Interaction Technology Based on 3D printing topographic sand table for Emergency Management

Dong Wu
Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences
No. 20 Datun Rd.,
Chaoyang Dist., Beijing, China
+86-17801040015
915001258@qq.com

Jianhua Gong
Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences
No. 20 Datun Rd.,
Chaoyang Dist., Beijing, China
+86-10-64849299
gongjh@radi.ac.cn

Yi Li*
Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences
No. 20 Datun Rd.,
Chaoyang Dist., Beijing, China
Key Laboratory of Poyang Lake Wetland and Watershed Research, Ministry of Education, Jiangxi Normal University
No.99 Ziyang Rd., Nanchang, China
+86-10-64849299
liyi@radi.ac.cn

ABSTRACT

Enhance Virtual Reality is an innovative technique to perform dynamic 3D visualization and emergency management based on sand table, which needs convenient human-computer interaction techniques. Traditional interaction techniques such as mouse and keyboard cannot realize convenient interaction between human and the dynamic virtual scene of sand table. Gesture interaction method has received increasing attention in human-computer interaction. This paper researches interaction techniques based on computer vision and deep learning, and convolutional neural network was used in gesture recognition. A prototype system based on 3D printing sand table is proposed to achieve dynamic scene demonstration. The experiment shows that gesture recognition accuracy is 99.51%, and the visualization system can perform convenient operations on 3D printing sand table.

CCS Concepts

• **Human-centered computing → Visualization → Visualization application domains → Geographic visualization**

Keywords

3D Printing Sand Table; Artificial Intelligence; Human-Computer Interaction; Gesture Recognition; Emergency Management

1. INTRODUCTION

Topographic Sand Table has important applications in teaching, disaster relief deployment, the deployment of military action and so on [1]. 3D printing technology has developed rapidly in recent years, and it has become possible to rapidly print 3D printing topographic sand table [2, 3]. However, 3D printing topographic sand table can only show a static geographical environment and can't change with the real-time changes such as disaster relief and battlefields. Combining the 3D printing topographic sand table and projection technology, this paper superimposes the virtual scene with the 3D printing topographic sand table to realize the fusion of virtual and reality, which can realize the dynamic change of the scene. It can play an important role in the development of the Internet of Things, data visualization interaction, virtual geographical environment [4], emergency management such as dam-break flood [5], etc.

BDIOT 2018, October 24–26, 2018, Beijing, China.

ACM ISBN 978-1-4503-6519-2/18/10…$15.00
https://doi.org/10.1145/3289430.3289441

The ultimate goal of human-computer interaction is to provide an intuitive and natural interaction experience for the operator [6]. The forms of human-computer interaction are both speech and non-speech [7]. Non-speech interaction forms include gestures, facial expressions, and postures [8]. How to use computers to accurately recognize gestures has important research value. The artificial neural network can autonomously complete the extraction of gesture features and complete the gesture classification to achieve end-to-end gesture recognition. In recent years, some scholars have tried to apply CNN to the field of gesture recognition. Yuan Li et al. [8] applied CNN to depth images (RGB-D) to implement static gesture recognition; Yi Sheng et al. [9] extended 2D CNN to perform dynamic gesture recognition. Some scholars have used RGB-D static gesture datasets and dynamic gesture datasets to achieve better recognition performance [10-12], but research based on RGB static gesture datasets is still relatively small.

This paper proposes and implements an projection enhancement visualization system based on 3D printing sand table, which combines deep learning and computer vision to realize dynamic scene interaction. Based on the RGB gesture dataset, the static gesture recognition using CNN is realized, and on the basis of this, a gesture interaction operation based on a sand table system is realized. A convenient and smooth gesture interaction is achieved.

