



Technical Report

선박평형수 내 동물플랑크톤(>50 μm 생물) 분류를 위한 인공지능 적용

황옥명¹ · 황세선² · 최규성² · 박혜진¹ · 권재형² · 이광호^{3,†}

¹㈜ 테크로스 중앙연구소 선임연구원

²㈜ 테크로스 중앙연구소 전임연구원

³㈜ 테크로스 중앙연구소 수석연구원

The Application of Artificial Intelligence to Classify Zooplankton (>50 μm) in Ballast Water

Okmyung Hwang¹, Seseon Hwang², Kyusung Choi², Hyejin Park¹,
Jaehyeong Kwon², and Gwangho Lee^{3,†}

¹Research Enginner, Techcross Inc, Busan 46758, Korea

²Assistant Research Enginner, Techcross Inc, Busan 46758, Korea

³Principal Research Enginner, Techcross Inc, Busan 46758, Korea

요 약

국제해사기구(IMO)에서는 해양 외래종 유입으로 인한 생태 및 경제적인 피해를 방지하기 위해 2004년 선박평형수 및 침전물의 배출규제와 관리에 관한 협약을 채택하여 선박평형수 처리를 요구하고 있다. 본 연구에서는 $\geq 50 \mu\text{m}$ 생물을 분류 및 판별하기 위한 영상 처리 기술(OpenCV)과 기계 학습 엔진(TensorFlow)을 접목하여 선박평형수 처리 장치를 검증하기 위한 도구로써 해당 기술의 적용 가능성을 확인 하였다. 2017년 3월에서 2018년 9월까지 이미지 수집 작업이 이루어졌고, 이를 바탕으로 2회에 걸쳐 기계학습을 진행하였다. 확보된 이미지는 1차시기에 29,032장, 2차 시기에 86,044장을 수집하였고, 인공 지능이 적용된 시스템의 평균 적중률은 각각 75.5%, 94.8%로 확인되었다. 2차 시기에 모든 분류군의 적중률이 90% 이상을 보였고, 전체적으로 정보량이 늘어남에 따라 동물플랑크톤에 대한 인식률은 증가하는 양상을 보였으나, 인공지능을 적용하는 과정에서 인지된 변수들을 원활하게 통제하기 위한 노력을 기울여야 할 것으로 본다.

Abstract – The International Maritime Organization (IMO) stipulates that the transportation of ship’s ballast water among geographically different regions is one of the greatest threats to the marine environment. In the efforts of international cooperation among IMO member states, the international convention for the control and management of ship’s ballast water and sediments, 2004(BWM convention 2004) was adapted to prevent ecological and economical damages caused by ship’s ballast water operation. This study focused on the verification of a device which can classify and identify the organism bigger than 50 μm ($\geq 50 \mu\text{m}$) by combining technologies between image processing(OpenCV) and mechanical training engine (TensorFlow). Image collection was carried out from March 2017 to September 2018. 29,032 images were obtained in the first period and 86,044 were obtained in the second period, and the accuracy rate for systems with artificial intelligence was 75.5% and 94.8%, respectively. In the second period, the accuracy rate of all Taxon showed more than 90%. As the overall learned information increased, the recognition rate for zooplankton appeared to be increasing, but it is necessity that the perceived variables should be controlled in the process of artificial intelligence application.

Keywords: Ballast water(선박평형수), Zooplankton(동물플랑크톤), Artificial Intelligence(인공지능), Machine Learning(기계학습), International Maritime Organization(IMO / 국제해사기구)

[†]Corresponding author: ghlee@techcross.com

1. 서 론

개방 경제의 확대는 전세계적으로 물동량의 증대를 야기 시켰고, 해상 운송은 대량의 물류를 비교적 저렴한 비용으로 수송할 수 있는 방법으로 세계 물동량의 80~90%를 담당하고 있다. 선박은 운항 및 화물 적재 과정에서 발생할 수 있는 불균형적인 문제를 내부 탱크에 물을 주입하거나 비움으로써 해결한다. 이때 채우는 물을 선박평형수(ballast water)라고 하며, 전 세계적으로 연간 100억톤 이상이 이동되는 것으로 추정하고 있다(Battle[2009]). 선박평형수를 주입하게 되면 물 뿐만 아니라 그 속에 존재하는 많은 수의 생물(plankton)도 함께 따라오게 되는데, 전 세계적으로 매년 7,000종 이상이 선박평형수를 통해 다른 해역으로 이동하는 것으로 추정되고 있다(Battle[2009]). 이러한 생물들은 선박이 이동함에 따라 탱크 내부에서 대부분 사멸하거나(Drake *et al.*[2002]; Gollasch and Leppäkoski [2007]), 선박평형수가 배출되는 지역의 물리학적, 생물학적 환경에 적응하지 못하고 자연 도태하게 된다(Gollasch *et al.*[2000]; Baek and Shin[2009]). 그러나 일부 생존 개체는 새로운 환경에 적응하여 기존의 생태계를 교란할 수 있는 가능성을 내포하고 있다(Baek and Shin[2009]). 생물학적 다양성에 기반한 자원의 중요성과 관심이 증대됨에 따라 외부에서 유입된 생물이 야기시키는 생태교란은 미래 자원 확보 및 보호라는 측면에서 국가적으로 매우 중대한 사안이다. 이에 국제연합(UN) 산하 국제해사기구(IMO)에서는 선박평형수를 세계 해양을 위협하는 심각한 4대 위협 중 하나로 규정하고, 해양 외래종 유입으로 인한 생태계 및 경제적 피해를 방지하기 위해 2004년 '선박평형수 및 침전물의 배출규제와 관리에 관한 협약'을 채택하여 선박평형수 처리 장비를 설치하고, 배출기준에 부합하도록 요구하고 있다(IMO[2004]).

선박평형수에 대한 법안이 발효(2017년 9월 발효, 기존선박은 2년 유예)됨에 따라 처리 기술의 개발뿐만 아니라 선박평형수 처리 장비의 정상 작동 유무를 판단하고, 선박평형수 배출 과정이 기준(D2)에 부합하는지 검증하는 단계도 중요해졌다. 해양 미소 생물에 대한 생사를 관찰함에 있어 $\geq 50 \mu\text{m}$ 생물(주로 동물플랑크톤)은 기본적으로 몸체, 편모, 또는 부속 지의 운동성에 기반하여 관찰하는 방법(Wright *et al.*[2010])을 들 수 있으며, 그 밖에 Neutral red(Elliott and Tang[2009]; Hyun *et al.*[2014]), Aniline blue(Bickel *et al.*[2009]; Baek and Shin[2015]) 등의 염색 시약을 이용한 관찰 방법들을 선행 연구에서 제시한바 있다. 또한 $\geq 10 \mu\text{m}$ and $50 < \mu\text{m}$ 생물(주로 식물플랑크톤)의 경우 자가형광방법과 FDA(Fluorescein Diacetate Assay), CMFDA(5-Chloromethylfluorescein diacetate) 등의 생체 염색법이 주로 이용되고 있다(Baek and Shin[2009]). 그러나 각 분류군에 따라 반응 특성이 다르고(Baek and Shin[2015]), 관찰자 개인의 주관적인 판단이 개입될 소지가 있으며, 실제 현장에서 바로 적용하기는 어렵다. 따라서 즉시 적용 가능하고, 전문적인 지식 없이(Matuszewski[2014])도 접근이 용이한 지표 분석 방법의 개발이 필요하다.

