Energy Proceedings
Vol 32, 2023
ISSN 2004-2965

Development of Augmented Reality Visualization Prototype System for Immersive Display of Building Carbon Emissions Using Urban 3D Models

Wenjie Gong[1], Ge Li [2*], Qi Chen[2], Qing Yu[3], Zhiling Guo[4], Haoran Zhang[5]

1 School of Information Management, Central China Normal University, Wuhan 430079, China

2 School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan 430074, China

3 Research Institute of Trustworthy Autonomous Systems, Southern University of Science and Technology, Shenzhen 518055, China

4 Center for Spatial Information Science, University of Tokyo, Kashiwa 277-8568, Japan

5 School of Urban Planning and Design, Peking University, Shenzhen 100871, China

(*Corresponding Author: rsge99@cug.edu.cn)

ABSTRACT

Monitoring the carbon emissions of urban buildings is of great significance for assessing the benefits of emission reduction and promoting the development of energy-saving technologies in buildings. Visualizing building carbon emissions helps intuitively understand carbon intensity and enhances public awareness of environmental protection and their willingness to participate in carbon reduction. However, existing building carbon emission visualization methods are generally limited to data statistics and map visualization, lacking immersive visualization solutions from the first-person perspective and integration with street scenes. To address this issue, this paper proposes an augmented reality (AR) visualization prototype system based on urban 3D models, which achieves an immersive display of building carbon emission data by accurately aligning each frame of a street-view video from a mobile phone with the backend urban 3D model. For an arbitrary input street-view video, we first locate its position in the urban 3D model based on its location information. We then use a scene segmentation model based on deep learning to obtain segmentation results for buildings, roads, and the background in each frame of the video. Finally, using segmentation results as prior information, we optimize the photographic parameters of the frame images to achieve high-precision alignment between each frame and the urban 3D model, thereby realizing the AR display of carbon emission data in the street-view video. The prototype system developed in this paper is expected to be deployed on mobile phones or wearable devices such as virtual reality glasses, which can be considered a promising visualization technology for observing building carbon emissions from a first-person perspective.

Keywords: building carbon emission, carbon emission visualization, urban 3D model, immersive display, image registration, augmented reality

1. INTRODUCTION

Global warming poses a significant threat to human survival, primarily due to the increase in global carbon emissions caused by human activities. [1] Urban buildings play a significant role in the emissions, [2] making monitoring the carbon emissions from urban buildings a key strategy for reducing greenhouse gas emissions and mitigating climate change.

Effective monitoring can provide policymakers and researchers with an insight into the urban building carbon emissions and inform targeted policies and practices that promote a more sustainable and resilient built environment. Furthermore, visualizing carbon emissions provides an easy-to-understand and intuitive way to help people gain a clearer understanding of the sources and drivers of emissions.

i) Statistical graphs. The most common and applicable visualization method is using basic charts such as bars, lines, and pie charts [3-6] to statistically analyze the size, composition, and trends of carbon emissions.

ii) Sankey diagrams. Sankey diagrams can represent the composition and flow of carbon emissions in a system through the size, width, and direction of arrows,

This is a paper for Applied Energy Symposium 2023: Clean Energy towards Carbon Neutrality (CEN2023), April 23-25, 2023, Ningbo, China.

providing a more intuitive visualization of the proportion and composition of emissions within a region. [7]

iii) Geological maps. Some scholars combine carbon emissions with 2D/3D maps to reflect the spatial distribution of carbon emissions. Specifically, they represent different carbon emissions sizes using different colors on a 2D map [8]; For 3D maps, colors, and model heights are used to simulate building carbon emissions, emphasizing high-energy consumption and high-emission areas. [9] These methods are more connected to reality, and oriented to the actual analysis requirements.

However, the above visualization methods tend to focus on presenting statistical information on urban building carbon emissions from a macro perspective, which is usually more suitable for experts and government decision-makers to display information. These methods are not attractive and immersive for the general public, making it difficult to arouse public attention.