2. GESTURE RECOGNITION TECHENIQUES

This paper attempts to use CNN to classify the RGB gesture dataset to realize gesture recognition, and use gestures to control the virtual image which is projected on the topographic sand table to realize natural and convenient human-computer interaction. Since the presence of the non-gesture background in the gesture image will affect the recognition accuracy of the gesture, the skin region detection method based on opencv is used to locate the area of the hand gesture in this paper firstly; the image rotation and other methods of data augmentation are used for preprocessing the original data; Using the trained CNN model to classify gesture images to achieve gesture recognition, the experiment confirmed that the method has good recognition effect, and the recognition accuracy of different gestures reached 99.51%. The experimental process is shown in Figure 1.

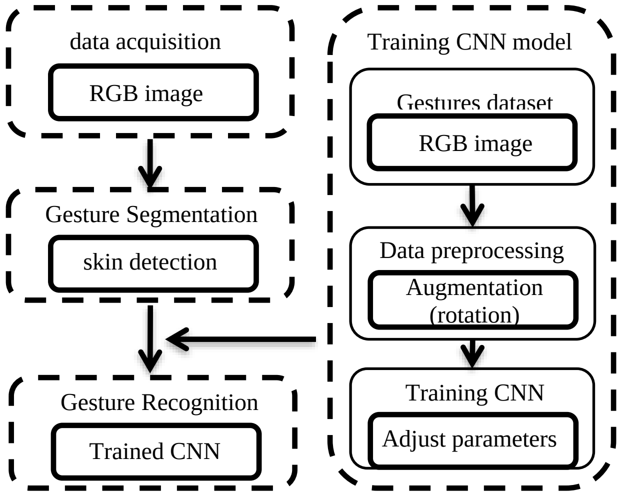

Figure 1. The process of gesture recognition

2.1 Hand gesture segmentation based on opencv

Because the background area of the non-human part of the gesture image will affect the gesture recognition, especially when the background area takes up a large area, the first step in the paper is to reduce the effect of the background area on gesture recognition by segmenting the gesture area in the image. The basic idea of gesture region segmentation selects the largest connected region that satisfies the skin color detection condition, and divides it with a rectangular frame to implement gesture region segmentation.

Because of the strong aggregation of skin color in color space, it is better to separate the gesture area from the background through the detection of skin color. Skin color detection can be carried out in different color spaces. The results of the study by the relevant scholars [13] show that the skin color has a stronger aggregation in the YCrCb color space than the HSV color space, and can reduce the effect of brightness on skin color detection.

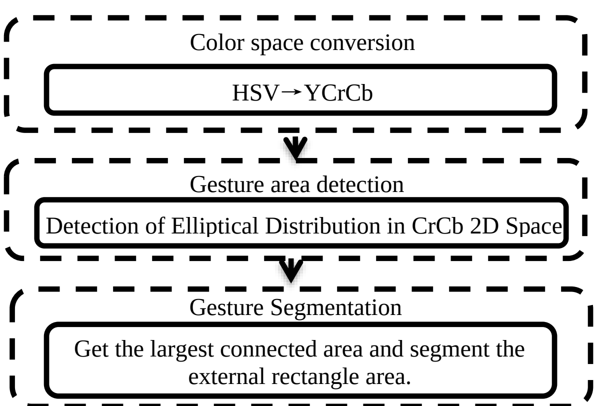

Figure 2.Gesture segmentation

In this paper, after the gesture image is converted to the YCrCb color space, the gesture region detection is performed using the ellipse detection model. The specific process is shown in Figure 2. Since there may be multiple areas in the image that are similar to the color of the skin of the hand, all the detected connected areas are compared in this paper to obtain the largest connected area to exclude the background area, and the gesture area is finally obtained.

2.2 Hand gesture recognition based on depth learning

As a feed-forward artificial neural network, convolutional neural networks have good performance for large-scale image processing. They are composed of multi-layer networks and usually include input layer, convolution layer, pooling layer, and fully connected layer. This paper adopts AlexNet model [14], the structure is shown in Figure 3.

