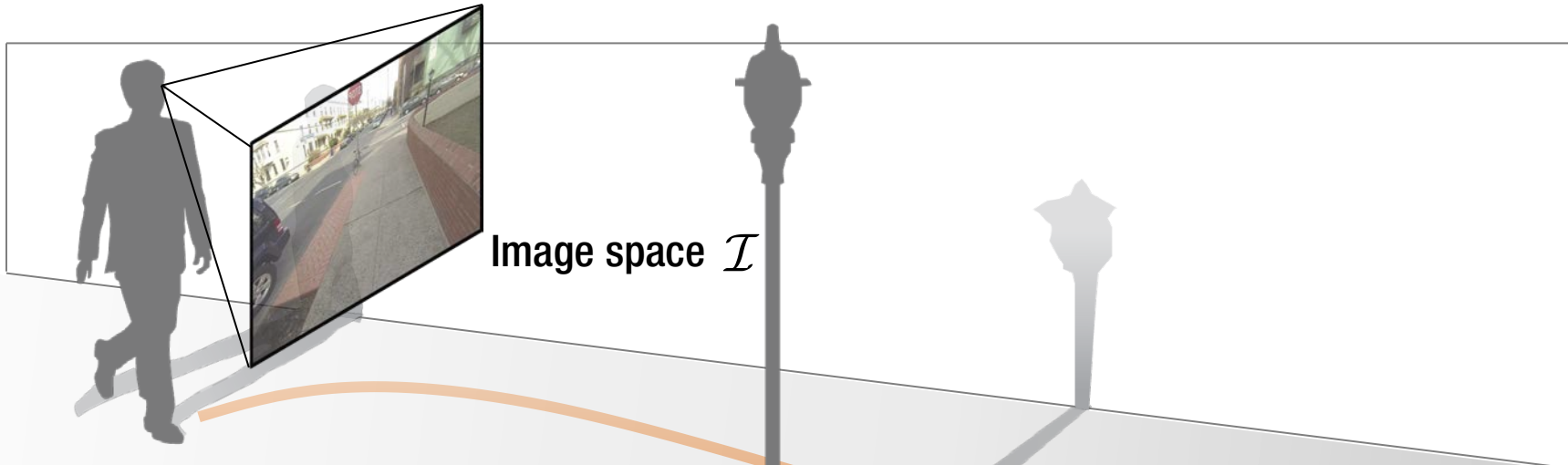
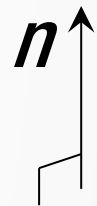
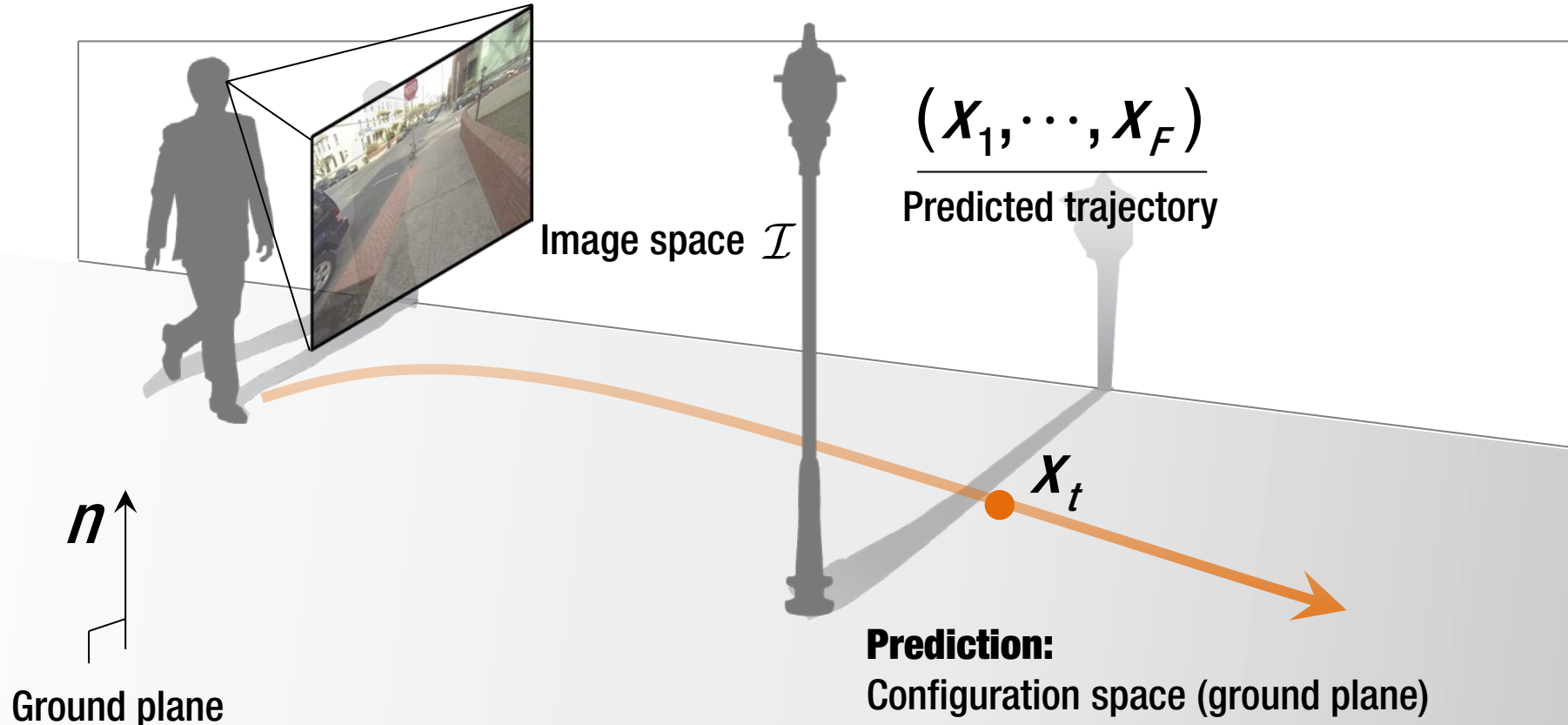


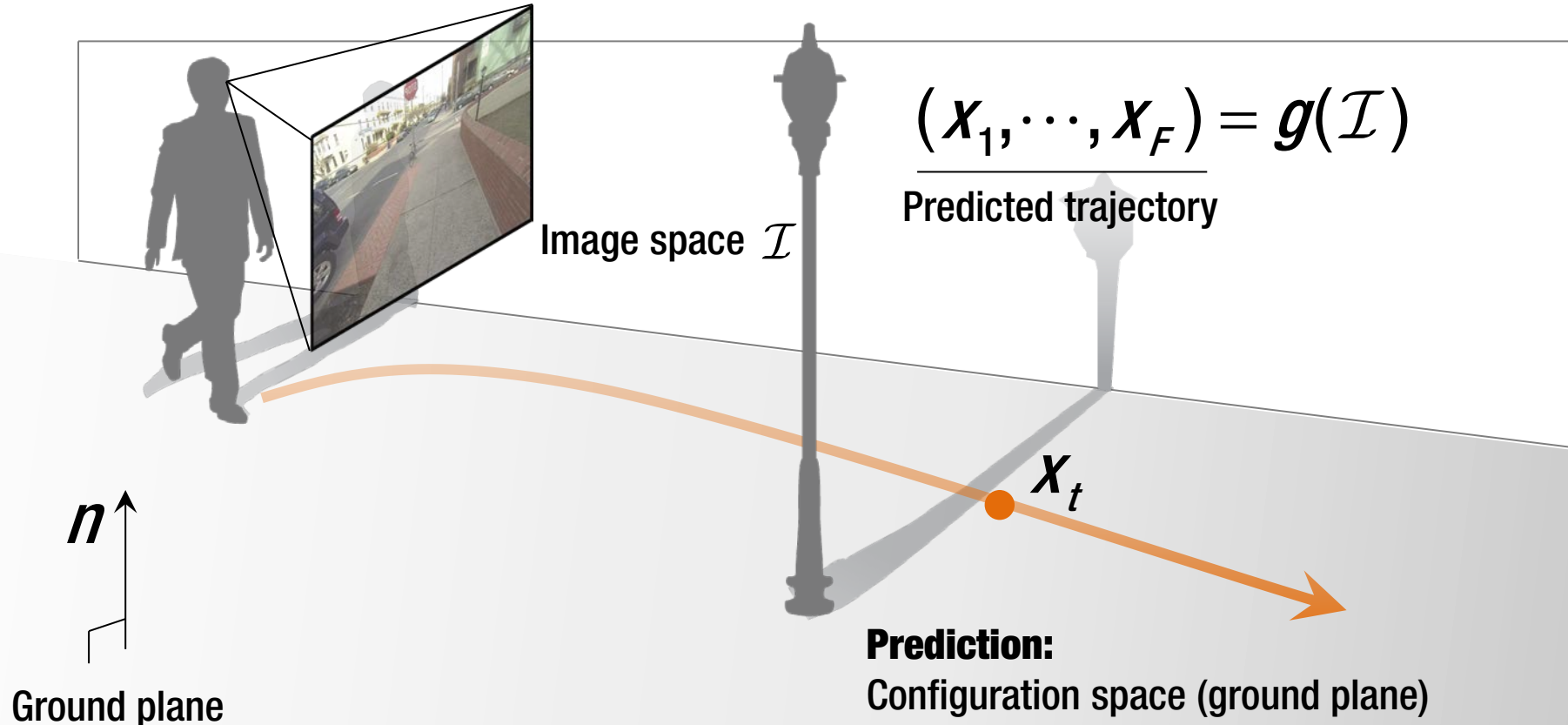
Image space \mathcal{I}



Ground plane

Prediction:
Configuration space (ground plane)





$$(x_1, \dots, x_F) = g(\mathcal{I})$$

Predicted trajectory

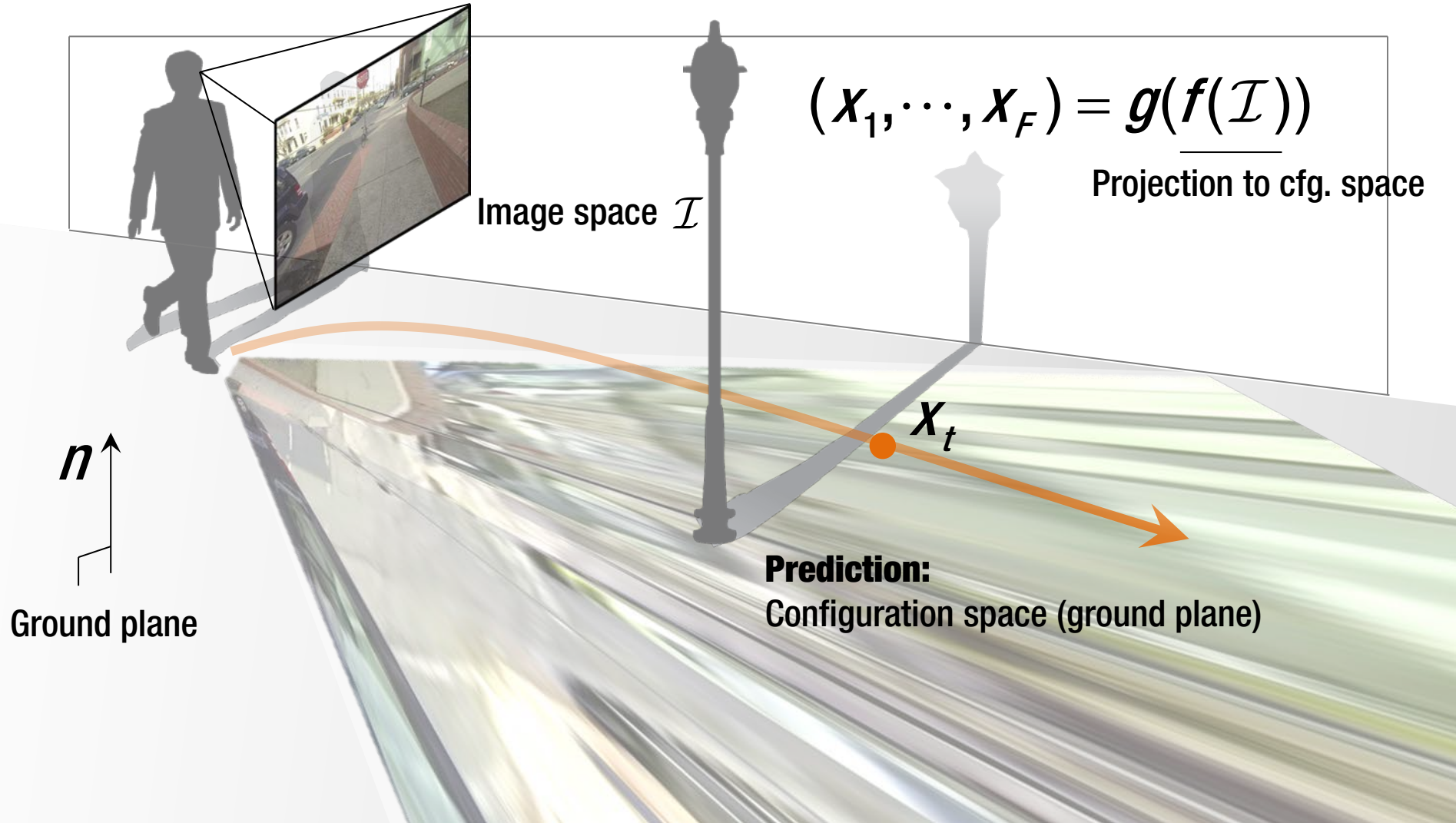
Image space \mathcal{I}

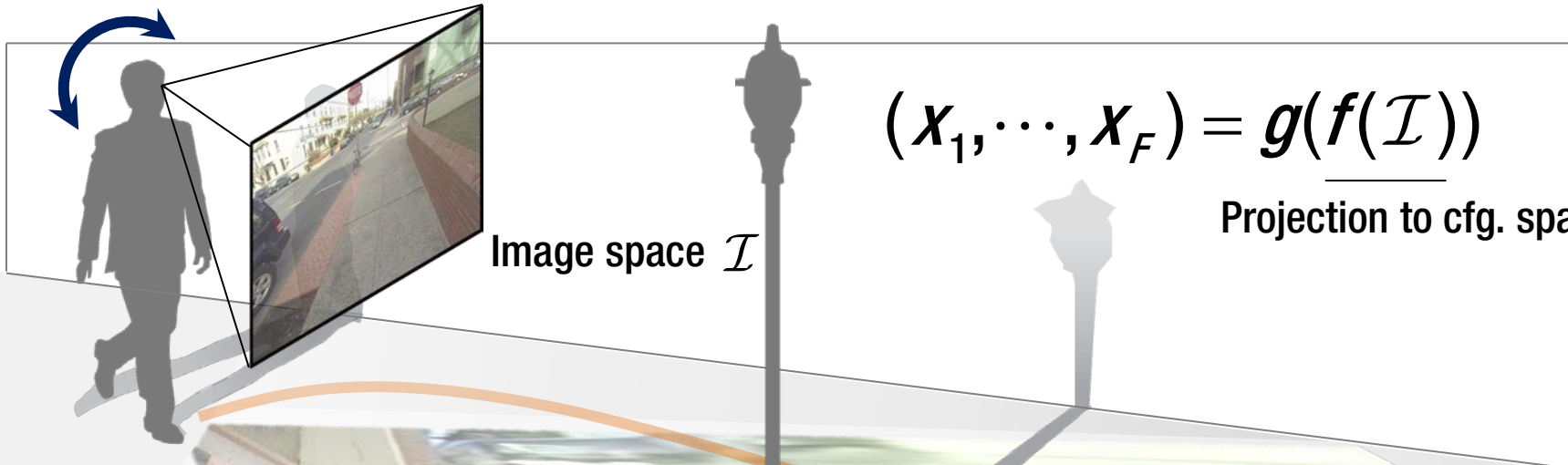
x_t

Prediction:
Configuration space (ground plane)

