IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 22, 2016, accepted October 8, 2016, date of publication October 13, 2016, date of current version November 18, 2016.

Digital Object Identifier 10.1109/ACCESS.2016.2617282

Hand Gesture Recognition Using Micro-Doppler Signatures With Convolutional Neural Network

YOUNGWOOK KIM, (Senior Member, IEEE), AND BRIAN TOOMAJIAN

Corresponding author: Y. Kim (youngkim@csufresno.edu)

Department of Electrical and Computer Engineering, California State University, Fresno, CA 93740, USA

ABSTRACT In this paper, we investigate the feasibility of recognizing human hand gestures using micro-Doppler signatures measured by Doppler radar with a deep convolutional neural network (DCNN). Hand gesture recognition using radar can be applied to control electronic appliances. Compared with an optical recognition system, radar can work regardless of light conditions and it can be embedded in a case. We classify ten different hand gestures, with only micro-Doppler signatures on spectrograms without range information. The ten gestures, which included swiping from left to right, swiping from right to left, rotating clockwise, rotating counterclockwise, pushing, double pushing, holding, and double holding, were measured using Doppler radar and their spectrograms investigated. A DCNN was employed to classify the spectrograms, with 90% of the data utilized for training and the remaining 10% for validation. After five-fold validation, the classification accuracy of the proposed method was found to be 85.6%. With seven gestures, the accuracy increased to 93.1%.

INDEX TERMS Hand gesture, micro-Doppler signatures, Doppler radar, deep convolutional neural networks.

I. INTRODUCTION

Recognizing human hand gestures can facilitate a number of important applications in the area of electronic device control, biomechanics research, computer gaming, and defense. Controlling devices without physical contact has the advantage of convenience for a user. In particular, hand gestures can be used as an input modality in automobiles where, for safe driving, physical contact with buttons is highly undesirable. In addition, it could be used in small-sized electronic devices, instead of a small button being employed. Replacing buttons in devices can also improve the reliability and design flexibility of products.

In the hand gesture recognition research area, several optical sensor methods, in which vision or depth camera are utilized, have been proposed [1]–[3]. Optical sensors have a high resolution that enables tracking and recognition of the motions of the finger and wrist. Temporal pattern recognition techniques such as hidden Markov model are usually employed to classify the gestures in these cases. Using a camera system, overall accuracies of more than 90% were obtained for 20 gestures [2]. With a depth camera sensor, recognition accuracy of 93.9% has been achieved for 10 motions [4]. Acoustic Doppler has also been used to detect hand motion, with accuracy of approximately 94% for five motions. However, it requires continuous transmission of audible waves, which is reported to be annoying [5]. Methods that recognize human hand gestures using radar have also been proposed [6]–[8]. Unlike optical applications, radar usage is not restricted by lighting condition. In addition, a miniaturized radar sensor can be embedded inside devices because radar has through-object capability. Embedding a sensor inside a device enables easier maintenance as well as robust operation as the possibility of buttons malfunctioning is avoided. Thus far, pulsed radar and frequency-modulated continuous-wave (FMCW) radar have been employed to measure the range to the fingers in order to track their motion. Using FMCW radar, the accuracy was approximately 89% for 10 gestures [7]. However, to the best of our knowledge, Doppler radar has never been applied for hand gesture recognition. Using Doppler sensors would result in a simple, cost effective, and easy approach to capturing radial velocity response.

