JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Implementation of an Autostereoscopic Virtual 3D Button in Non-contact Manner Using Simple Deep Learning Network

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

Abstract

This research presented an implementation of autostereoscopic virtual three-dimensional (3D) button device as non-contact style. The proposed device has several characteristics about visible feature, non-contact use and artificial intelligence (AI) engine. The device was designed to be contactless to prevent virus contamination and consists of 3D buttons in a virtual stereoscopic view. To specify the button pressed virtually by fingertip pointing, a simple deep learning network having two stages without convolution filters was designed. As confirmed in the experiment, if the input data composition is clearly designed, the deep learning network does not need to be configured so complexly. As the results of testing and evaluation by the certification institute, the proposed button device shows high reliability and stability.

1. Introduction

The pushbuttons on most machines have been around for a long time since the beginning of industrialization. However, with the rapid advancement of modern three-dimensional (3D) technology, it is time to replace it with new non-mechanical technology. Furthermore, advances in the latest biometric technologies such as fingerprint recognition have created a need for security of touch buttons. In other words, contact-type buttons have a problem in that their passwords are easily exposed because contact traces remain on the button surface, so it is necessary to popularize non-contact-type buttons. Especially from a hygiene perspective, contactless buttons are urgently needed to prevent infection by highly contagious viruses (e.g., COVID-19). Consequently, non-contact button is required as the most promising candidate to replace the pushbutton in terms of security and hygiene. After our society in East Asia went through the MERS (Middle East respiratory syndrome coronavirus) crisis in 2015, we started developing non-contact buttons.

To design the contactless button, we had to consider three requests in terms of actual demand. One is that the button should not have a big difference in manufacturing cost comparing to the mechanical type, the other is that it should be more durable than the mechanical type, and lastly, the consumer can have the feeling of pressing a button. In consideration of these points, a virtual button device was conceived

in this paper, which uses diode sensors and a neural network for sensing pointing position and shows virtual buttons to stereo vision by glasses-free stereoscopic method. The proposed device is composed of the light-emitting and receiving infrared diodes and the system board which can identify the pointing position through a two-stage deep learning network. It was certified by recording an accuracy of more than 99% in the evaluation at an accredited testing institute, and the production cost on mass production was also not high compared to the mechanical type, so that there were no problems in its distribution and commercialization.

A report for the virtual 3D button in non-contact manner was presented in 2020 [1] by our lab, which was focused to introduce the button concept. In this paper we present the empirical results for the detailed design of lenticular 3D representation process, the final hardware structure and assessment results.

2. Related Works with the Non-contact Button Technology

Until now, the development of the non-contact button has focused on the recognition technology that captures the pointing position. But our development focuses not only on the pointing recognition technology but also on implementing the expression technology which allows user to be able to feel the pointing in a virtual space. The non-contact button in this paper is composed of the display virtually expressed in 3D stereoscopic vision and the driving unit that recognizes the pointing position.

As a non-contact recognition technology, the technology using capacitive sensor [2] has been developed and applied to various devices. However, this is not advantageous over the mechanical button in terms of its relatively higher cost and the limitation that the distance between the button and the pointer must be close and the pointer must have electrostatic properties. Another contactless type using humidity data [3] was reported. However, due to the use of humidity data, the field of application is more limited than other types. The camera image recognition [4] is also actively reported, but the size of the control device including camera is larger than that of the mechanical type, and the implementation cost is higher than other types. Therefore, it is limitedly used for large screens as a touchless indicator.

For the recognition of the finger-pointing position, a simple two-stage deep learning network was designed and presented in this paper. The deep learning algorithm [5] has been developed since the 1980s and is based on a nested hierarchical structure of the middle layer extended from a simple neural network with one middle layer. Whereas the position of fingertip can be obtained using various linear equations, but it is vulnerable to environmental variance. Also, commercial IR (infrared ray) devices such as Kinect or Leap motion [6] can be used to obtain the position of fingertip but it is expensive and too big comparing with mechanical button device. In the recognition experiment there was no significant difference in recognition between the proposed network and a more complex neural network including convolutional neural network (CNN) filter [7-12]. The pointing recognition accuracy of the proposed device was over 99.0%, the lower limit set by us for product commercialization.

As a virtual 3D display, various glasses-free stereoscopic technologies have been developed and diversified from projection screen type to LCD panel type [13-15]. Among the technologies, the lenticular method [16] was used in this study. It is the most popular and inexpensive method to be implemented. For the proposed method to be an alternative to mechanical buttons, it must be advantageous in terms of cost and miniaturization mentioned above. To recognize the pointing position, a sensor system that recognizes the distribution of light input from a diode was devised. Another problem is a matter of

pressing, which should be checked if the feeling of pressing a button is visible. Hence a physical button could be substituted with the virtual button for which the glasses-free stereoscopic method was selected, whereas the holographic display as another virtual button was excluded in implementation cost and practicality.

