Sensors and Materials, Vol. 34, No. 1 (2022) 203–216
MYU Tokyo

S & M 2804

RGB-D Depth-sensor-based Hand Gesture Recognition Using Deep Learning of Depth Images with Shadow Effect Removal for Smart Gesture Communication

Ing-Jr Ding* and Nai-Wei Zheng

Department of Electrical Engineering, National Formosa University, Huwei Township, Yunlin 632, Taiwan, ROC

(Received May 26, 2021; accepted November 25, 2021)

Keywords: RGB-D depth sensor, shadow effect, serial binary image extraction, deep learning, hand gesture recognition

Recently, compound image sensor devices have been widely used to construct many next-generation human–machine interaction applications, including hand gesture action recognition. Such devices, generally known as RGB-D devices, contain an RGB color camera and a depth sensor set. The depth sensor in RGB-D devices is essentially a set of a specific type of sensor and comprises one IR projector and one IR camera. This structure of the depth sensor inevitably generates an undesired shadow effect, adversely affecting hand gesture recognition. To tackle this issue and alleviate the shadow effect on hand gesture recognition, we have developed a serial binary image extraction approach. The proposed approach is essentially composed of two consecutive computation phases, phase-1 and phase-2 binary image extraction. In this work, the Kinect compound sensor device is employed to capture hand gesture depth images. The deep learning model, a visual geometry group (VGG)-type convolutional neural network (CNN), i.e., the well-known VGG-CNN, is utilized to evaluate the recognition effectiveness of improved hand gesture depth images derived from serial binary image extraction. Ten hand gestures that are common in daily life are chosen to evaluate depth-sensor-based interactive action recognition. Experimental results show that the proposed serial binary image extraction can effectively eliminate the undesired shadow region in hand gesture depth images and significantly improve the recognition accuracy of VGG-type CNN hand gesture recognition. The proposed depth-sensor-based hand gesture recognition approach can benefit people requiring interaction action recognition and further promote smart gesture communication.

1. Introduction

The compound image sensors of RGB-D devices have promoted important developments in image processing techniques and creative human–machine interaction (HMI)[1–3] and 3D sensing recognition applications.[4–9] Well-known RGB-D sensor devices, such as Microsoft Kinect,[10] Leap Motion Control (LMC),[11] and ASUS Xtion,[12] contain two different modalities of sensors for image capture: RGB color and depth image sensors. Generally, the RGB sensor

*Corresponding author: e-mail: eugen.ding@gmail.com
https://doi.org/10.18494/SAM3557

ISSN 0914-4935
https://myukk.org/

cannot acquire standard images under low illumination. In contrast, the depth sensor has strong light tolerance and therefore its performance is satisfactory even in a dark environment. Therefore, depth-sensor-derived images have been employed to develop integrated applications or specific systems in most studies.[13–24]

Among the depth-sensor-based works, such as depth-based scene map construction for robot operations[13–16] and depth-based human body gesture image acquisitions for human body analysis,[17–19] little attention has been paid to the property of depth-grayscale images for recognition due to the relatively long sensor capture distance. However, for studies in which a depth sensor is used to develop a hand gesture recognition system,[20–24] the depth-sensor-based hand gesture recognition system will be extremely sensitive to the characteristics of depth-grayscale hand gesture images captured by the sensor. This is because a very short distance between the active hand and the depth sensor is required to obtain the complete hand gesture information. Such short-distance hand gesture image acquisition by the depth sensor will unavoidably cause a serious shadow effect. The shadow effect is not expected to appear in recognition systems (especially in deep-learning-based recognition systems) and adversely affects recognition accuracy, as described in detail in Sect. 2. Most depth-sensor-based hand gesture recognition systems have been targeted at developing improved deep learning models for deep neural networks (DNNs) and integrating intelligent systems in specific application fields.[20–24] An improved regression network for depth hand pose estimation was proposed by Xu *et al.*[20] Lai and Yanushkevich[21] developed a dynamic hand gesture recognition system where two different deep learning schemes based on a convolutional neural network (CNN) and a recurrent neural network are adopted to analyze both the skeleton information and the spatial information from depth images. Otberdout *et al.* proposed a hand pose estimation approach based on a deep-learning depth map for hand gesture recognition.[22] A CNN and a generative adversarial network were used for depth-sensor-based hand gesture recognition and hand pose estimation, respectively.[23,24]

However, the undesired phenomenon of the shadow effect caused by depth sensors deployed in an RGB-D device has scarcely been taken into consideration in current research on depth-sensor-based hand gesture recognition. To overcome the problem of the shadow effect resulting from the RGB-D depth sensor and further evaluate the effectiveness of depth sensor data without the shadow effect on the recognition performance of a DNN, we have developed an RGB-D depth-sensor-based interactive hand gesture recognition system using a typical visual geometry group (VGG)-type CNN with improved data of shadow effect removal derived from serial binary image extraction. The proposed depth-sensor-based hand gesture recognition with the improved depth gesture image with removed shadow regions can ensure more reliable action recognition results, and such hand gesture recognition with satisfactory performance can then be used in real-life applications that require interactive gesture recognition for interaction, communication, or operations (e.g., gesture interaction between disabled and able-bodied people, the communication of actions between miners underground in low light, and hand gesture control in a smart factory/car/home). In this work, the Kinect sensor device illustrated in Fig. 1 is employed to capture depth-grayscale hand gesture actions. As mentioned, the Kinect device, which is a compound image sensor, has separate RGB and depth sensors. These two