동물플랑크톤을 분류하기 위해 현재까지 가장 보편적으로 이용

하는 방법은 분류학적 특징을 숙지하고, 이를 활용하여 분류군을 나눌 수 있는 전문 인력에 의한 분석 방법이다(Benfield *et al.* [2007]). 그러나 이러한 작업은 시간이 오래 걸리고, 해당 생물에 대한 분류학적 전문 지식과 경험이 필요하다(Sieracki *et al.*[2010]; Schulze *et al.*[2013]). 이러한 문제점을 보완하기 위해 최근에는 기계 장비를 이용하여 자동으로 생물을 분류하는 방법들이 시도되고 있다. 이러한 시도 중 현재까지 가장 많이 이용되는 방법은 동물플랑크톤의 형태적 특징들을 추출하여 판별하는 이미지 기술이다(Hu and Davis[2005]; Sieracki *et al.*[2010]). ZooImage(Zoo/Bell and Hopcroft[2008]), VPR(Video Plankton Recorder/Hu and Davis[2005]; Leow *et al.*[2015]), ZOOVIS(ZOOplankton VISualization system/Bi *et al.*[2012]), 등 고속 연사 카메라를 이용하여 이미지를 분석하는 장비는 기존에 개발, 연구된 부분이나, 전후 처리에 인력의 도움을 받아야 한다. 또한 많은 연구들이 동물플랑크톤의 생태 및 개체 분포 현황 조사에 목적을 두고 있으며(Hu and Davis[2005] Bi *et al.* [2012] Bi *et al.*[2015]), 특히 국내에는 확보된 생물의 이미지를 기반으로 하는 시스템 연구가 미미한 실정이다. 본 연구에서는 영상 처리 기술과 최근에 부상하고 있는 인공지능에 의한 기계 학습(machine learning) 능력을 접목하여 한정된 공간에서 동물플랑크톤을 분류하고 개체를 확인하는데 있으며, 차후 선박평형수 처리 상태를 확인하기 위한 전, 후 처리(생물 채집 및 이송, 샘플 고정, 이미지 분류 등) 과정이 필요 없는 통합적이고, 자동화된 지표 분석 방법 개발에 기본 자료를 제공하는데 있다.

2. 도구 및 방법

2.1 동물플랑크톤 수집 및 처리

이미지 처리에 이용할 동물플랑크톤의 채집은 2017년 3~10월, 2018년 2~9월에 걸쳐 거제 장목(34°59'N, 128°40'E)과 부산 신항(35°4'N, 128°50'E), 마산 내만(35°12'N, 128°35'E)에서 실시하였다. 체포기 50 μm 보다 큰 생물을 채집하기 위해 망목 32 μm 원추형 네트(망구 30 cm, 길이 60 cm)를 이용하여 저층에서 표층까지 수직 예인 하였다. 채집된 시료는 500 ml 무균 채수병(pet)에 옮겨 담았고, 동물플랑크톤을 분류하기 위해 사용할 표층 해수 2리터를 채수 하였다. 온도 변화를 최소화하고, 생존 개체를 최대한 확보하기 위해 아이스 팩이 구비된 보관함에 넣어 신속히 실험실로 이동하였다.

동물플랑크톤을 분류하기 위해 사용될 해수는 불순물을 제거하기 위해 5 μm sieve 를 이용하여 거른 후 500 ml 유리병에 나누어 각각 보관하였다. 채집된 동물플랑크톤은 형태학적으로 동일한 요소를 가진 상위 그룹인 phylum 단위로 구분하였으며(Table 1 참조), 살아있는 개체를 대상으로 페트리디쉬에 분주하여 해부현미경(carl zeiss, stemi 200-c) 하에서 1 ml 스포이드를 이용하여 각 분류군에 따라 분류(sorting)하였다. 분류된 각각의 개체는 영상 촬영을 위한 챔버(zooplankton counting chamber)에 배분하였다. 생물에 대한 영상 촬영은 이미지의 잔상을 최소화 하기 위해 현미경용 고속 카메라(medicam-k/60fps, 2 megapixel)를 이용하였고, 노이즈를 줄이

Table 1. Classification groups used for machine learning

		Zooplankton groups		
Phylum	Subphylum	Class	Subclass	Order
Ciliophora				Tintinnida
Cnidaria		Hydrozoa		
Rotifera				
Annelida		Polychaeta		
Chaetognatha				
		Braanchiopoda		Cladocera
				Calanoida
Arthropoda	Crustacea	Maxillopoda	Copepoda	Cyclopoida Harpacticoida
				Cirripedia
		Malacostraca		Amphipoda Decapoda
Chordata				
Echinodermata				
Mollusca		Gastropoda Bivalvia		
Nematoda				
Platyhelminthes				

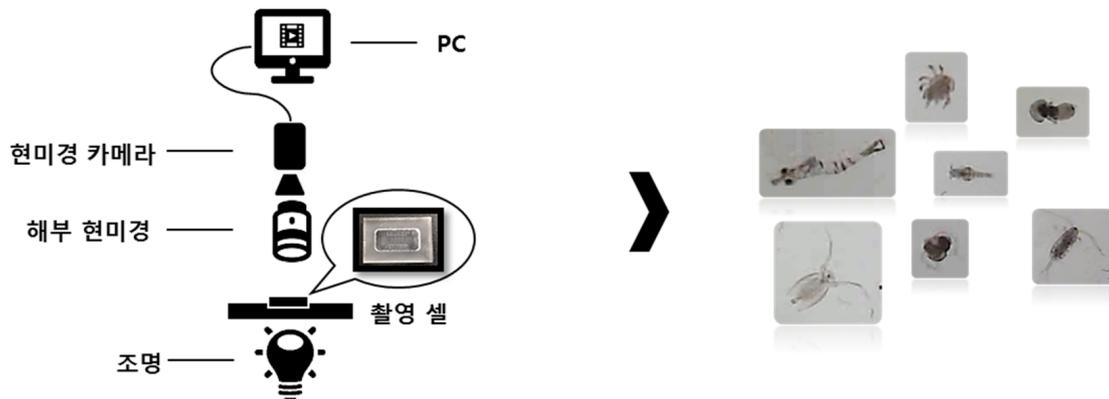


Fig. 1. Overview of image system design.

고, 경계선이 분명한 이미지를 확보하기 위해 백색 조명을 사용 하였다(Fig. 1 참조).