To increase public attention to statistical information on urban building carbon emissions, and further enhance public environmental awareness and participation in emission reduction supervision, it is necessary to consider adopting more intuitive and visually appealing visualization methods that are closer to the public's perspective. Therefore, this paper aims to develop an immersive visualization approach integrated with the street view from the public's first-person perspective.

Augmented reality (AR) technology, combining virtual information with real-world scenes, has become a mainstream technology that provides users with richer and more immersive experiences. It has been applied in many fields, such as medicine, industry, education, tourism, socializing, and gaming. [10] The popularity of various smartphone AR applications and wearable devices has provided a platform foundation for immersive visualization technology from a first-person perspective. Therefore, it is possible to present urban building carbon emissions data to individual users in an immersive way through AR. However, there has been no research or practical application of AR technology with carbon emissions to achieve immersive visualization targeting users. This paper proposes an AR visualization prototype system based on urban 3D models, which achieves an immersive display of urban building carbon emission data from a first-person perspective by accurately aligning the backend city 3D model with each frame of street view video.

2. MATERIAL AND METHODS

2.1 Material

The AR has two kinds of data, real data, and virtual data. In this paper, the 3D UrbanMOB project [11] was used as the virtual data provider, namely, to provide the carbon emission data.

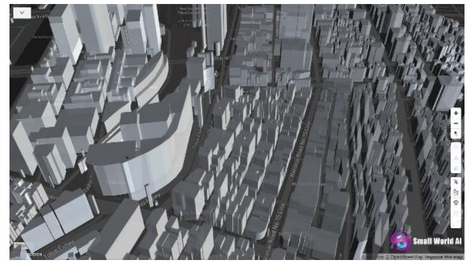


Fig. 1. Aerial view of the 3D UrbanMOB model

This project is based on a web platform and includes streets, buildings, and other realistic details as shown in Figure 1. We can set virtual camera parameters, including longitude, latitude, height, rotation angle, pitch angle, focal length, image width, and image height to render the model from a first-person perspective.

What we used as real data is a real-world street view video of Tokyo, Japan captured by a mobile camera. The video has a duration of 12 seconds and a frame rate of 29.75 frames per second.

2.2 Methods

The visualization system proposed in this paper consists of three steps (as shown in Figure 1): first, we process the street scene image, directly use the deep-learning-based segmentation model on each frame of the street view video, and present the segmented label map in grayscale mode; then, because of the mismatching state at first, we perform registration to match the real images and the backend urban model. We define a camera in the backend 3D space of the urban model, which can move freely in the space and take pictures. From the true street scene image, we get initial camera parameters and input them to the camera to render simulated street view images within the range. Evaluating the similarity between the label map and the rendered images, we find the optimized virtual camera parameters of each frame of the street view video. Finally, based on the accurate registration, we can obtain the carbon emissions of each building from the backend database and add dynamic model materials of carbon emissions in the backend urban 3D model, then generate the new building carbon emission layer of each frame, combine it with the original video to achieve the visual augmentation.

