

Single Image Human Proxemics Estimation for Visual Social Distancing (Supplementary Material)

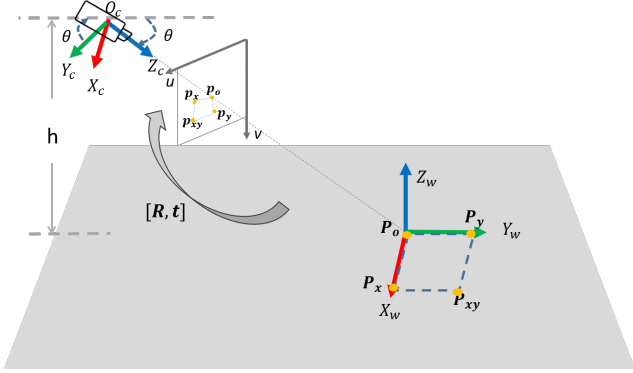


Figure 1. Illustration of the world and camera coordinate system under our assumptions that i) the camera only has a tilt angle θ with zero roll and pan angle, and ii) the camera has the height h from the ground and the world origin locates at where the camera principal axis intersects on the ground plane. The \mathbf{P} on the ground plane can be mapped to the pixel \mathbf{p} on the image through the homography matrix. On the ground plane the corner points on the unit square \mathbf{P}_o , \mathbf{P}_x , \mathbf{P}_y and \mathbf{P}_{xy} correspond to the pixel points \mathbf{p}_o , \mathbf{p}_x , \mathbf{p}_y and \mathbf{p}_{xy} , respectively.

We provide in this Supplementary Material more information about the experimental evaluation in the paper, in particular, related to Precision and Recall for the three datasets, *Epfl-Mpv-VSD*, *Epfl-Wildtrack-VSD*, and *OxTown-VSD*. In addition, we include the visualisation of heatmaps that shows intuitively the image areas where the violations appear more.

Finally, we provide short video clips for all the sequences in our dataset with the qualitative results of the proposed algorithm overlaid. It is worth to mention that the videos are generated given the best ρ_h and ρ_v by grid-search strategy (given in Table 1 of the main manuscript) and the best performing body part for each sequence (given in Table 2 of the main manuscript).¹

1. Relation between homography matrix to the 2 trapezoid ratios

Let us consider a pinhole camera model with the camera intrinsic matrix $\mathbf{K} \in \mathbb{R}^{3 \times 3}$ and extrinsic matrix $[\mathbf{R}|\mathbf{t}]$. Any 2D point in world ground plane $\mathbf{P} = [X, Y, 1]^T$ can then be projected to the image plane at pixel position $\mathbf{p} = [u, v, 1]^T$ as:

$$s\mathbf{p} = \mathbf{K} [\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}] \mathbf{P} = \mathbf{H}\mathbf{P}, \quad (1)$$

where s is the scale factor, \mathbf{H} is the *homography matrix* to project points on the ground plane to the image plane and $\mathbf{P}' = [X, Y, 1]^T$.

Based on our assumption of the world coordinate with the camera roll angle and pan angle to be 0° and the camera origin, we will have:

$$\mathbf{H} = \mathbf{K} [\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}] = \mathbf{K} \begin{bmatrix} 1 & 0 & \frac{o}{h} \\ 0 & -\cos(\theta) & -\frac{h}{\tan(\theta)} \\ 0 & -\sin(\theta) & h \end{bmatrix} \quad (2)$$

As we mentioned in the main paper, lines that are parallel to the X-axis in the ground plane remain horizontal in the projected image plane, and lines parallel to the Y-axis converge to the vanishing point along Y-axis in the image plane. Therefore, a rectangular area on the ground plane with a width W and a height H would be projected into an isosceles trapezoidal shape with short-based width W' and height H' , where the horizontal ratio $\rho_h = \frac{W'}{W}$ and vertical ratio $\rho_v = \frac{H'}{H}$ are purely related to the camera tilt angle θ and the camera height h .

To derive relation between the ρ_h , ρ_v and the camera parameters, we simplify the rectangular area with width W and height H on the ground plane to a square area with a unit length, therefore $W = H = 1$. Let us align the square at the corner as indicated as in Fig. 1. The corner points on the unit square are $\mathbf{P}_o = [0, 0, 1]^T$, $\mathbf{P}_x = [1, 0, 1]^T$, $\mathbf{P}_y = [0, 1, 1]^T$ and $\mathbf{P}_{xy} = [1, 1, 1]^T$. Their corresponding

¹We do not visualise the whole dataset images but only a fraction due to the overall size.