2.2 기계 학습에 이용된 도구 및 원리

본 연구에서는 동물플랑크톤을 인식하기 위한 기계 학습 구현을 위해 현재 인공지능 시스템 구축에 선호도가 높고, 다양한 분야와 환경에 응용되고 있어 풍부한 자원 환경(이용자 층, 적용 분야)을 보유하고 있는 Tensorflow(machine learning engine) 라이브러리(library)를 사용하였으며, 움직임 검출 및 이미지 추출을 위해 컴퓨터 비전 라이브러리 OpenCV(open source computer vision)을 사용하였다.

동물플랑크톤의 이미지 학습 및 추론을 위해 Tensorflow의 다양한 모델 중 이미지 인식에 성능이 좋은 합성곱 신경망(CNN - convolutional neural network) 모델을 적용하여 진행하였다. 이미지에는 경계선,

색상, 밝기 등 사물의 특징을 나타내는 다양한 정보들이 포함되어 있다. CNN 모델의 이미지 인식은 합성곱(convolution)으로 구성된 다수의 층에서 추출된 이미지의 서로 다른 특징을 감지하여 학습 모델을 만들어 내고, 제시된 이미지(test image)에서 특정 표현이 존재하는 비율에 따라 확률적으로 예측한다. 본 연구에서는 동물플랑크톤의 이미지를 phylum 단위로 분류하여 표지(labeling) 한 후 지도학습(supervised learning)을 통해 학습 모델을 생성하고, 이를 기반으로 제시된 동물플랑크톤의 분류군을 예측하였다(Fig. 2 참조). 또한 수 만장의 동물플랑크톤 이미지에 대한 학습을 빠르게 진행하기 위해 고성능 PC(CPU: i7-6850k, RAM: ddr4 64g, SSD: 850 pro 2tb)와 기계 학습의 병렬 데이터 연산 처리를 빠르게 수행하는 GPU(gtx1080ti)를 사용하였다.

기계 학습을 위한 이미지를 확보하기 위해 촬영용 챔버에 일정(수~수십) 개체의 살아있는 동물플랑크톤을 넣고 영상을 촬영하여

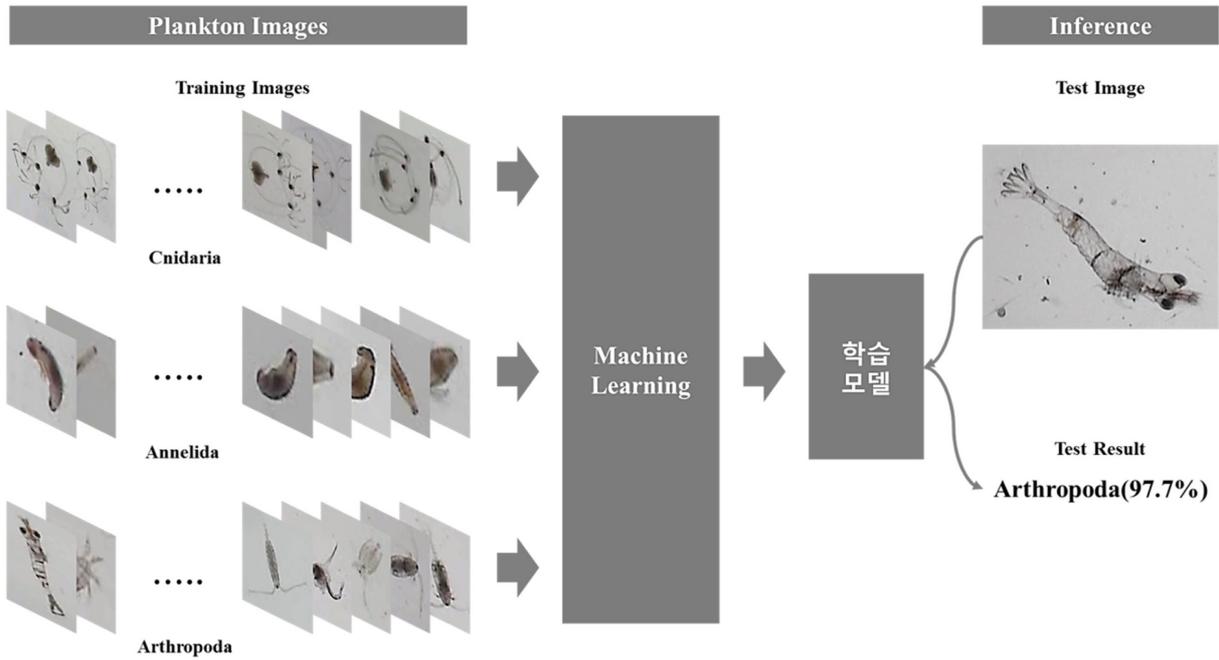


Fig. 2. Machine learning and inference concepts for identification objects.

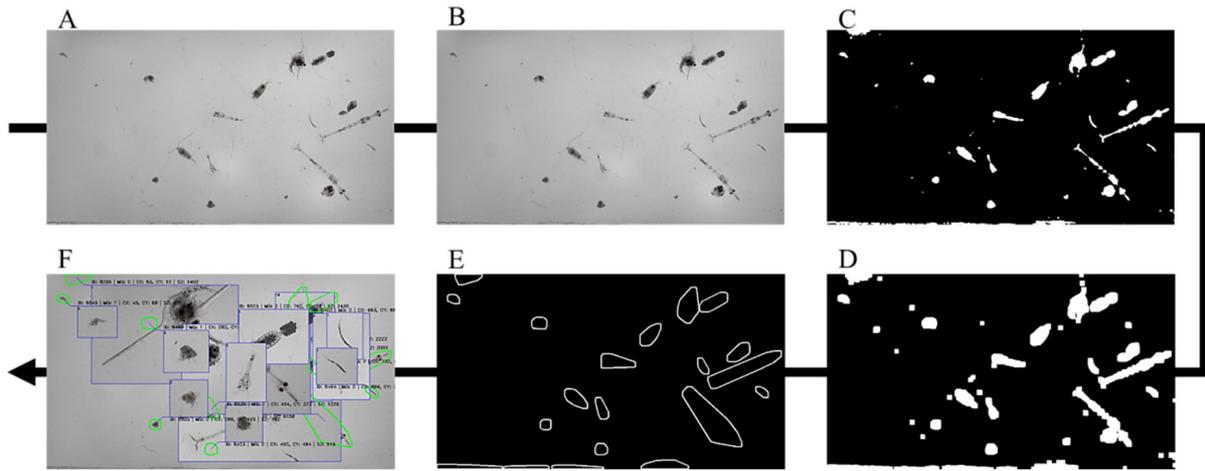


Fig. 3. Object detection process in image (A: Original image, B: Gray scale, C: Thresholding, D: Dilate, E: Contour, F: Organism detection).

