



Modeling pedestrian emotion in high-density cities using visual exposure and machine learning: Tracking real-time physiology and psychology in Hong Kong

Luyao Xiang^a, Meng Cai^{a,*}, Chao Ren^{b,c}, Edward Ng^{a,c,d}

^a School of Architecture, The Chinese University of Hong Kong, Hong Kong

^b Faculty of Architecture, The University of Hong Kong, Hong Kong

^c Institute of Future Cities, The Chinese University of Hong Kong, Hong Kong

^d Institute of Environment, Energy and Sustainability, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Keywords:

Ambulatory pedestrian emotion track
Environment perception
Street view
Isovist
Random forest
Deep learning

ABSTRACT

People's mental health has been deteriorating as a result of urban living, which also causes disability, pain, or even death among metropolitan residents. Understanding the interaction between human emotions and the built environment is therefore essential for developing strategies towards a psychologically friendly and socially resilient environment. This study aims to model the pedestrian emotion in high-density urban areas of Hong Kong using machine learning based on environmental visual exposure, which is one of the most significant factors affecting emotions. Using ambulatory sensing and portable technologies, two-dimensional emotional data from 99 pedestrians are retrieved from coupled data with respect to wearable arousal, subjective report preference, and the location at a high spatiotemporal resolution (4 m). The visual environment is quantified by isovist and street view factors. This study examines the impact of visual exposures using two multiple linear regression (MLR) models and establishes a predictive model using random forest (RF) (N = 548). The MLR models have R² values around 0.5 and suggest that in the high-density environment, exposure to more trees, visual volume, and drift magnitude can cause positive emotions. Conversely, areas with a view of sign symbols, object proportion, min-radial, and occlusivity can cause negative emotions. The resultant predictive model with nine visual exposure variables can explain 79% of the spatial variance of pedestrian emotion. Furthermore, the methodological framework provides opportunities for spatial, data-driven approaches to portable sensing and urban planning research. The findings are also applicable to optimize the infrastructure design of outdoor environments for more psychologically friendly experiences.

1. Introduction

Good mental health contributes to social and psychological well-being. Determinants of mental health and mental disorders include not only individual attributes but also social, cultural, economic, political, and environmental factors, such as national policies, social protection, living space, working conditions, and community social support [1]. A group of neuroscientists [2,3] have identified two typical urban environments that affect mental health: 1. an environment that features high complexity, high heterogeneity, and a high rate of change; 2. the shortened social distance and invasion of personal space. They thereby call for interdisciplinary cooperation to enhance the built environments and improve the deteriorating mental health conditions of metropolitan

residents.

1.1. Background of emotion

Emotion is an essential element of mental health and can be conceptualized by categorical [4] and dimensional [5] approaches. The categorical approach assumes that emotions are the basic and discrete components hard-wired in our brain and can be measured by self-report methods [6]. However, Bastiaansen, et al. [7] asserted that emotion can develop continuously over time, which may not be demonstrated by the categorical approach. In contrast, the dimensional models suggest that a common and interconnected neurophysiological system is responsible for all affective states [8]. It means that in everyday interactions, people tend to show non-basic, subtle, and quite complex affective states. This

* Corresponding author.

E-mail address: caimeng@link.cuhk.edu.hk (M. Cai).

<https://doi.org/10.1016/j.buildenv.2021.108273>

Received 14 May 2021; Received in revised form 16 August 2021; Accepted 17 August 2021

Available online 27 August 2021

0360-1323/© 2021 Elsevier Ltd. All rights reserved.

Nomenclature	
PAD	Pleasure – Arousal – Dominance
VAD	Valence-Arousal-Dominance
SC	Skin Conductance
EEG	electroencephalography
GEMA	Geographic Ecological Momentary Assessment
MLR	Multiple Linear Regression
RF	Random Forest
TST	Tsim Sha Tsui
HH	Hong Hum
SD	Standard Deviation
GPS	Global Positioning System
SCR	Skin Conductance Response
CDA	Continuous Deconvolution Analysis
CamVid	Cambridge-driving Labeled Video Database
FOV	Fields of Vision
SkyVF	Sky View Factor
BVF	Building View Factor
RVF	Road View Factor
SSVF	Sign symbol View Factor
CVF	Car View Factor
PVF	People View Factor
TVF	Tree View Factor
OOB	Out-of-Bag
MSE	Mean Squared Error
IoU	Intersection over Union
RMSE	Root Mean Square Error
MAE	Mean absolute error

dimensional perspective is the foundation for the measurement of verbal self-report: Pleasure – Arousal – Dominance (PAD) (or valence-arousal-dominance, VAD) scale. The valence dimension ranges from negative to positive, and the arousal dimension ranges from calm to excitement. The dimensional model can locate discrete emotions without a specific category definition, so it has been widely used in emotion recognition experiments [9].

The valence data can be obtained through subjective questionnaires or more advanced sensors, i.e., electroencephalography (EEG). The arousal dimension of emotion is often measured by skin conductance (SC), an unbiased and continuous measure of sympathetic activity via the electric impulses on the skin surface and sweat glands, which are innervated only by the sympathetic nervous system and is related to the increase in the arousal of a subject. The construction of valence and arousal dimensions can provide a more sufficient explanation for the structure of emotions [10,11].

1.2. Visual exposure

The human living environment is an important source of stimuli that can evoke different emotional responses [12]. Cognitive science research shows that human beings acquire external information mostly through vision [14], which is the main driver behind human perception [13] and consequent experience [14]. As human cognitive responses to urban space and form are closely related to their behaviors, visual exposure to urban space, therefore, is one of the most important factors that affect their perceived quality and wellbeing [13].

The concept of isovist has been widely used to describe and quantify visual exposure. Generally, it can be defined as “the set of all points visible from a given vantage point in space” [15]. However, isovist requires detailed 3D models of trees, buildings, and pedestrians, which are often unavailable or difficult to acquire. The recent development of street sensing techniques has made it possible to describe the visual environment by open-source street view images that are easily accessible. The necessary visual exposure information can be extracted from these images using certain photographic or deep learning techniques [16,17]. Thus, both isovist and street view images have been frequently adopted to quantify visual exposure.

1.3. Literature review

Understanding the effects of the visual environment on human emotions is essential for developing strategies towards psychologically friendly and socially resilient environments. Besides, prior knowledge of pedestrian emotion can facilitate informed control decisions of urban design. Accordingly, several studies have been carried out to analyze and predict the emotions of people from visual exposure in both

laboratories and outdoor environments.

Parametric studies conducted in the laboratories have achieved the best results by combining the methods of subjective questionnaires and showing the scenes with different techniques. Stamps [18] focused on visual responses to openness, and in the following studies [19,20], he considered four visual properties of the physical environment and established the correlation between isovists parameters and psychological feelings based on the prospect refuge theory [21]. Lindal and Hartig [22,23] performed a parametric study to examine the effects of streetscapes entropy, building height, and arrangement of trees on the evaluation of restoration possibilities. Other characteristics, such as orientation [24,25], and shape complexity [26], had also been examined in previous studies. Moreover, some studies adopted virtual reality techniques to simulate the real environment for better immersive feelings and easier control of the variables [27–29].

The development of mobile sensors, such as ambulatory sensing and portable technologies, has enabled researchers to measure emotions outdoors under moving conditions [29,30]. Field surveys, such as the Geographic Ecological Momentary Assessment (GEMA) [31–33], can link physiological measures of arousal with geographical location for a more realistic and comprehensive analysis of the interaction between humans and the environment [34–36]. The interest in emotion mapping has since grown within urban spatial analytics [29,30,35,37–43]. Most of these studies were based on urban walking experiences, and the measured data were usually interpreted in terms of proximity to urban features. In addition, the cycling experience has been examined in some studies with similar methods [30,44].

Further to mapping the emotion data, several studies [29,30,39–41] explored the statistical relationship between visual exposure and emotions. However, most of the statistical analysis of visual environment and emotion only measured the arousal dimension of emotions with the ambulatory assessment SC technique but did not include the measurement of valence, posing a limitation on their further implication because of the lack of positive/negative emotion information. Moreover, visual exposure has been considered an important source of emotion. While both the concept of isovist [29,39,40] and street views [30] can quantify visual exposure, the former emphasizes the geometric characteristics of the visual area while the latter contains information on the visible content. However, none of the previous studies have simultaneously applied the two concepts, leading to an incomplete understanding of the visual contents and a low R^2 value of the resultant model. In addition, most of them adopted multiple linear regression (MLR) to model emotions from visual exposure but the present study proposes the adoption of a more complicated model because their relationship may not be linear. Therefore, a machine learning algorithm, i.e., the random forest (RF), is used to model emotion. Finally, most of the previous studies were carried out in medium- and low-density cities and data from

high-density cities have been rare.

1.4. Objectives

Based on the research gaps, this study aims to understand the effects of visual environment on emotion in high-density environments more comprehensively and accurately, by quantifying two-dimensional emotion, using multi-level visual exposures from isovist and street view images, and modeling with machine learning. This study can inform practitioners across multiple disciplines, including urban planners, architects, psychologists, decision-makers, with cutting-edge and accurate urban planning schemes for a psychologically friendly environment.

2. Materials and methods

This study contains four steps. First, subtle and controlled experiments are conducted in high-density urban areas to collect two-dimensional emotion data in real-time with the ambulatory technique. Second, visual exposure variables from isovist and street view images are collected and calculated to characterize the visual environment that the participants see along the pre-defined routes. MLR is then used to examine how environmental visual exposure affects the emotions of pedestrians. Finally, a data-driven prediction model is established for pedestrian emotion estimation using random forest. The framework of this study is shown in Fig. 1.

2.1. Study area

Hong Kong is located on the southeastern coast of China and has a population density of more than 7000 people per square kilometer, ranking the third highest in the world [45]. Only 24% of the total land area of Hong Kong has been developed. The high-density planning makes the living environment cramped and impedes urban ventilation, generating heat islands and air pollution. In addition, the deteriorating living environment can affect mental health. Nearly 60% of the population has poor mental health according to the results of the Territory-wide Mental Health Index Survey 2020 [46].

Located in the central part of Hong Kong, the Kowloon Peninsula has the highest building density in the entire city (Fig. 2). Two representative districts with the highest building volume density of Kowloon have

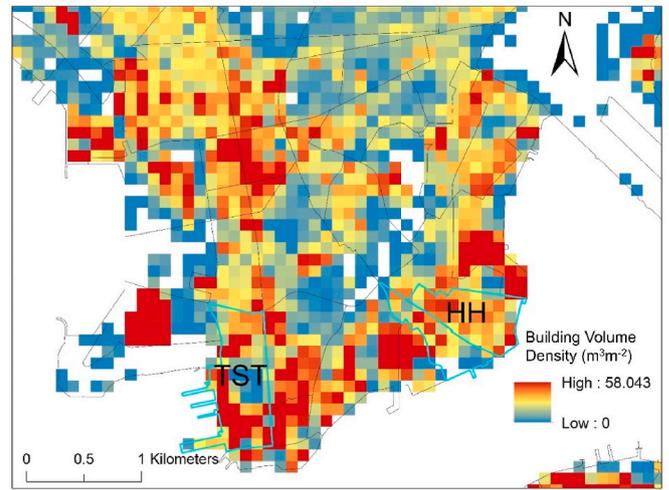


Fig. 2. Density situation of Kowloon.

been selected for this study. The first district is Tsim Sha Tsui (TST), which is famous for its commercial atmosphere and attracts a huge volume of visitors every day because of the abundance of restaurants and fashion retailers. Trees can hardly be found on the streets and the rich business atmosphere of TST makes it one of the most crowded places in Hong Kong. The other district is Hung Hom (HH), a residential area with a local living atmosphere. The first part of the HH route comprises mainly natural landscapes with just a few buildings. The atmosphere of this route is quiet and sparsely populated because there are no commercial activities. The second part of the HH route is of typical urban style with high-density and high-rise residential buildings, food markets, and communal facilities. The TST route is 1.1 km and the HH route is 1.3 km (Fig. 3).

