

Article

# Framework for Virtual Cognitive Experiment in Virtual Geographic Environments

Fan Zhang <sup>1,2</sup> , Mingyuan Hu <sup>2,3</sup>, Weitao Che <sup>2,3</sup> , Hui Lin <sup>2,3,4</sup> and Chaoyang Fang <sup>5,\*</sup>

<sup>1</sup> Institute of Remote Sensing and Geographical Information System, Peking University, Beijing 100871, China; zfan@pku.edu.cn

<sup>2</sup> Institute of Space and Earth Information Science, the Chinese University of Hong Kong, Hong Kong, China; humingyuan@cuhk.edu.hk (M.H.); cheweitao@cuhk.edu.hk (W.C.); huilin@cuhk.edu.hk (H.L.)

<sup>3</sup> Shenzhen Research Institute, the Chinese University of Hong Kong, Shenzhen 518057, China

<sup>4</sup> Department of Geography and Resource Management, the Chinese University of Hong Kong, Hong Kong, China

<sup>5</sup> Key Laboratory of Poyang Lake Wetland and Watershed Research Ministry of Education, Jiangxi Normal University, Nanchang 330022, China

\* Correspondence: fcy@jxnu.edu.cn; Tel.: +86-0791-8812-0539

Received: 31 October 2017; Accepted: 17 January 2018; Published: 22 January 2018

**Abstract:** Virtual Geographic Environment Cognition is the attempt to understand the human cognition of surface features, geographic processes, and human behaviour, as well as their relationships in the real world. From the perspective of human cognition behaviour analysis and simulation, previous work in Virtual Geographic Environments (VGEs) has focused mostly on representing and simulating the real world to create an ‘interpretive’ virtual world and improve an individual’s active cognition. In terms of reactive cognition, building a user ‘evaluative’ environment in a complex virtual experiment is a necessary yet challenging task. This paper discusses the outlook of VGEs and proposes a framework for virtual cognitive experiments. The framework not only employs immersive virtual environment technology to create a realistic virtual world but also involves a responsive mechanism to record the user’s cognitive activities during the experiment. Based on the framework, this paper presents two potential implementation methods: first, training a deep learning model with several hundred thousand street view images scored by online volunteers, with further analysis of which visual factors produce a sense of safety for the individual, and second, creating an immersive virtual environment and Electroencephalogram (EEG)-based experimental paradigm to both record and analyse the brain activity of a user and explore what type of virtual environment is more suitable and comfortable. Finally, we present some preliminary findings based on the first method.

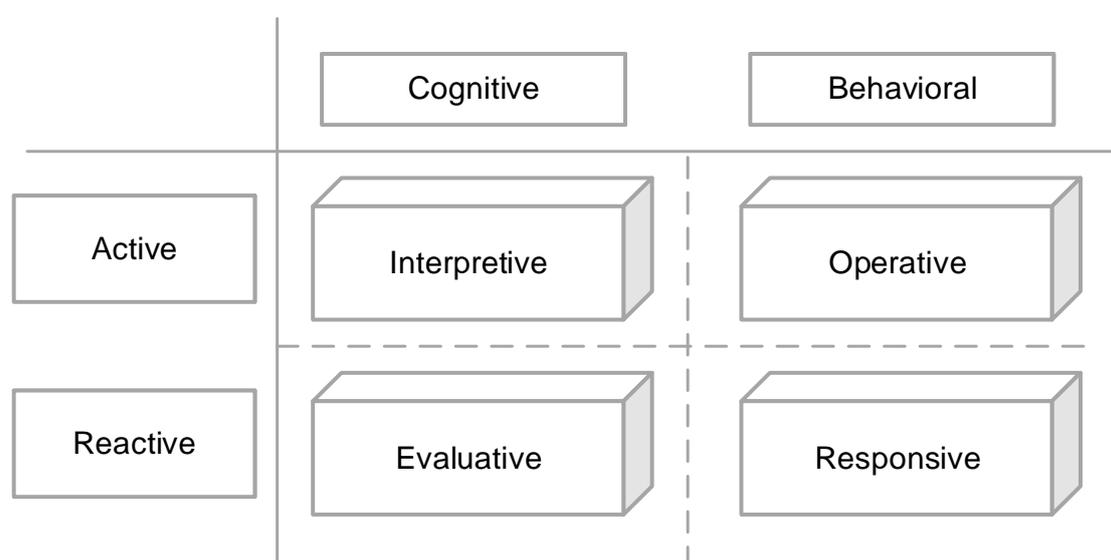
**Keywords:** virtual geographic environments; spatial cognition; brain-computer interface; street-level imagery; deep learning

## 1. Introduction

Human–environment relationships represent a traditional aspect in the study of geography [1,2]. Understanding the mechanisms underlying human–environment interactions and what factors in the urban built environment influence human perception and sentiments has long been of interest to a wide variety of research fields [3,4]. In previous studies, VGEs was initially proposed in the field of geography and geographic information, with the objective of integrating the virtual environment and the real world and ultimately exploring universal geographic processes, phenomena and laws of things [5–7]. Accordingly, virtual geographic environment cognition is the attempt to understand the human cognition

of the real-world phenomena of surface features, geographic processes, human behaviours and their relationships [8].

From the perspective of environmental psychology, the fundamental processes of human environment transaction include the following: the interpretive, evaluative, operative, and responsive modes of interacting with one's surroundings [9,10]. As Figure 1 depicts, 'interpretive' is an active-cognitive form of transaction, which involves the individual's cognitive representation of the surroundings; 'evaluative' is in a reactive and cognitive form, indicating the individual's attitudes and assessment [9]. For years, VGEs has been demonstrated as an effective tool to represent and simulate the real world [11–13]. Users can improve their cognition to geographic objects, processes and phenomena through VGEs. However, research on virtual geographic environment cognition still stops at the interpretive phases. How to collect a user's evaluative and responsive information concerning their surroundings and how to build a feedback loop from environmental controlled stimuli to the individual's assessment and reaction to the environment remain open issues.



**Figure 1.** Modes of human-environment transactions from the perspective of environmental psychology [9].

In the early stage, efforts were made using questionnaire, interviews and online surveys through case studies. These traditional experimental paradigms have been considered as costly, time consuming, and of low throughput [14,15]. Various studies conducted with these paradigms were faced with difficulty in terms of a lack of resolution, scale and throughput in understanding the human–environment interactions [3,16]. The last several years have witnessed the rapid development of sensor networks, map services and artificial intelligence, which have not only provided extensive amounts of geotagged data from around the world but also brought about approaches to handle these data [15]. The new available rich data and potential methods provide opportunities to better understand the mechanisms of human–environment reciprocal interactions in VGEs.

In addition, brain–computer interfacing and artificial intelligence have benefited from the rapid development of high-performance computing systems and the availability of large-scale annotated datasets. As a popular representative technology of artificial intelligence, deep neural networks, which were initially inspired by the neural networks of human brains, have drawn significant attention in many fields in recent years and have proved to be very effective and efficient for data feature extraction and representation [17–19]. In terms of brain-computer interfacing, devices such as EEG [20,21], Functional Magnetic Resonance Imaging (fMRI) [22,23], and eye-tracking [24,25] have

been widely used to simulate and model human brain activity and human behaviour. In particular, EEG and fMRI are used to map and record the electrical activity in the brain. Combined with state-of-the-art deep learning techniques, researchers have decoded the fMRI data for dynamic natural vision [26]. In addition, eye-tracking devices have been designed to measure eye position and eye movement, to acquire an individual's area of interest in their field of view (FOV), and to understand an individual's cognition processes [26]. These techniques prospectively advance the field of virtual cognitive experiments in VGEs.

VGEs attempts to represent the geographic subject and phenomena in the real world. Indeed, immersive 3D virtual environments have demonstrated their importance in the learning and education fields [11,27]. Conroy's work focused on employing immersive virtual environments to understand the behaviour of pedestrians and facilitated research in spatial navigation [28]. The effort made by Dias et al. aimed at detecting participants' emotional changes with virtual reality and biometric sensing techniques to inform architecture design [29,30]. Additionally, virtual environments can facilitate the exploration of human behaviour during indoor way-finding and evacuation scenarios [31,32].