In this letter, we investigate the feasibility of recognizing hand gestures using only micro-Doppler signatures, i.e., no range information is utilized. When a target has non-rigid body motions, micro-Doppler signatures are generated. These micro-Doppler signatures have served as features for recognizing humans and their motions [9]–[11]. However, micro-Doppler signatures are represented as overlapped signatures in the joint time-frequency domain when several scatterers, such as fingers, exist. As a result, they have to be carefully investigated in detail in order to distinguish revealing signatures associated with gestures. This study focused on analyzing micro-Doppler signatures in spectrograms from diverse hand gestures and investigated the feasibility of classifying them based on the measured signatures. Ten hand gestures, including swiping, rotating, pushing, and holding, were investigated. Because the spectrograms of these gestures only have subtle differences, instead of the conventional supervised learning approach, a powerful classifier is necessary. We employed deep convolutional neural network (DCNN) for spectrogram-based hand gesture classification. DCNN, which is inspired by the human visual cortex, is one of the most successful deep learning algorithms [12]–[14]. It has been effectively used in the field of image recognition. By training the convolutional filter and the fully connected multi-layer perceptrons, DCNN simultaneously extracts and classifies important features. DCNN does not require a handcrafted feature extraction process. Consequently, it can be employed in any image classification application. Because micro-Doppler signatures are represented as images in spectrograms, we were able to apply DCNN in this study as well. In this letter, we report on the measurement setup, measured data, DCNN, and classification results obtained.

II. EXPERIMENTAL SETUP AND MEASUREMENT

We employed Doppler radar in order to obtain micro-Doppler signatures of ten hand gestures from a single participant. Bumblebee Doppler radar (Samraksh Co. Ltd.), which operates at 5.8 GHz, was employed. This radar produces a quadrature output at an average output power of 4.5 dBm. It has an antenna beam width of 60 degrees, and responds to radial velocities between 2.6 cm/s and 2.6 m/s, which make it suitable for detecting hand gestures. We affixed the radar to a table and executed the hand motions in the main lobe of the radar antenna. The average distance from the radar to the hands was approximately 10 cm.

The ten hand gestures employed in this study were (a) swiping from left to right, (b) swiping from right to left, (c) swiping from up to down, (d) swiping from down to up, (e) rotating clockwise, (f) rotating counterclockwise, (g) pushing, (h) holding, (i) double pushing, and (j) double holding. The employed gestures are depicted in Fig. 1. The upper half of each picture is a snapshot of the starting posture, whereas the bottom half is that of the ending posture. Because double pushing and double holding are repeats of each motion, we omitted the corresponding pictures.

Because the Doppler device only detects radial velocity, the motions depicted in Figs. 1(a)–(d) required a slight variance from each other. Swiping left to right, Fig. 1(a), was a quick snap that involved the wrist and all fingers together. For swiping right to left, the wrist was no longer stationary and moved

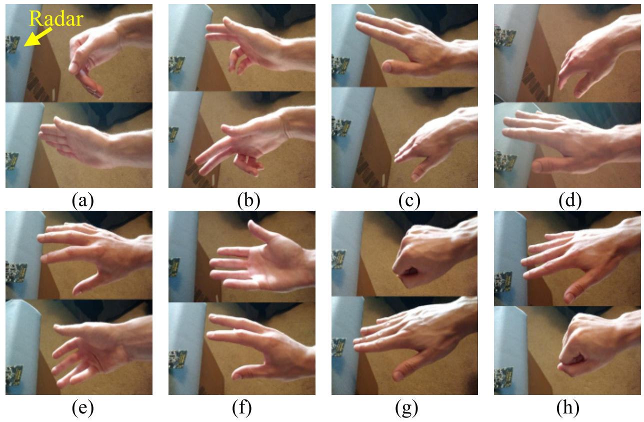

FIGURE 1. Eight of the ten hand gestures measured using Doppler radar: (a) swiping from left to right, (b) swiping from right to left, (c) swiping from up to down, (d) swiping from down to up, (e) rotating clockwise, (f) rotating counterclockwise, (g) pushing, and (h) holding.

a few inches during the motion. Another defining feature of this motion is the finger position, which involved only three fingers facing the Doppler instead of five. Similarly, the up to down swiping motion had the same finger positioning but the total distance traveled by the wrist was nearly twice as much as that from right to left. The swiping down to up motion involved all five fingers, but necessitated a more significant change in wrist positioning than all the other motions. Clockwise and counterclockwise are also very similar to each other; both have all five fingers towards the Doppler with the only changes stemming from the starting position of counterclockwise and ending position of clockwise. The counterclockwise rotation starts with the palm facing upwards and ends in the same starting point as clockwise. Finally, pushing and holding are opposite gestures, as shown in the figure. The differences between push and hold are simply based on whether the fingers are together or separated. In pushing and double pushing, the fingers are kept together through the full motion, whereas in holding and double holding the fingers are separated.