2.2. Experimental details

2.2.1. Participant recruitment

Participants were recruited via an online announcement detailing a reward of 60 HKD for each participant. Participants should be between 18 and 35 years old, should have resided in Hong Kong for no longer than five years, and have not been involved in severe accidents in the

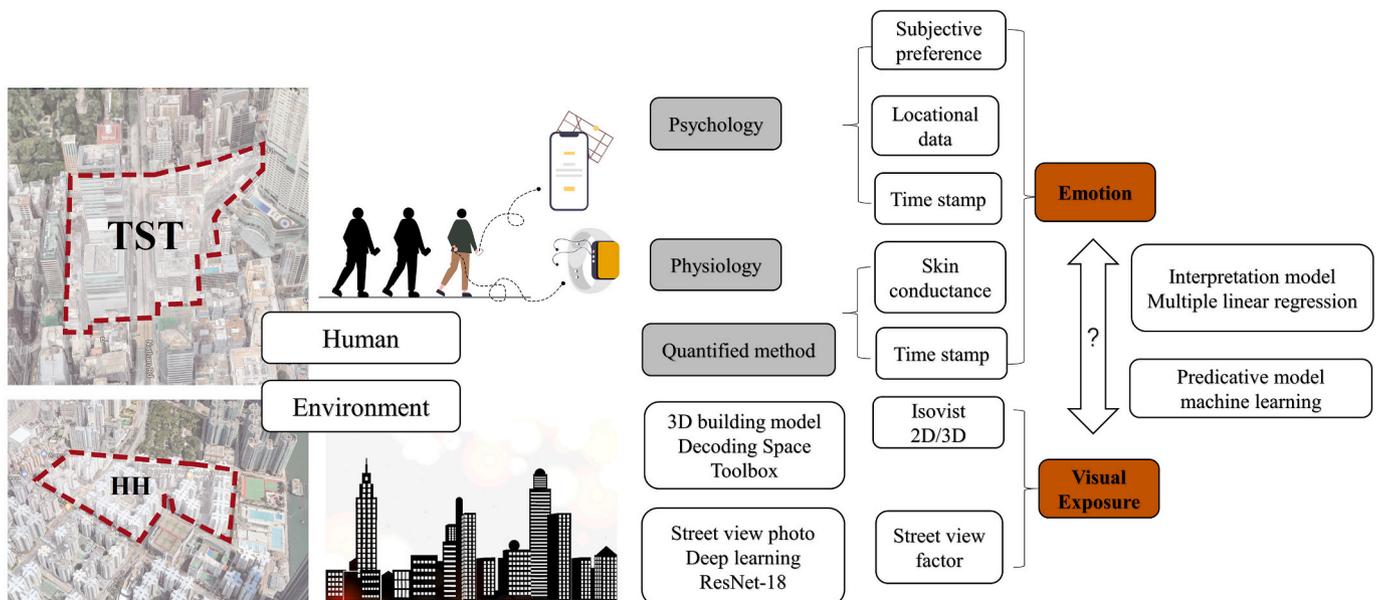


Fig. 1. The framework of this study.

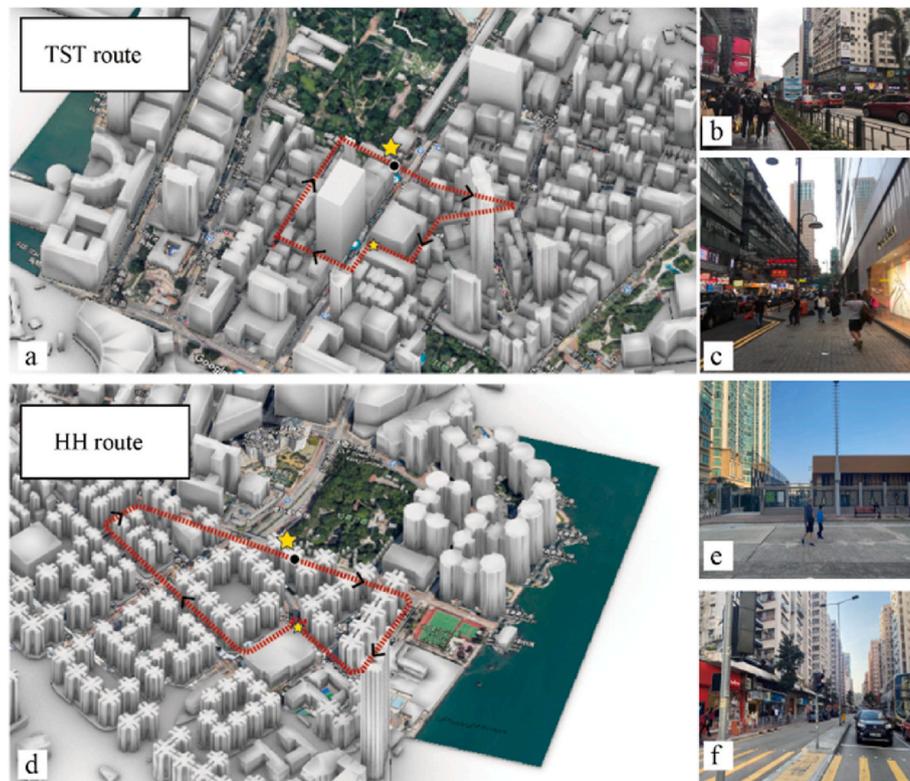


Fig. 3. (a&d) are route maps in TST and HH, respectively. The big star is the gathering point where the pre-walk test takes place. The data starts to be recorded at the black point. There is also a small star on the map, indicating the location for the mid-walk test; (b&c) are the representative scenes for the TST route - first and second part, respectively; (e&f) are the representative scenes for the HH route - first and second part, respectively.

past six months. A total of 103 participants enrolled, including 51 males and 52 females. The average age of the participants is 26 (Standard Deviation (SD) = 5). The participants were randomly allocated to the different experimental routes. Finally, 51 and 48 people took part in the TST route and HH route, respectively. 4 people terminated the experiment midway because of personal reasons.

2.2.2. Experiment time

The weather in Hong Kong can be extremely uncomfortable during the mid-summer months, i.e., from May to August, when the weather is hot and humid, and thus, no experiments were conducted during these periods. The current study lasted from December 2018 to May 2019. The weather for all experimental days was relatively comfortable, and the mean of the minimum air temperature was 17.67 °C, SD = 3.99 °C, the maximum was 24.08 °C, SD = 4.28 °C (data source: HKO, <https://www.hko.gov.hk/en/index.html>). In addition, experiments were only conducted on weekdays taking into account the differences in the street crowdedness and human activities on weekdays and weekends. Rush hours were also avoided to minimize the effects of crowding and congestion, such as by car horns. Therefore, the periods of 10:00–12:00 a.m. and 2:00–4:00 p.m. were chosen for conducting the experiments. The experiment would be canceled under rainy weather.

2.2.3. Experiment design

On each experimental day, a maximum of three participants took part in the experiment at the same time. No participants walked along the same route twice, and even for those who took both routes, the combination of the participants was different each time. The participants were sent an SMS reminder the night before each experimental day. All participants arrived at the gathering points by themselves. Upon arrival, the participants were given an introduction of the experiment after which they were asked to sign a written consent for their voluntary participation. The participants wore the devices to measure SC with the

help of the instructor. The pretreatment started with 100 s of relaxation, during which the SC baseline of the participant in the standing posture was recorded [47]. The first 10 min were spent standing quietly near the gathering points. The participants viewed the corresponding landscapes while the SC was recorded at the frequency of 10 Hz. The participants were then accompanied by an experimental instructor on a 40-min walk. At the outset, the instructor explained that conversation should be minimized to ensure consistency across subjects. To avoid distraction from the nearby participants, the participants were told to keep a 5-m distance from each other. The instructor led the participants by walking 10 m ahead of them at a slow pace. SC was recorded continuously during the entire course of the walk. Participants used their mobile phones to express their preference for the visual scenes along the street whenever they were emotional because of the streetscapes. Thereafter, they turned back to the starting points as the routes were designed as a circle and finished the post-test. The experimental procedure is summarized in Fig. 4.

2.3. Mobile data collection

2.3.1. Locational data

The geographic coordinates of the participants and the corresponding timestamps throughout their walk on the predefined route were collected by a dedicated Global Positioning System (GPS) device (GARMIN GPSmap 62s) attached to their shoulders with 1 Hz frequency. The locational data were then combined with the real-time SC data, street views, and urban form data.

2.3.2. Arousal

Shimmer 3 GSR + device (Shimmer, www.shimmersensing.com/) was used to measure SC (Appendix, Fig. A1). The main parts of the devices were worn on the wrist of the participants, attached by comfortable bands, comprising two 8-mm snap style finger type

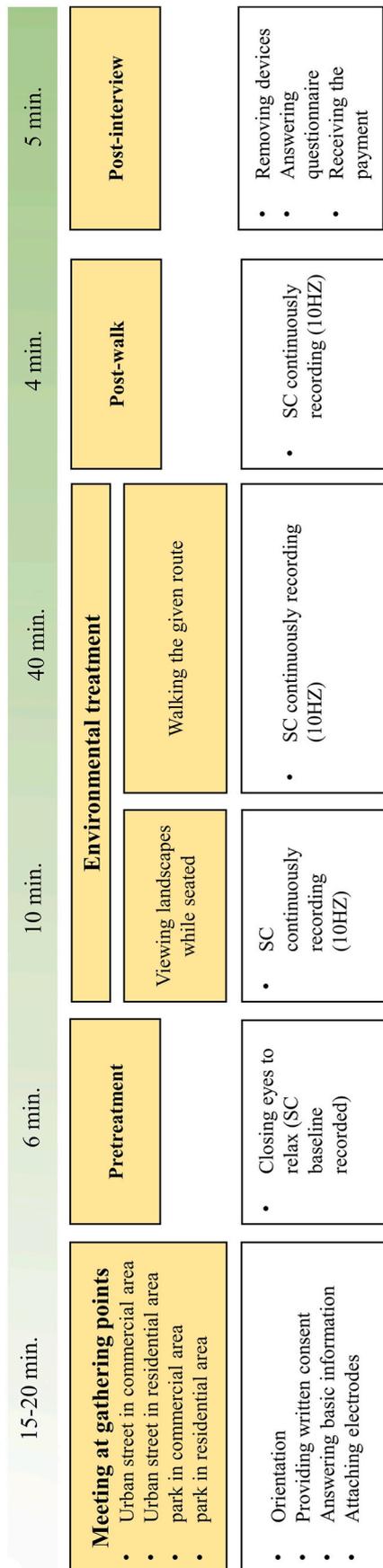


Fig. 4. Experiment procedure.

