International Journal of Computer & Software Engineering

Research Article   **Open Access**

Object Shape Classification Using Spatial Information in Myoelectric Prosthetic Control

Ryusei Shima[1], Yunan He[1*], Osamu Fukuda[1], Nan Bu[2], Hiroshi Okumura[1] and Nobuhiko Yamaguchi[1]

[1]Department of Information Science, Saga University, Saga, Japan
[2]Department of Control and Information Systems Engineering, NIT, Kumamoto College, Kumamoto, Japan

Abstract

This paper proposes a novel prosthetic hand control method that incorporates spatial information of target objects obtained with a RGB-D sensor into a myoelectric control procedure. The RGB-D sensor provides not only two-dimensional (2D) color information but also depth information as spatial cues on target objects, and these pieces of information are used to classify objects in terms of shape features. The shape features are then used to determine an appropriate grasp strategy/motion for control of a prosthetic hand. This paper uses a two-channel image format for classification, which contains grayscale and depth information of objects, and the image data is classified with a deep convolutional neural network (DCNN). Compared with previous studies based only on 2D color images, it is expected that the spatial information would improve classification accuracy, and consequently better grasping decision and prosthetic control can be achieved. In this study, a dataset of image pairs, consisting of grayscale images and their corresponding depth images, has been created to validate the proposed method. This database includes images of simple three-dimensional (3D) solid objects from six categories, namely, triangular prism, triangular pyramid, quadrangular prism, rectangular pyramid, cone, and cylinder. Image classification experiments were conducted with this database. The experimental results indicate that spatial information possesses high potential in classifying shape features of objects.

Introduction

Powered prosthetic hands can be intuitively controlled using surface electromyography (sEMG) signals, which may be measured at skin surface. This control strategy is called as myoelectric prosthetic control [1-3]. One of the two major methodologies for myoelectric control uses motion classification of sEMG signals and drives the prosthetic hand with classified motions [4]. The state of the art technology offers dexterous and multi-fingered prosthetic hands that are extremely advanced from a mechanical point of view and these hand devices can perform many different movements [5]. A variety of sEMG features and optimized pattern recognition techniques have been developed and extensive investigations have been conducted in order to improve accuracy of motion classification and to achieve reliable and robust control performance [6]. However, the classification capability is still limited to around ten motions, so that these multi-DoF and complicated hand devices are difficult for operations with enough dexterity and robustness and there is no commodity system yet that may perform effective and effortless multi-functional prosthetic control in applications of daily activities [7].

In order to make full use of the dexterity and high DoF in hand prostheses, alternative sources of information have been incorporated into myoelectric control system. For example, multimodal approaches based on fusion of biomechanical sensors signals (e.g., accelerometers and gyroscopes) with sEMG have been investigated to enhance stability, robustness, and accuracy of motion classification [8-10]. Biomechanical sensors can be embedded in prosthetic devices or attached on user's limb. These sensors provide information, such as postures and positions of user's limb, and integrating these user-related cues significantly improves motion classification accuracy. Fukuda et al. proposed a concept of Internet of Things (IoT)-based myoelectric control system that gathers information of a wide range of sensor data from sensor devices equipped in prosthetic hand, target objects, and even the operation environment [11]. Similarly, an RFID-based prosthetic controller has been developed that can change prosthetic hand's grips automatically when approaching it to an object embedded with an RFID tag [12]. There are a lot of

information sources available for prosthetic manipulation and such a rich information body would support prosthetic hands in achieving autonomous myoelectric control.

It is the inevitable trend to employ image sensors in prosthetic control system, since visual perception plays an important role for human being during object manipulation. More a control system is aware of the targets, more possible it is to plan a better grasping strategy and to achieve more reasonable and steady manipulating motions. Recently, several methods have been proposed to extract or recognize features of targets objects with image sensing techniques [13-18]. For example, Dosen et al. [13] built a cognitive vision system for control of dexterous prosthetic hands. The system uses image data obtained from a web camera and a distance sensor to estimate dimension features of a target object. The dimension features are used to determine the grasp type and size.

On the other hand, the progress in deep learning was amazing in the last few years and deep learning applications had led great revolution in the field of computer vision. Deep convolutional neural network (DCNN) is one of the cruces of deep learning methods for computer vision, which makes computers easier to understand images [19]. The authors' group proposed a myoelectric prosthetic control system equipped with a camera device and target object was recognized with image information from the camera [15]. A DCNN-based object classifier is used to recognize 22 categories of objects that are commonly used in daily life and a proper motion plan and corresponding control commands are selected according to the object category [16]. In the research studies of DeGol et al. [17] and Ghazaei et al.

*Corresponding Author: Yunan He, Department of Information Science, Saga University, Saga, Japan; E-mail: heyunan@live.com

[18], image data of objects are directly mapped into grasp patterns using DCNN and these grasp patterns may rule the following prosthetic control.