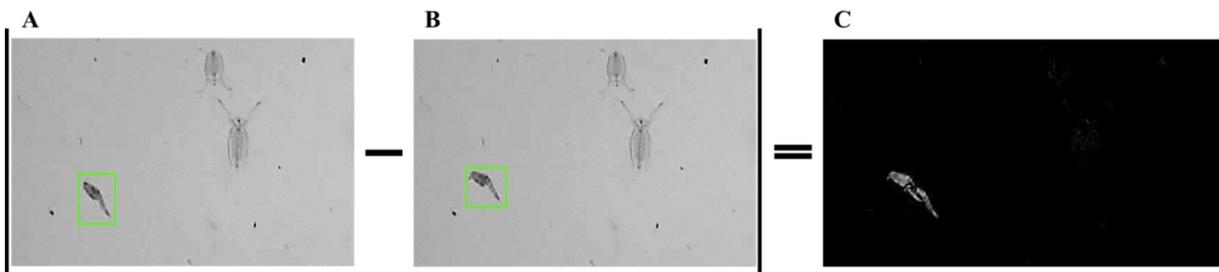


Fig. 4. Object tracking process in image(A: Previous frame, B: Current frame, C: Object tracking).

OpenCV를 활용한 프로그램을 작성하여 동물플랑크톤의 이미지를 추출하였다. 이미지 추출 과정에서 OpenCV의 이미지 연산 함수인 Gray scale, Blur, Thresholding, Contour 기능을 사용하여 한 프레

임에 존재하는 여러 개체 중 하나의 개체를 추출하였다(Fig. 3 참조). 추론 대상은 방향에 따라 형태가 다르게 보일 수 있기 때문에 프레임 간의 차이를 연산(Fig. 4 참조)하여 변화가 발생한 부분을 감



Fig. 5. Image taken part of the body.

지하고 이를 추출하여 객체(동물플랑크톤)에 대한 다양한 방향의 이미지를 확보하였다. 또한 이미지 학습과정에서 불분명 하거나 생물의 일부분만 촬영된 이미지는 제외하였다(Fig. 5 참조).

3. 결과 및 고찰

본 연구는 ‘자동화된 지표 분석 방법 개발’의 일환으로 진행되었으며, 선박평형수의 처리 상태를 검증할 수 있는 시스템 구축을 위한 인공지능 구현에 다양한 분야(보안분야, 산업분야, 농업분야, 경제 동향 예측, 자료 예측 등)에서 선호도가 매우 높은 Tensorflow를 이용하였다. 이러한 폭넓은 이용자 층은 운영상의 문제점이 발생했을 때 기존의 여러 사례들과 연계하여 효율적으로 대응 가능하고, 또한 향후 개선 방향 설정에 유리한 장점을 내포하고 있다. 자동화된 통합시스템을 구현하기 위해서는 샘플 수집, 이송 및 처리, 분류 등의 작업을 포함하고 있으나, 본 연구에서는 구현된 인공지능 하에서 50 μm 이상의 살아있는 생물(주로 동물플랑크톤)에 대한 분류 및 검출 능력을 확인하였다.

3.1 시스템에 적용된 이미지

개체를 분류하기 위해 동물플랑크톤의 형태적 특징들을 기계학습 시켜 1, 2차에 걸쳐서 인식을 결과를 확인하였으며, 시스템 상에서 분류 값을 출력하기 위한 동물플랑크톤은 1차 시기에 10개 항목으로 분류하였고, 2차 시기에는 출현 빈도가 낮은 분류군을 세분화하여 총 12개 분류군으로 나눌 수 있었다. 1차에서는 각 항목 별로 124~10,257, 총 29,032장의 이미지를 확보하여 기계학습 시켰고, 2차에서는 각 분류군 별로 124~51,719, 총 86,044장의 학습 이미

지를 확보하여 기계학습에 적용하였다. 확보된 이미지는 1차 시기에 연체동물(Mollusca)이 10,257개로 가장 많았고, 모악동물(Chaetognatha)이 124개로 가장 적었다(Table 2 참조). 2차 시기에 확보된 이미지는 절지동물(Arthropoda)이 51,719장으로 가장 많았으며, 극피동물(Echinodermata)이 124장으로 가장 적었다(Table 3 참조). 절지 동물과 같이 다양한 속 및 종이 출현하고, 샘플 시기에 주요 우점 분류군의 경우 많은 이미지를 확보할 수 있었으며, 상대적으로 출현 빈도가 낮아서 생존 개체를 확보하기 힘든 분류군은 적은 수의 이미지가 확보되었다.

3.2 시스템 적중률

1차 시기에는 총 29,032개의 이미지를 이용하여 기계학습 시켰으며, 이를 바탕으로 총 5,007개의 이미지에 대한 인식률을 알아보고, 전체적인 적중률은 평균 75.5%를 보였다(Table 2 참조). 이는 70~80% 적중률 범위를 보였던 기존 연구된 이미지 시스템 결과와 비교하여 다소 낮거나 유사한 확률을 보였다(Table 4 참조).

각 항목 별 적중률은 해당 해역에서 잘 발견되지 않은 분류군이 포함된 기타(Others) 항목과 복족류 및 이매패류가 포함된 연체동물이 각각 40.4%, 67.3%를 보여 상대적으로 적중률이 낮았고, 섬모충류(Ciliophora), 윤형동물(Rotifera), 척삭동물(Chordata) 세 분류군의 경우 각각 76.4%, 77.5%, 74.9%의 결과를 보여 평균 적중률과 유사한 값을 보였다. 자포동물(Cnidaria), 절지동물(Arthropoda), 모악동물(Chaetognatha), 환형동물(Annelida) 네 가지 분류군은 각각 83.7%, 80.6%, 96.8%, 82.1%로 상대적으로 적중률이 높았다(Table 2 참조).

2차 시기에는 1차 시기와 달리 인공지능에 의한 적중률 확인을

Table 2. Number of learning images for each classification group (first period, 2017)

Category	Training Images	Test Image	Accuracy rate (%)
Ciliophora	2,696	368	76.4
Cnidaria	925	338	83.7
Rotifera	1,738	889	77.5
Arthropoda	6,378	401	80.6
Chaetognatha	124	93	96.8
Chordata	1,090	399	74.9
Mollusca	10,257	1,513	67.3
Annelida	1,958	340	82.1
Others	3,866	666	40.4
sum/average	29,032	5,007	75.5

Table 3. Number of learning images for each classification group (second period, 2018)

Phylum	Training Images	Test Images	Accuracy rate (%)
Annelida	12,654	100	86
Arthropoda	51,719	100	92
Chaetognatha	2,517	100	100
Chordata	2,558	100	92
Ciliophora	2,975	100	95
Cnidaria	6,060	100	99
Echinodermata	124	100	99
Mollusca	3,949	100	84
Nematoda	205	100	100
Platyhelminthes	131	100	100
Rotifera	3,152	100	91
sum/average	86,044	1,200	94.8