In terms of building realistic three-dimensional (3D) environments, Juřík et al. compared the human ability to reckon altitude information in different settings of 3D visualization environments [33]; they also demonstrated the importance of interactions in the human cognition process in the virtual world—the flexibility of interaction can improve the extent and speed of users in acquiring knowledge in the virtual world. Research effort has also been made to understand the cognitive aspects with different levels of immersion [34]. Moreover, the choice of device type for use in interacting with the virtual environment also matters to the users [35]. The settings of the virtual environment should vary based on the given issue and case. Replicating increasingly more details of the real world should not be the aim of VGEs; the objective should be to provide a “realistic” feeling. An excessive amount of irrelevant details can possibly dilute the actual critical information and lead to cognitive overload for users [36]. Indeed, the trade-off between experimental control and mundane realism has long been discussed in the literature advanced by psychologists [37]. Additional cognitive experiments should be conducted to understand what types of elements and features in virtual environments are more sensitive to individuals, thus building an experimental-controlled and realistic experimental environment for specific research issues [36].

To advance the progress of virtual geographic environment cognition, from the “interpret” phase to the “evaluative” phase, this study proposes a new experimental framework under the concept of VGEs. For demonstration, two potential implementation methods are presented in this paper. First, by mining hundreds of thousands of street view images scored by individuals in terms of sense of safety, we explored what visual factors produce the different levels of perception for individuals. In this case, the urban visual environment is simplified and represented by the street view images to balance the experiment control and the sense of realism. Second, an EEG and immersive virtual environment-based approach is proposed to measure human perception and emotion from the perspective of physical-psychological-emotional mechanisms. The visual variables (enclosure, openness, and permeability) of the virtual environment can be controlled by adjusting the position and shape of the constructions. Accordingly, participants' psychological processes can be recorded by the EEG devices. The proposed method can be used to explore the type of virtual environment that is more suitable and comfortable.

## 2. Discussion on Future Geographic Cognitive Experiment in VGEs

In this era of big data and artificial intelligence, researchers from various fields and disciplines have been faced with the same issues: how to make good use of extensive data and information with limited data mining and knowledge discovery approaches; how to both incorporate the data from multiple sources and link the data together automatically and intelligently; and how to combine the Internet of Things, sensor networks and the real world for information acquisition and quick response to satisfy the requirements of human activities [38,39]. Indeed, by integrating the physical

geometry model, human behavioural model and geographic process model, the development of VGEs is facilitating the establishment of a geographic knowledge engineering and sharing system by incorporating geographic knowledge, brain-computer interfaces, and human social simulations, with the final goal of improving our abilities to understand, replicate natural and social phenomena and handle dig new geographic knowledge.

Embodied cognition suggests that cognition is not only a course of processing information in the brain but also the product of the reciprocal interactions among the cognition, body, and context of the subject [40]. Embodied cognition emphasizes understanding the relationship between physical experience and psychological processes. Benefiting from the rapid development of cognition science and cognitive experiment technology, embodied cognition has moved from philosophical speculation to an experimental and empirical research field [41,42].

Inspired by the theory of embodied cognition and enabled by the advancing cognition science, the development of VGEs in terms of human behavioural cognition and analysis is moving towards the combination of geographic environment, brain-computer interfacing, and physical-psychological integration, thus building up realistic and hyper-reality geo-spatial cognitive environments, facilitating research in geo-spatial cognition and further improving human understanding in the geographical space [7]. With respect to simulating real geographical environments, this framework will utilize typical features of VGEs—dynamic geographic process simulation—to present both static features and dynamic phenomena [13]. To model and simulate the mechanisms underlying human information acquisition processes from the surrounding environment, the framework should employ a multi-channel sensing mechanism [43]. Moreover, the framework should incorporate the traditional research paradigm of environmental psychology and cognitive science into the rapid development of artificial intelligence, cognitive techniques and affective computing, thereby facilitating the analysis of human perception, cognition and behaviour in virtual geographic cognitive environments [10]. Generally, future VGEs for conducting geographic cognitive experiments requires the following: (1) a special emphasize on the individual's sense of reality and immersive feeling, letting the individual interact with the virtual environment in a natural manner; (2) the introduction of geographic process simulations into the cognitive environment, where a sensor network will gather information about the surrounding conditions and the real-time changes in the real world to reconstruct the real world in the virtual environment dynamically; and (3) the simulation and analysis of human behaviour in virtual cognitive environments based on cognitive techniques and affective computing.

Based on the above general thinking and discussion for human behavioural simulation and analysis in VGEs, future research on VGEs suggests paying greater attention to the following aspects.

### *2.1. Building Virtual-Reality Geographic Subject in 3D VGEs*

A virtual geographic environment initially indicates an environment that includes the subjects, such as avatars, avatar groups, and avatar-based individuals, as well as all the objects that surround and support the existence of the subjects [5,44]. Compared with traditional geographic information systems, VGEs inherently suggests representing natural geographic objects, rather than interacting with static objects based on spatial geometry [8]. This feature will contribute to enhancing the user's sense of presence. On the other hand, it is possible to explore the user's interaction, collaboration, reactivity, mobility, etc. in a natural manner. Future VGEs is aimed at mapping the virtual world to the real world, including social relationships and information sharing between individuals and groups [11].

**Geographic behavioural subject modelling.** Geographic behavioural subject modelling is a general concept. The modelling of the subject will include all the possible forms in the intelligent space [11]. At the virtual level, the subject may include avatars, intelligent agents, and even robots; at the reality level, the subject refers to users, individuals, social groups, etc. These two levels will interact, communicate, and negotiate with each other to shape and form the virtual-reality world in the user's brain towards specific application scenarios.

**Information sharing, communication, and collaboration.** Future VGEs will put special emphasis on building communication networks through the Internet or Internet of Things. The network will enable users to share information and interoperate with each other [45]. On the one hand, the behavioural subjects (users, avatars, and geo-objects) can interact with each other and perform tasks collaboratively in the virtual world and real world simultaneously [46], which provide a potential solution to a challenging issue—online and offline among multiple ends. On the other hand, the real world can be mapped to the virtual world seamlessly through location-based information, which can not only support research on augmented reality in virtual environments but also facilitate research on augmented virtuality in real environments.

### *2.2. Multi-Dimensional Visual Representation and Multi-Modal Sensing*

VGEs provides the basic 3D spatial cognitive platform for users. It integrates numerous types of scenarios and scene details to give the users corresponding visual experiences. Moreover, VGEs is able to simulate dynamic geographic phenomena with real-time data acquired from sensor networks in the real world to recover the real environment to a large extent, thereby allowing users to “feel it in person” and “know it beyond reality” [5]. The users will perceive the environment actively and participate in decision making as an avatar in the virtual environment, whose behaviour and reaction will be observed and recorded for psychological and cognitive experiments [47,48]. More specifically, the multi-dimensional visual representation is aimed at building an environment with static scene information and dynamic geographic phenomena to present the real-world information from different levels of ontology for assisting in sensing static objects and monitoring dynamic events. For multi-modal sensing, the users are allowed to perceive the environmental information (such as temperature, vision, and sounds) in a more natural manner [49]. This is achieved by introducing numerous types of sensors and augmented reality devices. Based on multi-modal sensing mechanisms, the users will gain information and communicate with the environment seamlessly.

### *2.3. Behavioural Simulation and Analysis Based on Cognitive Psychology*

The human brain is one of the most complex organized structures [50] and is the centre of a human psychological and cognitive activities. Hence, to simulate human behaviours, preliminary studies on human brain simulation, perception modelling and formulation, and the inference and behavioural processes in individuals need to be conducted. As described above, VGEs provides a multi-modal sensing and real-time dynamic information collection environment, which supports human behaviour simulation and analysis from the perspective of realistic representations (context) and multi-channel information acquisition (individual). In addition, future VGEs will integrate cognitive psychology frameworks and build individual behaviour paradigms [51], thereby informing and guiding related behavioural cognitive experiments for dynamic processes. With controlled experiments, correlation analysis and optimization methods, future VGEs is set to build an ideal urban environment under different scenarios and scenes and towards different user groups. Furthermore, benefiting from affective computing [52] and deep learning [53] techniques, a virtual environmental-physical psychological emotion model can be designed and built to better understand and simulate the influencing and reciprocal mechanisms between real urban environments and human behaviour.

### *2.4. Framework of Virtual Cognitive Experiments*

Based on the thinking and discussion above, Figure 2 demonstrates the framework of a virtual cognitive experiment in VGEs. The framework follows the human-environment transaction modes in environmental psychology [9]: Interpretive, Evaluative, Operative, and Responsive. For each phase, the framework indicates the description, key issues and technologies.

