

Detection of Abnormal Vessel Trajectories with Convolutional Autoencoder

June-Hyoung Son* · Jun-Gun Jang** · Bongwan Choi* · Kyeongtaek Kim*[†]

*Department of Industrial Engineering, Hannam University

**Korea Aerospace Industries

합성곱 오토인코더를 이용한 이상거동 선박 식별

손준형* · 장준건** · 최봉완* · 김경택*[†]

*한남대학교 산업공학과

**한국항공우주산업(주)

Recently there was an incident that military radars, coastal CCTVs and other surveillance equipment captured a small rubber boat smuggling a group of illegal immigrants into South Korea, but guards on duty failed to notice it until after they reached the shore and fled. After that, the detection of such vessels before it reach to the Korean shore has emerged as an important issue to be solved. In the fields of marine navigation, Automatic Identification System (AIS) is widely equipped in vessels, and the vessels incessantly transmits its position information. In this paper, we propose a method of automatically identifying abnormally behaving vessels with AIS using convolutional autoencoder (CAE). Vessel anomaly detection can be referred to as the process of detecting its trajectory that significantly deviated from the majority of the trajectories. In this method, the normal vessel trajectory is gridded as an image, and CAE are trained with images from historical normal vessel trajectories to reconstruct the input image. Features of normal trajectories are captured into weights in CAE. As a result, images of the trajectories of abnormal behaving vessels are poorly reconstructed and end up with large reconstruction errors. We show how correctly the model detects simulated abnormal trajectories shifted a few pixel from normal trajectories. Since the proposed model identifies abnormally behaving ships using actual AIS data, it is expected to contribute to the strengthening of security level when it is applied to various maritime surveillance systems.

Keywords : Vessel Anomaly Detection, Convolutional Autoencoder, Trajectory, AIS

1. 서 론

2020년 6월에 중국 밀입국자들이 보트를 이용하여 충청남도 해안으로 상륙한 사건은, 레이더와 CCTV 등에 여러 번 노출 되었음에도 불구하고 밀입국선을 식별하지 못

했다는 사실로 큰 문제가 되었다. 심지어, 레이더와 CCTV 감시병의 근무태만이 원인의 하나로 의심을 받기도 했다. 그러나, 인간의 눈과 기억력으로 이러한 밀입국 선박을 식별하는 것은 한계가 있을 수밖에 없다. 레이더 화면의 깜박이는 점이나 CCTV 화면에 순간적으로 나타난 선박이 밀입국선 인지 여부를 판별하는 것은 불가능하다. 오랜 시간 동안 선박의 궤적을 추적해야만 이상거동 선박 여부를 판단할 수 있다.

선박항로와 관련된 국내의 대부분 연구는 경제성과

Received 26 November 2020; Finally Revised 20 December 2020;
Accepted 23 December 2020

[†] Corresponding Author : kkim610@gmail.com

최적화에 집중되어왔다[7, 8, 10]. 그러나, 선박의 궤적을 추적할 수 있게 만드는 AIS(Auto Identification System) 장비의 탑재가 의무화되면서 연구의 지평이 넓어졌다. AIS는 해상교통안전을 위해 도입된 선박 자동식별 시스템이다. AIS 데이터는 해양에서의 안전 및 해양으로부터의 위협을 확인할 수 있는 중요한 데이터로서 선박 안전 운항 정보 시스템 개발 등 다양한 분야의 연구에 활용되고 있다[9]. 무엇보다, 해양안보 측면에서 해양으로부터의 테러 및 공격 위협은 세계 각국의 주요 현안이므로 특정 지역에 접근하는 이상거동 선박을 식별하여 조치하는 것은 국가의 안전보장 측면에서 매우 중요하다. 본 논문에서는 딥러닝의 한 분야인 오토인코더를 이용하여 AIS 데이터를 분석함으로써 이상거동 선박을 자동 식별하는 모형을 제안한다.

2. 배경

2000년 6월 국제해사기구가 해상인명안전협약(SOLAS, International Convention for the Safety of Life at Sea)을 개정하여 AIS 장비 장착을 의무화하였다. 이에 따라, 국제적으로는 2002년 해상인명안전협약에 이 조항이 포함되었고, 국내에서는 2007년 선박 위치 발신장치의 설치기준 등에 관한 규정과 2008년 선박안전법이 개정되어 연해구역 이상을 항해하는 총톤수 10톤 이상 어선 및 50톤 이상 선박, 150톤 이상 여객선, 300톤 이상 국제항해 선박, 500톤 이상 국내 항해 선박에 AIS 장비를 탑재하고 있다.

이상거동 선박이란 정상적인 항로로 이동하는 선박들의 패턴을 벗어난 항로로 운항하는 선박을 뜻한다. 이상거동 선박에 관한 연구는 전통적인 방법부터 시작하여 최근에는 신경망을 이용한 방법까지 연구되고 있다.

2.1 전통적인 연구

이상거동 선박을 식별하기 위한 대부분의 연구는 정상 궤적을 먼저 정의하고, 정상 궤적과의 차이가 임계값을 넘으면 이상거동 궤적으로 결정하는 방식으로 진행되어왔다. 정상 궤적을 정의하기 위해서는 궤적 사이의 거리를 정의하고, 비슷한 궤적들끼리 군집화하는 방법이 전통적으로 널리 사용되었다. 이때 거리를 여러 가지 방법으로 정의할 수 있는데, 유클리디언거리(ED, Euclidean Distance), 동적시간왜곡거리(DTW, Dynamic Time Warping), 가장 긴 공통하위시퀀스거리(LCSS, Longest Common Subsequence), 실제시퀀스거리편집거리(EDR, Edit Distance on Real sequence), 실제페널티수정거리(ERP, Edit distance with Real Penalty) 등이 있다.

