International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2023)
A. Campbell, C. Krogmeier, and G. Young (Editors)

Landmark Recognition using Deep Learning in a Virtual Space

N. Mukai[1] T. Uematsu[1] and Y. Chang[1]

[1]Graduate School of Integrative Science and Engineering, Tokyo City University, Japan

Abstract

It is a very important issue to simulate human behavior in a virtual space for emergency evacuation in the real world. Humans take actions using their own eyes and memory. Then, the identification of the scene that virtual humans are looking at in a town is one of the key elements of the behavior, and image-based pattern matching is usually used; however, the accuracy is affected by the view angle and the length between the target object and the position at which the image is taken. This paper proposes a method to identify the images of landmarks that are placed at the corners in an intersection in a virtual space using a deep learning method and reports the relationship between the accuracy and the area rate that the landmark object occupies in the image.

CCS Concepts

• ***Computing methodologies*** → ***Instance-based learning;***

1. Introduction and Related Works

It is very useful to simulate human behavior in a virtual space for emergency evacuation in the real world such as earthquakes, conflagration, tornados, catastrophic floods, and so on. Then, there are some studies related to virtual human behavior. Mukai et al. proposed a model, where individual humans have their own eyes with which they can obtain the information necessary to take action. They also simulated crowd behavior just by applying the individual model to many people who form the crowd [MTC15]. In addition, Mukai et al. proposed a method for virtual humans to acquire their memory and to decide the direction to go by integrating the scene obtained through their own eyes with the information stored as the memory [MHC17]. Nakata et al. used deep neural networks (DNNs) to generate a human biological vision system [NCT18]. With the vision system, a virtual human can perceive a ball moving toward the human and can take action: reaching a hand to the ball and kicking the ball with the leg. On the other hand, Boiarov and Tyantov presented an approach to recognize landmarks in real images using deep metric learning [BT19].

In this paper, we propose a new approach to recognizing landmark objects in a virtual space using a deep learning method for virtual humans to understand the scene obtained through their own eyes by comparing it with the image stored in their memory.

2. Landmark Recognition

Figure 1 shows a birds-eye view of a virtual town generated according to the Building Standards Law in Japan. The areas surrounded by red, yellow, and green lines are low-rise residential zone, middle-rise residential zone, and commercial zone, respectively, and the blue line shows the road in the town.

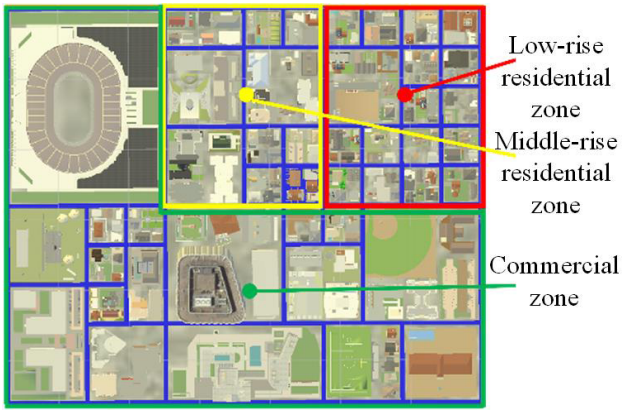

Figure 1: *Birds-eye view of a virtual town that has three kinds of zones.*

The virtual town shown in Figure 1 has 46 intersections including 25 three-way junctions and 21 crossroads. The landmark objects are defined as one building or house placed at every corner in all intersections. There are 159 $(= 25 \times 3 + 21 \times 4)$ landmarks in total, and 150 images are taken for one landmark. Then, $23,850$ $(= 159 \times 150)$ images are generated for the training of deep learning.

In this paper, we employ ResNet18 [HZRS16] as the deep learning model, and four-fold cross-validation is used for the validation.

DOI: 10.2312/egve.20231343

The validation results are shown in Table 1, where epoch, batch size, and learning rate were 32, 256, and 0.001, respectively.

Table 1: *Results of four-fold cross-validation*

#	Final loss	Validation Accuracy [%]
1	0.40	96.50
2	0.39	96.98
3	0.39	96.89
4	0.39	96.85

3. Simulation

The virtual town shown in Figure 1 has 25 three-way junctions and 21 crossroads, and there are three and four buildings and walking directions toward the center of the intersection for the three-way junction and the crossroad, respectively. Then, there are 561 ($= 25 \times 3 \times 3 + 21 \times 4 \times 4$) types of landmark images in total.

