

STCCrowd: A Multimodal Dataset for Pedestrian Perception in Crowded Scenes

Peishan Cong¹, Xinge Zhu², Feng Qiao³, Yiming Ren¹, Xidong Peng¹, Yuenan Hou⁴,
Lan Xu^{1,7}, Ruigang Yang⁵, Dinesh Manocha⁶, Yuexin Ma^{1,7†}

¹ShanghaiTech University ²The Chinese University of Hong Kong

³RWTH Aachen University ⁴Shanghai AI Laboratory

⁵University of Kentucky ⁶University of Maryland at College Park

⁷Shanghai Engineering Research Center of Intelligent Vision and Imaging

¹{congpsz, renyml, pengxd, mayuexin}@shanghaitech.edu.cn

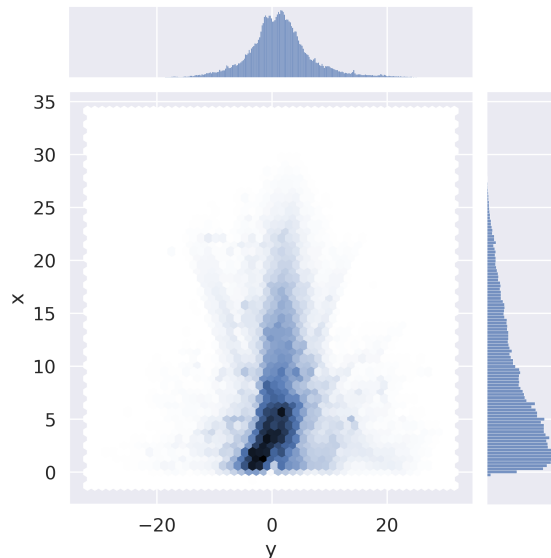


Figure 1. The pedestrian position distribution in BEV for point cloud. In STCCrowd, the pedestrians distribute from 0 to 35 meters in x-axis and -30 to 30 meters in y-axis.

1. Additional details of STCCrowd

Sensor setup. The vehicle is equipped with following sensors: Zed Sterelab2 stereo camera(15Hz frame rate, 1280×720 resolution and 60 field-of-view(FOV) for left and right, 70 FOV for center) and OS0 3D LiDAR with 128 beams (More details are provided in Table 1).

Sensor synchronization. To achieve good cross-modal data alignment between different sensors, the timestamp of the LiDAR is the time when the full rotation of current frame is achieved and the correspond of timestamps for different devices is achieved by special posture when recording data. We keep the common frequency as 5 Hz and annotate the frames per 0.4 second.

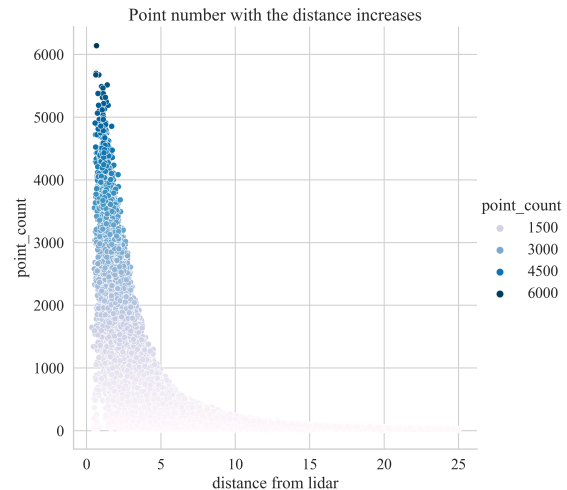


Figure 2. The number of points per instance decreases with the distance from LiDAR increases. The gradation of color represents the number of points.

Table 1. LiDAR specification.

Azimuthal FOV	360
Vertical FOV	90° ($\pm 45^\circ$)
Vertical resolution	0.7
Frequency	10Hz
range	55m
point per second	2,621,440

Pedestrian position distribution in BEV. We label the point cloud within the range of 180°. With the straight front of the camera as the reference direction, only the left and right 90° are labeled. The pedestrian position distribution in BEV for point cloud is presented in Fig 1.

Diverse instance-level densities and human poses . The density and position distribution is full of diversity which