To investigate micro-Doppler signatures, spectrograms of finger motions were observed through short-time fast Fourier transform (FFT). We set the size of the FFT to 256 ms and the time step of non-overlapping samples to 1 ms. Examples of the ten spectrograms obtained are shown in Fig. 2. It can be seen that the micro-Doppler signatures of a single hand motion occur within 200 ms–300 ms, and they exhibit marginally different features in the joint time-frequency domain. Each gesture was measured 50 times; consequently, we obtained 500 pieces of data in total. From a Doppler radar perspective, gestures (a), (b), (c), and (d) are almost similar because their radial velocities are analogous even though the directions of motion are different. However, because of the small variation in hand movements described in Section II, peculiar features that distinguish them can be observed. To differentiate those subtle differences, a powerful image recognition technique, rather than classification methods based on handcraft features, is necessary.

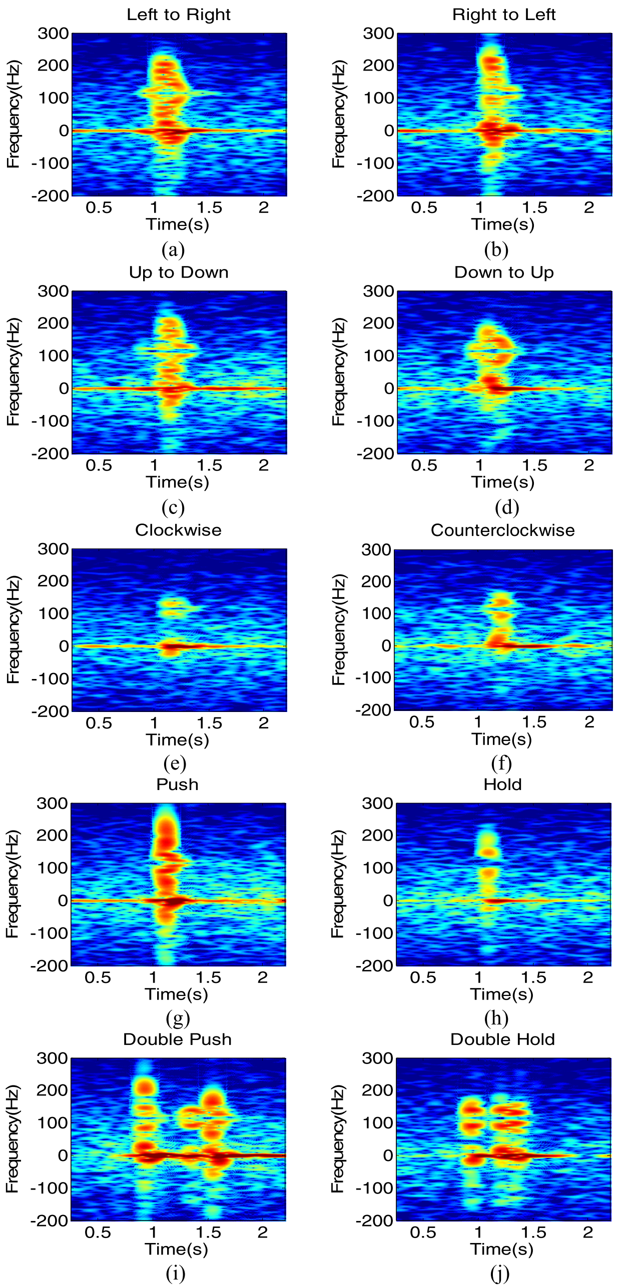

FIGURE 2. **Spectrograms of eight hand gestures: (a) swiping from left to right, (b) swiping from right to left, (c) swiping from up to down, (d) swiping from down to up, (e) rotating clockwise, (f) rotating counterclockwise, (g) pushing, (h) holding, (i) double pushing, and (j) double holding.**

