Active Sonar Target Classification with Power-Normalized Cepstral Coefficients and Convolutional Neural Network

Seungwoo Lee 1, Iksu Seo 1, Jongwon Seok 2, Yunsu Kim 2 and Dong Seog Han 3,*

1 Agency for Defense Development, Jinhae 51678, Korea; swlee16@knu.ac.kr (S.L.); seois@add.re.kr (I.S.)
2 Department of Information and Communication, Changwon National University, Changwon 51140, Korea; jwseok@changwon.ac.kr (J.S.); asd8780@gmail.com (Y.K.)
3 School of Electronics Engineering, Kyungpook National University, Daegu 41566, Korea
* Correspondence: dshan@knu.ac.kr; Tel.: +82-53-950-6609

Received: 29 September 2020; Accepted: 24 November 2020; Published: 26 November 2020

Featured Application: The underwater target classification algorithm proposed in this paper can be applied to an active sonar system to detect long-range targets.

Abstract: Detection and classification of unidentified underwater targets maneuvering in complex underwater environments are critical for active sonar systems. In previous studies, many detection methods were applied to separate targets from the clutter using signals that exceed a preset threshold determined by the sonar console operator. This is because the high signal-to-noise ratio target has enough feature vector components to separate. However, in a real environment, the signal-to-noise ratio of the received target does not always exceed the threshold. Therefore, a target detection algorithm for various target signal-to-noise ratio environments is required; strong clutter energy can lead to false detection, while weak target signals reduce the probability of detection. It also uses long pulse repetition intervals for long-range detection and high ambient noise, requiring classification processing for each ping without accumulating pings. In this study, a target classification algorithm is proposed that can be applied to signals in real underwater environments above the noise level without a threshold set by the sonar console operator, and the classification performance of the algorithm is verified. The active sonar for long-range target detection has low-resolution data; thus, feature vector extraction algorithms are required. Feature vectors are extracted from the experimental data using Power-Normalized Cepstral Coefficients for target classification. Feature vectors are also extracted with Mel-Frequency Cepstral Coefficients and compared with the proposed algorithm. A convolutional neural network was employed as the classifier. In addition, the proposed algorithm is to be compared with the result of target classification using a spectrogram and convolutional neural network. Experimental data were obtained using a hull-mounted active sonar system operating on a Korean naval ship in the East Sea of South Korea and a real maneuvering underwater target. From the experimental data with 29 pings, we extracted 361 target and 3351 clutter data. It is difficult to collect real underwater target data from the real sea environment. Therefore, the number of target data was increased using the data augmentation technique. Eighty percent of the data was used for training and the rest was used for testing. Accuracy value curves and classification rate tables are presented for performance analysis and discussion. Results showed that the proposed algorithm has a higher classification rate than Mel-Frequency Cepstral Coefficients without affecting the target classification by the signal level. Additionally, the obtained results showed that target classification is possible within one ping data without any ping accumulation.

Keywords: target classification; active sonar; MFCC; PNCC; convolutional neural network
1. Introduction

The attenuation of radio waves is more severe underwater compared to air, so that only very close distance targets can be detected. Therefore, sound waves rather than radio waves are used to detect underwater targets [1]; sound waves can be detected at relatively long distances, although the transmission distance of sound waves is also dependent on the underwater environment. The equipment used to detect underwater targets using sound waves is called sonar, which can be divided into two main categories: active sonar and passive sonar. In the latter case, the sound signal generated by the target is received and detected, while in the former case, the acoustic signal is transmitted and the echo returned from the target is detected. When the acoustic signal from an underwater target is low, it would be difficult to detect with a passive sonar, but active sonar could be used instead. As underwater targets become more and more quiet, they are difficult to detect using passive sonar, and so active sonar must be used. When using active sonar to detect underwater targets, some echoes are not only reflected from underwater targets, but also by the sea surface, sea bottom, sea bed topography, reefs, shoals of fish, and other ships that are not of interest to the sonar console operator. Signals reflected by a cause other than the target are called clutter. Distinguishing between target and clutter signals is very difficult because there are many clutter signals other than target signals while using active sonar to detect underwater targets [2–4]. Therefore, clutter degrades target detection performance in active sonar systems and makes target detection difficult for sonar operators performing anti-submarine warfare (ASW). In general, the detection of underwater targets is up to the decision of a trained sonar console operator. This underwater target detection method can be inaccurate because it requires the sonar operator to continuously monitor the console screen. In addition, it is difficult to continuously detect and classify the movement of a target in various underwater environments. Therefore, an effective detection and classification algorithm is required in these environments.