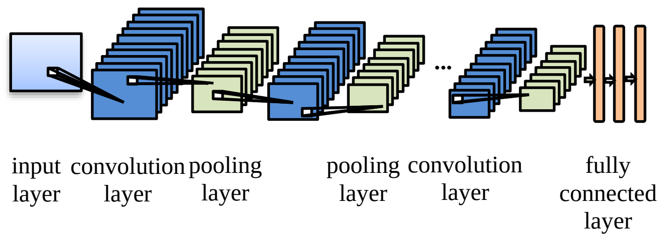

Figure 3. AlexNet

In order to enable the trained AlexNet to realize the recognition of static gestures, the data set used in the experiment must be a static gesture image, not a video. This article uses the data set The NUS hand posture datasets II[15], and the relevant information of the data is shown in Table 1. The data set includes 10 types of color gesture images. In this paper, the first 9 kinds of gestures are required.

Table 1. The NUS hand posture datasets II

Number of types of gestures	10
The number of each type of gesture image	200
The name of the gesture category	a,b,c,d,e,f,g,h,i,j
Image size	160×120
Image format	.jpg

In this paper, firstly, the CNN model is trained based on the original gesture image dataset. Because the data volume is too small, the training result is not ideal, and overfitting is produced. See the experimental results and analysis below. Therefore, in the data preprocessing stage, the data is augmented by rotating the original gesture image data, and the data amount is increased to 25 times of the original data amount. The experimental results show that the over-fitting is effectively avoided, and the classification accuracy of the model and the adaptability to gestures of different scales and rotation angles has been improved. The operation of data augmentation is to rotate the image clockwise and counterclockwise, and the rotation angle is 2°, 4°, 6°, 8°, 10°, 15°, 20°, 25°, 30°, 35°, 40°, 45 °.

Each class in the original gesture data set and the augmented data set is divided into a training dataset and a cross validation dataset according to the ratio of 4:1, and AlexNet is trained with the original dataset and the post-augmentation dataset respectively.

3. GESTURE INTERACTION DESIGN

3.1 Overall Structure

Based on 3D printing topographic sand table and projection technology, combined with image processing technology and deep learning, this paper constructs an interactive enhanced sand table system. Through projection, the virtual image is merged with the 3D printed topographic sand table, and gesture interaction is realized. The scene of the sand table system can be controlled through the hand gesture.

The conceptual diagram of the system is shown in Figure 4. Above the desktop is a projector. The map image is projected onto the desktop. The projection range is the entire desktop, which is divided into left and right sides. The left side is the active area of the three-dimensional printing sand table and the right side is the projection area of the three-dimensional scene. The system includes a total of two common USB cameras that capture sandbox images and gesture images, respectively. The camera that captures the image of the sandbox is placed on the projector; the camera that captures the gesture is placed behind the desk to capture gestures in front of the desk.

Figure 4. The concept map

3.2 Gesture interaction module

The model trained in this paper can recognize 9 types of gestures, namely a, b, c, d, e, f, g, h, i. Each of the nine gestures is given different semantics and corresponds to different functions, so as to achieve sand table gesture interaction. The corresponding semantics of various gestures are shown in Table 2.

Table 2. The corresponding semantics of various gestures

a	3D scene, moves forward
b	3D scene, moves backward
c	3D scene, moves left
d	3D scene, moves right
e	Stop moving scene
f	3D scene, Vertical clockwise rotation
g	3D scene, Vertical counterclockwise rotation
h	3D scene, decrease the height of the viewing angle
i	3D scene, increase the height of the viewing angle

The relationship between the functional modules involved in gesture interaction is shown in Figure 5. The CNN gesture recognition module and the gesture segmentation module of the server correspond to the scene control module of the client. First, the camera captures the gesture image and sends the image to the gesture segmentation module of the server. The gesture segmentation module divides the gesture region from the image based on the gesture segmentation algorithm. Then the image of the gesture area is used as input data. The CNN gesture recognition module uses the trained CNN to recognize and classify the gesture image to obtain a corresponding gesture ID.