However, detailed information (e.g., colors, textures, and boundaries) contained in object images is redundant for classifying simple grasp patterns and it is obvious that shape and dimension features of an object carry out more important duty in determining an appropriate grasp type. Spatial information can be estimated with image data obtained with RGB-D sensors. RGB-D sensor provides high resolution depth image and the corresponding RGB image. With the depth (distance) information, computer vision techniques based on traditional two dimensional (2D) images can be extended to their three dimensional (3D) version. Generally, reconstruction of 3D models from 2D image and depth image requires strict matching of points and time-consuming computation. Detailed and rich/redundant information of a 3D object model is also not an indispensable issue for object classification. If part of the 3D features related to the target object is sufficient for classification, the 3D reconstruction procedure can even be omitted from the data processing. Following this idea, some research studies have applied RGB-D sensor data to DCNN classifiers for object classification. RGB-D data possess four channels of raw images that are not directly applicable to the input of DCNN. A two-stream CNN architecture has been developed to deal with RGB-D images, where RGB image and depth image are handled separately with two DCNNs and the two streams are fused at the final layers for classification [20]. A similar multimodal method was developed with three stream DCNN structure [21]. In addition, a sophisticated 3D CNN architecture was utilized with a smart voxel representation that is converted from RGB-D images [22].

Alternatively, it would be more straightforward to achieve grasp pattern decision based on spatial information, such as shape features and placement directions, than categories of the objects since different objects with same shape features may share similar grasp patterns. The present paper attempts to develop a novel myoelectric control method using shape features of target object without exact and clear category information as well as a rigorous 3D model of the object. The object classification applies 2D and depth images directly to a DCNN and effective 3D features are extracted with the DCNN for classification of object shapes. By converting RGB true-color information to grayscale intensity image, the DCNN only has two input image channels, one grayscale and one depth, so that the network architecture is one-stream and this contribution may bring better optimization of the DCNN parameters with lower 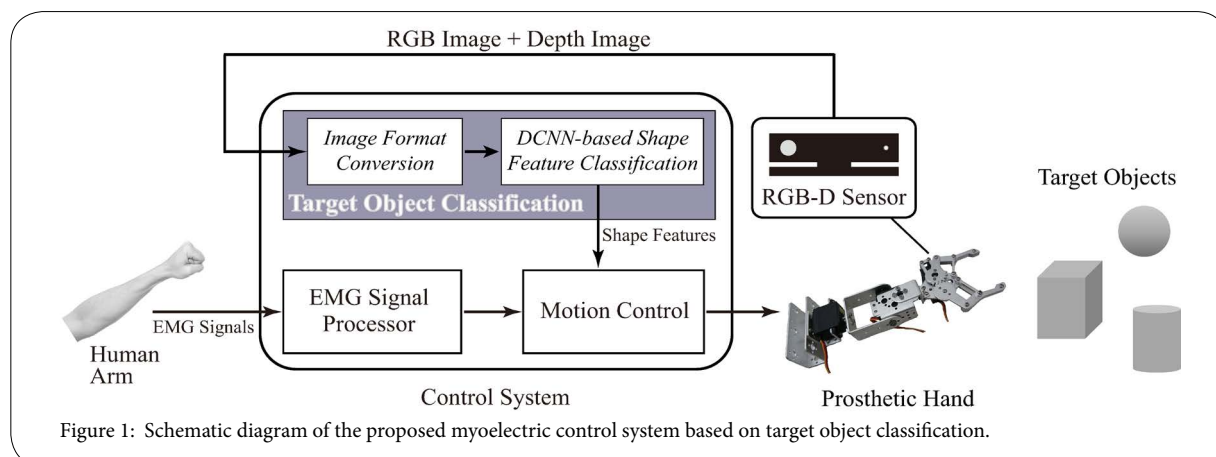computational burden. This end-to-end mapping provides shape cues for motion control in a myoelectric prosthetic control system, which accepts sEMG signals, depth and 2D RGB images as input and returns with corresponding grasp patterns as well as the control signals for prosthetic hands. Compared with the prosthetic control system with only RGB sensors, the proposed system focuses on the potential of spatial information in shape feature classification. Meanwhile, it can benefit from the characteristics of the RGB-D sensor, e.g. invariant to lighting conditions, adjustable sensing distance for depth perception. The proposed control system is expected to achieve much more reliable and robust manipulation and to realize high level autonomous prosthetic control.

The Proposed Myoelectric Control

Myoelectric Control with Object Shape Classification

A schematic view of the proposed myoelectric control system is shown in Figure 1. The system consists of three major parts: 1) Target object classification; 2) real-time EMG signal processing; and 3) motion control. Suppose a user is manipulating an object using a prosthetic hand, which is equipped with a RGB-D sensor. He/she may control the orientation of the hand while approaching to the object. In the meanwhile, the control system obtains RGB images and depth images simultaneously from the RGB-D sensor. The image streams are fed into a DCNN-based target object classifier, mapping image information of the objects to their shape features. For each shape, a grasp pattern is predefined with step-by-step control commands for motors.