Detecting a maneuvering target underwater is difficult for the following reasons:

- Target detection is a complex pattern classification problem due to changes over time and various underwater environments. The complexity of the acoustic transmission environment leads to loss of signal information, distortion of the acoustic signal waveform, and incomplete receipt of acoustic signals.
- Once a target is detected it will take evasive action; therefore it is necessary to continuously classify and track weak target echoes.
- Since the detection of long-range targets using low-frequency active sonar is low-resolution data, feature extraction algorithms are required.
- Long pulse repetition intervals (PRIs) are used to detect long-range targets, which results in relatively small data accumulation over time.
- It is very difficult to obtain data from various sea experiments on underwater targets.

Various algorithms have been developed and applied in the field of active sonar detection and classification. These include methods that detect targets from reverberation like morphological and statistical approaches to improve detection [5], a contrast box detector based on the statistical features of reverberation [6], and detection using Markov random field [7]. The morphological detector distinguishes between the characteristics of the target signal and the reverberation signal, and processes them under the condition that the target signal is in an isolated area and the reverberation signal has multiple clutter distributions. This method is effective in removing reverberation, but it has the disadvantage of slowing down the target detection rate with a single ping. A classification method using temporal and spatial features of targets and clutters from multiple pings has also been proposed [8,9]. However, using multiple ping data causes a reduction in the classification rate for a single ping. Moreover, using multiple pings causes the unfiltered signal in a single ping to accumulate across multiple pings.

There are many approaches to detecting mines using a side-scan sonar with high-resolution data [10–12]. However, these approaches are mainly employed for short-range detection purposes.
using high-resolution image data. This makes it difficult to apply to low-frequency active sonar for long-range target detection using low-resolution data. A variety of studies have been conducted to assess different properties of the sonar feature of the mine depending on its location on the seabed and angle of incidence of the ping. Seo et al. [13] used the spectral feature information on the bottom of the sea to separate the target from the clutter. In [13], the target signals were generated using the mathematical model of a cylindrical object proposed by Ye [14], and the clutter signals were generated based on the K-distribution reverberation model introduced by Abraham and Lyons [15,16]. To evaluate the classification performance, a logistic regression model trained with the simulated data was applied to the experimental data. It is very difficult to obtain experimental data for training in an underwater environment, so this approach may be an alternative method in situations where experimental data are scarce. In addition, a method based on the time-reversal technique has been proposed to improve the detection of cylindrical objects on the seafloor [17]. Although many studies have been conducted on the short-range detection of fixed objects on the seafloor such as mines, studies reported on long-range detection of moving targets in an underwater environment are insufficient.

The level of underwater noise is dependent on environmental factors such as sea state and surrounding vessels. These factors influence the intensity and pattern of the underwater target’s echo. Additionally, the strength of signals reflected from the target also depends on the type of target and the angle of incidence of the transmitted signal. These factors affect the echo strength and echo pattern of the target. This leads to a variable signal-to-noise ratio (SNR) in the signal processing of the sonar and make it more difficult to continuously detect and classify the target. The target signal with low SNR does not provide enough features to distinguish it from clutter. For this reason, many researchers have tried to solve classification problems for echoes above a preset threshold by the sonar console operator. These matched filter output data above a preset threshold are selected by the sonar console operator for detection, tracking, classification, and console display. The process of selection of these signals is not accurate, as the threshold is adjusted according to the sonar console operator’s experience. If the SNR of the target’s echo is below the threshold set by the sonar console operator, the sonar console operator cannot detect and classify the target. Therefore, the matched filter output data is very important as it affects the performance of the sonar console operator’s manual detection, tracking and classification. It is very important to continuously detect, track and classify targets, especially in sonars used for military purposes. To detect long-range targets using active sonar, we use long PRI, meaning that fewer data can be obtained over time, and it is difficult to quickly notice changes in environmental noise and target strength. To overcome this problem, we propose a target classification algorithm that can be applied to all signals above the noise level regardless of the preset threshold by the sonar console operator and SNR of the signal in one ping.