n

Ground plane





n
Ground plane

Prediction:
Configuration space (ground plane)
Head orientation invariant

x_t

n
Ground plane

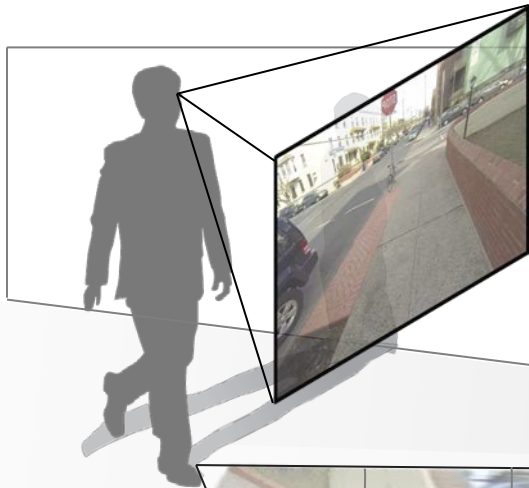



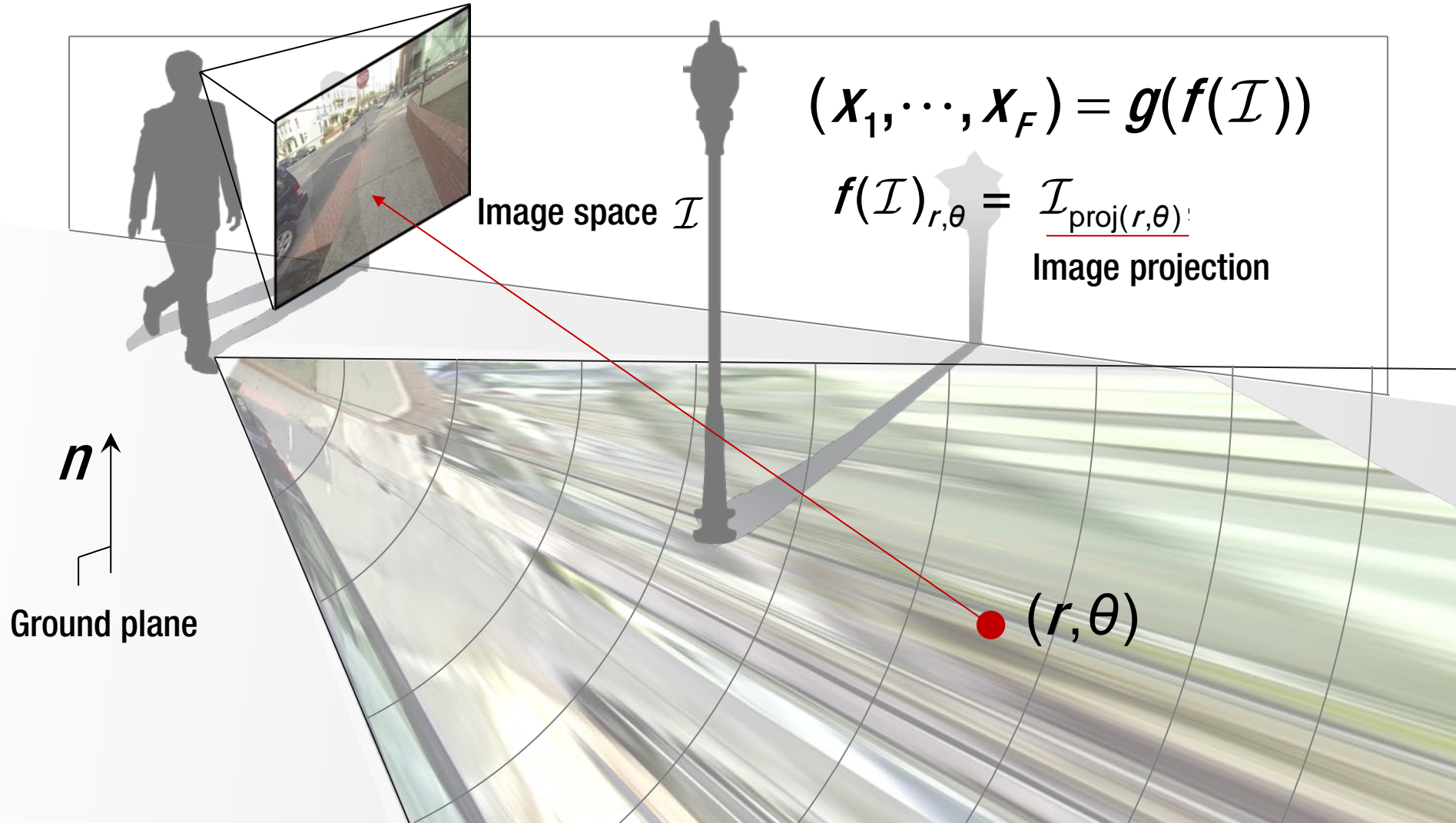
Image space \mathcal{I}

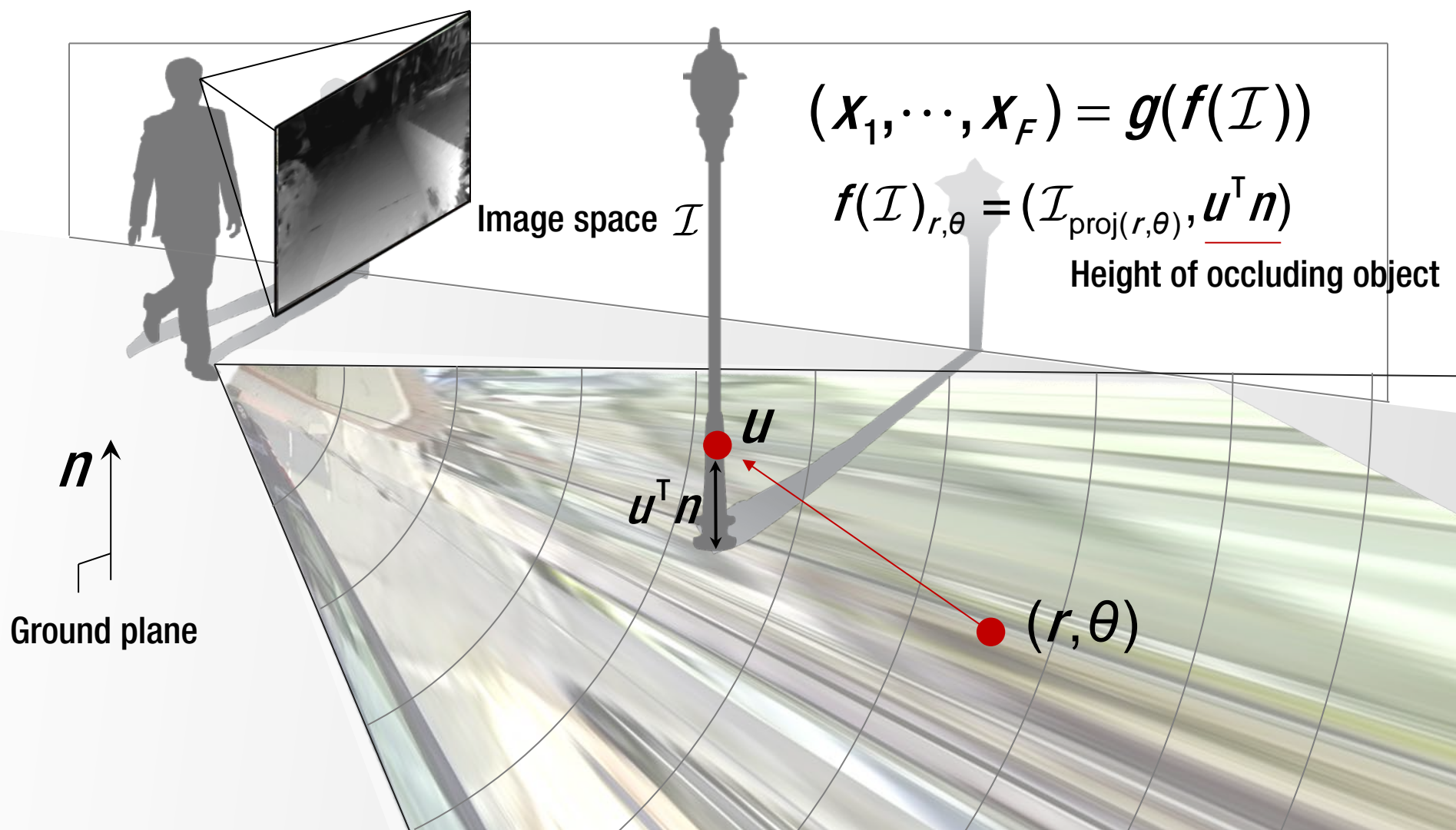
$$(x_1, \dots, x_F) = g(f(\mathcal{I}))$$

$$f(\mathcal{I})_{r,\theta}$$



(r, θ)

