# Hand Gesture Recognition Using a Radar Echo I–Q Plot and a Convolutional Neural Network

AUTHOR(S):

Sakamoto, Takuya; Gao, Xiaomeng; Yavari, Ehsan; Rahman, Ashikur; Boric-Lubecke, Olga; Lubecke, Victor M.

CITATION:

Sakamoto, Takuya ...[et al]. Hand Gesture Recognition Using a Radar Echo I–Q Plot and a Convolutional Neural Network. IEEE Sensors Letters 2018, 2(3): 7000904.

ISSUE DATE:

2018-09

URL:

http://hdl.handle.net/2433/259194

Sensor signals processing

# Hand Gesture Recognition Using a Radar Echo I–Q Plot and a Convolutional Neural Network

Takuya Sakamoto[1,2,3*], Xiaomeng Gao[4,5,6**], Ehsan Yavari[4**], Ashikur Rahman[1,7**], Olga Boric-Lubecke[1†], and Victor M. Lubecke[1†]

[1]*Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, HI 96822 USA*
[2]*Graduate School of Engineering, University of Hyogo, Himeji 671-2280, Japan*
[3]*Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan*
[4]*Adnoviv LLC, Honolulu, HI 96822 USA*
[5]*University of California, Davis, CA 95616 USA*
[6]*Cardiac Motion LLC, Sacramento, CA 95817 USA*
[7]*Aptiv PLC, Kokomo, IN 46902 USA*
[*] *Senior Member, IEEE*
[**] *Member, IEEE*
[†] *Fellow, IEEE*

Manuscript received June 4, 2018; revised July 7, 2018 and August 2, 2018; accepted August 18, 2018. Date of publication August 21, 2018; date of current version September 6, 2018.

Abstract—We propose a hand gesture recognition technique using a convolutional neural network applied to radar echo in-phase/quadrature (I/Q) plot trajectories. The proposed technique is demonstrated to accurately recognize six types of hand gestures for ten participants. The system consists of a low-cost 2.4-GHz continuous-wave monostatic radar with a single antenna. The radar echo trajectories are converted to low-resolution images and are used for the training and evaluation of the proposed technique. Results indicate that the proposed technique can recognize hand gestures with average accuracy exceeding 90%.

Index Terms—Sensor signals processing, gesture recognition, machine learning, neural network, radar.

## I. INTRODUCTION

Automatic gesture recognition, as represented by Google Soli [1], is an active field of research having various applications, including man-machine interfaces. Different approaches have been proposed for gesture recognition; e.g., the use of wearable devices [2]–[8], computer vision, and depth cameras [9]–[12]. The wearable devices allow the accurate and reliable measurement of human posture and motion, although the frequent wearing of such devices might be inconvenient and interfere with daily life. In contrast, computer vision techniques with RGB and depth cameras offer a noncontact measurement and more convenience to users. Nonetheless, the use of camera-based systems in a private space can cause privacy concerns.

Hand gesture recognition using radar and wireless sensors has attracted interest recently. Google Soli [1] uses a 60-GHz ultrawideband radar with a $2 \times 4$ multiple-input multiple-output array, and its outstanding performance has been demonstrated, although such a radar system could be costly. Fan *et al.* [13] developed a low-cost continuous wave (CW) radar system with two receivers and succeeded in measuring target position and motion. Molchanov *et al.* [14] proposed a technique for measuring gestures by combining a depth camera and frequency-modulated CW radar. Kim and Toomajian applied a convolutional neural network (CNN) to spectrogram images containing micro-Doppler information for the recognition of gestures [15]. Similar techniques using machine learning with spectrogram images have been used for a radar target classification [16], [17].

In the case of real-time systems, however, time-domain approaches are preferable because they do not require time-consuming time-frequency analysis. Kim *et al.* applied the CNN to the time-domain signals of an impulse-radio radar and the recognized gestures with accuracy exceeding 90% [18]. Gao *et al.* proposed an alternative approach of using barcode-like patterns generated from zero-crossing points of the time-domain waveform [19]. In this article, we propose a new time-domain gesture recognition technique using a low-cost 2.4-GHz CW radar and CNN. The proposed method applies CNN to in-phase/quadrature (I/Q) trajectory patterns of radar echoes and recognizes six types of hand gesture. The performance of the proposed method is evaluated using experimental radar data for ten participants. A preliminary result of this study has been reported in [20].

