Motor Focus: Fast Ego-Motion Prediction for Assistive Visual Navigation

Hao Wang, Jiayou Qin, Xiwen Chen, Ashish Bastola, John Suchanek, Zihao Gong, and Abolfazl Razi

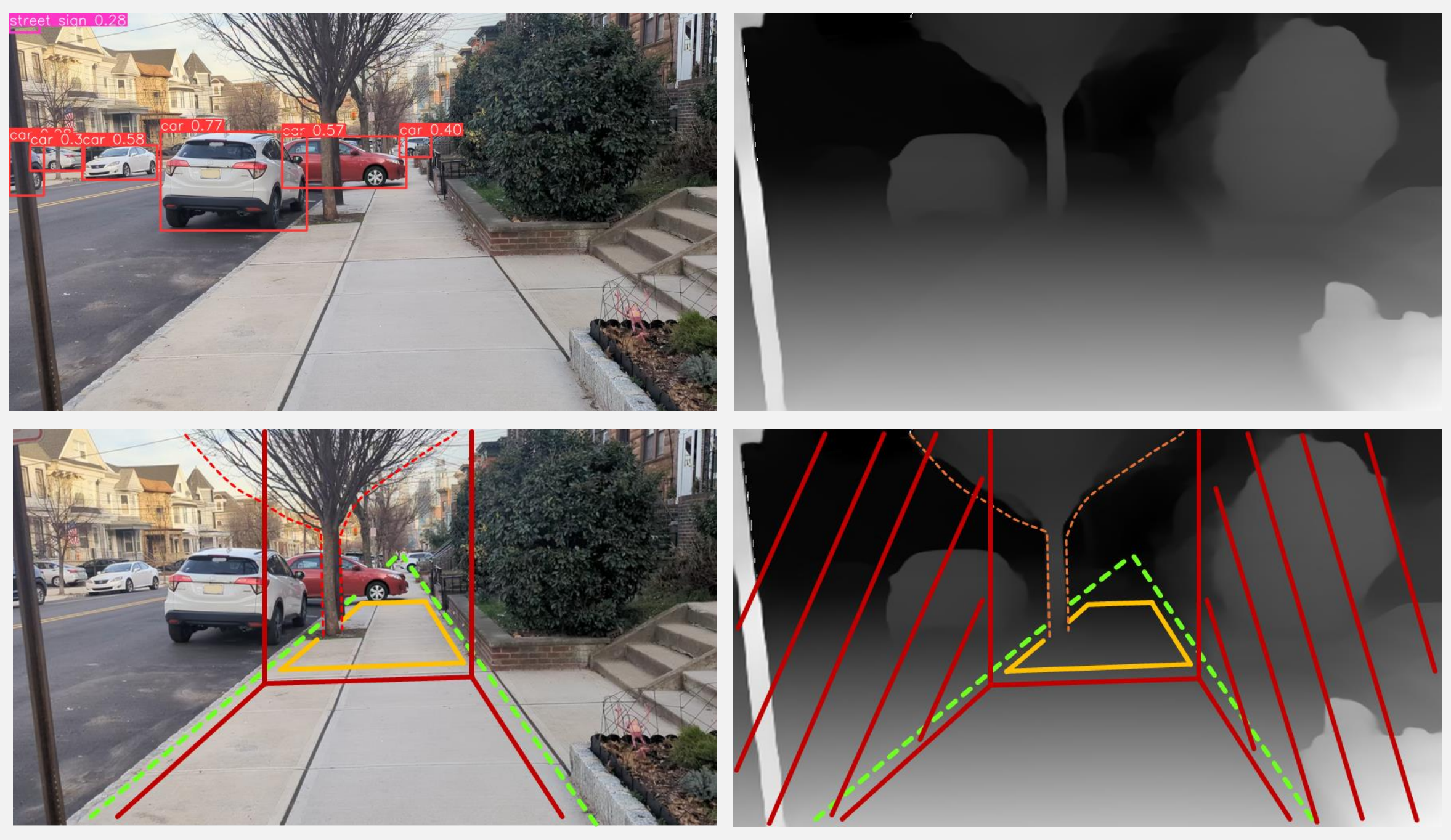
Abstract

Assistive visual navigation systems for visually impaired individuals have become increasingly popular thanks to the rise of mobile computing.
Most of these devices work by translating visual information into voice commands. In complex scenarios where multiple objects are present, it is imperative to prioritize object detection and provide immediate notifications for key entities in specific directions. This brings the need for identifying the observer's motion direction (ego-motion) by merely processing visual information, which is the key contribution of this project.