Ag–AgCl dry electrodes (GSR electrodes, Shimmer Sensing), with a constant voltage (0.5 V) attached to the intermediate phalanges of their left middle and index fingers, and GSR sampling at 10 Hz. The device has been clinically tested and has shown high validity and accuracy [48].

2.3.3. Subjective preference

The participants installed a mobile application provided in advance. At the very beginning of the experiment, the participants were given the following instructions: ‘if you like the viewing scenery in front of you and you feel pleased, please push the green button on your mobile phone screen/if you dislike the viewing scenery in front of you and you feel displeased, please push the yellow button on your mobile phone screen/if you have no feeling of both above choices, please push the grey button on your mobile phone screen. Please make a choice enthusiastically whenever you have any above feelings along the route’. Accordingly, both the preference and information on the level of pleasure of the participants were obtained, which can be used to describe valence.

The green button (like/pleased) was marked as 1, the grey button (neutral) as 0, and the yellow button (dislike/displeased) as -1. The participants were required to maintain a high-frequency reporting rhythm.

2.4. Emotional data processing

2.4.1. Extract skin conductance response (SCR) with ledalab, continuous deconvolution analysis (CDA) method

Each record was analyzed using the MATLAB 2020b, Ledalab V3.4.8 toolbox (www.ledalab.de). Before importing the raw SC data into Ledalab, a moving window of 10 s was applied on the unprocessed SC data to identify potential motion artifacts defined as three standard deviations or more. To visually inspect whether the deviations failed to confirm a standard, physiologically plausible shape for an SCR (a decline that lasted at least 3 or more seconds after a peak) was replaced with linear interpolation. Using Ledalab, the data was first downsampled to 1 Hz and high and low filters were used to reduce the results caused by motion artifacts. Ledalab implements the method of CDA to separate the phasic and tonic components of the raw data. The default setting of CDA was used to export the event-related activation with an SCR window of 1–4 s (start–end) with 4-s intervals. The minimum amplitude criterion of SCRs was set as 0.04 mμS, larger than the default setting (0.01 mμS), to achieve a balance between the sensitive detection of SCRs and to minimize the effects of movement artifacts [49].

2.4.2. Spatial, temporal, and emotional aggregative data

2.4.2.1. Standardize SCR data. We obtained SCR data every 4 s. The raw SCR data need to be transformed into a standardized and corrected form that can facilitate individual-difference analysis. This study adopted the most common method that assumes the minimum value to be zero and the maximum value to be the result of some stimulus. Each SCR was standardized for individual differences by dividing it by the participant’s maximum SCR [49].

2.4.2.2. Fill in the missing subjective preference data. Since the participants reported preference at random incidences along the routes, the dataset built at 4 m resolution may contain some points with no data because no participants made any choice at that place. 0 was assigned as the preference value assuming they felt neutral about those points when they did not make a choice.

2.4.2.3. Combine the SCRs and subjective preference data into emotional data. In order to distinguish the different states of valence and arousal, we quantified emotion based on the vector model [50] by multiplying the SCR and subjective preference for each participant at each point (Equation (1)) to obtain two-dimensional emotional responses,

assuming that there is always an underlying arousal dimension, and that valence determines the direction in which a particular emotion lies. Previous studies [29,51] also used the method of multiplying these two dimensions to obtain the emotion value for further analysis. In this model, high arousal states are differentiated by their valence, whereas low arousal states are more neutral and are represented near the meeting point of the vectors.

$$\text{Emotion}_{ij} = \text{SCR}_{ij} * \text{Preference}_{ij} \tag{1}$$

where *i* refers to the participants, *j* refers to the location points.

2.5. Semantic segmentation of street view images by deep learning

To obtain the street view photos at 4-m intervals, a Canon 6D-mark2, EF 24–70 mm f/2.8L II USM was used. The average focal length of each photo is 43 mm, which is the closest to the human eyes. There were 548 street view photos in total.

Semantic segmentation with deep learning has been widely adopted to extract information from images [52]. In this study, the semantic segmentation technique is applied to the 548 street view photos for the extraction of visual exposure information such as the exposure to trees, buildings, and sign symbols (Fig. 5). First, the photos were resampled from 6240*4160 pixels to 1260*840 pixels to reduce computation time while retaining the maximum spectral information. Second, we adopted a pre-trained network based on the Cambridge-driving Labeled Video Database (CamVid) [53] and the Deeplab v3+ network [54] with

weights initialized from a pre-trained Resnet-18 network (<https://www.mathworks.com/supportfiles/vision/data/deeplabv3plusResnet18CamVid.mat>). The pre-trained Deeplab v3+ network was applied to the 548 street view photos to classify them into 32 classes of the CamVid. Thereafter, the photos in 32 classes were reclassified into seven classes that include the most common and basic elements in urban areas, namely, sky, building, road, sign symbol, car, people, and tree.

The view factor of each class was determined as the percentage of the class pixels and the total pixels of the picture (1260*840). Finally, 15 out of 548 photos were used to validate the segmentation results. We manually labeled 15 photos as independent ground-truth with three classes, i.e., tree, sky, and urban (an aggregation of the other five classes), and cross-compared the labeled photos with the corresponding segmentation images. The semantic segmentation and accuracy assessment were performed on the platform of MATLAB 2020b.

2.6. Isovist calculation

The 2D and 3D-volumetric isovists were calculated in Rhino 6, Grasshopper Decoding Spaces Toolbox developed by the team of Computer Science in Architecture at Bauhaus University (Fig. 6). The building models used for calculation were adopted from a previous study [55].

The first step was to determine the length of the visual line and the standard visual height. The maximum visual line length is also used to determine the calculation limit. There is no strict standard for this

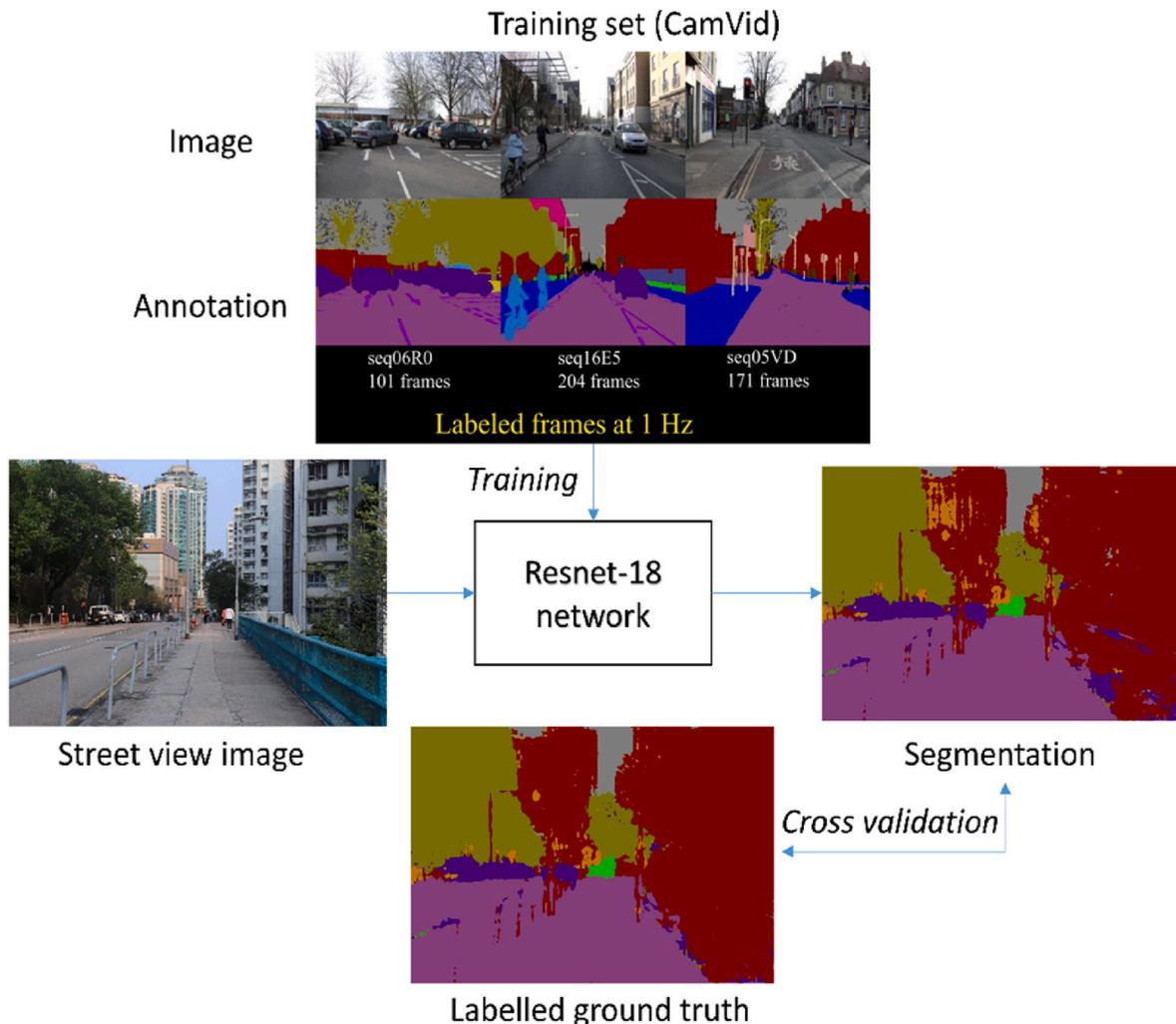


Fig. 5. Semantic segmentation workflow.