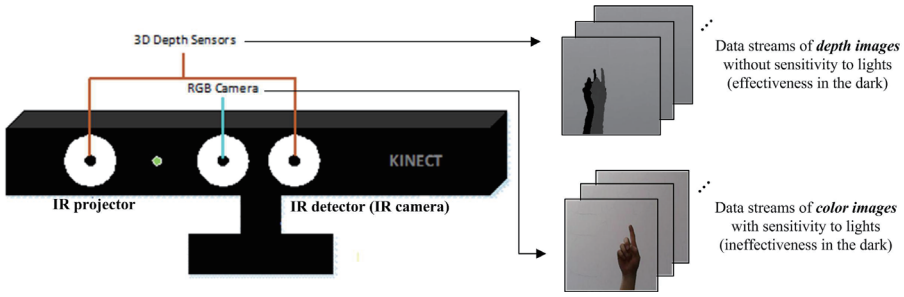


Fig. 1. (Color online) Popular Kinect RGB-D image sensor device with set of depth sensors comprising IR projector and IR camera.

image sensors, which are collectively referred to as a CMOS image sensor, and the time of flight can be used to capture color-RGB and depth-grayscale images simultaneously. Another feature of Kinect image acquisition is that either the RGB sensor or depth sensor is excited and can operate alone. In this work on RGB-D depth-sensor-based hand gesture recognition, only the Kinect depth sensor channel is operated. As can be seen in Fig. 1, the depth sensors in Kinect essentially include one IR projector and one IR camera, resulting in an adverse shadow effect, as described in detail in Sect. 2.

The main contributions of this study are summarized as follows:

(1) Removal of undesired shadow effect of hand gesture depth images from the RGB-D depth sensor set comprising one IR projector and one IR camera by the proposed serial binary image extraction.
(2) Effectiveness demonstrations of improved hand gesture depth images without shadow effect on recognition of CNN deep learning.
(3) Development of RGB-D depth-sensor-based hand gesture recognition by VGG-type CNN incorporated with serial binary image extraction for smart gesture communication.

2. Undesired Shadow Effect of Typical Depth Sensors in Gesture Recognition

A typical depth sensor uses both an IR projector and an IR camera to capture the depth image of a target. As mentioned before, popular RGB-D sensor devices generally have RGB and depth sensors. The depth sensor in the RGB-D sensor device is composed of one IR projector and one IR camera. During image capture or recording, both the IR projector and the IR camera are simultaneously used to capture the depth image of the target. Figure 2 depicts the projection and reception of Kinect depth-sensor-derived IR light beams. As can be seen in Fig. 2, the captured depth image representing the target object has a shadow region with black pixels. This is the so-called shadow effect of the depth sensor, and the shadow region that appears in the captured depth image mainly consists of black pixels. According to the analysis of Danciu *et al.*,[25] there are two causes of the shadow region: (1) unexpected object occlusion and (2) optical refraction and reflection. The rationale behind the first cause is that an unexpected object can act as an obstacle to the target object located behind it. Regarding the second cause, reflection or refraction will occur when the IR light emitted by the IR projector reaches the target object. In this case, depending on the material and the degree of light absorbance of the target object,

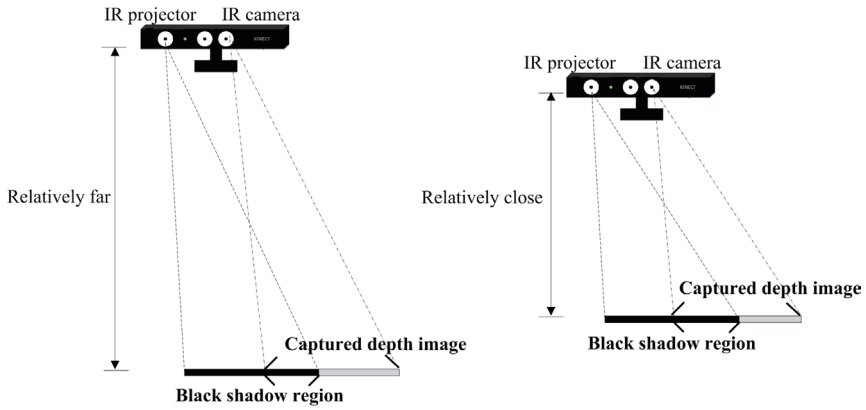


Fig. 2. Illustrations of the undesired shadow effect of the depth sensor set (greater effect at a shorter distance).

different reflection or refraction phenomena will occur. Owing to the design limitations of such depth sensor devices, when constructing a depth sensor image-based application system, black shadow regions will inevitably exist in captured depth images. The effect of the black shadow region on the captured image will significantly depend on the relative distance between the depth sensor device and the desired target object. When the distance is relatively large (such as human body or skeleton detection for human gesture recognition applications), the effect of the black shadow region will be small. However, when a relatively small distance is required for fine image capture (e.g., recognition of hand gesture communication actions carried out in this study), the effect of the unwanted black region in the overall captured depth image will greatly increase. Figure 3 illustrates the unwanted black shadow segment located around a real hand segment with depth-grayscale information. It can be clearly seen in Fig. 3 that the black hand shadow region has a large effect on the intended depth-grayscale hand region, which will be a serious problem in hand gesture recognition with high recognition accuracy.