On the other hand, sEMG signals are obtained from muscles of user's arm. After full wave rectification and low pass filtering, linear envelopes are derived as amplitude levels of the sEMG channels and the amplitude levels can be interpreted as user's intention [2,4]. For example, in our previous studies [15,16], two channels of sEMG signals are measured from a pair of muscles, one for an extensor and the other for a flexor. When amplitude level of the extensor is higher than that of the flexor, the control system can explain it as that the user intends to grasp the object. The user intention can be interpreted as relaxing the grip on the object when amplitude level of the flexor is higher. In addition, if the amplitude of either channel does not exceed a predefined threshold and it will be treated as no motion. Otherwise, the prosthetic hand will execute the control commands generated by the target classification module. It is worth noting that even the same object, the grasp strategy can be different according to the intention of the operator. For example, whether you want to use scissors or you

Figure 1: Schematic diagram of the proposed myoelectric control system based on target object classification.

want to give scissors to someone. In this case, multiple grasp strategies should be predefined. The selection of the grasp strategies can be realized by discriminating more patterns from sEMG signals.

Image information from an RGB-D sensor

RGB-D sensor data contains a true-color RGB image and its corresponding depth image. It is worth noting that information from the depth image has a superiority in applications like prosthetic control. The depth sensing works on the principle that the sensor builds a depth map through projecting a pattern of infrared light on objects, so that the depth information is invariant to lighting and color conditions. In addition, the depth mapping can be configured to discard sensor data outside of a certain distance range. This will help the object classification module to easily remove background and keep focus on the target object.

The true color RGB image is converted to a grayscale intensity image by eliminating the hue and saturation information while retaining the luminance. The grayscale image is stacked on the depth image and a new image format is created with two image channels. The motivation of format conversion is twofold. Firstly, the depth image may have equal volume of information as that of 2D grayscale image in the converted format. Then, a two-channel image is readily available for application of DCNN and the computational burden may be relatively smaller than a four-channel format. The two-channel image is input into a DCNN for shape feature classification.

DCNN-based shape feature classification

This part classifies target objects in terms of shape features using grayscale and depth information extracted from RGB-D image. Several levels of shape features may be used as classes in the final layer, such as, shape categories of 3D solids, some common characteristics of 3D shapes, and 3D solid shapes with placement directions. This paper considers six categories of shapes, namely, triangular prism (TPR), triangular pyramid (TPY), quadrangular prism (QPR), rectangular pyramid (RPY), cylinder (CLD), and cone (CON).

The DCNN is developed based on the network architecture of AlexNet, which is famous because of its remarkable performance in ILSVRC 2012 [23]. In order to avoid overfitting in some extent, the local response normalization of AlexNet is replaced with batch normalization [24]. This DCNN consists of five convolutional layers followed by max-pooling layers, and three fully-connected layers with a final **N**-way softmax. The number of outputs, **N**, corresponds to the number of classes. Figure 2 illustrates an example of the DCNN-based classifier structure for six classes of shape names. It needs to be noted that the DCNN classifier is one-stream so that it is easier for training than previous studies, which use multi-stream structures to process RGB image and depth image separately [20,21].

Experiments

Three object shape classification experiments were carried out to evaluate the proposed target object classification module. The validation of whole myoelectric control system will be reported in our future studies.

Dataset

A small dataset is prepared with respect to the size of DCNN structure and the database contains RGB-D images of 3D solid objects from six categories. For each shape, RGB-D images were taken from different angles/directions and distances. The image dataset was created using a Kinect sensor device (Microsoft Corp.) as shown in Figure 3. The Kinect sensor is featured with an RGB camera and a 3D depth sensor. Resolution of RGB images is 1920 × 1080 with a frame rate limit as 30 fps. The viewing angle is 84.1 deg in the horizontal direction and 53.8 deg in the vertical direction. The measurement range of the depth sensor is [0.05, 8.0] m, while higher limit of the guaranteed depth range is 4.5 m. The resolution of the depth image is 512 × 424 and the frame rate is 30 fps, the viewing angles are 70 deg and 60 deg in the horizontal and the vertical directions.