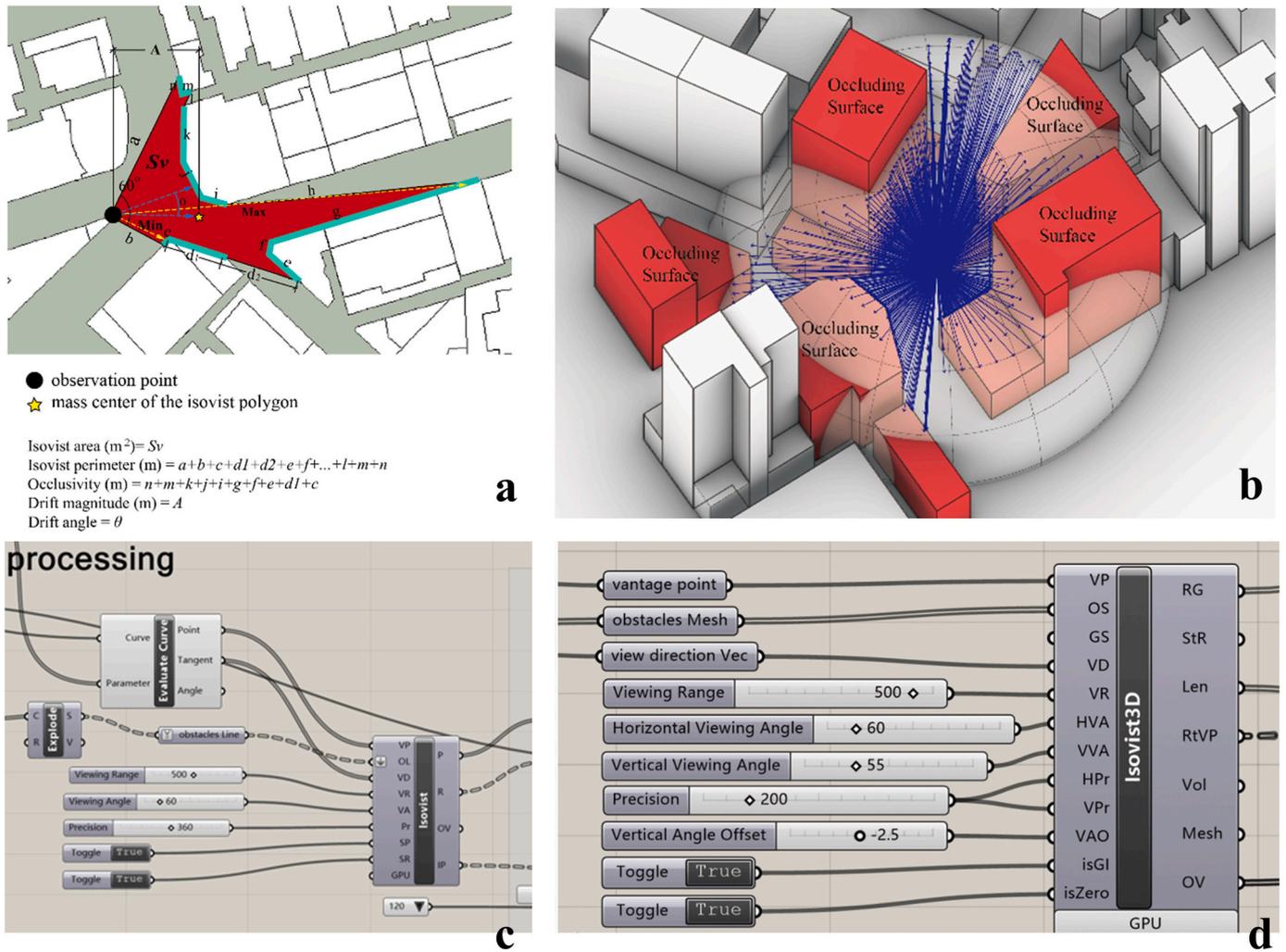


Fig. 6. (a) An example of 2D-isovist calculation point; (b) an example of 2D-isovist calculation point; (c) the basic setting for calculation 2D-isovsit; (d) the basic setting for calculation 3D-isovsit.

number [56]. The size of the blocks in the study areas varied from 50 m*200 m–250 m*250 m, and the maximum length of the experimental route was 500 m. Considering these aspects, the study adopted 500 m as the maximum visual length. The visual height adopted was 1.6 m, which is close to the average height of the participants.

The attention and perception of people were related to different fields of vision (FOV), and for both 2D and 3D isovist calculation, we set FOV_{vertical} as 60°, which is the center field of FOV. The standard sightline is horizontally forward, and the normal sight line while standing is 10° downwards.

2.7. Building the dataset of human movement, physiology, psychology & environment

By combining the spatial and emotional data of 99 participants, the emotionally evocative points in the streets were identified. The streets were divided into segments with a 4 m resolution to analyze the data. There were 211 points for the TST and 337 points for the HH route (548 points in total). For each point, we calculated the average emotion value of all participants (Equation (2)).

$$Emotion_j = \frac{\sum_{i=1}^{99} Emotion_{ij}}{99} \quad (2)$$

The visual exposure variables from isovist and street view are summarized in Table 1 as potential predictors. The variable ‘change

direction’ that describes the movement of the participants was also incorporated as a potential predictor since it may influence the visual environment as well as the emotions of the participants. Finally, the 29 potential predictors were geo-referenced and matched with the averaged emotion for each point.

2.8. Statistical analysis

2.8.1. Identifying the effects of visual exposure using multiple linear regression

The statistical analysis is based on a classic MLR model (Equation (3)):

$$Emotion_j = \alpha_1 Var_1 + \alpha_2 Var_2 + \dots + \alpha_n Var_n + \gamma + \epsilon \quad (3)$$

where $Emotion_j$ is the averaged Emotion value of all participants at the aggregation point j . The model includes n visual exposure metrics as the predictor variables. $\alpha_1 \dots, \alpha_n$ are the coefficient estimates of the metrics Var_1, Var_n at the aggregation point j . γ is the model intercept, and ϵ is the residual. The MLR model was constructed based on 29 candidate variables subset in Table 1, among which ‘change direction’ is the control variable, the other 28 visual exposure variables from street view and isovist categories are explanatory variables. The current study was conducted MLR in SPSS 16.0. Stepwise linear regression was applied. It performs multiple regression several times, removing the weakest correlated variable each time, and at the end, all significant variables are

Table 1
The potential predictors in statistical analysis.

Variable	Theoretical meaning	Calculation method ^a
Change direction	whether the participants change direction	1 for change, 0 for no change
Street view		
Sky view factor (SkyVF)	percentage of sky pixels of the street view image	$SkyVF = \frac{SkyPixel}{P}$
Building view factor (BVF)	percentage of building pixels of the street view image	$BVF = \frac{BldgPixel}{P}$
Road view factor (RVF)	percentage of road pixels of the street view image	$RVF = \frac{RoadPixel}{P}$
Sign symbol view factor (SSVF)	percentage of sign symbol pixels of the street view image	$SSVF = \frac{SSPixel}{P}$
Car view factor (CVF)	percentage of car pixels of the street view image	$CVF = \frac{CarPixel}{P}$
People view factor (PVF)	percentage of people pixels of the street view image	$PVF = \frac{PeoplePixel}{P}$
Tree view factor (TVFi)	percentage of tree pixels of the street view image	$TVFi = \frac{TreePixel}{P}$
2d_ismovist		
Isovist area (Sv)	the area of visual polygon area	N.A.
Isovist perimeter (Pv)	the total length of the visual boundary	N.A.
Compactness (Cv)	describes the compact and simple degree of the visual area	$Cv = \frac{4\pi A_v}{P_v^2}$
Circularity (Ci)	describes how well a space approximates a circle.	$Circularity = \frac{\pi r ^2}{A_v}$
Convex Deficiency (Cd)	The ratio of the area of the dent over the area of the hull	$C_d = \frac{P_v^2}{A}$
Elongation (ψ)	the ratio of the radius of an idealized circle associated with the actual area of the isovist to the radius of an idealized perimeter from the actual perimeter	$\psi = \frac{A_v^{1/2}}{P_v}$ $\psi = \frac{\pi}{2\pi}$
Occlusivity (Ov)	the difference between isovist perimeter and the overall length of the solid boundaries within the isovist area	$O_v = \frac{k}{nP_v} \sum_{i=1}^n \sum_{l=occ}^2 L_i$
Min-radial	the minimum visual radial line	N.A.
Max-radial	the maximum of visual radial line	N.A.
Mean-radial	the mean of visual radial line	$Q_v = \frac{1}{n} \sum_{i=1}^n L_i$
SD-radial	the SD of visual radial line	N.A.
variance-radial	the mean of the square of deviation between all radial lengths and average radial length of an isovist	$T_v = \sqrt{\frac{1}{n} \sum_{i=1}^n L_i - Q_v ^2}$
Skewness	the mean of the cube of deviation between all radial lengths and average radial length of an isovist	$S_v = \sqrt[3]{\frac{1}{n} \sum_{i=1}^n L_i - Q_v ^3}$
Dispersion	the difference between the values of the mean and the standard deviation of the isovist radial lengths	$D_v = M_v - SD_v$
Drift-magnitude	the distance between the observation points and the mass center of an isovist polygon	$M_v = \frac{(\sum_{i=1}^n L_i^2 \cdot R_i)}{(\sum_{i=1}^n L_i^2)}$
Drift-angle	the angle between the direction facing occupant and the mass center of an isovist polygon	N.A.
3d_ismovist		
Sky proportion	the proportion of visual lines to the sky and all visual lines	$Sky3D = \frac{S_i}{n}$
Object proportion	the proportion of visual lines to the objects and all visual lines	$Obj3D = \frac{O_i}{n}$
Ground proportion	the proportion of visual lines to the ground and all visual lines	$Grd3D = \frac{G_i}{n}$
Visual volume	the total length of all visual lines	$V_v = \sum_{i=1}^n L_i$
Log visual volume	The base 10 logarithms of the visual volume	$LogV_v = Lg V_v$

^a 'Li' provides radial length, 'n' is the total number of radials sampled, 'k' is the number of samples in one 360-degree cycle, 'Ei_occ' is the fraction of occluded edge detected, 'Qv' the average radial length from V, 'Ri' is the

coordinate of the nearest radial intersection, 'Xm and Ym' are the coordinates of 'M', 'Xv and Yv' are the coordinates of 'V', 'Si' is the number of visual lines reaching the sky, 'Oi' is the number of visual lines reaching the objects, 'Gi' is the number of visual lines reaching the ground.

included. We refined the independent variables by adopting the following rules: p-value < 0.05 and VIF <5 to include significant and non-collinear variables. To include as many variables as possible in the MLR model, collinear variables would be put into separate models to run the MLR.

2.8.2. Establishing a predictive model for pedestrian emotion using random forest

The RF model [57], an ensemble statistical algorithm composed of multiple decision trees, was utilized to establish the emotion prediction model. It is superior to linear and other machine learning algorithms in several aspects. First, it is suitable for dealing with a large number of variables because its random feature selection property makes it reliable and robust against correlated predictors. A random subset of predictors is used as split candidates for each tree. Second, the problem of overfitting is rare because it conducts bagging where multiple trees are generated using bootstrap samples and the majority votes are taken. Besides, RF can generate unbiased estimates without cross-validating or using an independent test set. It can be evaluated internally so that unbiased estimates of errors can be established during the generation process. The out-of-bag (OOB) error is the prediction error based on the trees that do not use a specific sample for training. Moreover, the results are relatively easier to explain because it can provide the importance of each variable. The RF regression is therefore a reliable, convenient, and moderately interpretable model that is suitable for the pedestrian emotion prediction in this study.

To establish the prediction model based on the visual exposure variables, all 548 observations were used to train the RF. First, the variable selection was performed using an RF-based approach to remove redundancy and determine the appropriate split predictor [58]. The selection process ranks the importance of all the variables and adds the variables one by one based on their importance to reach the accuracy of the full model. The importance is measured by the incremental mean squared error (MSE) because of the permuting OOB observations across each input variable. All 29 variables were selected for each split to avoid underestimating the importance of some variables.

Based on the selected predictor variables in this study, the set of parameters of the RF was tuned using Bayesian Optimization to identify the optimal model that minimizes the upper confidence interval of the MSE target model. In the tuning process, the minimum terminal node ranged from 1 to 274 (half of the 548 observations). The number of predictors to sample at each node started from 1 through all the selected predictors. The number of trees started from 50 and increased by 50 each time until it reached 500. Moreover, 10-fold cross-validation was conducted for the observations to protect against overfitting. The predictive model can be therefore established using the knowledge of the predictors and optimal parameter values. The RF regression was performed on the platform of MATLAB 2020b.

3. Results

3.1. Descriptive statistics

3.1.1. Emotional data: SCR and preference

In total, there were 99 trials of SC raw data, 23930 times of SC records, and 30067 times of self-report preference data from all participants. After processing the raw data through the method mentioned in section 2.4, the emotional dataset included 54252 (99*548) processed data. Finally, there were 548 emotion datasets after aggregating all data from participants at each point.