In this paper, we propose a method of obtaining feature information from human auditory characteristics. The sonar console operator cannot distinguish between a target and clutter from the sonar console display but it can distinguish between them from the audio signal. Therefore, this paper proposes an approach for extracting features using Power-Normalized Cepstral Coefficients (PNCC) for active sonar with real sea trial data and Mel-Frequency Cepstral Coefficients (MFCC) are used to compare the classification results for feature extraction. PNCC has been recently developed and has superior speech recognition performance compared to MFCC, which is widely used in speech recognition. PNCC is also stronger than MFCC in noisy environments [18]. PNCC is advantageous for sonar operating environments in which noise is time-varying depending on the characteristics of the underwater environment. The MFCC has been applied to active sonar target classification studies [19], but there are no studies that have applied active sonar target classification using PNCC. These feature extraction results are imaged and used as an input for a classifier.

As artificial intelligence technology advances, recent identification studies show that deep learning has good performance in various fields [20–25]. Convolutional Neural Network (CNN), a field of deep learning, is showing good performance in the field of image recognition. Recently, Choo et al. [26] studied active sonar target classification using a spectrogram and CNN. In [26], the beamforming result
was converted into a spectrogram, and then the spectrogram image was classified using CNN. In this paper, we propose a CNN structure suitable for active sonar data and use it as a classifier. The results of feature extraction for real sea trial data are used as the CNN input. It is difficult to collect target data from the real sea environment resulting in target data being very small compared to clutter data. Therefore, a data augmentation technique is required to increase the target data [24].

In this paper, the performance of the proposed algorithm is evaluated using the classification rate and the proposed CNN model is used to classify feature vectors. The classification results indicate that the proposed algorithm is better than the case of extracting features using MFCC. In addition, the classification result using the proposed algorithm showed better performance than the classification result using the spectrogram. Classifying targets with the proposed algorithm is a huge advantage for the sonar console operators in anti-submarine warfare because classification is possible without setting a threshold within a single ping. The paper is sectioned as follows. Section 2 introduces MFCC and PNCC. While Section 3 describes the proposed algorithm. Then, in Section 4, experiments are described, and the results are discussed. Finally, Section 5 summarizes the findings of the study.

2. Introduction to Acoustic Feature Extraction Methods

Low-frequency active sonar utilizes a frequency that can be heard by humans, and the sonar console operator can classify between echoes from the target and clutter. The probability of target classification is increasing with the recent advancement of technology using acoustic feature extraction for underwater target classification [27]. In this paper, a method of extracting acoustic features for low-frequency active sonar, a target and clutter classification using MFCC and PNCC, which are voice signal processing techniques similar to human hearing processes, is used.

2.1. Mel-Frequency Cepstral Coefficients

Feature extraction for the recognition of acoustic signals, such as voice and audio, is used by a nonlinear distribution of filter banks that transform basic frequencies into the mel-scale focusing on human hearing ability. Voice vectors obtained by these filter banks are called MFCC [28]. The procedure for extracting the features of MFCC is as follows; a pre-emphasis is performed after extracting the voice signal frame by frame. Then, the frame signal is applied to the fast Fourier transform to obtain a power spectrum. Afterward, this result is passed by the mel-scale filter bank and converted back to the frequency scale again. Finally, the logarithm is taken to reflect the perceived characteristics of the frequency, and MFCC is extracted through both the discrete cosine transform (DCT) and mean normalization process. Figure 1 shows the process of MFCC.

![Figure 1. Block diagram of the Mel-Frequency Cepstral Coefficients (MFCC).](image)

A brief description of each block in Figure 1 is as follows.

- **Pre-Emphasis**: This performs the role of a kind of high-pass filter and emphasizes high frequency components.
- **STFT**: The length of the frame is divided into about 20 to 40 ms (short time), and a frequency component is obtained by performing Fourier transform for each frame.
- **Magnitude Squared**: Finds the energy for each frame.
- **Triangular Frequency Integration**: For each frame, a triangular filter bank is applied.
Logarithmic Nonlinearity: This takes a logarithm of the integration result.
DCT: Performs discrete cosine transform operation to make the cepstral coefficient.
Mean Normalization: Takes an average to reduce the impact on fast-changing components.

