

Conversation activity recognition using interaction video sequences in pedestrian groups

Introduction

We propose a method for recognizing conversation activity in a pedestrian group by extracting an interaction video sequence as a feature, which has high recognition accuracy and can be visually confirmed by human observers.

Conversation activity label

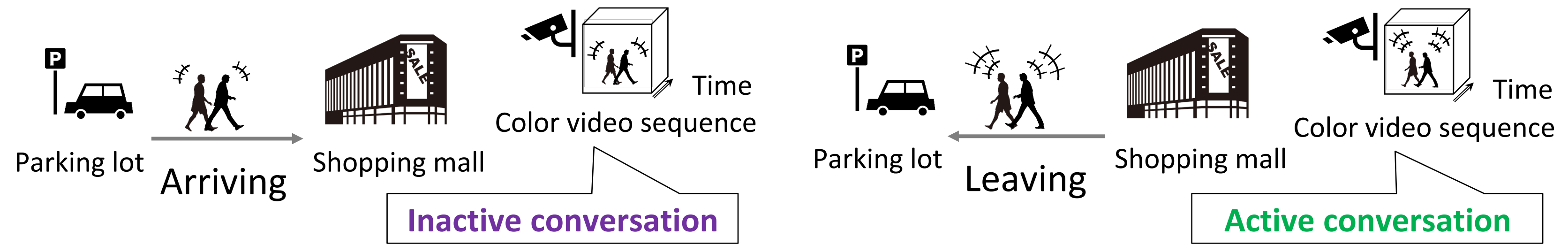
Active: The group is having a lively conversation on topics of mutual interest.

Inactive: The group is not having a lively conversation on topics of no interest.

None: No conversation is occurring.

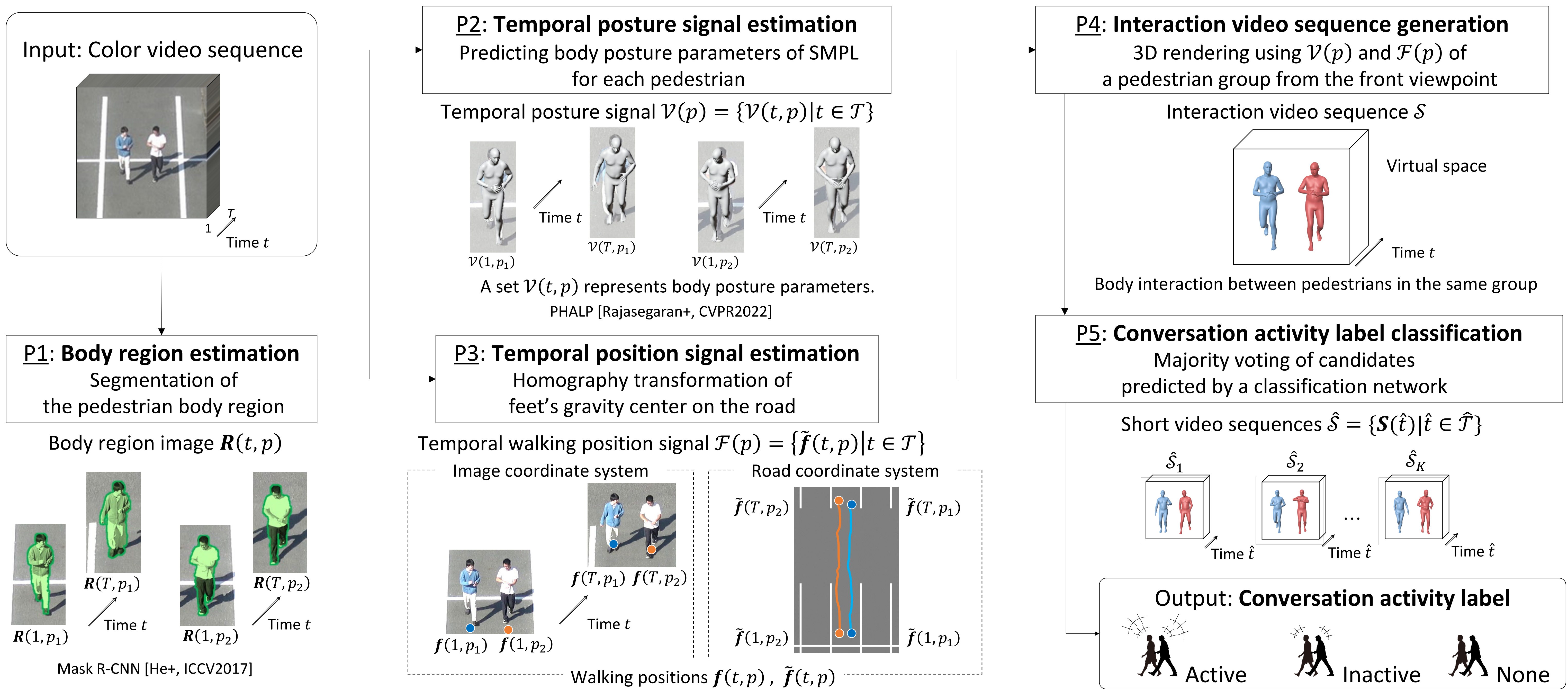


An application can determine whether pedestrian groups are satisfied with their visit by comparing the number of pedestrian groups engaged in active conversation between visitors that are arriving and leaving.



