Contemporary Engineering Sciences, Vol. 9, 2016, no. 20, 953 - 960
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ces.2016.66107

Feature Based Weighted Neural Network for Hand Gesture Recognition

HyeYeon Cho, Hyo-Rim Choi and Taeyong Kim*

Dept. of Advanced Imaging Science, Chung-Ang University
Heukseok-dong, Dongjak-gu, Seoul, 156-756, Republic of Korea
*Corresponding author

Abstract

Recently, developing interaction techniques that allow gesture recognition for home application and game control is one popular field in Human-Computer-Interaction (HCI) research. In this paper, we proposed a method for hand gesture recognition using artificial neural network algorithm commonly used in HCI. Previous studies have generally used distinguished distribution datasets. However, in the real world, gesture data are imbalanced. In addition, if gesture distribution data are imbalanced, it is difficult to classify gesture. To solve these problems, we present a gesture distribution feature based weighted Neural Network (FWNN) after adding trajectory distribution data to input layer before network training. Specific hand trajectory coordinate and the number of extrema of hand trajectory are added to the trajectory distribution data. Our experimental results demonstrate that our method is much more accurate than the method of using only sequence data.

Keywords: Neural Network, Kinect, Gesture recognition, HCI, Hand signals

1 Introduction

Gesture recognition is an interesting part of recently study. Camera based gesture recognition has been used in a wide variety of applications such as games, medical systems, hand signals, and sign language [1]. We can make a gesture using various parts of the body. Hand is used more widely than other parts of the body for making gestures. The overall goal of hand gesture recognition is to interpret hand pose and its meaning. Previous studies have commonly used 2-D image for hand gesture recognition. Due to the lack of robustness in changes in various environments, depth camera based studies are taking the spotlight recently [2]. Kinect that provides tracking data and depth data is convenient for gesture recognition research. Balanced and distinguished distribution datasets have been mostly used in experiments of previous studies. However, everyday gesture distribution data are imbalanced [3]. Gesture recognition accuracy becomes lower with imbalanced data or increased data. Various studies have been performed to overcome such disadvantages [4]. In general, hand gesture recognition needs complicated calculations which is hard to implement algorithm. In this paper, we used a method with simple calculation so that users can make easy feed-forward neural networks structure with a learning algorithm using back-propagation. First we used tracking algorithm with a Kinect sensor and extract hand trajectory. We added the trajectory distribution data onto the input layer to improve the accuracy. The trajectory distribution data used a very simple calculation. Through experiment, our method was found to be able to reduce the number of weight updates and improve the accuracy of hand gesture recognition. These concepts are included in the research of Feature based hybrid neural network for hand gesture recognition [14]. We included more details in this paper.

2 Hand gesture recognition and Neural Networks

2.1 Hand gesture recognition

Many studies have been performed on hand gesture recognition using Hidden Markov Models (HMM) [5], Dynamic Time Warping (DTW) [6], Support Vector Machine (SVM) [7], and Neural Networks (NN) [8]. Although HMM has rich mathematical structure with statistical model of sequence data that are suitable for various application fields such as speech signal and gesture recognition, it has discomfort in discretization of multi-dimensional data converted to one-dimensional data [5]. DTW is an algorithm for measuring time and speed difference between two sequences with weakness in data increase and operation increase [6]. SVM is a supervised learning model for gesture recognition. However, its speed and size have limit in the learning and testing phase [7]. We focused on Artificial Neural network. Neural network trains and updates until the output and target are matched. Once network is trained, classification and recognition can be used in verification step very quickly. It has problem of overfitting. However, there are many ways to fix overfitting. Neural network is a non-parametric model that is easier and faster to adopt than other models. Its application program and field are making progress for more than 15 years [8] [9] [10].

2.2 Neural Networks

Neural networks are models of biological neural structures. They are composed of multiple layers. Feed-forward neural network is a popular neural network structure with a learning algorithm widely using back-propagation. Neural networks have already been proven to be efficient in recognition and classification. They have several advantages. They can learn from themselves and handle complex data. In addition, they are robust in changing environment. Moreover, they can be built in less time [11]. Furthermore, it is easy to materialize and apply them. They can be used in many applications. However, it has a problem of overfitting when there the number of parameters is increased. According to gesture distribution feature, if data distributions between classes are highly overlapped, recognition performance will decline. To fix this problem, many suggestions have been provided [9] [12] [13]. We have suggested a gesture distribution feature based weighted Neural Network for hand gesture recognition using pre-extracted gesture distribution feature as input layer to recognize many diverse gestures [14]. It can be processed effectively and speedily.