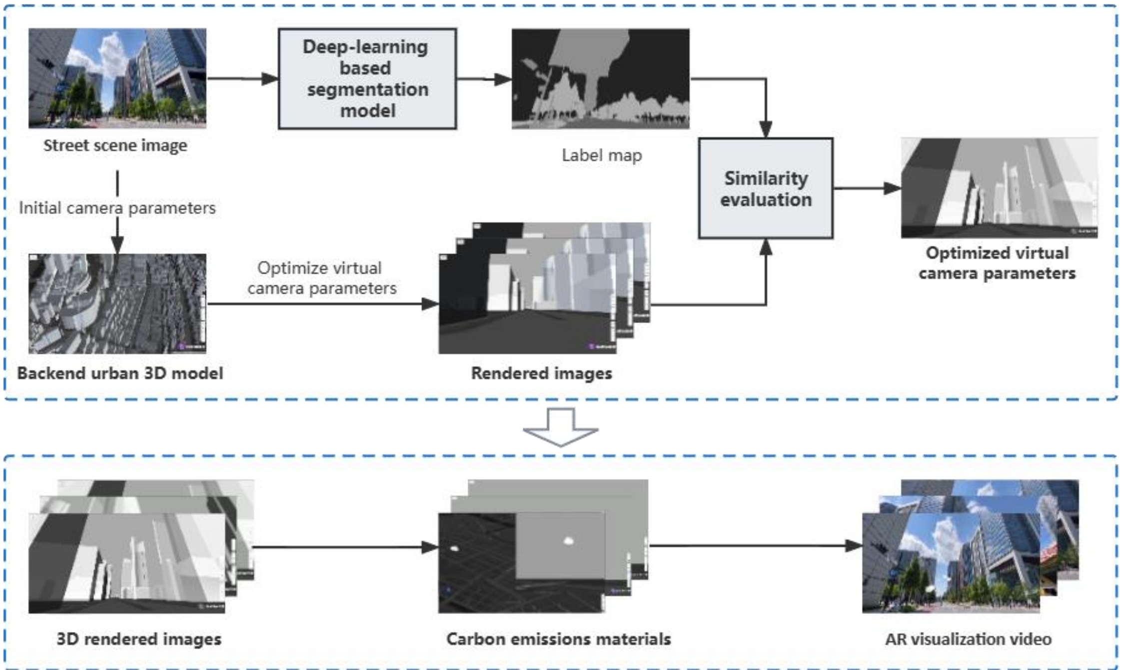


Fig. 2. The framework of proposed system

2.2.1 Street scene image segmentation

Real-world urban street view images contain various environmental components, including sky, roads, buildings, vegetation, vehicles, and pedestrians. To identify the specific position and pose of the video shoot, different objects in the scene must first be recognized through image segmentation.

The segmentation model used in this paper employs the deep learning method and utilizes the high-resolution convolutional neural network HRNet-W48 as the base model. We directly use the semantic segmentation model trained on the MSeg semantic segmentation dataset to segment the video image. [12] Then, apply this model to segment every frame of the street view video and obtain semantic segmentation labels of various objects such as the sky, buildings, and roads. Since only the position contour information of the image is needed for subsequent registration, we display the semantic segmentation label map in grayscale mode to avoid generating redundant data.

2.2.2 Registration between street scene and backend 3D model

Our system aims to integrate carbon emission data as virtual information into real street videos. To achieve this, we intend to align the real-world street view images with the urban 3D model that includes the carbon emission data of each building, enabling visualization of the carbon emissions of buildings in the image.

However, the initial localization of devices using various positioning satellite networks may result in imprecise and coarse camera parameters for shooting locations and poses, leading to a mismatch with the backend urban 3D model. This means that we cannot accurately locate the buildings in the street view images, and perform operations such as road and building recognition in the backend, thus accurately enhancing the user's reality on the client side. Therefore, we need to perform registration between the street view images and the urban 3D model to achieve precise analysis of the entire street view video and accurately project the information from the backend onto the user's view.

Since the street view image is presented in 2D while the urban model is presented in 3D, the street view matching problem is essentially a 2D/3D matching problem. To match data of different dimensions, we need to first match one data to the dimension of the other data.

We can access the semantic information of the backend urban model and adjust the position and pose of the camera in the backend to generate a corresponding 2D image with semantic information from the 3D model. Through iterating through camera parameters, we can compare the generated simulated images with the real street view images to continuously optimize the parameters and search for the simulated image that best matches the real image.

Optimization Parameters

We define a camera o in the 3D space of the urban model, which can move freely in space and take pictures. The virtual camera's position and orientation are determined by a combination of six parameters: latitude, longitude, height, azimuth angle, pitch angle, and focal length. These parameters constrain the imaging results of the camera's shooting and represent the optimization parameters that we continuously adjust to optimize the simulated image of the 3D model to match the real image. Therefore, the search space for optimization is defined by these six parameters.

