Deep Single Image Camera Calibration by Heatmap Regression to Recover Fisheye Images Under Manhattan World Assumption

Nobuhiko Wakai[1] Satoshi Sato[1] Yasunori Ishii[1] Takayoshi Yamashita[2]
[1] Panasonic Holdings Corporation [2] Chubu University
{wakai.nobuhiko,sato.satoshi,ishii.yasunori}@jp.panasonic.com takayoshi@isc.chubu.ac.jp

Abstract

A Manhattan world lying along cuboid buildings is useful for camera angle estimation. However, accurate and robust angle estimation from fisheye images in the Manhattan world has remained an open challenge because general scene images tend to lack constraints such as lines, arcs, and vanishing points. To achieve higher accuracy and robustness, we propose a learning-based calibration method that uses heatmap regression, which is similar to pose estimation using keypoints, to detect the directions of labeled image coordinates. Simultaneously, our two estimators recover the rotation and remove fisheye distortion by remapping from a general scene image. Without considering vanishing-point constraints, we find that additional points for learning-based methods can be defined. To compensate for the lack of vanishing points in images, we introduce auxiliary diagonal points that have the optimal 3D arrangement of spatial uniformity. Extensive experiments demonstrated that our method outperforms conventional methods on large-scale datasets and with off-the-shelf cameras.

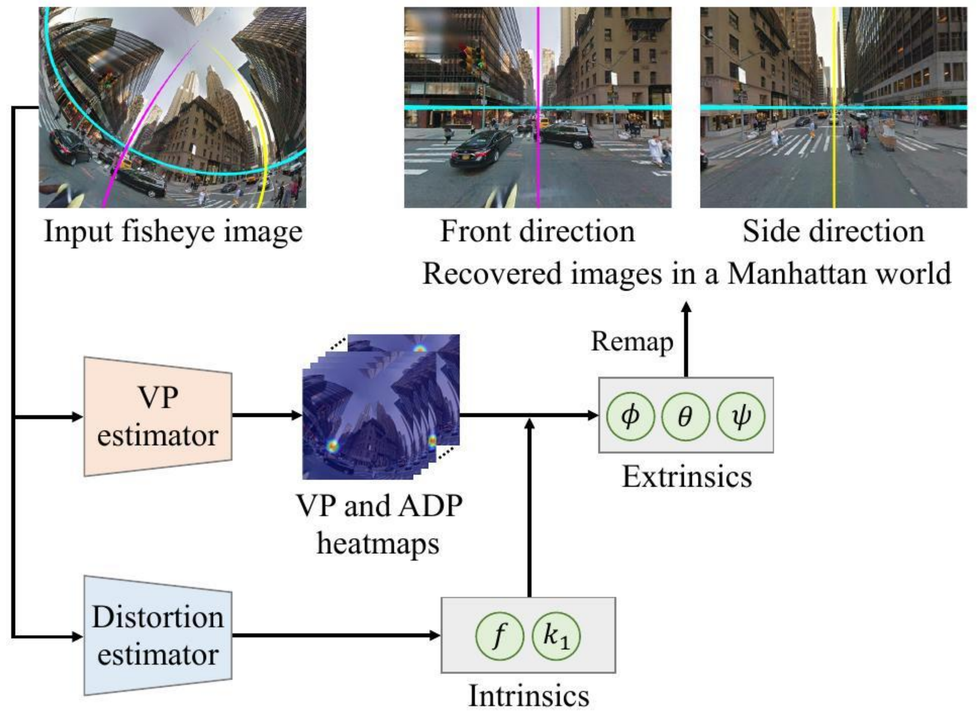

Figure 1. Our network estimates extrinsics and intrinsics in a Manhattan world from a single image. Our estimated camera parameters are used to fully recover images by remapping them while distinguishing the front and side directions on the basis of the Manhattan world. Cyan, magenta, and yellow lines indicate the three orthogonal planes of the Manhattan frame in each of the images. The input image is generated from [38].

1. Introduction

In city scenes, image-based recognition methods are widely used for cars, drones, and robots. It is desirable to recognize the directions in which roads exist for navigation, self-driving, and driver assistance. To avoid colliding with cars and pedestrians, it is more important to detect these objects in front of a vehicle rather than at the sides in Figure 1. We can obtain the travel direction from odometry, gyroscopes, or accelerator sensors using these specific devices. However, for cars, drones, and robots, image-based angle estimation of the travel direction without these devices is better for miniaturized and lightweight design. To determine the origin of angles, a Manhattan world [12] defines orthogonal world coordinates along cuboid buildings and a grid of streets. Although this image-based angle estimation is a long-studied topic in areas of geometric tasks [1, 4, 51], accurate and robust angle estimation has remained an open challenge because general scene images tend to lack constraints such as lines, arcs, and vanishing points (VPs).

To control cars, drones, and robots, images for recognition are needed that have a large field of view (FOV). Fisheye cameras have a larger FOV than other cameras, but fisheye images are highly distorted. After fisheye distortion has been removed, we can use various learning-based recognition methods, such as object detection [25, 27], semantic segmentation [10, 26], lane detection [15, 61], action recognition [50, 59], and action prediction [6, 20]. To recover fisheye images, performing camera calibration before the recognition tasks mentioned above is desirable.

Geometry-based calibration methods can estimate the camera rotation and distortion from a distorted image [3, 32, 41, 57]. However, it is difficult for geometry-based methods to calibrate cameras from images that contain few artificial