ED 방식은 비교하는 두 개의 궤적 길이가 동일한 경우에는 문제가 없으나 궤적들의 길이가 일정하지 않으면 결과가 좋지 않다[1]. DTW 방식은 서로 다른 길이를 갖는 궤적들간의 유사도를 측정하기 위해 궤적 내 특정 시공간 좌표값을 임의의 수만큼 반복시키는 것을 허용하는 방식이나, 정확히 같은 방식으로 움직이는 두 시계열 궤적이라도 속도가 다른 경우 결과에 큰 영향을 주는 단점이 있다[4]. LCSS 방식[17]은 잡음(노이즈)에 민감할 뿐 아니라 유사한 궤적 모양 사이에 다양한 공간의 차이를 고려하지 않는 방식이므로 정확도가 낮다. EDR 방식[3]은 잡음에 의한 민감도를 감소시키기 위해 유사 궤적 내의 시공간 좌표값의 차이가 주어진 허용치보다 작으면 두 좌표가 일치된 것으로 간주하는 방식이나 서브 궤적에 대한 유사도 측정 과정을 반복 수행해야 하므로 이로 인한 성능 저하 문제가 발생할 수 있다. 실제시퀀스거리편집 방식과 동적시간왜곡거리 방식을 병합한 ERP[2] 방식은 반복을 허용하는 방법이지만 이동궤적이 연속적으로 동일하게 반복 출현하는 경우가 발생하지 않으면 실제페널티수정거리 방식의 적용은 어렵다.

K-means는 거리 기반 군집화 방법으로 전통적으로 널리 사용되어왔다. 그러나, 거리기반 군집화 방법을 해상 항로 군집화에 사용하기에는 어려움이 많다. 군집화는 궤적의 유사도를 기반으로 하며, 유사도는 거리로 표현된다. 앞에서 언급한 여러 가지 종류의 거리는 궤적 모양(shape)의 유사도를 계산하므로, 육상에서처럼 고정된 도로를 따라 움직이면 구속력 있는 규칙에 따라 일정한 궤적을 보여주지만, 선박은 안전을 위해 다른 선박과 멀리 떨어져 이동해야 하는 데다, 해상은 육상보다 상대적으로 자유 공간에서 움직이므로 동일한 목적지로 향하는 동일한 선박이라 하여도 해상 교통 상황에 따라 일정하지 않은 궤적을 나타낼 가능성이 높다. 이로 인해 전통적인 방식의 거리 기반 클러스터링 알고리즘을 해상 궤적에 직접 적용하기는 어려웠다. 그래서, 해상 응용 분야에서는 DBSCAN(Density-Based Spatial Clustering of Applications with Noise)이 군집의 개수를 미리 명시할 필요가 없으며, 자유로운 형상의 클러스터를 생성하는 능력을 가지고 있어, 이상거동 선박 식별에 널리 응용되어 오고 있다[5, 14, 15]. 하지만, DBSCAN을 적용하기 위해서는 군집의 반지름을 나타내는 임실론과 군집 내에 있는 최소 포인트의 갯수를 사용자가 제공하여야 하는 제한이 있다.

2.2 신경망을 이용한 연구

인공 신경망을 이용한 이상(anomaly) 식별 방법론은 크게 오토인코더를 이용하는 방법과 GAN(Generative Adversarial Network)을 이용하는 방법으로 분류할 수 있으나, 이상거동

선박 식별 분야에서는 오토인코더를 사용한 연구들이 보고되고 있다.

오토인코더는 데이터를 압축하거나 이미지를 축소 또는 생성시킬 경우 네트워크의 입력을 조절하여 입력과 유사한 출력을 재구성하는 방법을 학습하는 네트워크이다. 오토인코더는 데이터를 압축하는 인코더와 복원하는 디코더로 구성되며 전체는 뉴런의 레이어로 상호 연결되어 있다.

인코더는 입력 데이터를 보다 낮은 차원의 공간으로 맵핑(mapping)하며, 디코더는 낮은 차원의 공간에 투사된 데이터로부터 원래 데이터로의 회복을 시도한다. 이 두 네트워크의 파라미터들은 복원 손실 함수(reconstruction loss function)를 통하여 학습한다. 전체적인 복원 오차를 최소화하기 위하여, 저차원으로 맵핑되었다. 원래 차원으로 회복하는 과정에서 파라미터에 배정되는 값은 지배적인 입력 패턴(즉 정상 패턴)과 밀접한 관련을 갖게 된다. 결과적으로, 입력된 데이터가 지배적인 패턴에서 벗어나는 경우 복원 오차가 크게 나타나게 된다. 따라서, 복원 오차는 비정상성의 정도를 나타내는 척도로의 사용이 가능하다.

Nguyen et al.[11]은 VRNN(Variational Recurrent Neural Network) 모델을 이용하여 선박의 궤적을 예측하고, 이상거동을 식별하였다. VRNN은 변형오토인코더(Variational Autoencoder)를 확장하여 RNN에 적용한 모형이다. Iltanen [6]은 오토인코더의 복원 오차와 밀도 기반 이상치 스코어를 결합하여 이상거동 궤적을 식별하는 모형을 발표하였다. Yu et al.[18]은 선박 궤적을 적당한 길이의 세그먼트(segment)로 잘라 1차원 데이터로 만들어 오토인코더에 입력한 후 복원 오차를 사용하여 이상거동 여부를 판별하였다.

