Twitter: **T**opdown **WIT**h **T**ransformation Using **E**gocentric **R**GB Images

Sundar, Akash Goel, Ayush Xie, Jason

Venkatesh, Sharanya

Project Summary

In video games, we often view the scene from the top as if we are birds observing the environment. We can also do this in the real world with some clever camera tricks. In fact, many automobile manufacturers in recent-years have been using top-down 360 views to assist the driver in parking. Top down views can also be used for downstream automation tasks such as parking assist and autonomous driving. Thus, this project pertains to the transformation of multiple images captured in the world to a single bird-eye view.

Goals and Objectives

- Given a single image from a egocentric RGB camera, transform the image to its corresponding top-down view.
- From the top-down view of multiple images, stitch these images and apply blending to form a single coherent top-down view.

Related Works

There are plenty of research for Advanced Driver Assistance Systems (ADAS) using both classical and neural approaches. [1] provides an overview of the hardware and software involved in implementing driver assistance systems. The authors in that paper also discuss in detail on applying vision for downstream tasks such as blindspot monitoring, parking spot recognition, and pedestrian tracking. One work that is directly related to our project proposal is [2], where birds-eye-views are generated via a Generative Adversarial Network trained on egocentric/top-down image pairs taken from GTA5.

Proposed Approach

We propose a classical approach to the problem of BEV generation using homographies. We will be using the Kitti dataset for our images.

Let the original camera frame be $\mathbf{C_0}$, and the virtual top-down camera frame be $\mathbf{C_v}$. We define the transformation from $\mathbf{C_0}$ to $\mathbf{C_v}$ to be $[\mathbf{R^{C_0}_{C_v}}|\mathbf{t^{C_0}_{V_v}}]$. Let $\mathbf{P}$ be an imaged point in world coordinates. The two camera projection equations are:

$$\mathbf{q_{C_0}} \sim \mathbf{K_{C_0}}[\mathbf{R^{C_0}_{W}}|\mathbf{t^{C_0}_{W}}]\mathbf{P} \tag{1}$$

$$\mathbf{q_{C_v}} \sim \mathbf{K_{C_v}}[\mathbf{R^{C_v}_{W}}|\mathbf{t^{C_v}_{W}}]\mathbf{P} \tag{2}$$

Since we do not need to operate in world space, we fix the world frame to be the same as $\mathbf{C_0}$. Then the world point $\mathbf{P}$ can be described as:

$$\mathbf{P} \sim \mathbf{K_{C_0}}^{-1}\mathbf{q_{C_0}} \tag{3}$$

Substituting this into equation 2, we see that the re-projected top down image is described by:

$$\mathbf{q_{C_v}} \sim \mathbf{K_{C_v}}[\mathbf{R^{C_v}_{C_0}}|\mathbf{t^{C_v}_{C_0}}]\mathbf{K_{C_0}}^{-1}\mathbf{q_{C_0}} \tag{4}$$

Following re-projection of images into the top-down view, we can attempt to stitch images together using feature matching. Further blending can be done to the stitched images to enhance quality.

Timeline

- Nov. 13th: Finish single-image re-projection
- Nov. 22nd: Midterm report, Multiple image stitching
- Dec. 1st: Image blending
- Dec. 15th: Submit presentation and final report

Divide & Conquer

- Akash: Implement single-view transformations
- Ayush: Explore Kitti Dataset
- Jason: Explore Kitti Dataset
- Sharanya: Implement image stitching & blending

References

[1] M. Heimberger, J. Horgan, C. Hughes, J. McDonald, and S. Yogamani, "Computer vision in automated parking systems: Design, implementation and challenges," *Image and Vision Computing*, vol. 68, pp. 88–101, 2017, automotive Vision: Challenges, Trends, Technologies and Systems for Vision-Based Intelligent Vehicles. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885617301105

[2] "Generating bird's eye view from egocentric rgb videos," *Generative Adversarial Networks for Multi-Modal Multimedia Computing*, vol. 2021, 2021. [Online]. Available: https://www.hindawi.com/journals/wcmc/2021/7479473/