III. DEEP CONVOLUTIONAL NEURAL NETWORKS (DCNNS)

To classify targets based on spectrograms, a process comprising feature extraction followed by classification is usually employed. There are five main approaches to feature extraction: i) handcraft features [10], ii) empirical mode decomposition [11], iii) linear predictive code [15], iv) principal component analysis [16], and v) DCNNs [17], [18]. Of these approaches, DCNNs exhibit the best classification accuracy. Consequently, we employed DCNNs to classify our hand gestures. Whereas most classification algorithms utilize a normalization process, which is quite cumbersome and can be subjective, DCNN does not require signature normalization as it can recognize a target regardless of the location.

Among deep learning algorithms, DCNN is regarded as one of most powerful classifiers and it has been successfully used in image recognition applications. Compared to other conventional machine learning algorithms, DCNN employs a multi-layer structure to improve generalization and abstraction performance. Because of limited computational resources and diminishing backpropagation error, in the past it was not feasible to train such a multi-layer structure. However, with the development of graphics processing units that can carry out parallel processing and improve training algorithms, deep learning models can now be properly trained.

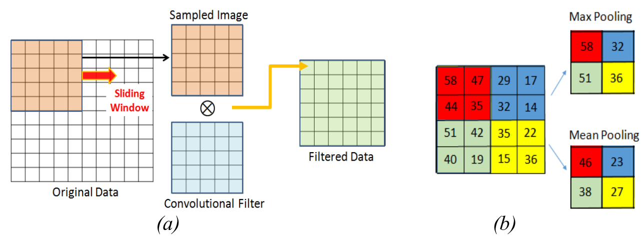

FIGURE 3. **(a) Process of applying a 5-by-5 convolution filter to the input data (in orange) to generate the output (in green). (b) Examples of 2-by-2 pooling (max-pooling or mean-pooling) that reduces the data dimension by half.**

A DCNN primarily consists of a convolutional filter, an activation function, and a pooling layer. The combination of convolution filters, activation function, and pooling constitutes one layer and multiple layers are consecutively connected in a DCNN. The convolutional filter extracts the features of a spectrogram through its convolution process, as shown in Fig. 3(a). The coefficients of the convolutional filter are trained by a given dataset. The number of convolution filters is determined empirically; for spectrogram recognition, this number is usually in the range five to twenty. The convolutional filter is followed by an activation function. This activation function is highly nonlinear such that it can describe the nonlinear relationship between inputs and outputs. Instead of the sigmoid function, deep learning employs restricted linear units, $f(x) = \max(0, x)$, because it can achieve better empirical results [19] owing to its piecewise linear characteristics. The third stage is a pooling layer that is used for data dimension reduction. This pooling layer enables the final output to be more robust to noise. Pooling can be performed by selecting a maximum value or a mean value, as shown in Fig. 3(b).

Finally, a general perceptron is connected in the last layer for classification purposes. Fig. 4 shows the architecture of a simple DCNN, which consists of three convolution layers and one final fully connected layer. A DCNN also omits hidden nodes through a predetermined probability that is independent of the test samples. The dropout [20] of these nodes is

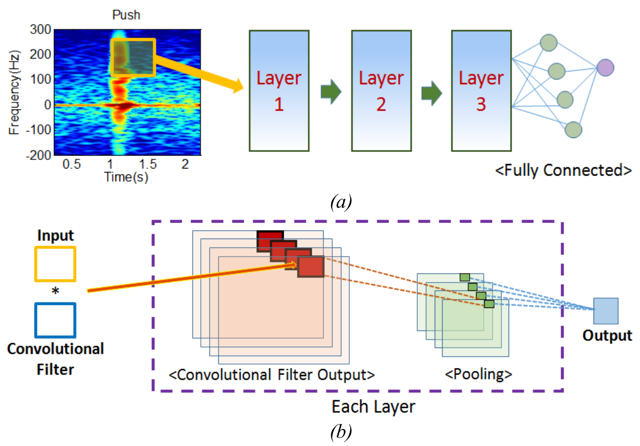

FIGURE 4. **(a) Structure in DCNN with three layers, and (b) Structure of each layer.**