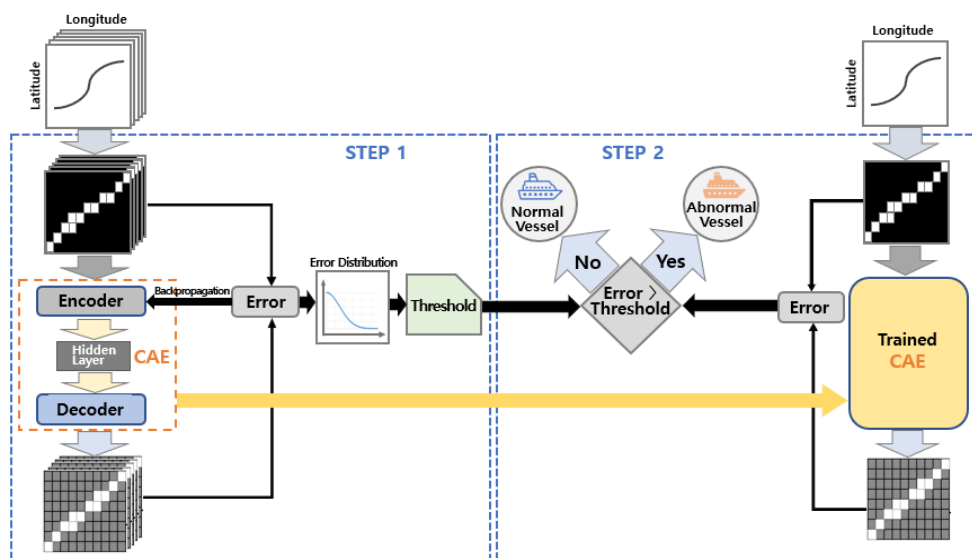
신경망을 이용한 이상거동 항적 식별에 관한 국내 연구는 극히 제한적이다. 오재용 등[12, 13]은 오토인코더를 사

용하여 입력 데이터를 저차원으로 축소한 후 복원하는 과정을 통하여 구한 복원 오차와 축소된 차원 데이터를 군집화하고, 이에 따른 군집 분류오차 등을 이용하여 이상치 값을 구한 다음, 이상거동 궤적의 판별에 사용하였다. 최근에 박재민과 김성일[16]은 경도, 위도, 속도, 방향, 시간을 각각 원-핫 인코딩(One-Hot Encoding) 형태로 전처리를 한 후에 변형오토인코더(Variational Autoencoder)와 관리도를 결합하여 이상거동 궤적을 식별하였다.

국내외를 막론하고 대부분의 연구에서는 이상거동 선박의 궤적 중 몇몇 사례들을 설명하고, 제안된 모형들이 효율적으로 작동하고 있음을 보여주고 있으나, 객관적인 성능 지표를 제시하고 있지는 않다. 또한, 대부분의 연구는 3분 또는 10분 단위와 같은 일정 시간의 궤적을 바탕으로 이상거동 여부를 판별한다. 이는 일정 시간 내의 관점에서 궤적오류를 판별하는 장점이 있으나, 그 시간 이전의 정보는 소실되는 단점을 가진다.

3. 이상거동 선박 식별 모델

본 논문에서 제시하고 있는 이상거동 선박을 식별하는 절차를 <Figure 1>에 나타내었다. 먼저 입력된 데이터를 사전 프로세싱(preprocessing) 한 다음, Step 1에서 복원 오차를 최소화하도록 합성곱 오토인코더로 훈련하여 네트워크 파라미터 값을 고정하고, 복원 오차로부터 이상거동 여부를 판별하는 기준이 되는 임계값(threshold)을 결정한다. 새로운 항로가 나타나면 Step 2의 훈련된 합성곱 오토인코더에 입력하여 복원 오차를 구하고, 임계값 이상이면 이상거동 궤적으로 판정한다.



<Figure 1> Flowchart of the Suggested Methodology

3.1 사전 프로세싱

주어진 AIS 데이터 셋을 선박식별번호별, 날짜별, 시간순으로 위도 및 경도에 배열한다. 선박식별번호와 시간은 데이터를 그룹화하고 그룹 내에서 시간적인 순서를 정하기 위해 사용될 뿐 분석에는 영향을 미치지 않는다. 감시 대상 해역 상하좌우의 경계에 해당하는 위도와 경도를 확인하고, 합성곱 오토인코더에 입력할 이미지의 해상도를 결정한 후 궤적 이미지를 생성한다. 생성된 이미지에서, 선박의 궤적은 1로 표현되며 궤적이 없는 영역은 0으로 표현된다.

3.2 합성곱 오토인코더

일반적인 선형 오토인코더 아키텍처는 불필요한 잡음을 제거하고 데이터를 압축하는 장점이 있으나, 모든 뉴런이 연결되어 있어 학습을 위해 수많은 매개변수를 조정해야 하는 문제가 발생하며, 저차원으로 압축하는 과정에서 데이터의 지역적 특성을 유지하지 못한다. 이러한 단점을 해결할 수 있는 것이 합성곱 신경망이다. 합성곱 신경망은 하나의 공간을 이동해 가며 여러 필터를 사용해 입력데이터의 특징적인 패턴을 식별하는 이상적인 필터 가중치를 찾고, 풀링 레이어를 통하여 출력값을 점진적으로 줄여간다.

