Real-time Social Distancing Detection System with Auto-calibration using Pose Information

Gaku Nakano
g-nakano@nec.com
Central Research Labs, NEC Coporation
Kawasaki, Kanagawa, Japan

Shoji Nishimura
nishimura@nec.com
Central Research Labs, NEC Coporation
Kawasaki, Kanagawa, Japan

ABSTRACT

In this demo, we present a real-time social distancing system with auto-calibration using human pose information. Our system first calculates geometric parameters of a camera in 3D space, *i.e.* position and rotation, then, measures absolute distances between pedestrians. Detection results are visualized as 3D circles in input images and a bird's eye view. All processing steps are completely automatic, therefore, our system can be applied for uncalibrated surveillance cameras, which are already installed in town. A demo video is available in the supplemental material.

CCS CONCEPTS

• **Computing methodologies** → **Scene understanding**; **Camera calibration**.

KEYWORDS

Social distancing, Real-time detection, Surviellance camrea, Camera calibration, Human pose estimation

1 INTRODUCTION

Social distancing has been a simple but an effective measure to reduce risks of infectious diseases caused by a virus, *e.g.* COVID-19. WHO (World Health Organization) recommends maintaining at least a 1-meter distance between people [4]. Also, the government of the United Kingdom reports that relative infection risk is 2–10 times higher at 1-meter distance than 2 meters, and shows a guideline to mitigate risks where a 2-meter distance is not viable [1].

One of a promising approach to detect social distancing violations is to use surveillance cameras that are already installed in town. Several methods [2, 6], which use an object detection with deep learning, *e.g.* YOLO and R-CNN, have been proposed for measuring physical distances between two pedestrians in a surveillance video. Those methods have two drawbacks for practical use: 1) Camera calibration data is manually given. 2) Distance measurement is inaccurate due to bounding box based methods.

To overcome those issues, we propose a social distancing detector including an automatic camera calibration using a human pose detector. Figure 1 illustrates the overview of our system consisting of two stages: an offline stage for camera auto-calibration, and an online stage for social distancing detection. At the offline stage, we calibrate a surveillance camera using human joints to determine the 3D position and rotation of the camera. Then, we detect pedestrians frame-by-frame, and estimate pedestrians' 3D location using the estimated calibration data at the online stage. Finally, social distancing is visualized in the input frame and a bird's eye view. As shown in Fig. 2, a 1-meter distance of pedestrians is marked by a blue circle whereas social distancing violation is colored in red.

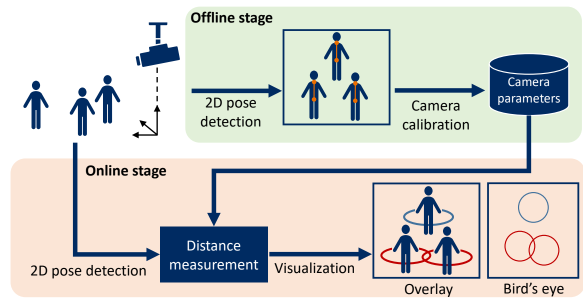

Figure 1: System overview. Camera calibration is first conducted at the offline stage. Physical distances between pedestrians are measured and visualized frame-by-frame at the online stage. The calibration data can be reused until the camera setting changes.

