Development of a Non-contact Autostereoscopic 3D Button Using Artificial Intelligence

Sang-Hee You, Min Hwang, Ki-Hoon Kim, and Chang-Suk Cho

Abstract This study presents the results of the development of a contactless button device that represents the button as an autostereoscopic vision and visually captures the position of the fingertip that presses the virtual button. To this end, a 3D stereoscopic expression module, a pointing location display module, an FPGA design for driving the modules, and an artificial neural network algorithm were developed. The pointing accuracy of the developed button device showed 99% accuracy in the recognition test in the laboratory.

Keywords Stereoscopic · Contactless · Virtual button · 3D · AI

1 Introduction

Pushbuttons, which have been used on most machines, have been around for a long time since industrialization began, but as modern three-dimensional technology rapidly evolves, it is time to replace them with new non-mechanical technologies. As a new button technology, the contactless button as the most promising candidate to replace the existing mechanical button is required in terms of security and hygiene. With the development of modern biometric technology, the need for security of touch-type buttons has emerged with the development of fingerprint recognition technology. In other words, the touch-type button has a problem that the password is easily exposed because there remains a trace of the contacts on the button, therefore there occurs a need to spread the non-contact type. In terms of hygiene, there is a desperate need for contactless buttons to prevent infection from highly infectious viruses or bacteria, such as the MERS virus in 2015 and the Corona virus that brought the 2020 pandemic. After experiencing the MERS crisis, which caused...
Korean society to be shocked by the virus epidemic, we started to develop a contactless button. In order to design such a non-contact button, it was necessary to consider three points in terms of real demand. One was that the non-contact type should not have much difference compared to the mechanical type in terms of manufacturing cost, and the other was that the durability should be better than the mechanical type, and the last was that the consumer should be able to feel the point of pressing the button non-contactly. In consideration of the above points, the expression of a virtual button by autostereoscopic vision was envisioned in order to provide a feeling that the push button is on the button plate even in a contactless manner. To reduce costs and ensure stability, light-emitting and receiving sensors composed of infrared diodes were used, and the data obtained from the sensors are identified through a simple deep learning network.

A non-contact button using humidity sensor was reported [1] but its application field is more restricted than our vision button due to using humidity information. Various types of autostereoscopic vision technology have been developed since the 1970s, and their implementation has also diversified from cinema screens to LCD [2–4]. The most popular method is the implementation of autostereoscopic vision by making a multi-view using a lenticular lens, which is simple to implement and inexpensive. The lenticular method [5] was also used in this study. The recognition of the pointing position was performed using a simple deep learning network [6–8]. The structure and concept of deep learning was introduced in earnest from the 1980s, and is based on the construction of the middle layer extending from the initial artificial neural network. In the recognition experiment with this experimental device, there was no significant difference in recognition accuracy between a simplified deep learning network and a more complex deep learning network including CNN filters. Based on this, a deep learning network with a simple structure was used to recognize the pointing position. The recognition success rate of the non-contact virtual button was recorded over 99%, so that commercial success can be expected.

2 Button Device with Lenticular Display

The lenticular method distributes left and right images using the refraction of a lens which has semi-cylindrical columns arranged vertically on a screen. Figure 1 shows the principle of autostereoscopic vision using lenticular method. That is, when the convex lens-shaped semi-cylindrical lenses are arranged on a transparent plate vertically and the left and right images are alternately arranged under the plate, the images are distributed to the left and right eyes as shown in Fig. 1. Since the distributed left and right images are combined in the brain as a 3D form, the object is seen in stereoscopic vision. The lenticular method is the most widely used method in the field of autostereoscopic vision, and its implementation cost is also the cheapest, whereas if the target screen is slightly out of the main viewing angle, stereoscopic vision is not well recognized. The depth information of the virtual 3D by the lenticular lens is proportional to the thickness of the lenticular lens plate. The thickness of the lens plate
used in this experiment is 4 mm, so the maximum depth of the virtual 3D is 1.8 cm. On the other hand a touch panel key system with lenticular lens was reported as an autostereoscopic button in Fujitsu (Patent application number: 10–2019-0,038,210, Korea). When a button in the system is expressed, then the touch panel detects the input as touch panel sensor. This method is a heavy and expensive system due to using touch screen, and the problem of recognizing the pointing position of a finger as non-contact has not been solved. Comparing to our proposed method the autostereoscopic touch panel has not been put into practical use yet. Comparing to those methods our proposed button does not require LCD screen or special sensor system, so that it has the merits of very low implementation cost, lightweight and contactless style.