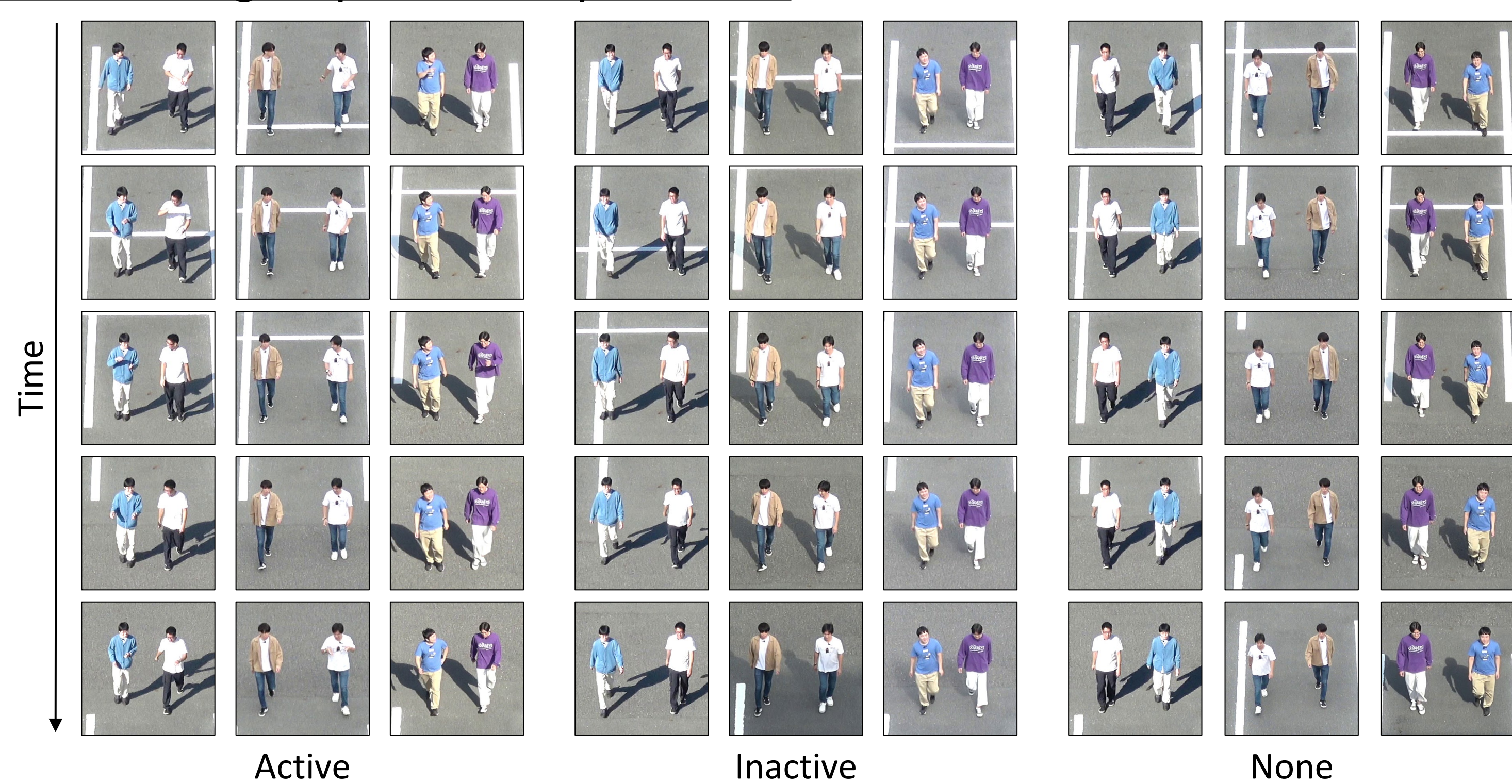
Our method (Conversation activity recognition)

- Extract an informative feature using an interaction video sequence rendered in a virtual space by fixing the viewpoint position of the virtual camera in front of a pedestrian group.
- Design a visible feature that allows human observers to directly see physical body interaction performed in a group.



Experiments

Pedestrian group video sequence \mathcal{R}'



Interaction video sequence \mathcal{S}



Human observers can visually confirm the posture among pedestrians, such as arm bending and face orientation, from the interaction video sequences.

Basic performance Comparison of the recognition accuracy

Collected an outdoor pedestrian dataset of 624 video sequences in 52 groups. Used C3D [Tran+, CVPR2015] in P5.

Interaction video sequence \mathcal{S}	76.2 ± 0.7%	Temporal posture signal $\mathcal{V}(p)$	72.9 ± 0.9%
Pedestrian group video sequence \mathcal{R}'	57.3 ± 1.3%	Temporal posture and walking position signal $\mathcal{V}(p), \mathcal{F}(p)$	74.1 ± 0.7%

Our interaction video sequence recognized conversation activity more accurately than alternative techniques.

Evaluation of different virtual camera viewpoints

We set six positions of the virtual camera viewpoints.

Front	Back	Overhead	Underfoot	Right side	Left side
Acc. (%): 76.2 ± 0.7	74.8 ± 0.3	70.2 ± 0.8	72.0 ± 0.9	40.0 ± 1.5	48.2 ± 2.6

Placing the virtual camera viewpoint in front of a pedestrian group effectively improved recognition accuracy.