For depth-sensor-based applications, the above-mentioned shadow effect adversely affects the system performance and is undesired. In this work, the Kinect depth sensor is used to recognize hand gesture communication actions, and a large black shadow region in the captured hand gesture depth image will clearly reduce the recognition accuracy of the established system. To alleviate the shadow effect, a serial binary image extraction approach is proposed to obtain improved hand gesture depth data for VGG-type CNN deep learning and recognition, as described in detail in Sect. 3.

3. Deep Learning of Depth Images with Shadow Effect Removal for RGB-D Depth-sensor-based Interactive Hand Gesture Recognition

Figure 4 shows the overall framework of the constructed hand gesture communication action recognition system using the Kinect sensing device with depth sensors comprising one IR projector and one IR camera. Using the developed recognition system, 10 hand gesture communication actions that are commonly used for interactions can be categorized. When performing recognition on a dynamic hand gesture action, continuous-time hand gesture images (frames) are obtained from the Kinect depth sensor. As shown in Fig. 4, for the acquired depth

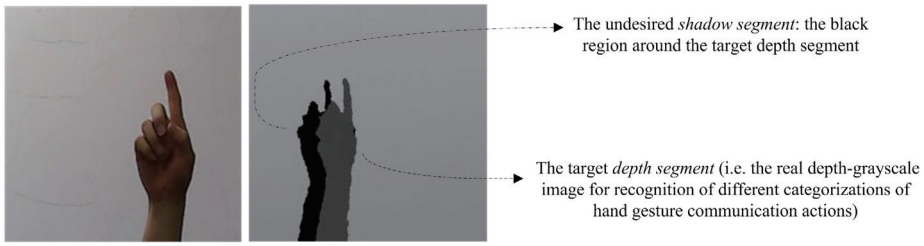


Fig. 3. (Color online) RGB and depth images with the undesired shadow region obtained with the Kinect RGB-D device (interactive hand gesture "what?").

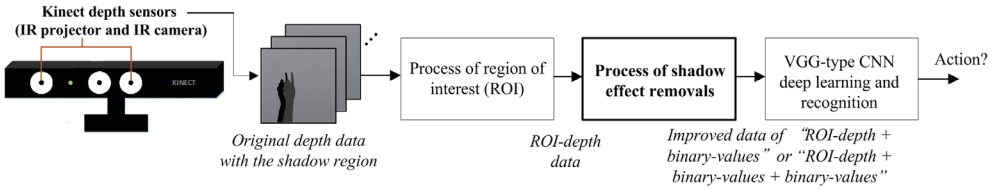


Fig. 4. (Color online) Framework of depth-sensor-based interactive hand gesture recognition with deep learning of depth images and shadow effect removal.

sensor images, the region of interest (ROI) is first extracted to derive the significant image segment of active hand gesture regions, referred to as the *ROI-depth* image. As mentioned in the previous section, such depth-sensor-derived images are not ideal at this stage and have a relatively large shadow region. The proposed serial binary image extraction scheme is effectively applied to the *ROI-depth* image to eliminate most of the shadow region and alleviate the shadow effect. After removing the shadow from the *ROI-depth* image, the improved depth sensing data, referred to as *ROI-depth* + *binary-values* or *ROI-depth* + *binary-values* + *binary-values*, will further be used for the training and recognition test of the VGG-type CNN deep learning model, as described in Sects. 3.1 and 3.2.

3.1 Shadow effect removal of depth images using serial binary image extraction

Each pixel in the *ROI-depth* image has a grayscale value between 0 and 255, which indicates the degree of closeness between the depth sensor and the point. Assuming that the size of the *ROI-depth* image is n (width) by m (height), there will be $n \cdot m$ depth degree values determined directly from the Kinect software development kit (Kinect-SDK). Each of these $n \cdot m$ values represents the depth-grayscale value of the corresponding pixel (x, y), denoted as *Depth degree value*(x, y). Equation (1) shows the calculation of *Depth degree value*(x, y). Note that the value of *Threshold* is generally set as 4000, corresponding to the maximum sensing distance of the Kinect depth sensor of 4000 mm (i.e., 4 m). When the distance of the object is greater than 4 m, the value of all pixels in this object will be 255 ('white'). Conversely, if the object is much closer to the depth sensor (i.e., an extremely small value of *Distance*), then *Depth degree value*(x, y). will be close to 0 ('black'). The depth-grayscale degree varies linearly in the range of Threshold.

$$Depth\ degree\ value(x,y)=\begin{cases} Distance\times\left(\dfrac{255}{Threshold}\right), & \text{if } Distance\le Threshold \\ 255, & \text{if } Distance > Threshold \end{cases} \quad (1)$$

