Classification of human activity on water through micro-Dopplers using deep convolutional neural networks

Youngwook Kim,*a Taesup Moonb

aElectrical and Computer Engineering, California State University, Fresno, 2320 E San Ramon Ave EE/94, Fresno, CA, USA 93720; bDepartment of Information and Communication Engineering, DGIST, 333 Techno Jungang-daero, Daegu, South Korea, 42988

ABSTRACT

Detecting humans and classifying their activities on the water has significant applications for surveillance, border patrols, and rescue operations. When humans are illuminated by radar signal, they produce micro-Doppler signatures due to moving limbs. There has been a number of research into recognizing humans on land by their unique micro-Doppler signatures, but there is scant research into detecting humans on water. In this study, we investigate the micro-Doppler signatures of humans on water, including a swimming person, a swimming person pulling a floating object, and a rowing person in a small boat. The measured swimming styles were free stroke, backstroke, and breaststroke. Each activity was observed to have a unique micro-Doppler signature. Human activities were classified based on their micro-Doppler signatures. For the classification, we propose to apply deep convolutional neural networks (DCNN), a powerful deep learning technique. Rather than using conventional supervised learning that relies on handcrafted features, we present an alternative deep learning approach. We apply the DCNN, one of the most successful deep learning algorithms for image recognition, directly to a raw micro-Doppler spectrogram of humans on the water. Without extracting any explicit features from the micro-Dopplers, the DCNN can learn the necessary features and build classification boundaries using the training data. We show that the DCNN can achieve accuracy of more than 87.8% for activity classification using 5-fold cross validation.

Keywords: micro-Dopplers, target classification, human activity, target on water, deep learning, convolutional neural networks

1. INTRODUCTION

Increased demand for security, border patrols, and rescue operations has accelerated research into human detection and classification of their activities [1-2]. Human detection in urban areas can be applied to security and law enforcement. For border patrols, the automatic detection of a human presence is necessary to monitor large areas. In addition, detecting and localizing a human subject in rescue operations is critical.

There has been diverse and extensive research into ground detection, including human detection, activity classification, gait recognition, and motion analysis based on micro-Doppler signatures [3-5]. When human limbs are illuminated by electromagnetic waves, each body part generates its own Doppler signal, resulting in modulated components on the Doppler from a torso. Because the micro-Doppler signatures are unique to the human activity, they provide significant information for classifying such activities. Compared to optical sensors, electromagnetic sensors can operate in all weather conditions, regardless of lighting. However, human detection and activity classification on water have not yet been addressed.

In this paper, we investigate the micro-Doppler signatures of humans to classify their activities on water. Human detection and classification on water can be applied to border patrols and rescue operations. In particular, since monitoring ocean at night or on foggy day using optical sensors is challenging, automatic detection using radar is desirable. In this study, we measure a human subject performing activities on water using Doppler radar. Doppler radar can detect a moving object that has a radial velocity while suppressing clutter. Human motions on water are captured by Doppler radar, and the micro-Doppler signatures are investigated through a joint time frequency analysis. The activities measured are a swimming person, a swimming person pulling a floating object, and a rowing person in a small boat. The
measured swimming styles include free stroke, backstroke, and breaststroke. Compared to the micro-Doppler signatures from human activity on ground, it is more difficult to design handcrafted features in this case because there are more variations in human motions on water. In addition, the signatures sometimes become unclear due to water waves and water drops. Therefore, to classify human activities, we employ a deep convolutional neural network (DCNN) [6]. DCNN has been successfully used to recognize objects in images and outperformed against most other algorithms in image recognition. After training, DCNN finds a convolutional filter that effectively detects objects included in the training data. Because the spectrogram of measured human motions on water can be regarded as an image, we can apply DCNN. We designed DCNN with three convolutional layers, three sub-sampling layers, and multi-layer perceptron. Based on the designed DCNN, we conducted training using the measured data, and classification accuracies were found.

2. MEASUREMENTS

We measured a human subject in a swimming pool to detect human motion only. In a sea or lake, water waves significantly affect the Doppler signal, making it difficult to discriminate them from human motion. Because we only want to focus on human signatures in this study, we do not intend on considering the water wave. When a human subject approaches a radar system, Doppler radar captures human motions in water. The Doppler radar we used operates at 7.25GHz, with an output power of 15dBm. The radar consists of two Vivaldi antennas for transmission and reception. The received signals are directly converted to Doppler signal through mixing with a carrier signal. The radar has an \( I \) channel and a \( Q \) channel to capture the complex characteristics of the received signal. The down converted output is sampled at a speed of 1Ksps in accordance with the National Instrument Data Acquisition Board (DAQ-6001). We used vertical polarization assuming that the human motion, especially arm motion, effectively interacts with illuminated EM waves. The block diagram of Doppler radar and its picture are shown in Figure 1.