3 Gesture spotting and calculation of weighted node

Here, we suggest a method to make diverse hand gestures effectively recognized based on feed-forward neural network classifier. If there are many overlaps between gestures classes, recognition performance will decline [3]. To have balanced data distribution, distribution of gesture can be extracted with simple calculation by composing a weighted node and adding it to previous node made of only hand tracking data as shown in Figure 1. Accuracy of recognition is raised by learning weighted node added onto the new input layer. Gesture spotting and preprocessing using distribution feature of track to extract are used to compose the weighted node.

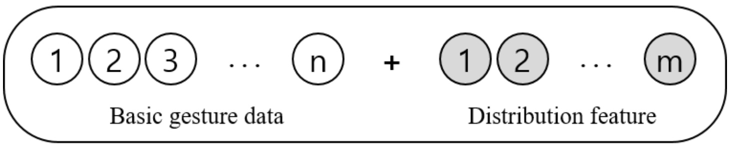

Fig 1. New input layer

3.1 Gesture spotting and preprocessing

In this paper, neural network as a classification algorithm is used to recognize hand gesture. Neural network for classification is a supervised learning algorithm that needs inputs and outputs. It requires the same length of input while learning. Therefore, in learning and classification, data have to be fixed in size. Spotting phase is also required in order to extract gesture matter. To suggest a method, we considered various body conditions that can normalize joint coordinate to relative coordinate. To designate a particular motion to preparation motion for hand gesture spotting, if Kinect starts joint tracking, then preparation motion is taken. Starting from a live hand moving from the preparation motion, the trajectory of hand gesture is captured. Every frame is captured except the first and the last frame to delete sudden change in trajectory. Captured gesture trajectories are extracted until the motion stops or certain parts of captured trajectories are cut if trajectories are long enough to extract hand gesture trajectories. However, if motion is too short, it will exclude trajectories from recognition. The trajectories of extracted hand gesture can be fixed to the same length with linear interpolation. Fixed coordinates of trajectories constituted the input layer.

3.2 Weighted node calculation by using distribution of trajectory

Besides trajectory used for gesture classification, to compose weighted node that can also affect classification, we used a simple method to calculate gesture distribution feature data for approximate classification and extrema of trajectory. We composed training dataset and test dataset to spotted hand gesture in order to compose additional node. Datasets were repeatedly collected gesture sequence data classified into different gesture classes. Represented distribution feature is extracted from training dataset to represent each class. Represented distribution feature can be used to verify hand gesture trajectory feature for approximate classification. Each distribution feature of a hand gesture trajectory is processed as part of the training dataset for extracting represented distribution feature.

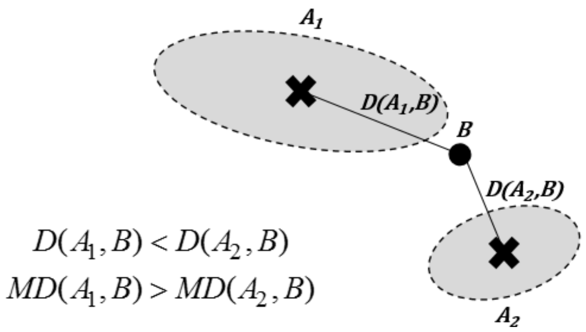

Fig 2. Euclidean distance and Mahalanobis distance considering the probability of distribution. A_k is kth class' representative location information. B is individual location information. $D(A_k, B)$ is Euclidean distance between A_k and B. $MD(A_k, B)$ is Mahalanobis distance between A_k and B

The represented distribution feature of gesture is then calculated. It is expressible by coordinates and covariance for processed distribution feature in each gesture class. Calculated represented distribution feature is centroid coordinate A_k and covariance S_k where k is the class of gesture. Here, we used two values of trajectory (the average coordinate of trajectory and the end coordinate of trajectory) for distribution feature. Those calculated dates are class reference data for approximate classification. Individual distribution feature B is extracted from a hand gesture composed of training and test datasets. Such method is used to process the distribution feature from the training dataset. At approximate classification, Mahalanobis distances between extracted A_k and B of a hand gesture in each k class are then repeatedly calculated. As shown in Figure 2, even if Euclidean distance is closer, calculating probability distribution considering Mahalanobis distance can result in accurate approximate classification. The following formula can be used to calculate Mahalanobis distance *MD* between each represented distribution feature and an individual distribution feature:

$$MD(A_k, B) = \sqrt{(A_k - B)S_k^{-1}(A_k - B)^T} \tag{1}$$

where A_k is the k[th] represented distribution feature data of each class, *B* is an individual distribution feature extracted from a gesture. S_k is covariance matrix in k[th] class and *T* is conversion matrix. After calculating index of gesture class that satisfies k, the shortest Mahalanobis distance is given by equation (2) as shown below:

