Table 3. Number of panoramic images and fisheye image patches

Dataset	Panorama		Fisheye	
	Train	Test	Train	Test
SL-MH	55,599	161	555,990	16,100
SL-PB	57,840	167	578,400	16,700
SP360	19,038	55	571,140	16,500
HoliCity	6,235	18	561,150	16,200

Table 4. Distribution of the camera parameters for training sets

Parameter	Distribution	Range or value[1]
Pan ϕ	Uniform	$[0, 360)$
Tilt θ, Roll ψ	Mix Normal Uniform	Normal 70%, Uniform 30% $\mu = 0, \sigma = 15$ $[-90, 90]$
Aspect ratio	Varying	{1/1 9%, 5/4 1%, 4/3 66%, 3/2 20%, 16/9 4%}
Focal length f	Uniform	$[6, 15]$
Distortion k_1	Uniform	$[-1/6, 1/3]$
Max angle $\eta_{\max}$ [2]	Uniform	$[84, 96]$

[1] Units: ϕ, θ, ψ, and $\eta_{\max}$ [deg]; f [mm]; k_1 [dimensionless]

[2] Max angle $\eta_{\max}$ is the maximum incident angle

geometric calculations, is irrelevant to training. The principle of the estimation is to fit two sets of world coordinates and is known as the absolute orientation problem [55]. One set of world coordinates consists of the 3D VP/ADPs projected by backprojection [53] using the camera parameters. The other set consists of the 3D points corresponding to these VP/ADPs along the orthogonal Manhattan world coordinates shown in Figure 3. In our case, we focus on rotation without translation and scaling because all 3D points are on a unit sphere. In this fitting, which has a lower computational cost than methods based on a singular value decomposition, we use the optimal linear attitude estimator [34, 39], which calculates the Rodrigues vector [13] expressing the principal axis and angle. This Rodrigues vector is compatible with a quaternion, and we can obtain pan, tilt, and roll angles from this quaternion. Note that we regard undeterminable angles as 0° when point detection fails or an image has one or no axes from the VP/ADPs. The supplementary presents details of the rotation estimation.

4. Experiments

To demonstrate the validity and effectiveness of our approach, we employed extensive experiments using large-scale synthetic datasets and off-the-shelf fisheye cameras.

4.1. Datasets

Panoramic image datasets. We used three large-scale datasets of outdoor panoramas, the StreetLearn dataset [38], the SP360 dataset [9], and the HoliCity dataset [64], as listed in Table 3. In StreetLearn, we used the Manhattan 2019 subset (SL-MH) and the Pittsburgh 2019 subset (SL-PB). These panoramic images are the equirectangular projection using calibrated cameras. Assuming practical conditions following [33, 52, 53], we regarded the vertical center of the panoramic images as the ground. Moreover, the tilt angle is 0° because the height of a camera mounted on a car is sufficiently small with respect to the distance between the camera and other objects. The horizontal center of the panoramic images corresponds with the travel direction of cars: the pan angle is 0°.

Fisheye-image and camera-parameter generation. For a fair comparison with the state-of-the-art method [53] in the estimation of tilt and roll angles, we used the generic camera model [53] to generate fisheye images. Following the procedure for dataset generation and capture [53], we generated fisheye images from panoramic images using the generic camera models with the ground-truth camera parameters in Table 4, and we captured outdoor images in Kyoto, Japan, using six off-the-shelf fisheye cameras. To generate the test set, we replaced the mixed and varying distributions in the training sets with a uniform distribution. Therefore, our generated fisheye images and ground-truth camera parameters were used for training and evaluation.

4.2. Vanishing-point annotation

Vanishing-point label ambiguity. As shown in Table 2, we annotated the VP/ADPs of the image coordinates and labels on the basis of panoramic-image width and height. We found that some generated fisheye images had label ambiguity; that is, we cannot annotate unique VP/ADP labels for these images. For example, we cannot distinguish one image with a 0°-pan angle from another with a 180°-pan angle because we cannot determine the direction of travel of the cars from one image. In other words, we cannot distinguish front labels from back labels in Table 2. Similarly, we cannot distinguish left labels from right labels.

Removal of label ambiguity. Considering generalized cases of label ambiguity, we annotated the image coordinates of VP/ADPs as follows. We 180°-rotationally align all labels based on two conditions: 1) the images have back labels without front labels, and 2) the images have right labels without front and left labels. Details of the number of labels can be found in the supplementary materials.

Label ambiguity also affects conventional methods in a Manhattan world. For example, it is often the case that three orthogonal directions can be estimated using the Gaussian sphere representation of VPs [63]; however, the representation does not regard the difference between front and back directions. For a fair comparison in the evaluation, we selected the errors with the smallest angles from among the candidate ambiguous angles in both the conventional methods and our method. Therefore, the estimated pan-angle ranges from $-90°$ to $90°$.