![Diagram of Doppler radar](image)

Figure 1: (a) Diagram of Doppler radar, and (b) picture of the radar.

We measured five activities: a swimming person doing the free stroke, backstroke, and breaststroke, a swimming person pulling a floating object, and a rowing person in a small boat. The five captured motions are shown in Figure 2. In the free stroke, the subject rotated his head for each stroke. In the back stroke, the subject tried to stretch his arm as straight as possible. In the breast stroke, the arms and legs were mainly under the water. When pulling a boat, the subject performed the breast stroke with a single arm while another hand held the boat. In the rowing activity, the subject alternated between rowing on left and right sides.
The received signal was processed with the joint-time frequency analysis to investigate its time-varying micro-Doppler characteristics. To observe the Doppler signatures, spectrograms were generated, examples of which are presented in Figure 3. We used the short time Fourier transform. When performing the spectrogram, the FFT size was set at 256, and the non-overlapping step size was 20 msec. Each activity was observed to possess unique micro-Doppler signatures. The spectrogram of the free style showed sharp micro-Dopplers from arm motions. In the back stroke, the arm motion was slower than that of free style, so the maximum Doppler frequency is lower. The breast stroke did not show significant micro-Dopplers due to the difficulty detecting limb motions under the water. Mainly, the speed of the subject’s head is captured. We observed that the speed is periodic with fluctuation due to the periodic arm motions. When the subject used the breast stroke while pulling the boat, the Doppler pattern was similar to that of the breast stroke. The magnitude of the signal was high because the RCS of a boat is higher than a human head. In the rowing activity, the paddles’ movement generated micro-Dopplers.

To construct the training and test data sets, we measured a single subject five times for each activity. From each measurement, we extracted five spectrograms for 2 seconds (100 pixels) with overlapping time windows. In the cropped spectrogram, the Doppler frequency was between 0Hz and 500Hz (256 pixels). The negative frequency does not contain significant information because the human subject was approaching to the radar in the measurement.
Figure 3: Spectrograms of (a) free stroke, (b) back stroke, (c) breast stroke, (d) swimming person pulling a floating boat, and (e) rowing person.

3. CLASSIFICATION USING DEEP CONVOLUTIONAL NEURAL NETWORKS

3.1 Deep convolutional neural networks

Deep learning is a sub-field of machine learning that has recently had tremendous impact on several application areas such as speech recognition, image recognition, natural language processing, and games [7-10]. In contrast to the conventional supervised learning approach, which makes predictions based on hand-crafted features extracted from data, deep learning algorithms find both features and prediction boundaries by optimizing a certain objective function. For this purpose, these algorithms employ a deep neural network structure, which stacks multiple layers of simple neural network architecture, to automatically extract hierarchical features from the data. While the main idea of deep learning was proposed in the 1980s, it regained the spotlight in the last decade due to its impressive results in the abovementioned applications, significantly outperforming previously state-of-the-art schemes. There are two main reasons for such a renaissance: the advent of high-performance computing processors such as the graphic processing unit (GPU) and continued algorithmic innovations. A more detailed overview of deep-learning algorithms can be found in [12] and the references therein.

Proposed by LeCun [13] in the late 1980s, DCNN is one of the most successful deep learning algorithms for processing images. It is a supervised learning algorithm that attempts to learn mapping between the input data point and its corresponding label, which is provided by a human annotator. It was shown to be very effective, particularly for image classification. DCNN’s hierarchical structure is inspired by the visual cortex of a human brain that efficiently recognizes objects. The convolutional filer is designed to be analogues to the visual cortex. The DCNN consists of three main components. The first component is the convolution filter, which works on small local receptive fields of input data in a sliding-window fashion. Each filter can be thought of as a specific feature detector, and multiple convolution filters exist in each layer that work in parallel. The second component is the nonlinear activation function imposed on the
convolution filter output, which enables nonlinear transformation of a data space so that discrimination among classes can become easier. The conventional choice for the activation function is the sigmoid function: \( f(x) = \frac{1}{1 + \exp(-x)} \).

More recently, restricted linear units (ReLU), \( f(x) = \max(0, x) \) were shown to achieve better empirical results when used as an activation function [14]. The third component is the pooling, which reduces the data size by taking the minimum or mean values of small patches, for example, 2-by-2 patch. The final prediction can become robust to the translation of input data through pooling. The combination of convolution filters, nonlinear activation, and pooling are regarded as one layer in DCNN, and multiple layers (sometimes more than 100) are consecutively employed in modern CNNs; hence, the term deep CNN. When the DCNN is used as a classifier, the final few layers usually assume the form of a fully connected multi-layer perceptron (MLP), that is, all of the layer’s input nodes are connected to all of the output nodes. Fig. 4 shows a simple DCNN architecture with two convolution layers and one final fully connected layer.