Table 4. Results of different(various) image systems

System	Classification group	Accuracy rate (%)	Reference
VPR	Hydroida, Copepoda	74.7%	Hu and Davis [2005]
ZooImage	53 categories	81.7%	Bell and Hopcroft [2008]
ZooSCAN	20 categories	78.0%	Gorsky <i>et al.</i> [2010]
ZooVIS	Hydroida, Copepoda, Chaetognatha	>80.0%	Bi <i>et al.</i> [2015]

위해 각 분류군 별로 기계학습을 거치지 않은 100개의 객체(Test image)를 선정하여 결과 값을 도출하였다(Table 3 참조). 2차 시기에는 84%~100% 적중률 범위를 확인할 수 있었고, 전체 적중률은 평균 94.8%로 나타났다. 12개 분류군 중 10개 분류군이 90% 넘는 적중률을 보여 1차 시기에 비해 전체적인 적중률이 높은 것으로 확인되었다. 연체동물 및 환형동물은 각각 84%, 86%의 적중률을 보여 1차 시기와 마찬가지로 다른 항목에 비해 비교적 낮은 적중률을 나타냈고, 가장 많은 학습 이미지를 확보한 절지동물은 비롯하여 척삭동물, 윤형동물은 각각 92.0%, 92.0%, 91.0% 를 보여 전체 평균보다 다소 낮은 적중률을 보였다. 모약동물, 선형동물, 편형동물은 131~2,517개의 이미지를 확보하여 환형동물, 절지동물 등 다른 분류군에 비해 상대적으로 학습된 이미지는 적었으나, 100%의 적중률을 확인할 수 있었다(Table 3 참조).

3.3 시스템의 인식 오류 및 분류군 특성

이미지 시스템의 경우 자동화 측정에 사용된 렌즈의 심도 한계로 한정된 영역을 관찰할 수 밖에 없다. 이러한 점은 큰 개체의 경우 몸체의 일부 구역만 관측되거나(Benfield *et al.*[2007]), 중복 계수되는 문제점으로 작용한다(Bi *et al.*[2015]). 또한 Bi *et al.*[2015]은 다른 개체에 비해 상대적으로 큰 자포동물(Hydroid medusa), 모약동물(Chaetognatha)의 관측 오류가 30~40%에 이를 수 있음을 보고한바 있다. 본 연구에서도 모약동물, 자포동물(Hydroid medusa), 절지동물(*Evadne tergestina*, *Calanus sinicus*) 등 다른 개체에 비해 상대적으로 큰 개체의 경우 이미지를 학습하는 과정에서 부분적으로 인식되거나, 인식 부위에 따라 중복 관측되는 양상을 확인할 수 있었다(Fig. 6, 7 참조). 그러나, 본 연구에서 모약동물 및 자

포동물은 다른 분류군과 비교하여 높은 적중률을 보였다. 개방된 공간(sea)에서 이동하는 큰 개체의 인식은 하드웨어적으로 많은 제약(심도, 촬영 범위)이 따를 것으로 판단되나, 반대로 관찰 구역을 일정한 공간(zooplankton counting chamber)으로 제한하여 개체의 활동성을 구속한다면 오히려 하드웨어적(lens)으로 관찰이 용이한 특성을 보이기 때문으로 풀이된다.

모약동물(700~2500 µm)은 종에 따라 일부 특징적인 모습을 보이나, 전체적으로 긴 화살모양으로 비교적 일정한 형태를 보인다. 1차 시기에는 확보된 정보량(image)이 많지 않음에도 96.8%의 적중률을 보였고, 2차 시기에는 100% 인식률이 확인되어 1, 2차 시기 모두 다른 분류군과 비교하여 높은 수치를 보였다. 시스템 상에서 자포동물의 적중률은 1차, 2차 시기를 통틀어 각각 83.7, 99.0%로 확인되어 전체 평균보다 높게 나타났다. 자포동물은 성장 단계에 따라 형태적 특징이 다르고, 일부 개체(Planula larva)의 경우 길고 둥근 형태적 유사점으로 인해 연체동물, 환형동물 등으로 오판하는 경우도 있는 것으로 나타났다(Table 5 참조).

척삭동물의 주요 분류군으로 확인된 *Oikopleura* sp.는 크게 몸체(trunk)와 투명하고 긴 척삭(notocord and tail)으로 구성되어 있다. 이동할 때 순간적으로 매우 빠른 척삭의 움직임을 확인할 수 있으며, 전체적으로 반투명한 척삭이 휘한 형태에 따라 기계적으로 인식되기 어려운 조건을 내포하고 있다(Fig. 8 참조). 또한 부유물 등의 이물질이 흡착되는 경우가 많아 개체에 대한 정확한 특성 확인이 어려웠을 것으로 예상된다(Tang *et al.*[1998]; Bell and Hopcroft[2008]). 척삭동물의 적중률은 1차, 2차 시기에 각각 74.9%, 92.0%의 적중률로 전체 시스템의 평균 적중률에 비해 다소 낮은 결과를 보였다. 체폭 100 µm 이하의 개체가 다수 발견된 1차 시기에는 섬모충류,