Six objects were prepared with similar dimensions and volumes for six shape categories, respectively. These solid objects were all in blue color and no meaningful texture information is included in the database. For image acquisition, the objects were hung from the ceiling and were turned during image data recording. Video sequence from the Kinect sensor was captured and RGB and depth images were taken from the video randomly in order to collect data from various directions. The sensor was mounted at different heights to acquire RGB-D data of the objects from different angles and distances. The RGB and depth images were recorded synchronously at 30 Hz. Figure 3(b) shows an example of the acquired RGB and depth information. The RGB images were converted to grayscale images. Since there are differences in resolution and viewing angle between the grayscale
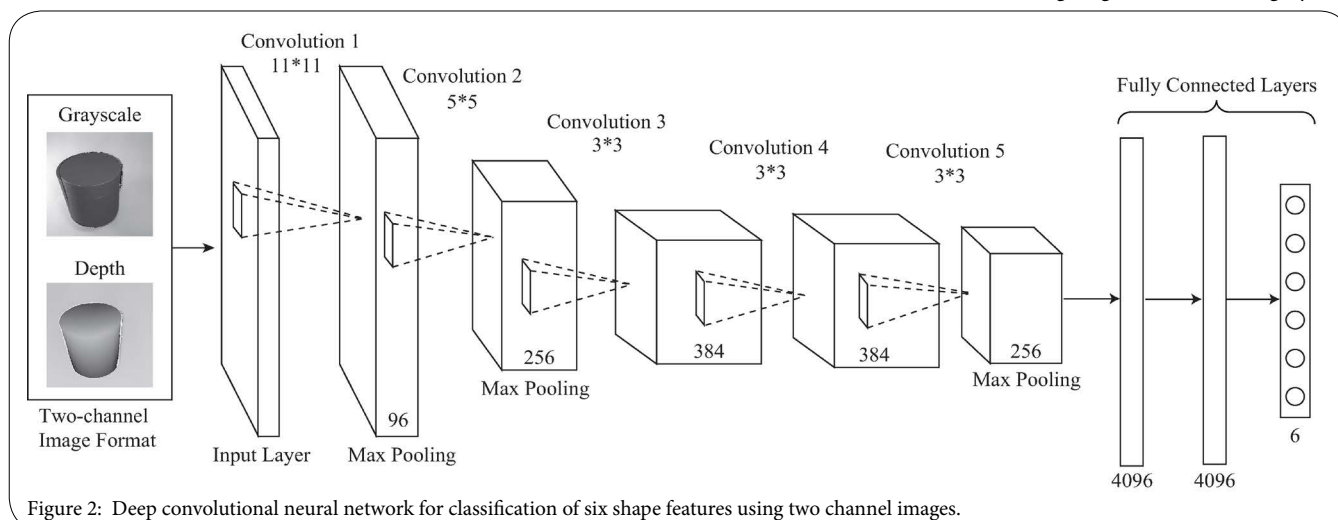

Figure 2: Deep convolutional neural network for classification of six shape features using two channel images.

image and the depth image, simple alignment of the image resolutions and positions of the object were made to form an image pair. These image pairs were stored to construct the 3D object image database. Examples of (grayscale and depth) image pairs for six shape categories are shown in Figure 4.

Experimental setup

The DCNN was implemented on a Caffe deep learning framework [25]. The computation was accelerated with an NVIDIA TESLA K40 GPU. With the 3D object image database, image pairs of objects can be grouped into classes according to the class definition. In the experiments, samples were randomly selected from each class in order to train the network structure, while different sample data were used for test. Initialization of the weight coefficient was achieved with the Xavier uniform initializer and an all zero array was used as bias. Optimization based on stochastic gradient descent was used in order to minimize the cross-entropy loss. In the image classification experiments, the training sets were input into the DCNN classifier and the network updated its weights through training process.

Shape classification with images of constant direction

First, shape category classification experiment was conducted with image data taken from same angle. Here, images taken from the horizontal direction were picked out (50 pairs per shape; 300 pairs in total) to train the DCNN classifier and the number of classes, **N** was six. Another 90 image pairs (15/shape) were used for test the classification accuracy. Examples of depth images for each shape category can be found in Figure 5.
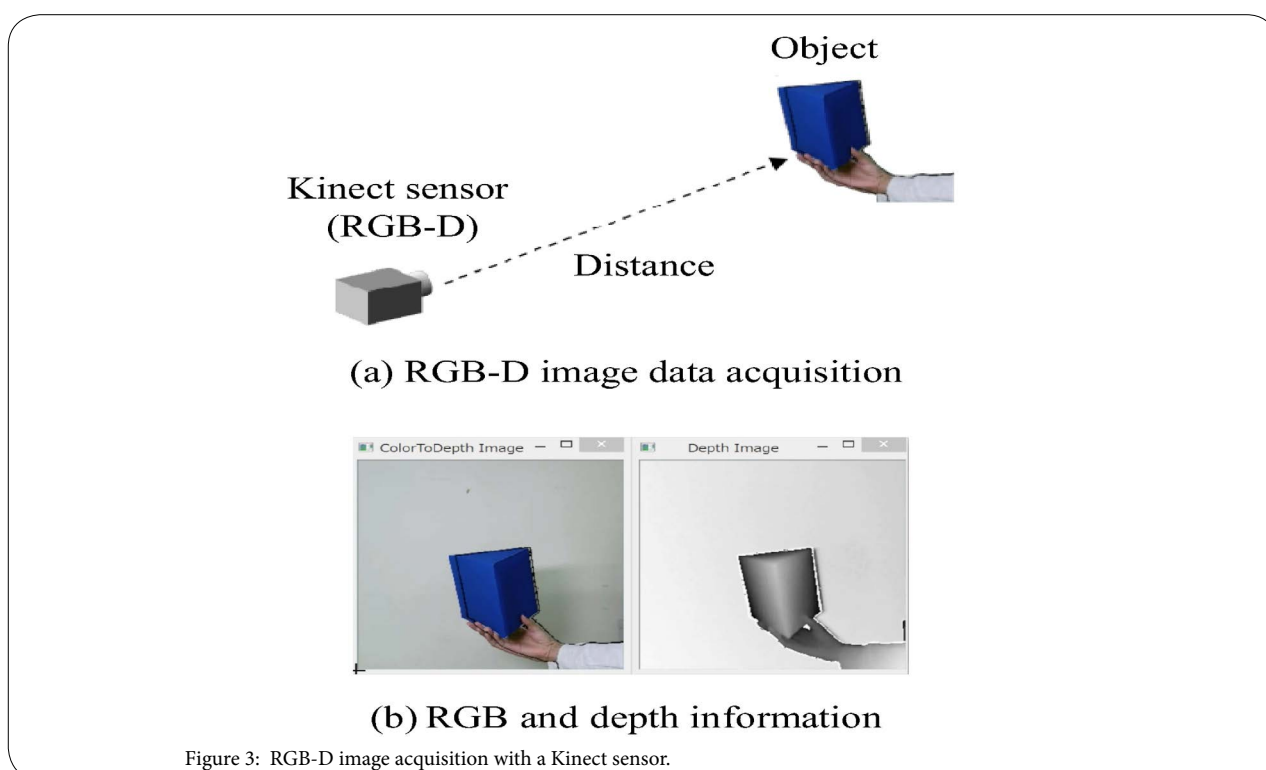