## II. SYSTEM MODEL

### A. Radar System

We use a monostatic CW radar system with an operating frequency of 2.4 GHz and transmitting power of 10.0 dBm. This radar system uses a fixed frequency of 2.4 GHz without modulation. The same antenna is used for transmitting and receiving, where the transmitting/receiving signals are isolated using a hybrid coupler. The antenna has a gain of 8.0 dBi, vertical polarization, and respective E- and H-plane beamwidths of 60.0° and 80.0°. The received signal

Corresponding author: Takuya Sakamoto (e-mail: t-sakamo@i.kyoto-u.ac.jp).
*(Xiaomeng Gao, Ehsan Yavari, Ashikur Rahman, Olga Boric-Lubecke, and Victor M. Lubecke contributed equally to this work.)*
Associate Editor: F. Costa.
Digital Object Identifier 10.1109/LSENS.2018.2866371

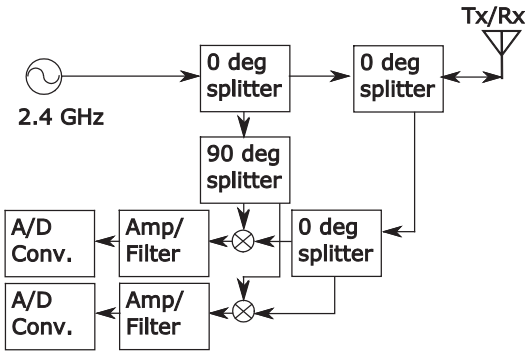


Fig. 1. Block diagram of the measurement setup.


Fig. 2. Measurement setup and a participant seated in an anechoic chamber.

is mixed with in-phase and quadrature signals and low-pass filtered, and analog-to-digital (A/D) converted to obtain in-phase (I) and quadrature (Q) signals, where the sampling frequency is 1.0 kHz.

The A/D converter is connected to the signal cable through dc coupling, and the A/D converted data contain dc components that are removed through dc subtraction in postprocessing. The dc subtraction does not distort I–Q plots, and thus, does not affect even slow-moving movements such as respiration and head movements. These slight movements can negatively affect the gesture recognition accuracy. A block diagram of the measurement setup is shown in Fig. 1.

### B. Measurement of Hand Gestures

We measured radar echoes from ten participants. The received signals contained mainly echoes from the arm and hand of the participants because echoes from stationary body parts were rejected by dc subtraction. Each participant was instructed to perform each of six gestures while remaining seated with his/her arm approximately 120.0 cm from the antennas. Each measurement took 2.0 s, and each gesture was repeated 150 times. The measurement setup is shown in Fig. 2. We denote by $s^p_{i,j}(t)$ the complex-valued time-domain signal from the $j$th measurement ($j = 1, 2, \ldots, N_0$) of the $i$th type of gesture ($i = 1, 2, \ldots, N_g$) performed by the $p$th participant ($p = 1, 2, \ldots, N_p$), where $N_0 = 150$, $N_g = 6$, and $N_p = 10$.

| | | | |
|---|---|---|---|
| (1) Stand-up gesture | | | |
| (2) Palm rotating | | | |
| (3) Fist and palm | | | |
| (4) Palm back and forth | | | |
| (5) Bye gesture | | | |
| (6) Push gesture | | | |

Fig. 3. Examples of radar-echo I–Q plot JPEG images.