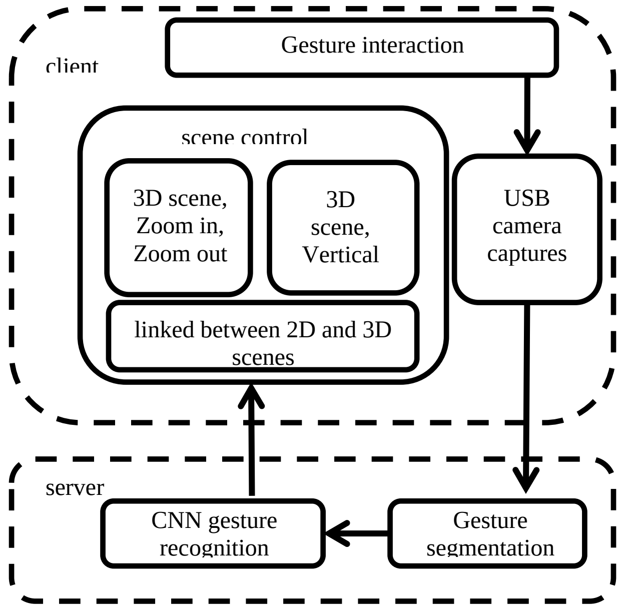

Figure 5. Gesture interaction service

After the gesture recognition is completed, the server sends the gesture ID to the scene control module of the client. The function in the scene control module is based on linked between 2D and 3D scenes to realize the movement, scaling, and rotation of the projected scene. The scene control module activates the corresponding function according to the received gesture ID, and implements gesture interaction of the system.

4. PROTOTYPE SYSTEM AND EXPERIMENTS

4.1 Prototype System

Finally developed a system based on a 3D printing sand table, see Figure 6. The overall structure of the system is basically the same as that of Figure 7. The left side of the projection area is the sand table area, and the projection image overlaps with the 3D printing sand table so as to realize the fusion of virtual image and sand table; the right side is a 3D scene area, the hand gesture can control the change of the scene and realize the roaming in the scene, as shown in Figure 7.

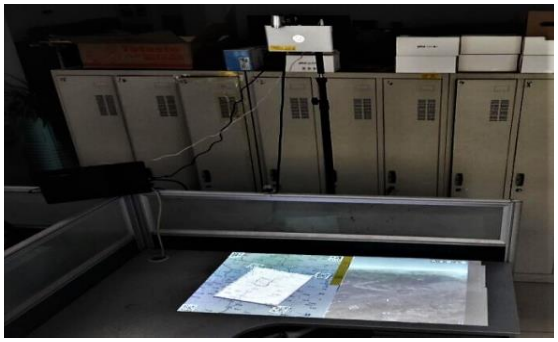

Figure 6. Prototype system

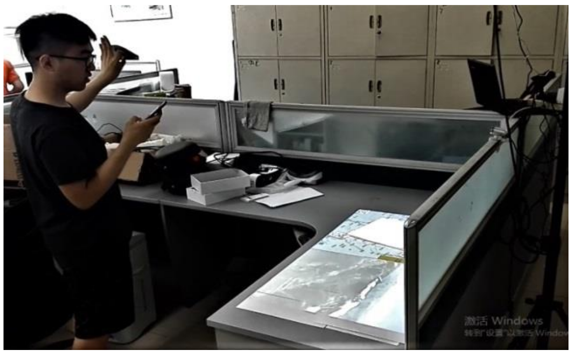

Figure 7. Gesture interaction

4.2 Experimental Results and Analysis

Figure 8 shows the classification accuracy based on the original gesture dataset. Figure 11 shows the classification accuracy based on the preprocessed gesture dataset. The vertical axis represents the classification accuracy, and the horizontal axis represents the number of iterations. Figure 9 shows the loss based on the original gesture dataset. Figure 10 shows the loss based on the preprocessed gesture dataset. The vertical axis represents the loss value and the horizontal axis represents the number of iterations. The evaluation of experimental results is based on cross-validation accuracy. The classification accuracy based on the original data has reached 93.00%, and the classification accuracy based on preprocessed data has reached 99.51%.