pixel points are \mathbf{p}_o , \mathbf{p}_x , \mathbf{p}_y and \mathbf{p}_{xy} , respectively:

$$\begin{aligned} s\mathbf{p}_o &= \mathbf{K}\mathbf{t} \\ s\mathbf{p}_x &= \mathbf{K} [\mathbf{r}_1 \mathbf{t}] \\ s\mathbf{p}_y &= \mathbf{K} [\mathbf{r}_2 \mathbf{t}] \\ s\mathbf{p}_{xy} &= \mathbf{K} [\mathbf{r}_1 \mathbf{r}_2 \mathbf{t}], \end{aligned} \quad (3)$$

The horizontal ratio ρ_h and vertical ratio ρ_v is related to the camera model as following:

$$\rho_h = \frac{\|\mathbf{p}_x - \mathbf{p}_o\|}{\|\mathbf{p}_{xy} - \mathbf{p}_y\|} \quad (4)$$

$$\rho_v = \|\mathbf{p}_y - \mathbf{p}_o\|. \quad (5)$$

Since pixels \mathbf{p}_o , \mathbf{p}_x , \mathbf{p}_y and \mathbf{p}_{xy} are only related the tilt angle θ and camera height h , ρ_h and ρ_v are therefore only related to these two parameters.

2. Further evaluation results

In the main paper, due to restrictions in space, we solely reported the results in terms of F1-score, in this supplementary material, we provide the complete results, including the Precision and Recall values. Table 1 corresponds to the results given in Table 1 of the main manuscript to compare the performance achieved using our proposed method for computing Homography versus the automatic approach proposed in state-of-the-art.

Tables 2, 3, 4, 5 provides the Precision and Recall values corresponding to all experiments reported in Table 2 of the main manuscript, intending to investigate ‘Leg length’, ‘Arm length’, ‘Torso length’, and the entire body in form of ‘BBX height’ as the metric reference choice, using a fixed homography.

3. Social Distancing Violations Heatmaps

In this section, we demonstrate the heatmap of social distancing violations over a short clip extracted from each dataset, as well as the footstep of the people on the ground shown in green dots. The correspondence of the density of footsteps and red area in the heatmap for all the examples can be appreciated.

4. Body-Part Joints

In Table 6, we list the index of joints we used for defining a body part. This indexing indeed corresponds to the OpenPose 25 joint model output and might need adaption if any other pose detector is used.

Table 1. Investigating the estimation of Homography matrix \mathbf{H}

| Dataset | Seq. | Proposed VSD - Grid search | | | | | AutoRect \mathbf{H} | | | Monoloco | | |
|----------------|------|----------------------------|----------|-----------|--------|----------|-----------------------|--------|----------|-----------|--------|----------|
| | | ρ_h | ρ_v | Precision | Recall | F1-score | Precision | Recall | F1-Score | Precision | Recall | F1-Score |
| EPFL-mpv | C0 | 0.7 | 0.5 | 78.26 | 78.26 | 77.90 | 70.04 | 81.05 | 73.47 | 70.56 | 78.42 | 73.46 |
| | C1 | 0.5 | 0.6 | 74.45 | 77.03 | 75.39 | 51.27 | 87.72 | 61.17 | 68.54 | 73.87 | 70.19 |
| | C2 | 0.6 | 0.6 | 77.13 | 81.20 | 78.67 | 74.27 | 85.74 | 74.14 | 74.78 | 81.10 | 77.20 |
| | C3 | 0.5 | 0.6 | 74.27 | 78.47 | 75.86 | 51.45 | 75.95 | 58.36 | 70.72 | 76.15 | 72.60 |
| EPFL-wildtrack | C1 | 0.8 | 0.7 | 87.97 | 86.83 | 86.31 | 59.81 | 71.38 | 61.80 | 58.25 | 92.22 | 70.07 |
| | C2 | 0.8 | 0.6 | 69.01 | 84.95 | 85.57 | 63.99 | 54.59 | 57.27 | 56.12 | 92.73 | 68.31 |
| | C3 | 0.8 | 0.7 | 90.58 | 87.91 | 87.96 | 53.86 | 40.93 | 45.21 | 54.87 | 90.66 | 66.73 |
| | C4 | 0.6 | 0.8 | 89.35 | 84.37 | 85.54 | 42.84 | 29.12 | 35.06 | 53.84 | 85.97 | 64.29 |
| | C5 | 0.8 | 0.8 | 59.86 | 91.38 | 69.91 | 39.56 | 81.10 | 50.99 | 40.78 | 91.93 | 54.88 |
| | C6 | 0.8 | 0.8 | 60.92 | 76.06 | 65.27 | 49.97 | 36.58 | 39.54 | 36.25 | 89.31 | 49.64 |
| | C7 | 0.5 | 0.7 | 88.17 | 87.73 | 86.96 | 41.37 | 94.23 | 55.63 | 57.57 | 88.17 | 68.44 |
| OxTown | - | 0.5 | 0.8 | 82.98 | 82.30 | 81.04 | 37.38 | 95.16 | 51.78 | 42.46 | 83.78 | 54.57 |