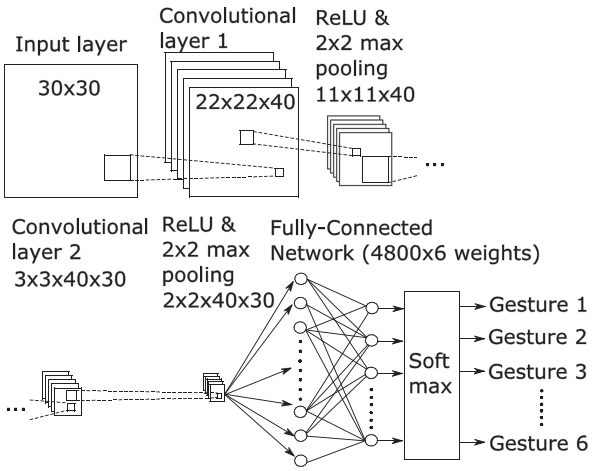


Fig. 4. Block diagram of the CNN.

## III. GESTURE RECOGNITION AND THE CNN

For gesture recognition, our proposed method uses the trajectory image of the I–Q plot of received signals $s^p_{i,j}(t)$ that are normalized so that $\max_{t,i,j}|s^p_{i,j}(t)| = 1$ for each $p$. The complex signal changes not only in phase but also in amplitude during the measurement. The trajectory images are converted to low-resolution JPEG images with a size of $N_s \times N_s$, where $N_s = 30$ pixels. For each participant, we measured each gesture $N_0$ times, generating $N_0 = 150$ JPEG images. Fig. 3 shows three example trajectory images for each of the six gestures.

Fig. 4 is a block diagram of the CNN used in this study. The input image with a size of $N_s \times N_s = 30 \times 30$ pixels is convoluted with 40 types of filters having a size of $5 \times 5$, resulting in 40 images having a size of $22 \times 22$. These images go through a rectified linear unit (ReLU) and max-pooling with nonoverlapping $2 \times 2$ pixels, resulting in 40 images with a size of $11 \times 11$. In the second convolution layer, these images are convoluted with 30 types of $5 \times 5$ filters, then subjected to an ReLU and max-pooling layers, and finally connected to six output neurons with a dense (fully connected) network, whose weights are optimized using the stochastic gradient descent with momentum (SGDM) optimization algorithm to minimize the difference between the training and output labels.

The CNN in Fig. 4 applies convolution and max-pooling twice for each, resulting in the final image size of $2 \times 2$, which means that the initial image size cannot be smaller than $30 \times 30$. Thus, we selected this image size as an input to make the CNN size small. Note that the CNN structure in this article is not optimized, and it will be important to optimize the CNN structure in future studies.

Table 1. Accuracy of the Proposed Method With the CNN Trained and Tested Using Nonoverlapping Data for the Same Single Participant (Columns 2, 3, and 4) and Accuracy of the Proposed Method With the CNN Trained Using Data for all Participants Except One and Tested on the Excluded Participant (Column 5).

| | Accuracy of the proposed method (%) | | | |
|---|---|---|---|---|
| Training &testing | Same subject | | | Different subjects |
| Training data size | $N_g N_{tr} = 90$ (10%) | $N_g N_{tr} = 450$ (50%) | $N_g N_{tr} = 810$ (90%) | $(N_p - 1)N_g N_{tr} = 8100$ |
| Subj. 1 | 84.1 | 89.8 | 91.6 | 51.3 |
| Subj. 2 | 83.7 | 89.6 | 90.4 | 39.0 |
| Subj. 3 | 88.8 | 93.2 | 94.9 | 32.7 |
| Subj. 4 | 77.9 | 85.7 | 88.2 | 36.5 |
| Subj. 5 | 88.2 | 92.6 | 94.7 | 36.4 |
| Subj. 6 | 94.7 | 97.6 | 98.6 | 43.6 |
| Subj. 7 | 80.6 | 85.8 | 89.0 | 45.1 |
| Subj. 8 | 82.5 | 90.0 | 93.1 | 40.7 |
| Subj. 9 | 78.6 | 85.9 | 85.0 | 42.6 |
| Subj. 10 | 73.7 | 83.2 | 87.1 | 16.8 |
| Average | 83.3 | 89.4 | 91.3 | 38.5 |
| $\alpha/\beta = 2$ | 82.4 | 88.9 | 91.3 | |
| $\alpha/\beta = 5$ | 78.8 | 85.9 | 88.7 | |
| | Accuracy of the single-layered CNN in [20] (%) | | | |
| Average | 78.8 | 84.9 | 87.6 | |

*Note:* Training and test datasets do not overlap, and are randomly selected multiple times for averaging accuracies.