3. Implement on Non-contact Button Device

3.1 Autostereoscopic 3D Display Technique with Lenticular Lens

Humans get a 3D information by looking at the same object in different directions with the left and right eyes at the same time. The final information processing result in the brain is one accurate image integrated from the two input images. The left eye image is processed by the right half of the brain, and the right eye image is processed by the left half of the brain.

In the lenticular method as mentioned in Section 2, left and right images for an object were divided by the refraction of a lens plate whose structure is composed of semi-cylindrical columns vertical on a screen. The principle of the lenticular method is shown in Fig. 1. In Fig. 1, the sliced left and right image columns ordered vertically are arranged alternatively under the lens, each image columns refracted are distributed to the left and right eyes, respectively. Since the left and right image columns distributed to the two eyes are combined to a 3D image in the brain, the image can be understood as a 3D object.
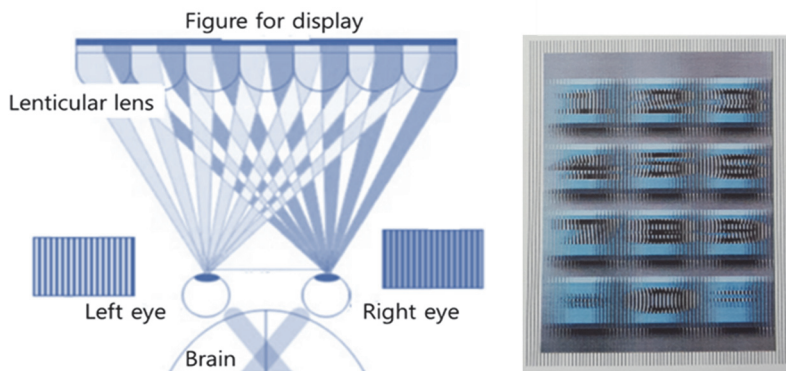

Fig. 1. The principal of lenticular lens and the button display.

For the visual realization of a 3D button using a lenticular lens, a button image was created in which the design of the created button and the background of the button were mapped in 3D for each field-of-view in 20 directions as shown in Fig. 2. That is, after dividing the viewing angle of 180° into 20 fields-of-view, 3D mapping of the button was performed for each of 20 fields. Stereoscopic vision was constructed by refracting this with a lenticular lens to synthesize it in the brain. The depth of the image reproduced virtually by the lens is proportional to the thickness of the lens. The thickness of ours is 4 mm, so that the maximum depth of the reproduced image is 1.8 cm.

To design the virtual button, it is important to consider three layers structure in Fig. 2. The button design is wanted to be displayed at the closest distance from the eye, the picture of the button object should be positioned in front of the eye in 3D coordinates. As shown in Fig. 2, button object picture was

positioned in the higher layer than background layer. The button picture which was reorganized by projecting a button image as a multi-viewpoint is attached under the lenticular lens, and the backlight is placed under each button.

As another virtual 3D button using lenticular lens, an LCD button device was registered in patent (P.N: 10-2019-0038210, Korea). The device recognizes the pointing position using touch panel sensor when a button expressed in the panel was touched. However, it is not an inexpensive device because it requires an LCD touch panel, and furthermore, it cannot solve the problem of non-contact implementation. The touch panel method cannot be commercialized yet as a non-contact stereoscopic button. The method proposed in this paper has the advantage of low manufacturing cost and non-contact use because it does not use LCD panels or expensive devices.
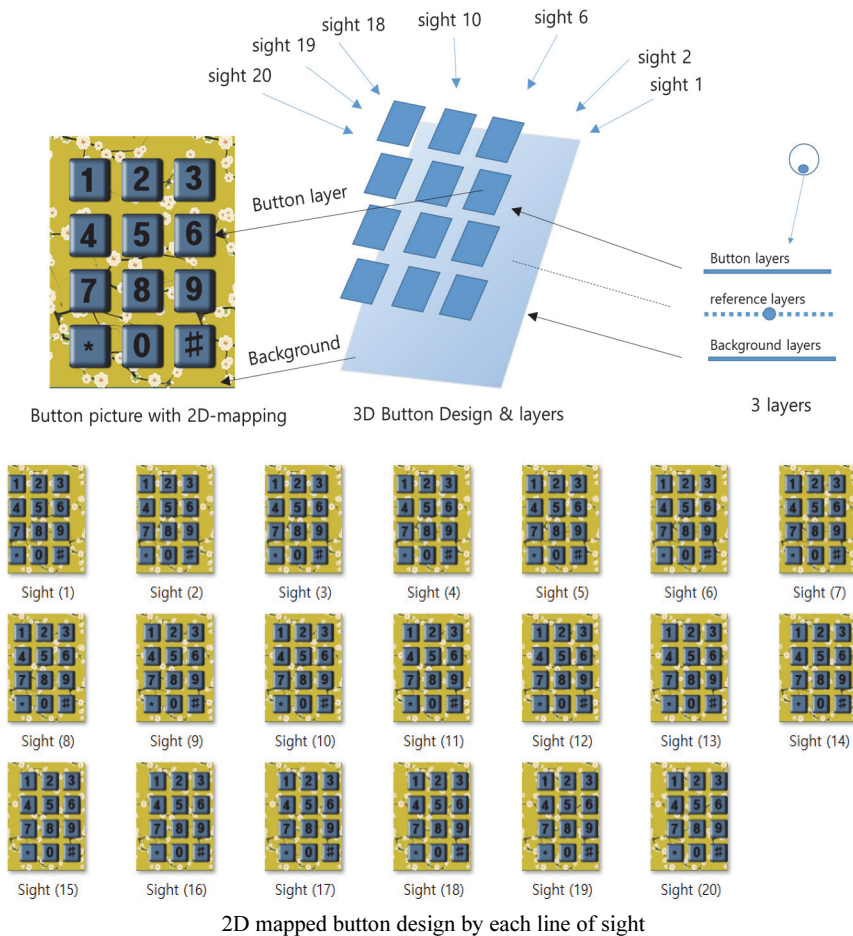