Motor Focus, a lightweight image-based framework that predicts the ego-motion—the **humans**' (and humanoid machines') **movement intentions** based on their visual feeds, while filtering out camera motion without any camera calibration.

Current Challenge

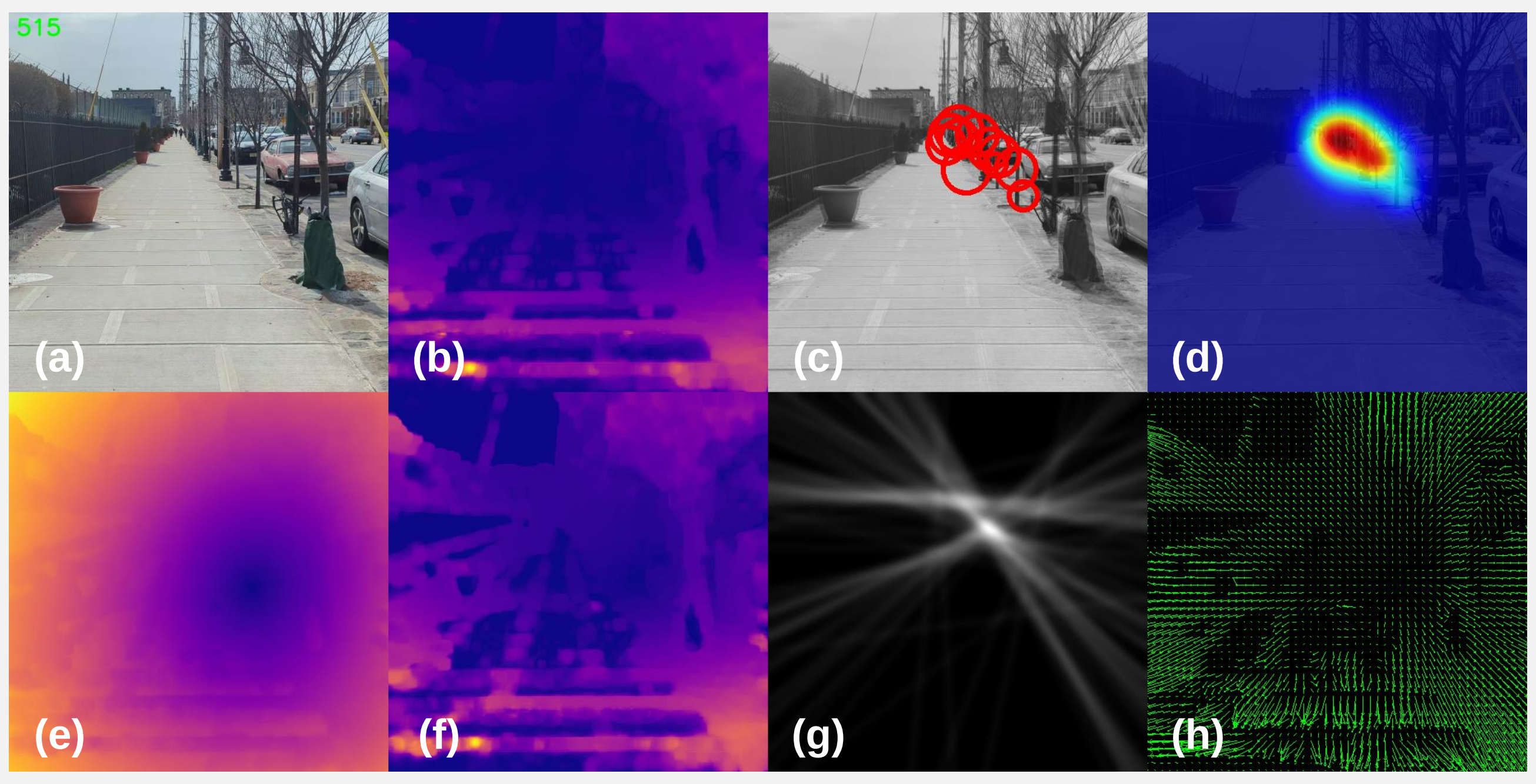
General object detection models report all detected objects for users, widely used in various applications.



