Visual Camera Orientation Estimation using Machine Learning

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2024

Martin Kubička
Supervisor: doc. Ing. Martin Čadík Ph.D.

TASK

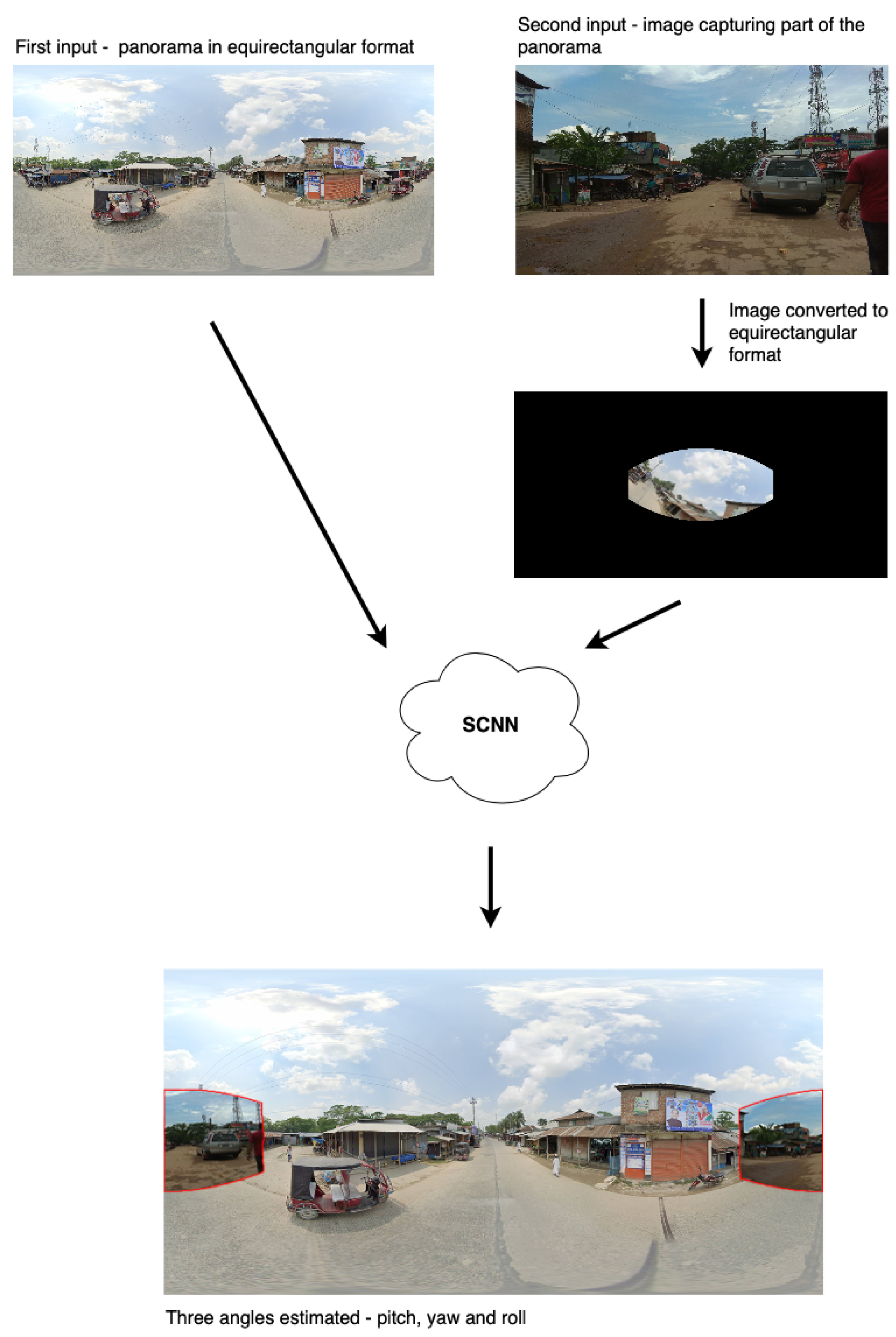

Figure 1. Task pipeline

Advantages of Spherical CNNs

Rotation equivariance or invariance

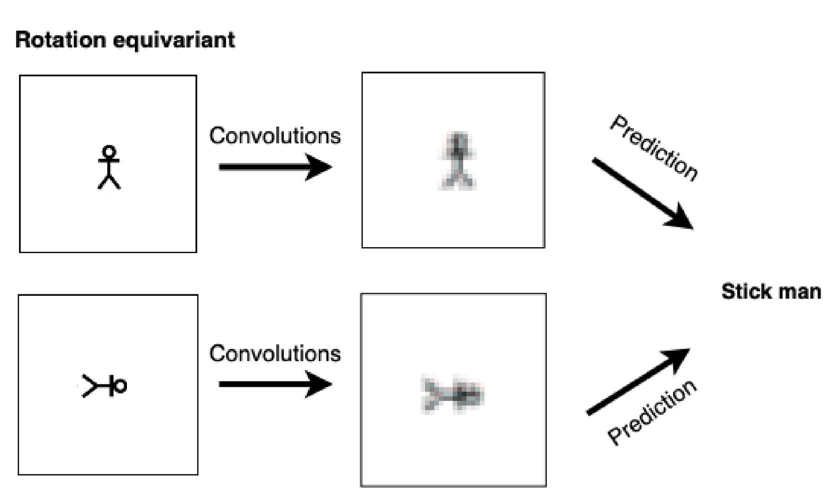

Figure 2. Rotation equivariance visualization

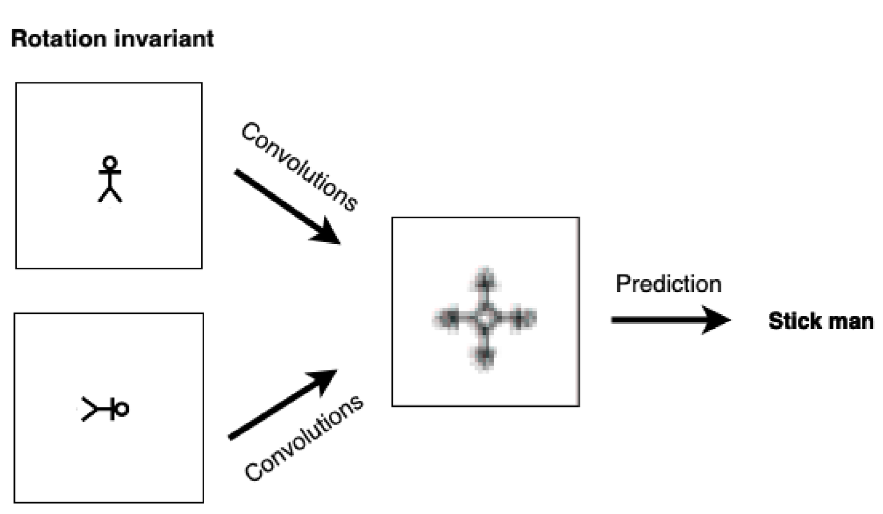

Figure 3. Rotation invariance visualization

Effectively handles distortions

Figure 4. Distorted car after transformation from sphere to plain

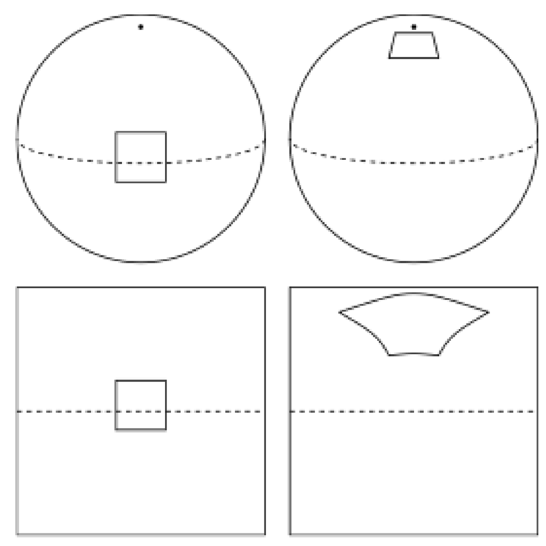

Figure 5. Visualization how spherical convolutions pass through photo.

RESULTS

For PACNoRoll/GeoPose3K dataset it was possible to estimate the pitch angle with an accuracy of 39/76% if we tolerate an error of 10 degrees, 53/92% if we tolerate an error of 15 degrees and 61/96% if we tolerate an error of 20 degrees and for yaw angle, if we tolerate and an error of 10 degrees accuracy is 6/16%, if we tolerate an error of 15 degrees accuracy is 10/28% and if we tolerate and error of 20 degrees accuracy is 16/42%.

Figure 6. Visualization of prediction and ground truth error

Conclusion

Several models were created for resolving this task, where each has a different idea behind it and brings various results. The best results were shown by the direction of semantic segmentation, where for each angle, a separate model is created where initially, the pitch angle is estimated and then the yaw angle in case we are estimating only these 2 angles. During the development of the model architecture, we encountered the limits of current hardware, specifically the size of GPU memory. Another challenge was the fact that there are no pre-trained models for this type of convolution, so we trained from scratch. In addition to the created models, several datasets were developed, which are a significant contribution, as there are not many datasets that contain panoramas in equirectangular format and even fewer datasets that contain panoramas in stereographic format. Part of the work also includes programs that generate these datasets, and thus they can be expanded.