**Fig. 2.** Three-dimensional (3D) virtual button designs for lenticular lens.

3.2 Button Device with Infrared Diodes

The autostereoscopic button device consists of a display unit equipped with a lenticular lens, an infrared sensor unit, a backlight unit, and a driving unit including an artificial intelligence (AI) control system. Fig. 3 shows the construction of the button unit, and Fig. 4 shows the assembly sequence of all parts. As

shown in Fig. 3, infrared diodes for detecting pointing position are arranged at the horizontal and vertical edges of the device frame. The surface frame has IR band pass filters that only pass from near-infrared to far-field, so that the IR receiving sensor can only detect pure IRs from all lights. The IR sensor frame supports the infrared light receiving sensors and light-emitting sensors to detect the pointing position of a fingertip, and the protection film is to protect lenticular lens surface. The infrared sensor part is composed of a pair of sensors with a light-emitting sensor and a light-receiving sensor. And it is arranged in three pairs of top and bottom and four pairs of left and right.
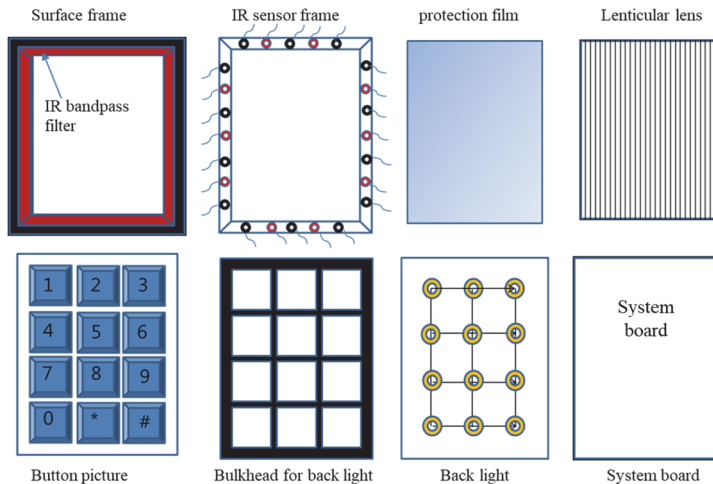
**Fig. 3.** Non-contact 3D virtual button device.
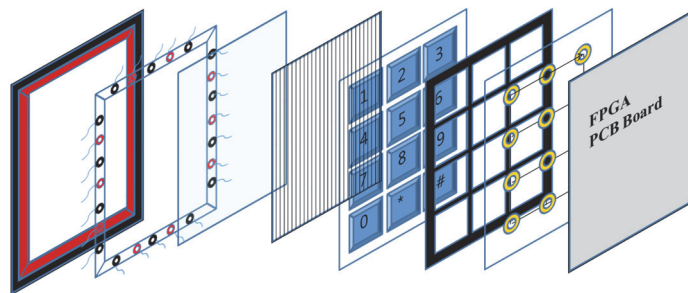
**Fig. 4.** Assembling order for the button device.

The IR emitting diode in this device emits the light including wavelengths from near-infrared to far-field. Fourteen IR emitting diodes are placed on the top, bottom, left, and right of the button unit. The IR receiving diodes also detect IRs from near-infrared to far-field. Whenever the IR emitting diodes arranged along the button frame emit infrared light sequentially in the clockwise direction, all the IR receiving diodes simultaneously detect the amount of light from the IR emitting diode. In spite that the light-emitting is done sequentially, it appears as a simultaneous emitting on the photo in Fig. 5, since it is instantaneous when observed.

