Crowd Tracking Techniques for Evaluation of Adherence to Social Distancing Measures

Computer Vision Final Project Proposal

Jacob Desman, Sandy Shi, Emily Chang, Andrew Cornelio

The discovery and subsequent rapid spread of COVID-19 has brought with it a variety of regulations aimed to prevent further transmission of the disease. One such guideline is social distancing: putting approximately 6 feet of distance between oneself and other people. Unfortunately, in-person enforcement of social distancing guidelines on a broad scale is difficult and poses risk of infection to the monitors. A better approach is to analyze video feeds from network cameras and gauge social distancing remotely, as shown in Ghodgaonkar, et al. (2020) [1]. We aim to explore and analyze different crowd tracking techniques, obtaining metrics for how well they perform when determining crowd size and interpersonal distance.

To evaluate the extent of proper social distancing in crowds, we propose two main approaches: one based on analysis of bounding box movement and one based on the calculation of homographies between the camera's view and the ground plane:

Detection. The first goal of social distancing evaluation is crowd tracking to detect both the presence and location of individuals. This can be done using many different methods, such as histogram of oriented gradients (HOG) [2], graph cuts [3], and deep learning [4]. For our purposes, we will evaluate different proven deep learning person detection systems, such as YOLO [5] and AlphaPose [6], to automatically process the images at various scales and generate a list of key points that correspond to a person's positioning.

Bounding box approach. Using the list of key points, we will generate a bounding box for each frame. Counting the number of bounding boxes returns the number of people within the scene, and distance can be gauged using an assumption where every human is the average height of 5.4 feet [1]. We also hope to explore more complex methods of distance estimation.

Homography-based approach. We first want to determine if we can compute image geometry in known rooms to determine depth of detected people. We will then evaluate methods of determining the ground plane to calculate the homography that converts the scene to a birds eye view for interpersonal distance detection. Past papers have explored such methods [7, 8, 9].

Tracking. As a stretch goal, we hope to extend our image-based solution to video. Baseline tracking would consist of frame-by-frame mapping of keypoints, which would later be developed into more sophisticated tracking techniques, such as analysis of optical flow.

Performance will be analyzed based on comparison to the ground truth. For that purpose, we have selected datasets that include calibrated sets of labeled images/videos, such as the kaggle crowd counting dataset.

In terms of defining project roles, we have assigned each member to be the "lead" of one aspect of the project and make sure each component remains on track, but we will still collaborate amongst ourselves and contribute where help is needed. The roles are as follows:

- **Jacob Desman:** Lead for calibration/homography determination
- **Sandy Shi:** Lead for baseline tracking
- **Emily Chang:** Lead for relative sized bounding box approach
- **Andrew Cornelio:** Lead for Deep Learning and homography determination

[1] https://arxiv.org/pdf/2008.12363.pdf
[2] https://ieeexplore-ieee-org.proxy1.library.jhu.edu/document/1467360
[3] https://cvg.ethz.ch/teaching/cvl/2012/grabcut-siggraph04.pdf
[4] https://arxiv.org/pdf/1807.05511.pdf
[5] https://pjreddie.com/darknet/yolo/
[6] https://github.com/MVIG-SJTU/AlphaPose
[7] https://arxiv.org/pdf/2005.04813.pdf
[8] https://arxiv.org/pdf/1905.02231.pdf
[9] https://arxiv.org/pdf/2011.02018.pdf