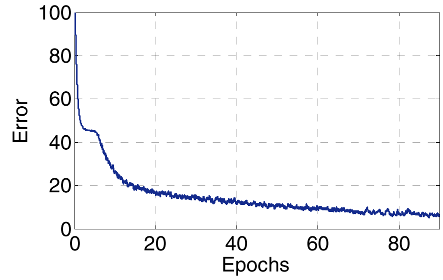

FIGURE 5. **Test error curve with epochs.**

TABLE 1. **Accuracies of the DCNN for each fold and their average.**

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average
88%	84%	90%	84%	82%	85.6%

TABLE 2. **Confusion matrix (%).**

Ac *Es*	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)
(a)	76	20	0	0	0	0	0	0	4	0
(b)	8	48	8	0	0	0	0	0	4	4
(c)	4	16	80	0	0	0	0	0	0	0
(d)	4	0	0	92	12	0	0	0	0	0
(e)	0	0	4	4	88	0	0	0	0	0
(f)	0	0	0	0	0	100	0	0	0	0
(g)	0	0	0	0	0	0	100	0	0	0
(h)	4	0	0	4	0	0	0	100	4	0
(i)	4	0	4	0	0	0	0	0	80	4
(j)	0	16	4	0	0	0	0	0	8	92

used to prevent overfitting in a regularization scheme. This enables the neural network to prevent co-adaption among its nodes.

The coefficients of convolution filters and weighting values of the final fully connected layers, are trained via a dataset. A backpropagation algorithm with a stochastic gradient descent (SGD) is usually used as a training algorithm. Trained convolutional filters work as a feature extractor, and the fully connected perceptron functions as a classifier. Even though DCNNs are predominantly applied to RGB images, they are also effective in spectrogram recognition applications [17], [18].

IV. HAND GESTURE CLASSIFICATION

In our study, to generate the training dataset, a two-second time window was used to crop the spectrograms. Then, each spectrogram was resized to 60-by-60 and the values normalized from zero to one. Among the 500 pieces of data measured from the single participant, we used 90% as training data and 10% as testing data. We used 5-fold validation to obtain valid accuracy by dividing the measured data into five different training datasets and test datasets. For the DCNN structure, we used three layers, with five, four, and two convolutional filters. At each layer, the convolutional filter had dimensions 5-by-5 and the reduction ratio was 2:1 in all pooling layers. The number of layers and filters were empirically optimized for the highest accuracy. In the training process, the iteration was set to 90 because the error saturates. The batch size was two. The training error curve for Fold 1 is shown in Fig. 5.

For the investigated ten activities, the averaged classification accuracy was 85.6%. The 5-fold validation accuracies are shown in Table I. For the analysis of misclassification, the confusion matrix for all 5-folds is presented in Table II. The values in the table are for classification accuracy (%). From the table, it can be seen that gestures (a) and (b) each have a high inaccuracy, as expected, because there were similarities between activities (a), (b), (c), and (d). We evaluated the performance of DCNNs again with a total of eight activities by omitting activities (a) and (b). With the eight activities, the accuracy increased to 91.4%. With seven gestures, by omitting (a), (b), and (c), the accuracy reached 93.1%.

V. CROSS-VALIDATION THROUGH DIVERSE SCENARIOS

In the previous measurements, it was assumed that the DCNN is trained by data from a particular user and are used to recognize that user's hand gesture in a controlled environment. The hand gestures of a specific user should be recorded first to train the machine before the system is actually used. How the trained DCNN would effectively recognize hand gestures in uncontrolled environments is an interesting question. Therefore, we set up four practical scenarios. For each scenario, we measured data three times for seven gestures and investigated the classification accuracy. The four scenarios included measurements i) with different incident angles, ii) with different aspect angles, iii) with different distances, and iv) with a second participant.