The autostereoscopic button device is composed of a display part with lenticular lens, an infrared sensor part, a backlight part for each button, and a driving unit including artificial intelligence. Figure 2 shows the configuration of the button device. The button picture reconstructed by dividing a button photo into multi-viewpoints

Fig. 1 Lenticular autostereoscopic vision and the button set in lenticular display

Fig. 2 Non-contact 3D virtual button device
is mounted under the lenticular lens, and a backlight is disposed on the back side of the picture. The backlight serves to indicate the position where a virtual button is pointing. Without such a backlight a user can’t recognize whether a button has been pressed or not because it is a non-contact method with virtual button. At the each edge of the horizontal and vertical axis of the button device, infrared diodes are arranged.

Figure 3 shows the cross-section and pointing situation of the button device. When a fingertip enters into the virtual button floated into the air by the autostereoscopic vision, the AI system recognizes the button as pressed. Artificial intelligence is used to determine whether a finger has entered into the virtual button image. Figure 3 shows a cross-section of the button. The infrared diodes consist of light emitting sensors and light receiving sensors, and are arranged in three pairs on the upper and lower sides and four pairs on the left and right sides. In the case of the infrared light emitting diode, a total of 14 are arranged on the top, bottom, left, and right of the button device, and light is sequentially turned on when emitting. The infrared light emitting diode here emits light including wavelengths from near infrared to far infrared. The light-receiving diode does not selectively select a specific wavelength, but detects it integrally, including infrared wavelengths from near infrared to far infrared. The infrared light emitting sensors arranged for each position of the button device sequentially emit light clockwise one by one, so that the light receiving sensors detect light emission simultaneously from every light emission. Even if the light is emitted sequentially, it is instantaneous when observed, so it appears as a simultaneous light on the photo.

3 Pointing Algorithm Using Deep Learning

The deep learning network learns the $14 \times 14$ pixel map of images obtained by the receiving sensors shown in Fig. 4. The control board sequentially lights up 14 light emitting diodes and obtains a $14 \times 14$ pixel gray image composed of 14 gains from light-receiving diodes. The obtained gray image is learned in a deep learning
Development of a Non-contact Autostereoscopic 3D Button …

network on the control board and outputs the pressed button information. That is, when one light-emitting diode is turned on, 14 light-receiving diodes around the button set simultaneously receive infrared light to output the detected value. The light-receiving sensor in the area blocked by a fingertip outputs a small value and the sensor unblocked does a large value. By this principle, it is possible to capture the position of the fingertip. In Fig. 4 it can be seen a $14 \times 14$ image composed of light emission and light reception for each sensor. Figure 5 shows the lighting control and

Fig. 4 Gains from 14 receiving diodes and image map obtained

Fig. 5 The lighting control and learning system on PCB board
learning system on PCB board. The light-receiving diodes blocked by the fingertips are relatively clearly distinguished. Thus it can be determine the fingertip position by establishing a linear equation, so that the button pressed can be obtained without a deep learning network. However the problem is that the use of a linear relational expression is very vulnerable to changes in the lighting environment, so that we decided to use a deep learning network in spite of its simple input pattern.

As described above in Sect. 3, the gray image for input is an image created by 14 light emitting diodes sequentially emitting and 14 light receiving diodes recording the input value. Therefore, a relatively simple network with two middle layers was used because the input image has relatively simple and clear features.