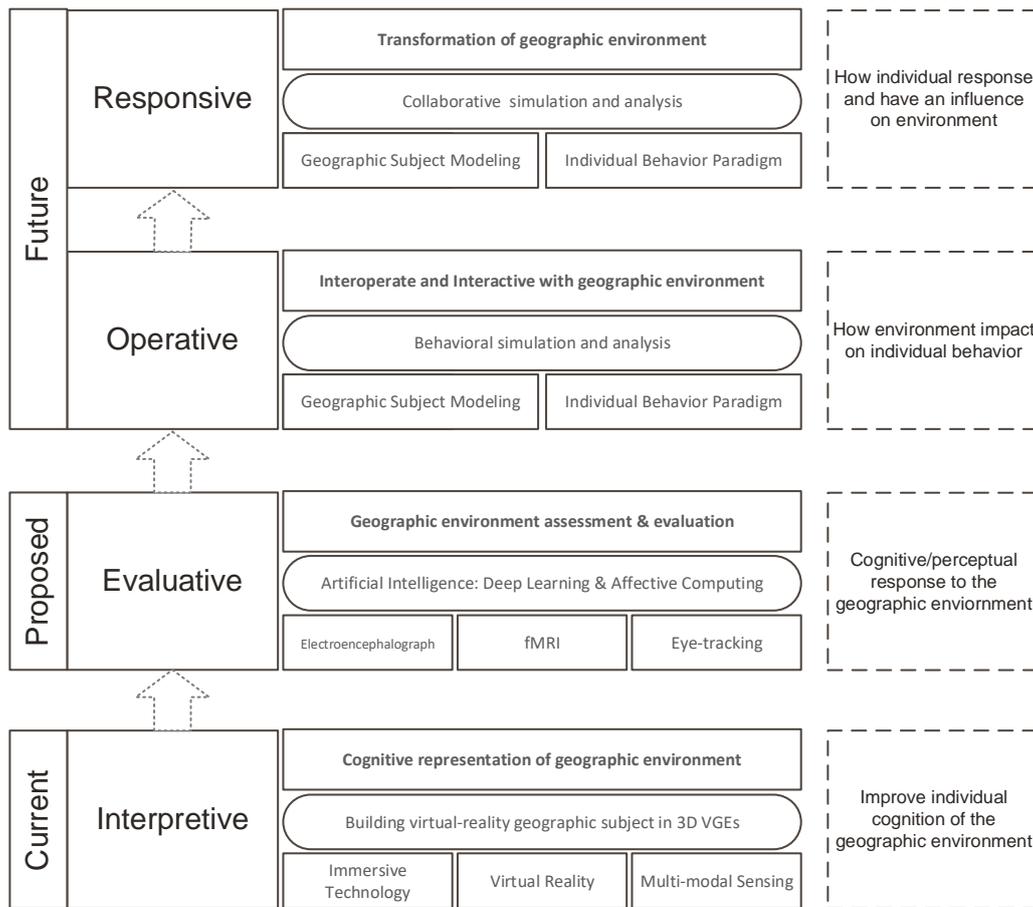


Figure 2. Framework of virtual cognitive experiments in vges.

Current studies on virtual cognitive experiments pay greater attention to improving the cognition of individuals to the virtual environment by involving immersive technologies, realistic 3D environment modelling, and multi-modal sensing to build geographic subjects in 3D VGEs [7]. As discussed above, the future development of virtual cognitive experiments should integrate individual and social behavioural models to understand how people interact, operate, and influence the geographic environment by combining theory and simulation modelling in environmental psychology and social psychology.

In this study, we focus on the evaluative phase, and we propose two potential implementation methods. The evaluative phase aims at recording an individual’s cognitive and perceptual reactions to environmental stimuli. To obtain and quantify the response, EEG, fMRI can be employed to record human brain activity; eye tracking can be used to record an individual’s focus of attention [25]. In addition, with deep-learning-based data mining methods and human evaluation data collected from large-scale online surveys, it is also possible to explore the latent human preference knowledge.

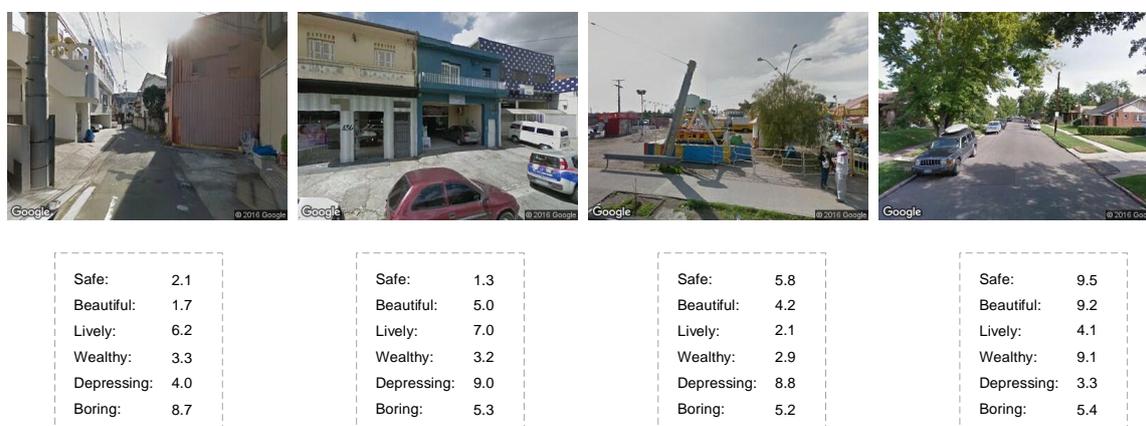
### 3. Computer-Vision-Based Cognition Experiment Using Street-Level Images

In this section, a deep-learning-based urban cognition-perception experiment is presented. The objectives of this experiment are two-fold: modelling and predicting an individual’s sense of safety in certain urban visual environments and exploring which visual elements produce the sense of safety. For the urban scenes, the experiment simply employs street-level images. Millions of street-level images and their online rating scores from volunteers have been obtained from the MIT Place Pulse dataset [54]. A Deep Convolutional Neural Network (DCNN) model [19] is trained with the dataset

to understand and evaluate new street view images from a human perspective and to analyse how different visual elements impact an individual's perception.

### 3.1. Human-Rated Street View Image Database

The MIT Place Pulse project was launched in 2013. In this project, online data collection was performed to collect the reactions to different urban images from volunteers. In addition, these datasets contained 110,988 street view images captured between 2007 and 2012, covering 56 cities from 28 countries on six continents. Figure 3 shows four image samples with their corresponding six perceptual scores: the degree of safety, depressing, boring, beautiful, lively and wealthy. These ratings have expressed the different characteristics of these images and could potentially be used to both reflect people and train the human rating prediction model. This dataset was also employed due to its high data quality: the locations for the images were dense and random, the meta-data of these images were provided, and 1,169,078 pairwise image comparisons had been collection by October of 2016 [55]. More detailed information can be accessed through the project page (<http://pulse.media.mit.edu>).

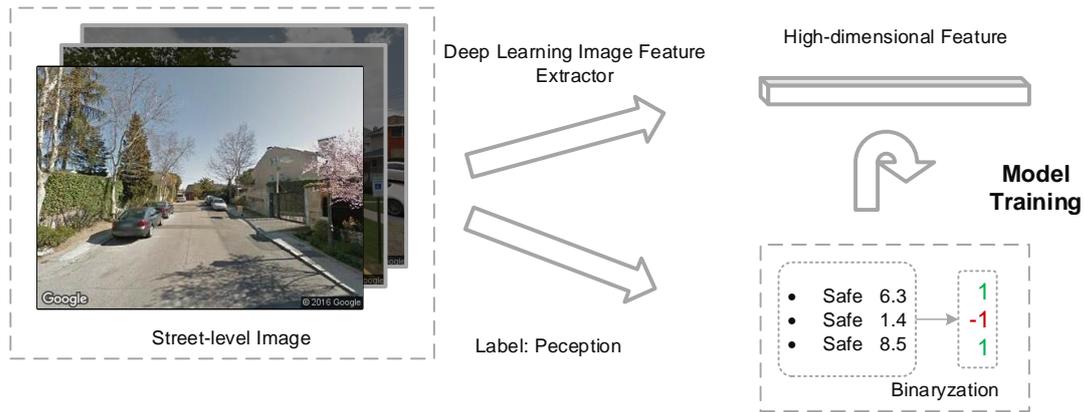


**Figure 3.** Image samples in the MIT Place Pulse dataset with their perceptual score along the six dimensions.

### 3.2. Human Perception Modelling and Prediction

In the first part of this case, we train a model to predict the score of the sense of safety from a street view image. In a previous study, a large number of image feature representation methods, such as DeCAF features [14], Dense SIFT [56], were used. In addition, Support Vector Machine, Linear Regression [14], and RankingSVM [57] have been used to model the process with the image features. Compared with previous work, the recently developed DCNN outperformed these traditional methods in combining feature extraction and task modelling [19,55]. This study also introduced a state-of-the-art DCNN architecture—Deep Residual Network (ResNet) [58]—to predict human perceptions.