In the first scenario, we measured the hand gestures of the participant when the incident angle to the radar was −45 degrees and +45 degrees, as shown in Fig. 6(a). In this case, the Doppler signatures did not change, as it is the same situation as the previous case from the Doppler radar point of view. However, we noticed that the signal to noise ratio (SNR) of the Doppler signal decreased because the hand was not

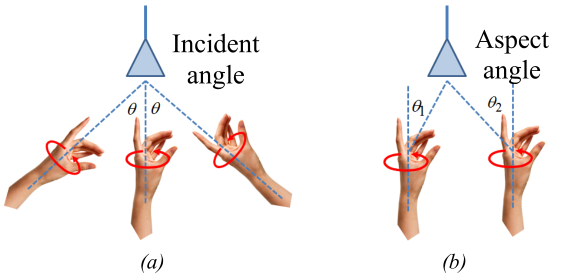

FIGURE 6. (a) With zero aspect angle, (b) With non-zero aspect angle.

TABLE 3. Classification accuracy with distance (%).

	15 cm	30 cm	45 cm	60 cm
Accuracy	85.8%	90.5%	71.5%	66.7%

aligned with the main lobe of the radar antenna. The antenna has a limited beam width of 60 degrees such that the received power decreases with the incident angle.

In the second scenario, we measured hand gestures for non-zero aspect angles. In this case, the hand had an offset in position to make the aspect angle. When $\theta_1$was 15 degrees and θ_2 45 degrees, we measured the data from the participant and calculated the accuracy. The accuracies were 81.12% and 57.5%, respectively, when the previously trained DCNN was utilized. The aspect angle causes variations in the micro-Doppler signatures because Doppler corresponds to radial velocity only, which is calculated by the cosine term [21], [22]. In addition, the signature became attenuated due to the low SNR as the hand was not aligned to the antenna's main lobe.

In the third scenario, we measured the gestures with different distances when the aspect angle of the hand was zero. The measured distances were 15 cm, 30 cm, 45 cm, and 60 cm. As shown in Table III, in general, the classification accuracy decreased with distance because of the low SNR and the variation of micro-Doppler signatures.

In the fourth scenario, we measured the seven gestures from the second participant with the same conditions as before. When the previously trained DCNN was used, the classification accuracy was 71.5%. This accuracy reduction is a result of the gesture difference between users. Consequently, when the data from the second participant was included in the training dataset, the accuracy increased to 90.48% with a newly trained DCNN. This test implies the possibility of implementing a user-independent classifier trained by a massive dataset comprising data from multiple human subjects.

VI. CONCLUSION

In this study, we investigated the feasibility of a proposed method that classifies human hand gestures using micro-Doppler signatures with a DCNN. Ten hand gestures were measured using Doppler radar and their spectrograms analyzed. The DCNN was employed to classify the micro-Doppler signatures of the hand gestures. The classification accuracy of the proposed method was found to be 85.6% for ten gestures. With seven gestures, the accuracy increased to 93.1%. Because high classification accuracy is required in practical applications, it would be reasonable to use the suggested seven gestures when Doppler radar is employed. However, it is possible to include more gestures if they produce unique signatures in a spectrogram.

However, it should also be noted that micro-Doppler signatures can vary depending on aspect angle and distance to the radar, as shown in our experiments. For robust and practical operation, data from diverse scenarios should be included in the measurement process. In addition, multiple human subjects should be measured to construct a user-independent classifier. As the diversity and complexity of data continue to increase, the use of DCNN as a classifier will become more suitable as a result of its powerful learning capability. In future work, we plan to measure many gestures from various human subjects to train a DCNN for general-purpose hand gesture recognition.

REFERENCES

[1] B.-W. Min, H.-S. Yoon, J. Soh, Y.-M. Yang, and T. Ejima, "Hand gesture recognition using hidden Markov models," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, vol. 5. Oct. 1997, pp. 4232–4235.

[2] F.-S. Chen, C.-M. Fu, and C.-L. Huang, "Hand gesture recognition using a real-time tracking method and hidden Markov models," *Image Vis. Comput.*, vol. 21, no. 8, pp. 745–758, Aug. 2003.