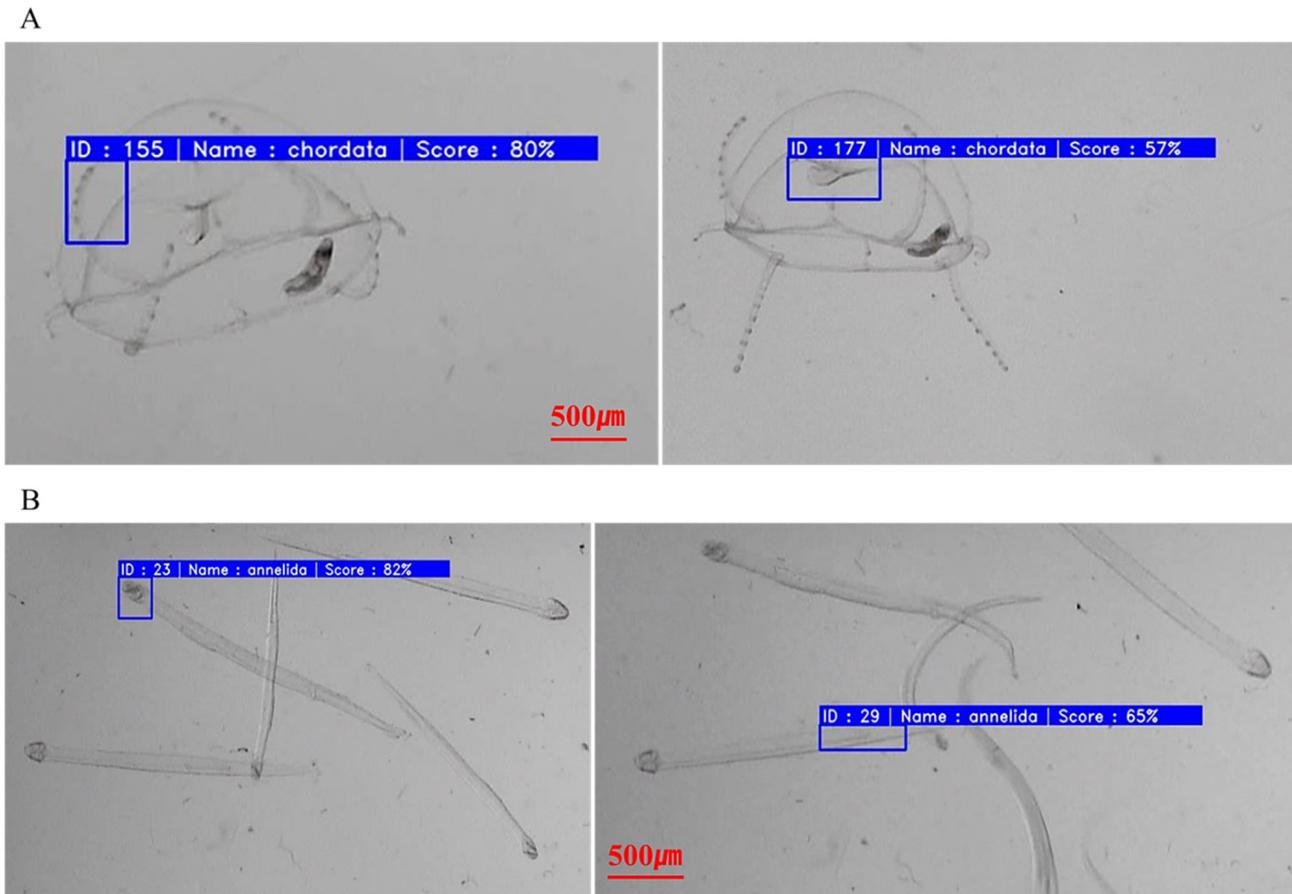


Fig. 6. large objects recognized as part of body (A: Cnidaria, B: Chaetognatha).

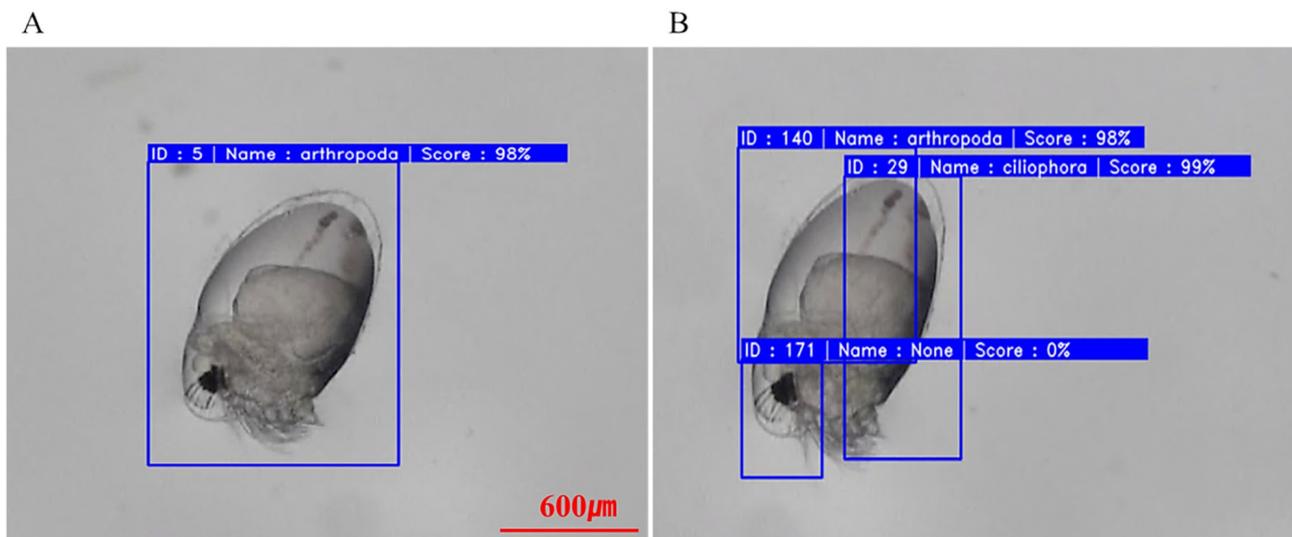


Fig. 7. Image of object recognized as multiple taxonomic groups (A: single reading, B: multiple reading).

소형 절지동물(*Oithona* sp. 등) 등으로 오판하는 비율이 20% 내외로 나타났고(Table 5, Fig. 9 참조), 체폭 200 μm 이상의 개체가 주를 이룬 2차 시기에는 환형동물과, 자포동물 등으로 잘못 인식하는 것으로 나타났는데 척삭동물을 인식함에 있어 앞서 언급했던 여러 특

성들 외에 개체의 크기도 부분적으로 영향을 미치는 것으로 판단된다(Table 6 참조).

동물플랑크톤을 채집한 장목과 부산 신행에서 잘 발견되지 않는 분류군이 포함된 1차 시기의 기타(Others) 항목을 제외하면 두 시

Table 5. Confusion matrix of dataset system evaluation (first period, 2017)

		Confusion group								
		Ann.	Arth.	Chae.	Chor.	Cilio.	Cnid.	Mollu.	Roti.	Others
Target group	Annelida	82.1	2.1	0.0	0.0	0.6	1.8	5.9	2.9	4.6
	Arthropoda	7.7	80.6	1.0	0.5	3.2	0.0	0.3	3.2	3.5
	Chaetognatha	1.1	1.0	96.8	1.1	0.0	0.0	0.0	0.0	0.0
	Chordata	0.5	13.0	0.8	74.9	6.8	2.5	0.0	1.3	0.2
	Ciliophora	1.4	0.5	0.3	0.2	76.4	0.3	0.2	20.7	0.0
	Cnidaria	3.8	0.9	0.0	0.0	2.7	83.7	5.3	2.1	1.5
	Mollusca	5.3	2.9	0.4	1.7	0.9	6.3	67.3	0.7	14.5
	Rotifera	1.6	0.9	0.2	0.8	18.7	0.1	0.2	77.5	0.0
	Others	6.6	3.5	0.2	0.0	25.7	4.8	3.3	15.5	40.4

Table 6. Confusion matrix of dataset system evaluation (second period, 2018)