The distribution of the SCR and preference levels can be visualized

for each route through violin plots (Fig. 7). The ranges of data density are different between the two routes. Data are more concentrated between 0.3 and 0.5, and two relatively dense ranges are identified between 0.8-1.0 and 1.2-1.4 for the TST route. For the HH route, data is more concentrated from 0 to 0.2 and relatively uniform from 0.4 to 1.0.

In terms of preference, the means for both routes lie in the interval of 0–0.08 ($F(1,97) = 1.424, p < 0.01$), while the standard deviations are in the range 0.205–0.228. For the TST route, data are located most densely between 0 and 0.2, and the rest of the data are evenly distributed on both sides. For the HH route, data are located relatively scattered.

The distribution of emotion data (Fig. 8 (a)) shows that there are three points with frequent negative emotions for the TST route. The negative emotions in HH mainly occur at the big street crossings with heavy traffic. The positive emotions in the TST route mainly appear near the man-made park. The photos of the participants during the experiment are shown in Fig. 8(b–d).

3.1.2. Street view and isovist

The street view images were categorized into seven classes using semantic segmentation with deep learning (Fig. 9). Accuracy assessment from independent validation samples indicates an accuracy of 88.9%, 95.7%, and 87.7% for the sky, tree, and urban classes, respectively. The mean Intersection over Union (IoU) is 0.820. The validation results demonstrate that the semantic segmentation method adopted in this study is reliable and sufficient to detect necessary information from street view photos, especially for tree extraction.

Fig. 10 shows the selected predictors from street view and isovist for each point in TST and HH, including Tree View Factor, compactness, min-radial, and visual volume. For the Tree View Factor maps (Fig. 10 (a)&(b)), the points in HH generally have a higher tree proportion compared to those in TST, indicating greater exposure to trees along this route. The majority of the points in TST have a tree proportion lower than 10%. Only the points close to the park are exposed to a high tree coverage. For the min-radial maps (Fig. 10 (c)&(d)), the variation of HH points is larger than the ones in TST. Min-radial always has a strong fluctuation on each corner where the participants change their walking direction. Compactness, which describes the compact and simple degree of the visual area, is another important 2D isovist variable. For the TST route (Fig. 10 (e)), the smallest points are located beside a huge high-rise tower, and other points have significant differences compared to this part of the route. For the HH route (Fig. 10 (f)), the simplest area is identified near the sea. These conditions are in line with reality, which highlights the reliability of our calculated results. In terms of 3D isovist, we visualized visual volume data in Fig. 10(g)&(h). It refers to the degree of visual openness of the points. For both routes, the broadest areas are near the parks and the HH routes have significantly higher visual volume than TST.

3.2. Multiple linear regression model results

According to the VIF test for the 29 candidate variables, collinearity is identified between occlusivity and drift magnitude. We put these two variables in separate models to run the MLR and keep both models when reporting the results to incorporate as many variables as possible. The residuals conform to the normal distribution. The final models are therefore effective and stable.

Table 2 shows the resultant MLR models. Model 1 (with occlusivity) already explains almost 44.5% of the variation in the measured emotion while model 2 (with drift magnitude) explains almost 47.6% of the variation. In both models, emotions are regressed on eight visual exposure variables while also statically accounting for the effect of change direction. The results indicate the significance of street view and isovist in explaining the spatial variation of emotion. Both models have eight common predictors, three of which influence emotion in a positive way, namely Tree View Factor, visual volume, and change direction. The remaining 5 predictors are all related to negative emotion. Occlusivity has the largest standardized coefficient (−0.433) while the change direction has the lowest one (0.107) in model 1. In model 2, drift-magnitude has the largest standardized coefficient (−0.554) while the isovist area has the lowest one (−0.075).

3.3. Random forest model results and evaluation

To minimize the prediction time, it is necessary to create a predictive model using as few predictors as possible. Therefore, the variable selection was performed based on the importance ranking of all the predictors using an RF model with 50 trees. According to the variable importance (Fig. 11), the Tree View Factor is the most important

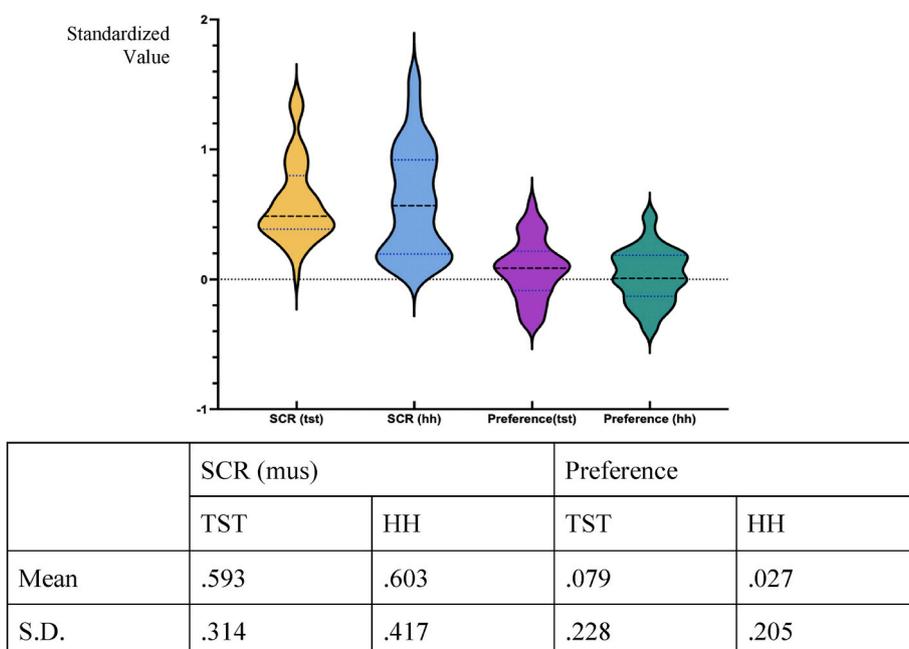


Fig. 7. Violin Plots showing the distribution of gathered responses.

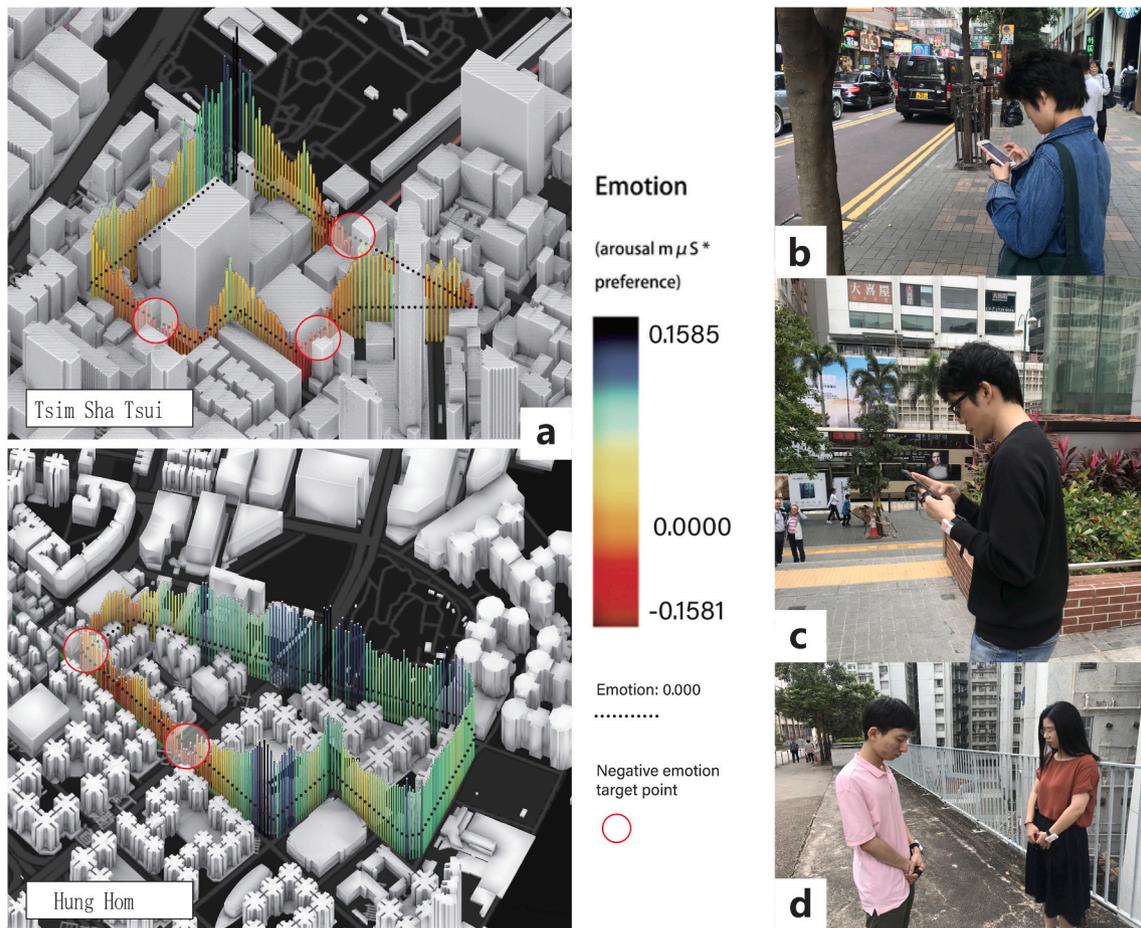


Fig. 8. (a) Visualization maps of the final emotion results for the two experimental routes; (b,c) the participants were finishing the tasks; (d) the participants were relaxing, and the SC baselines were recorded during this period.