But for **visually impaired individuals**, prioritize object detection and provide immediate notifications for key entities (hazards) in specific directions is necessary.

Our Solution

Motor Focus: a novel framework for predicting how users physically **move** and **orient** themselves in space. Specifically, we introduce an **optical flow-based** pixel-wise temporal analysis that can predict the **movement direction** of users and simultaneously **filter** out the unintended and noise-like **camera motion** without any camera calibration.



(a) is the raw RGB image, **(b)** is the compensated optical flow, **(c)** is the attention area of 10 consecutive frames, **(d)** is the attention map, aggregated by the gaussian distributions of 10 recent frames, **(e)** is the camera noise, **(f)** is the original optical flow, **(g)** is the probability map of focus center, and **(h)** is the compensated optical flow field.

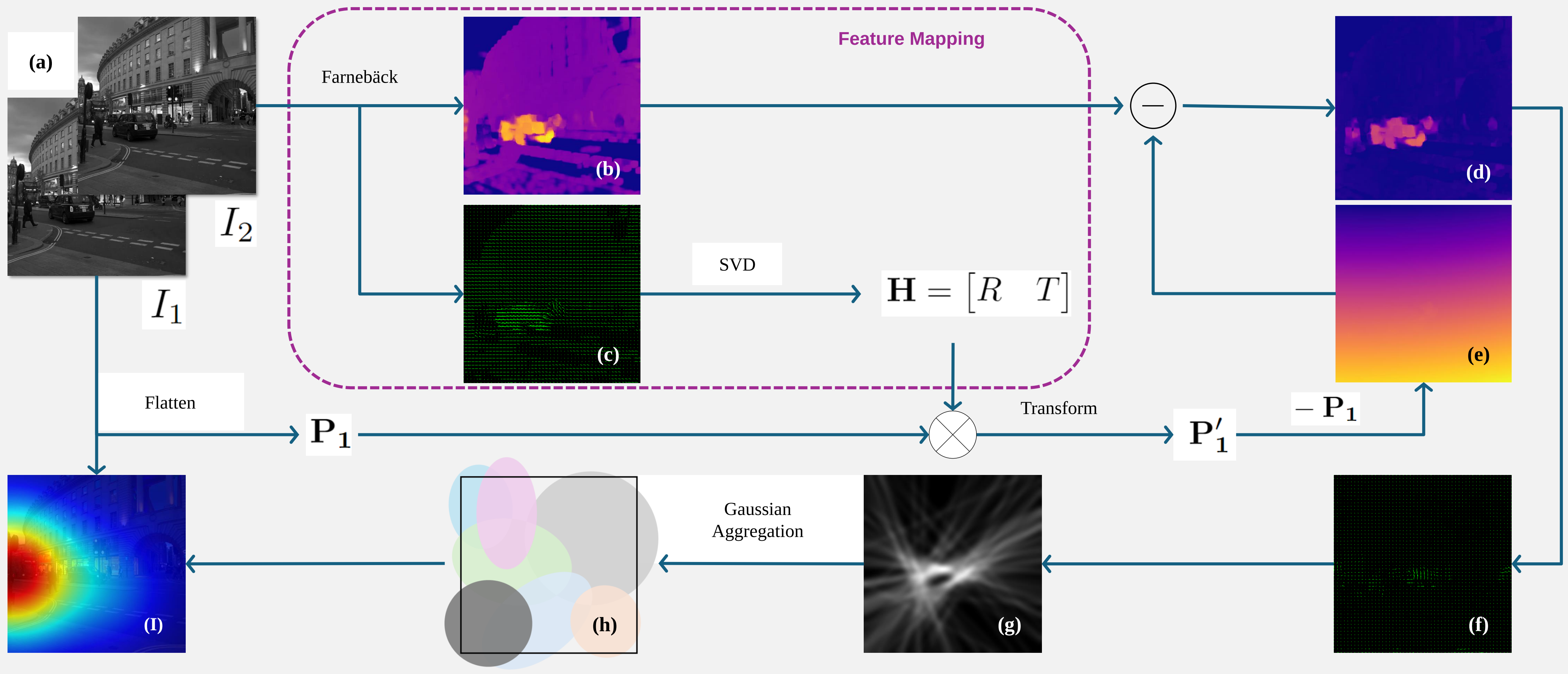
Collected Dataset

Assistive visual navigation we collected a dataset that is specialized for visual navigation including **50 clips** of various scenarios. In advance, each video clip is observed by three researchers frame by frame, and a pixel location (x, y) of moving direction is annotated for each frame.