2.2. Power-Normalized Cepstral Coefficients

PNCC is a recently developed voice recognition technology that exploits 50–120 ms frames for additional medium-time processes for existing single-segment spectra using 20–30 ms frames, making it stronger for noise, channel distortion, and reverberation. A medium-time process asymmetrically suppresses the noise in the voice signal. Figure 2 shows the PNCC feature extraction block diagram.

A brief description of each block in Figure 2 is as follows.

- **Pre-Emphasis**: This performs the role of a kind of high-pass filter and emphasizes high frequency components.
- **STFT**: The length of the frame is divided into about 20 to 40 ms (short time), and a frequency component is obtained by performing Fourier transform for each frame.
- **Magnitude Squared**: Finds the energy for each frame.
- **Gammatone Frequency Integration**: For each frame, a gammatone filterbank is applied.
- **Medium-Time Power Calculation**: The energy of the spectrum is obtained for a frame of 50 to 120 ms (medium-time).
- **Asymmetric Noise Suppression with Temporal Masking**: The predicted signal is subtracted from the input signal by predicting the level of background noise for each frame through asymmetric-nonlinear filtering.
- **Weight Smoothing**: This goes through the smoothing process for the transfer function.
- **Time-Frequency Normalization**: Takes the normalization process for time-frequency.
- **Mean Power Normalization**: This goes through the normalization process for the average power.
- **Power Function Nonlinearity**: This goes through a process of non-linearization of power.
- **DCT**: Performs the discrete cosine transform operation to make the cepstral coefficient.
- **Mean Normalization**: Takes an average to reduce the impact on fast-changing components.

PNCC performs pre-emphasis when a voice signal is received and weighs the short-time Fourier transform (STFT) output to positive frequencies by a frequency response associated with gammatone filter banks. This helps to obtain power from the spectrum in the 40 analytical bands. If background noise and channel distortion are present compared to the existing single-segment spectrum, the performance is increased by predicting the level of background noise for each frame through asymmetric-nonlinear filtering and then subtracting the predicted signal from the input signal. The time-frequency, average power, and power function normalization process are followed by DCT and subsequently, the average normalization to extract PNCC on signals that have undergone a medium-time process.
3. The Proposed Algorithm

Generally, active sonar systems display signals exceeding a preset threshold by the sonar console operator from matched filter outputs for beamforming results. These signals are displayed cumulatively and the results revealed to the sonar console operator, who then performs target classification. Figure 3 shows a procedure of target classification by the sonar console operator using active sonar systems.

![Figure 3. Target classification procedure by the sonar console operator.](image)

The feature vectors were extracted from the beamforming output of the sonar system. Figure 4 shows the block diagram of the proposed algorithm. The target classification algorithm is performed as a classification-before-detection concept with the beamforming output. The results of this classification are displayed on the sonar console for each ping. These results are very important for automatic tracking and detection by the sonar operator.

![Figure 4. The proposed classification diagram.](image)

Data received from the sensor of the cylindrical hull-mounted array used the Hanning window during the beamforming process. The delay and sum beamforming was performed in the time domain while the interpolation rate was 32. Compensation for sensor position changes was done to reduce the influence of the movement of the ship during the beamforming phase. The change in position of the sensor array was recalculated from the ship’s pitch and roll data. Therefore, beamforming was performed on the compensated position of the sensor array. In addition, the sound velocity value was used for beamforming.

PNCC is used to extract features from the beamforming result. Compared to MFCC, PNCC is superior in feature discrimination in the sonar operation environment where ambient noise changes time varying to the characteristics of an underwater environment, unlike Gaussian noise [29]. Therefore, auditory features were extracted using PNCC, which is advantageous for the sonar operating environment.
environment. The results were classified using CNN, and the results of MFCC feature extraction were compared with those classified using CNN. Figure 5 shows the block diagram for comparison with PNCC. The CNN model in Figure 5 is the same as in Figure 4.