본 논문에서 제안하는 합성곱 오토인코더(CAE, Convolutional Autoencoder) 모형의 구조는 64×64 크기의 입력 이미지를 사용하는 경우 <Figure 2>와 같다. 제안된 모형에서 AIS 데이터로 표현된 항로를 이미지로 변환하는 이유는, 선박의 속도에 따라 변하는 AIS 데이터 발신 주기에 영향을 적게 받도록 하기 위함이다. CAE 모형은 하나의 이미지에 한 선박의 한 항차 입항 또는 출항 궤적을 모두 표시하며, CAE의 합성곱 레이어로 인해 기본적인 오토인코더를 사용할 때보다 격자형 선박 항로의 기하학적 특징을 추출하는데 더 적합하다.

제안된 모형의 네트워크 구조는 크게 인코더 및 디코

더로 구성된다. 인코더의 각 합성곱 레이어의 필터 크기와 수는 각각 64×64 , 32×32 , 16×16 과 16, 16 및 8로 설정되어 있으며, 활성화 함수로는 Relu가 사용된다. 2×2 그리드의 최대 풀링 레이어가 각 합성곱 레이어 다음에 위치한다. 합성곱 레이어와 풀링 레이어를 모두 통과하게 되면, 데이터 크기는 $8 \times 8 \times 8 = 512$ 로 축소된다. 축소된 데이터는 1차원 데이터로 변환된 후 완전 연결 레이어로 들어간다. 완전 연결 레이어를 통하여 데이터의 크기는 512로부터 64를 거쳐 8까지 축소된다.

인코더의 결과인 크기가 8인 데이터는 디코더에 입력이 되어, 완전 연결 레이어를 통하여 크기가 64, 512로 확대된다. 이후 다시 $8 \times 8 \times 8$ 의 3차원 이미지로 변환되며, 계속해서 업 샘플링 레이어와 합성곱 레이어를 통하여 원래의 크기 64×64 인 이미지를 결과로 배출한다. 이때 마지막 합성곱 레이어를 제외한 모든 합성곱 레이어에서 활성화 함수로 Relu가 사용되며, 마지막 합성곱 레이어에서는 활성화 함수로 Sigmoid가 사용된다.

훈련을 모두 마치고 나온 CAE의 출력 이미지와 입력 이미지 사이의 binary cross entropy를 복원 오차로 정의하며, 식 (1)과 같다.

$$Error = \frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) - (1 - y_i) \log(1 - p_i)] \quad (1)$$

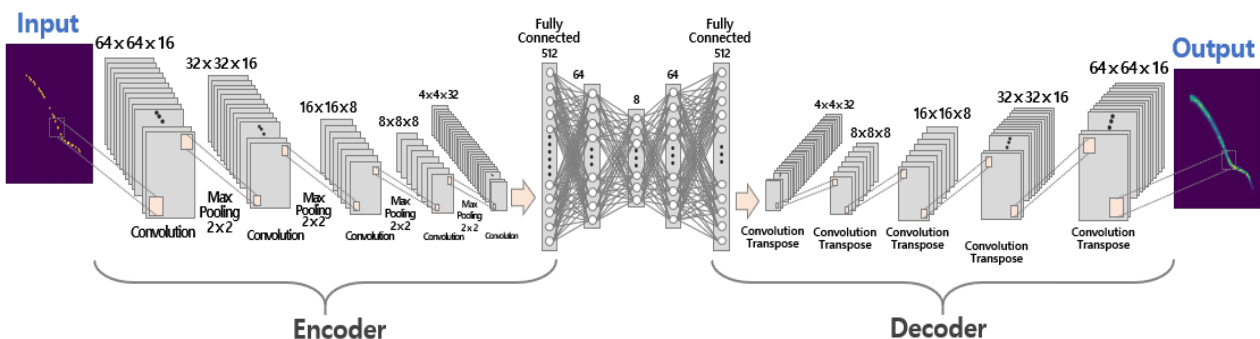
Error : 복원 오차

N : 출력 이미지의 픽셀 수

y_i : 입력 이미지의 i 번째 픽셀과 이에 대응하는 출력 이미지의 i 번째 픽셀의 이진값이 같은지를 나타내는 진위 값(0 또는 1)

p_i : 출력 이미지의 i 번째 픽셀의 예측값($0 \leq p_i \leq 1$)

모형이 적합하면 이 오차들은 0에 가까운 값을 갖는 분포를 보이게 된다. 이 복원 오차의 분포로부터 정상 궤적인지 이상거동 궤적인지 구별하는 기준으로 사용할 임계값을 설정한다.



<Figure 2> Proposed CAE Architecture

3.3 새로운 궤적에 대한 이상거동 식별

<Figure 1>의 Step 2에서 보는 바와 같이, 새로운 선박의 AIS 데이터가 입력되면 규정된 크기의 이미지로 변환되어, 훈련된 CAE 모형의 입력으로 들어간다. CAE 모형은 새로운 선박에 대한 복원 오차를 출력으로 배출하며, 이 오차가 임계값 이상이면 이상거동 선박으로 판정한다.

4. 실험

본 연구에서 제시한 모델을 검증하기 위해, 1개월간 평택항으로 입출항하는 선박들의 AIS 데이터를 활용하여, <Figure 1>에 나타난 절차에 따라 실험을 진행하였다.