This figure shows the sample of the collected dataset, where colored points represent the annotation from different researchers. The ground truth is calculated by the average of three different pixel locations. Specifically, (a) is a **biking** scene, (b) is a scooter **riding** scene, (c) (d) is **walking** scenes.

Framework

We implement an **optical flow-based** pixel-wise temporal analysis method to compensate for the camera motion with a **Gaussian aggregation** to smooth out the movement prediction area. We apply Singular Vector Decomposition (**SVD**) instead of classical feature mapping to reduce the computation load.

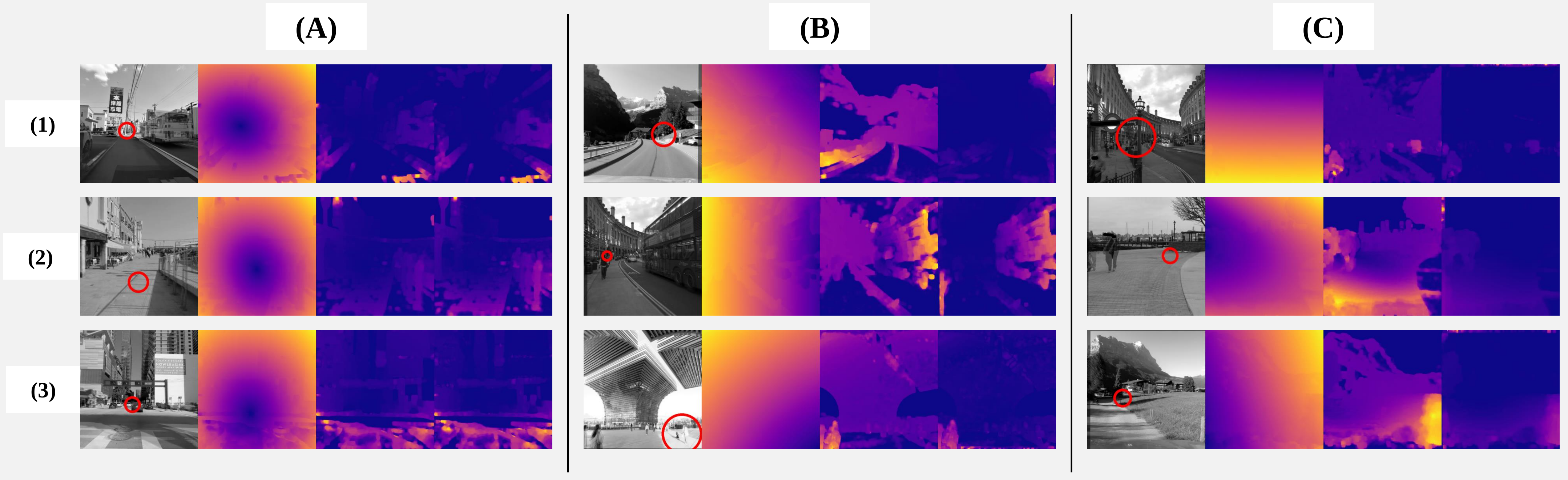


The proposed framework as shown above, **(a)** is a two consecutive frame pair, **(b)** is the original optical flow map (magnitude), **(c)** is the original optical flow field (vector), **(d)** is the compensated optical flow map, **(e)** is the camera motion ϵ, **(f)** is the compensated optical flow field, **(g)** is the probability map of attention point for I2, **(h)** is the aggregated gaussian distribution of attention points from (g), and **(i)** is the attention map for motor focus of frame I2.