![Block Diagram](image)

**Figure 5.** The classification diagram for comparison considering MFCC as baseline.

The feature extraction example for the target data using MFCC and PNCC is shown in Figure 6. Figure 6a represents the beamforming output. In Figure 6a, a target exists between 2.7 and 2.8 s. Figure 6b is the feature extraction output of MFCC and Figure 6c is the feature extraction output of the PNCC.

![Feature Extraction Results](image)

**Figure 6.** Feature extraction results for target: (a) time signal of target with ambient noise; (b) MFCC feature extraction result of target; (c) PNCC feature extraction result of target.

The feature extraction example for clutter data using MFCC and PNCC is shown in Figure 7. Figure 7a is the beamforming output. In Figure 7a, the target exists between 2.2 and 2.3 s. Figure 7b is the feature extraction output of MFCC and Figure 7c is the feature extraction output of the PNCC.
Figure 7. Feature extraction results for clutter: (a) time signal of clutter with ambient noise; (b) MFCC feature extraction result of clutter; (c) PNCC feature extraction result of clutter.

According to the feature extraction results in Figures 6 and 7, it can be seen that the PNCC is more discernable than the MFCC. Comparing Figure 6b,c, it is possible to distinguish between frame indexes 270 to 280, which is the part where the target exists, as PNCC looks more discernable than MFCC. In addition, in Figure 7b,c, it can be seen that PNCC looks more discernable than MFCC between frame indexes 220 and 230, where the clutter is located. In Figures 6 and 7, the cepstrum index means the order of cepstrum coefficients for each frame. Coefficient is 13th order, delta is 13th order, and delta-delta is 13th order, and therefore it has a total cepstrum index of 39th order.

Since the target data from the real sea environment is very small compared to clutter data, more target data are generated using the data augmentation technique. Data augmentation is a technique that increases the number of training data when the number of training data is insufficient by adding transformed data [24]. In image classification, the amount of training data can be augmented using methods such as flipping, cropping, rotating, sampling, scaling, inversion, and noise addition of the original image. These augmentations can have a positive effect on performance. Figure 8 shows the target data generated using the data augmentation technique. The position of the window (red square) in Figure 8 is moving forward and backward randomly in the frame and generating more target data for CNN input.

Figure 8. Target data generation using data augmentation.
CNN is a neural network modeling the visual information processing of animals and shows good performance in image classification. When visual information is input, the stimulus is not transmitted to all the nerve cells, but the stimulus is received from the cells of the receiving area. This is a neural network designed for image processing by expression as a neural network structure. The results of feature extraction using PNCC are used to classify targets and clutter via CNN. Figure 9 shows the proposed CNN model structure. The proposed CNN model structure is designed to be suitable for classifying low-frequency active sonar data used in experiments.

4. Experimental Results and Discussion

In this section, the proposed approach was validated with real sea trial data. The use of an active hull-mounted sonar system was considered to detect a moving underwater target.

4.1. Experiments Data

As previously stated, the active sonar real sea trial data used in the experiments are based on the beamforming of the reflected signal in the East Sea of South Korea. The target size was less than 100 m in length and the depth of the East Sea of South Korea was about 500 to 3000 m. The transmitting signal is a linear frequency modulation (LFM) signal with a sampling frequency of 31.25 kHz, a center frequency of 3.9 kHz, a bandwidth of 400 Hz, and a pulse length of 50 ms. The real sea trial data consists of 128 beams for 360 degrees omnidirectional and has 29 ping data obtained by transmitting with a PRI of about 13 s. This sea trial environment is shown in Figure 10.
Figure 10. The environment of the experiments.

From the received data of 29 pings, 361 target data and 3351 clutter data were extracted for classification. The extracted data were above the noise level. The example of the target data from the beamforming output and spectrogram is shown in Figure 11. Figure 11a shows 4-s beamforming output data and Figure 11b shows a spectrogram of Figure 11a. In Figure 11a, the LFM signal reflected from the target exists between 2.7 and 2.8 s and in Figure 11b, the LFM signal reflected from the target exists between 2.7 and 2.8 s and a center frequency of 3.9 kHz. Feature information for target classification is extracted from the beamforming result in Figure 11a.

Figure 11. Example of the target data spectrogram: (a) beamforming output; (b) spectrogram.