The backlight part functions to point the position of the indicated button. Without the backlight, user cannot feel whether a button worked or not because it is a contactless type. Fig. 6 shows the pointing situation and the cross-section of the button set. When a fingertip in Fig. 6 goes through the virtual button

image floated on front of the eyes by the virtual 3D vision, the deep learning program decides the button as pressed. The deep learning works to determine whether a fingertip has pushed the virtual button image and detect the pressed button.
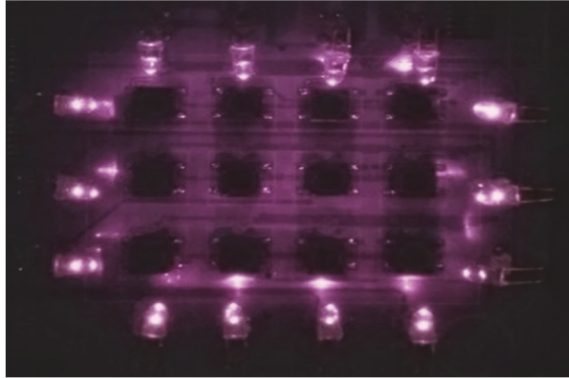
Fig. 5. Emitting from infrared light emitting diodes.
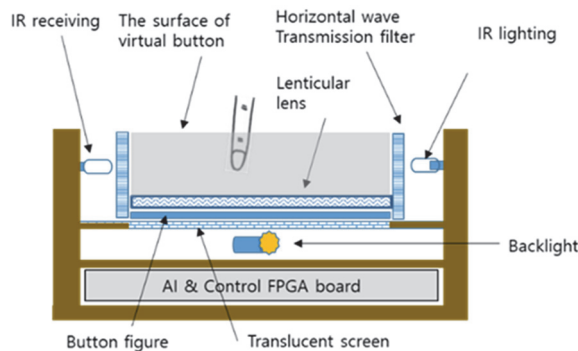
Fig. 6. The button device cross-section.

# 4. Positioning Algorithm of Fingertip Using Deep Learning

4.1 Image Data

To arrange the sensors effectively a graphic simulation was performed as shown in Fig. 7. The light-emitting and the light-receiving sensors were fixed to the device up and down at the same position.

As shown in the simulation image (Fig. 7), the IR receiving diodes occluded by a finger can be clearly recognized. Even if deep learning algorithm is not used, the fingertip position can be determined by setting up some linear equations. Since the linear equations are very weak to variations in environment, deep learning algorithm was selected to be used in spite that its output shows linear pattern. The control board in Fig. 7 turns on a total of 14 IR emitting diodes sequentially and creates an intensity map of 14×14 pixels consisting of 14 gains obtained through the analog-to-digital converter (ADC) from the IR receiving diodes. The intensity map is inputted to the deep learning network in the control board, which outputs the pressed position. Fig. 8 shows an intensity map consisting of the IR receiving sensor values obtained for each IR emitting sensor. When one IR emitting sensor is turned on, 14 IR receiving sensors