$$\mathrm{N1} = \arg\min_k |MD(A_k, B)| \tag{2}$$

where *N1* is an index of the 1[st] nearest gesture class. As a gesture belongs to certain class, Mahalanobis distance of class will be shortened to compose *N1* as a weighted node. Another weighted node with extrema of trajectory is composed. If gesture holds a large number of waving, it may invade other classes that can fall off in recognition performance. Extrema of trajectory is a good feature showing gesture characteristics. With a number of extrema of gesture trajectory, you can earn a hint about how waving the gesture is with a simple calculation. Unnecessary extrema can be detected on fine hand shake, even with less waving. In order to reduce this problem, Gaussian filter kernel for smoothing is used to extract hand trajectory from a gesture. We used first derivative of smoothed trajectory to calculate the number of local extrema and compose extrema as weighted node from the calculation. If there are overlaps in data distribution for each class of gesture, recognition performance will fall off [13]. However, simple calculation can make data balanced by composing weighted node with approximate classification data and the numbers of extrema.

4 Experiments

For the experiment, of 20 joints data with Kinect, functional game was performed for bicycle hand signal using only wrist data to recognize hand gesture. Datasets are shown in Figure 3. We studied left and right hand separately with Kinect. We made 5 hand gestures while riding a bike. Three gestures composed of one motion (stop, go left or go right, beware). Two gestures had the waving motion (lead the way, slow down). A total of 500 learning data were established for 5 men. There were 20 input data in a hand per gesture, resulting in a total of 100 input data per man. A total of 150 testing data used for recognition were from 3 persons who had established learning data and made 10 input data per gesture (50 testing data per man). Input node was composed of 40 hand trajectory points and 3 weighted nodes. The hidden layer was composed of 26 layers targeting 5.

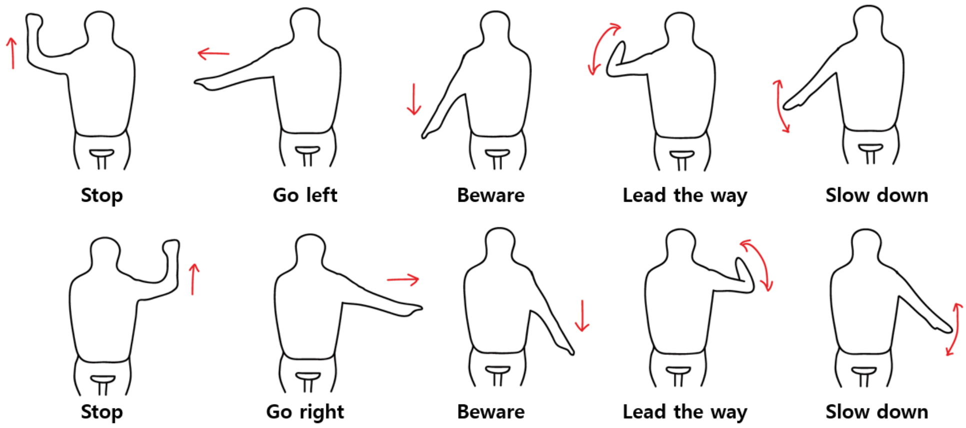

Fig 3. Bicycle hand signal of both hands. Upper image is for left hand signal. Lower image is for right hand signal

The results of performance analysis of FNN, FWNN, and DTW are shown in Fig. 4. Test data were composed of data of different participants from the learning data experiment, including T1 (widely distributed classes), T2 (widely distributed end-point but not classes), and T3 (widely distributed classes but not end-point). 1NN-DTW is dynamic time warping compared to one-nearest-neighbor. KNN-DTW is dynamic-time-warping compared to k-nearest neighbor.