By using the estimated value of each *Depth degree value*(x, y) of the *ROI-depth* image, the proposed serial binary image extraction method removes the undesired shadow region in the *ROI-depth* image. The proposed process of serial binary image extraction for shadow effect removal is composed of two stages of binary value estimation, called phase-1 and phase-2. Phase-1 binary image extraction is performed using the following equation:

$$Binary\ value(x,y)_{phase-1}=\begin{cases} 0, & \text{if } Hand\ gray\ level < Depth\ degree\ value(x,y) < 255 \\ Depth\ degree\ value(x,y), & \text{otherwise} \end{cases} \quad (2)$$

In the phase-1 extraction, the data of the *ROI-depth* image with each *Depth degree value*(x, y) determined by Eq. (1) are further converted to binary-valued information. As can be seen in Eq. (2), the grayscale-degree value of each pixel in the *ROI-depth* image is transformed to an estimated binary value. *Hand gray level* in Eq. (2) represents the depth-grayscale value of the hand region of the *ROI-depth* image, and its value is computed using the algorithm in Fig. 5 expressed in pseudo-code. In Eq. (2), the value of 255 represents a white pixel, i.e., no degree of

```
Initialize:
I = a depth sensor-derived hand gesture image with a size of n by m;
Set each of Gray-0, Gray-1,..., and Gray-255 to be zero;
for x = 1 to n do
    for y = 1 to m do
        Vote the I[x][y] to one of Gray-0, Gray-1, ..., and Gray-255 according to the value of I[x][y];
    end for
end for
//Check effectiveness of each of Gray-0, Gray-1, ..., and Gray-255
for i = 0 to 255 do
    if the value of Gray-i is larger than the predefined pixel number then
        Set Gray-i as a valid and effective vote box, Effective Gray-i;
    end if
end for
Discard the vote box Effective Gray-0 (denoting 'black' numbers) and find vote box of hand region:
Hand_region_vote_box = minimum of labels (grayscale values) of all remaining effective vote boxes;
Hand_gray_level = the label (the grayscale value) of the vote box, Hand_region_vote_box;
return Hand_gray_level;
```

Fig. 5. Algorithm to determine *Hand gray level* (grayscale value of hand regions) in shadow effect removal of depth images by serial binary image extraction.

grayscale appears in this pixel. The proposed algorithm for determining *Hand gray level* is conceptually simple and computationally fast. In the algorithm, the input is the *ROI-depth* image, and the main purpose of this algorithm is to find significant regions including the hand region, each of which has numerous pixels with the same depth-grayscale value. The depth-grayscale value of the estimated hand region is finally returned. As can be seen in Fig. 5, the values of *Gray*-0, *Gray*-1, …, *Gray*-255 are set to zero in the initialization step. *Gray*-0, *Gray*-1, …, *Gray*-255 denote the vote boxes that accumulate votes of pixel grayscale values of the input *ROI-depth* image. When serial binary image extraction is performed on an *ROI-depth* image with a size of n by m, a total of $n \cdot m$ pixels are required to vote among *Gray*-0, *Gray*-1, …, *Gray*-255. Note that the pixel grayscale values are in the range of 0 to 255. If an image pixel has a grayscale value of i, the vote box of *Gray*-i is then increased by one. Also note that in Eq. (2), if the value of *Depth degree value*(x, y) is located between *Hand gray level* and 255, this pixel is directly set to 0. The operations in Eq. (2) transform the *ROI-depth* image to a binary-valued image. In the transformed image, the hand region retains the same depth-grayscale value, and the pixels in the other regions of the image are black. The black shadow region is therefore not distinguishable and is discarded (see Fig. 6).

Equation (3) describes phase-2 binary image extraction. The main operation of phase-2 binary image extraction is to further convert the output binary image derived from phase-1 binary image extraction to different categorizations of binary images. As can be seen in Fig. 5, by setting *Gray degree* = 255 in Eq. (3), all pixels in the hand region become white. The transformed image is a binary-valued black-white image. Note that *Gray degree* in Eq. (3) is variable and adjustable. To achieve the optimal recognition performance of the recognition system with the depth sensor, using a trial-and-error approach, a grayscale value between *Hand gray level* and 255 (white) is appropriately chosen and given to *Gray degree*. Images with various settings of *Gray degree* are illustrated in Fig. 7.

$$Binary\,value(x,y)_{phase-2} = \begin{cases} Gray\ degree, & \text{if } Binary\,value(x,y)_{phase-1} \neq 0 \\ Binary\,value(x,y)_{phase-1}, & \text{otherwise} \end{cases} \quad (3)$$

As a result of the proposed serial binary image extraction, the *ROI-depth* data with the shadow effect removed can be obtained, which are then used in the training and recognition of

Fig. 6. Shadow effect removal of depth images by serial binary image extraction (from left to right, original depth with shadow regions and improved depths after phase-1 and phase-2 binary image extraction).

Fig. 7. Depth-grayscale values of 80, 140, 170, and 255 (from left to right) set in the hand region in the phase-2 binary image extraction approach.