## IV. PERFORMANCE EVALUATION

This section presents the application results of the proposed technique and evaluates the accuracy of the technique. We first investigate the gesture recognition accuracy when the CNN is trained using signals only from a single participant and is tested on a different subset of signals from the same participant, where datasets for training and testing do not overlap. For this purpose, we used $N_g N_{tr}$ images to train the CNN, where $N_{tr}$ is the training data size for each participant and gesture; the remaining $N_g(N_0 - N_{tr})$ images were used to evaluate the performance, where $N_{tr}(\leq N_0)$ was set to different values to see how the accuracy is affected by the training data size. In the training process, $N_g N_{tr}$ images were used to optimize the weights in the fully-connected network. The number of iterations of the SGDM optimization algorithm was empirically set to be 300.

The second, third, and fourth columns of Table 1 show the accuracies of the proposed method with the CNN trained using 10%, 50%, and 90% of all data (900 measurements) from a single participant and tested on the remaining data for the same participant. When training the CNN using 90% of the dataset of each participant, the average accuracy of the proposed method was 91.3%. It is noted that the accuracy depends on the training data size; the more data used for training, the higher the accuracy obtained.

We next investigate the applicability of the proposed method trained and tested using data for different people. The CNN was trained using data from $N_p - 1$ participants (i.e., all but one participant) and tested on the excluded participant, giving a training data size $(N_p - 1)N_g N_{tr} = 8100$. The accuracy in this scenario is shown in the rightmost column of Table 1. Although the accuracy is higher than that of random selection from the six gestures (1/6 = 16.7%), the average accuracy was only 38.4%, which is much lower than the accuracy of the proposed method trained and tested on the same single participant.

This result suggests that I–Q plots of the same gesture performed by multiple participants can appear to be different and that the CNN was unable to be trained well enough to recognize the gestures correctly, possibly because participants interpreted our instructions on how to perform gestures differently; the participants performed gestures in different ways, although they were given the same instruction. Therefore, the proposed system is suitable for personal use only with a single user; the system is not intended to be shared by multiple users.

We also investigate the performance of the proposed method when I/Q channels have unbalanced gains $\alpha$ and $\beta$. The average accuracies of the proposed method for $\alpha/\beta = 2$ and 5 are shown in Table 1, indicating that this method can tolerate a relatively large imbalance, especially when the training data size is sufficiently large.

Finally, we apply a single-layered CNN [20] for comparison instead of the multiple-layered CNN used above. The single-layered CNN uses input images of $16 \times 12$ pixels convoluted with ten types of $3 \times 3$ filters, which is followed by a ReLU, $2 \times 2$ max-pooling, and a fully connected network. Its average accuracies are shown in Table 1, which indicates that the multilayered CNN adopted in this study achieves a higher accuracy than the single-layered one [20].

Table 2. Accuracy of an Existing Method [15] Using the Time-Frequency Distribution With the CNN Trained and Tested for a Single Participant.

| | Accuracy of CNN in classifying gestures (%) | | |
|---|---|---|---|
| Training data size $N_g N_{train}$ | 90 (10% of all data) | 450 (50% of all data) | 810 (90% of all data) |
| Subject 1 | 83.3 | 89.4 | 91.9 |
| Subject 2 | 90.9 | 94.6 | 95.7 |
| Subject 3 | 89.3 | 94.2 | 95.7 |
| Subject 4 | 72.4 | 80.0 | 81.8 |
| Subject 5 | 92.6 | 96.6 | 96.9 |
| Subject 6 | 94.0 | 96.8 | 98.4 |
| Subject 7 | 68.6 | 80.9 | 86.6 |
| Subject 8 | 88.4 | 93.2 | 95.8 |
| Subject 9 | 92.7 | 95.9 | 97.2 |
| Subject 10 | 83.2 | 92.2 | 94.7 |
| Average | 85.6 | 91.4 | 93.5 |