on the button set simultaneously detect the infrared light, and transferred the values to the deep learning network on the board. The IR receiving sensors in the shadow occluded by a finger output low values and the other sensors in unblocked area do high values. By this operational principle, the position of the pressed button can be decided. The gray image of the intensity map in Fig. 8 is the image created by the sequential emissions of 14 light-emitting diodes and the gains of 14 receiving diodes. Hence a relatively simple network with two middle layers can be applied in this development because the data structure for analysis is designed to have clear feature.
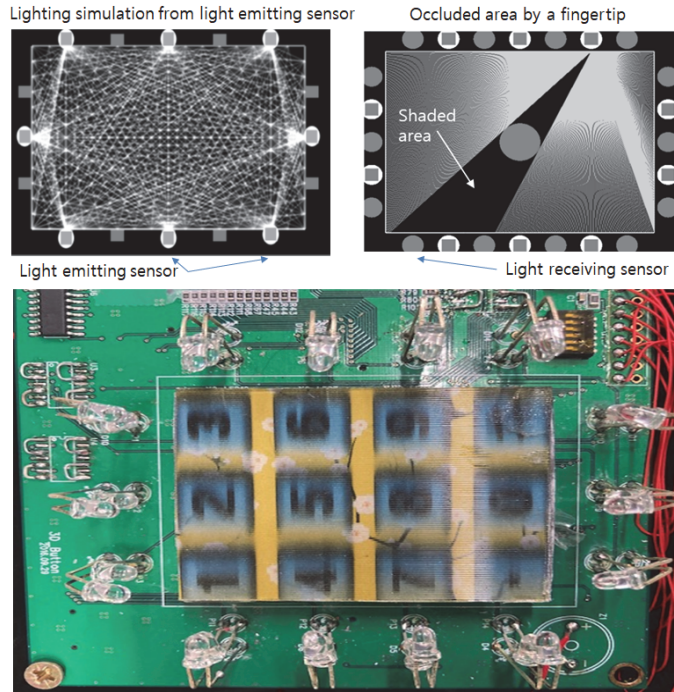
Fig. 7. Simulation for sensor arrangement and the device arrangement.
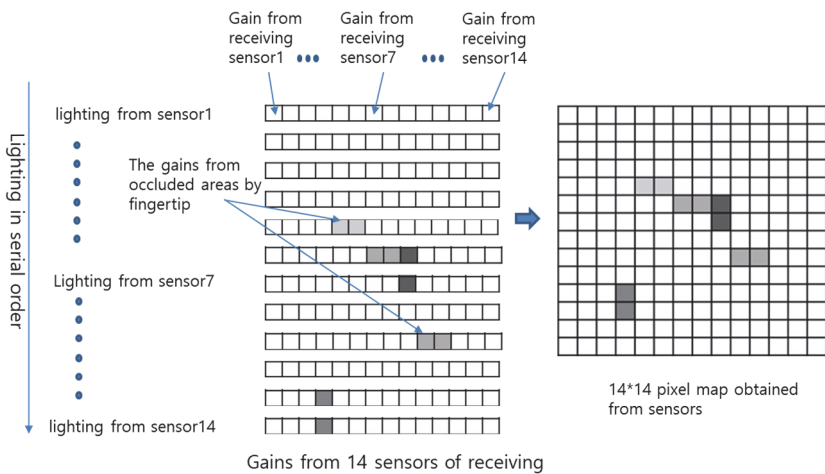
**Fig. 8.** Gains from 14 receiving diodes and intensity map obtained.

4.2 Deep Learning Network for Identifying Pointing Position

Generally, the CNN is a suitable method for detecting objects (persons/cars, etc.) in images, and is highly effective when the object to be detected has correlation between pixels. However, the data received from IR-LED as shown in Fig. 8 do not have correlation unlike the image obtained from the image sensor. Therefore multi-layer neural network (MNN) is more efficient algorithm in terms of this experimental values rather than using CNN. For the reasons mentioned above, this network includes no convolutional filters in layers due to its simple intensity pattern. The IR sensors are connected with neural network engine directly in Fig. 9 and the engine controls the backlights using the output from the engine. The deep learning algorithm used in the virtual 3D button is suggested in Fig. 10. Since the neural network in Fig. 10 receives an image with 14×14 pixels to one pressing of the button, the number of input nodes is 196 as 14×14, and the two middle layers have 100 and 50 nodes, respectively, and the output layer has 12 nodes. The number of input nodes to the neural network is 196 which is the number of 14 diodes around the button device multiplied with its sequential inputs 14 times. The arrangement of these sensors was determined by simulation experiments and the cost of the sensor configuration.

The network in Fig. 10 adopts the sigmoid function as an activation function, but the output layer does not have an activation function (i.e., Bypass). The outputs of the output nodes are converted to probability from 0.0 to 1.0 in softmax using the cross-entropy method. The softmax function is an activation function widely applied in the output layer of a deep learning network for classification of 3-class or higher. The deviation of each value is magnified so that large values are relatively large and small values are relatively small, and then normalized. Here the output layer has 12 nodes because the number of buttons consists of 0 to 9 and 2 special characters. As a cost function for back-propagation, Eq. (1) was used.

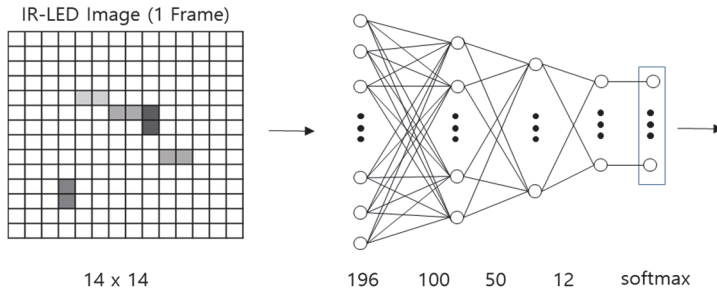$$J(\theta) = -\sum_j y_j log\,(p_j) \tag{1}$$