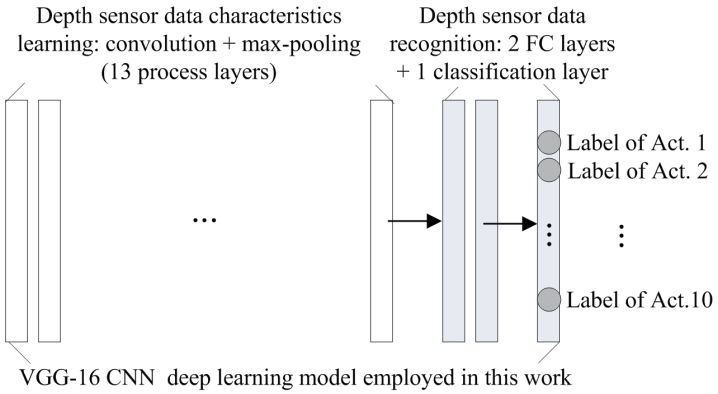


Fig. 8. Typical VGG-16 CNN used in depth-sensor-based hand gesture recognition for deep learning of depth images with shadow effect removal.

the deep learning VGG-type CNN model. For simplicity, the improved image data after phase-1 and phase-2 binary image extraction calculations are referred to as *ROI-depth* + *binary-values* and *ROI-depth* + *binary-values* + *binary-values*, respectively.

3.2 Deep learning and recognition of interactive hand gestures of depth images without shadow effect

We employ the VGG-16 CNN[26] to construct a system for hand gesture communication action recognition using the depth sensor. As mentioned, the improved data of the depth-sensor-derived image without the shadow effect, *ROI-depth* + *binary-values* or *ROI-depth* + *binary-values* + *binary-values*, are used to perform training and recognition in the VGG-16 CNN. As shown in Fig. 8, the deep learning model of the VGG-16 CNN is composed of 16 processing layers: 13 layers for the deep learning and extraction of color characteristics of input image data and three layers for the classification of the extracted image feature. Note that these 13 layers for image feature estimation mainly perform a series of computations of convolutions and max-pooling. The final three layers in the VGG-16 CNN, which act as a general artificial neural network, are composed of two fully connected (FC) layers and one classification layer.

For typical VGG-16 CNN deep learning calculations, the size of the input image is fixed at 224 by 224. As can be seen in Fig. 8, in this work, the final layer of the VGG-16 CNN is designed

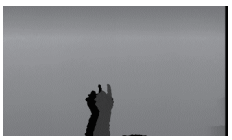
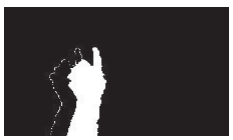
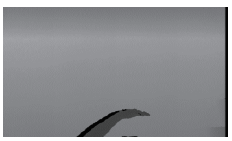
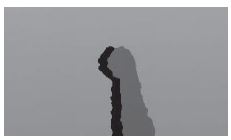
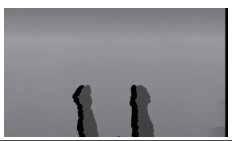
to have 10 calculation nodes, each of which denotes the categorization scores of the corresponding type of hand gesture communication actions (as mentioned above, 10 different classes of common hand gesture actions for interactions in daily life are considered in this study).

4. Experiments and Results

Experiments on depth-sensor-based interactive hand gesture recognition are carried out in a laboratory environment. The test user is requested to make 10 specific hand gestures that are widely used for interaction. These 10 different actions are labeled as *Action*-1, *Action*-2,… and *Action*-10. Table 1 shows each of these 10 hand gesture actions. For briefness and readability of the paper, only the first frame (image) of the specific action is shown as the representative in each of these 10 continuous-time gesture actions in this work, which can be clearly seen in Table 1. When performing VGG-16 CNN recognition with improved data of depth-sensor-derived images, all images contained in each of these 10 continuous-time actions will still be considered. A hand gesture action database established in the laboratory with the Kinect sensor device contains 3000 images, half of which are used for training of the VGG-16 CNN, and the other half of which are used for testing the constructed VGG-16 CNN model. For the experimental database of 3000 images, the test user is requested to make 50 actions for each of these 10 different categorizations of hand gesture actions. Each action contains 60 frames (2 s), in which the frame rate of the Kinect depth sensor is 30 (i.e., 30 depth sensing frames are captured by Kinect per second).

As shown in Table 1, for each of these 10 different types of hand gesture actions, the original depth image obtained from Kinect, the *ROI-depth* image, the improved data of *ROI-depth* + *binary-values*, and the improved data of *ROI-depth* + *binary-values* + *binary-values* are provided. It can be observed that, by using the proposed serial binary image extraction approach, the undesired black shadow region existing in *ROI-depth* is greatly reduced in both *ROI-depth* + *binary-values* and *ROI-depth* + *binary*-values + *binary-values*. Figures 9 and 10 depict *ROI-depth* + *binary-values* and *ROI-depth* + *binary-values* + *binary-values* frame sequences of a specific hand gesture categorization, *Action*-3 (denoting "Goodbye"), for training and recognition of the VGG-type CNN, respectively. Table 2 gives the recognition accuracy of VGG-type CNN hand gesture communication action recognition using depth sensing data with and without shadow effect removal. Three recognition performance results are compared, training accuracy, validation accuracy, and test accuracy, in the phases of model training, model validation, and model testing, respectively, with images of *ROI-depth*, *ROI-depth* + *binary-values* and *ROI-depth* + *binary-values* + *binary-values*. It can be clearly seen that the proposed serial binary image extraction approach has a positive effect on the recognition performance. Without the removal of black shadow regions, the average accuracy of VGG-type CNN recognition with *ROI-depth* is only 72.95%. In contrast, the improved data *ROI-depth* + *binary-values* derived from phase-1 binary image extraction and the improved data *ROI-depth* + *binary-values* + *binary-values* obtained from phase-2 binary image extraction give better performances for VGG-type CNN recognition of 78.54 and 78.01%, respectively. Note that in the phase-2 binary

Table 1
Ten common hand gesture actions with various image types for deep learning and recognition.