Optimization Function

By providing the backend 3D model with the six parameters that describe the image, we can generate a first-person simulated street view of the model under those parameters. This simulation includes only three categories of objects: sky, buildings, and roads. In real scenes, buildings and roads can often be occluded, making it difficult to obtain their precise feature information. Therefore, we use the background sky of the image as a representative feature for the entire image.

Using the street view image segmentation label grayscale image as prior information, we compare the similarity between the sky shape in the street view segmentation image and that in the screenshot of the urban 3D model, which is used to replace the similarity between the two images.

The similarity between the background sky shapes SI_{bg} in the real street view image and RI_{bg} in the modeled simulated image is defined as:

$$E = sim(SI_{bg} \cdot RI_{bg})$$

where $sim(\cdot)$ selects the Jaccard index to measure the degree of overlap between two shapes.

Optimization Method

To obtain the optimized virtual camera parameters, we start with the initial six camera parameters and search within a certain range. Two optimization strategies are used in the search space: exhaustive search and ant colony optimization method. The exhaustive search traverses the search space and can obtain the optimized virtual camera parameters but requires a long time and a large number of model images. On the other hand, the ant colony optimization method uses a non-traversal optimization search strategy, which can search for the global optimal solution and has a smaller computational cost compared to the traversal search.

The street view video, as a definitive series of images, cannot obtain corresponding parameters for each frame when the image is taken. The automatic location information obtained from the video file is only valid for one frame, and using it as initial optimization parameters for each frame image will increase the time and space consumption of registration.

Due to the continuity of each frame image in the street view video, there is overlapping information between frames that can be inherited. Therefore, we can use a sequential processing method to match the video frame by frame in registration. We use the rough location information obtained by the device that recorded the video as the initial parameters for optimization when processing the first frame image, and use the optimal parameters of the previous frame image as the initial optimization parameters for subsequent frames.

2.2.3 AR visualization of building carbon emission

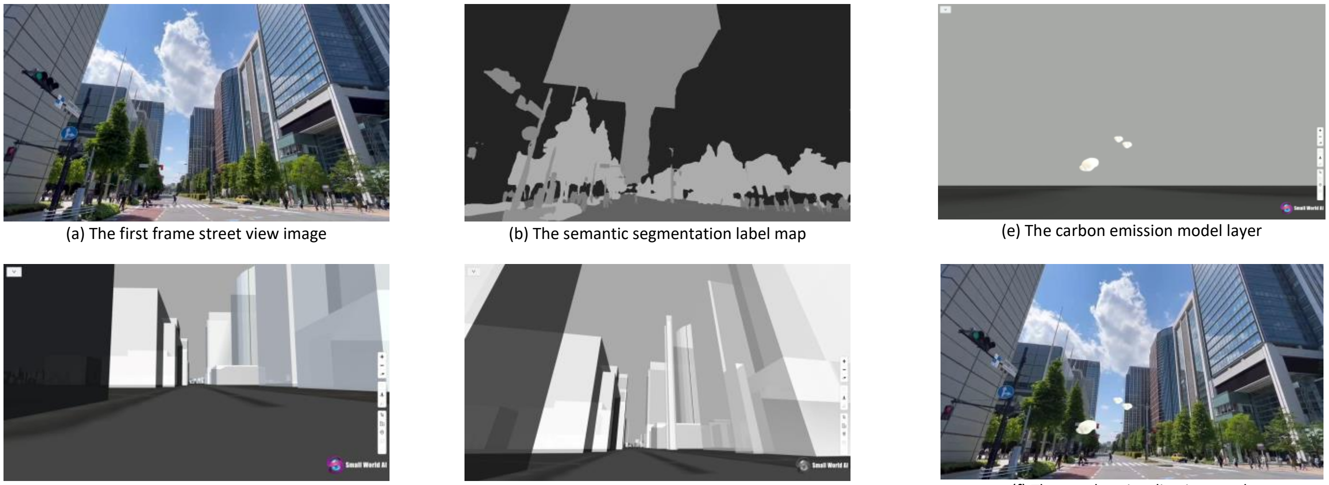
In this paper, we assume that the carbon emissions of each building in the backend urban model have been calculated and stored in the building information through a specific channel, but the calculation of carbon emissions is not within the scope of this research.