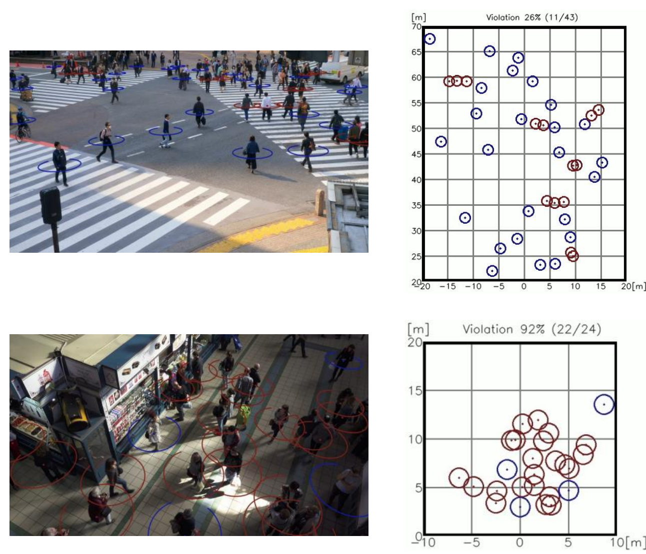

Figure 2: Visualization of social distancing detection. Each circle represents a 1-meter distance of a pedestrian. Distancing violations are marked by red. The right images are bird's eye views of the left images. *Top*: An intersection in Shibuya, Japan. *Bottom*: A market in Budapest, Hungary.

In the following sections, we describe the details of our system and report an experimental result conducted at a shopping district in Japan.

2 PROPOSED SYSTEM

2.1 Camera auto-calibration using human joints

Since the use of calibration patterns requires stopping traffic around a surveillance camera, we adopt a camera auto-calibration method that can estimate camera parameters from 2D observations of pedestrians in video frames.

Let $\mathbf{A}$ and $\mathbf{B}$ be 3D points of a pedestrian's hip and neck, respectively. Their 2D projections $\mathbf{a}, \mathbf{b}$ can be written by

$$\mathbf{a} \propto \mathsf{K}(\mathsf{R}\mathbf{A} + \mathbf{t}), \quad \mathbf{b} \propto \mathsf{K}(\mathsf{R}\mathbf{B} + \mathbf{t}), \tag{1}$$

where $\mathsf{K}, \mathsf{R}, \mathbf{t}$ are camera's intrinsic parameter matrix, rotation matrix, and translation vector, respectively. Note that $\propto$ represents equality up to scale, and $\mathbf{a}, \mathbf{b}$ are represented by the 3×1 homogeneous coordinates. 2D-3D point correspondences of a neck–hip connection, $\{\mathbf{A}, \mathbf{B}\} \leftrightarrow \{\mathbf{a}, \mathbf{b}\}$, is illustrated by orange lines in Fig. 1.

First we detect human joints by NeoPose [5], which is a deep-learning-based 2D pose detector robust against low image resolution of people observations. Next, assuming that all pedestrians are the same height, *e.g.* mean height of adults, we utilize 2D observations of a neck–hip connection of pedestrians as vertical line segments for the calibration method. To deal with significant height differences between people, *e.g.* an adult and a child, we remove wrong line segments by incorporating the calibration method with RANSAC (RANdom SAmple Consensus). Although the camera parameters are obtainable from at least two line segments [3], we found that more than 100 lines are required for stable estimation in practice. Once calibration is done, the parameters are stored in a storage and can be used again until the camera configuration is modified.

2.2 Social distancing measurement and visualization

At the online stage, we first detect human poses in a video frame and estimate the 3D location of neck–hip connections. Using the estimated camera calibration data, we can calculate (x, y) position of a pedestrian and the length ℓ of his/her neck–hip connection by solving a least square problem:

$$\begin{bmatrix} \mathbf{c} \times \mathbf{r}_1 & \mathbf{c} \times \mathbf{r}_2 & \mathbf{0} \\ \mathbf{d} \times \mathbf{r}_1 & \mathbf{d} \times \mathbf{r}_2 & \mathbf{d} \times \mathbf{r}_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ \ell \end{bmatrix} = \begin{bmatrix} \mathbf{c} \times \mathbf{t} \\ \mathbf{d} \times \mathbf{t} \end{bmatrix}, \tag{2}$$

where $\mathbf{r}_k$ represents the k-th column of R, and $\mathbf{c} = \mathsf{K}^{-1}\mathbf{a}$, $\mathbf{d} = \mathsf{K}^{-1}\mathbf{b}$. The estimated values x, y, ℓ can be further optimized by performing bundle adjustment on Eq. (1) if lens distortion is not negligible.