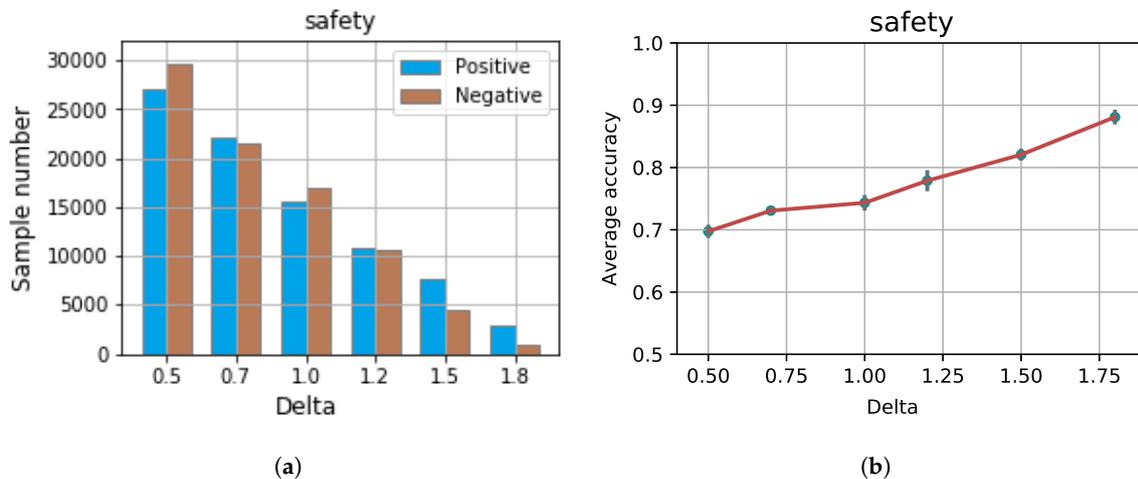
In Figure 4, we demonstrate the methodology. We formulate the image score prediction problem as a binary classification task. Using the binary classification model instead of a regression model is more generalizable for human-perception-related tasks since human perceptions are inherently unstable and uncertain, however, we can still obtain a continuous score through the probability of label predictions [59–61]. To organize the training dataset, we use a threshold to determine the selection of positive and negative samples. In the training phase, we will conduct the training task with different threshold values. The image feature is extracted by Places365-ResNet, which is a deep learning feature extractor trained on the Places2 database [62]. With the high-dimensional deep image features, we use the Radial Basis Function (RBF) kernel Support Vector Machine (SVM) [63], which is a binary classifier, to perform the classification task.



**Figure 4.** An overview: Modelling human perception to urban visual environment. First, we extract the image features using DCNN and annotate the image with a binary label. Second, an SVM classifier is trained to predict the human perception of a new region.

### 3.2.1. Experiment and Results

In the experiment, we trained the model using the pipeline described above. Figure 5 shows the training samples (a) and prediction accuracy (b) with different positive-negative thresholds. The model was validated using five-fold cross-validation. We can see that the prediction performance varies with the size of the training set.



**Figure 5.** Sample number (a) and the average accuracy (b) in the experiment. The vertical bars in the left figure show the positive and negative samples used in the training task with different threshold values. The red curves in the right figure indicate the average accuracy with different training sample sizes.

As we can see from Figure 5b, we achieved 88.2% accuracy in predicting whether a street view image looks safe. As described in the methodology section, we can obtain a continuous score by referring to the probability of each prediction. For instance, a high probability of one image that is considered to be safe indicates a high safety score of the image.

In addition, we employed the well-trained model to predict the distribution characterizing the sense of safety around a new area. In this experiment, the downtown area of Chengdu, China has been selected as the research area. We collected approximately 420,000 street view images for Chengdu

from Tencent Maps (<http://map.qq.com>). All images were predicted by the model with a safety score. Next, we aggregated all the images into their corresponding streets to obtain the safety score of each street. Figure 6 presents the results.



**Figure 6.** Spatial distribution of the sense of safety in Chengdu. Green indicates higher values of the sense of safety. Streets in red mean low safety values.

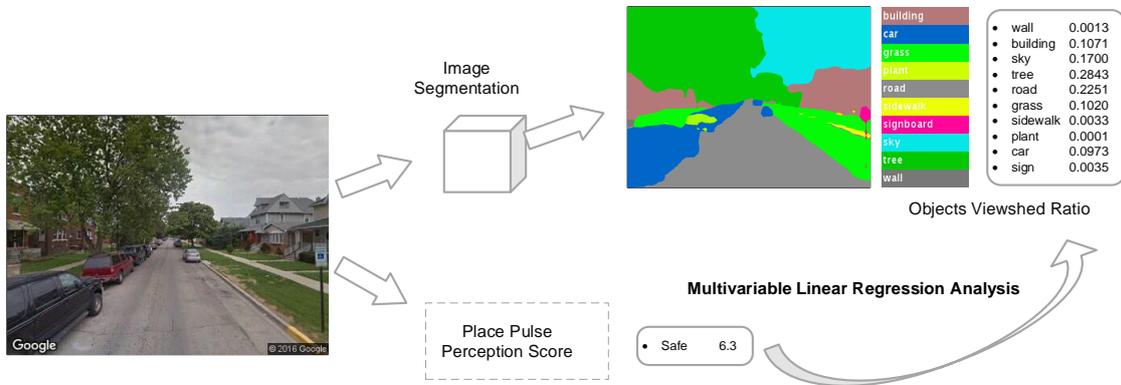
The experiment demonstrates the feasibility and effectiveness of the proposed method in modelling human perception of street view images or urban scenes. By applying the model to large-scale geotagged street view image data, we are able to plot the spatial distribution of human perception of a specific area. More in-depth studies can be conducted by integrating other social-economic data and analysing spatial patterns.

### 3.3. Exploring the Visual Features that Produce Human Sense of Safety

Second, DCNN was employed to explore how different visual features impact the sense of safety, which would influence daily life significantly. For example, people may choose to take a different route if a neighbourhood is believed to be unsafe [64]. Previous traditional studies have noted that the personalization of property, the presence of street lights, and private plantings would produce a safer feeling and that litter, graffiti and poorly maintained buildings would make a street appear much less safe. In this study, semantic scene parsing techniques will be used to facilitate a more quantitative and comprehensive analysis.

The dataset used in this experiment was also from MIT Place Pulse. Basically, we obtained the perceptual rating scores of each image from the calculation. Meanwhile, to represent the street scene elements, we employed semantic scene parsing techniques to calculate the area ratio of semantic objects in the Field of View (FOV). Semantic scene parsing is a key technique in scene understanding [65] and aims at recognizing and segmenting object instances in a natural image. Given an input image, the model is able to predict one class label for each pixel. The state-of-the-art scene parsing model, PSPNet, has achieved 79.70% pixel-wise accuracy in classifying 150 object categories [66] and has been employed in this study.

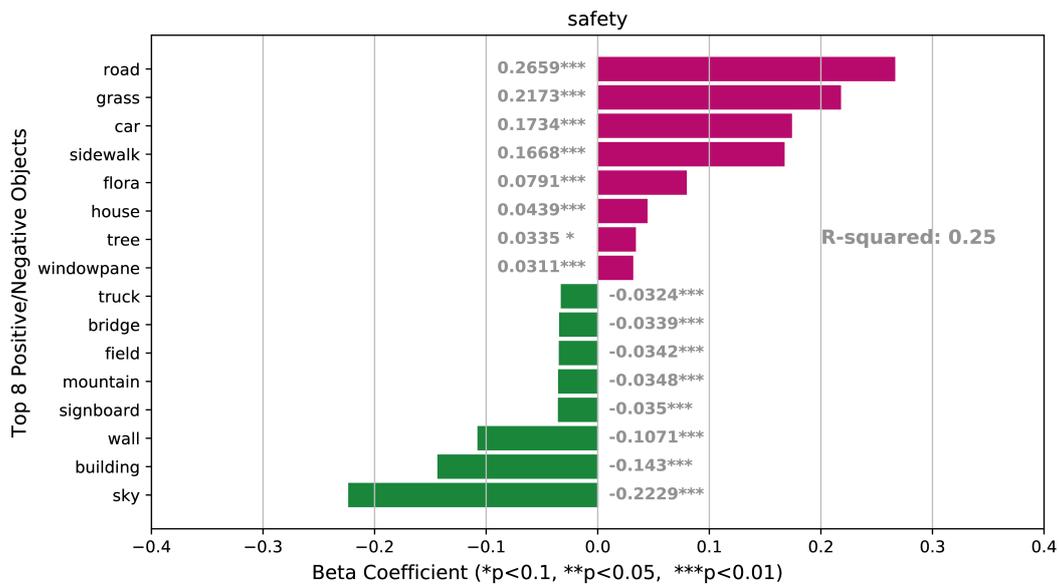
Figure 7 is an overview of the multivariate regression analysis. The safety score of each image sample was obtained from the Place Pulse dataset, and the FOV ratio of each object category in the image was calculated by counting the number of pixels in the segmentation mask.