(a) RGB-D image data acquisition

(b) RGB and depth information

Figure 3: RGB-D image acquisition with a Kinect sensor.

Triangular prism Quadrangular prism Cylinder

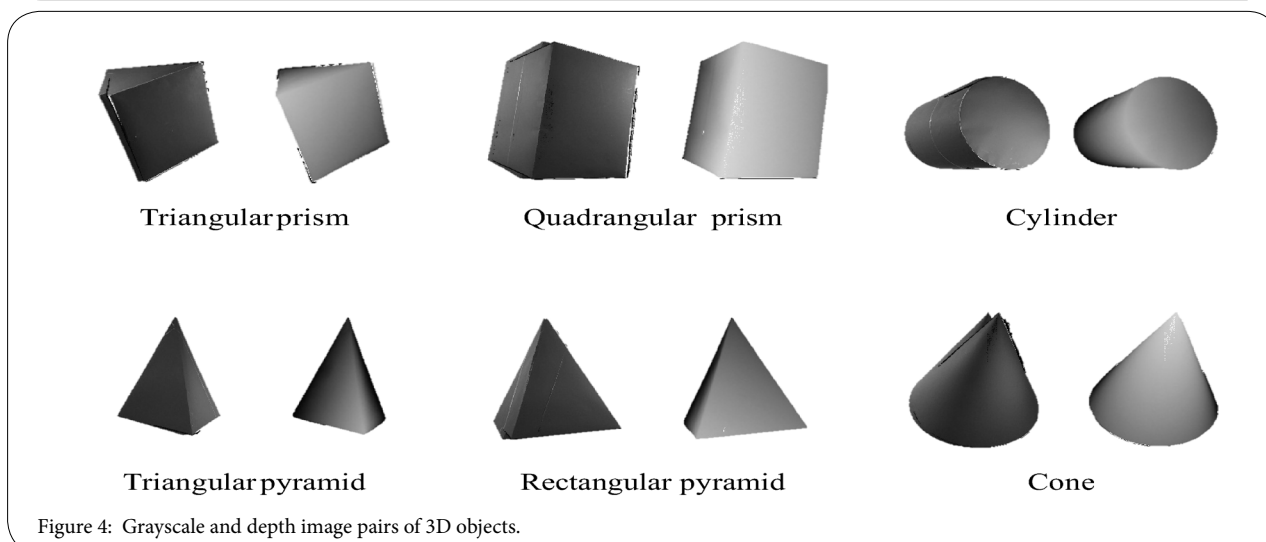Triangular pyramid Rectangular pyramid Cone

Figure 4: Grayscale and depth image pairs of 3D objects.

After 30 epochs of training on the training set, the loss value stopped increase and the overall training accuracy was achieved around 98.9%. The classification accuracies for six classes are shown in Table 1. It is clear from the depth image examples shown in Figure 5 that classification of shape categories with images from constant direction is relatively easy for DCNN.

Shape classification with images of different directions

Image classification experiment was also carried out with data taken from different directions. Directions of the images are all random selected. For training data, 300 image pairs are used for each shape category (1,800 pairs in total) and the number of classes, **N** = 6. Another 90 image pairs (15/shape) were used for assess the classification accuracy. Examples of depth images with different directions are given in Figure 6. The overall classification accuracy of this experiment was 92.2% and details of classification results are listed in Table 2.