먼저 데이터 전처리를 통해 AIS 데이터를 이미지 데이터로 변환하여 CAE 모형에 입력한다. 입력된 이미지들을 통하여 CAE 모형을 학습시키고, 복원 오차 분포로부터 임의의 임계값을 선택한다. 정상 궤적으로부터 평행 이동 거리를 늘려가며 비정상 궤적을 만들어, 제시된 모형이 거리 차이에 따라 비정상 궤적을 얼마나 잘 식별하는지를 살펴본다. 여러 해상 감시 여건과 식별률에 따라 임계값을 확정한다.

4.1 데이터 전처리

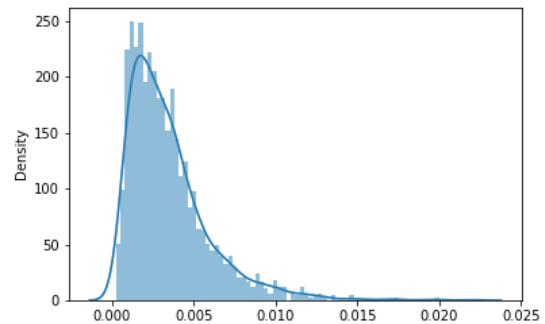
AIS는 시간에 따라 누적되는 데이터이므로 AIS 수신 시간에 따라 데이터를 정렬하면 선박 항로를 알 수 있다. 본 실험에서는, 평택항에 2019년 5월 한 달간 입출항한 선박의 약 1100만 개의 AIS 데이터로부터 선박식별번호, 위도, 경도, 시간을 사용하여 2718개의 항적을 도출하였다. 각 항적에서 정렬된 AIS 데이터를 크기 32×32 , 64×64 , 128×128 인 이미지로 변환하였다. 이는 어떤 이미지 크기가 이상거동 궤적 식별률이 높은지 알아보기 위함이다. 본 실험에서 1픽셀의 실제 모양은 정사각형으로, 32×32 이미지에서는 1.5km, 64×64 이미지에서는 0.75km, 그리고 128×128 인 이미지에서는 0.375km의 가로 및 세로의 길이를 나타낸다.

4.2 CAE 훈련

제시된 CAE 모형은 Python과 Keras를 이용하여 구현하였다. 전처리를 통하여 변환된 이미지를 입력으로 사용하였으며, 입력된 이미지의 80%를 훈련데이터로 사용하고, 20%를 검증 데이터로 사용하였다. 입력된 데이터에 대하여, Adam optimizer, learning rate 0.01, loss 함수 binary cross entropy, epoch 100을 사용하여 Google Colab에서 GPU 옵션으로 프로그램을 실행하였다.

64×64 입력 이미지의 경우, <Figure 3>과 같은 복원 오차 분포를 얻었으며, 32×32 및 128×128 경우도 복원 오차는 유사한 모양의 분포를 보였다.

임계값 선택에 따른 식별률의 차이를 보여주기 위하여,

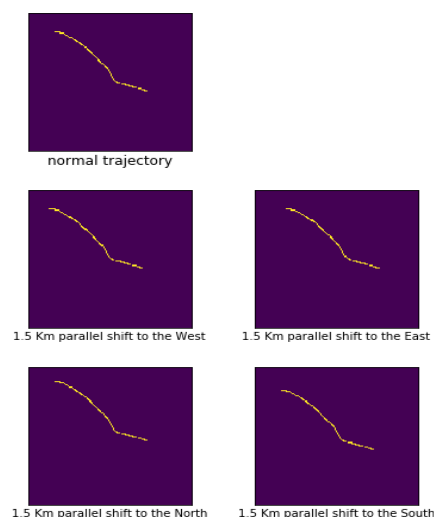


<Figure 3> Histogram and KDE Plot for Error Distribution for 128×128 Normal Trajectories

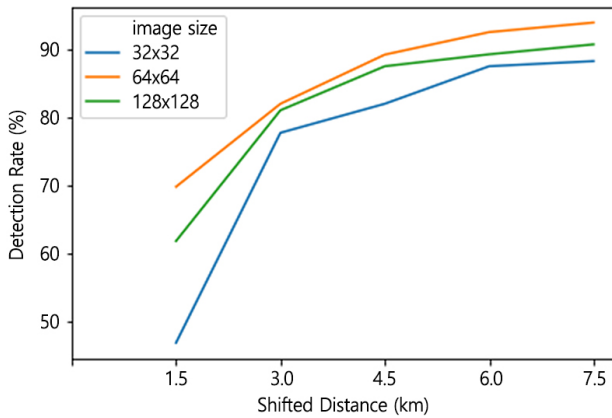
본 논문에서는 복원 오차 분포로부터 품질관리에서 널리 쓰이는 (평균+2 시그마) 및 (평균+3 시그마)를 임의의 임계값으로 설정하여 실험을 진행하였다.