**Figure 7.** Overview of multivariable linear regression analysis between perceptual scores and object FOV ratio. The image samples were selected from the Place Pulse dataset with the perceptual scores. The object FOV ratio was calculated from the image using the image semantic segmentation model.

A multivariate regression analysis was conducted to investigate the dependence between multiple variables. In this case, the safety score was taken as the dependent variable, and the 150 object categories were treated as the predictors. The contribution of each object to a specific perceptual attribute was compared by observing the standardized coefficient of that object in the regression analysis.

In Figure 8, we present the results, where the top eight objects that positively (red bar) or negatively (blue bar) contributed to each perceptual indicator are ranked and listed. The length of the bar indicates the value of beta—standardized coefficients—and \* indicates the significance level.



**Figure 8.** Results of multivariable regression analysis between scene elements and perception scores. For each pair, the pixel number of a particular object category and perception score along a specific dimension are given. The top eight objects that positively/negatively contributed to each of the six perception types are shown.

The positive objects could be separated into two categories sparsely. One category indicates the presence of human activities, including roads, cars, side walks, houses and windowpanes. The other type is natural elements such as grass, flora and trees. These elements can increase the sense of safety mainly because they are more vivid and creative. The negative group contains sky, mountains, fields, and buildings, walls, and bridges, which are intuitively more closed and depressed. Although street lights and traffic signs were labelled as important safety indicators, they have not been identified in this research. A major reason for this could most likely be their low frequency in the sample images. Indeed, the results are consistent with classical theories. For example, greenery was considered to provide greater quietness and peacefulness [67].

This case demonstrated the possibility of using human-rated street view images and deep learning techniques to explore human cognition and perception and the reciprocal mechanism with the surroundings. In future work, the pipeline and deep-learning-based methodology can potentially be used in realistic 3D scenarios in VGEs, thus allowing one to explore more factors (other than visual factors) that produce different human perceptions.

#### 4. EEG-Based Cognitive Experiment in 3D VGEs

Human environmental perception is a comprehensive experience when stimulated by circumstances such as an individual's specific activity, event or physical environment (as depicted in Figure 9). Actually, the process starting from context to human physical responses, human psychological responses and further to human perception and cognition is a rather complex dynamic. Understanding the mechanisms and relationships among these aspects provides the potential of exploring the dynamic emotion process as well as the key factors in spatial-temporal serial events that affect human psychological activities. Among the factors, the visual quality of the environment has been identified as an important dimension of the human environmental perception in urban design [68,69]. It may evoke strong emotional responses such as the aesthetic experience of the surroundings. More specifically in urban spaces, the visual quality is generated by spatial variables (e.g., openness) and is reflected on the psychological state of individuals (e.g., frustration), leaving a gap in between. The majority of empirical studies on the visual quality of surrounding environments have mostly focused on the users' perceived visual experience of the spatial variables rather than the change in their psychological state.

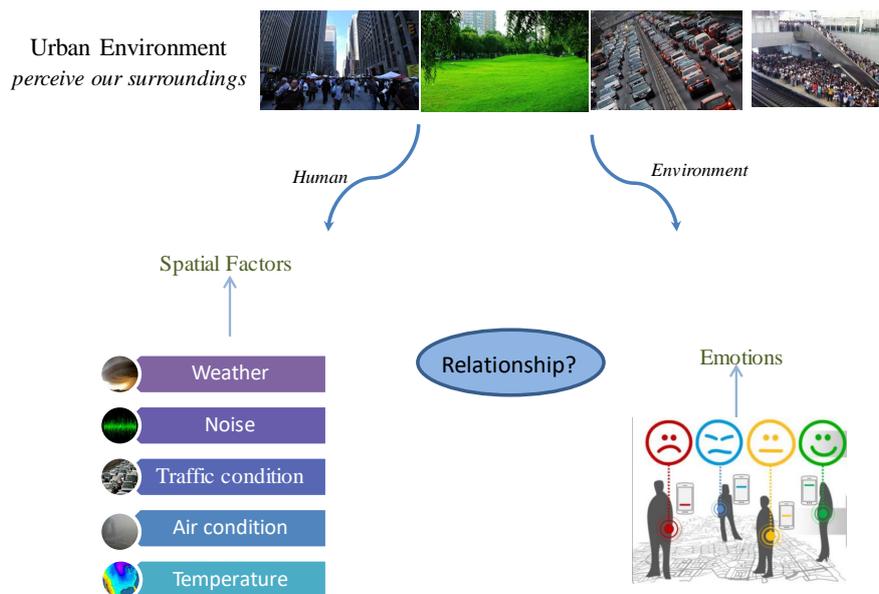
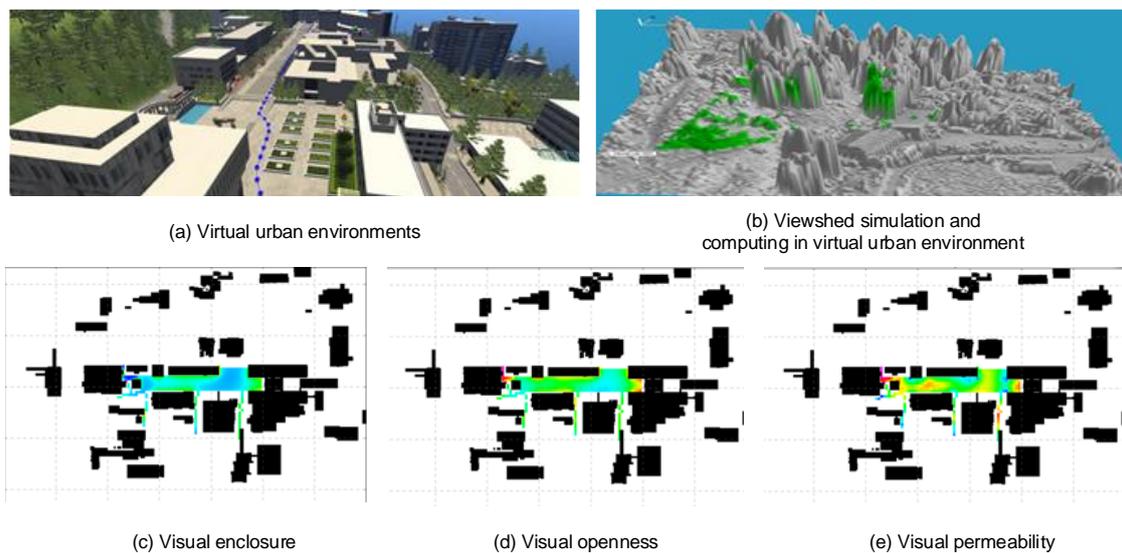


Figure 9. Multi-modal sensing for understanding human environmental perception.