Experiments

We applied the proposed method on various videos including our collected clips and the online videos. The compensated optical flow can **filter** the motion caused by camera shifting, **distinguishing** the relatively **moving objects**. Meanwhile, nearby objects' **moving speed** and direction can also be estimated from the compensated optical flow map. For instance:

- In group **A**, when the camera moving direction and the user movement are aligned straightforwardly, the vanilla optical flow and the compensated optical flow become identical, and the camera motion map tends to overlap with the motor focus area
- In group **B**, the compensated optical flow and camera motion indicated the camera moving potential is different from the actual body movement.
  - In B1 and B3, the camera moves toward the left corner, while the user moves toward the right side.
  - In B2, both the camera and the user move toward the left.
- In group **C**, the camera in both C2 and C3 are turning right, while the user in C2 is moving right and the user in C3 is moving left.
  - In C1, the user stands statically, while the camera motion suggests that the camera is slowly pitching up.



This figure shows the **Visualization** of ego-motion compensation, each image consists of four cells, from left to right: grayscale image with predicted moving direction, the magnitude of camera motion (ego-motion), raw optical flow (vanilla dense-optical flow), and optical flow with ego-motion compensation.

Results

We applied the proposed method to predict the center of **moving direction** and calculate the Mean Absolute Error (**MAE**) and Mean Squared Error (**MSE**) to compare with the annotated ground truth. Furthermore, we used the Signal-to-Noise Ratio (**SNR**) to compare the **vertical motion** along the time in a selected scene to test the **video stabilization** performance.