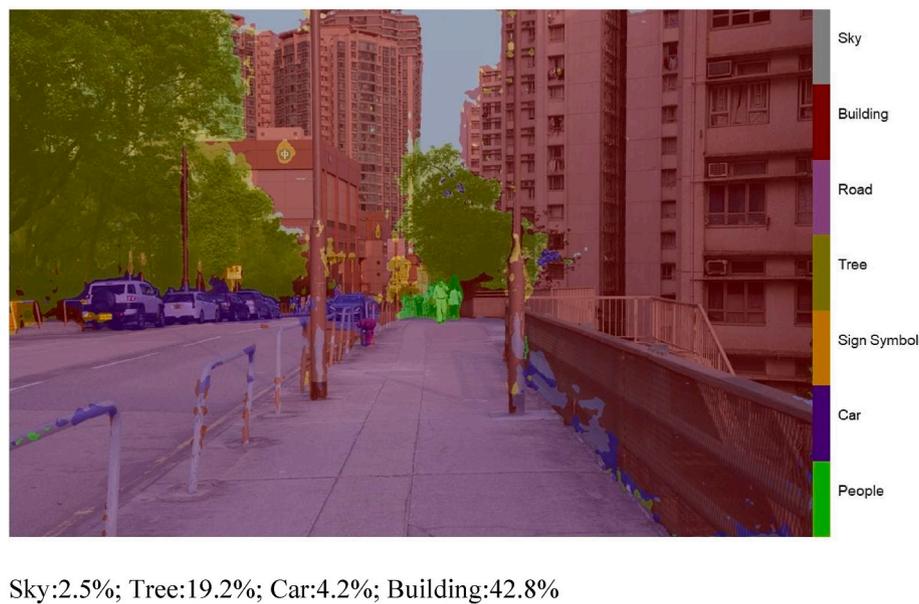


Fig. 9. Overlay of a classified image and the original street view image in HH Sky:2.5%; Tree:19.2%; Car:4.2%; Building:42.8%.

predictor variable, followed by drift angle, and sky proportion. It is identified that the nine variables with the highest importance, i.e., in order, Tree View Factor, Drift Angle, Sky Proportion, Min Radial, Area,

Ground Proportion, Object Proportion, Convex Deficiency, and BVF are sufficient to predict the pedestrian emotion since the reduced model containing only the nine variables has an R^2 value of 0.776, which

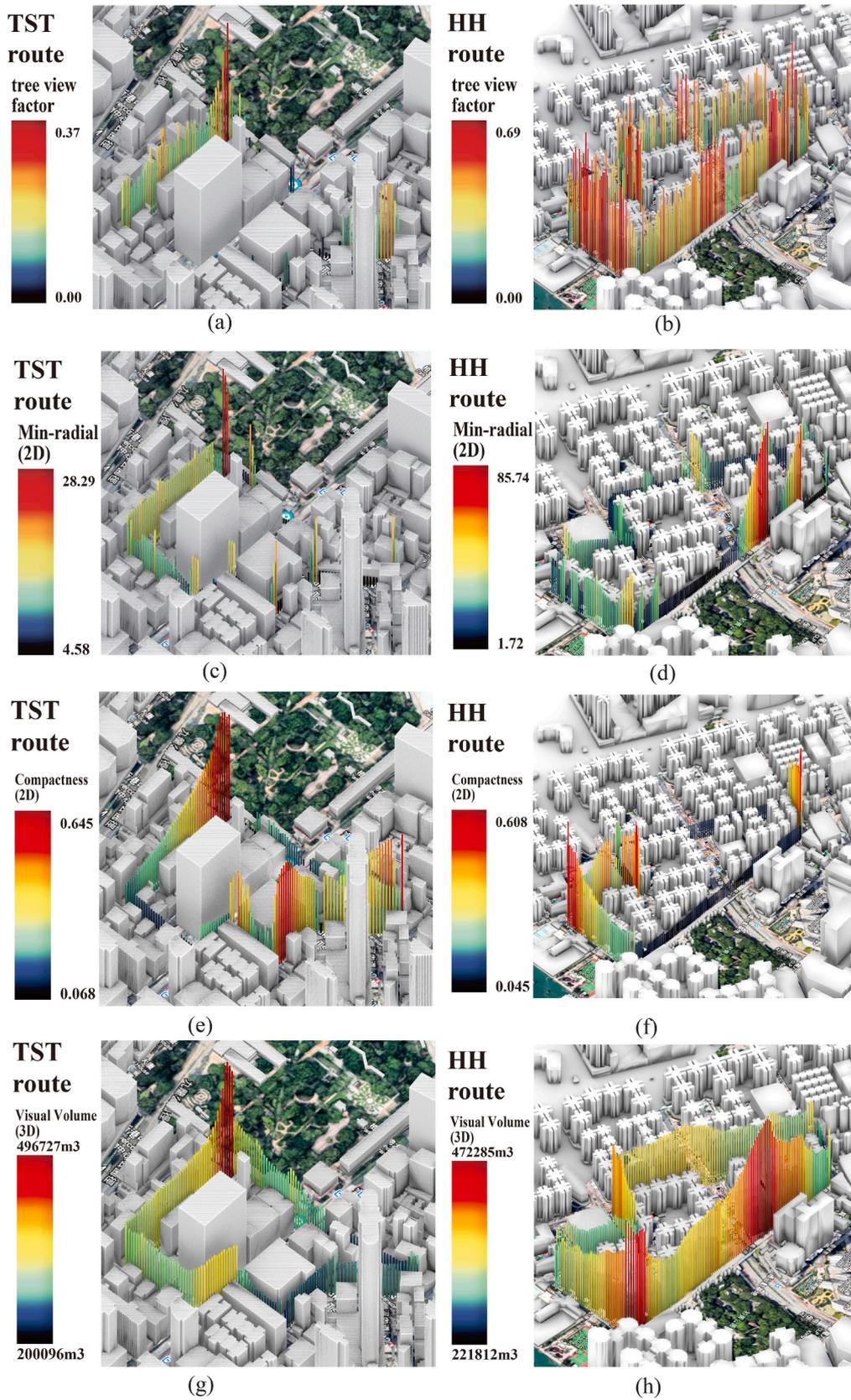


Fig. 10. Visualized maps of selected predictors, (a) Tree View Factor in TST route; (b) Tree View Factor in HH route; (c) 2D_isovist min-radial in TST route; (d) 2D_isovist min-radial in HH route; (e) 2D_isovist compactness in TST route; (f) 2D_isovist compactness in HH route; (g) 3D_isovist visual volume in TST route; (h) 3D_isovist visual volume in HH route.

Table 2
The performance and structure of the resultant MLR models of emotion states.

Model 1 ($R^2 = 0.445$)			Model 2 ($R^2 = 0.476$)		
Predictor variables	standardized coefficients	VIF	Predictor variables	standardized coefficients	VIF
Tree View Factor	.145***	1.569	Tree View Factor	.196***	1.473
Object proportion (3D)	-.248***	2.276	Sign Symbol View Factor	-.219 ***	1.296
Occlusivity (2D)	-.433 ***	4.997	Object proportion (3D)	-.228 ***	2.109
Visual volume (3D)	.419 ***	4.047	Drift magnitude (2D)	.554 ***	3.297
Min-radial (2D)	-.230 ***	1.619	Visual volume (3D)	.430 ***	3.912
SSVF	-.231 ***	1.295	Min-radial (2D)	-.292 ***	1.588
Compactness (2D)	-.209 ***	3.482	Compactness (2D)	-.226 ***	2.869
change direction	.107 ***	1.131	change direction	.128 ***	1.059
Isovist area (2D)	-.124*	2.834	Isovist area (2D)	-.075*	4.417

Note: Significance * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. The descriptions of the variables can refer to Table 1.

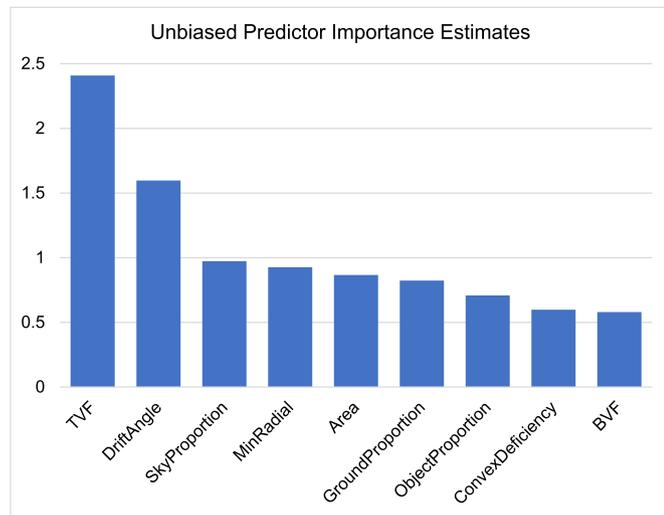


Fig. 11. Unbiased predictor importance estimates.

already exceeds the accuracy of the full model containing all the 29 potential variables ($R^2 = 0.767$). Moreover, the nine selected variables were reused to conduct the experimentation to determine the parameter values of the RF model. The tuning results reveal that in the optimal model, the minimum terminal node is 1, the number of predictors to sample is 6, and the number of trees is 400.

The final predictive model includes 9 predictors and 400 trees. It has an R^2 value of 0.790, a Root Mean Square Error (RMSE) of 0.022, and a Mean Absolute Error (MAE) of 0.015. The final model can explain 79.0% of the spatial variance of pedestrian emotion in Hong Kong. The model has demonstrated a stronger ability to predict pedestrian emotion in Hong Kong compared to the MLR model in section 3.2. Besides, the results show that the final RF model has achieved good precision and stability. Therefore, the established model with the visual exposure features identified with predictive potential can be applied to predict pedestrian emotion for new subjects with a satisfactory accuracy of over 79%.

4. Discussions

4.1. Model interpretation

The understanding of the effects of the visible environment on the emotional experiences of an individual during his/her high-density urban walking is essential for creating a psychologically friendly built environment. In this study, a multiple linear regression model is used to determine the relationship between different visual exposures and emotions. A more complicated model that can consider the non-linear relationship – random forest modeling – is also for predicting

pedestrian emotion. These two models have advantages in distinct aspects. The MLR has stricter requirements for the raw data that all variables in the final model should not be collinear, while the RF model does not require specific input data types. MLR can reveal whether the independent variable has a negative or positive effect on the dependent variable, while RF can only show the importance of each variable. Therefore, the results from MLR were used to obtain a basic understanding of the linear relationship between visual exposure and emotion, and the RF was then used to build the predictive model.

According to the MLR, visual exposure from 2D_isovist, 3D_isovist, and street view influences emotion simultaneously, and thus, none of them have been completely excluded in the final model. The findings indicate that the method of including these three aspects of visual exposure is effective in explaining the variations of emotion. As a result of including all visual exposure elements and distinguishing the positive and negative emotions, the R^2 values (0.445 and 0.476) in the developed MLR models are higher than in previous studies [29]. Therefore, it is suggested that isovists and street view factors should be considered at the same time for further studies.

Moreover, all 2D_isovist variables, including occlusivity, min-radial, compactness, and isovist area, in the resultant MLR models have a negative influence on emotion. Visual volume, as a comprehensive variable from 3D_isovist calculation, remains in both final MLR models and has a positive influence on emotion with the largest standardized coefficient among all the variables. Objective proportion and min-radial are the other two variables from 3D_isovist which show a moderately negative relationship with emotion.

The street view factors can be divided into two types: stationary (tree, sign symbol, sky, building, road) and dynamic (car and people). Two stationary variables (Tree View Factor and Sign Symbol View Factor) remained in the resultant models, providing more detailed information on the real environment that cannot be reflected by isovist calculations. This indicates the isovist and street view factors are complementary and compensate for the shortcomings of each other. For instance, the isovist supplements the geometric characteristics of a visual area, whereas the street view factor provides extra data that are difficult to model in the isovist calculation. Car View Factor and People View Factor have been excluded because the raw street view photos were taken after the experiments all at once, and can hardly reflect the real-time dynamic state on the street when the participants were walking.

The RF model has a better prediction ability (79.0%) than the MLR model ($R^2 = 0.476$). Also, it is slightly higher than the predictive models of Bielik et al. [29] ($R^2 = 0.67$, $R^2 = 0.79$ for their two tested routes, respectively). The nine most important variables consist of visual exposure factors from both isovist and street view. The Tree View Factor is the most important variable to predict emotion, followed by two variables (drift angle and min-radial) from 2D_isovist. Besides, Building View Factor also ranks the top among all the variables. The predictive model has the highest prediction ability when the emotion value is in the range of 0–0.05 and becomes less predictive when the observation value

extends to the two extremes (Fig. 12).