The feature extraction process uses a total of 26 mel-scale filter banks considering only the frequency band up to 8 kHz. The 13th MFCC and PNCC were extracted through DCT and lifting. To account for the change over time, delta and delta-delta MFCC and PNCC were obtained. In total, 39 feature vectors were extracted.

As mentioned, it is difficult to collect target data from the real sea experiment. The number of target data is very small compared to clutter data, so target data are generated using the data augmentation technique. The data augmentation technique of the target data is shown in Figure 12. The position of the window (red square) on the feature extraction output is moving forward and backward in the frame (the same meaning as time). This means that the temporal position moves while...
maintaining the characteristics of the target. In the experiments, 3610 target data were generated from 361 original target data.

**Figure 12.** Target data generation of MFCC and PNCC output using data augmentation: (a) MFCC data augmentation; (b) PNCC data augmentation.

The generated target data using the data augmentation technique and clutter data are used as CNN inputs. Figure 13 shows the entire experimental process. In this figure, the proposed CNN model takes an image of size 39 × 395 as input and has 3 convolution layers and 2 outputs. In the proposed CNN model, the size of the convolution filter is fixed at 3 × 3, and maxpooling uses a 2 × 2 window. Moreover, the rectified linear unit is used as the activation function and the result is obtained using a softmax function in the last layer. The software used in the experiments is Python 3.6.0 with TensorFlow 2.0 and Keras 2.3.0. The hardware specifications in the experiments are as follows: graphic card is GeForce RTX 2080 Ti and CPU is AMD Ryzen 7 2700X (8 core processor).

**Figure 13.** Structure for the experiment.

Table 1 lists CNN input data numbers of target, clutter, training and testing. Eighty percent of the total data were used for training, while the remaining 20% were used for testing.
Table 1. Input data numbers.

| Item                              | Number of Data |
|----------------------------------|----------------|
| Target (data augmentation output)| 3610           |
| Clutter                          | 3351           |
| Training                         | 5568           |
| Testing                          | 1393           |

Table 2 lists the parameters of CNN.

Table 2. Experiment parameters.

| Parameter            | Content  |
|----------------------|----------|
| Epochs               | 50       |
| Training rate        | 0.001    |
| Batch size           | 32       |
| Weight initialization| He       |

4.2. Results and Discussion

The classification performance was evaluated between target and clutter data. Figure 14 shows the accuracy for training and testing of MFCC and PNCC for 50 epochs. In Figure 14, compared to MFCC, PNCC converges faster during training and testing.

Figure 14. Accuracy depending on learning epochs: (a) MFCC accuracy; (b) PNCC accuracy.

The results of classifying targets and clutters using CNN are shown in Table 3. It can be observed that the classification rate is higher when the feature is extracted using PNCC rather than MFCC, and it is classified using CNN.
Table 3. Results of classification.

| Feature Extraction | Class      | Results of Classification |
|--------------------|------------|---------------------------|
| MFCC               | Target     | 97.23%                    |
|                    | Clutter    | 2.766%                    |
| PNCC               | Target     | 98.617%                   |
|                    | Clutter    | 1.383%                    |

In the case of target classification, PNCC has a higher classification rate of 1.383% than MFCC, and for clutter classification, PNCC has a 0.597% higher classification rate than MFCC. From the classification results, Table 4 summarizes the analysis results of precision, recall, and F-measure using these two criteria (precision and recall). For all three analysis results in Table 4, PNCC shows better performance than MFCC.

Table 4. Precision-Recall analysis.

|         | Precision | Recall | F-Measure |
|---------|-----------|--------|-----------|
| MFCC    | 98.49%    | 97.23% | 97.86%    |
| PNCC    | 99.10%    | 98.62% | 98.86%    |

The result of converting the sea experiment data into a spectrogram and classifying the target using CNN showed that the classification rate of the test was 94% [26]. The sea experiment data used in [26] is the same as in this paper. Compared with the results of [26], the result using PNCC showed about 4.6% higher classification rate than the result using the spectrogram.

In the proposed algorithm result, target echoes are well classified. Furthermore, this result shows that the PNCC has a better performance compared with the MFCC in terms of the classification rate. It can also greatly help sonar operators when detecting a target manually.