where θ is the deep learning model to be trained (i.e., weight and bias), and y_j is the j-th element of the correct answer vector, and p_j is the j-th output value of the softmax function. In this device, j has an integer value from 0 to 11 since there are 12 buttons.
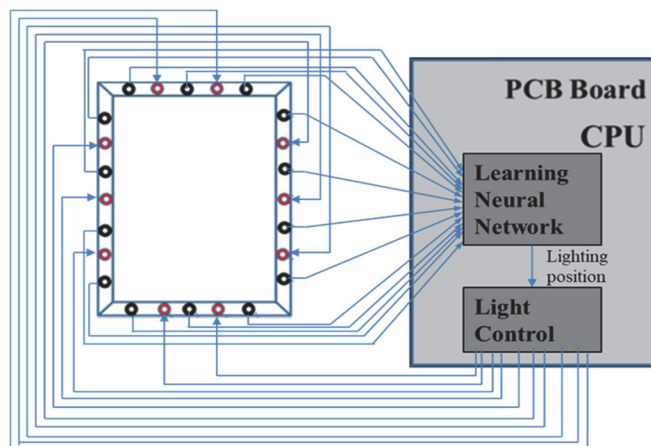
**Fig. 9.** The light control system and neural network on PCB board.

Fig. 10. The proposed network for recognizing fingertip position.

5. Experimental Result

To detect the button position pressed, two stage neural network was used in this development. Fig. 11 shows the two-stage network. For training the network, 200 images were extracted per position by placing a fingertip on the 12 key positions of the virtual button. As a learning experiment, we designed an auxiliary push button set for manual input of correct answers, which is connected to the button device only when learning. When the network is learning, the position pressed on the button device is identified as the correct position entered in the auxiliary set.

Reliability is particularly important for the button, so it must go through two verification processes as shown in Fig. 11. In other words, if the two-recognition results match, the recognized button value may be output. If not, the recognition process must be performed again. There is no problem in using this repetitive step because the time delay is too small to be felt it.
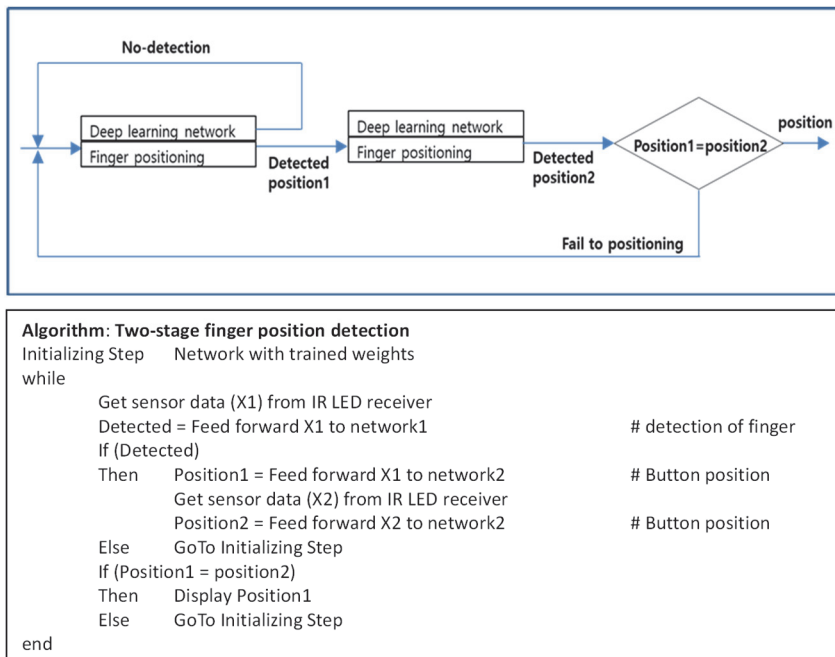
Fig. 11. Two-stage deep learning network to improve reliability.

The button set was trained using 2,400 training data in an office environment, with 800 training iterations in the experiment showing 99.7504% correct answers. To optimize the learning process in time cost, Adam Optimizer [17] was used for accelerating the gradient in back-propagation. Fig. 12 shows the cost according to 600 training iterations, where the cost was computed in Eq. (1).
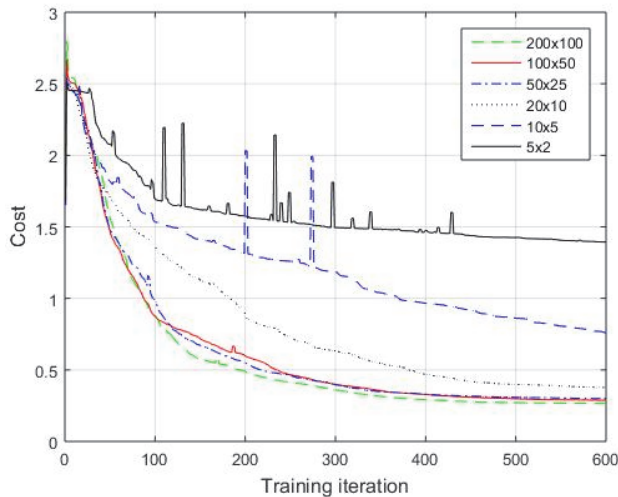
Fig. 12. Cost function for error according to learning iteration by each middle layer structure.