4.3 비정상 궤적을 입력한 결과 및 분석

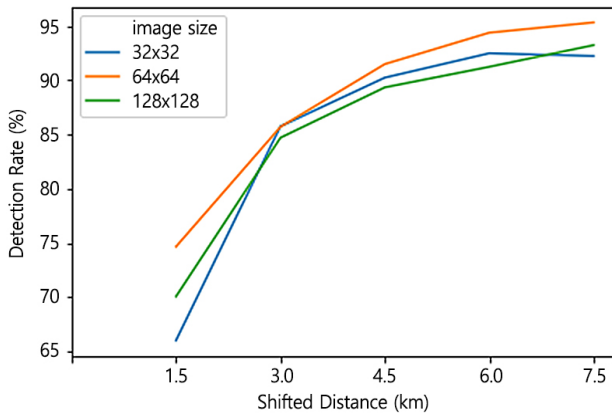
해상을 통하여 불법적으로 국내 육지에 접근한 선박의 경로에 관한 과거 데이터는 극히 드물 뿐 아니라 데이터 획득도 어렵다. 따라서, 기존의 정상 경로를 벗어난 이상거동 선박의 궤적을 만들어 제안된 모형에 입력하고 이 선박을 이상거동 선박으로 분류하는지를 실험한다. 본 논문에서는 정상 궤적 100개를 무작위로 선택하여 동서남북 방향으로 일정 거리를 평행하게 이동한 비정상 궤적을 평행 이동 거리(32×32 이미지 기준 1픽셀-5픽셀)별로 400개씩 만들고, 제시된 CAE 모형에 입력하여 해상도 32×32 , 64×64 , 128×128 에 따라 이상거동 궤적인지를 구별하는 식별률을 구하였다. <Figure 4>는 정상 궤적의 이미지와 이 이미지로부터 4픽셀(실제 거리 1.5km)만큼 동서남북으로 평행 이동하여 만들어낸 이상거동 궤적의 128×128 이미지이다.



<Figure 4> Abnormal Trajectory Images Generated from a 128×128 Normal Trajectory Image

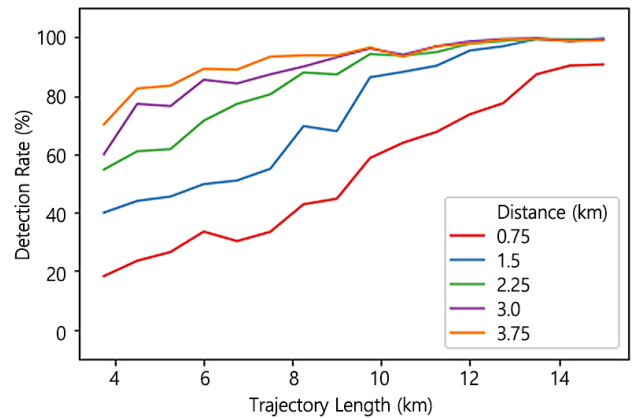


<Figure 5> Detection Rate According to the Parallel Shifted Distance for 3 Sigma Threshold



<Figure 6> Detection Rate According to the Parallel Shifted Distance for 2 Sigma Threshold

<Figure 5>는 복원 오차의 임계값을 (평균+3 시그마)로 하였을 때, 이상거동 궤적을 생성할 때 사용한 평행 이동한 거리와 각 해상도별 식별률과의 관계를 나타낸다. 선박의 궤적을 32×32 이미지로 나타내는 경우가 64×64 및 128×128 이미지로 나타내는 경우보다 식별률이 낮음을 보여준다. 전체적인 True Positive Rate(비정상 궤적을 비정상 궤적으로 판단하는 비율, 본 논문의 식별률)는, 32×32, 64×64 및 128×128 이미지에 대하여 각각 76.5%, 85.5%, 82%를 보였으며, False Positive Rate(정상 궤적을 비정상 궤적으로 판단하는 비율)는 각각 1.6%, 1.8%, 1.8%로 나타났다. <Figure 6>은 복원 오차의 임계값을 (평균+2 시그마)로 하였을 때, 이상거동 궤적을 생성할 때 사용한 평행 이동한 거리에 따른 각 해상도별 식별률을 보여준다. 이 경우에도 32×32 이미지가 약간 낮은 식별률을 보인다. 전체적인 True Positive Rate는, 32×32, 64×64 및 128×128 이미지에 대하여 각각 83.25%, 88.25%, 85.75%를 보였으며, False Positive Rate는 4.1%,

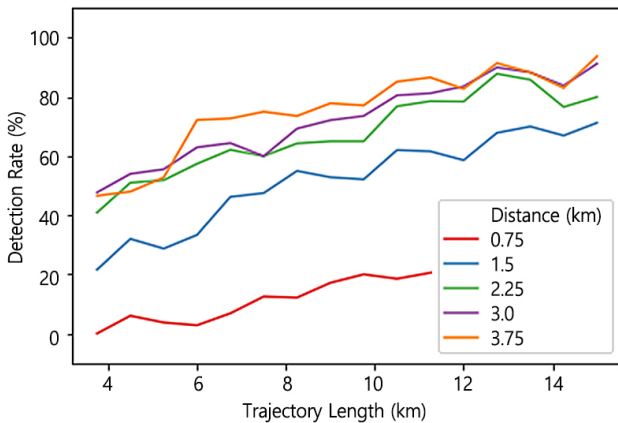


<Figure 7> Relationship Among Detection Rate, Trajectory Length and Shifted Distance for 64x64 Input Image Size and 3 Sigma Threshold

4.3%, 4.4%로 나타났다. 또한, <Figure 5>와 <Figure 6>에서 알 수 있듯이 임계값을 (평균+2 시그마)로 하는 것이 (평균+3 시그마)로 할 때보다 이상거동 선박의 식별률을 높일 수 있음을 알 수 있다. 그러나, 임계값을 (평균+2 시그마)로 하는 경우, 정상 선박을 이상거동 선박으로 잘못 식별하는 False Positive Rate가 늘어나게 되므로, 두 임계값 사이에 상충관계가 존재한다. 본 논문에서 제시된 모형을 실제로 적용할 때에는, 주어진 해상 감시 영역의 크기와 해상 감시 영역에 드나드는 선박의 수, 해상 감시 자원(모니터링 요원 수, 검문 선박 척수와 인원 등), 그리고 비상/정상 근무 여부 등을 고려하여 각 해상 감시 단위별로 결정되어야 한다. 무엇보다도, 해상 보안 측면에서, 이상거동 선박의 식별에 실패했을 때의 유무형의 비용이 반드시 고려되어야 한다.