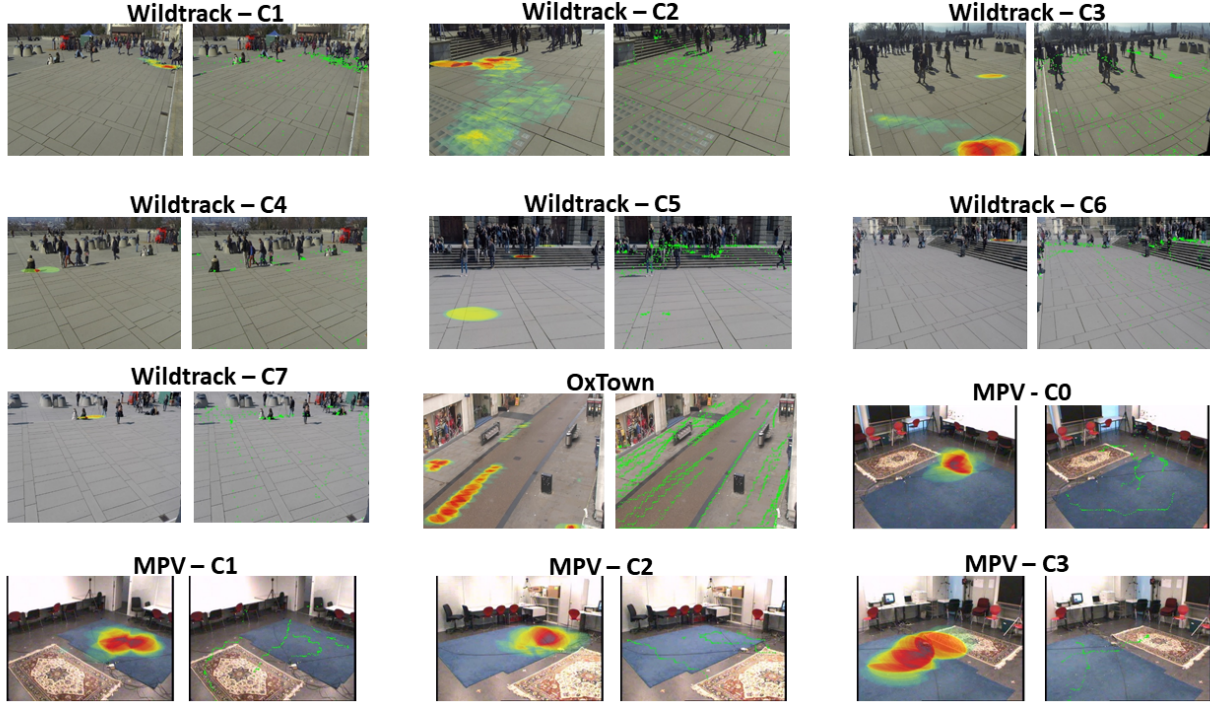


Figure 2. Image samples from all the sequences of the three datasets with the heatmap of social distancing violations overlaid.

Table 2. Investigating ‘Leg’ body part as the metric reference choice, using the fixed homography

| Dataset | Seq. | ρ_h | ρ_v | Body parts - Leg length | | |
|----------|------|----------|----------|-------------------------|--------|----------|
| | | | | Precision | Recall | F1-score |
| EPFL-mpv | C0 | 0.6 | 0.5 | 74.91 | 80.67 | 77.14 |
| | C1 | 0.5 | 0.6 | 72.63 | 77.97 | 74.80 |
| | C2 | 0.8 | 0.5 | 77.30 | 80.45 | 78.43 |
| | C3 | 0.8 | 0.5 | 73.31 | 74.54 | 73.41 |
| EPFL-WT | C1 | 0.5 | 0.8 | 76.69 | 95.12 | 83.93 |
| | C2 | 0.7 | 0.5 | 72.12 | 96.12 | 83.93 |
| | C3 | 0.5 | 0.8 | 76.28 | 95.56 | 83.57 |
| | C4 | 0.6 | 0.8 | 86.94 | 87.69 | 85.95 |
| | C5 | 0.8 | 0.8 | 56.89 | 93.13 | 68.40 |
| | C6 | 0.8 | 0.7 | 51.09 | 89.08 | 62.98 |
| | C7 | 0.6 | 0.8 | 94.10 | 82.85 | 87.20 |
| OxTown | - | 0.5 | 0.8 | 80.61 | 87.62 | 82.59 |

