3D Geometry for Panorama

Jianxiong Xiao
Massachusetts Institute of Technology

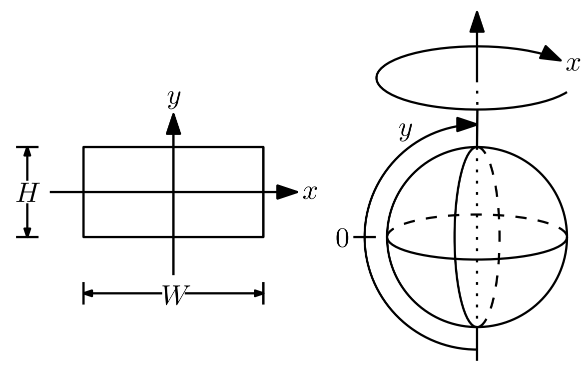
Abstract

This document briefly explains the geometry of full-view panorama image we used in [7].

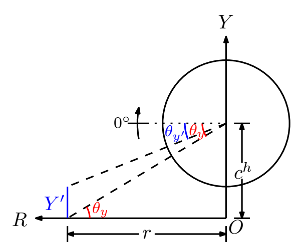
Here, we describe how to reconstruct the rough 3D model for each single spherical image. As in [2, 4, 1], we assume that a scene is composed of a number of objects standing on a ground plane (dubbed standing objects), with the objects represented as piecewise-connected planes oriented orthogonally to the ground plane. To obtain the ground line for each spherical image, we ask people to label the ground plane polygon using LabelMe [4].

As shown in Figure 1(a), let's denote the resolution of the spherical image to be $W \times H$ pixels and represent the panorama in Equirectangular projection [6, 5]. Because the spherical images covers 360 degree field of view horizontally and 180 degree field of view vertically, we can know that $W = 2H$, and the focal length is $W/2\pi$ pixels. For spherical projection images, because of the full coverage of field of view, people normally put the camera in a position that the top of the projection sphere is pointing to the sky. Therefore, we can safely assume that the horizontal vanishing line of the ground plane is at 0 height of the image coordinate $[-\frac{W}{2}, \frac{W}{2}] \times [-\frac{H}{2}, \frac{H}{2}]$. As in [1, 3, 4], we further assume that the camera is at the height of $c^h = 1.7$ meter away from the ground plane.

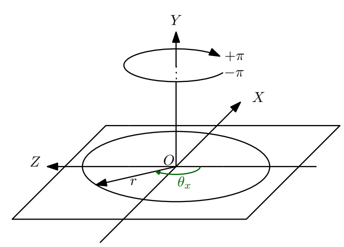
With these conditions, the geometry transformatioin for 3D reconstruction becomes very straightforward and elegant. For a 3D point $\mathbf{P} = (X, Y, Z)$ on the ground plane with image coordinate (x, y), we have $Y = 0$ since it is on the ground. And $\theta_x = 2\pi x/W$, $\theta_y = \pi y/H$. From Figure 1(b), we can see $r = c^h |\cot \theta_y|$. Finally, from Figure 1(c), we can see that $X = r\cos(-\theta_x + \frac{\pi}{2})$ and $Z = r\sin(-\theta_x + \frac{\pi}{2})$. For a 3D point $\mathbf{P'} = (X', Y', Z')$ not on the ground plane with image coordinate (x, y'), which is assumed to be on a dubbed standing billboard plane, we first find out a point $\mathbf{P} = (X, Y, Z)$ on the ground plan with same x image coordinate, then we have $X' = X$, $Z' = Z$, and $r' = r = c^h |\cot \theta_y|$. The only difference is $Y' \neq Y = 0$. From Figure 1(b), we can see that $Y' = c^h + r \tan \theta_{y'}$.



(a) Spherical Image Plane.



(b) Slice of 3D to calculate r and Y.



(c) X and Z are calculated from r and θ_x.

Figure 1. Geometry derivation for 3D reconstruciton of spherical image using ground line.

References

[1] D. Hoiem, A. Efros, and M. Hebert. Putting objects in perspective. In *CVPR*, 2006. 1

[2] D. Hoiem, A. A. Efros, and M. Hebert. Geometric context from a single image. In *ICCV*, 2005. 1

[3] D. Hoiem, A. A. Efros, and M. Hebert. Closing the loop on scene interpretation. In *CVPR*, 2008. 1

[4] B. Russell and A. Torralba. Building a database of 3d scenes from user annotations. In *CVPR*, 2009. 1

[5] J. P. Snyder. *Flattening the Earth: Two Thousand Years of Map Projections*. 1993. 1

[6] Wikipedia. Equirectangular projection, 2004. http://en.wikipedia.org/wiki/Equirectangular_projection. 1

[7] J. Xiao, K. Ehinger, A. Oliva, and A. Torralba. Recognizing scene viewpoint using panoramic place representation. In *CVPR*, 2012. 1