		Confusion group										
		Ann.	Arth.	Chae.	Chor.	Cilio.	Cni.	Ech.	Mollu.	Nema.	Platy.	Roti.
Target group	Annelida	86	3	0	0	1	0	0	5	0	0	5
	Arthropoda	1	92	0	0	3	0	0	2	0	0	2
	Chaetognatha	0	0	100	0	0	0	0	0	0	0	0
	Chordata	4	0	0	92	0	4	0	0	0	0	0
	Ciliophora	0	1	0	0	95	0	0	0	0	0	4
	Cnidaria	0	0	0	0	0	99	0	0	0	0	1
	Echinodermata	0	0	0	0	0	1	99	0	0	0	0
	Mollusca	2	4	0	0	1	0	0	84	0	0	9
	Nematoda	0	0	0	0	0	0	0	0	100	0	0
	Platyhelminthes	0	0	0	0	0	0	0	0	0	100	0
	Rotifera	1	1	0	0	4	0	0	3	0	0	91

Ann: Annelida, Arth: Arthropoda, Chae: Chaetognatha, Chor: Chordata, Cilio: Ciliophora, Cni: Cnidaria, Ech: Echinodermata, Mollu: Mollusca, Nema: Nematoda, Platy: Platyhelminthes, Roti: Rotifera

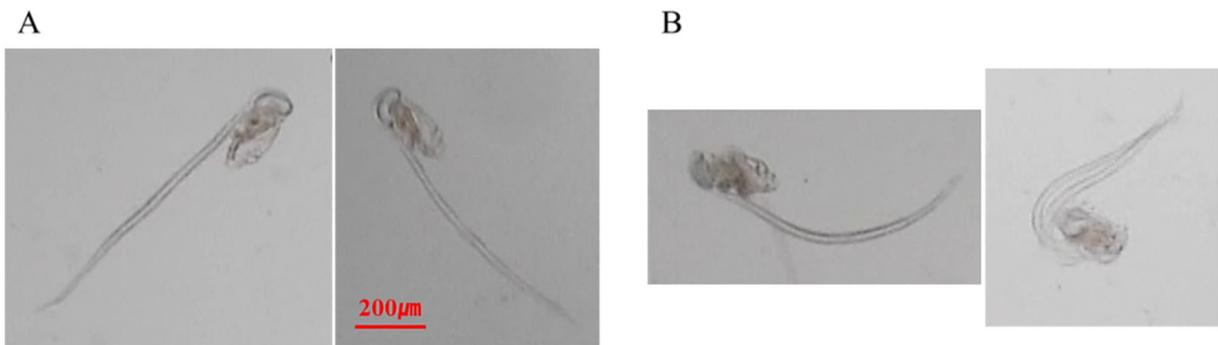


Fig. 8. *Oikopleura* sp. images according to the shape of the notocord (A: straight posture, B: bent posture).

기를 통틀어 연체동물의 적중률이 각각 67.3%, 84.0%로 가장 낮은 것으로 나왔다(Table 5, 6 참조) 연체동물은 폐각 또는 조가비를 보유하고 있어 전형적인 형태는 일정하나, 수중에서 유영하는 개체는 그 과정에서 다양한 단면을 보일 수 있다. 이러한 특징들로 인해 순간의 단편적인 이미지만으로는 해당 개체의 분류학적 특징을 발견하기 어려울 수 있다(Benfield *et al.*[2007]). 연체동물의 경우 수중에서 유영하는 과정에서 다리 및 섬모 등의 움직임으로 인해 제시된 이미지에 따라 자포동물, 환형동물 등과 형태적인 유사점을 보일 수도 있으며, 이는 시스템이 대상을 인식함에 어려움으로 작용

할 것으로 판단된다(Fig. 10 참조).

다모류(Polychaeta) 유생이 포함된 환형동물은 1, 2차 시기에 각각 82.1%, 86.0%의 적중률을 보여 상대적으로 차이가 적었다. 2차 시기에 12,654장의 이미지를 학습하여 1차 시기에 비해 6배가 넘는 학습량을 보였으나, 실제 차이는 크지 않았다. 환형동물은 개체들의 모양이 단순하여 적중률이 높을 것으로 예상되었으나, 상대적으로 작은 개체의 경우 하드웨어적인 한계로 이미지 상으로 형태적 구분이 확실하지 않았다(Rodenacker *et al.*[2006]; Benfield *et al.*[2007]). 또한 개체의 유영과정에서 이미지의 방향과 위치가 임

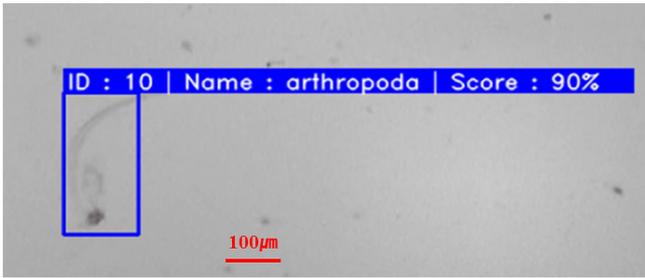


Fig. 9. Image of Chordata with small body width.

의적인 모습을 보이고, 특성상 형태적 변형이 용이하여(Tang *et al.*[1998]) 연체동물 등으로 오판하는 비율이 다소 나타난 것으로 판단된다(Fig. 10 참조).

크기가 상대적으로 작은 섬모충류, 유행동물의 경우 체폭 100 µm 내외의 크기를 보유하고 있으며, 원뿔 모양으로 형태적인 유사성이 많아 판독에서 상호 오류를 일으킬 것으로 예상되었다(Culverhouse *et al.*[2003]; Schulze *et al.*[2013]). 이러한 점은 1차 시기 섬모충류 판독에서 유행동물로 오판한 비율이 20%, 유행동물 판독에서 섬모충류로 오판한 비율이 18%로 확인되어 시스템상으로 두 분류군의 구분이 확실하지 않음을 나타내는 것으로 이는 기계 학습에 대한 정보량이 늘어난 2차 시기에도 수치적으로 차이는 있었으나, 각각의 분류군을 판별함에 있어 상호 유사성이 가장 높게 나타난 것을 확인할 수 있었다(Table 5, 6 참조).

절지동물은 전체적으로 몸체가 길고, 대부분 외골격으로 둘러 쌓여 있어 형태적 변동성은 적은 편이나, 크기 및 유영하는 형상(Benfield *et al.*[2007])에 따라 일부 개체가 환형동물 등으로 판독한 것으로 보이며(Table 5, 6 참조), 이미지의 해상도에 따라 섬모충류, 유행동물 등으로 오류 판정을 받는 것으로 보인다. 특히 본 연구에서는 부속지가 상대적으로 짧은 Harpacticoida, *Oithona* sp. 등이 환형동물 등으로 오판하는 경우가 다수 확인되었다(Fig. 11 참조).