이상거동 선박은 초기에 식별할수록 상황에 대처할 시간이 늘어난다. 따라서, 궤적의 길이에 따른 이상거동 식별률을 살펴보는 것은 의미가 있다. 이를 위하여 궤적의 길이가 3.75km부터 15km로 변할 때 매 0.75km(128×128 기준 4픽셀, 64×64 기준 2픽셀)마다 50개의 정상 궤적을 무작위로 선택하여 동서남북 방향으로 평행으로 이동한 비정상 궤적을 각 200개를 만들었다. 정상 궤적의 수가 50개 미만인 일부 궤적 길이의 경우, 정상 궤적을 모두 사용하여 비정상 궤적을 만들었으며, 따라서 비정상 궤적이 200개 미만인 경우도 있다.

<Figure 7>은 복원 오차의 임계값을 (평균+3 시그마)로 하였을 때, 64×64 이미지를 입력으로 하는 경우, 궤적의 길이가 3.75km(5픽셀)부터 15km(20픽셀)로 변할 때, 평행 이동한 거리에 따른 이상거동 선박의 식별률을 보여준다. 이 경우 정확한 궤적의 길이는 이미지만다 다르나, 궤적이 남북으로 또는 동서로만 되어있다고 가정하여 계산하였다.



<Figure 8> Relationship Among Detection Rate, Trajectory Length and Shifted Distance for 128x128 Input Image Size and 3 Sigma Threshold

<Figure 8>은 복원 오차의 임계값을 (평균 + 3 시그마)로 하였을 때, 128x128 이미지를 입력으로 하는 경우, 궤적의 길이가 3.75km(10픽셀)부터 15km(40픽셀)로 변할 때, 평행 이동한 거리에 따른 이상거동 선박의 식별률을 보여준다. <Figure 7>과 <Figure 8>로부터 64x64 이미지를 사용하는 경우가 128x128 이미지를 사용하는 경우보다 식별률이 높음을 알 수 있다. 궤적의 크기가 작을 때에도 64x64 이미지를 사용하는 경우가 이상거동 선박의 식별률이 상대적으로 높았다. 아울러 이미지의 크기에 관계 없이 궤적의 길이가 늘어날수록 이상거동 선박의 식별률이 높아지는 것을 알 수 있다. 복원 오차의 임계값이 (평균 + 2 시그마)인 경우에도 전체적인 경향이 유사하였다.

본 실험에서 항로를 64x64 크기의 이미지로 나타낼 때가, 다른 두 크기의 이미지로 나타낼 때보다 식별률이 높았으나, 이를 일반화 시킬 근거는 없다. 따라서, 각 해안경계 단위마다, 주어진 입력에 따른 모형의 실험 결과에 사용할 이미지의 크기를 결정하여야 한다. 아울러, 기후의 변화, 어장의 변화 등에 따라 정상 항로가 달라질 수 있으므로 CAE 모형을 주기적으로 새롭게 훈련할 필요가 있다.

5. 결 론

본 논문에서는 선박이 발송하는 AIS 데이터 중에서 선박식별번호, 위도, 경도, AIS 데이터 발송 시각을 입력 데이터로 사용하여, 이상거동 선박을 식별하는 모형을 제안하였다. 정상 항로를 픽셀 수에 따라 평행 이동시킨 이상거동 궤적을 만들고, 이를 입력 이미지 크기와 임계값을 변화시켜 식별률을 구하였다. 이러한 시뮬레이션을 통해 입력 이미지의 크기와 임계값은 해상 감시 단위마다 상황에 따라 다르게 결정될 수 있음을 알 수 있었다.

본 연구에서는 이상거동 궤적을 체계적으로 발생시켜 평행 이동한 거리에 따른 식별률을 객관적인 성능 지표로 제시하였다. 물론, 발생시킨 이상거동 궤적은 발생 가능한 이상거동 궤적의 한 종류에 해당한다는 한계는 있다. 이러한 한계에도 불구하고, 지금까지 이상거동 선박 탐지 연구에서 객관적인 성능 지표가 전혀 없으므로, 본 논문에서 제시한 평행 이동 거리별 성능평가가 앞으로 성능평가의 객관적인 척도가 될 수 있을 것으로 기대한다.

신경망을 이용한 이상거동 선박을 탐지하는 대부분의 연구는 3분 또는 10분 단위와 같은 일정 시간의 궤적을 바탕으로 이상거동 여부를 판별한다. 이는 일정 시간 내의 관점에서 이상거동을 식별하는 장점이 있으나, 그 시간 이전의 정보는 거의 소실되는 단점을 가진다. 그러나 본 연구에서는 판별이 요구되는 선박의 AIS 신호가 최초로 수신된 시점부터 사전 프로세싱이 시작되는 시점까지 전체 궤적 관점에서 이상거동을 식별하므로 다른 연구와 차별된다.