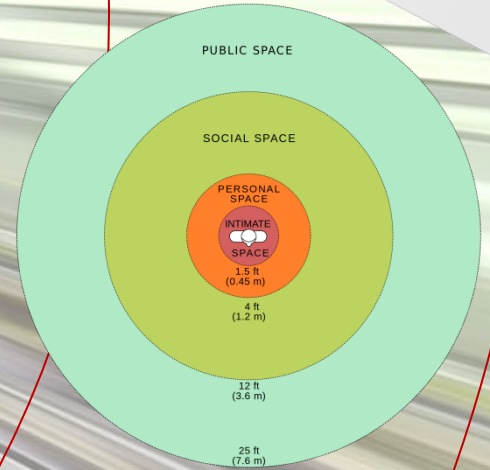




Retinal representation

$$\Delta r \propto \log \frac{1}{D} \text{ where } D \text{ is depth.}$$

n
↑
Ground plane



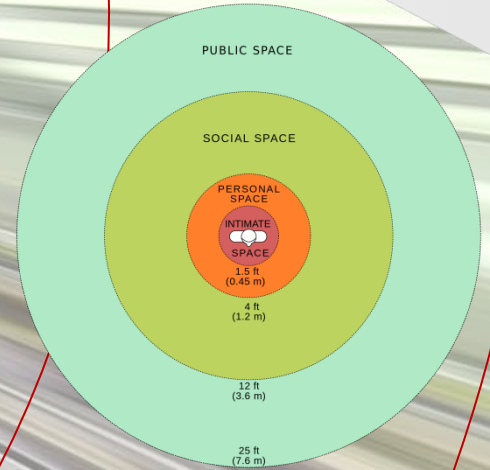
Cf) Proxemics

Retinal representation

Persistent to 2D and 3D distance

$$\Delta r \propto \log \frac{1}{D} \text{ where } D \text{ is depth.}$$

n
Ground plane

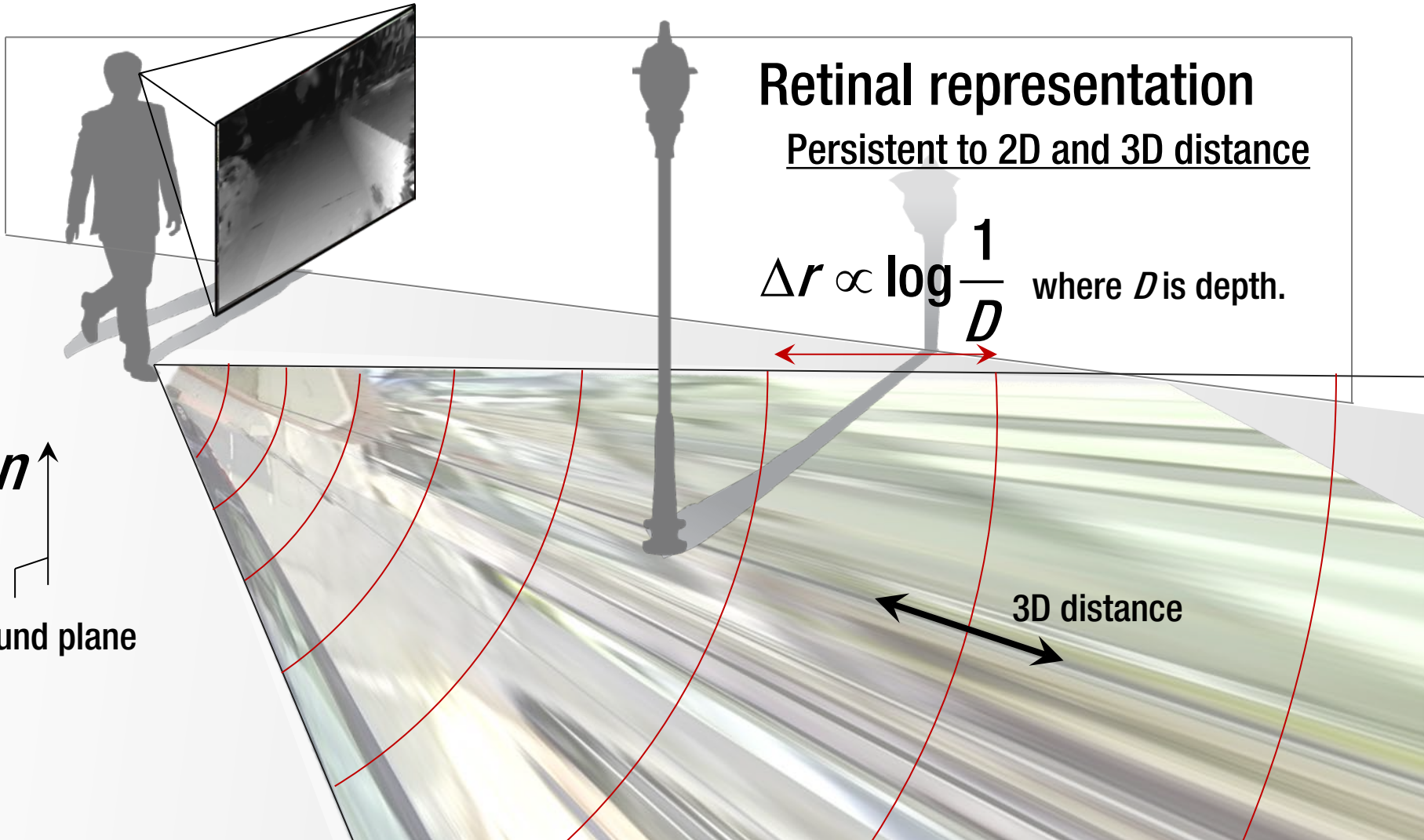


Retinal representation

Persistent to 2D and 3D distance

$$\Delta r \propto \log \frac{1}{D} \text{ where } D \text{ is depth.}$$

n
Ground plane



Retinal representation

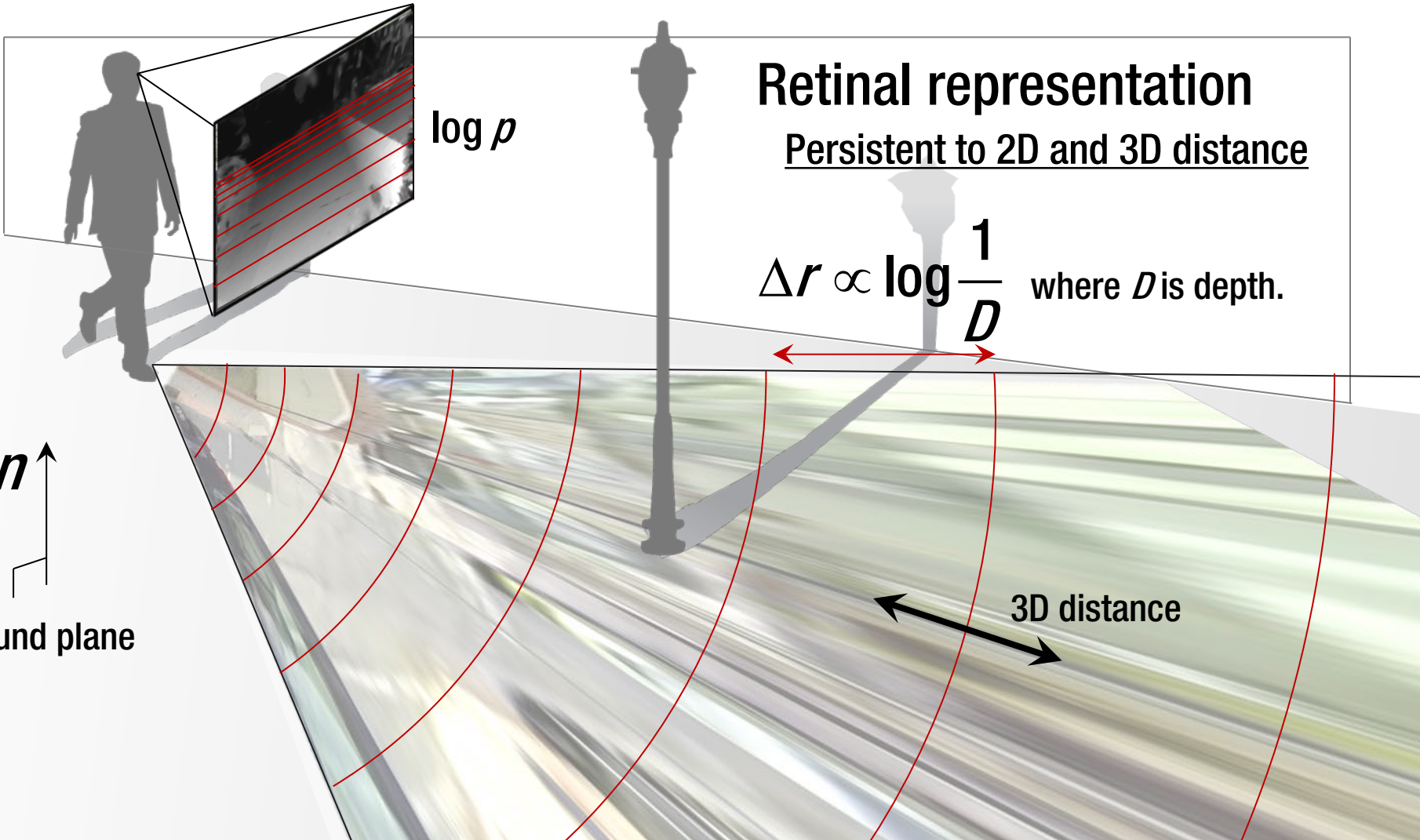
Persistent to 2D and 3D distance

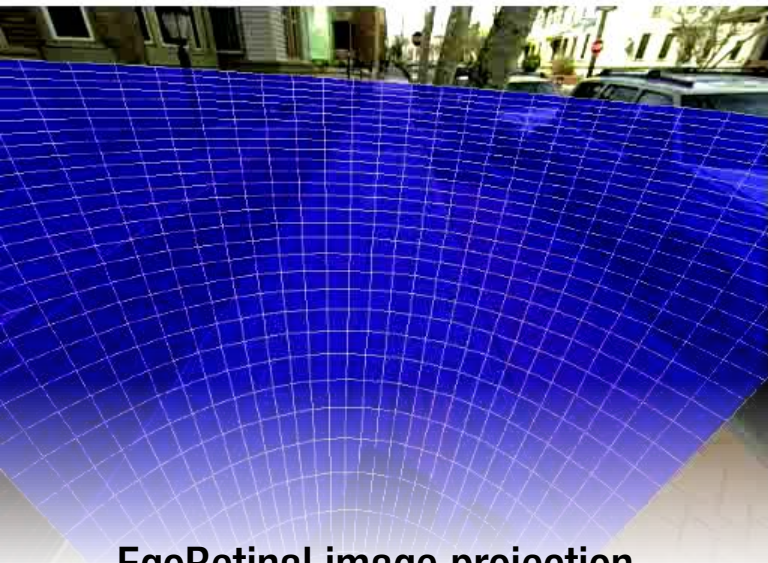
$$\Delta r \propto \log \frac{1}{D} \text{ where } D \text{ is depth.}$$

$\log \rho$

n

Ground plane

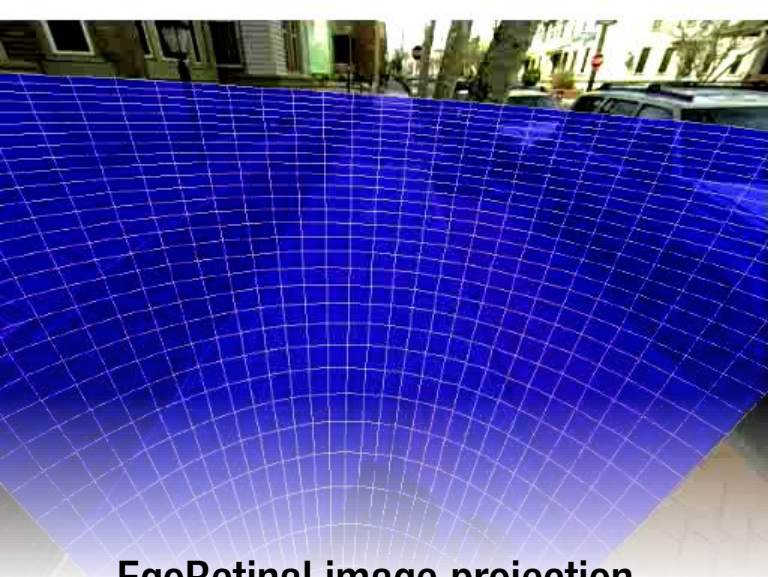




EgoRetinal image projection



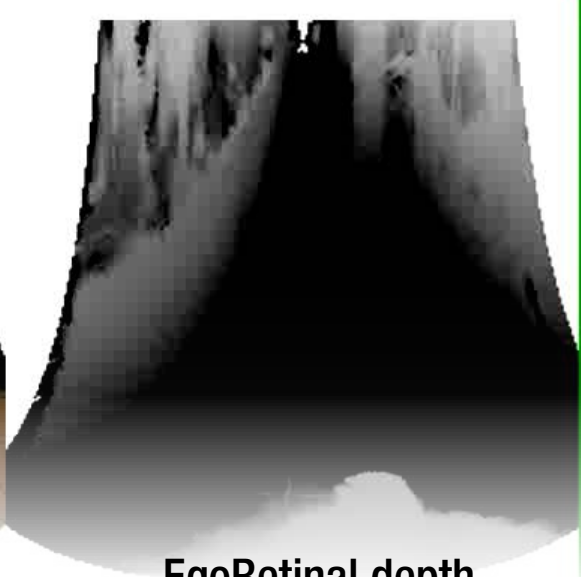
EgoRetinal RGB



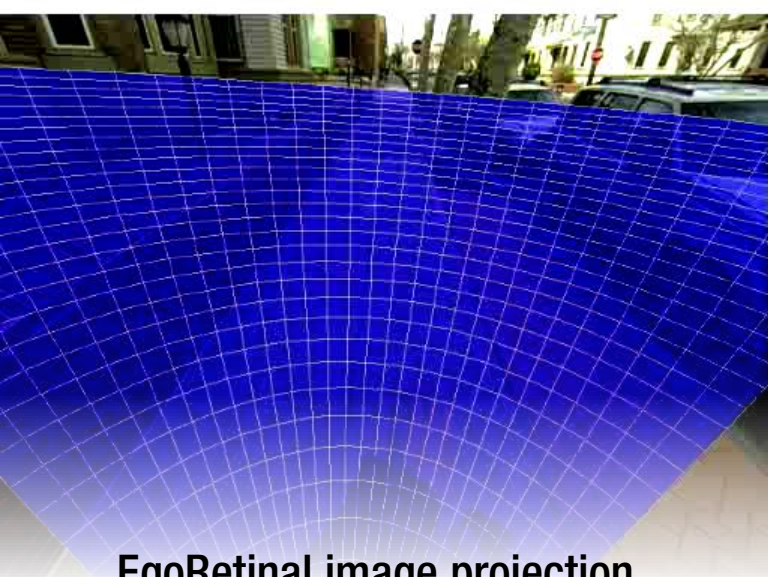
EgoRetinal image projection



EgoRetinal RGB



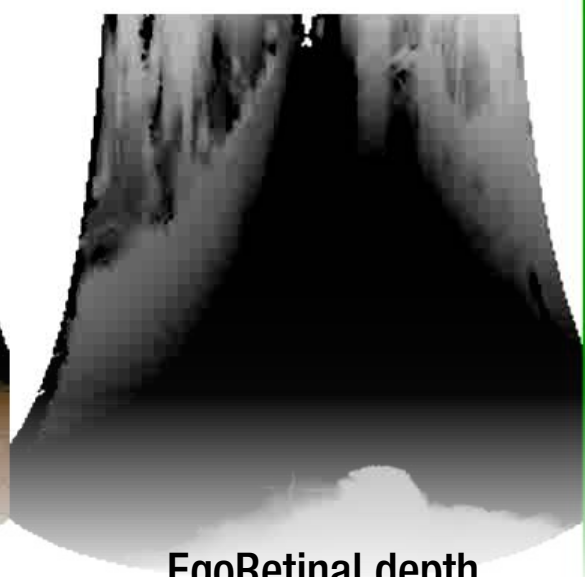
EgoRetinal depth



EgoRetinal image projection



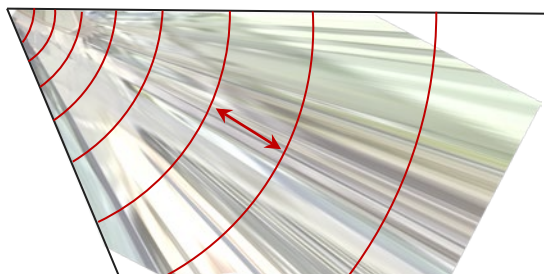
EgoRetinal RGB



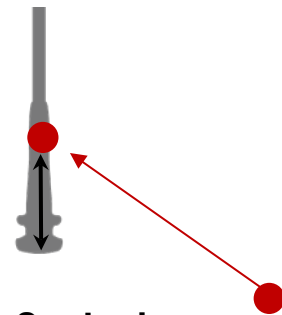
EgoRetinal depth



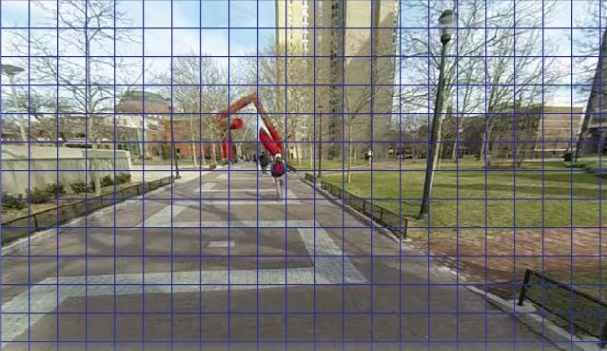
P1: Head orientation invariant



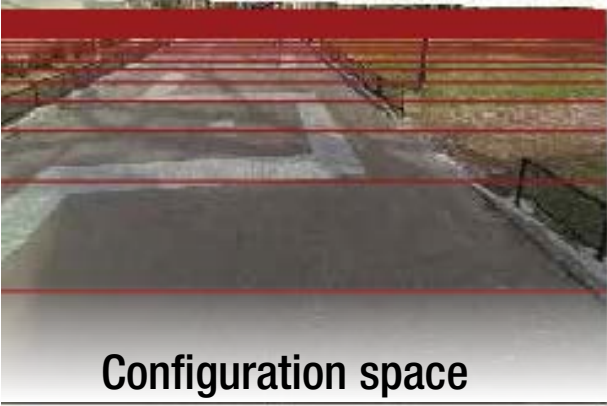
P2: 2D and 3D persistent



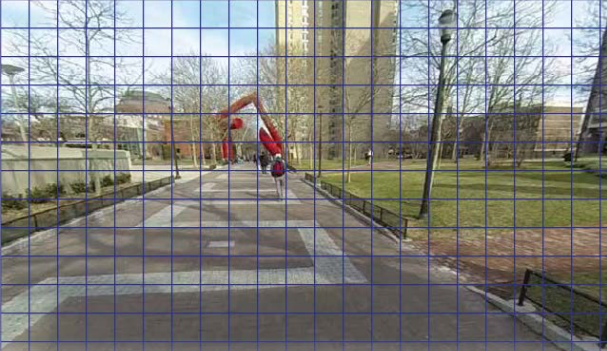
P3: Occlusion reasoning



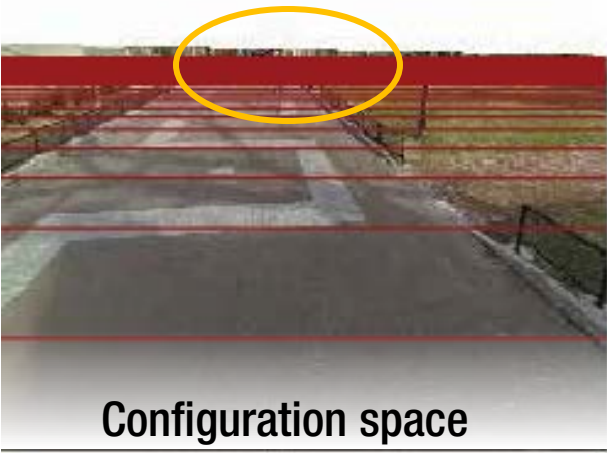
Cartesian in image (LeCun et al.)



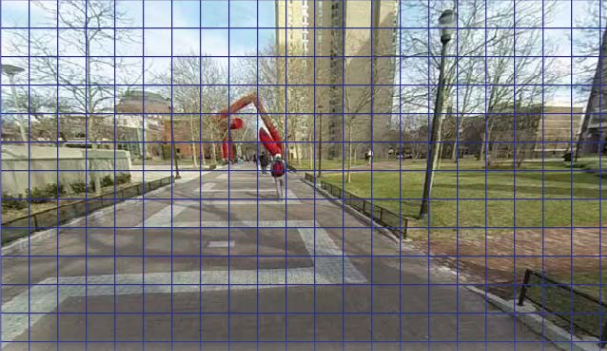
Configuration space



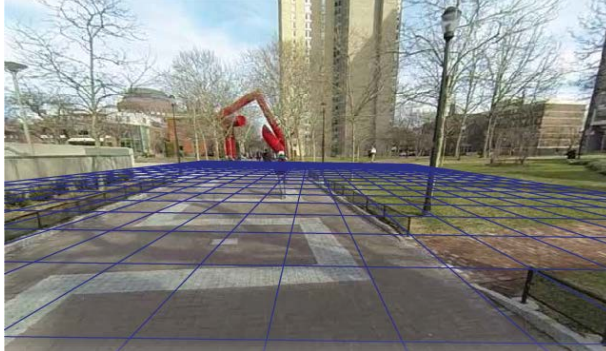
Cartesian in image (LeCun et al.)



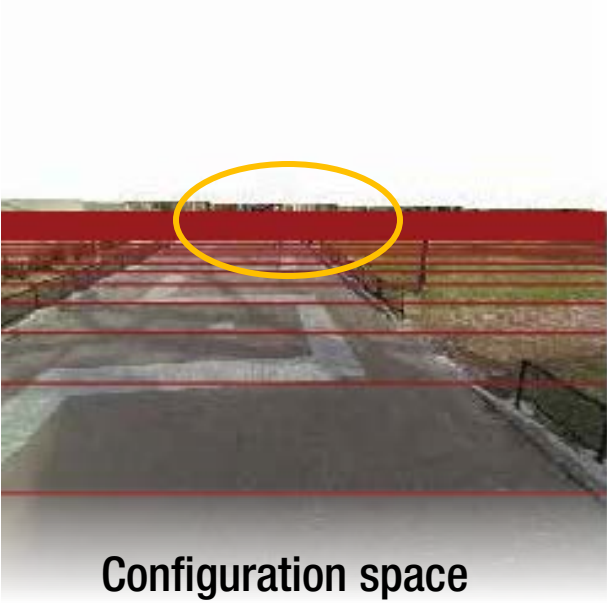
Configuration space



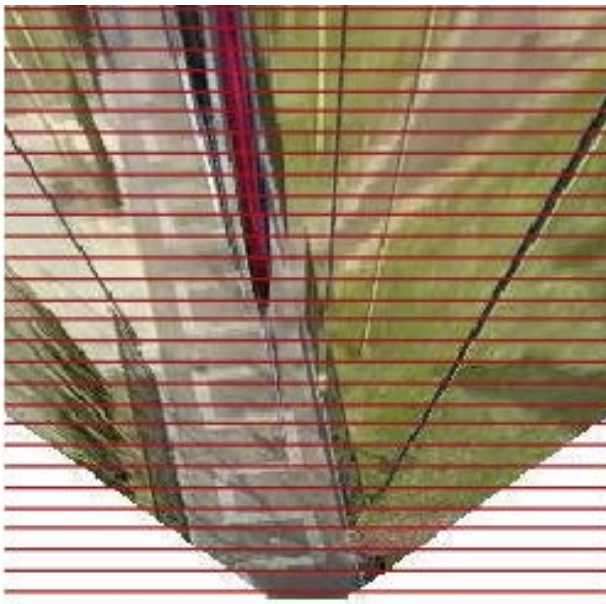
Cartesian in image (LeCun et al.)