There are some common variables of the final resultant models of the MLR and RF. The TVF, min-radial, isovist area, and object proportion are the overlapping variables that exist in both the final MLR models and the RF model. The RF also demonstrates the non-linear relationship between emotion and drift angle, sky proportion, ground proportion, BVF, convex deficiency. The effects of these variables require further investigation to quantify the non-linear correlations.

4.2. Comparison with relevant studies

The current study was conducted in a real outdoor environment. When compared to those that were strictly controlled in laboratories, the current results are more representative of the real conditions in terms of the environment and how the people move. When compared to the previous field studies, such as the GEMA [31–33], which often have longer experiment periods (always from days to weeks), the current results are likely more accurate because the participants were guided by an instructor during the whole experiment process to ensure that they walked the pre-defined route, and the visual exposure data were consistent with the real condition.

There are several common grounds and different results when comparing the current MLR results with previous studies conducted in different countries. First, in model 1, occlusivity is negatively correlated with emotion in all the studies that mentioned isovist [39–41]. Second, occlusivity was excluded in model 2 because of collinearity while the drift magnitude takes a heavy share in a positive influence on emotion. This result is in line with a previous study in Hong Kong [39], which indicated that open space with a visual target set at a distance is the dominant factor in creating positive emotions. Third, compactness is another variable significant in all the studies. However, it has different associations with emotions from the previous studies. Results from Refs. [39–41] indicated that the simpler the visual area, the more likely it led to positive emotions. Knöll et al. [59] contradicted that visual complexity would contribute to less stress and more positive states. The current study gives the same results as the latter. It is hard to draw a definite conclusion on the relationship between visual complexity and emotion because compactness could have opposite results even for studies conducted in similar environments and cultural contexts.

Therefore, it is proposed that the relationship between visual simple degree and emotion may be curvilinear rather than linear. Moreover, the visual perception of complexity is not only determined by the geometric characteristics of the visual area but also relevant to the specific context. Lastly, object proportion (highly related to the BVF) and TVF show a significant correlation with emotion, which is in line with [30].

4.3. Implications for city design

The urban environment, particularly in the high-density city, is found to be associated with a series of physical and mental diseases [57]. Data-driven and evidence-based approaches are indispensable to healthy city development intending to enhance the quality of life and reduce the adverse effects of urban living [57]. Therefore, the methodological framework, statistical model, and the data-driven approach provided in this study demonstrate new opportunities to evaluate and implement a healthy city for planners and decision-makers. Moreover, the variable selection can inform researchers and urban designers on the environmental factors critical for mental health. The current study recognized the importance of visual exposure on emotion and recommends the consideration of street-view factors and isovists in creating healthy cities.

In this study, *Visual Volume (3D)*, also expressed as Spatial Openness Index [56], shows the positive influence on triggering a positive emotion. Therefore, the designers could control the visual volume value from the pedestrians' visual height of the neighborhoods at the district and building levels. Efficient methods to increase visual volume value include controlling the total building volume rate at the district level, designing the setback spaces at the building podium, adopting different building forms and arrangements, and decreasing the obstacle area by setting up holes at proper locations at the building level.

In addition, *Drift-magnitude*, which refers to the distance between the testing point and the mass center of the isovist polygon, has the most positive relationship with emotion value. Therefore, setting a visual target at a proper distance can be helpful in increasing the emotion value. Also relating to this strategy, *occlusivity* shows the most negative influence on emotion. It offers implications for the design of some important urban nodes that it is better to guide the visual sight to the target with boundary.

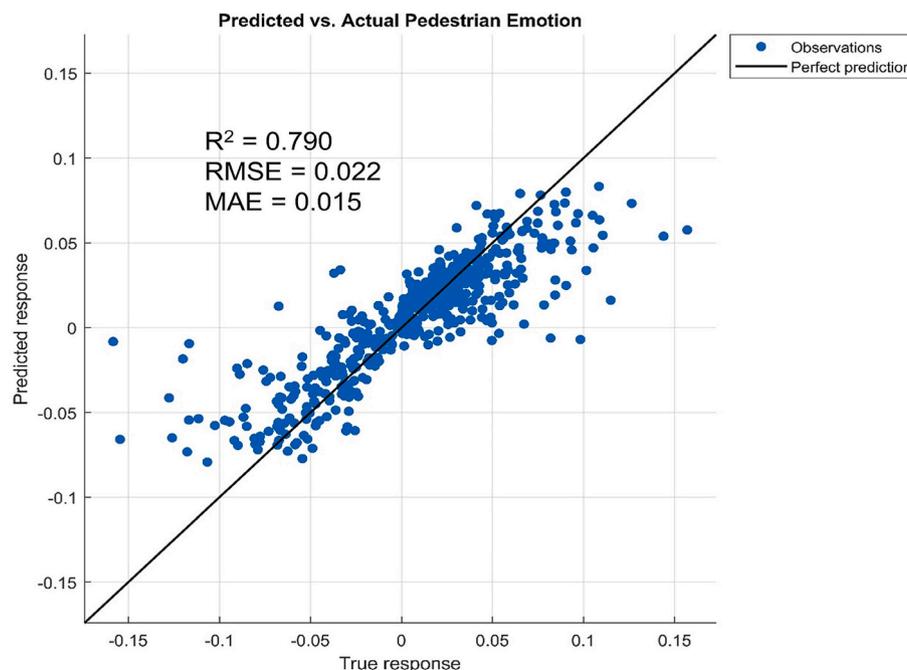


Fig. 12. Predicted and the actual pedestrian emotion from the final predictive model.

Moreover, the tree view factor and sign symbol view factor can significantly influence emotion. Previous studies [60–65] have shown that trees or greening can increase the restorative capability and psychological friendliness of the landscape. Setting up street trees at proper locations and decreasing the number of billboards will contribute to triggering better emotional states.

Therefore, the effects of the visual exposure variables explored in this study could be referred to in urban design practice. With such information, urban planners and policymakers can perform better environmental control and intervention to optimize the spatial planning scheme of each factor and design a healthy city from a psychologically friendly environment perspective.

4.4. Limitations and future studies

The current study has several limitations, mostly stemming from the technical complexity and precision of the data collection over space and time. First, the time of the street view photo is different from the experiment. Therefore, the exposure to moving objects, such as cars and people during the experiment may not be truthfully reflected by the street view images. This limitation has already been reflected in the results that no dynamic factors from street view factors were included within the final models. Although these dynamic factors (cars, crowding, etc.) are important elements in an urban outdoor space, they are not related directly to the visual urban form and are difficult to be controlled in the urban design of stationary elements, i. e. the shapes and layout of buildings and streets. Therefore, this limitation does not jeopardize the practical implications of this study. Second, the building model for the isovist calculation is primitive, containing only the basic information of the massing, but not other information from the calculation model (such as billboards, semi-outdoor spaces, and main holes in the building façades). The missing information has been supplemented by real street view images in this study. Finally, the participants in the experiments were all young, healthy adults, which would lead to selection bias and model deviation. The sample bias has been minimized by standardizing the SCR.

Several future research plans can be listed based on the current study. First, different models can be developed based on the current data. Factors from the environment and individuals, such as spatial variations and the fixed effects from different participants can be considered using geographically weighted models or panel data. Time-series analysis can also be conducted to investigate the dynamic of emotion over time. Second, we plan to further explore and quantify the effects of the visual environment for urban design implications, for example, the exact interval of each visual target, the proper amount of street trees, the effects of different building forms on emotion, etc. in underway. Finally, more advanced ways to acquire valence, such as the EEG, can be adopted to improve the accuracy of self-report data.

5. Conclusion

To facilitate informed control decisions of urban design, this study analyzed the impact of visual exposure on pedestrian emotion using MLR and established a predictive model using RF. The current approach is novel in four aspects. First, it is the first to incorporate and quantify multiple levels of environmental exposure from isovist (2D and 3D) and street view factors for a complete visual environment. Besides, it attempts to quantify two-dimensional emotion by combining subjective self-report data and ambulatory arousal assessment simultaneously. In addition, this study more adequately models the emotion from environmental factors with machine learning techniques. Furthermore, it compensates for the lack of reference data in high-density cities.

The following conclusions can be drawn from the present study: 1. both street view and isovist show a significant correlation with pedestrian emotions; 2. the MLR models have an R^2 value of around 0.5 and suggest that in a high-density environment, exposure to a larger tree

view factor, visual volume, and drift magnitude can cause positive emotions. Conversely, areas with the view of sign symbols and object proportion, min-radial, and occlusivity incite negative emotions; 3. the RF model demonstrates a strong ability to predict pedestrian emotion and can explain 79.0% of the spatial variance of the pedestrian emotion using only nine visual exposure variables; 4. TVF, drift angle, and min-radial are the dominant influential variables in the RF model.

This study demonstrates the potential and effectiveness of monitoring the emotions of pedestrians with wearable sensors in high-density built environments. The regression models built based on this approach can be adopted to better understand human emotions over space and time under different visual exposure. Such provides urban planners and designers with accurate, concrete, quantifiable, and actionable insights and evidence that can be used to make decisions and investments that aim at transforming our environment into not only a more beautiful place to be visited and traveled through but also a functional space that contributes to emotional and physical well-being.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The study is supported by a PGS studentship from The Chinese University of Hong Kong. The research gains ethics approval from The Chinese University of Hong Kong (Certificate No. S24601960). The authors thank Miss Kwok Yu Ting and Mr. Cyrus Wong for proofreading.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.buildenv.2021.108273>.

References

- [1] WHO, Comprehensive Mental Health Action Plan 2013–2020, 2013.
- [2] F. Lederbogen, P. Kirsch, L. Haddad, F. Streit, H. Tost, P. Schuch, S. Wust, J. C. Pruessner, M. Rietschel, M. Deuschle, A. Meyer-Lindenberg, City living and urban upbringing affect neural social stress processing in humans, *Sci. Rep.* 474 (7352) (2011) 498–501.
- [3] H. Tost, City living marks the brain, *Nat. Newslett.* 474 (2011).
- [4] C.E. Izard, Human Emotions, Plenum Press, New York, 1977.
- [5] A. Mehrabian, J.A. Russell, An Approach to Environmental Psychology, the MIT Press, 1974.
- [6] P. Ekman, Basic emotions, in: Handbook of Cognition and Emotion, Wiley, UK, 1999.
- [7] M. Bastiaansen, X.D. Lub, O. Mitas, T.H. Jung, M.P. Ascensão, D.-I. Han, T. Moilanen, B. Smit, W. Strijbosch, Emotions as core building blocks of an experience, *Int. J. Contemp. Hospit. Manag.* 31 (2) (2019) 651–668.
- [8] J. Posner, J.A. Russell, B.S. Peterson, The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology, *Dev. Psychopathol.* 17 (3) (2005) 715–734.
- [9] I.B. Mauss, M.D. Robinson, Measures of emotion: a review, *Cognit. Emot.* 23 (2) (2009) 209–237.
- [10] J. Wirtz, J.E. Bateson, Consumer satisfaction with services: integrating the environment perspective in services marketing into the traditional disconfirmation paradigm, *J. Bus. Res.* 44 (1) (1999) 55–66.
- [11] J.A. Russell, A circumplex model of affect, *J. Pers. Soc. Psychol.* 39 (6) (1980) 1161.
- [12] K. Kirillova, X. Fu, X. Lehto, L. Cai, What makes a destination beautiful? Dimensions of tourist aesthetic judgment, *Tourism Manag.* 42 (2014) 282–293.
- [13] J.J. Brakus, B.H. Schmitt, L. Zarantonello, Brand experience: what is it? How is it measured? Does it affect loyalty? *J. Market.* 73 (3) (2009) 52–68.
- [14] J.W. Kalat, Biological Psychology, Cengage Learning, 2015.
- [15] L.S. Davis, M.L. Benedikt, Computational models of space: isovists and isovist fields, *Comput. Graph. Image Process.* 11 (1) (1979) 49–72.
- [16] X. Li, C. Ratti, I. Seiferling, Quantifying the shade provision of street trees in urban landscape: a case study in Boston, USA, using Google Street View, *Landsc. Urban Plann.* 169 (2018) 81–91.