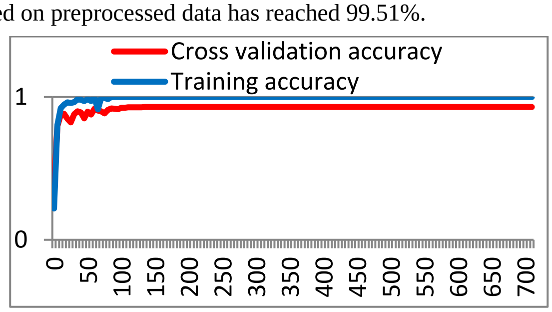

Figure 8. Classification Accuracy (original dataset)

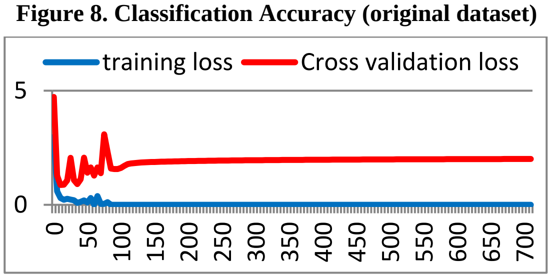

Figure 9. Loss (original dataset)

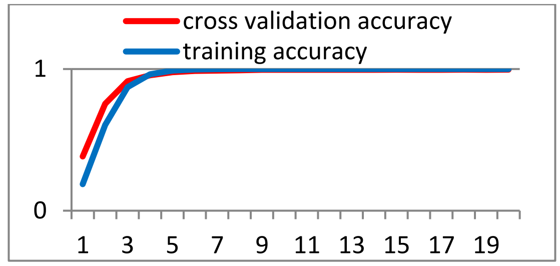

Figure 10. Classification Accuracy (preprocessed dataset)

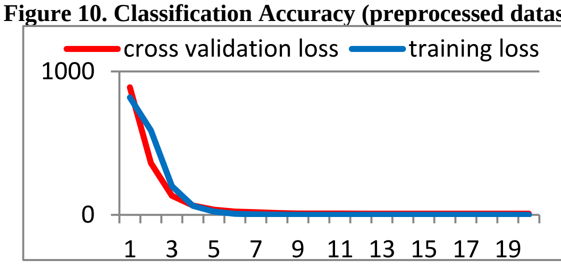

Figure 11. Loss (preprocessed dataset)

Figure 8 shows that as the number of iterations increases, the training accuracy and cross-validation accuracy generally increase and tend to be stable, and the trends of the two curves are consistent; Figure 9 shows that as the number of iterations increases, the loss function value of the training dataset generally shows a downward trend and eventually approaches zero, but the loss of the cross-validated dataset shows a U-shape, and a typical overfitting occurs. The training effect of the model is poor; Figure 11 shows that the loss of training data and cross-validation data sets shows a downward trend and approaches zero. The consistency of the two curves is good; Figure 10 shows that the training accuracy and cross-validation accuracy are on the rise as a whole and approach 1, and the trends of the two curves are in good agreement. After comparison, it can be found that after the original data was augmented, the overfitting problem was effectively avoided during the training process, and the classification accuracy was significantly improved, from 93.00% to 99.51%.

Based on the same gesture dataset[15], some scholars have used other meth ods to implement gesture recognition in recent years, and the results are listed in Table 3.

Table 3. Hand gesture recognition accuracy

Methods	Accuracy(%)
Our methods	99.51
Yuelong Chuang et al.[16]	95.27
Pisharady et al.[15]	94.36
Triesch et al.[17]	69.80

From the above analysis, we can find that the classification accuracy of the CNN model based on the preprocessed gesture data is the highest, reaching 99.51%. Therefore, in this system, this model is used for gesture recognition to achieve gesture interaction.

