

## Appendix – Movement Analysis Parameters and Calculations

Based on the semantic trajectory data and semantic point information generated in the previous steps, several types of movement analysis and classification can be performed, which help to understand general or location-specific area characteristics in case study research.

### Movement speed coefficient

Following the basic principle to calculate speed by dividing distance by time, we calculate the movement speed of each pedestrian based on the trajectory points as shown in Eq. (1):

$$speed_j = \frac{\sum_{i=1}^{len(t)} |p_i - p_{i-1}|}{len(t)} \quad (1)$$

where  $t_j = \{p_1, p_2, \dots, p_{len(t)}\}_j$  represents the  $j$ -th trajectory consisting of  $len(t)$  trajectory points, and  $|p_i - p_{i-1}|$  is the Euclidean distance between two consecutive points derived from their geolocation coordinates.

We also calculate the average of these movement speeds as shown in Eq. (2):

$$sp = \frac{\sum_{j=1}^n speed_j}{n} \quad (2)$$

where  $n$  is the number of trajectory segments.

### Staying detection

To detect staying behaviours such as resting or talk to others, we evaluate both the speed of movement and the duration of time in which this condition occurs. Duration-based methods allow for the use of a threshold, which improves the accuracy of staying detection (Yan & Forsyth, 2005). The calculation of staying time is shown in Eqs. (3) and (4):

$$d_i = \begin{cases} e_i - s_i & t_i \in T_{stay} \\ 0 & t_i \in T_{moving} \end{cases} \quad (3)$$

where  $T_{stay}$  is the trajectory set of staying people,  $t_i$  is the  $i$ -th trajectory in the set,  $e_i, s_i$  are the end frame and start frame of  $t_i$ ,  $d_i$  is the staying duration of  $t_i$ .

$$dur = \sum_{i=1}^{len(T_{stay})} d_i \quad (4)$$

where  $dur$  is the duration of stay.

### Movement categorization

To categorize pedestrian movements into categories of “divergent” and “non-divergent” movements, we calculate the maximum deviation ( $de$ ) of each trajectory. Deviation is defined by measuring the distance between trajectory segments and a straight line representing the theoretical shortest path.

For each trajectory:

$$de_i = \sum_{j=1}^n \frac{(x_i - x_1) \times (x_i - x_n)}{(x_n - x_1)} \quad (5)$$

Where  $de_i$  is the distance from point  $i$  in the trajectory to the straight line connecting the start and end points,  $x_1$  and  $x_n$  are the start and end point,  $n$  is the number of trajectory segments.

The  $maxde_i$  is the maximum distance in one trajectory. We set a  $threshold_i$ :

If  $maxde_i > threshold_i, tra_j \in TRA_{divergent}$

If  $maxde_i \leq threshold_i, tra_j \in TRA_{non-divergent}$

### **Behaviour intensity mapping**

Heat maps were produced by identifying the social condition detection points in each analysis grid and referencing the distance between the two persons creating this condition. If multiple detection points were found inside the same grid cell, the one with the lowest value was selected (Eq. 6).

For each behavior intensity detection:

$$Prox_x = distance(A, B)_{min} \quad (6)$$

Where  $Prox_x$  is the proximity value of detected behavior in grid  $x$ ,  $distance(A, B)$  is the proximity distance between trajectory  $a$  and  $b$ ,  $distance(A, B)_{min}$  is the minimum proximity value of two trajectories in grid  $x$ .  $In_x$  is the intensity of detected behavior in grid  $x$ .

### **Detection intensity analysis**

The average density of detections per minute is derived as:

$$In_x = \frac{\sum_{i=1}^{len(t)} (N_i + N_{i+1} + \dots + N_{len(t)})}{len(t)} \quad (7)$$

where  $In_x$  is the intensity of detected behaviour in grid  $x$ ,  $(N_i + N_{i+1} + \dots + N_{len(t)})$  is the accumulative detected behaviours from timestep  $i$  to  $len(t)$ ,  $len(t)$  is the detected time duration within grid  $x$ .