Based on the optimized virtual camera parameters and rendered images, we can obtain carbon emissions of the specific building from the backend database. Then, we can select a location to add dynamic 〖CO〗_2 materials to the corresponding buildings in the backend urban model to represent carbon emissions. Next, we input the optimized parameters sequence into the model and render the model frame by frame from a first-person perspective to obtain the building carbon emission layer. Finally, we overlay this layer with the street view images to achieve carbon emissions visualization augmentation in the street view images.

3. RESULTS

In the Linux system, we used Python and C programming system implementation. After inputting the street view video into our visualization system, the segmentation and registration results of its first frame image are shown in Figure 3(a-d).

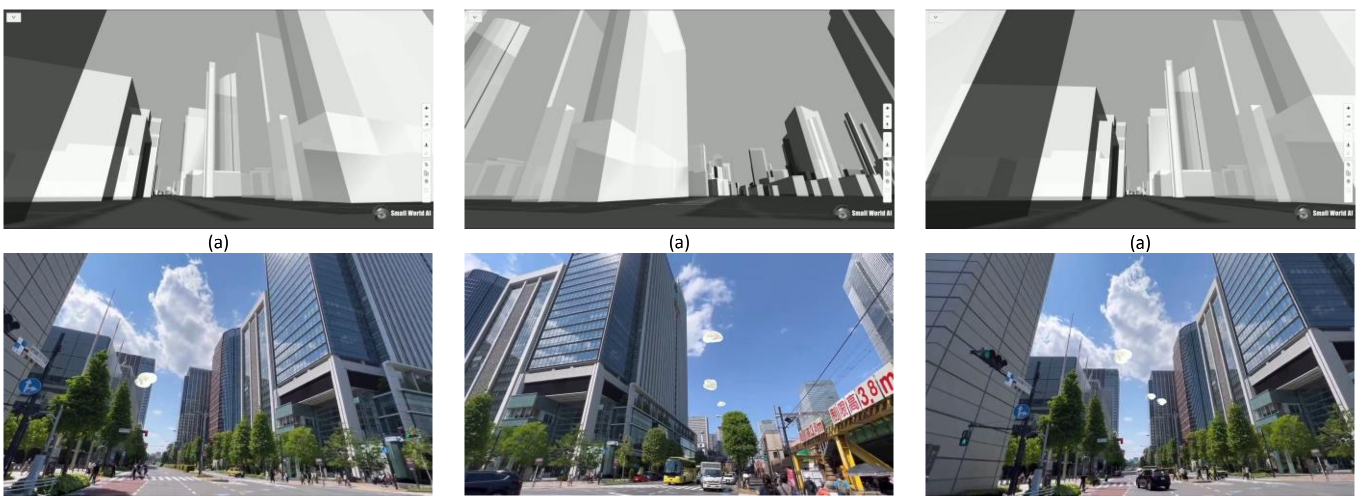
The street view video was segmented and registered frame by frame to obtain a registered 3D rendering video. Adding carbon emission models to the corresponding buildings in the backend urban model, and the visualization results in the current street view video were obtained by overlaying them. The carbon emission layer and the overlay visualization effect image of the first frame of the video are shown in Figure 3(e, f).



(a) The first frame street view image
(b) The semantic segmentation label map
(e) The carbon emission model layer
(c) The rendered image with initial camera parameters
(d) The rendered image after registration.
(f) The overlay visualization result

Fig. 3. The segmentation and registration results of the first frame image

The registration and rendering results of the frames for the front, middle, and back segments of the video, as well as the AR visualization effect, are shown in Figure 4.