5. CONCLUSION

The 3D printing sand table system proposed in this paper can play an important role in the development of virtual geographical environment, battlefield situation deduction, tilt photography model display, and urban planning. This paper mainly studies the key techniques of gesture interaction that realize the interaction part of the system. Although this paper has done some research on the gesture interaction of the terrain sand table, there are still some shortcomings, which need further research: how to better integrate the 3D printing sand table and dynamic data visualization is one of the important research directions in the future. In terms of the visualization of the 3D scene of the system, the current 3D scene is visualized by projection on a 2D plane. We can try combining VR, AR technology or naked eye 3D technology to achieve stereoscopic visualization. Naturally and conveniently interacting with the system is the ultimate goal of interaction research in terms of system interaction. There are too many gestures for the current static gesture interaction, and it is difficult to remember in the actual application of the system. Therefore, static gesture interactions can be extended to dynamic gesture interactions to achieve a more natural and convenient way of interaction.

6. ACKNOWLEDGMENTS

This research is supported by the National Key Research and Development Program of China (2016YFB0502502); the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDA 20030302); the National Natural Science Foundation of China (41471341, 41571393); the Opening Fund of Key Laboratory of Poyang Lake Wetland and Watershed Research (Jiangxi Normal University), Ministry of Education (PK2016002).

7. REFERENCES

[1] Xie Y. Design and Implementation of Three-dimensional GIS Wisdom Scenic Spot［J］.Geomatics & Spatial Information Technology, 2017, 40(10):89-91..

[2] Wang S, Lin L, Ma P G, et al. Based on the 3D Product to Produce Three Dimensional Terrain Data Model [J]. Geomatics & Spatial Information Technology, 2017(8):129-130.

[3] Zhao Z, Zhang T. Three-dimensional printing based on FreeForm and 3ds Max modeling[J]. Journal of Computer Applications, 2016.

[4] Lin H, Gong J H. On Virtual Geographic Environments[J]. Acta Geodaetica Et Cartographic Sinica, 2002, 31(1):1-6.

[5] Li Y, Gong J, Song Y, et al. Design and key techniques of a collaborative virtual flood experiment that integrates cellular automata and dynamic observations[J]. Environmental Earth Sciences, 2015, 74(10):7059-7067.

[6] Yi J G, Cheng J H, Ku X S. Review of Gestures Recognition Based on Vision [J]. Computer Science, 2016, 43(s1):103-108.

[7] Kaur H, Rani J. A review: Study of various techniques of Hand gesture recognition[C]// IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems. IEEE, 2017:1-5.

[8] Li Y, Wang X, Liu W, et al. Deep Attention Network for Joint Hand Gesture Localization and Recognition Using Static RGB-D Images[J]. Information Sciences, 2018.

[9] Yi S, Liang H, Ru F. Hand Gesture Recognition Based on Multi-column Deep 3D Convolutional Neural Network [J]. Computer Engineering, 2017, 43(8):243-248.

[10] Mueller F, Bernard F, Sotnychenko O, et al. GANerated Hands for Real-time 3D Hand Tracking from Monocular RGB[J]. 2017.

[11] Li Y, Wang X, Liu W, et al. Deep Attention Network for Joint Hand Gesture Localization and Recognition Using Static RGB-D Images[J]. Information Sciences, 2018.

[12] Karami A, Zanj B, Sarkaleh A K. Persian sign language (PSL) recognition using wavelet transform and neural networks[J]. Expert Systems with Applications, 2011, 38(3):2661-2667.

[13] Qiu D. Research About Skin Color Detection Base on HSV and YCrCb Color Space [J]. Computer Programming Skills & Maintenance, 2012(10):74-75.

[14] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.

[15] Pisharady P K, Vadakkepat P, Ai P L. Attention Based Detection and Recognition of Hand Postures Against Complex Backgrounds[J]. International Journal of Computer Vision, 2013, 101(3):403-419.

[16] Chuang Y, Chen L, Chen G. Saliency-guided improvement for hand posture detection and recognition[J]. Neurocomputing, 2014, 133(8):404-415.

[17] Triesch J, Malsburg C V D. A System for Person-Independent Hand Posture Recognition against Complex Backgrounds[J]. Pattern Analysis & Machine Intelligence IEEE Transactions on, 2001, 23(12):1449-1453.