[3] Y. Yao and C.-T. Li, "Hand gesture recognition and spotting in uncontrolled environments based on classifier weighting," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 3082–3086.

[4] Z. Ren, J. Meng, and J. Yuan, "Depth camera based hand gesture recognition and its applications in human-computer-interaction," in *Proc. Int. Conf. Inf., Commun. Signal Process.*, Dec. 2011, pp. 1–5.

[5] S. Gupta, D. Morris, S. Patel, and D. Tan, "SoundWave: Using the Doppler effect to sense gestures," in *Proc. ACM SIGCHI Conf. Human Factors Comput. Syst.*, May 2012, pp. 1911–1914.

[6] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Short-range FMCW monopulse radar for hand-gesture sensing," in *Proc. IEEE Radar Conf.*, May 2015, pp. 1491–1495.

[7] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Multi-sensor system for driver's hand-gesture recognition," in *Proc. IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, May 2015, pp. 1–8.

[8] P. Hagler, M. Geiger, and C. Waldschmidt, "RCS measurements of a human hand for radar-based gesture recognition at E-band," in *Proc. German Microw. Conf.*, Mar. 2016, pp. 259–262.

[9] D. Tahmoush and J. Silvious, "Radar micro-Doppler for long range front-view gait recognition," in *Proc. IEEE 3rd Int. Conf. Biometrics, Theory, Appl., Syst.*, Washington, DC, USA, Sep. 2009, pp. 1–6.

[10] 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.

[11] D. Fairchild and R. Narayanan, "Classification of human motions using empirical mode decomposition of human micro- Doppler signatures," *IET Radar, Sonar, Navigat.*, vol. 8, no. 5, pp. 425–434, Jun. 2014.

[12] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Proces. Syst.*, vol. 25. 2012, pp. 1090–1098.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[15] R. J. Javier and Y. Kim, "Application of linear predictive coding for human activity classification based on micro-Doppler signatures," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1831–1834, Oct. 2014.

[16] Y. Kim, "Detection of eye blinking using Doppler sensor with principal component analysis," *IEEE Antennas Wireless Propag. Lett.*, no. 14, pp. 123–126, 2015.
[17] Y. Kim and T. Moon, "Human detection and activity classification based on micro-doppler signatures using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 2–8, Jan. 2016.
[18] Y. Kim and T. Moon, "Classification of human activity on water through micro-Dopplers using deep convolutional neural networks," *Proc. SPIE*, vol. 9829, p. 982917, Apr. 2016.
[19] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, Fort Lauderdale, FL, USA, 2011, pp. 315–323.
[20] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[21] F. Fioranelli, M. Ritchie, and H. Griffiths, "Aspect angle dependence and multistatic data fusion for micro-Doppler classification of armed/unarmed personnel," *IET Radar, Sonar Navigat.*, vol. 9, no. 9, pp. 1231–1239, Dec. 2015.
[22] D. P. Fairchild and R. M. Narayanan, "Multistatic micro-Doppler radar for determining target orientation and activity classification," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 1, pp. 512–521, Feb. 2016.

YOUNGWOOK KIM (S'03–M'08–SM'14) received the B.S. degree in electrical engineering from Seoul National University, South Korea, in 2003, and the M.S. and Ph.D. degrees in electrical and computer engineering from The University of Texas at Austin, USA, in 2005 and 2008, respectively. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, California State University, Fresno. He has authored over 50 technical papers. His research interests are in the area of radar signal processing, antenna design, and RF electronics. His primary topic of research lies in radar target classification with machine-learning techniques. Currently, remote detection, monitoring, and analysis of human motions using EM wave have been focused in his research. He is a recipient of the Provost's New Faculty Award from the California State University, the A. D. Hutchison Fellowship from The University of Texas at Austin, and the National IT Fellowship from the Ministry of Information and Communication, South Korea.

BRIAN TOOMAJIAN is currently pursuing the bachelor's degree in electrical and computer engineering from California State University, Fresno. His current research interests include electromagnetic sensors, sensor measurement, and sensor signal processing. Especially, he is researching on the joint time-frequency analysis of sensor signals.