## V. COMPARISON WITH AN EXISTING TECHNIQUE

This section compares the proposed method with an existing method [15], which we refer to as Kim's method in this article. Kim's method uses a spectrogram (time-frequency power distribution) as input data of a CNN. We use the same CNN architecture shown in Fig. 4 for both the proposed method and Kim's method. In Kim's method, a spectrogram is obtained using the short-time Fourier transform with a window size of $T_{FFT} = 256$ ms, and the spectrogram is normalized to its maximum value and converted to a decibel-scale image with a color range from $-10$ to 0 dB, which is resized to 30-by-30 pixels.

We applied Kim's method to the same data used in the previous section and evaluated its accuracy, as shown in Table 2, where the CNN was trained and tested using data of the same single participant. When

90% of data were used for training, the average accuracy of Kim's method was 93.5%, which was 2.2% higher than that of the proposed method. This is because spectrogram images used in Kim's method contain information of time and the Doppler frequency, whereas the I–Q plot images used in the proposed approach contain only amplitude and phase without a temporal information. Nonetheless, the difference in accuracies of the proposed method and Kim's method was less than 3%, while an advantage of the proposed technique is that the received signal can be directly used as an input of the CNN, whereas Kim's method requires preprocessing.

Because signals are sampled every $\Delta t = 1$ ms over $T_{\text{obs}} = 2.0$ s, to obtain a spectrogram when using Kim's method, the fast Fourier transform with a length of $N_{\text{FFT}} = 256$ must be applied $(T_{\text{obs}} - T_{\text{FFT}})/\Delta t + 1 = 1745$ times, which requires $1.8 \times 10^6$ complex-valued multiplications using the Cooley–Tukey algorithm. The proposed method can avoid such processing and still recognize gestures with accuracy higher than 90%. This means that the proposed approach can avoid preprocessing for time-frequency analysis, and thus, it is suitable for real-time applications. The computational time for generating a spectrogram image and an I–Q plot image were 1.4 and 0.10 ms, respectively, on a 64-bit Windows computer with Intel Core i7-4600U processor and 16 GB RAM.

Although we compared different algorithms using the same hardware system (and the same data) above, it will be necessary to also compare different hardware systems (e.g., different frequencies, modulation waveforms, and antenna types) for gesture recognition in future work. In the future, more comprehensive analysis will be needed to clarify the difference between single-user and multiuser results, including the special case when a user imitates another user's gesture.

## VI. CONCLUSION

We proposed a radar-based hand gesture recognition technique, which applies a CNN-based machine learning algorithm to time-domain I–Q plot trajectory images. The measurement data were analyzed to evaluate the accuracy in recognizing six different hand gestures for the ten participants. The proposed technique achieved average accuracy of 91.3% for the ten participants, which suggests the feasibility of gesture recognition using computationally inexpensive time-domain signal representation. Nonetheless, additional studies considering existing micro-Doppler-based techniques will be necessary to assess its real-time performance. In addition, a neural network itself can be computationally expensive, which must be also considered in such applications.

## ACKNOWLEDGMENT

This work was supported in part by the KAKENHI grants from the Japan Society for the Promotion of Science under Grant 25249057, Grant 15K18077, and Grant 15KK0243 and in part by the Center of Innovation Program of Kyoto University. Experiments were conducted according to the University of Hawaii Committee on Human Studies under Protocol Number 14884.