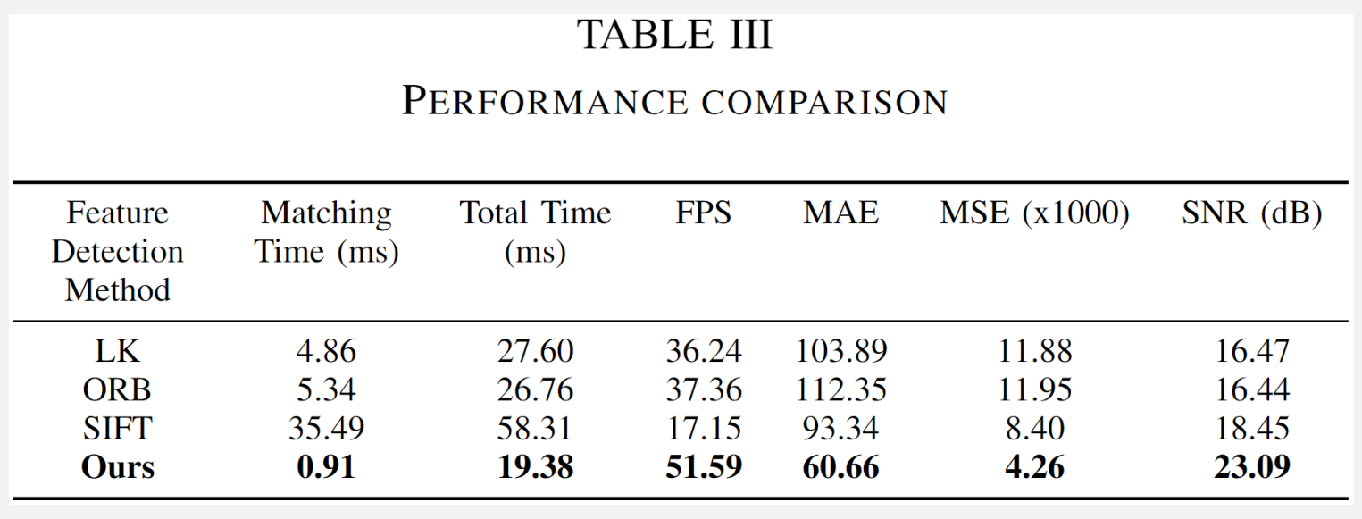
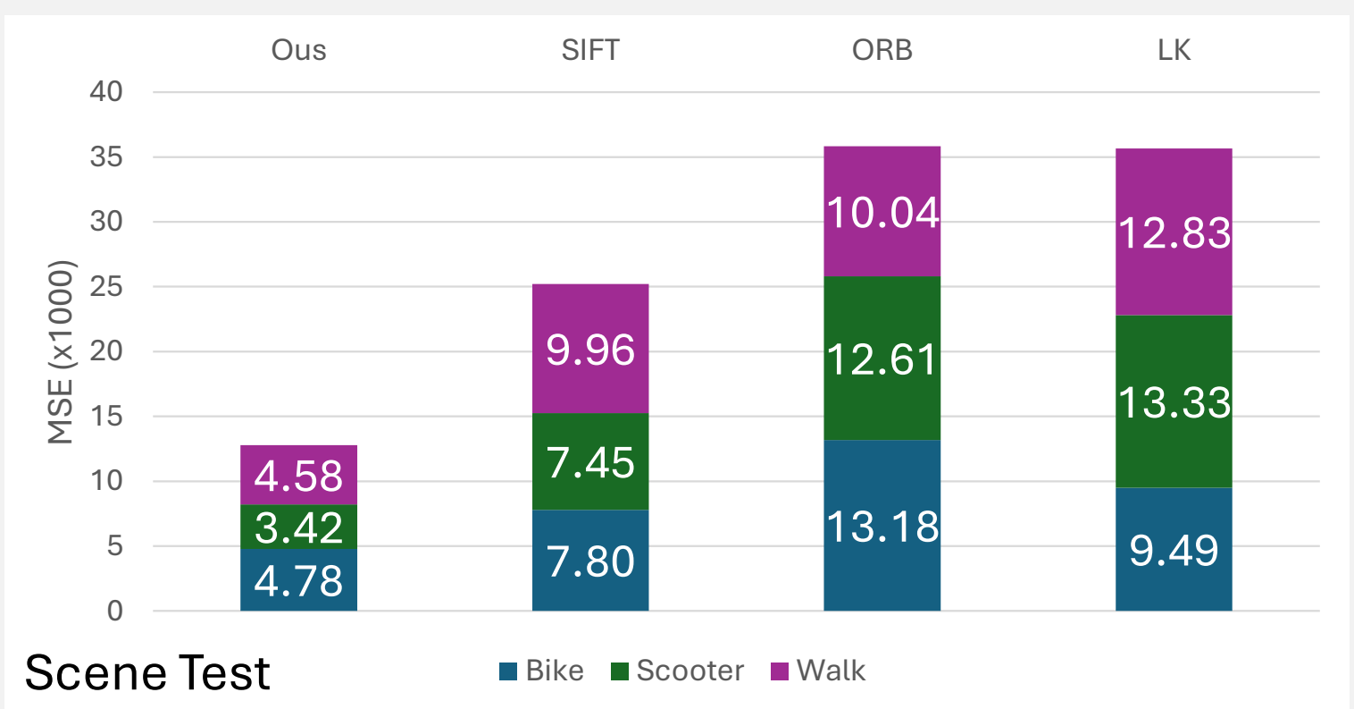
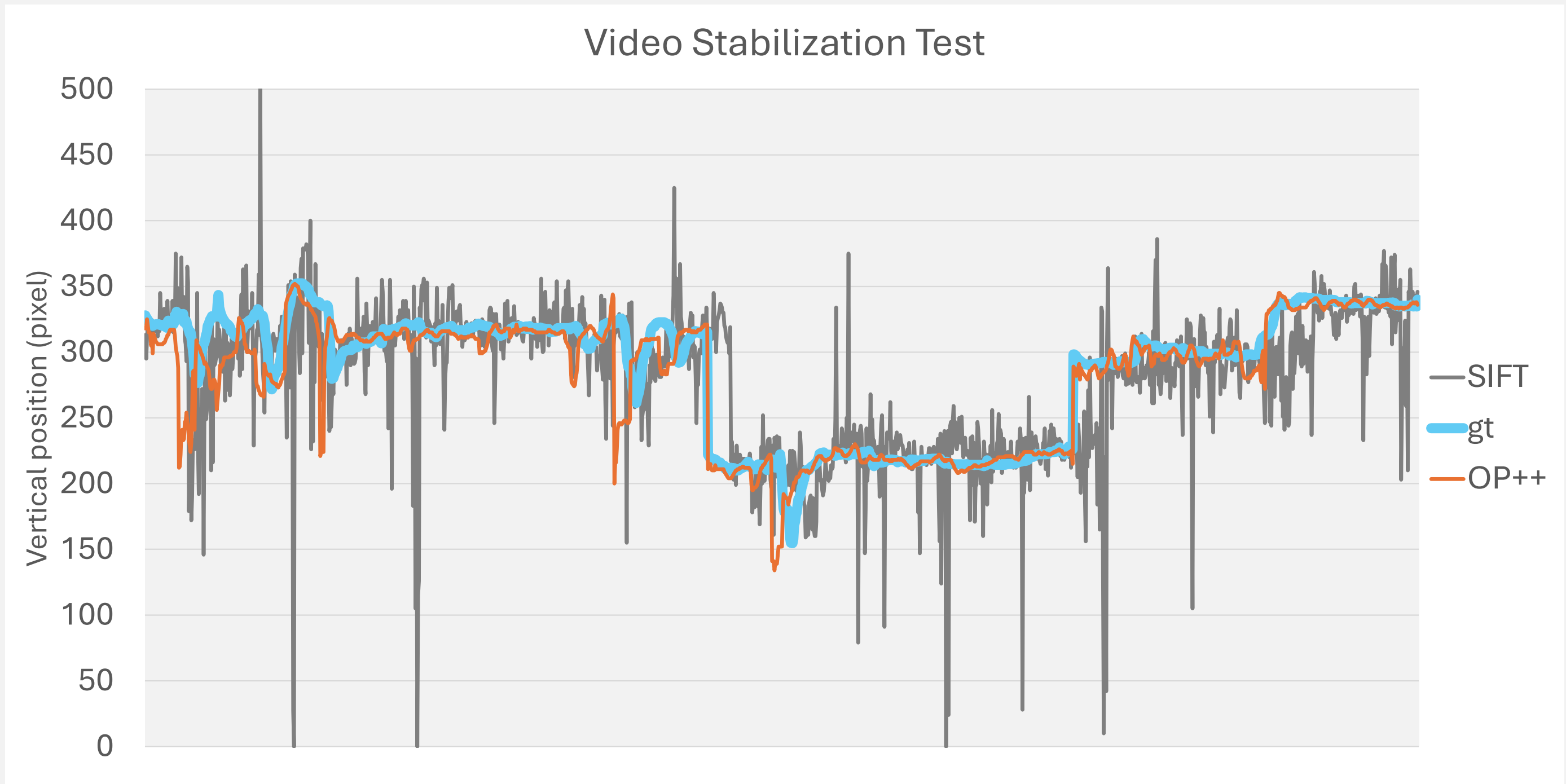


TABLE III
PERFORMANCE COMPARISON

Feature Detection Method	Matching Time (ms)	Total Time (ms)	FPS	MAE	MSE (x1000)	SNR (dB)
LK	4.86	27.60	36.24	103.89	11.88	16.47
ORB	5.34	26.76	37.36	112.35	11.95	16.44
SIFT	35.49	58.31	17.15	93.34	8.40	18.45
Ours	**0.91**	**19.38**	**51.59**	**60.66**	**4.26**	**23.09**





In summary, our method obtained: **(1)** the lowest prediction error of motor focus in both **MAE** and **MSE**, **(2)** and highest **FPS** due to less matching time, and **(3)** the better **accuracy** (alignment with ground truth) and **stability** (fluctuations) compared with other methods like SIFT and ORB.

Project

Paper

Code

IEEE
BSN 2024
Oct. 15 – 17, Chicago, USA