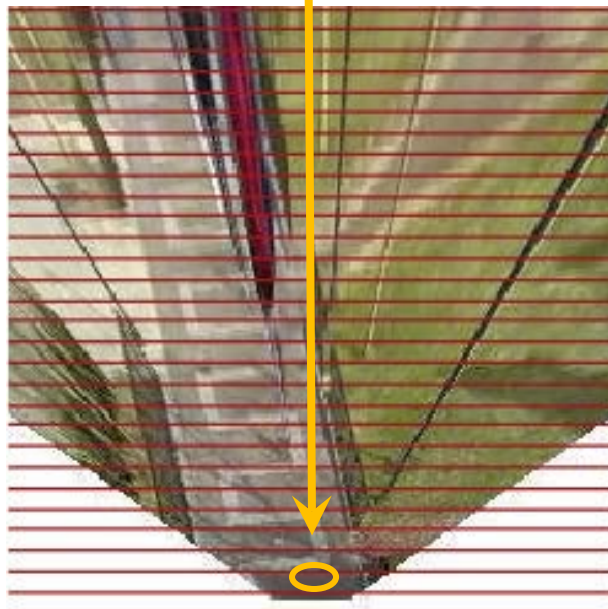
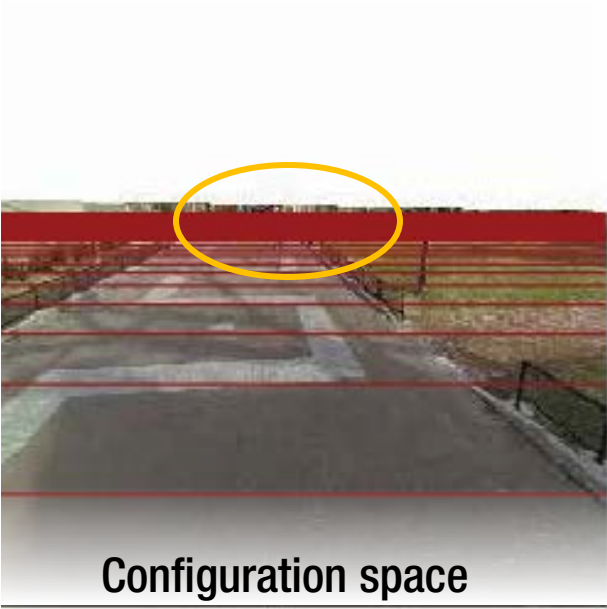


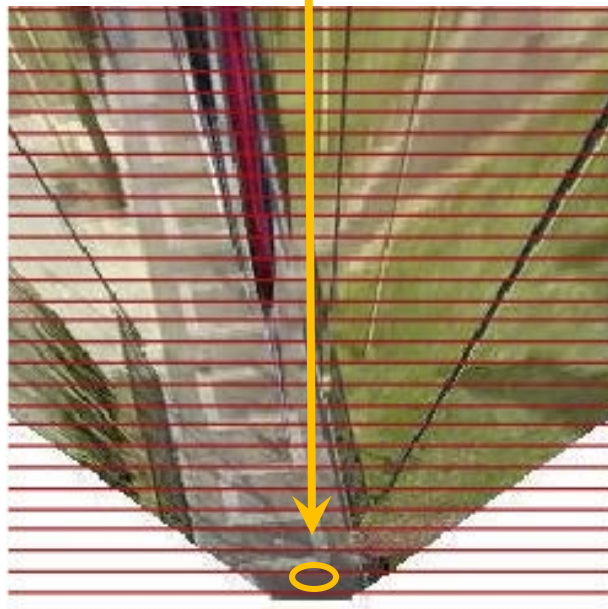
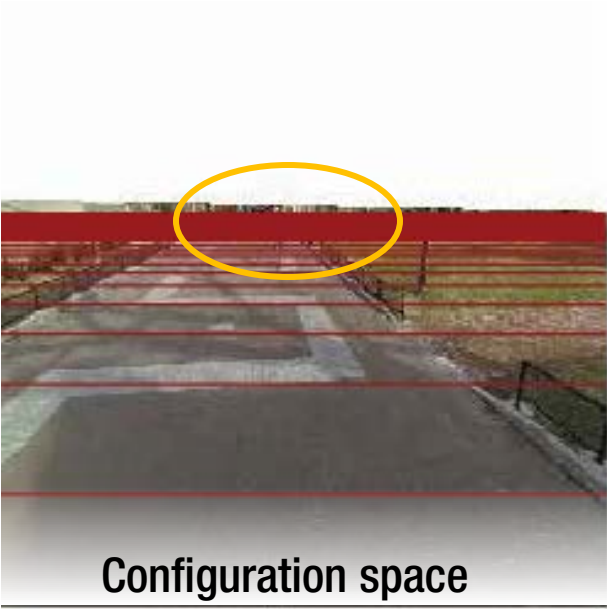
Cartesian in ground plane



Configuration space









Cartesian in image (LeCun et al.)



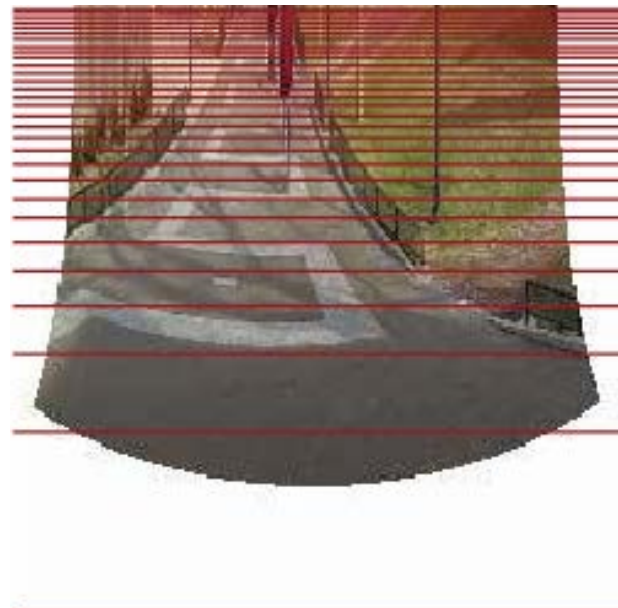
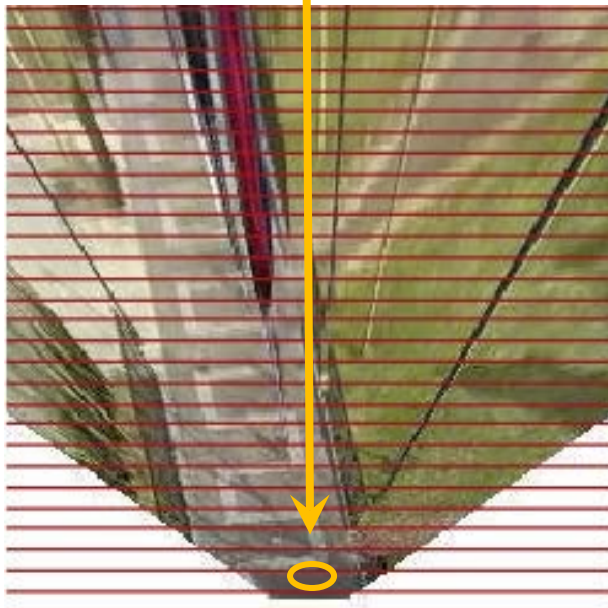
Cartesian in ground plane



EgoRetinal space



Configuration space



EgoMotion Dataset (outdoor)



EgoMotion

OUT
STOP

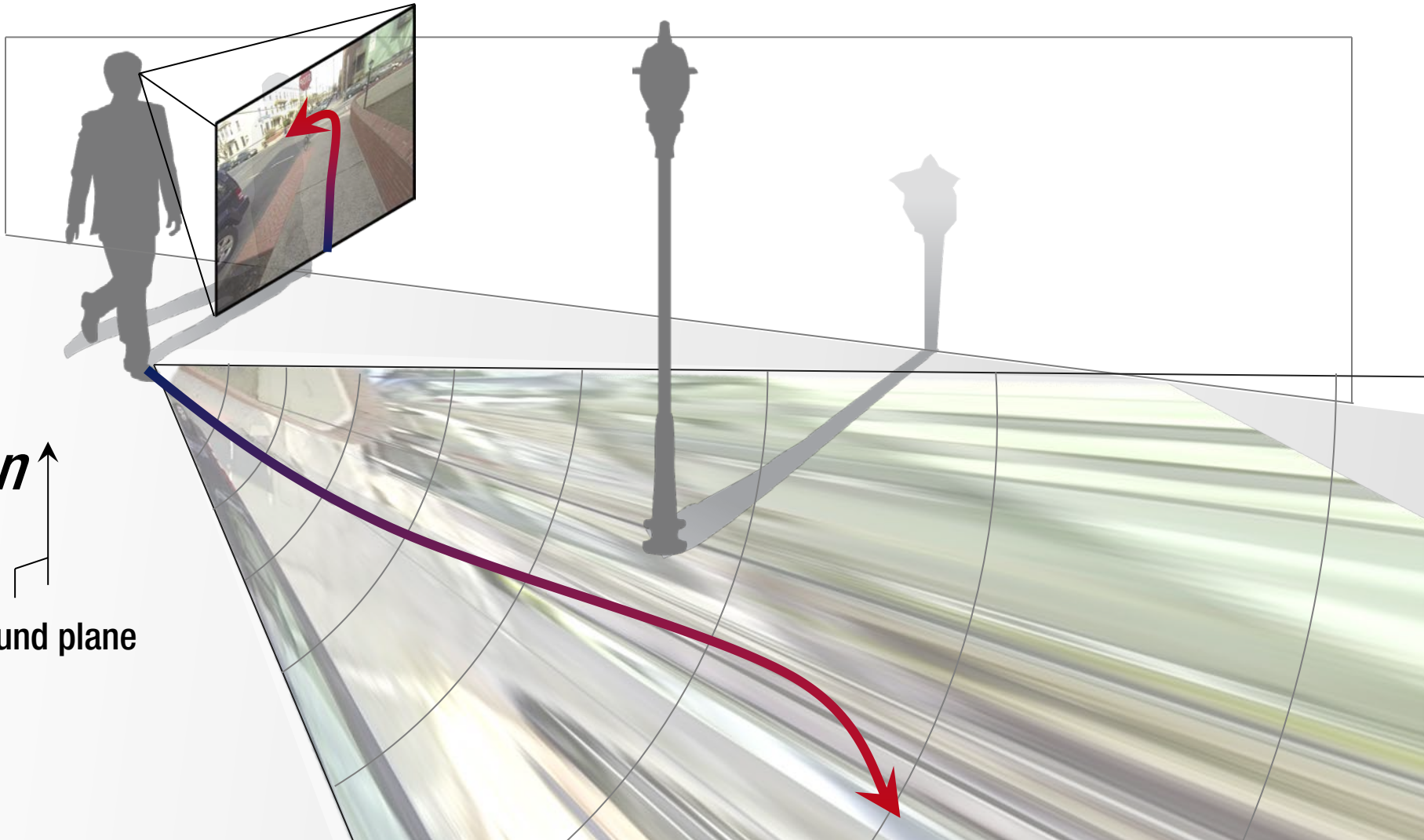
Image Disparity								
Scene	IKEA	Costco	Mall	Park	School1/2	Downtown1/2	Grocery1/2/3	Bus1/2
Frames	966	577	2683	3088	3754/3736	2856/3405	2858/2892/2834	2292/1850
Duration	08:03	04:49	22:22	25:44	31:17/31:08	23:48/28:23	23:49/24:06/23:37	19:06/15:25
Image Disparity								
Scene	Campus1/2/3	CVS1/2	Train Sta.1/2	River1/2	Dep. store	Library	Apartment	Caffe
Frames	2607/1884/1975	2359/3337	4034/2568	3378/2250	2250	1255	2050	1550
Duration	21:44/15:42/16:28	19:40/27:49	33:37/21:24	28:09/18:45	13:20	10:30	17:05	13:00

Dataset summary

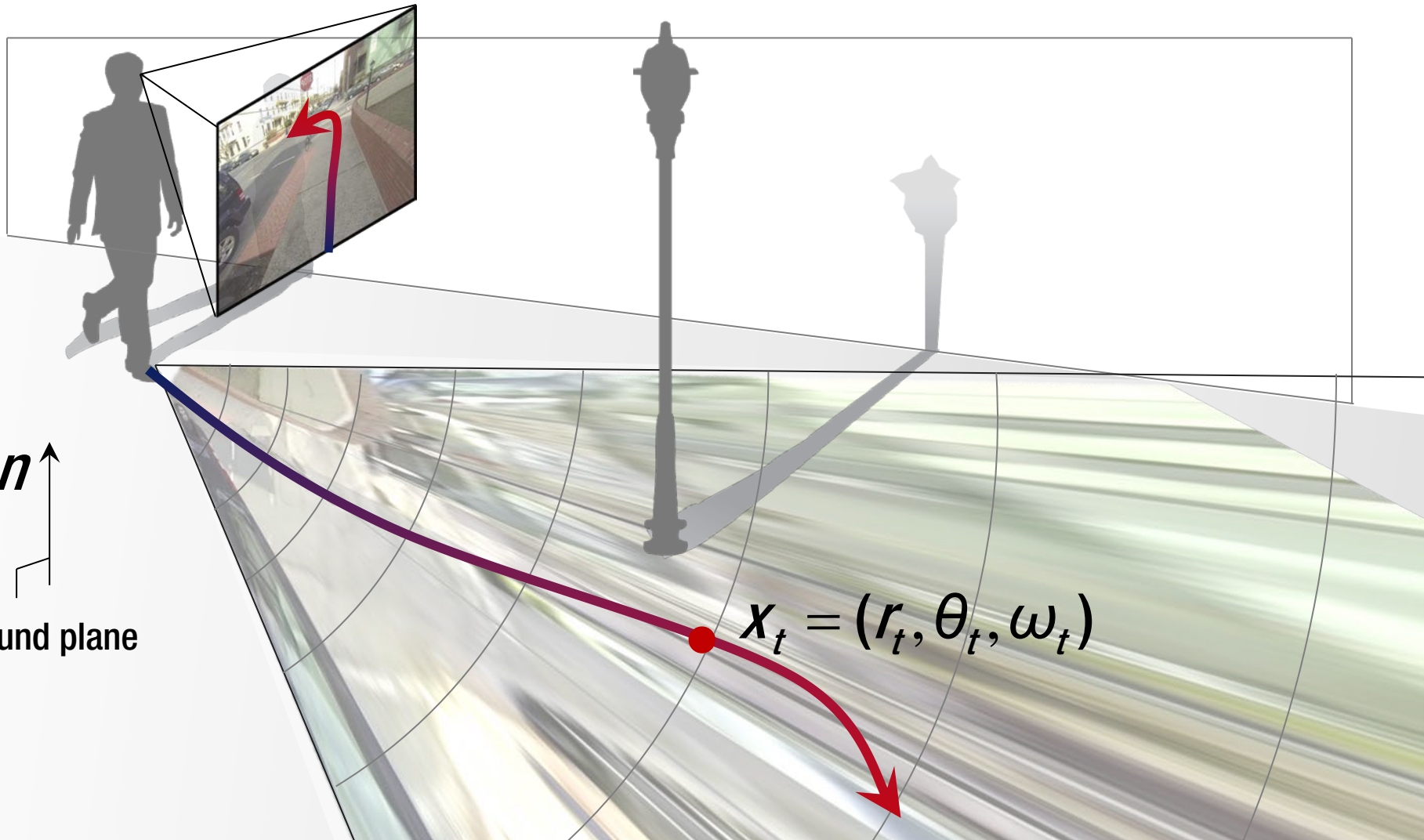
- 1280x960 stereo (100mm baseline, ~15m depth resolution)
- 26 scenes (13 indoor, 13 outdoor)
- 65.5k frames (9.1 hours) via SfM

The trajectory is projected onto the ground plane and its time is color coded.