Figure 6 shows the neural network used in this study. Since the deep learning network in Fig. 7 receives a single 14 × 14 pixel image as an input, the number of input nodes is 196, and the middle layers are composed of 100 and 50 nodes respectively, and the output layer is composed of 12 nodes. This network includes

![Fig. 6 Simple deep learning network for the fingertip-pointing](image)

![Fig. 7 Error according to learning iteration](image)
Development of a Non-contact Autostereoscopic 3D Button … 75

no convolution filters in layers due to its simple inputs. The two hidden layers used sigmoid function as an activation function but did not use the activation function in output layer (i.e. Bypass). The value of the output node was converted to a stochastic value (between 0 and 1) by using SOFTMAX in the cross-entropy method. The SOFTMAX function is an activation function generally used in the output layer of a deep learning model for classification of 3-class or higher. The SOFTMAX function expands the deviation of each value when there are K values making the larger value relatively larger and the smaller value relatively smaller, and then normalizing it. The final output layer here is 12 because the number of buttons consists of 0–9 and 2 special characters. As a cost function, MSE (Mean Square Error) between the output value and the correct answer was used.

4 Experimental Result

For the experiment, a deep learning network was trained with 100 learning images extracted per position by placing a finger at the position of 12 key areas of the 3D virtual button. As a learning environment, a teaching button set for inputting manually target value was designed for easy learning and connected to the stereoscopic button device during learning. When the network is being learned, a position pointed on the stereoscopic button is identified to the target position inputted on the teaching button set. If the button set is learned for 600 epochs which have 1200 images respectively under an environment, it shows 99% or more correct answers in our experiments. Figure 8 shows the error (MSE: Mean Square Error) according to 600 iterations. The MSE was computed in Eq. 1.

\[
MSE = \frac{\sum \delta(i)^2}{n}, \left\{ \begin{array}{ll} 
\text{if } i = T & \delta(i) = 0 \\
\text{else } i \neq T & \delta(i) = 1 
\end{array} \right. 
\]

\(i\) : Inputted value, \(T\) : Target  \(1\)

Reliability is very important in the button, so we have to go through two verification processes as shown in Fig. 8. That is, if the results of the two recognitions match, the results are output, and if they do not match, the recognition process is

Fig. 8  Double verification method to improve reliability
performed again. There is no problem in using this repetitive process because the
time difference is small enough that the user cannot feel it.

5 Conclusion and Future Work

This study presented the design of a stereoscopic virtual button as non-contact, the
configuration of a button sensor system, and the design of a simple deep learning
network to recognize it. If there is a clear design of the input data composition, it
proved that the deep learning network needs not be so complicated. The production
of contactless devices according to the recent social distance is expected to gradually
expand its scope. As a result of this study, the proposed new concept button has a
feature of high reliability of over 99% and simple hardware configuration with rela-
tively low cost. In the future, this device will be released after additional verification
processes for commercialization.

Acknowledgements This work was partly supported by the Hansin University Grant and
Technology development Program of MSS [S2758698] and of MSS [S2721154].

References

1. Mondal S, Kim SJ, Choi C-G (2020) Honeycomb-like MoS2 nanotube array-based wearable
sensors for noninvasive detection of human skin moisture. ACS Appl Mater Interfaces 17029–
17038
2. Dodgson NA (2005) Autostereoscopic 3D displays. IEEE Comput 38(8):31–35
3. Tyler CW, Clarke MB (1990) Autostereogram. In: Proceedings SPIE 1256, stereoscopic displays
and applications
4. Ezra D, Woodgate GI, Omar BA, Holliman NS, Harrold J, Shapiro LS (1995) New autostereo-
scopic display system. In: Proceedings SPIE 2409, stereoscopic displays and virtual reality
systems II
5. Bourke P (2014) Autostereoscopic lenticular images. Lenticular Prints. https://paulbourke.net/
stereographics/lenticular/
6. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional
neural networks. Advances in neural information processing systems, vol 25. NIPS
7. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image
recognition. In: Computer vision and pattern recognition, vol 6
8. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. IEEE CVPR,
pp 770–778