Previous approaches have relied on interviews and questionnaires to collect urban users' physical responses and the related emotional changes to which they are confined, making quantitative analysis difficult to conduct. This study presents an EEG and immersive virtual environment-based framework for exploring the quantitative relationships among the spatial features of urban environments and the psychological responses of individuals, thereby facilitating visual quality assessment in urban design. The framework consists of the following three components:

*(a) Multi-Modal Immersive 3D Urban Virtual Environments*

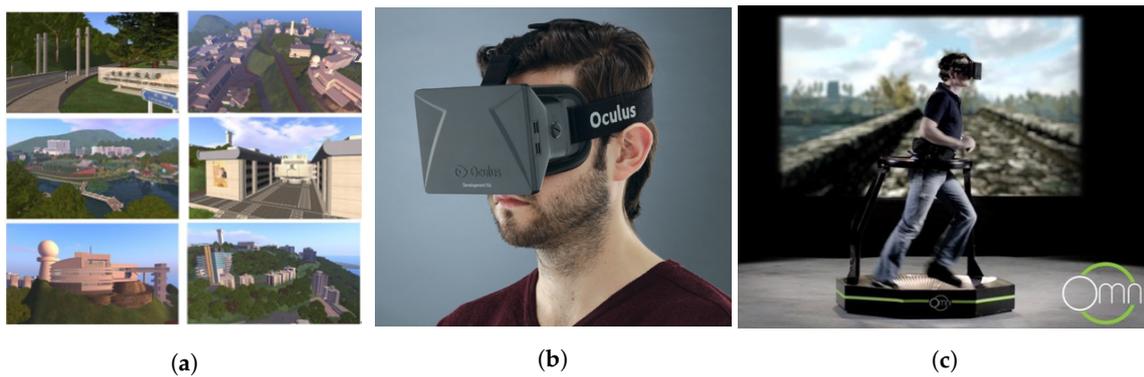
VGEs is capable of simulating real environments at different scales to satisfy the requirements of human cognitive experiments. To provide a life-like geographical experience to the users, VGEs implements multi-dimensional representations of urban spaces and multi-modal immersive devices for human perception in a virtual-reality-mixed manner. A collaborative modelling approach to creating 3D contents at multiple levels of detail (LODs) will be provided to support the whole scene of the virtual urban environment, from a simple bounding appearance to real 3D interior layouts. Figure 10 depicts the virtual urban environment designed and controlled by three spatial variables: enclosure, openness, and permeability.



**Figure 10.** Controlled spatial relationships in virtual urban environments.

In addition, a set of immersive virtual-reality devices will be seamlessly integrated into the virtual urban environment platform to create an immersive experience for participants and eliminate unnecessary natural effects simultaneously. As shown in Figure 11, a virtual-reality headset and an immersive movement platform are provided for the participant. The virtual-reality headset offers a basic visual environment, and the immersive movement platform provides a basic interactive environment for users.

The immersive virtual environment with 3D spatial features and immersive device utilized by participants provide a potential solution for incorporating individuals' feelings from the visual assessments into a dynamic simulation of complex urban environments.



**Figure 11.** Implementation of Immersive Virtual Geographic Cognitive Environments. (a) Virtual urban environment (Virtual CUHK Campus as a case study); (b) Immersive virtual-reality headset (Oculus Rift from Samsung); (c) Immersive movement platform (Virtuix Omni).

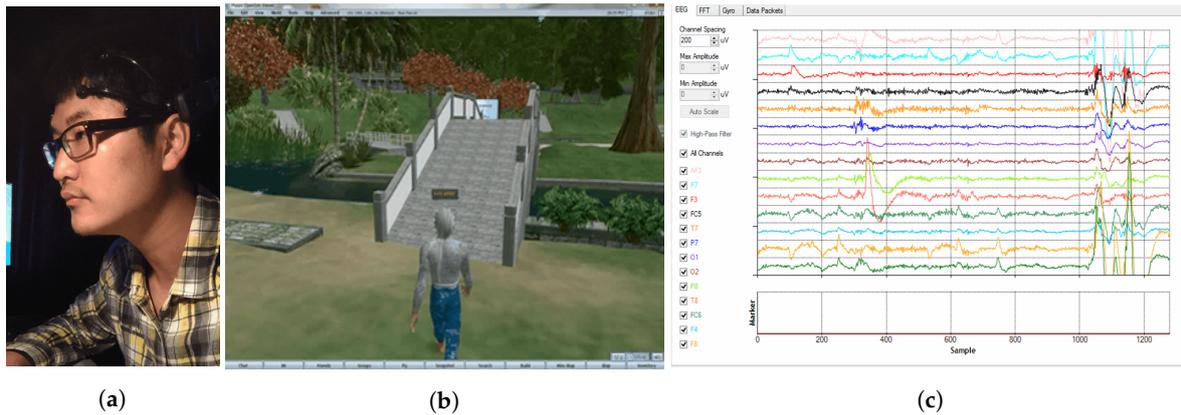
*(b) Real-Time Collaborative Sensing with Mobile EEG Device*

EEG is an electrophysiological monitoring method to record electrical activity of the brain. To measure the human response under different virtual environment settings, this study employs a type of mobile EEG device—the EMOTIV ([www.emotiv.com](http://www.emotiv.com)) (EMOTIV) mobile neuroheadset [70] (as shown in Figure 12). The EMOTIV mobile neuroheadset is used to monitor and record the brain activities of participants during the virtual cognitive experiment. Compared with a traditional EEG recorder, which is highly sensitive to the participants' movements during the experiment, EMOTIV allows motion by the user to a certain extent [71]. Compared with the traditional EEG device in neuro-science, this feature enables the participants to move or even walk around to a certain extent. Using the mobile EEG recorder, and combined with the immersive class and walking platform, it is possible for users to enter the virtual world immersively and interact with the virtual environment not only visually but also using interoperation as in the real world.

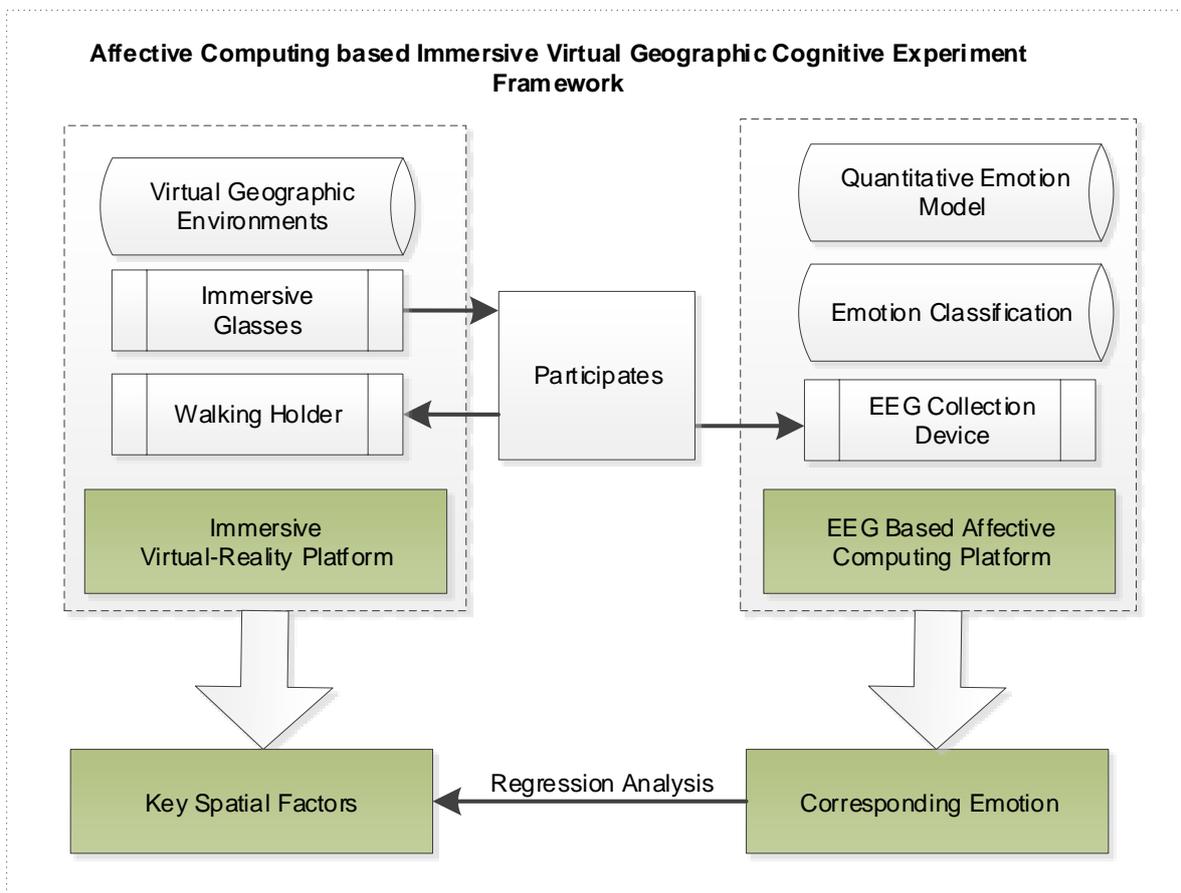


**Figure 12.** EMOTIV mobile neuroheadset for measuring an individual's brain activities in virtual cognitive experiments.