In order to verify the classification efficiency of neural networks, the number of nodes in the middle layers were tested in six configurations: having the 200 and 100 nodes, respectively, the 100 and 50 nodes, the 50 and 25 nodes, the 20 and 10 nodes, the 10 and 5 nodes, the 5 and 2 nodes as shown in Fig. 12. As a result, the node configuration of middle layer shows good result in the structures having the 200 and 100 nodes, the 100 nodes and 50 nodes and the 50 and 25 nodes in terms of accuracy. In the case of the node configurations of 20 and 10, 10 and 5, and 5 and 2, the results were inferior in accuracy. The three configurations show almost same results but in consideration of the load and stability we decided the configurations having 100 and 50 nodes. As shown in Fig. 10, the middle layer of this experiment consists of the two layers having 100 and 50 nodes, respectively, and the output layer consists of 12 nodes. Based on this result, we constructed a neural network for button recognition by selecting the middle layer node configuration of 100 and 50.

In Table 1, the assessment result for the prototype of button was reported. The assessment was performed to four items by the National IT Industry Promotion Agency (NIPA in Korea) which is a national institute that specializes in testing prototypes and new technologies. The response speed of the button prototype was recorded 0.075 seconds per one pressing. The respond time was estimated by an oscilloscope, which could measure the interval between the rising edge and falling edge in signal. Another important function that buttons should have is how to judge when the user presses the middle boundary between buttons. In this button, when the middle boundary is pressed, the ratio of fingertips is judged to determine that the button with the higher occupancy ratio is pressed. In the case of half-and-half share, it was not judged. This function was evaluated by visual judgment and the result was pass. The limitation of miniaturization of buttons is a commercially important factor. The goal in this development was to minimize the size of one button to 5 mm in width and height, and a functional test was performed with a

prototype that was miniaturized. The miniaturization result was pass. Finally, the accuracy of the button's response is the most important evaluation factor.

Table 1. Button device performance evaluation result[a]

Evaluation item (main performance spec)	Results	Assessment methods
Button response speed (s)	0.075	Measure response speed after pressing
Determination of button press position (when pressing the border between buttons)	Pass	Whether the overlaid button can be recognized when the finger is positioned at the button boundary
Button miniaturization	Pass	Confirm the operation of each 5 mm square button horizontally and vertically
Button pointing accuracy (%)	99	Confirm button pointing accuracy (success rate when pointing the same button 100 times)

[a] This assessment was performed by the National IT Industry Promotion Agency (NIPA) (Accreditation No. S161109-009).

In the evaluation, the same button was pressed 100 times to evaluate the coincidence with the response. The result showed 99% identified result and was judged as a success. The reason the result is not 100% but 99% is that the decision of the result is ambiguous because there are many cases where the border between buttons is pressed. When the button's center was pressed, the correct response always came out. For implementation of the button, Xylinx FPGA (Zynq7000) was used, which has built-in CPU (ARM core) and gate array to design logic. Since the network structure in this development is so simple that the convolution filters in the network were removed, the built-in CPU only was used instead of the gate array for acceleration. The CPU in FPGA has built-in Dual ARM9 Cores, but only one was used in this study. The operating frequency of the CPU is 666 MHz and two DDR3 used as memory. The operating frequency is 533 MHz.

6. Conclusion

This research presented an implementation of autostereoscopic virtual 3D button device as non-contact style. The proposed device has several characteristics about visible feature, non-contact use and AI engine. In visible feature, the device was designed to be used as non-contact and as autostereoscopic virtual 3D display. To specify the button pressed virtually by fingertip pointing, a simple deep learning network having two stages without convolution filters was designed. The deep learning network does not have to be so complex if the composition of the input data has a clear design, as demonstrated in the experiments. As the results of testing and evaluation by the certification institute, the proposed button shows high reliability and stability. As the results for the assessment, the response speed of the button prototype was recorded 0.075 seconds per one pressing. As a stability of the device, when the middle boundary between buttons is pressed, the ratio of fingertips is used to determine that the button with the higher occupancy ratio is pressed. The size of the button can be made to the minimum size of one button to 5 mm in width and height. With accuracy, the button's response shows more than 99% match results. The reason why the accuracy is not 100% is that the 99% includes the ambiguous cases in which the border between buttons is pressed. The supply and sales of the device are expected to gradually expand its scope according to the recent hygiene trend and to the low manufacturing cost of simple hardware

configuration. In the future, this button set will be released through an additional learning process for commercialization. As a further study, we are planning to increase the number of button keys keeping the number of sensors arranged in the device to a minimum.