![Fig. 4. An example structure of a DCNN](image)

The DCNN parameters—the coefficients of convolution filters and the final fully connected layers—can be trained by back-propagation with stochastic gradient descent (SGD). To speed up the optimization process, a momentum method [15] is used as a standard choice. In addition, to prevent over-fitting, dropout [16] is widely used as a regularization scheme, typically in the fully connected layers.

While the DCNN has recently revolutionized the image recognition community by significantly outperforming the previous state-of-the-art methods, most of the advanced DCNN developed so far has focused on natural RGB image recognition problems. To the best of our knowledge, [11] was the first to apply DCNN to micro-Doppler signatures by treating them as an image and training a classifier. Moreover, by comparing Figure 3 and Figure 7 of [11], we can see that the spectrograms of human activities in water is different from those on ground. Hence, if we were to follow the conventional hand-crafted feature-based supervised learning methods in [3], we need to go over the feature engineering steps to identify informative features that are critical in this problem. However, in the next section, we show that an almost identical DCNN structure used in [11] can be applied again to classify activities on water and still achieve high accuracy.

### 4.2 Training with DCNN

For the experiments, we follow the approach of [11]. As described above, we have a total of 125 data samples (i.e., spectrograms), which consist of 5 actions with 25 samples for each action. The dimension of each spectrogram was 252 (frequency) by 100 (time). We randomly split the data into 5 folds, and each fold selected 5 random data samples from each of the 5 actions. Therefore, the training and test set in each fold contained 100 and 25 samples, respectively. We used almost the same model configuration in [11]: We used three convolution layers, in which each convolution layer had 20 filters of 5-by-5 in size, ReLU activation functions, and 2-by-2 max pooling layers. Moreover, we used one fully connected layer with 500 hidden nodes, followed by a softmax classifier. We used mini-batch SGD with a learning rate of 0.001 and a batch size of 100. The parameter of the momentum method was 0.9, and a dropout rate at the fully connected layers was set at 0.5. The maximum iteration of the SGD update was 1000. In our model configuration, we used padding on the boundary data to utilize the boundary data points and in the convolution process.

Since the model’s hyper-parameters—size of convolution filters, number of convolution filters, number of layers, and so forth—were fixed, we employed cross validation (CV) for the early stopping parameter. Namely, we picked the SGD iteration that gave the best average test set score across all the folds. We carried out 10 independent experiments, and Table II summarizes the result. From the table, we can observe that the DCNN can achieve an accuracy of 87.8% for activity classification on water, without any extensive tuning of hyper-parameters.
TABLE II
ACCURACY OF THE DCNN FOR EACH FOLD AND THE AVERAGE, EVALUATED AT THE BEST SGD ITERATION STEP SELECTED BY CROSS-VALIDATION. THE VALUES ARE THE AVERAGES OF 10 INDEPENDENT EXPERIMENTS.

<table>
<thead>
<tr>
<th>Fold 1</th>
<th>Fold 2</th>
<th>Fold 3</th>
<th>Fold 4</th>
<th>Fold 5</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>87.2%</td>
<td>94.8%</td>
<td>90.0%</td>
<td>79.2%</td>
<td>88%</td>
<td>87.8%</td>
</tr>
</tbody>
</table>

In Figure 5, we provide visualizations of a sample input spectrogram data as it passes through each convolution layer. Figure 5(a) shows the original input spectrogram we used. Each row of Figure 5(b) shows the convolution filter outputs for each layer. We picked 6 filters in each layer that showcase most different outputs. Note that the figure resolution reduces by half as the spectrogram data passes through each convolution layer since we used 2-by-2 max pooling.

![Input Data](image)

![Conv1 output](image)

![Conv2 output](image)

![Conv3 output](image)

(a) (b)

Figure 5. (a) Sample spectrogram, and (b) visualization of outputs of each convolution layer for a sample input spectrogram data (‘breast stroke’).

From the visualization, we can observe that the DCNN successfully learns convolution filters at each layer, capturing quite different aspects of the input spectrogram; for example, some filters focus on textures, whereas others focus on the edges and so forth. We see that such a tendency is well-preserved at the higher layer (conv3), which is fed into the fully connected layer and the final classification.

4. CONCLUSIONS

We proposed to classify human activities on water using DCNN. Five human activities on water were measured using Doppler radar. After training, DCNN could recognize the activities with an accuracy of 87.8%. In particular, the visualization of output of the convolution filters showed that they capture different aspects of the spectrogram such as micro-Doppler patterns and noise patterns. These patterns recognition is analogues for humans to recognize different images. Based on the classification accuracy and visualization of the filter outputs, it is reasonable to say that DCNN is capable of automatically learning a set of necessary features for the activity classification on water. This result shows that our DCNN approach is promising for further extensions on the micro-Doppler based activity classification problems.
ACKNOWLEDGMENT

Authors would like to thank Jinhee Park for her help on the data processing and training experiments.

REFERENCES