Although more samples were used in this experiment, the classification accuracy dropped slightly because more directions were included in this problem. In addition, it should be noted that decrease in the overall classification accuracy is not significant, so is for the individual classification rates. For objects of rectangular pyramid, twelve among the 15 test samples were correctly classified and the rest three samples were classified as cone. This result meets the common sense that misclassification is easy to occur between objects with similar shapes. For some special directions, objects are difficult to be correctly recognized even by human beings. This experiment indicates the validation of the proposed object shape classification method and the authors would like to improve performance of the proposed method in the future works.
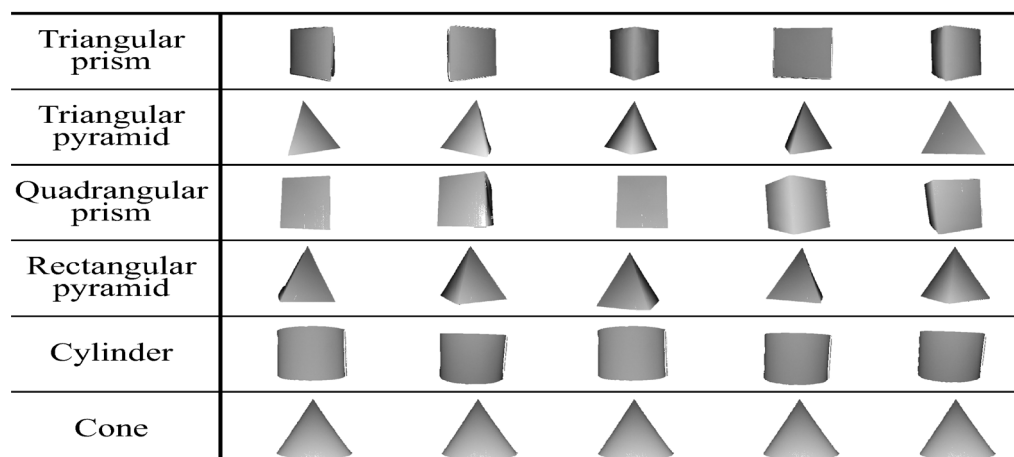

Figure 5: Examples of depth images for six shape categories that were taken from constant direction.

Object	TPR	TPY	QPR	RPY	CLD	CON
Accuracy	93.3	100	100	100	100	100

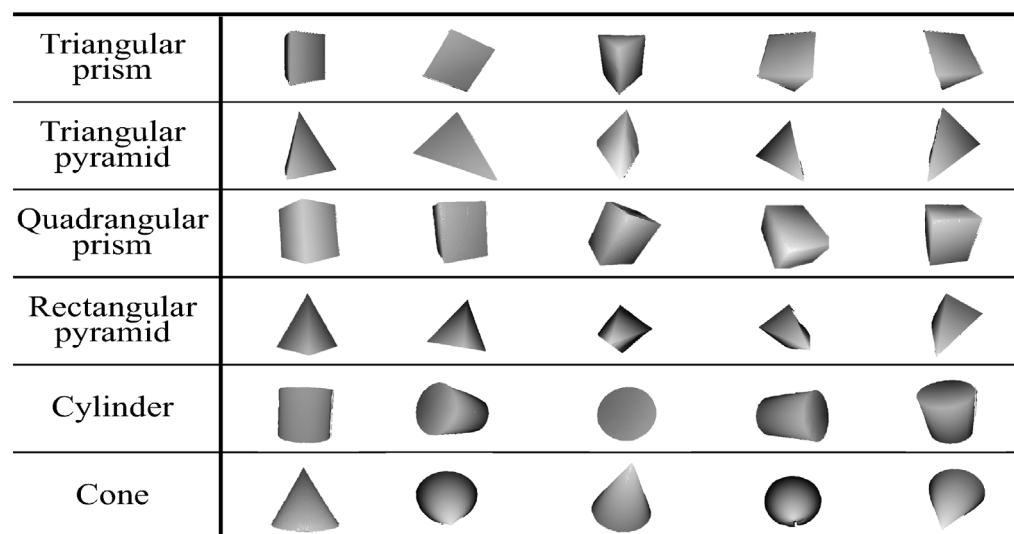Table 1: Classification accuracies of six classes with image data taken from constant direction [%].

Figure 6: Examples of depth images for six shape categories that were taken from different directions.

Classification for objects with smooth and angular edges

Finally, a two class image classification experiment was conducted to classify objects with smooth edges from those with angular edges. Figure 7 shows examples of training data for each class in this experiment. From the test database, objects with shapes of cone and cylinder, which have smooth edges (viewed from the sensor), were grouped together and all the other objects were grouped as another class labelled as angular edge. For each class, 600 samples were randomly selected for DCNN training and another 15 samples/class were used for test. The confusion matrix between two classes is depicted in Table 3. The classifier missed only one sample for each class in the experiment. With these experimental results, it is proven that the proposed method can classify not only shape categories, as the results shown above, but also abstracted shape features. Besides the edge feature, many other feature information, such as, surface smoothness, edge numbers, etc., can be considered with the proposed method. This functionality is of great importance for prosthetic control since the shape features of target objects may largely affect a control strategy.

Visualization of the process results in all convolution and pooling layers also provides an insight on neuron responses and the classification ability of the proposed DCNN classifier. Figure 8 shows examples of visualization results of three convolution layers (the first, third and fifth) and the pooling layer just before the fully connected layers. In this figure, visualization results in the top row are for a cone object and those in the bottom row are for a triangular pyramid. Although the difference in intensities of each layer is not extremely large, it can be recognized between two objects and these differences represent that a well trained DCNN classifier processes objects from each class in two distinct manners. Remember that a DCNN has a deep-layer structure, the data processing and feature extraction functionality are distributed over the whole network architecture in DCNN.