References

[1] S. H. You, M. Hwang, K. H. Kim, and C. S. Cho, "Development of a non-contact autostereoscopic 3D button using artificial intelligence," in *Advanced Multimedia and Ubiquitous Engineering*. Singapore: Springer, 2021, pp. 69-76.

[2] Y. Oishi and Y. Yamashiro, "Control circuit of electrostatic capacitive sensor and electronic device using the same," U.S. Patent 9838527, Dec 5, 2017.

[3] S. Mondal, S. J. Kim, and C. G. Choi, "Honeycomb-like MoS2 nanotube array-based wearable sensors for noninvasive detection of human skin moisture," *ACS Applied Materials & Interfaces*, vol. 12, no. 14, pp. 17029-17038, 2020.

[4] N. Zengeler, T. Kopinski, and U. Handmann, "Hand gesture recognition in automotive human–machine interaction using depth cameras," *Sensors*, vol. 19, no. 1, article no. 59, 2019. https://doi.org/10.3390/s19010059

[5] Y. S. Jeong and J. H. Park, "Advanced big data analysis, artificial intelligence & communication systems," *Journal of Information Processing Systems*, vol. 15, no. 1, pp. 1-6, 2019.

[6] S. Park, S. Cho, J. Park, K. Huang, Y. Sung, and K. Cho, "Infrared bundle adjusting and clustering method for head-mounted display and Leap Motion calibration," *Human-centric Computing and Information Sciences*, vol. 9, article no. 8, 2019. https://doi.org/10.1186/s13673-019-0169-6

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097-1105, 2012.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, San Diego, CA, 2015.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 2016, pp. 770-778.

[10] M. T. N. Truong and S. Kim, "A tracking-by-detection system for pedestrian tracking using deep learning technique and color information," *Journal of Information Processing Systems*, vol. 15, no. 4, pp. 1017-1028, 2019.

[11] W. Song, L. Zhang, Y. Tian, S. Fong, J. Liu, and A. Gozho, "CNN-based 3D object classification using Hough space of LiDAR point clouds," *Human-centric Computing and Information Sciences*, vol. 10, article no. 19, 2020. https://doi.org/10.1186/s13673-020-00228-8

[12] D. Cao, Z. Chen, and L. Gao, "An improved object detection algorithm based on multi-scaled and deformable convolutional neural networks," *Human-centric Computing and Information Sciences*, vol. 10, article no. 14, 2020. https://doi.org/10.1186/s13673-020-00219-9

[13] N. A. Dodgson, "Autostereoscopic 3D displays," *Computer*, vol. 38, no. 8, pp. 31-36, 2005.

[14] C. W. Tyler and M. B. Clarke, "Autostereogram," in *Proceedings of SPIE 1256: Stereoscopic Displays and Applications*. Bellingham, WA: International Society for Optics and Photonics, 1990, pp. 182-197.

[15] D. Ezra, G. J. Woodgate, B. A. Omar, N. S. Holliman, J. Harrold, and L. S. Shapiro, "New autostereoscopic display system," in Proceedings of SPIE 2409: *Stereoscopic Displays and Virtual Reality Systems II*. Bellingham, WA: International Society for Optics and Photonics, 1995, pp. 31-40.

[16] P. Bourke, "Autostereoscopic lenticular images," 2014 [Online]. Available: http://paulbourke.net/stereo graphics/lenticular/.

[17] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, San Diego, CA, 2015.

Sang-Hee You https://orcid.org/0000-0001-5660-838X

He received M.E. degree in Sungkyunkwan University in 2009. Since March 2019, he is with the Department of Information & Telecommunication from Hanshin University as a Ph.D. candidate. His current research interests include AI & image processing. He is working in IVSYS Company as CEO.

Min Hwang https://orcid.org/0000-0001-9892-8453

He received M.E. degrees in Sungkyunkwan University in 2012. Since March 2019, he is with the Department of Information & Telecommunication from Hanshin University as a Ph.D. candidate. His current research interests include AI & Embedded system. He is working in IVSYS Company.

Ki-Hoon Kim https://orcid.org/0000-0002-5296-8393

He received M.E. degrees in Sungkyunkwan University in 2008. His current research interests include AI & mechatronics. He is working in IVSYS Company as a 3D graphic algorithm Engineer.

Chang-Suk Cho https://orcid.org/0000-0002-0929-3028

He received Ph.D. degrees in Keio University of Japan in 1995. From 1985 to 1995, he walked as a researcher in ETRI of Korea. His current research interests include AI & Virtual Reality. He is working in Division of Information & Telecommunication of Hanshin University as a professor.