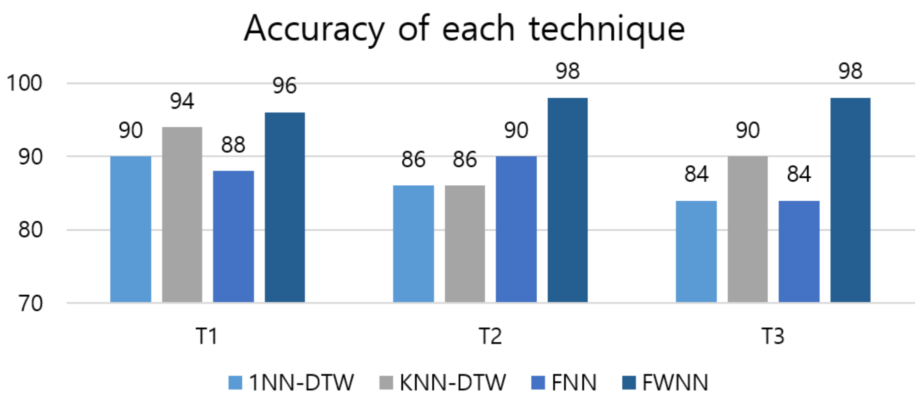

Fig 4. Accuracy of each hand gesture recognition technique

FNN is feed-forward neural network. It is the most popular neural network. In this paper, FWNN is the suggested gesture distribution feature based weighted Neural Network. It is composed of extracted feature from gesture distribution to weighted node with added feature of input layer to learn. Input parameters are extracted from basic hand gesture trajectory data. Weighted nodes are composed of numbers of extrema calculated from hand trajectory, an index of approximated classified gesture by using the extended position of hand trajectory, and an index of approximated classified gesture by using the last point of hand trajectory for FWNN. Generally, DTW and FWNN are more accurate than FNN. The accuracy depends on how much weighted node types are added.1NN-DTW is fast but less accurate as shown in Fig. 4. KNN-DTW is more accurate than 1NN-DTW. Although it takes 1.503 sec for FWNN to learn, once learned, it can categorize test dataset fast. The suggested method is 0.027 sec faster than other methods. As a result, recognition rate for FNN is 95.4% and the number of updates for weight is 31. Recognition rate for FWNN is 97.6% and the number of updates for weight is 18. Therefore, FWNN has higher recognition rate with less updates.

5 Conclusion

In this section, after adding extracted extended feature to input layer, FWNN is more accurate in classifying various hand gestures with faster recognition speed that is fit for live classification. FWNN can recognize not only simple hand gestures, but also waving hand gestures. However, z-axis leaned complex dynamic gestures need more experiments. Further studies are needed so that a method can be used to control home appliances or serious games with more hand gesture recognition abilities.

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science, and Technology (NRF-2015R1D1A1A01058394) and by the Chung-Ang University research grant in 2016.

References

[1] Siddharth S. Rautaray and Anupam Agrawal, Vision based hand gesture recognition for human computer interaction: a survey, *Artificial Intelligence Review*, **43** (2015), no. 1, 1-54. http://dx.doi.org/10.1007/s10462-012-9356-9

[2] Kajal Sharma, Kinect Sensor based Object Feature Estimation in Depth Images, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, **8** (2015), no. 12, 237-246. http://dx.doi.org/10.14257/ijsip.2015.8.12.23

[3] Nathalie Japkowicz, Learning from imbalanced data sets: a comparison of various strategies, AAAI Workshop on Learning from Imbalanced Data Sets, **68** (2000).

[4] Jingjing Zhang, et al., Least Squares Fuzzy One-class Support Vector Machine for Imbalanced Data, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, **8** (2015), no. 8, 299-308. http://dx.doi.org/10.14257/ijsip.2015.8.8.31

[5] Lawrence R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, **77** (1989), no. 2, 257-286. http://dx.doi.org/10.1109/5.18626

[6] Eamonn Keogh and Chotirat Ann Ratanamahatana, Exact indexing of dynamic time warping, *Knowledge and Information Systems*, **7** (2005), no. 3, 358-386. http://dx.doi.org/10.1007/s10115-004-0154-9

[7] Christopher J.C. Burges, A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, **2** (1998), no. 2, 121-167. http://dx.doi.org/10.1023/a:1009715923555

[8] Klimis Symeonidis, Hand gesture recognition using neural networks, Neural Networks, 13,5,1, 1996.

[9] Christopher M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

[10] Jack V. Tu, Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes, *Journal of Clinical Epidemiology*, **49** (1996), no. 11, 1225-1231. http://dx.doi.org/10.1016/s0895-4356(96)00002-9

[11] Tin Hninn Hninn Maung, Real-time hand tracking and gesture recognition system using neural networks, *World Academy of Science, Engineering and Technology*, **50** (2009), 466-470.

[12] Nitish Srivastava, et al., Dropout: A simple way to prevent neural networks from overfitting, *The Journal of Machine Learning Research*, **15** (2014), 1929-1958.

[13] Cristiano L. Castro and Antônio P. Braga, Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data, IEEE transactions on neural networks and learning systems, **24** (2013), no. 6, 888-899. http://dx.doi.org/10.1109/tnnls.2013.2246188

[14] HyeYeon Cho, Hyo-Rim Choi and Taeyong Kim, Feature Based Hybrid Neural Network for Hand Gesture Recognition, *Advanced Science and Technology Letters*, **129** (2016), 124-128. http://dx.doi.org/10.14257/astl.2016.129.25

Received: May 1, 2016; Published: August 12, 2016