The results of computational demands MFCC and PNCC are shown in Table 5 [18]. Table 5 shows that PNCC has about 34.6% more computation than MFCC.

Table 5. Results of computational demands.

| Item                        | MFCC  | PNCC  |
|-----------------------------|-------|-------|
| Sum of Multiplications and Divisions | 13,010 | 17,516 |

In the case of feature extraction using PNCC, a better classification result can be obtained through an increase of only 34.6% of the computational amount compared to MFCC, and the addition of this amount of computation due to the development of computing power will not cause problems in real-time processing.

5. Conclusions

In this paper, we studied whether target and clutter can be classified by feature extraction and CNN. The classification performance of the proposed algorithm was analyzed by applying data above the noise level without a preset threshold by the sonar console operator. As a result of the evaluation, it was confirmed that targets and clutter can be classified by the proposed algorithm in a real underwater environment. Therefore, the proposed algorithm is of potential use for classifying underwater targets and can be helpful to the sonar console operators. These results are based on data from sea experiments obtained by applying an actual active sonar system. Although the sea
experimental data do not represent all the characteristics of the underwater environment, the possibility of applying the proposed algorithm to an active sonar system has been vindicated.

This paper shows that the PNCC can be used as feature vectors while CNN can be used as a classifier that leads to a higher classification rate in an active sonar system. The classification results also indicate that the proposed approach is better than MFCC used as feature vectors and the proposed approach is better than using the spectrogram. The fact that this is possible without setting a threshold when classifying targets is of great help to the sonar console operators in performing ASW. The proposed algorithm can be applied to all signals above the noise level without a preset threshold by the sonar console operator, so signals that are ignored in the process using a preset threshold can be detected. It is also possible to classify targets in one ping without accumulation of pings. Therefore, the proposed algorithm can greatly improve the detection, tracking and classification capabilities of the sonar console operators. Since real sea trial data of active sonar systems operating on naval ships were used in real underwater environments, the proposed algorithm is applicable to real active sonar systems.

In future work, the fusion of different feature extraction methods can be a useful approach for active sonar systems. Furthermore, data from real sea trials in more diverse environments will be useful to compare the classification rates for each case.

Author Contributions: Conceptualization and methodology, S.L.; software, Y.K.; validation, I.S. and J.S.; formal analysis, J.S. and Y.K.; resources, S.L.; data curation, S.L. and Y.K.; writing—original draft preparation, S.L. and J.S.; writing—review and editing, S.L. and D.S.H.; visualization, Y.K.; supervision, D.S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Urick, R.J. *Principles of Underwater Sound*, 3rd ed.; McGraw-Hill: New York, NY, USA, 1983.
2. Harrison, R.; Yang, C.; Lin, C.F.; Politopoulos, T.; Chang, E. Classification of Underwater Targets with Active Sonar. In Proceedings of the First IEEE Regional Conference on Aerospace Control Systems, Westlake Village, CA, USA, 25–27 May 1993; pp. 534–538.
3. Kelly, J.G.; Carpenter, R.N.; Tague, J.A.; Haddad, N.K. Optimum Classification with Active Sonar: New Theoretical Results. In Proceedings of the ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing, Toronto, ON, Canada, 14–17 April 1991; Volume 2, pp. 1445–1448.
4. Young, V.W.; Hines, P.C. Perception-based Automatic Classification of Impulsive Source Active Sonar Echoes. *J. Acoust. Soc. Am.* 2007, 122, 1502–1517. [CrossRef] [PubMed]
5. Ginolhac, G.; Chanussot, J.; Hory, C. Morphological and Statistical Approaches to Improve Detection in the Presence of Reverberation. *IEEE J. Ocean. Eng.* 2005, 30, 881–899. [CrossRef]
6. Choi, J.; Yoon, K.; Lee, S.; Kwon, B.; Lee, K. Four Segmentalized CBD Algorithm Using Maximum Contrast Value to Improve Detection in the Presence of Reverberation. *J. Acoust. Soc. Korea* 2009, 28, 761–767. (In Korean)
7. Laterveer, R. *Single Ping Clutter Reduction: Segmentation Using Markov Random Fields*; SACLANT undersea Research Centre: La Spezia, Italy, 1999.
8. Design Disclosures for Interacting Multiple Model and for the Multi Ping Classifier; Technical Report; Applied Research Lab: State College, PA, USA, 1999.
9. Carlson, B.D.; Evans, E.D.; Wilson, S.L. Search radar detection and track with the Hough transform, Part I: System concept. *IEEE Trans. Aerosp. Electron. Syst.* 1994, 30, 102–108. [CrossRef]
10. Blondel, P. *The Handbook of Sidescan Sonar*; Springer Science & Business Media: New York, NY, USA, 2010.
11. Wang, X.; Lie, X.; Japkowicz, N.; Matwin, S.; Nguyen, B. Automatic Target Recognition using multiple-aspect sonar images. In Proceedings of the 2014 IEEE Congress on Evolutionary Computation (CEC), Beijing, China, 6–11 July 2014; pp. 2330–2337.
12. Zerr, B.; Stage, B. Three-dimensional reconstruction of underwater objects from a sequence of sonar images. In Proceedings of the 1996 International Conference on Image Processing, Lausanne, Switzerland, 19 September 1996; IEEE: Piscataway, NJ, USA, 1996; Volume 3, pp. 927–930.