True Category	Predicted Shape Category					
	TPR	TPY	QPR	RPY	CLD	CON
TPR	93.3					6.7
TPY		93.3			6.7	
QPR			100			
RPY				80.0		20.0
CLD					100	
CON		13.3				86.7

Table 2: Confusion matrix of the classification result using image data taken from different direction [%].

True Feature	Predicted Shape Feature	
	Smooth	Angular
Smooth	93.3	6.7
Angular	6.7	93.3

Table 3: Confusion matrix of the two class classification for objects with smooth and angular edges [%].

Conclusion

This paper proposed an object shape classification method using DCNN in the framework of a novel myoelectric prosthetic hand control system, which is based on image information of target objects. Spatial information obtained from RGB-D sensor is the key for the proposed shape feature classification method. A two-channel image format and the one-stream DCNN architecture is one of the contributions developed in this paper. This combination facilitates the training process and the classification performance. The DCNN classifier can classify image data of objects in terms of either their shape categories or shape features that are shared among
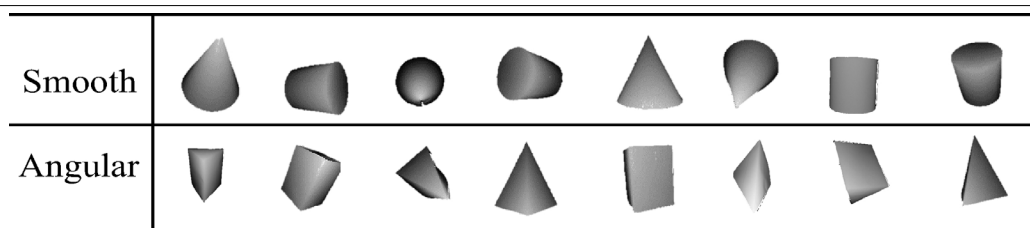

Figure 7: Examples of depth images for objects with smooth edges and angular edges.
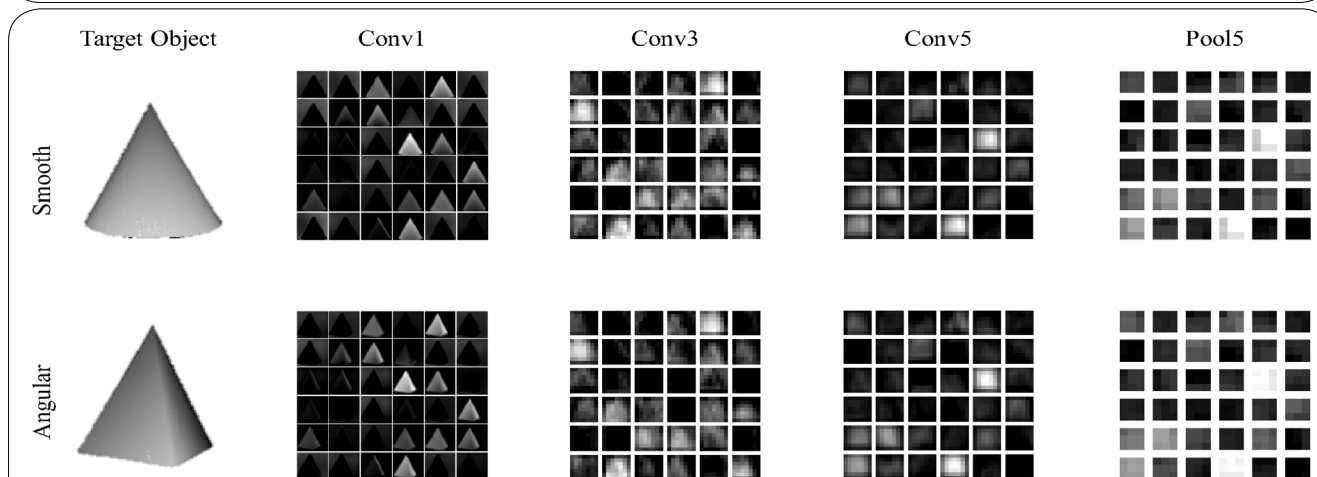

Figure 8. Visualization of the process results in convolution and pooling layers for a cone object and a triangular pyramid.

several categories. In the scenario of myoelectric control for object manipulation, both shape categories and shape features are important for control strategy decision. In addition, the authors had developed object recognition methods to classify object categories from 2D image data in our previous studies [15,16]. All these cues extracted from image information can be summed together to achieve more reliable and effective autonomous prosthetic control.

A 3D object image dataset has been created with an RGB-D (Kinect) sensor to verify the proposed method. Three image classification experiments have been conducted and the overall classification accuracies all exceed 90%. These experimental results indicate that spatial information has high potential in shape feature classification of target objects.

In the future research, we would like to apply the proposed shape classification method to real objects. In addition, we are going to take advantage of the time-series characteristic of the RGB-D images since the time-series images contain not only distance information but also the spatial information from different perspectives. Furthermore, direct usage of RGB image with depth image, in the four channel format, needs to be considered when fusing the object category classification with the proposed method.