Interaction actions	Original RGB-D depth	*ROI-depth*	*ROI-depth + binary-values*	*ROI-depth + binary-values + binary-values*
Action-1 (1st frame)				
Action-2 (1st frame)				
Action-3 (1st frame)				
Action-4 (1st frame)				
Action-5 (1st frame)				
Action-6 (1st frame)				
Action-7 (1st frame)				
Action-8 (1st frame)				
Action-9 (1st frame)				
Action-10 (1st frame)				

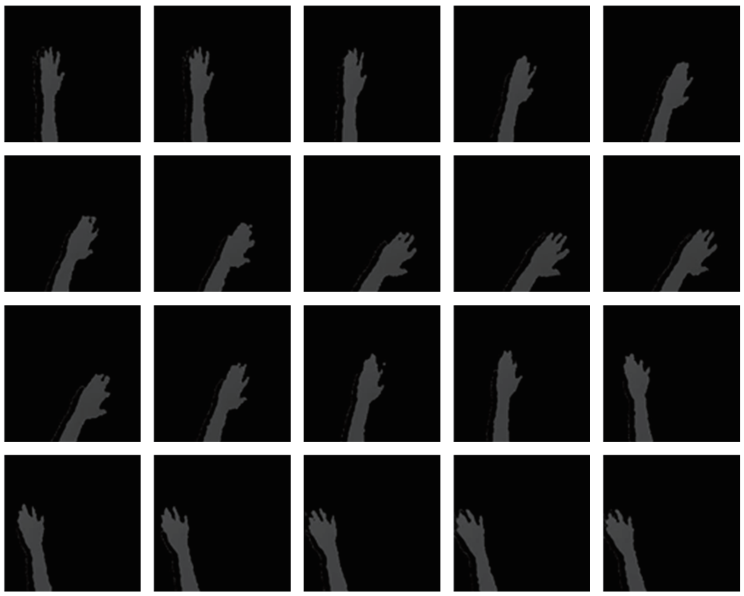


Fig. 9. Continuous-time data streams of interaction hand gestures (Action-3, "Goodbye") with improved data of *ROI-depth* + *binary-values* for deep learning and recognition.

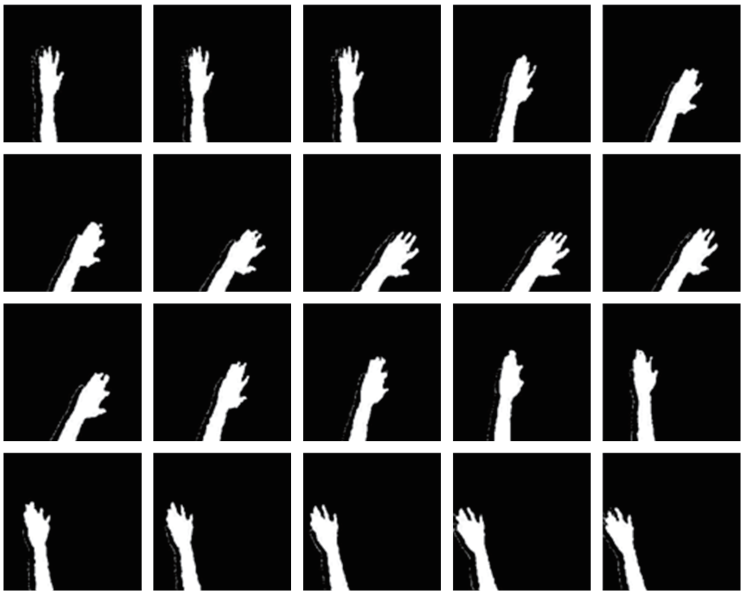


Fig. 10. Continuous-time data streams of interaction hand gestures (Action-3, "Goodbye") with improved data of *ROI-depth* + *binary-values* + *binary-values* for deep learning and recognition.

Table 2
Recognition accuracy comparisons of VGG-type CNN hand gesture recognition using depth images with and without shadow effect removal.

Data categorization	Data with shadows	Data with shadow effect removal	
Data for VGG-type CNN	ROI-depth (%)	*ROI-depth* + *binary-values* (%)	*ROI-depth* + *binary-values* + *binary-values* (%)
Training accuracy	100	100	100
Validation accuracy	100	100	100
Test accuracy	72.95	78.54	78.01

image extraction, the item *Gray degree* in Eq. (3) is set to 255, which corresponds to all white pixels distributed in the hand region. As mentioned, the depth-grayscale degree of the hand region is adjustable by appropriately setting *Gray degree*, and the optimal recognition accuracy can be acquired by trial and error. In addition, the recognition accuracy and loss value curves of VGG-16 CNN recognition with *ROI-depth*, *ROI-depth* + *binary-values*, and *ROI-depth* + *binary-values* + *binary-values* in the training phase of the deep learning models are also observed, as shown in Figs. 11–13, respectively. It can be observed by comparing the figures that the improved depth data of *ROI-depth* + *binary-values* + *binary-values* have the most satisfactory performance, i.e., both the highest recognition accuracy and the lowest loss function value after a small number of iterations of model learning, followed by the improved depth data of *ROI-depth* + *binary-values*. The *ROI-depth* images, which still have the undesired black shadow regions, have the lowest performance.