(a) (a) (a)
(b) (b) (b)
Frame 30 **Frame 210** **Frame 350**

Fig. 4. The registration and visualization results of the frames 30, 210 and 350. (a) the first-person simulation street view rendered by the 3D model through registration; (b) the AR visualization of the corresponding frame image's carbon emissions

4. CONCLUSIONS

This paper proposes an AR visualization prototype system that visualizes carbon emissions data from a first-person perspective by segmenting actual street scene images and registering them with a backend urban 3D model. Experiments using Tokyo street scene videos prove the accuracy and feasibility of our research.

By registering with the urban model, the visualization system eliminates the need for building recognition, monolithic segmentation, simulated depth of field, and other operations on street scene images. This ensures the feasibility and stability of the visualization method. However, it is worth noting that the video image matching sequence method cannot be performed in multiple threads, resulting in limited operating speed, which is a potential field of future research.

Despite the limitations, the real-time visualization prototype system developed in this paper is still a promising technology for first-person perspective observation of building carbon emissions. It is expected to be deployed on mobile phones or wearable devices such as virtual reality glasses through technological optimization, thus achieving true real-time AR.

REFERENCE

[1] Huisingh D, Zhang Z, Moore JC, Qiao Q, Li Q. Recent advances in carbon emissions reduction: policies, technologies, monitoring, assessment and modeling. Journal of Cleaner Production 2015; 103: 1–12.

[2] United Nations Environment Programme. 2022 Global Status Report for Buildings and Construction: Towards a Zero‐emission, Efficient and Resilient Buildings and Construction Sector. Nairobi: 2022.

[3] Shao L, Chen GQ, Chen ZM, Guo S, Han MY, Zhang B, Hayat T, Alsaedi A, Ahmad B. Systems accounting for energy consumption and carbon emission by building. Communications in Nonlinear Science and Numerical Simulation 2014; 19: 1859–73.

[4] Yang Z, Gao W, Han Q, Qi L, Cui Y, Chen Y. Digitalization and carbon emissions: How does digital city construction affect china's carbon emission reduction? Sustainable Cities and Society 2022; 87: 104201.

[5] Zhang Y, Yan D, Hu S, Guo S. Modelling of energy consumption and carbon emission from the building construction sector in China, a process-based LCA approach. Energy Policy 2019; 134: 110949.

[6] Huang L, Krigsvoll G, Johansen F, Liu Y, Zhang X. Carbon emission of global construction sector. Renewable and Sustainable Energy Reviews 2018; 81: 1906–16.

[7] Abdelalim A, O'Brien W, Shi Z. Visualization of energy and water consumption and GHG emissions: A case study of a Canadian University Campus. Energy and Buildings 2015; 109: 334–52.

[8] Liu H, Yan F, Tian H. A Vector Map of Carbon Emission Based on Point-Line-Area Carbon Emission Classified Allocation Method. Sustainability 2020; 12: 10058.

[9] Wu Y, Sharifi A, Yang P, Borjigin H, Murakami D, Yamagata Y. Mapping building carbon emissions within local climate zones in Shanghai. Energy Procedia 2018; 152: 815–22.

[10] Mora CE, Martín-Gutiérrez J, Añorbe-Díaz B, González-Marrero A. Virtual Technologies Trends in Education. EURASIA Journal of Mathematics, Science and Technology Education 2017 Jan 18; 13.

[11] Yu Q, Shang W-L, Chen J, Zhang H. Web-based spatio-temporal data visualization technology for urban digital twin. Handbook of Mobility Data Mining 2023; 185–201.

[12] John Lambert, Zhuang Liu, Ozan Sener, James Hays, and Vladlen Koltun. MSeg: A Composite Dataset for Multi-domain Semantic Segmentation. In CVPR 2020.