Figure 13 demonstrates the implementation of the EEG-based cognitive experiment. The participant (Figure 13a) enters the 3D virtual world as an avatar (Figure 13b). As the participants "feel" and experience the visual changes in the virtual environment during their movement, their brain activity will be recorded in real time (Figure 13c). The experiment will help to understand which visual features in urban environments positively/negatively impact human perception.



**Figure 13.** Implementation of EEG-based Virtual Geographic Cognitive Experiment. (a) Participant in the real world; (b) Avatar in the virtual environment; (c) Participant’s real-time EEG wave.



**Figure 14.** Affective-Computing-based Immersive Virtual Geographic Cognitive Experiment Framework.

*(c) Affective Computing and Cognition Techniques*

With the EEG data recorded in the virtual cognitive experiment, the next step is to analyse and interpret the brain activity and understand human emotion and perception changes during the experiment. Affective computing involves new technologies and theories that advance the basic understanding of affective phenomena and their role in human experience. It is the study and development of systems and devices that can recognize, interpret, process, and simulate human

affects [52,72]. An affective-computing-based machine should interpret the emotional states of humans and adapt its behaviour to them, therein giving an appropriate response to those emotions. Enabled by the proliferation of artificial intelligence technology and deep learning, interpreting complex human psychological activity has moved from an elusive goal to a recent multi-discipline focus [72]. In this study, the framework will follow the paradigm in affective computing research to explore participants' emotional changes in different visual variable settings of the virtual urban environment. Furthermore, by employing a series of statistic tools, it is feasible to further quantify the connections between visual features in the urban visual environment and different perceptions of individuals. Figure 14 demonstrates the overall framework of affective-computing-based immersive virtual geographic cognitive experiments.

## 5. Conclusions

In this era of big data and artificial intelligence, researchers from various fields and disciplines have been facing the same issues: how to make good use of extensive data and information with the limited data mining and knowledge discovering approaches; how to incorporate the data from multiple sources and link the data together automatically and intelligently; and how to combine the Internet of Things, sensor networks and the real world for information acquisition and quick response to satisfy the requirements presented by human activities. Actually, by integrating physical geometry models, human behavioural models and geographic process models, the development of VGEs facilitates the incorporation of geographic knowledge, brain-computer interfacing, and human social simulation, therein building a geographic knowledge engineering and sharing system and improving the ability for handling and mining new geographic knowledge and understanding natural and social phenomena.

From the perspective of human cognition-behaviour analysis and simulation, previous work in VGEs has focused mostly on representing and simulating the real world to create an 'interpretive' virtual world and improve an individual's active cognition. In this paper, we discussed the outlook for VGEs in terms of human environment perception and cognition, and we proposed a framework for virtual cognitive experiments. Moreover, two potential implementation methods were proposed. Using deep-learning-based data mining methods, new knowledge can be potentially discovered with large-scale survey data. In addition, by integrating brain-computer interfacing techniques and affective computing technology, we are able to conduct cognitive experiments in immersive virtual environments to explore the key factors in suitable and sustainable environments.

Nevertheless, the scope that this paper discussed on future VGEs in terms of human cognition and behaviour analysis is limited by the specific domain and scenario. The implementation framework and case study introduced in this paper are set to enlighten future studies and gain further insight into bringing VGEs—the new generation of geographic cognition and analysis tools—into wider and more in-depth research areas.

**Acknowledgments:** This work was funded by the National Natural Science Foundation of China (Grant Nos. 41671378, 41371388 and 41401459) and the Hong Kong Research Grants Council under GRF grant number 14606715, which are gratefully acknowledged.

**Author Contributions:** Fan Zhang conceived of and designed the experiments, analysed the results and wrote the paper. Mingyuan Hu and Weitao Che developed the program and analysed the results. Hui Lin and Chaoyang Fang conceived of the paper and revised the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Pattison, W.D. The four traditions of geography. *J. Geogr.* **1964**, *63*, 211–216.
2. Smith, B.; Mark, D.M. Geographical categories: An ontological investigation. *Int. J. Geogr. Inf. Sci.* **2001**, *15*, 591–612.
3. Lynch, K. *The Image of the City*; MIT Press: Cambridge, MA, USA, 1960; Volume 11.

4. Hull, R.B.; Lam, M.; Vigo, G. Place identity: Symbols of self in the urban fabric. *Landsc. Urban Plan.* **1994**, *28*, 109–120.
5. Lin, H.; Gong, J.H. On virtual geographic environments. *Acta Geod. Cartogr. Sin.* **2002**, *31*, 1–6.
6. Lin, H.; Chen, M.; Lu, G.; Zhu, Q.; Gong, J.; You, X.; Wen, Y.; Xu, B.; Hu, M. Virtual geographic environments (VGEs): A new generation of geographic analysis tool. *Earth-Sci. Rev.* **2013**, *126*, 74–84.
7. Chen, M.; Lin, H.; Lu, G. *Virtual Geographic Environments*; Wiley: Hoboken, NJ, USA, 2017.
8. Lin, H.; Huang, F.; Lu, X.; Hu, M.; Xu, B.; Wu, L. Preliminary study on virtual geographic environment cognition and representation. *J. Remote Sens.* **2010**, *14*, 822–838.
9. Stokols, D. Environmental psychology. *Annu. Rev. Psychol.* **1978**, *29*, 253–295.
10. Bechtel, R.B. *Environmental Psychology*; Wiley Online Library: Hoboken, NJ, USA, 2002.
11. Hu, M.; Lin, H.; Chen, B.; Chen, M.; Che, W.; Huang, F. A virtual learning environment of the Chinese University of Hong Kong. *Int. J. Digit. Earth* **2011**, *4*, 171–182.
12. Lin, H.; Chen, M.; Lu, G. Virtual geographic environment: A workspace for computer-aided geographic experiments. *Ann. Assoc. Am. Geogr.* **2013**, *103*, 465–482.
13. Chen, M.; Lin, H.; Kolditz, O.; Chen, C. Developing dynamic virtual geographic environments (VGEs) for geographic research. *Environ. Earth Sci.* **2015**, *74*, 6975–6980.
14. Ordonez, V.; Berg, T.L. *Learning High-Level Judgments of Urban Perception*; Springer: Berlin, Germany, 2014.
15. Liu, Y.; Liu, X.; Gao, S.; Gong, L.; Kang, C.; Zhi, Y.; Chi, G.; Shi, L. Social sensing: A new approach to understanding our socioeconomic environments. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 512–530.
16. Nasar, J.L. *The Evaluative Image of the City*; Sage Publications: Thousand Oaks, CA, USA, 1997.
17. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
18. Sainath, T.N.; Mohamed, A.R.; Kingsbury, B.; Ramabhadran, B. Deep convolutional neural networks for LVCSR. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC, Canada, 26–31 May 2013; pp. 8614–8618.
19. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.
20. Campbell, A.; Choudhury, T.; Hu, S.; Lu, H.; Mukerjee, M.K.; Rabbi, M.; Raizada, R.D. NeuroPhone: Brain-mobile phone interface using a wireless EEG headset. In Proceedings of the Second ACM SIGCOMM Workshop on Networking, Systems, and Applications on Mobile Handhelds, New Delhi, India, 30 August 2010; pp. 3–8.
21. Ang, K.K.; Chua, K.S.G.; Phua, K.S.; Wang, C.; Chin, Z.Y.; Kuah, C.W.K.; Low, W.; Guan, C. A randomized controlled trial of EEG-based motor imagery brain-computer interface robotic rehabilitation for stroke. *Clin. EEG Neurosci.* **2015**, *46*, 310–320.
22. Ruiz, S.; Buyukturkoglu, K.; Rana, M.; Birbaumer, N.; Sitaram, R. Real-time fMRI brain computer interfaces: Self-regulation of single brain regions to networks. *Biol. Psychol.* **2014**, *95*, 4–20.
23. Suk, H.I.; Wee, C.Y.; Lee, S.W.; Shen, D. State-space model with deep learning for functional dynamics estimation in resting-state fMRI. *NeuroImage* **2016**, *129*, 292–307.
24. Popelka, S.; Dedkova, P. *Extinct Village 3D Visualization and Its Evaluation with Eye-Movement Recording*; Springer: Berlin, Germany, 2014.
25. Herman, L.; Popelka, S.; Hejlova, V. Eye-tracking Analysis of Interactive 3D Geovisualization. *J. Eye Mov. Res.* **2017**, *10*, 2.
26. Wen, H.; Shi, J.; Zhang, Y.; Lu, K.H.; Liu, Z. Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision. *arXiv* **2016**, arXiv:1608.03425.
27. Juřík, V.; Šašínska, Č. Learning in Virtual 3D Environments: All About Immersive 3D Interfaces. In Proceedings of the International Conference on Education and New Learning Technologies, Barcelona, Spain, 4–6 July 2016.
28. Conroy, R.A. Spatial Navigation in Immersive Virtual Environments. Ph.D. Thesis, University of London, London, UK, 2001.