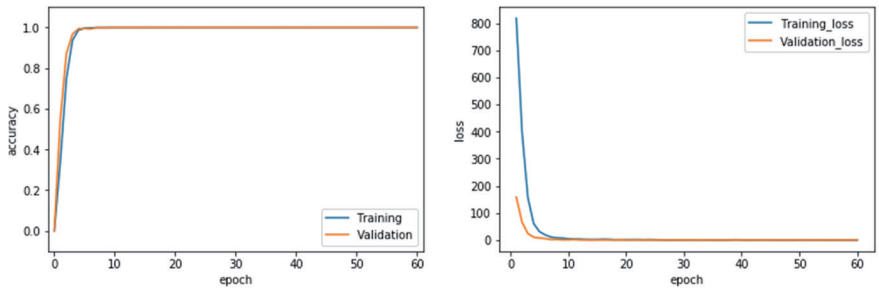


Fig. 11. (Color online) Accuracy and loss value curves of VGG-type CNN hand gesture recognition using *ROI-depth* over 60 iterations of model training in deep learning.

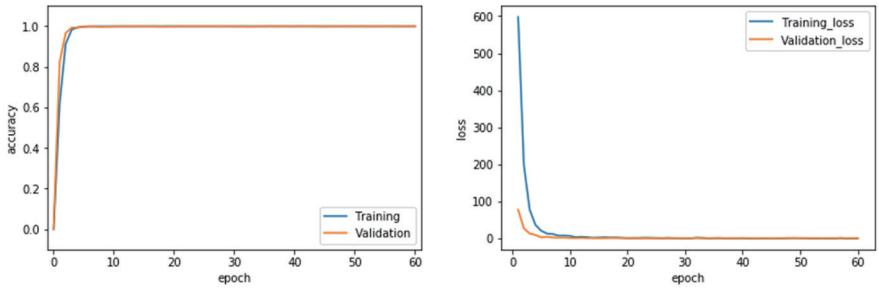


Fig. 12. (Color online) Accuracy and loss value curves of VGG-type CNN hand gesture recognition using *ROI-depth* + *binary-values* over 60 iterations of model training in deep learning.

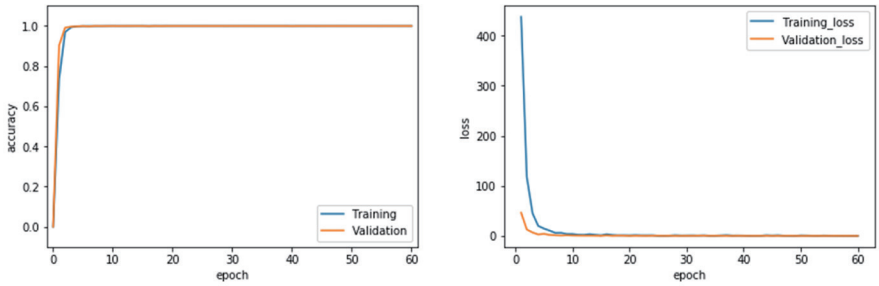


Fig. 13. (Color online) Accuracy and loss value curves of VGG-type CNN hand gesture recognition using *ROI-depth* + *binary-values* + *binary-values* over 60 iterations of model training in deep learning.

5. Conclusions

Hand gesture depth images obtained from the RGB-D depth sensor include an undesired shadow effect, adversely affecting hand gesture recognition. In this study, a serial binary image extraction approach containing two consecutive phases of well-designed binary value image process schemes was employed to effectively remove the black shadow regions in hand depth images. Depth-sensor-based hand gesture interaction action recognition was developed by VGG-type CNN deep learning with improved depth data with removed shadow regions. Experimental results clearly show that in terms of the recognition accuracy of 10 common hand gesture interaction actions by VGG-type CNN deep learning, the improved hand depth data derived from the binary image extraction calculations of the two phases are superior to the original depth images acquired from the depth sensor.

Acknowledgments

This research is partly supported by the Ministry of Science and Technology (MOST) of Taiwan under Grant MOST 109-2221-E-150-034-MY2.