Then, 561 images were tested in the simulation, and 428 images were correctly recognized. Then, the accuracy was 76.3%. The accuracy of landmark recognition is affected by the area rate that the landmark object occupies in the image. Then, we call it the "area rate of landmark" in this paper.

Figure 2 shows the relationship between the area rate of the landmark and the average accuracy of the landmark, which is the average of the probability, with which each landmark is recognized. In the figure, there is a feature that the image, in which the area rate of the landmark is higher, has a higher accuracy since there is much information about the landmark in the image.

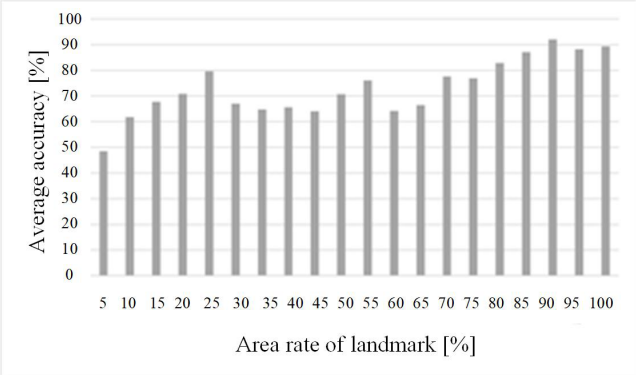

Figure 2: *Relationship between the area rate of the landmark and the average accuracy.*

On the other hand, Figure 3 shows the relationship between the area rate of the landmark and the number of identified images. From the figure, there are so many images, in which the area rate of the landmark is low, while there are few images, in which the area rate of the landmark is high. The reason is that many landmark images are taken from a distance. Then, the average accuracy would be increased if the landmark is recognized at a place nearer to the corner because the area rate of the landmark becomes higher.

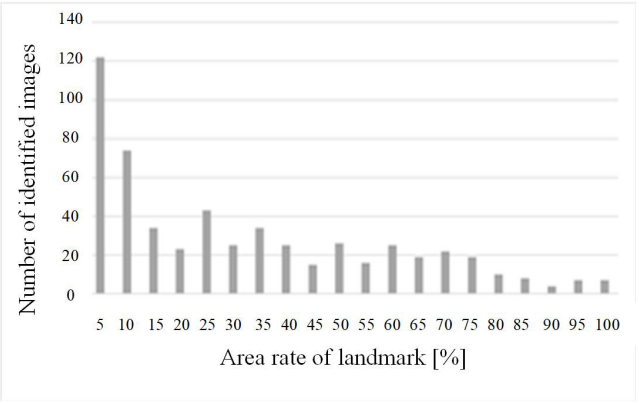

Figure 3: *Relationship between the area rate of the landmark and the number of identified images.*

4. Conclusion and Future Works

In this paper, we have proposed a new method to identify landmark objects using a deep learning method in a virtual space to investigate human behavior in the real world. For the simulation, we have created a virtual town according to the Building Standards Law in Japan and also generated landmark images at all junctions. After the learning process, the simulation was performed and the accuracy was investigated. As a result of the experiment, the average accuracy of the image, in which the area rate of the landmark is higher, became higher. However, there are many landmarks, in which the area rates are low, and this decreases the average accuracy because the area rate of the landmark is small due to the image being taken from a distance.

In the real world, people do not recognize the landmark objects when they perceive the junctions, and they approach the corner a little more to confirm if the landmark they are looking at is the same scene they saw in the past. In the future, we have to reconsider the position at which virtual people recognize landmarks to improve the recognition accuracy.

References

[BT19] BOIAROV A., TYANTOV E.: Large scale landmark recognition via deep metric learning. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (2019), pp. 169–178. 1

[HZRS16] 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* (2016), pp. 770–778. 1

[MHC17] MUKAI N., HAYASHI Y., CHANG Y.: Virtual human simulation on memory acquisition and walking with the memory. In *Proceedings of International Congress on Modelling and Simulation* (2017), pp. 354–360. 1

[MTC15] MUKAI N., TANAKA K., CHANG Y.: Crowd simulation by applying individual human model with vision. In *Proceedings of International Conference on Cyberworlds* (2015), pp. 210–215. 1

[NCT18] NAKADA M., CHEN H., TERZOPOULOS D.: Deep learning of biomimetic visual perception for virtual humans. In *Proceedings of the 15th ACM Symposium on Applied Perception* (2018), no. 20, pp. 1–8. 1