29. Dias, M.; Eloy, S.; Carreiro, M.; Proença, P.; Moural, A.; Pedro, T.; Freitas, J.; Vilar, E.; d'Alpuim, J.; Azevedo, S. Designing Better Spaces for People. In *Rethinking Comprehensive Design: Speculative Counterculture, Proceedings of the 19th International Conference on Computer-Aided Architectural Design Research*; Kyoto Institute of Technology: Kyoto, Japan, 2014.
30. Dias, M.; Eloy, S.; Carreiro, M.; Vilar, E.; Marques, S.; Moural, A.; Proença, P.; Cruz, J.; d'Alpuim, J.; Carvalho, N.; et al. Space perception in virtual environments-on how biometric sensing in virtual environments may give architects users's feedback. In *Proceedings of the 32nd eCAADe Conference, Newcastle upon Tyne, UK, 10–12 September 2014; Volume 2*, pp. 271–280.
31. Vilar, E.; Rebelo, F.; Noriega, P.; Duarte, E.; Mayhorn, C.B. Effects of competing environmental variables and signage on route-choices in simulated everyday and emergency wayfinding situations. *Ergonomics* **2014**, *57*, 511–524.
32. Vilar, E.; Rebelo, F.; Noriega, P. *Comparing Three Stimulus Presentation Types in a Virtual Reality Experiment to Human Wayfinding Behavior During Emergency Situation*; Springer: Berlin, Germany, 2017.
33. Juřík, V.; Herman, L.; Šašinka, Č.; Stachoň, Z.; Chmelík, J. When the display matters: A multifaceted perspective on 3D geovisualizations. *Open Geosci.* **2017**, *9*, 89–100.
34. Juřík, V.; Herman, L.; Kubíček, P.; Stachoň, Z.; Šašinka, Č. Cognitive Aspects of Collaboration in 3D Virtual Environments. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 663–670.
35. Špriňarová, K.; Juřík, V.; Šašinka, Č.; Herman, L.; Štěřba, Z.; Stachoň, Z.; Chmelík, J.; Kozlíková, B. Human-Computer Interaction in Real-3D and Pseudo-3D Cartographic Visualization: A Comparative Study. In *Cartography-Maps Connecting the World*; Springer: Berlin, Germany, 2015; pp. 59–73.
36. Voinov, A.; Çöltekin, A.; Chen, M.; Beydoun, G. Virtual geographic environments in socio-environmental modeling: A fancy distraction or a key to communication? *Int. J. Digit. Earth* **2017**, 1–12. doi:10.1080/17538947.2017.1365961.
37. Blascovich, J.; Loomis, J.; Beall, A.C.; Swinth, K.R.; Hoyt, C.L.; Bailenson, J.N. Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* **2002**, *13*, 103–124.
38. Goodchild, M.F. In the World of Web 2.0. *Int. J.* **2007**, *2*, 24–32.
39. Wu, X.; Zhu, X.; Wu, G.Q.; Ding, W. Data mining with big data. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 97–107.
40. Anderson, M.L. Embodied cognition: A field guide. *Art. Intell.* **2003**, *149*, 91–130.
41. Wilson, M. Six views of embodied cognition. *Psychon. Bull. Rev.* **2002**, *9*, 625–636.
42. Shapiro, L. *The Routledge Handbook of Embodied Cognition*; Routledge: London, UK, 2014.
43. Kern, N.; Schiele, B.; Schmidt, A. *Multi-Sensor Activity Context Detection for Wearable Computing*; Springer: Berlin, Germany, 2003.
44. Gong, J.; Zhou, J.; Zhang, L. Study Progress and Theoretical Framework of Virtual Geographic Environments. *Adv. Earth Sci.* **2010**, *25*, 915–926.
45. Jarmon, L.; Traphagan, T.; Mayrath, M.; Trivedi, A. Virtual world teaching, experiential learning, and assessment: An interdisciplinary communication course in Second Life. *Comput. Educ.* **2009**, *53*, 169–182.
46. Harris, H.; Bailenson, J.N.; Nielsen, A.; Yee, N. The evolution of social behavior over time in second life. *Presence* **2009**, *18*, 434–448.
47. Oinas-Kukkonen, H.; Harjumaa, M. Persuasive systems design: Key issues, process model, and system features. *Commun. Assoc. Inf. Syst.* **2009**, *24*, 28.
48. Turban, E.; Liang, T.P.; Wu, S.P. A framework for adopting collaboration 2.0 tools for virtual group decision making. *Group Decis. Negot.* **2011**, *20*, 137–154.
49. Hart, J.K.; Martinez, K. Environmental sensor networks: A revolution in the earth system science? *Earth-Sci. Rev.* **2006**, *78*, 177–191.
50. Hämäläinen, M.; Hari, R.; Ilmoniemi, R.J.; Knuutila, J.; Lounasmaa, O.V. Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.* **1993**, *65*, 413.
51. Neisser, U. *Cognitive Psychology: Classic Edition*; Psychology Press: Oxon, UK, 2014.
52. Picard, R.W.; Picard, R. *Affective Computing*; MIT Press: Cambridge, MA, USA, 1997.
53. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.R.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N.; et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97.

54. Salesses, P.; Schechtner, K.; Hidalgo, C.A. The collaborative image of the city: Mapping the inequality of urban perception. *PLoS ONE* **2013**, *8*, e68400.
55. Dubey, A.; Naik, N.; Parikh, D.; Raskar, R.; Hidalgo, C.A. *Deep Learning the City: Quantifying Urban Perception at a Global Scale*; Springer: Berlin, Germany, 2016; pp. 196–212.
56. Naik, N.; Philipoom, J.; Raskar, R.; Hidalgo, C. Streetscore-predicting the perceived safety of one million streetscapes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 779–785.
57. Porzi, L.; Rota Bulò, S.; Lepri, B.; Ricci, E. Predicting and understanding urban perception with convolutional neural networks. In Proceedings of the 23rd ACM international conference on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 139–148.
58. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
59. Datta, R.; Joshi, D.; Li, J.; Wang, J.Z. *Studying Aesthetics in Photographic Images Using a Computational Approach*; Springer: Berlin, Germany, 2006; pp. 288–301.
60. Datta, R.; Li, J.; Wang, J.Z. Algorithmic inferencing of aesthetics and emotion in natural images: An exposition. In Proceedings of the 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. 105–108.
61. Dhar, S.; Ordonez, V.; Berg, T.L. High level describable attributes for predicting aesthetics and interestingness. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011; pp. 1657–1664.
62. Zhou, B.; Lapedriza, A.; Khosla, A.; Oliva, A.; Torralba, A. Places: A 10 million Image Database for Scene Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *PP*, 1.
63. Joachims, T. *Text Categorization with Support Vector Machines: Learning with Many Relevant Features*; Springer: Berlin, Germany, 1998; pp. 137–142.
64. Southworth, M. Designing the walkable city. *J. Urban Plan. Dev.* **2005**, *131*, 246–257.
65. Zhou, B.; Zhao, H.; Puig, X.; Fidler, S.; Barriuso, A.; Torralba, A. Scene Parsing through ADE20K Dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
66. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
67. Ashihara, Y. *The Aesthetic Townscape*; The MIT Press: Cambridge, MA, USA, 1983.
68. Sitte, C. *City Planning According to Artistic Principles*; Rizzoli: 1986.
69. Gosling, D. *Gordon Cullen: Visions of Urban Design*; John Wiley & Son Ltd.: Hoboken, NJ, USA, 1996.
70. Aspinall, P.; Mavros, P.; Coyne, R.; Roe, J. The urban brain: Analysing outdoor physical activity with mobile EEG. *Br. J. Sports Med.* **2015**, *49*, 272–276.
71. Vala, N.; Trivedi, K. Brain computer interface: Data acquisition using non-invasive Emotiv Epoc neuroheadset. *Int. J. Softw. Hardw. Res. Eng.* **2014**, *2*, 127–130.
72. Cambria, E. Affective computing and sentiment analysis. *IEEE Intell. Syst.* **2016**, *31*, 102–107.