References

1 M. B. Shaikh and D. Chai: Sensors **21** (2021) 4246. https://doi.org/10.3390/s21124246
2 Y. Liu, R. Ma, H. Li, C. Wang, and Y. Tao: J. Sens. **2021** (2021). https://doi.org/10.1155/2021/8864870
3 I. J. Ding and N. W. Zheng: Sens. Mater. **32** (2020) 2329. https://doi.org/10.18494/SAM.2020.2881
4 S. Mori, O. Erat, W. Broll, H. Saito, D. Schmalstieg, and D. Kalkofen: IEEE Trans Vis Comput Graph **26** (2020) 2994. https://doi.org/10.1109/TVCG.2020.3003768
5 C. Zhang, T. Huang, and Q. Zhao: Sensors **19** (2019) 5082. https://doi.org/10.3390/s19235082
6 M. Zollhöfer, P. Stotko, A. Görlitz, C. Theobalt, M. Nießner, R. Klein, and A. Kolb: Comput. Graphics Forum **37** (2018) 625. http://dx.doi.org/10.1111/cgf.13386
7 I. J. Ding and Z. G. Wu: IEEE Sens. Lett. **19** (2019) 8432. https://doi.org/10.1109/JSEN.2018.2873490
8 I. J. Ding and C. M. Ruan: J. Imaging Sci. Technol. **63** (2019) 50402-1. https://doi.org/10.2352/J.ImagingSci.Technol.2019.63.5.050402
9 I. J. Ding and Y. J. Chang: Neurocomputing **262** (2017) 108. https://doi.org/10.1016/j.neucom.2016.11.089
10 I. Tashev: IEEE Signal Process Lett. **30** (2013) 129. https://doi.org/10.1109/MSP.2013.2266959
11 R. McCartney, J. Yuan, and H.-P. Bischof: Proc. 2015 Int. Conf. Image Processing, Computer Vision, & Pattern Recognition (2015) 3–9. https://scholarworks.rit.edu/other/857
12 H. Gonzalez-Jorge, B. Riveiro, E. Vazquez-Fernandez, J. Martínez-Sánchez, and P. Arias: J. Int. Meas. Confed. **46** (2013) 1800. http://dx.doi.org/10.1016/j.measurement.2013.01.011
13 Z. Jiang, Q. Zhao, and Y. Tomioka: Proc. 2019 Int. Conf. Machine Learning and Cybernetics (IEEE, 2019) 1–6. https://doi.org/10.1109/ICMLC48188.2019.8949186
14 A. Staroverov, D. A. Yudin, I. Belkin, V. Adeshkin, Y. K. Solomentsev, and A. I. Panov: IEEE Access **8** (2020) 195608. https://doi.org/10.1109/ACCESS.2020.3034524
15 J. Jhou, Y. Lin, and Y. Leu: Proc. 2018 Int. Automatic Control Conf. (IEEE, 2018). https://doi.org/10.1109/CACS.2018.8606747
16 A. Cofield, Z. El-Shair, and S. A. Rawashdeh: Proc. 2019 IEEE National Aerospace and Electronics Conf. (IEEE, 2019) 437–442. https://doi.org/10.1109/NAECON46414.2019.9057808
17 D. Jiang, G. Li, G. Jiang, D. Chen, and Z. Ju: Proc. 2018 IEEE Int. Conf. Systems, Man, and Cybernetics (IEEE, 2018) 4041–4046. https://doi.org/10.1109/SMC.2018.00685
18 E. Coupeté, F. Moutarde, and S. Manitsaris: Procedia Manuf. **3** (2015) 518. https://doi.org/10.1016/j.promfg.2015.07.216
19 H. Takimoto, J. Lee, and A. Kanagawa: Int. J. Mach. Learn. Comput. **3** (2013) 245. https://doi.org/10.7763/IJMLC.2013.V3.312

20 L. Xu, C. Hu, J. Tao, J. Xue, and K. Mei: IEEE Trans. Circuits Syst. Video Technol. **31** (2021) 890. https://doi.org/10.1109/TCSVT.2020.2991987
21 K. Lai and S. N. Yanushkevich: Proc. 2018 Int. Conf. Pattern Recognition (IEEE, 2018) 3451–3456. https://doi.org/10.1109/ICPR.2018.8545718
22 N. Otberdout, L. Ballihi, and D. Aboutajdine: Proc. 2017 Intelligent Systems and Computer Vision (2017) 1–8. https://doi.org/10.1109/ISACV.2017.8054904
23 D. Tasmere, B. Ahmed, and S. Das: Int. J. Comput. Appl. **174** (2021) 28. http://dx.doi.org/10.5120/ijca2021921040
24 W. He, Z. Xie, Y. Li, X. Wang, and W. Cai: Sensors **19** (2019) 2919. https://doi.org/10.3390/s19132919
25 G. Danciu, S. M. Banu, and A. Caliman: Proc. 2012 Int. Conf. System Theory, Control and Computing (IEEE, 2012) 1–6.
26 K. Simonyan and A. Zisserman: Proc. Int. Conf. Learning Representations (2015). https://arxiv.org/abs/1409.1556

About the Authors

Ing-Jr Ding received his B.S. degree from Chang-Gung University in 1999, his M.S. degree from National Central University in 2001, and his Ph.D. degree from National Chiao-Tung University in 2008. He is currently a professor in the Department of Electrical Engineering, National Formosa University, Taiwan. His major research interests include speech processing and recognition, pattern recognition, artificial intelligence, innovative sensor technology, and multimedia techniques.

Nai-Wei Zheng received his M.S. degree from the Department of Electrical Engineering, National Formosa University in 2021. He is currently fulfilling his mandatory military service. His research interests are deep learning, image processing, and hand gesture recognition.