- [17] F.-Y. Gong, Z.-C. Zeng, F. Zhang, X. Li, E. Ng, L.K. Norford, Mapping sky, tree, and building view factors of street canyons in a high-density urban environment, *Build. Environ.* 134 (2018) 155–167.
- [18] A.E. Stamps, Advances in Visual Diversity and Entropy, *Environ. Plann. B: Plann. Design* 30 (3) (2003) 449–463.
- [19] E. Stamps, Isovists, enclosure, and permeability theory, *Environ. Plann. Plann. Des.* 32 (2005).
- [20] A.E. Stamps 3rd, Some findings on prospect and refuge theory: II, *Percept. Mot. Skills* 107 (1) (2008) 141–158.
- [21] D.E. Berlyne, Attention, perception and behavior theory, *Psychol. Rev.* 58 (2) (1951) 137.
- [22] P.J. Lindal, T. Hartig, Effects of urban street vegetation on judgments of restoration likelihood, *Urban For. Urban Green.* 14 (2) (2015) 200–209.
- [23] P.J. Lindal, T. Hartig, Architectural variation, building height, and the restorative quality of urban residential streetscapes, *J. Environ. Psychol.* 33 (2013) 26–36.
- [24] R.C. Dalton, The secret is to follow your nose: route path selection and angularity, *Environ. Behav.* 35 (1) (2003) 107–131.
- [25] J.M. Wiener, S. Hölscher C Fau - Büchner, L. Büchner S Fau - Konieczny, L. Konieczny, Gaze Behaviour during Space Perception and Spatial Decision Making, (1430-2772 (Electronic)).
- [26] G. Franz, J.M. Wiener, From space syntax to space semantics: a behaviorally and perceptually oriented methodology for the efficient description of the geometry and topology of environments, *Environ. Plann. Plann. Des.* 35 (4) (2008) 574–592.
- [27] J. Yin, S. Zhu, P. MacNaughton, J.G. Allen, J.D. Spengler, Physiological and cognitive performance of exposure to biophilic indoor environment, *Build. Environ.* 132 (2018) 255–262.
- [28] Q. Huang, M. Yang, H.-a. Jane, S. Li, N. Bauer, Trees, grass, or concrete? The effects of different types of environments on stress reduction, *Landsc. Urban Plann.* 193 (2020).
- [29] M. Bielik, S. Schneider, S. Kuliga, D. Griego, V. Ojha, R. König, G. Schmitt, D. Donath, Examining trade-offs between social, psychological, and energy potential of urban form, *ISPRS Int. J. Geo-Inf.* 8 (2) (2019).
- [30] G.C. Millar, O. Mitas, W. Boode, L. Hoeke, J. de Kruijf, A. Petrasova, H. Mitasova, Space-time analytics of human physiology for urban planning, *Comput. Environ. Urban Syst.* 85 (2021).
- [31] J. Mennis, M. Mason, A. Ambrus, Urban greenspace is associated with reduced psychological stress among adolescents: a geographic ecological momentary assessment (GEMA) analysis of activity space, *Landsc. Urban Plann.* 174 (2018) 1–9.
- [32] T.R. Kirchner, S. Shiffman, Spatio-temporal determinants of mental health and well-being: advances in geographically-explicit ecological momentary assessment (GEMA), *Soc. Psychiatr. Epidemiol.* 51 (9) (2016) 1211–1223.
- [33] L. Kou, Y. Tao, M.-P. Kwan, Y. Chai, Understanding the relationships among individual-based momentary measured noise, perceived noise, and psychological stress: a geographic ecological momentary assessment (GEMA) approach, *Health Place* 64 (2020) 102285.
- [34] C. Nold, *EmotionalCartography: Technologies of the Self*, 2009.
- [35] P. Zeile, S. Höffken, G. Papastefanou, Mapping people – the measurement of physiological data in city areas and the potential benefit for urban planning, *REALCORP 2009* (2009).
- [36] G. Sagl, B. Resch, T. Blaschke, Contextual Sensing: Integrating Contextual Information with Human and Technical Geo-Sensor Information for Smart Cities, *Sensors*, Basel, Switzerland, 2015, pp. 17013–17035.
- [37] A. Birenboim, K.H. Reinau, N. Shoval, H. Harder, High-resolution measurement and analysis of visitor experiences in time and space: the case of Aalborg zoo in Denmark, *Prof. Geogr.* 67 (4) (2015) 620–629.
- [38] Linda Dörrzapf, Zeile Peter, Günther Sagl, Sudmanns Martin, Anja Summa, Bernd Resch, Urban Emotions - an Interdisciplinary Interface between Geoinformatics and Urban Planning, *Planning Support Systems and Smart Cities*, 2015.
- [39] L. Xiang, G. Papastefanou, E. Ng, Isovist Indicators as a Means to Relieve Pedestrian Psycho-Physiological Stress in Hong Kong, *Environment and Planning B: Urban Analytics and City Science*, 2020.
- [40] X. Li, I. Hijazi, R. Koenig, Z. Lv, C. Zhong, G. Schmitt, Assessing essential qualities of urban space with emotional and visual data based on GIS technique, *ISPRS Int. J. Geo-Inf.* 5 (11) (2016) 218.
- [41] I.H. Hijazi, R. Koenig, S. Schneider, X. Li, M. Bielik, G.N.J. Schmit, D. Donath, Geostatistical analysis for the study of relationships between the emotional responses of urban walkers to urban spaces, *Int. J. E Plann. Res.* 5 (1) (2016) 1–19.
- [42] Y.S. Noam Shoval, Maya Tamir, Real-time measurement of tourists' objective and subjective emotions in time and space, *J. Trav. Res.* 57 (1) (2018) 3–16.
- [43] C. Ellard, C. Montgomery, TESTING, TESTING! A Psychological Study on City Spaces and How They Affect Our Bodies and Minds, 2013.
- [44] M. Schmidkunz, O. Schroth, P. Zeile, U. Kias, Road Safety from Cyclist's Perspective, *REAL CORP 2019– IS THIS THE REAL WORLD? Perfect Smart Cities vs, Real Emotional Cities*, 2019, pp. 597–604.
- [45] The World Bank, Population Density (People Per Sq. Km of Land Area) - Hong Kong SAR, 2018. China, https://data.worldbank.org/indicator/EN.POP.DNST?location_s=HK.
- [46] Working Group of the Preparatory Committee for Mental Health Month 2020, Territory-wide Mental Health Index Survey 2020, 2020.
- [47] S.A. Chatterjee, J.J. Daly, E.C. Porges, E.J. Fox, D.K. Rose, T.E. McGuirk, D. M. Otzel, K.A. Butera, D.J. Clark, Mobility function and recovery after stroke: preliminary insights from sympathetic nervous system Activity, *J. Neurol. Phys. Ther.* 42 (4) (2018) 224–232.
- [48] E. Gatti, E. Calzolari, E. Maggioni, M. Obrist, Emotional ratings and skin conductance response to visual, auditory and haptic stimuli, *Sci. Data* 5 (2018) 180120.
- [49] D.G.W. Jason, J. Braithwaite, Robert Jones, Mickey Rowe, A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments (revised version: 2.0), in: S.o.P. Behavioural Brain Sciences Centre, University of Birmingham, UK, 2015.
- [50] M.M. Bradley, M.K. Greenwald, M.C. Petry, P.J. Lang, Remembering pictures: pleasure and arousal in memory, *J. Exp. Psychol. Learn. Mem. Cogn.* 18 (2) (1992) 379.
- [51] Z. Chen, S. Schulz, M. Qiu, W. Yang, X. He, Z. Wang, L. Yang, Assessing affective experience of in-situ environmental walk via wearable biosensors for evidence-based design, *Cognit. Syst. Res.* 52 (2018) 970–977.
- [52] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, J. Garcia-Rodriguez, A Review on Deep Learning Techniques Applied to Semantic Segmentation, 2017 arXiv preprint arXiv:1704.06857.
- [53] G.J. Brostow, J. Fauqueur, R. Cipolla, Semantic object classes in video: a high-definition ground truth database, *Pattern Recogn. Lett.* 30 (2) (2009) 88–97.
- [54] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with Atrous Separable Convolution for Semantic Image Segmentation, 2018, pp. 801–818.
- [55] C. Ren, M. Cai, X. Li, Y. Shi, L. See, Developing a rapid method for 3-dimensional urban morphology extraction using open-source data, *Sustain. Cities Soc.* 53 (2020) 101962.
- [56] D. Fisher-Gewirtzman, I.A. Wagner, Spatial openness as a practical metric for evaluating built-up environments, *Environ. Plann. Plann. Des.* 30 (1) (2003) 37–49.
- [57] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32.
- [58] R. Genuer, J.-M. Poggi, C. Tuleau-Malot, Variable selection using random forests, *Pattern Recogn. Lett.* 31 (14) (2010) 2225–2236.
- [59] M. Knöll, K. Neuheuser, T. Cleff, A. Rudolph-Cleff, A tool to predict perceived urban stress in open public spaces, *Environ. Plann. B: Urban Anal. City Sci.* 45 (4) (2017) 797–813.
- [60] C. Ward Thompson, J. Roe, P. Aspinall, R. Mitchell, A. Clow, D. Miller, More green space is linked to less stress in deprived communities: evidence from salivary cortisol patterns, *Landsc. Urban Plann.* 105 (3) (2012) 221–229.
- [61] G.N. Bratman, J.P. Hamilton, K.S. Hahn, G.C. Daily, J.J. Gross, Nature experience reduces rumination and subgenual prefrontal cortex activation, *Proc. Natl. Acad. Sci. U. S. A.* 112 (28) (2015) 8567–8572.
- [62] Hartig Terry, Marlis Mang, Gary W. Evans, Restorative Effects of Natural Environment Experiences, *Environ. Behav.* 23 (1) (1991) 24.
- [63] T.H. Kim, G.W. Jeong, H.S. Baek, G.W. Kim, T. Sundaram, H.K. Kang, S.W. Lee, H. J. Kim, J.K. Song, Human brain activation in response to visual stimulation with rural and urban scenery pictures: a functional magnetic resonance imaging study, *Sci. Total Environ.* 408 (12) (2010) 2600–2607.
- [64] R. Kaplan, S. Kaplan, *The Experience of Nature: A Psychological Perspective*, CUP Archive, 1989.
- [65] R.S. Ulrich, R.F. Simons, B.D. Losito, E. Fiorito, M.A. Miles, M. Zelson, Stress recovery during exposure to natural and urban environments, *J. Environ. Psychol.* 11 (3) (1991) 201–230.