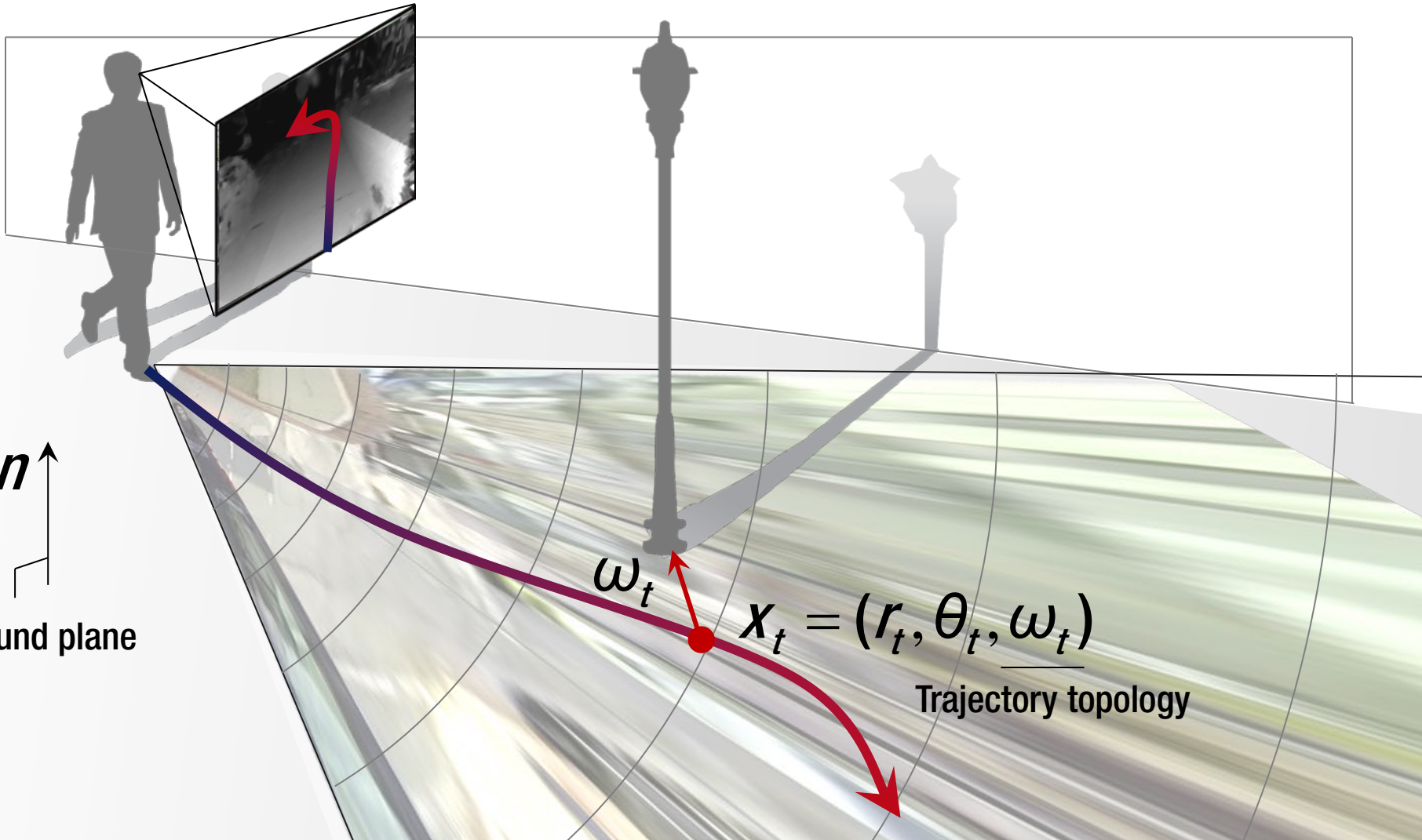
n
Ground plane



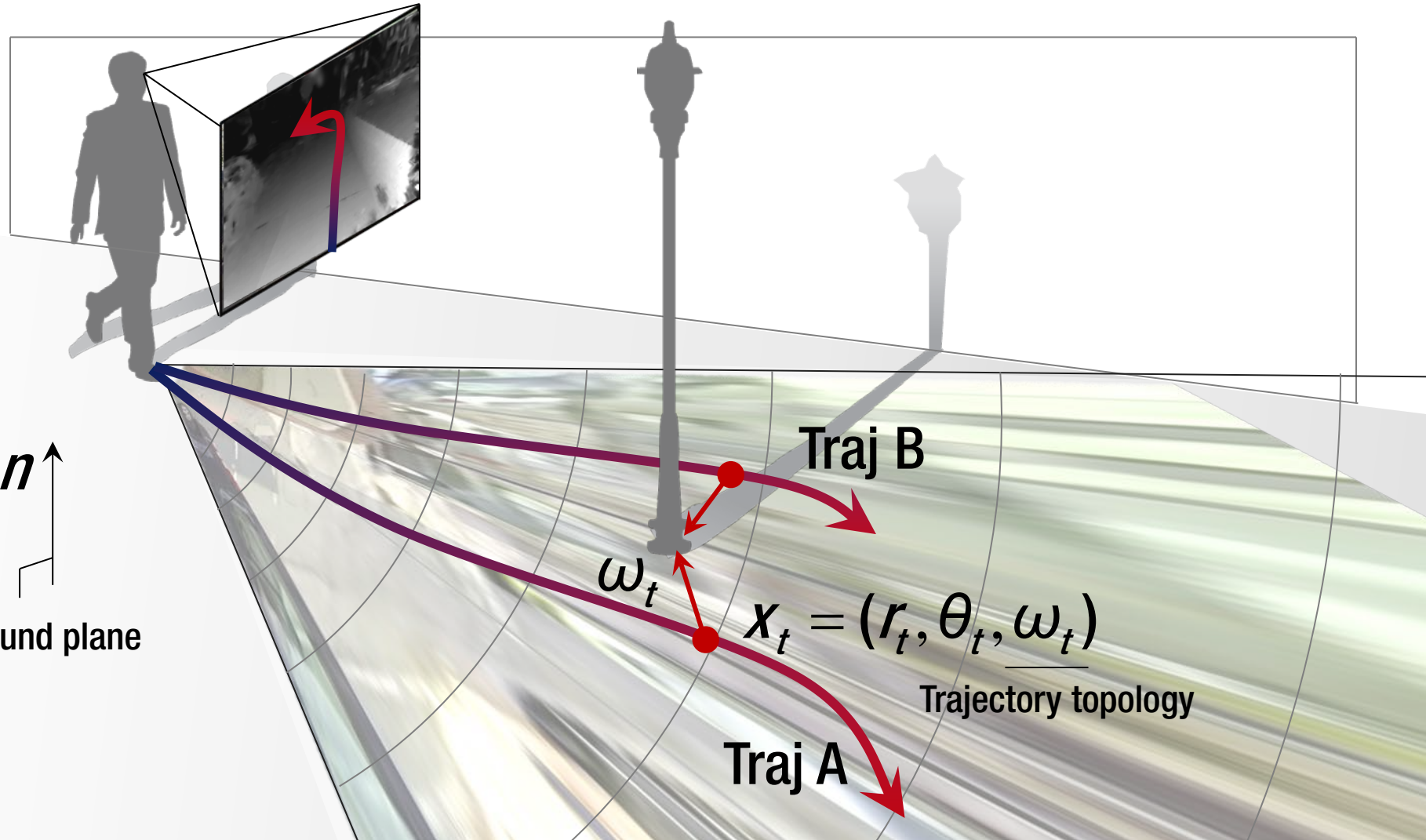
n
Ground plane



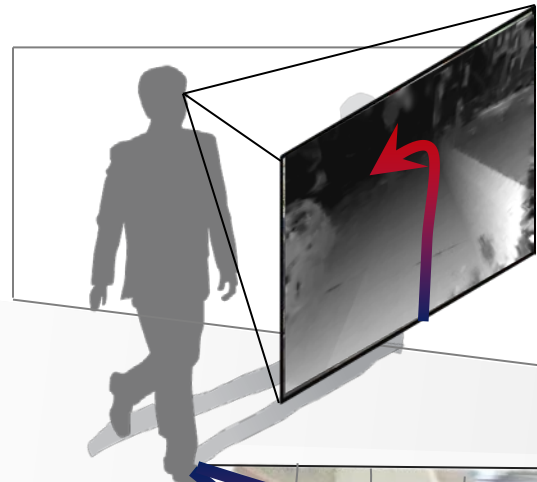
n
Ground plane



n
Ground plane



n
Ground plane



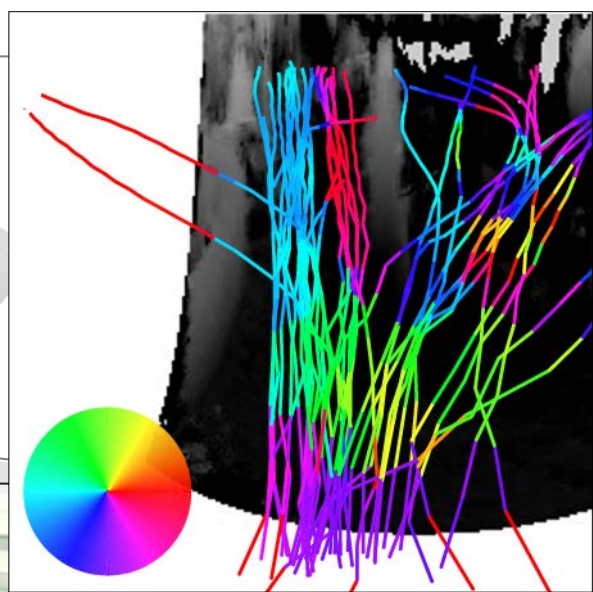
Traj B

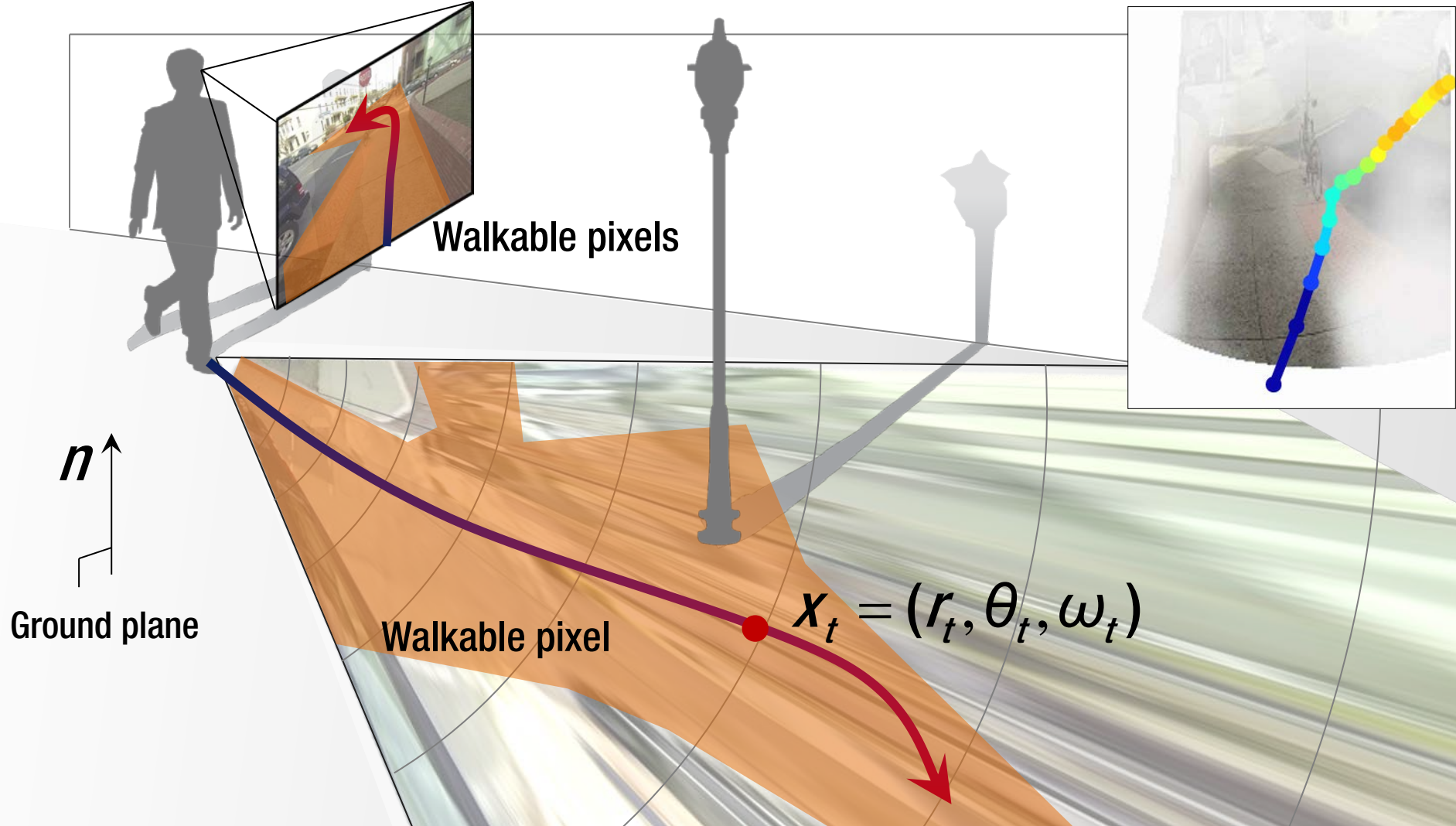
ω_t

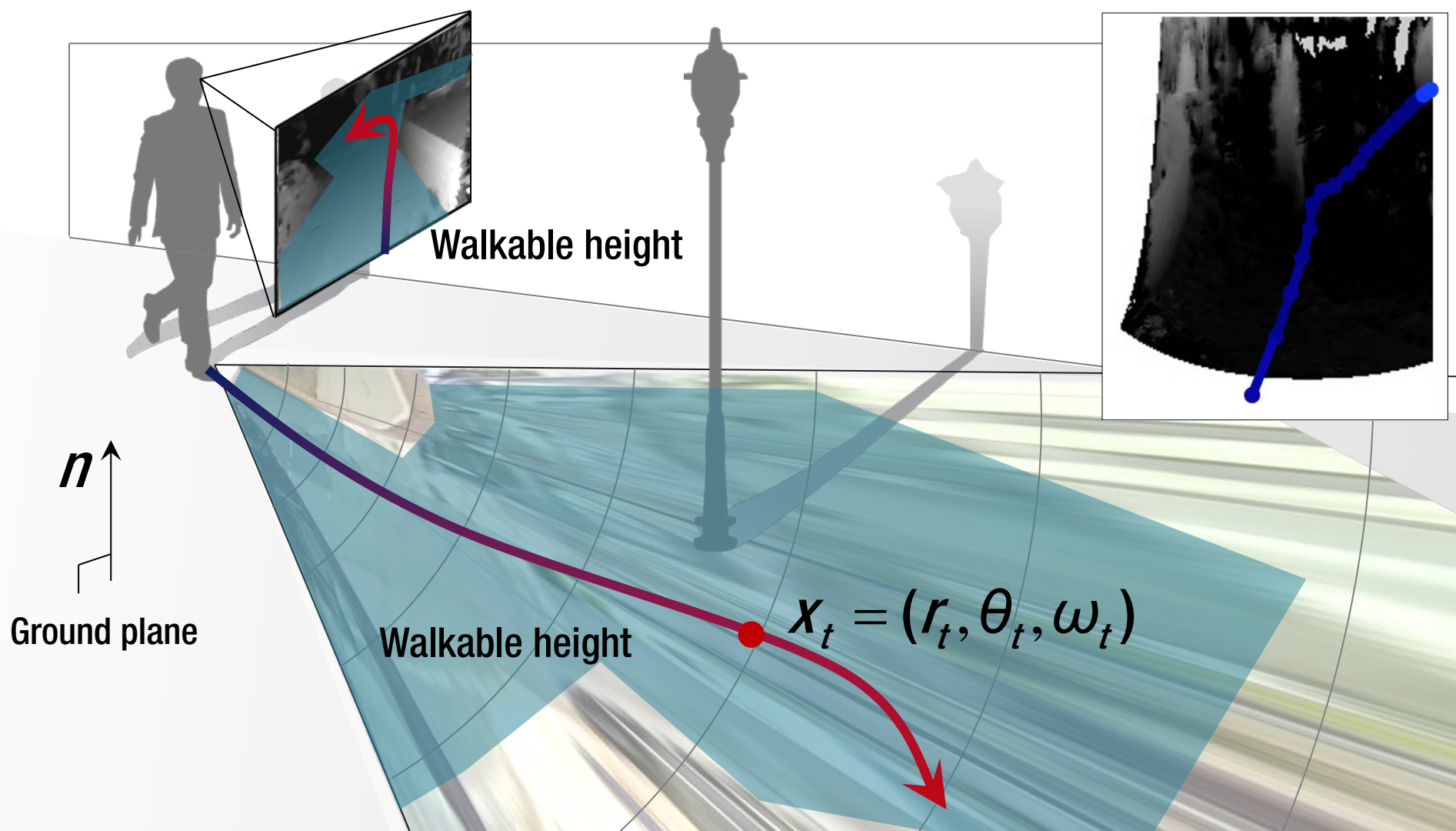
$x_t = (r_t, \theta_t, \omega_t)$

Traj A

Trajectory topology





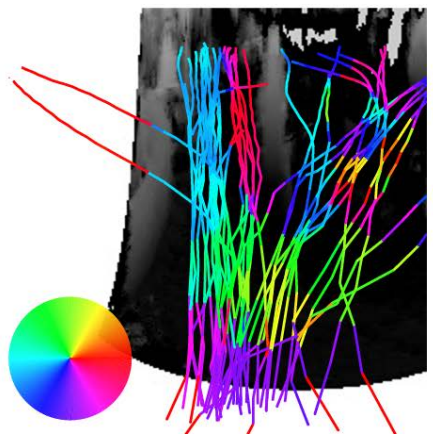




Testing image



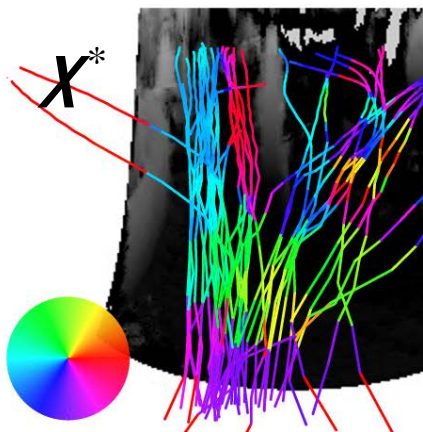
Testing image



Trajectory retrieval



Testing image



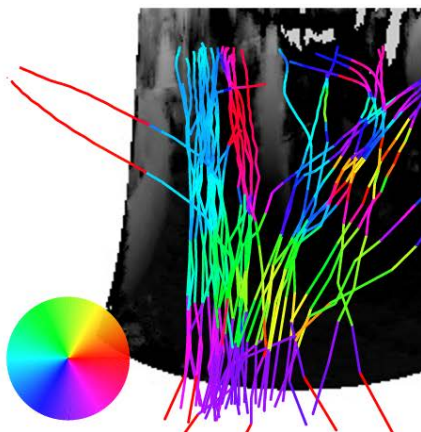
Trajectory retrieval

$$\underset{X}{\text{minimize}} E_D + E_{\text{RGB}} + \lambda \frac{\|X - X^*\|^2}{\text{Data cost}}$$