The human pose detector sometimes wrongly detects neck–hip connections from the background image. We find those outliers by validating the length ℓ of each line segment. For example, a neck-hip connection is removed if its length ℓ is not a realistic value, *i.e.* too short $\ell < \ell_{\min}$ or too long $\ell > \ell_{\max}$.

Then, the physical distance $d_{i,j}$ between an i-th and j-th pedestrians is measured by

$$d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \tag{3}$$

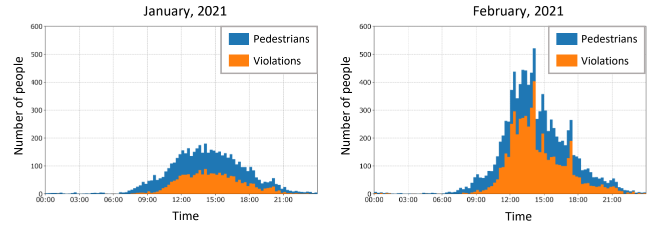

Figure 3: Experimental results at a shopping district in Utsunomiya City, Japan. In conjunction with lifting the state of emergency, pedestrians and distancing violations increase from January to February.

Thus, a distance violation can be determined by $d_{i,j} \leq \delta$ where δ is the distance threshold, which can be arbitrarily set by users, *e.g.* 1–2 meters.

Finally, we visualize detection results in the input frame and a bird's eye view. In Fig. 2, red circles indicate *violations* and blue ones represent *good distancing*, *i.e.* they are far enough away from each other.

3 EXPERIMENTS

We implemented the proposed system on a PC with Core i7-9750H and Quadro RTX5000. All programs were written in Python and Cython. The 2D human pose detector, NeoPose, was trained using COCO dataset on FP32, and was optimized to an FP16 model by TensorRT. The inference speed of FP16-NeoPose is 15–20 fps for 1920×1080 images. The camera calibration step spends 3–5 minutes including the pose detection from video frames and the camera parameter estimation. We uniformly selected video frames for the calibration up to 2000 neck–hip connections obtained.

In cooperation with MLIT (the Ministry of Land, Infrastructure, Transport and Tourism) and Utsunomiya City in Japan, we demonstrated our system in the Utsunomiya shopping district from January to February 2021. Figure 3 shows the average number of pedestrians and social distancing violations detected on Saturdays and Sundays during the experiment. Since the state of emergency was lifted in February, it was observed that the number of pedestrians and social distancing violations both increased from January.

REFERENCES

[1] HM Government. 2020. *Review of two metre social distancing guidance.* Technical Report. HM Government. https://www.gov.uk/government/publications/review-of-two-metre-social-distancing-guidance

[2] Yew Cheong Hou, Mohd Zafri Baharuddin, Salman Yussof, and Sumayyah Dzulkifly. 2020. Social distancing detection with deep learning model. In *2020 8th International Conference on Information Technology and Multimedia (ICIMU).* IEEE, 334–338.

[3] Gaku Nakano. 2021. Camera Calibration Using Parallel Line Segments. In *2020 25th International Conference on Pattern Recognition (ICPR).* IEEE, 1505–1512.

[4] World Health Organization. 2021. *Coronavirus disease (COVID-19) advice for the public.* World Health Organization. Retrieved July 1, 2021 from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public

[5] Yadong Pan and Shoji Nishimura. 2019. Multi-person pose estimation with mid-points for human detection under real-world surveillance. In *Asian conference on pattern recognition.* Springer, 239–253.

[6] Dongfang Yang, Ekim Yurtsever, Vishnu Renganathan, Keith A Redmill, and Ümit Özgüner. 2021. A vision-based social distancing and critical density detection system for covid-19. *Sensors* 21, 13 (2021), 4608.