Competing Interests

The authors declare that they have no competing interests.

References

1. Farina D, Jiang N, Rehbaum H, Holobar A, Graimann B, et al. (2014) The extraction of neural information from the surface EMG for the control of upper-limb prostheses: emerging avenues and challenges. IEEE Transactions on Neural Systems and Rehabilitation Engineering 22: 797-809.

2. Fukuda O, Tsuji T, Kaneko M, Otsuka A (2003) A human-assisting manipulator teleoperated by EMG signals and arm motions. IEEE Transactions on Robotics and Automation 19: 210-222.

3. Parker P, Englehart K, Hudgins B (2006) Myoelectric signal processing for control of powered limb prostheses. Journal of Electromyography and Kinesiology 16: 541-548.

4. Oskoei M A, Hu H (2007) Myoelectric control systems—A survey. Biomedical Signal Processing and Control 2: 275-294.

5. Atzori M, Cognolato M, Muller H (2016) Deep Learning with Convolutional Neural Networks Applied to Electromyography Data: A Resource for the Classification of Movements for Prosthetic Hands. Front Neurorobot 10: 9.

6. Xie HB, Guo T, Bai S, Dokos S (2014) Hybrid soft computing systems for electromyographic signals analysis: a review. Biomedical Engineering Online 13: 8.

7. Fougner A, Stavdahl O, Kyberd PJ, Losier YG, Parker PA (2012) Control of upper limb prostheses: terminology and proportional myoelectric control—a review. IEEE Transactions on neural systems and rehabilitation engineering 20: 663-677.

8. Fang Y, Hettiarachchi N, Zhou D, Liu H (2015) Multi-Modal Sensing Techniques for Interfacing Hand Prostheses: A Review. IEEE Sensors Journal 15: 6065-6076.

9. Ohnishi K, Kajitani I, Morio T, Takagi T (2013) Multimodal sensor controlled three Degree of Freedom transradial prosthesis. Proc IEEE Int Conf Rehabilitation Robotics.

10. Xiloyannis M, Gavriel C, Thomik AAC, Faisal AA (2017) Gaussian Process Autoregression for Simultaneous Proportional Multi-Modal Prosthetic Control With Natural Hand Kinematics. IEEE Transactions on Neural Systems and Rehabilitation Engineering 25: 1785-1801.

11. Fukuda O, Takahashi Y, Bu N, Okumura H, Arai K, et al. (2017) Development of an IoT-based Prosthetic Control System. Journal of Robotics and Mechatronics 29: 1049-1056

12. Trachtenberg MS, Singhal G, Kaliki R, Smith RJ, Thakor NV, et al. (2011) Radio frequency identification—An innovative solution to guide dexterous prosthetic hands. Proc Int Conf IEEE Engineering in Medicine and Biology Society 2011: 3511-3514.

13. Dosen S, Cipriani C, Kostic M, Controzzi M, Carrozza MC, et al. (2010) Cognitive vision system for control of dexterous prosthetic hands: experimental evaluation. Journal of Neuroengineering and Rehabilitation 7: 42.

14. Markovic M, Dosen S, Popovic D, Graimann B, Farina D (2015) Sensor fusion and computer vision for context-aware control of a multi degree-of-freedom prosthesis. J Neural Engineering 12: 066022.

15. Bando Y, Bu N, Fukuda O, Okumura H, Arai K (2017) Object classification using a deep convolutional neural network and its application to myoelectric hand control. Proc 22nd Int Symp on Artificial Life and Robotics.

16. Bu N, Bandou Y, Fukuda O, Okumura H, Arai K (2017) A semi-automatic control method for myoelectric prosthetic hand based on image information of objects. Proc Int Conf Artificial Intelligence, Robotics and Human-Computer Interaction.

17. DeGol J, Akhtar A, Manja B, Bretl T (2016) Automatic grasp selection using a camera in a hand prosthesis. Proc IEEE 38th Ann Int Conf Engineering in Medicine and Biology Society 2016: 431-434.

18. Ghazaei G, Alameer A, Degenaar P, Morgan G, Nazarpour K (2017) Deep learning-based artificial vision for grasp classification in myoelectric hands. J Neural Eng 14: 036025.

19. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521: 436-444

20. Eitel A, Springenberg JT, Spinello L, Riedmiller M, Burgard W (2015) Multimodal deep learning for robust RGB-D object recognition. Proc IEEE/RSJ Int Conf Intelligent Robots and Systems.

21. Rahman MM, Tan Y, Xue J, Lu K (2017) RGB-D object recognition with multimodal deep convolutional neural networks. Proc IEEE Int Conf Multimedia and Expo.

22. Zia S, Yuksel B, Yuret D, Yemez Y (2017) RGB-D object recognition using deep convolutional neural networks. Proc IEEE Int Conf Computer Vision.

23. Krizhevsky A, Sutskever I, Hinton GE (2012) Image Net classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25: 1097-1105.

24. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. Proc Int Conf Machine Learning.

25. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, et al. (2014) Caffe: Convolutional Architecture for Fast Feature Embedding. Proc ACM Int Conf Multimedia.