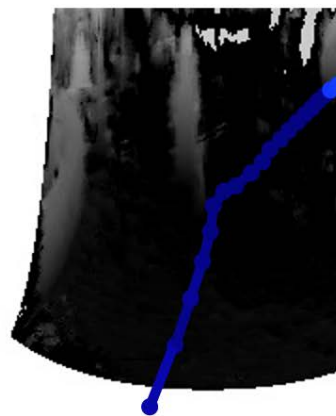
X^* : retrieved trajectory



Testing image



Trajectory retrieval



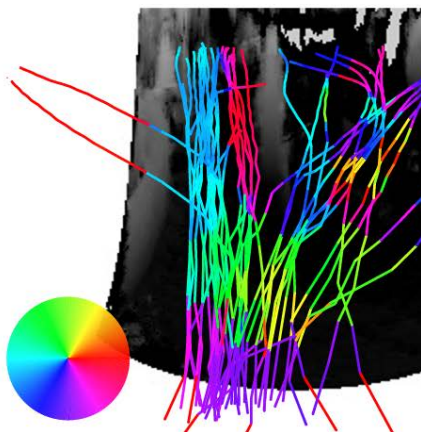
Depth cost

$$\underset{X}{\text{minimize}} \quad \underbrace{E_D + E_{\text{RGB}}}_{\text{Depth walking preference}} + \lambda \|X - X^*\|^2$$

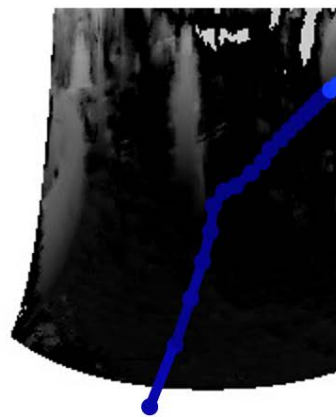
X^* : retrieved trajectory



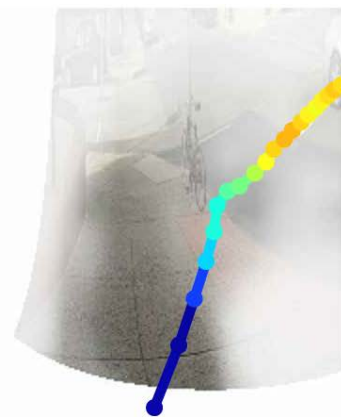
Testing image



Trajectory retrieval



Depth cost

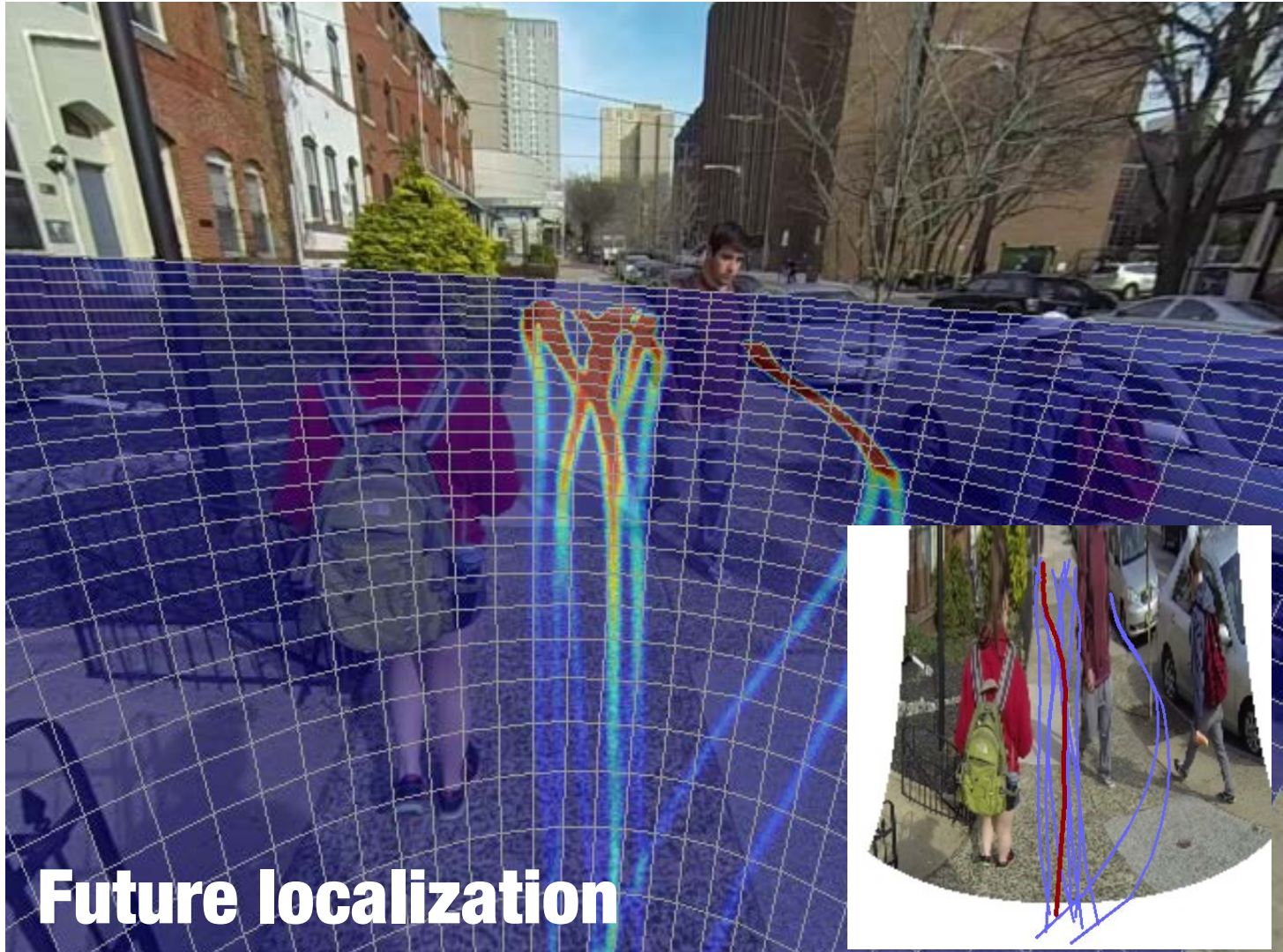


RGB cost

$$\underset{X}{\text{minimize}} \ E_D + \underline{E_{\text{RGB}}} + \lambda \left\| X - X^* \right\|^2$$

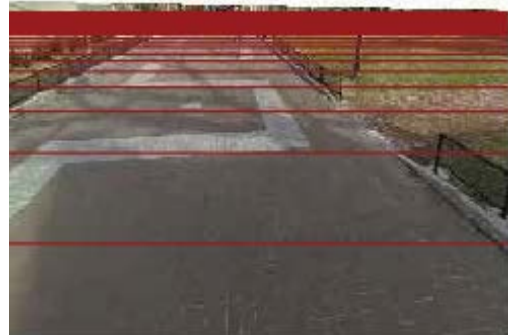
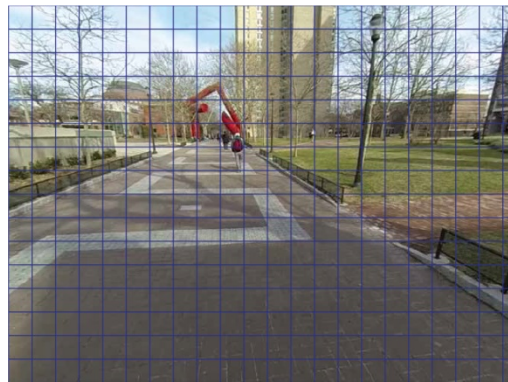
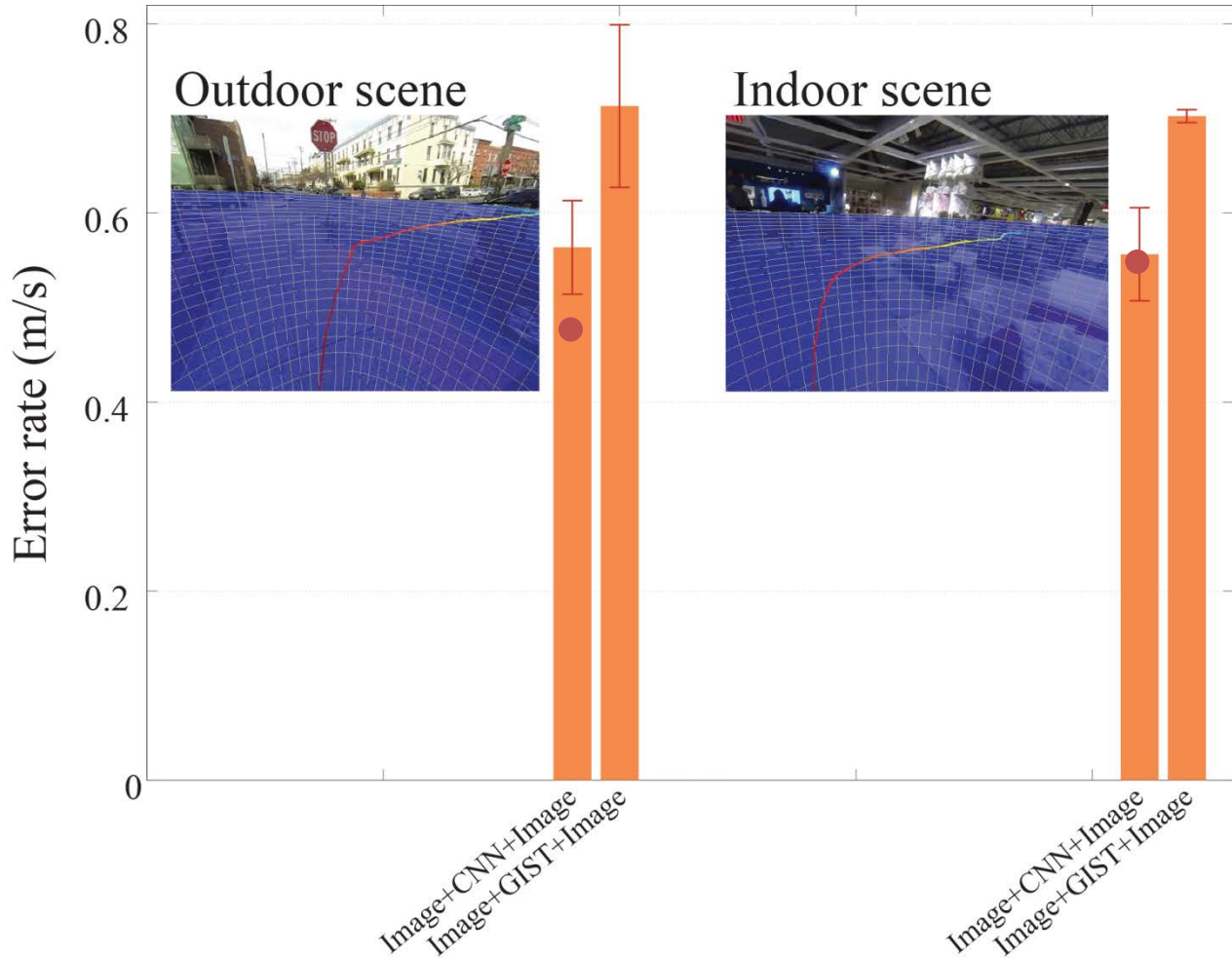
Walking preference

X^* : retrieved trajectory

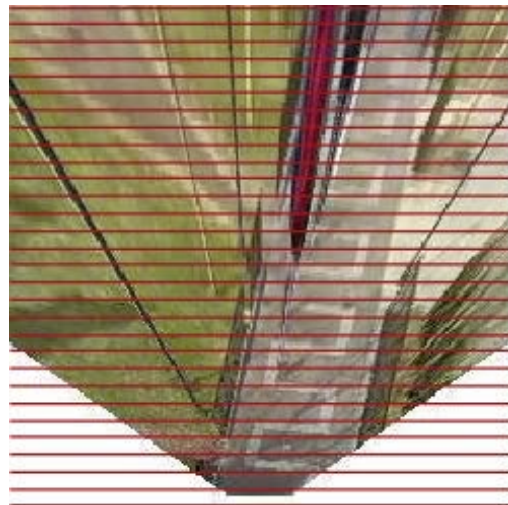
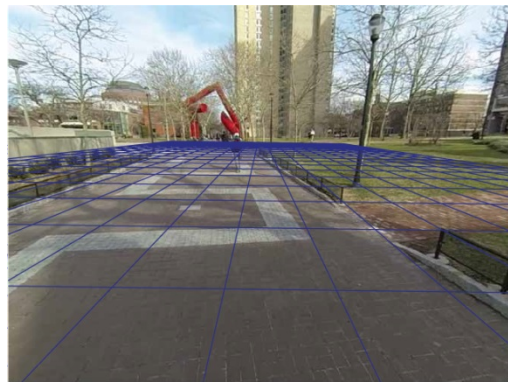
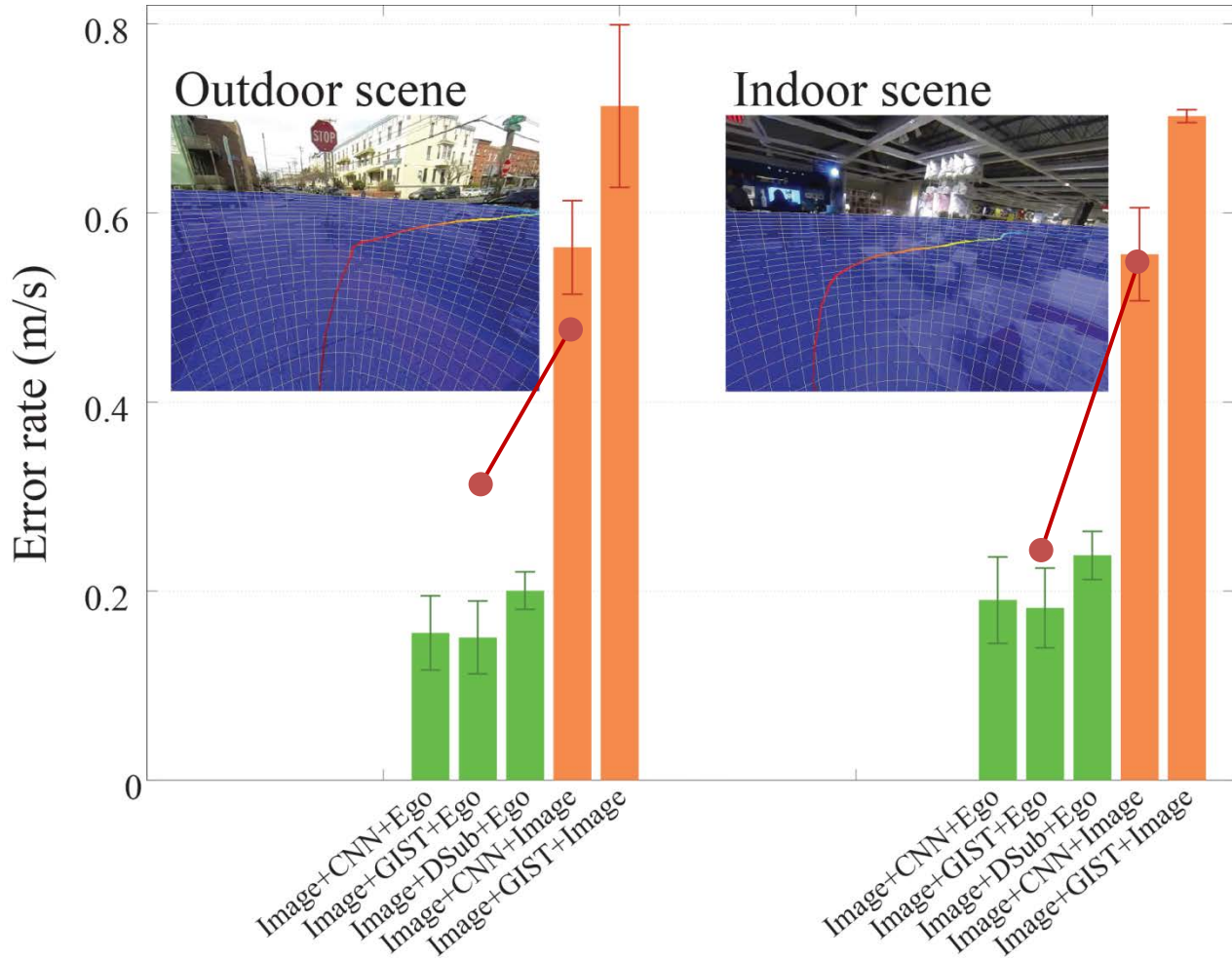