비정상 궤적을 생성하여 수행한 실험의 결과는 본 논문이 실제 상황에서 이상거동 선박을 식별하는 시스템으로 활용 가능함을 보여주었다. 그러나, 식별률이 현장에서 바로 사용할 만큼 높지 않았다. 그 이유 중 하나는 AIS 데이터의 일부만 사용함으로써 속도와 선박의 종류 등 여러 유용한 정보를 모형에 반영하지 않았다는 점이다. 예를 들어, 정상 속도가 아닌 선박이 정상 궤적으로 운항한다 하여도 제시된 모형은 이를 이상거동 선박으로 분류하지 못한다. 아울러, 본 논문의 실험에서 입출항을 분리하여 CAE를 훈련하지 않았고, 선박의 종류에 따라 분리하여 훈련하지도 않았다. 입출항별, 선박종류별로 각기 다른 CAE를 훈련하여 적용한다면 식별률이 높아질 것으로 기대되며, 따라서 실용성도 높아질 것으로 예상된다.

장기적인 연구과제로, 선박의 추후 항로를 예측하는 기능을 가진다면 선박의 궤적이 늘어나는 효과와 연관되므로 이상거동 선박의 식별률에 영향을 줄 것으로 기대한다.

References

- [1] Besse, P., Guillouet, B., Loubes, J., and Royer, F., Review and Perspective for Distance Clustering of Vehicle Trajectories, *IEEE Transactions on Intelligence Transportation Systems*, 2016, Vol. 17, No. 11, pp. 3306-3317.
- [2] Chen, L. and Ng, R., On the Marriage of Lp-norms and Edit Distance, *Proceeding of the International Conference on Very Large Data Bases*, 2004, pp. 1040-1049.
- [3] Chen, L., Ozsu, M., and Oria, V., Robust and Fast

- Similarity Search for Moving Object Trajectories, *Proceedings of the International Conference on Management of Data*, 2005, pp. 491-502.
- [4] Fu, P., Wang, H., Liu, K., Hu, X., and Zhang, H., Finding Abnormal Vessel Trajectories Using Feature Learning, *IEEE Access*, 2017, Vol. 5, pp. 7898-7909.
- [5] Han, H., Armenakis, C., and Jaddi, M., DBSCAN Optimization for Improving Marine Trajectory Clustering and Anomaly Detection, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, 2020, Vol. XLIII-B4-2020, pp. 445-461.
- [6] Iltanan, H., Maritime Anomaly Detection using Autoencoders and OPTICS-OF [Mastre's Thesis], [Helsinki, Finland] : University of Helsinki, 2020.
- [7] Kim, H. and Kim, J., A Heuristic Algorithm for a Ship Speed and Bunkering Decision Problem, *Journal of Society of Korea Industrial and Systems Engineering*, 2016, Vol. 39, No. 2, pp. 19-27.
- [8] Kim, H. and Kim, J., Determining Economic Ship Speeds and Fleet Sizes Considering Greenhouse Gas Emissions, *Journal of Society of Korea Industrial and Systems Engineering*, 2011, Vol. 34, No. 2, pp. 49-59.
- [9] Kim, K., Development of Ship Safety Navigation Alarm System using AIS, *Journal of Korean Institute of Information Technology*, 2013, Vol. 11, No. 4, pp. 69-75.
- [10] Kwon, S.H. and Oh, H.S., Reduction of Simulation Number for Ship Handling Safety Assessment, *Journal of Society of Korea Industrial and Systems Engineering*, 2012, Vol. 35, No. 1, pp. 101-106.
- [11] Nguyen, D., Vadaine, R., Hajduch, G., Garelo, R., and Fablet, R., A Multi-task Deep Learning Architecture for Maritime Surveillance using AIS Data Streams, *IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 2018.
- [12] Oh, J-Y., Kim, Y-J., and Park, S-K., Detection of Ship Movement Anomaly using AIS Data : A Study, *Journal of Korean Navigation and Port Research*, 2018, Vol. 42, No. 4, pp. 277-282.
- [13] Oh, J-Y., Kim, Y-J., and Park, S-K., Development of a Decision Support System based on Autoencoder for Vessel Traffic Service, *KIISE Transactions on Computing Practices*, 2018, Vol. 24, No. 12, pp. 642-648.
- [14] Palliotta, G. and Joussetme, A-L., Data-driven Detection and Context-based Classification of Maritime Anomalies, *The 18th International Conference on Information Fusion*, 2015, pp. 1152-1159.
- [15] Pallotta, G., Vespe, M., and Bryan, K., Traffic Knowledge Discovery from AIS data, *Proceedings of the 16th International Conference on Information Fusion*, 2013, pp. 1996-2003.
- [16] Park, J. and Kim, S., Maritime Anomaly Detection Based on VAE-CUSUM Monitoring System, *Journal of the Korean Institute of Industrial Engineers*, 2020, Vol. 46, No. 4, pp. 432-442.
- [17] Vlachos, M., Kollios, G., and Gunopulos, D., Discovering Similar Multidimensional Trajectories, *Proceeding of the International Conference on Data Engineering*, 2002, pp. 673-684.
- [18] Yu, J.Y., Sghaier, M.O., and Grabowiecka, Z., Deep Learning Approaches for AIS Data Association in the Context of Maritime Domain Awareness, *IEEE 23rd International Conference on Information Fusion*, 2020, pp. 1-8.

ORCID

June-Hyoung Son | <http://orcid.org/0000-0002-5021-8900>
 Jun-Gun Jang | <http://orcid.org/0000-0002-8787-2922>
 Bongwan Choi | <http://orcid.org/0000-0002-9609-1714>
 Kyeongtaek Kim | <http://orcid.org/0000-0002-5636-2487>