Table 3. Investigating ‘Arm’ body part as the metric reference choice, using the fixed homography

| Dataset | Seq. | ρ_h | ρ_v | Body parts - Arm length | | |
|----------|------|----------|----------|-------------------------|--------|----------|
| | | | | Precision | Recall | F1-score |
| EPFL-mpv | C0 | 0.6 | 0.5 | 75.61 | 74.88 | 74.68 |
| | C1 | 0.5 | 0.6 | 72.09 | 71.26 | 71.19 |
| | C2 | 0.8 | 0.5 | 72.30 | 69.05 | 70.02 |
| | C3 | 0.8 | 0.5 | 69.74 | 64.61 | 66.38 |
| EPFL-WT | C1 | 0.5 | 0.8 | 84.62 | 81.17 | 81.46 |
| | C2 | 0.7 | 0.5 | 81.36 | 82.80 | 80.47 |
| | C3 | 0.5 | 0.8 | 87.16 | 79.28 | 81.00 |
| | C4 | 0.6 | 0.8 | 86.87 | 66.59 | 73.30 |
| | C5 | 0.8 | 0.8 | 66.14 | 86.28 | 72.58 |
| | C6 | 0.8 | 0.7 | 59.61 | 73.57 | 63.53 |
| | C7 | 0.6 | 0.8 | 91.99 | 68.44 | 76.75 |
| OxTown | - | 0.5 | 0.8 | 82.66 | 69.01 | 73.03 |

Table 4. Investigating ‘Torso’ body part as the metric reference choice, using the fixed homography

| Dataset | Seq. | ρ_h | ρ_v | Body parts - Torso length | | |
|----------|------|----------|----------|---------------------------|--------|----------|
| | | | | Precision | Recall | F1-score |
| EPFL-mpv | C0 | 0.6 | 0.5 | 76.44 | 79.88 | 77.64 |
| | C1 | 0.5 | 0.6 | 74.43 | 77.03 | 75.38 |
| | C2 | 0.8 | 0.5 | 77.11 | 77.88 | 77.12 |
| | C3 | 0.8 | 0.5 | 72.24 | 71.22 | 71.19 |
| EPFL-WT | C1 | 0.5 | 0.8 | 79.39 | 91.80 | 84.10 |
| | C2 | 0.7 | 0.5 | 76.79 | 92.38 | 82.64 |
| | C3 | 0.5 | 0.8 | 82.06 | 91.89 | 85.37 |
| | C4 | 0.6 | 0.8 | 89.58 | 86.16 | 86.68 |
| | C5 | 0.8 | 0.8 | 59.47 | 91.82 | 69.82 |
| | C6 | 0.8 | 0.7 | 53.86 | 84.63 | 63.72 |
| | C7 | 0.6 | 0.8 | 93.07 | 78.36 | 84.03 |
| OxTown | - | 0.5 | 0.8 | 82.98 | 82.30 | 81.04 |

Table 5. Investigating ‘BBX height’ as the metric reference choice, using the fixed homography

| Dataset | Seq. | ρ_h | ρ_v | Body parts - BBX height | | |
|----------|------|----------|----------|-------------------------|--------|----------|
| | | | | Precision | Recall | F1-score |
| EPFL-mpv | C0 | 0.6 | 0.5 | 73.44 | 81.59 | 76.71 |
| | C1 | 0.5 | 0.6 | 70.85 | 78.39 | 73.90 |
| | C2 | 0.8 | 0.5 | 76.63 | 81.65 | 78.63 |
| | C3 | 0.8 | 0.5 | 73.38 | 77.24 | 74.69 |
| EPFL-WT | C1 | 0.5 | 0.8 | 68.05 | 97.77 | 79.18 |
| | C2 | 0.7 | 0.5 | 71.97 | 96.56 | 81.18 |
| | C3 | 0.5 | 0.8 | 68.10 | 98.34 | 79.13 |
| | C4 | 0.6 | 0.8 | 74.15 | 94.55 | 81.52 |
| | C5 | 0.8 | 0.8 | 50.64 | 96.81 | 64.85 |
| | C6 | 0.8 | 0.7 | 54.26 | 93.09 | 59.73 |
| | C7 | 0.6 | 0.8 | 87.93 | 87.16 | 86.62 |
| OxTown | - | 0.5 | 0.8 | 63.18 | 90.12 | 72.38 |

Table 6. Body part joint index correspondence

| Body part | Corresponding joints |
|-----------|----------------------|
| Right arm | 5,6,7 |
| Left arm | 2,3,4 |
| Right leg | 12,13,14,19 |
| Left leg | 9,10,11,22 |
| Torso | 1,8 |