Future localization

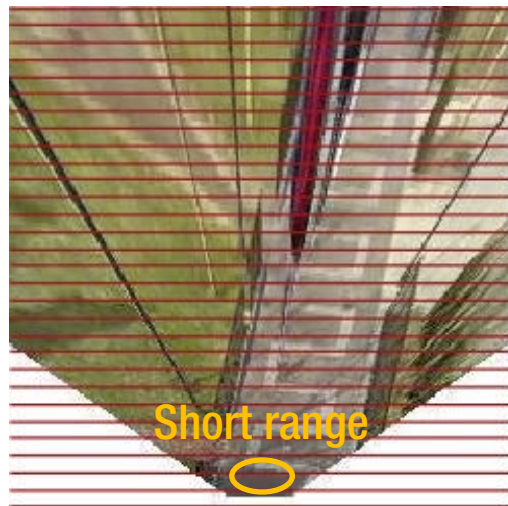
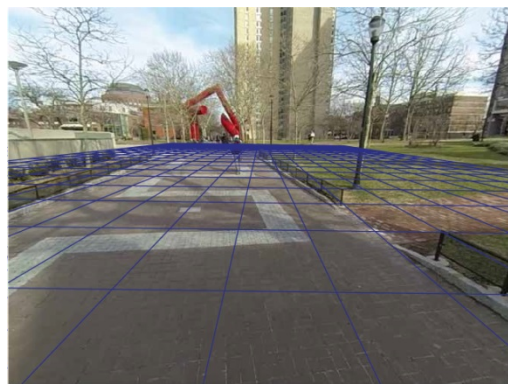
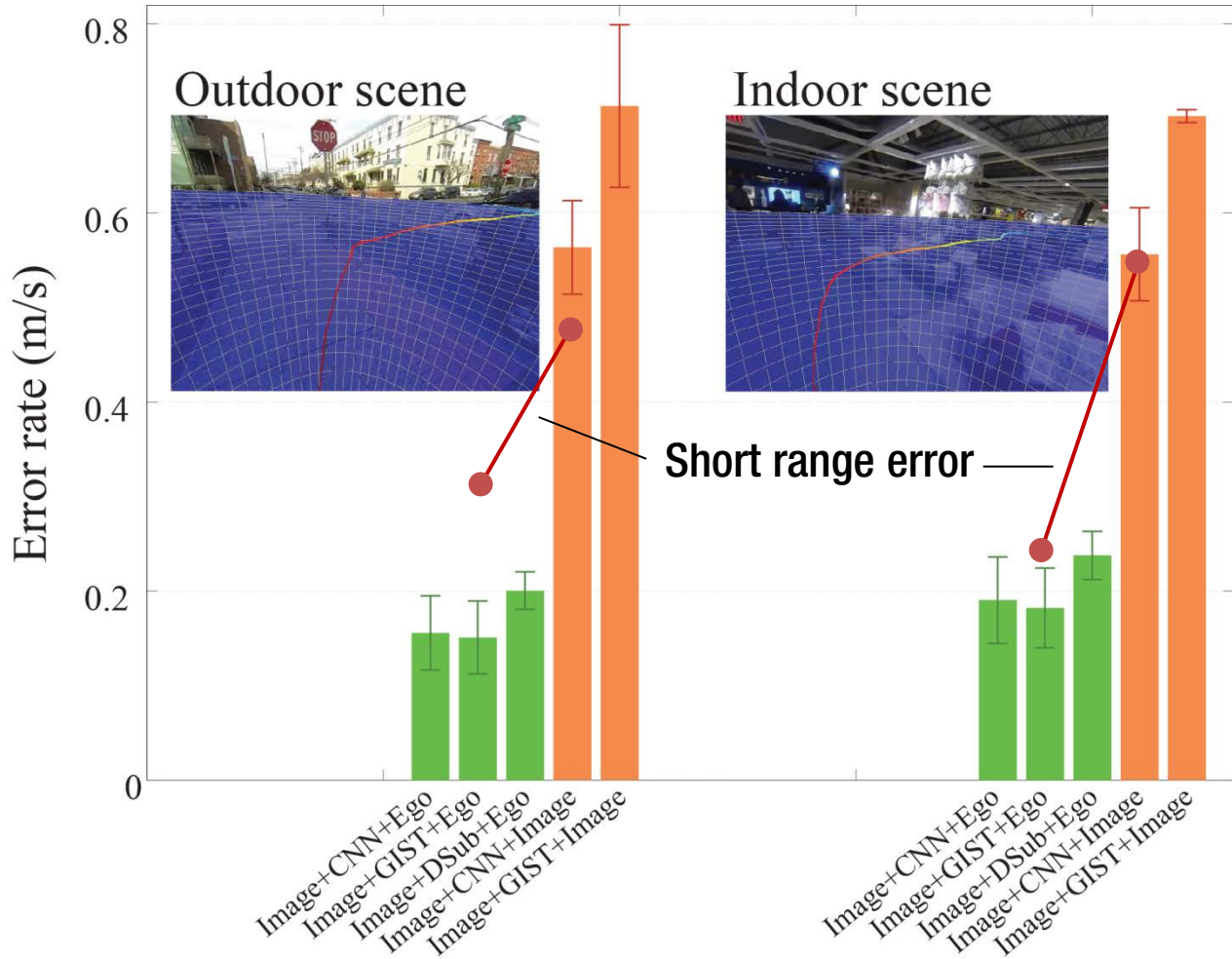




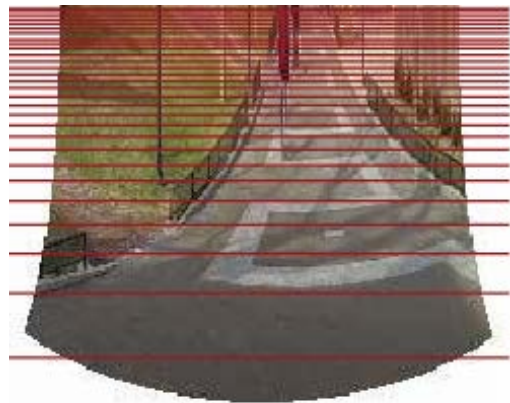
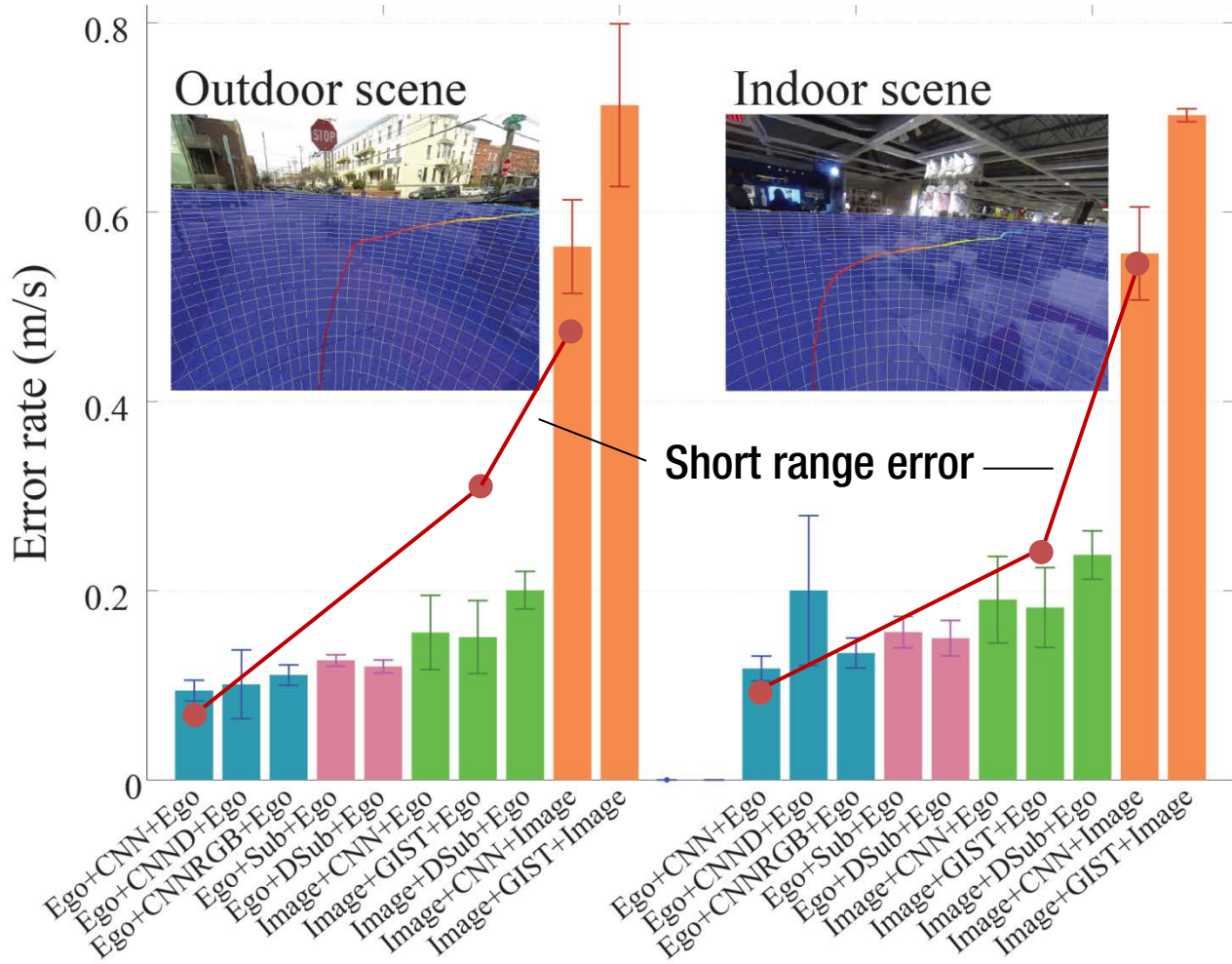
Cartesian in image (LeCun et al.)