13. Seo, Y.; On, B.; Im, S.; Shim, T.; Seo, I. Underwater Cylindrical Object Detection Using the Spectral Features of Active Sonar Signals with Logistic regression Models. Appl. Sci. 2018, 8, 116. [CrossRef]

14. Ye, Z. A novel approach to sound scattering by cylinders of finite length. J. Acoust. Soc. Am. 1997, 102, 877–884. [CrossRef]

15. On, B.; Kim, S.; Moon, W.; Im, S.; Seo, I. Detection of an Object Bottoming at Seabed by the Reflected Signal Modeling. J. Inst. Electron. Inf. Eng. 2016, 53, 55–65.

16. Abraham, D.A.; Lyons, A.P. Novel physical interpretations of K-distributed reverberation. IEEE J. Ocean. Eng. 2002, 27, 800–813. [CrossRef]

17. On, B.; Im, S.; Seo, I. Performance of Time Reversal Based Underwater Target Detection in Shallow Water. Appl. Sci. 2017, 7, 1180. [CrossRef]

18. Kim, C.; Stern, R.M. Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition. IEEE/ACM Trans. Audio Speech Lang. Process. 2016, 24, 1315–1329. [CrossRef]

19. Seok, J. Active Sonar Target/Non-target Classification Using Real Sea-trial Data. J. Korea Multimed. Soc. 2017, 20, 1637–1645. (In Korean)

20. Hinton, G.E.; Osindero, S.; Teh, Y.W. A fast learning algorithm for deep belief nets. Neural Comput. 2006, 18, 1527–1554. [CrossRef] [PubMed]

21. LeCun, Y.; Bengio, Y.; Hinton, G.E. Deep learning. Nature 2015, 521, 436–444. [CrossRef] [PubMed]

22. Schmidhuber, J. Deep learning in neural networks: An overview. Neural Netw. 2015, 61, 85–117. [CrossRef] [PubMed]

23. LeCun, Y.; Bengio, Y.Convolutional networks for images, speech, and time series. Handb. Brain Theory Neural Netw. 1995, 3361, 1995.

24. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Commun. ACM 2017, 60, 84–90. [CrossRef]

25. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.

26. Choo, Y.; Hong, W.; Seong, W.; Lee, W.; Seo, I. Discrimination of Target Signals Using Convolution Neural Network. Proc. Acoust. Soc. Korea 2018, 37, 63. (In Korean)

27. Zhang, L.; Wu, D.; Han, X.; Zhu, Z. Feature Extraction of Underwater Target Signal Using Mel Frequency Cepstrum Coefficients Based on Acoustic Vector Sensor. J. Sensors 2016, 2016, 7864213. [CrossRef]

28. Davis, S.; Mermelstein, P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans. Acoust. Speech Signal Process. 1980, 28, 357–366. [CrossRef]

29. Zhang, Y.; Xu, K.; Wan, J. Robust Feature for Underwater Targets Recognition Using Power-Normalized Cepstral Coefficients. In Proceedings of the 2018 14th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 12–16 August 2018; pp. 90–93.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).