## REFERENCES

[1] J. Lien, "Soli: Ubiquitous gesture sensing with millimeter wave radar," in *Proc. 43rd Int. Conf. Exhib. Comput. Graph. Interactive Techn.*, 2016, vol. 35, Art. no. 142.
[2] Z. Lu, X. Chen, Q. Li, X. Zhang, and P. Zhou, "A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices," *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 2, pp. 293–299, Apr. 2014.
[3] A. Nelson, G. Singh, R. Robucci, C. Patel, and N. Banerjee, "Adaptive and personalized gesture recognition using textile capacitive sensor arrays," *IEEE Trans. Multi-Scale Comput. Syst.*, vol. 1, no. 2, pp. 62–75, Apr./Jun. 2015.
[4] P. G. Jung, G. Lim, S. Kim, and K. Kong, "A wearable gesture recognition device for detecting muscular activities based on air-pressure sensors," *IEEE Trans. Ind. Informat.*, vol. 11, no. 2, pp. 485–494, Apr. 2015.
[5] P. Pławiak, T. Sośnicki, M. Niedźwiecki, Z. Tabor, and K. Rzecki, "Hand body language gesture recognition based on signals from specialized glove and machine learning algorithms," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 1104–1113, Jun. 2016.
[6] H. P. Gupta, H. S. Chudgar, S. Mukherjee, T. Dutta, and K. Sharma, "A continuous hand gestures recognition technique for human-machine interaction using accelerometer and gyroscope sensors," *IEEE Sensors J.*, vol. 16, no. 16, pp. 6425–6432, Aug. 2016.
[7] Y. Wu, K. Chen, and C. Fu, "Natural gesture modeling and recognition approach based on joint movements and arm orientations," *IEEE Sensors J.*, vol. 16, no. 21, pp. 7753–7761, Nov. 2016.
[8] K. van Volkinburg and G. Washington, "Development of a wearable controller for gesture-recognition-based applications using polyvinylidene fluoride," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 4, pp. 900–909, Aug. 2017.
[9] H. Cheng, L. Yang, and Z. Liu, "Survey on 3D hand gesture recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1659–1673, Sep. 2016.
[10] N. Rossol, I. Cheng, and A. Basu, "A multisensor technique for gesture recognition through intelligent skeletal pose analysis," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 3, pp. 350–359, Jun. 2016.
[11] D. Wu *et al.*, "Deep dynamic neural networks for multimodal gesture segmentation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1583–1597, Aug. 2016.
[12] G. Zhu, L. Zhang, P. Shen, and J. Song, "Multimodal gesture recognition using 3-D convolution and convolutional LSTM," *IEEE Access*, vol. 5, pp. 4517–4524, 2017.
[13] T. Fan *et al.*, "Wireless hand gesture recognition based on continuous-wave Doppler radar sensors," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 11, pp. 4012–4020, Nov. 2016.
[14] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Short-range FMCW monopulse radar for hand-gesture sensing," in *Proc. Int. Conf. IEEE Radar*, 2015, pp. 1491–1496.
[15] Y. Kim and B. Toomajian, "Hand gesture recognition using micro-Doppler signatures with convolutional neural network," *IEEE Access*, vol. 4, pp. 7125–7130, 2016.
[16] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, May 2009.
[17] Y. Lang, C. Hou, Y. Yang, D. Huang, and Y. He, "Convolutional neural network for human micro-Doppler classification," in *Proc. Eur. Microw. Conf.*, 2017, pp. 497–500.
[18] S. Y. Kim, H. G. Han, J. W. Kim, S. Lee, and T. W. Kim, "A hand gesture recognition sensor using reflected impulses," *IEEE Sensors J.*, vol. 17, no. 10, pp. 2975–2976, May 2017.
[19] X. Gao, J. Xu, A. Rahman, E. Yavari, A. Lee, V. Lubecke, and O. Boric-Lubecke, "Barcode based hand gesture classification using AC coupled quadrature Doppler radar," in *Proc. IEEE MTT-S Int. Microw. Symp.*, 2016. doi: 10.1109/MWSYM.2016.7540013.
[20] T. Sakamoto, X. Gao, E. Yavari, A. Rahman, O. Boric-Lubecke, and V. Lubecke, "Radar-based hand gesture recognition using I-Q echo plot and convolutional neural network," in *Proc. Int. Conf. IEEE Antenna Meas. Appl.*, 2017, pp. 393–395, doi: 10.1109/CAMA.2017.8273461.