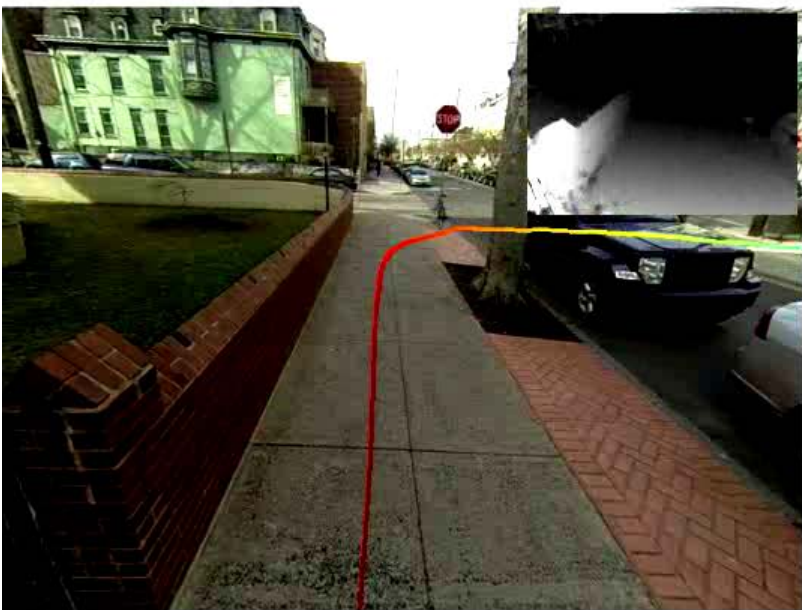
Cartesian on ground plane



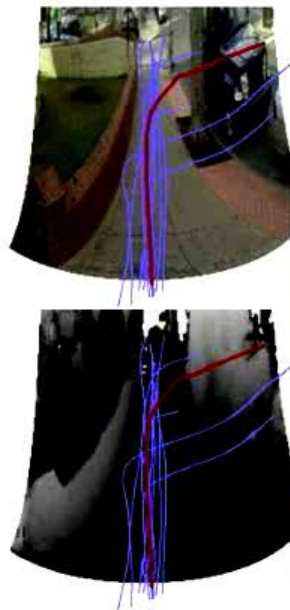
Cartesian on ground plane



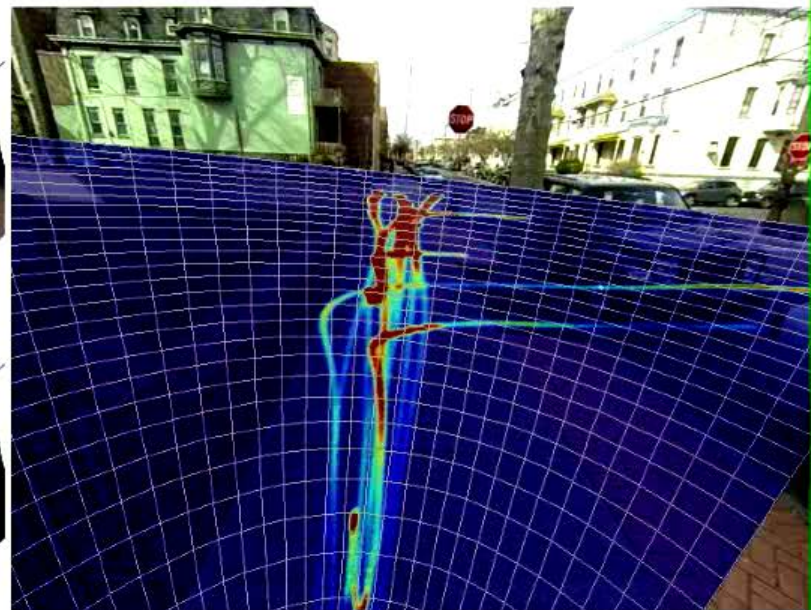
Ours



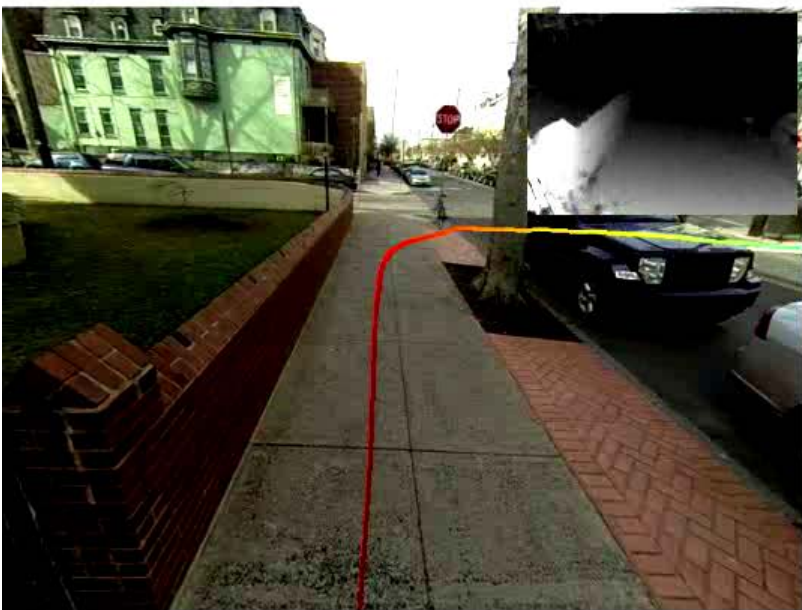
Ground truth



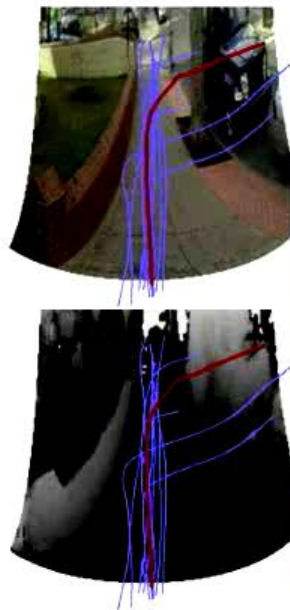
EgoRetinal map



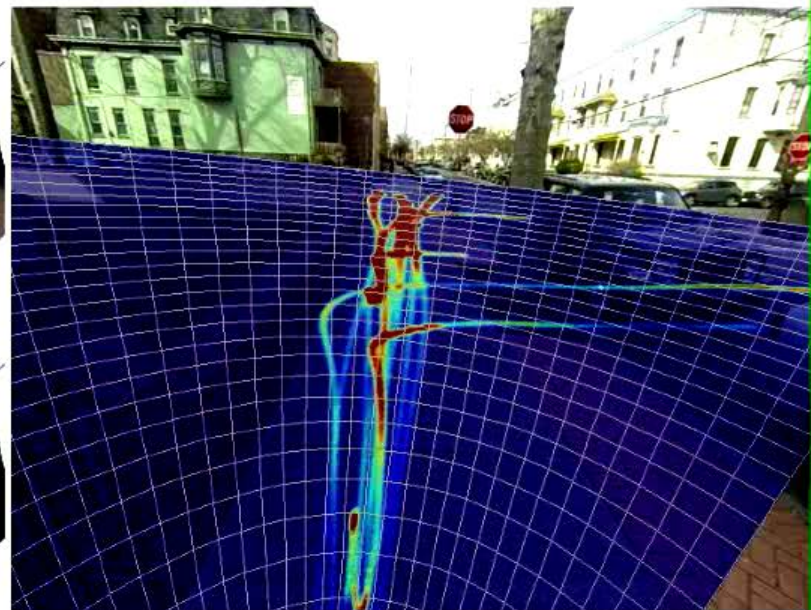
Predicted trajectories



Ground truth



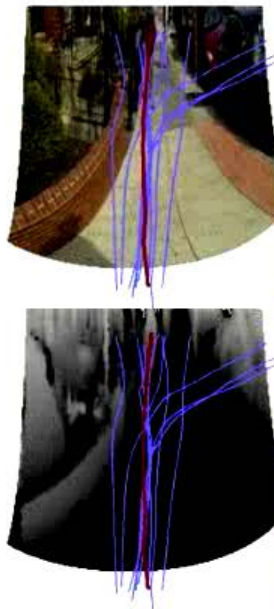
EgoRetinal map



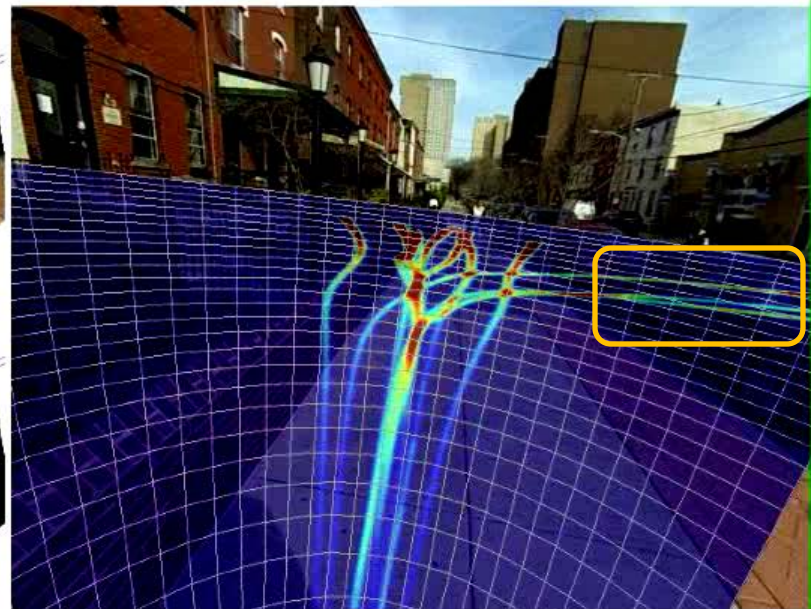
Predicted trajectories



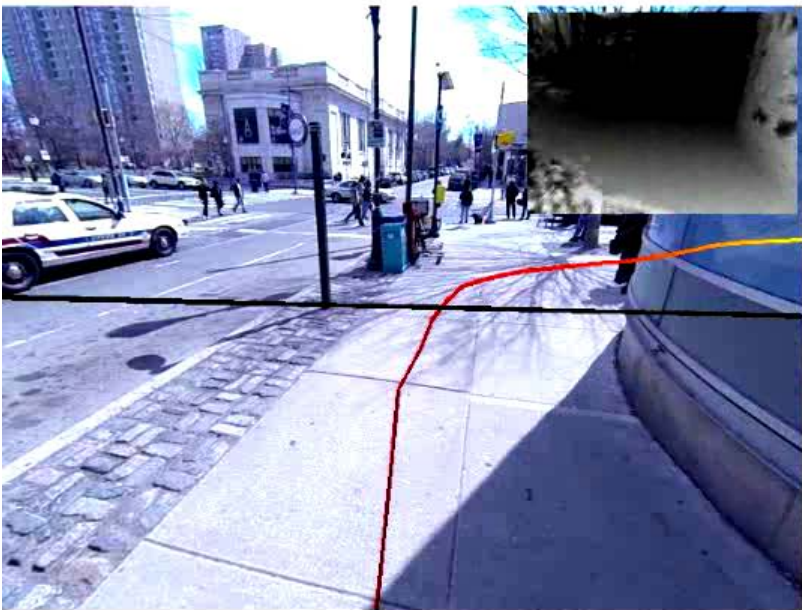
Ground truth



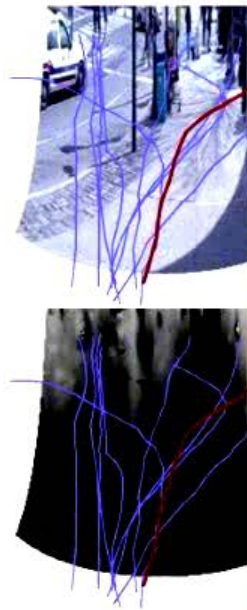
EgoRetinal map



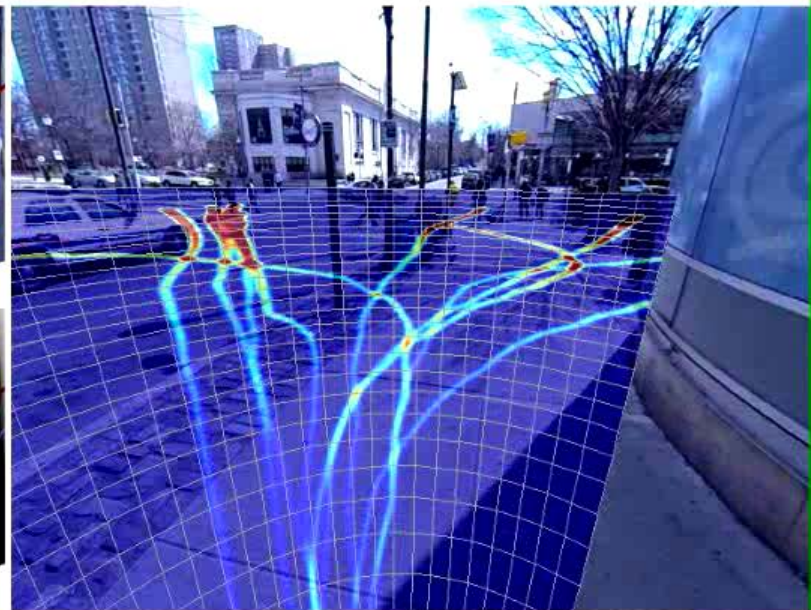
Predicted trajectories



Ground truth



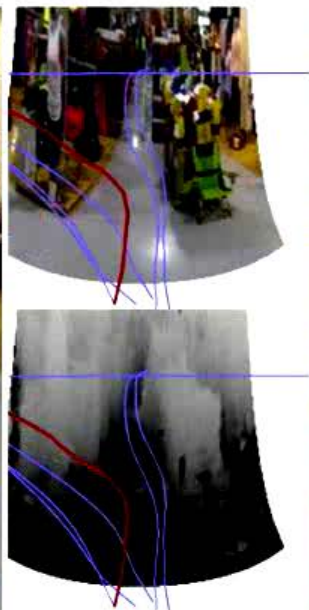
EgoRetinal map



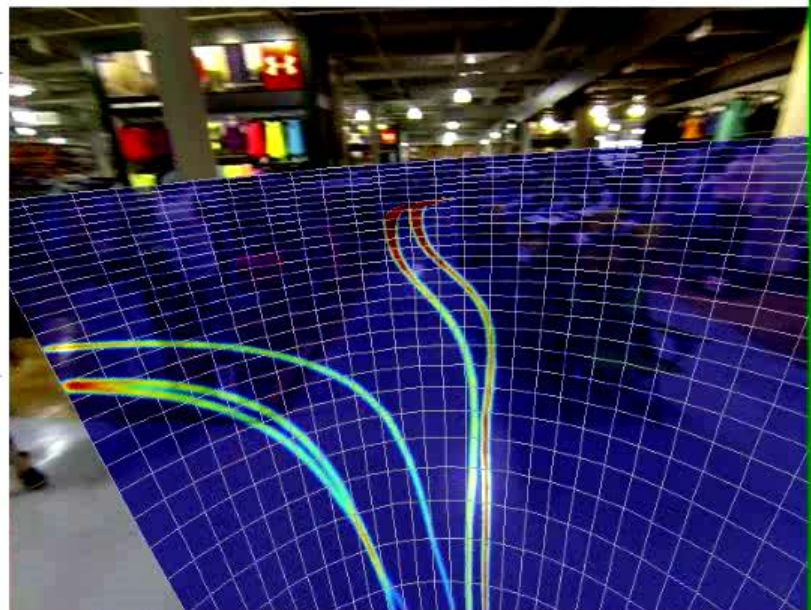
Predicted trajectories



Ground truth



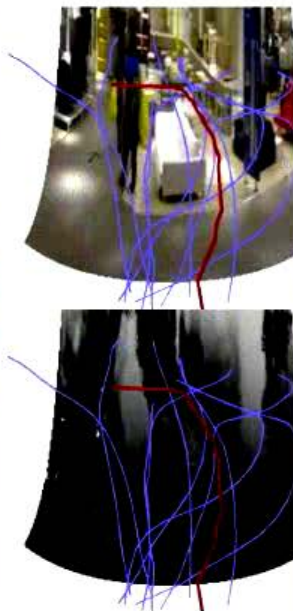
EgoRetinal map



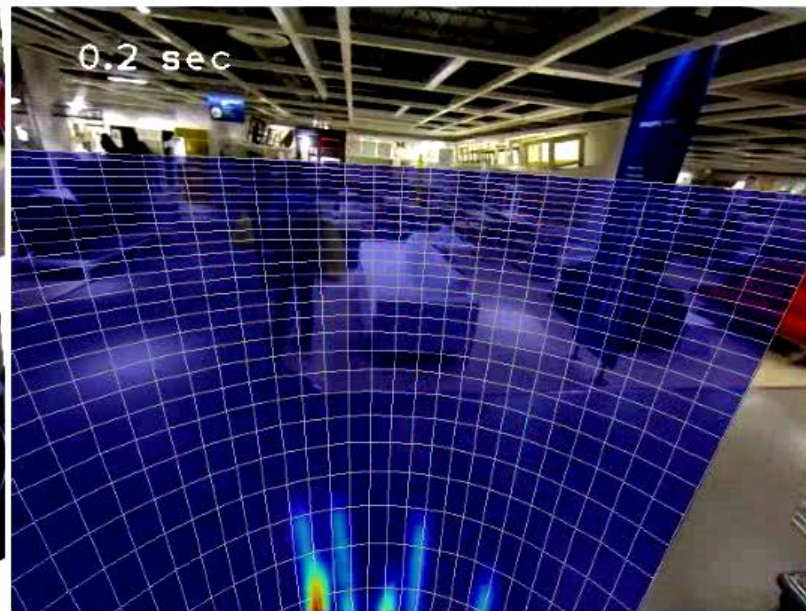
Predicted trajectories



Ground truth



EgoRetinal map

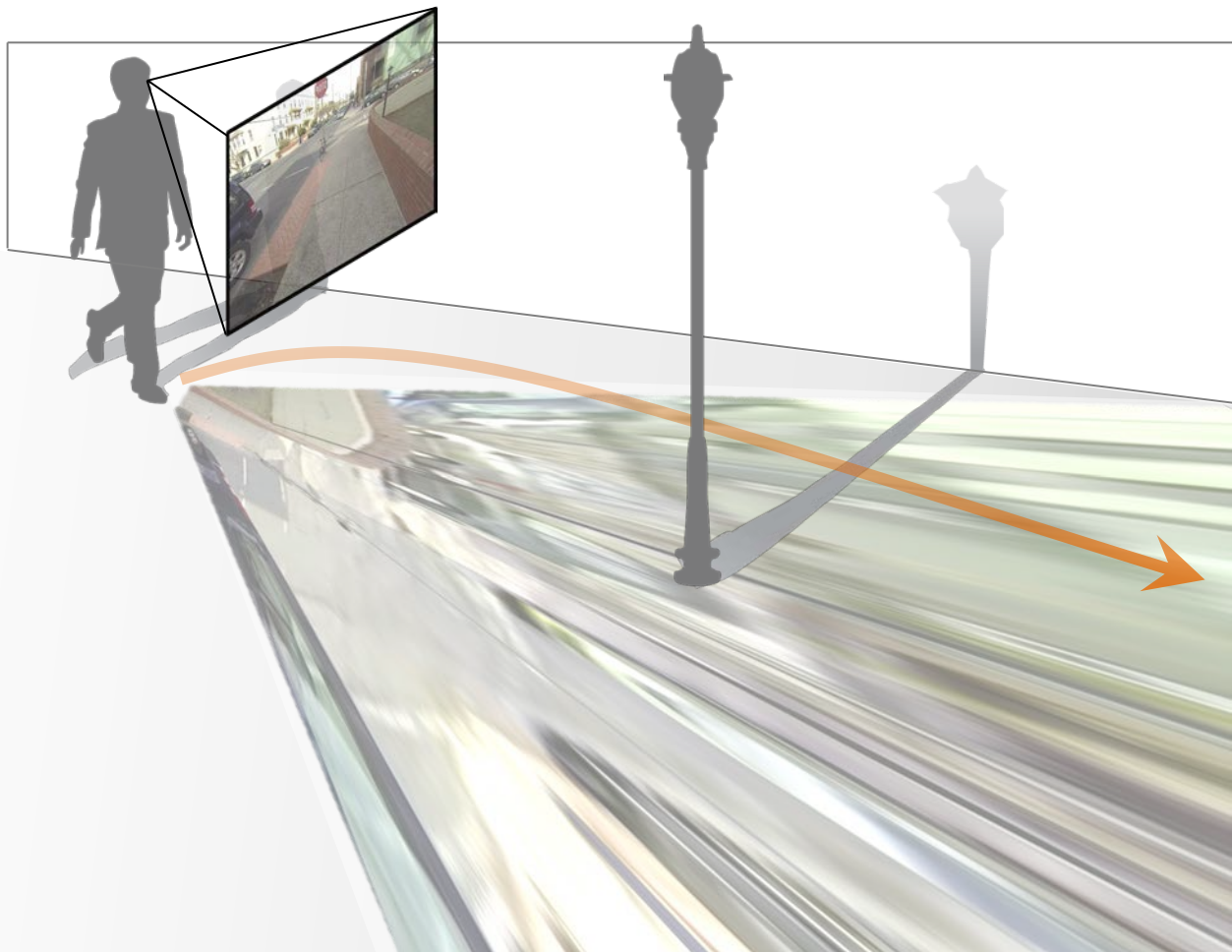


Predicted trajectories

Putting yourself in his shoes



EgoRetinal map



Egocentric Future Localization

Website: http://www.seas.upenn.edu/~hypar/future_loc.html