다른 분류군과 비교하여 상대적으로 출현 빈도가 낮은 분류군은 확보할 수 있는 정보량(image)이 적으며, 이에 따라 기계학습에 의한 학습량도 적을 수 밖에 없다. 그러나 적중률 값은 분류군의 특성에 따라 서로 다른 결과를 보였다. 1차 시기의 Others 항목은 40.4%로 이시기 가장 낮은 적중률을 보였다. 추형동물 유생(Actinotroch), 치어(Fish larva) 등을 포함하고 있으며, 서로간에 유사성이 부족하다. 특히 추형동물 유생은 다수의 촉수를 보유하여 형태적으로 매우 독특한 형상을 보이고, 환형동물과 마찬가지로 유영과정에서 수축과 팽창을 반복하는 특성을 보여 시스템 상으로 다른 분류군과 구분하기 어려울 것으로 판단된다(Table 5 참조). 2차 시기에는 출현빈도가 낮은 분류군을 Echinodermata, Nematoda, Platyhelminthes 등으로 세분화하여 시스템에 대한 적중률을 확인하였다. 각각의 학습된 이미지는 124, 205, 131장으로 다른 분류군에 비해 적었으나, 전체적인 적중률은 99~100%로 매우 높은 값을 보였다. 상대적으로 움직임이 둔하거나 형태적인 변형성이 적으며

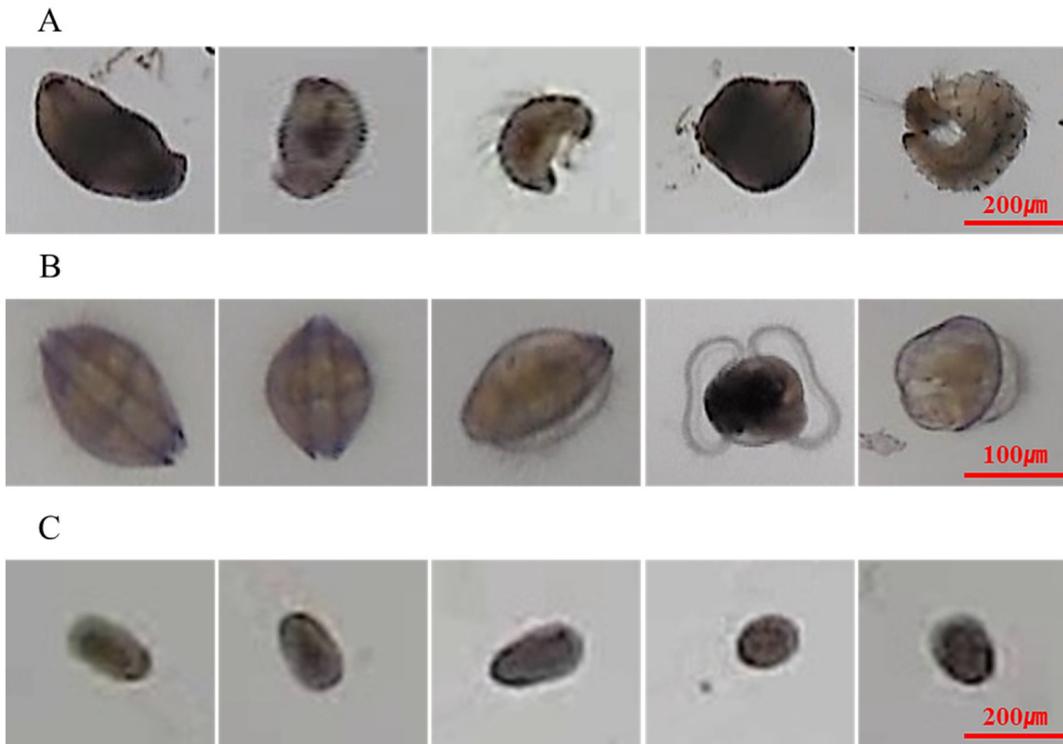


Fig. 10. Image showing morphological similarity according to angle direction (A: Annelida, B: Mollusca, C: Cnidaria).

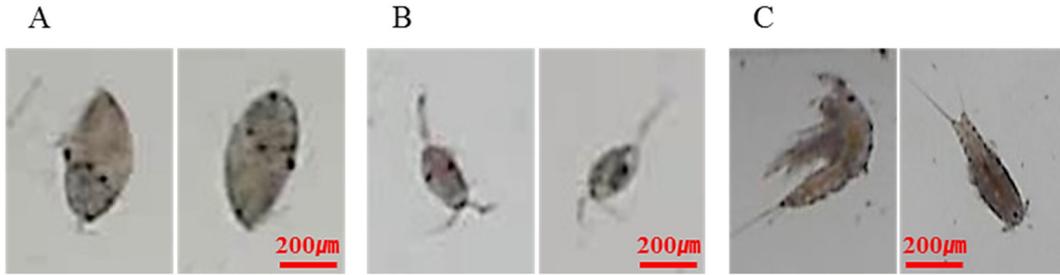


Fig. 11. Object images misinterpreted as Annelida (A: Arthropoda/Cyprid), B: Arthropoda/Oithona sp.), C: Arthropoda/Harpacticoida).

(Echinodermata), 분류군의 특징이 확실한 개체(Nematoda, Platyhelminthes)는 상대적으로 적은 학습량에도 높은 적응률을 보이는 것으로 확인되었다(Table 6 참조).

3.4 시스템의 총족 요건

본 연구에서는 일정한 광원과 관찰 구역(zooplankton counting chamber)이 한정적인 최적화된 실험실적인 환경으로 차후 실제로 현장에 적용하는 과정에서 여러 문제점들이 발생할 수 있다. 따라서 실제 현장에서 플랑크톤을 분류하기 위해서는 다음 사항들에 대해 우선적으로 고려되어야 한다. (1) 플랑크톤은 수 많은 분류군이 존재하며, 생애 주기에 따라 형태적 특성이 변동적이고(Tang *et al.*[1998]), 같은 분류군이라 하더라도 각각의 형태적 특징이 상이할 정도로 매우 다양하다(Sieracki *et al.*[2010]; Schulze *et al.*[2013]). 또한 지역과 시기에 따라 생물의 출현 특성이 다르게 나타나며(Matuszewski[2014]), 출현 빈도가 낮은 분류군의 경우 해당 생물의 이미지 확보가 매우 제한적이다. 따라서 자동 분류 시스템은 플랑크톤의 이러한 넓은 대역을 판별할 수 있는 환경(분류군의 특성에 따른 정보량 확보) 및 성능이 확보되어야 한다. (2) 선박평형수 처리 장비를 검증함에 있어 전통적인 방법은 실험실 수준의 장비를 요구하고, 시간이 비교적 오래 걸리는 관계로 현장에서 직접 대응하기 힘들었다. 이러한 점은 선박의 운항에 차질을 발생시킬 수 있는 부분이다. 따라서 자동 분류 장치는 대량의 자료를 신속하게 처리할 수 있어야 하며, 연속적이고, 적용이 용이한 특성을 보유해야 한다(Matuszewski[2014]). 따라서 이미지에 대한 부분 분류 작업(Flowcam 등), 채집된 생물의 처리과정(Zooscan - formalin 고정) 등을 포함한 일련의 과정이 동반(Zoimage 등)되는 부분 자동화를 구현하고 있는 기존의 시스템들은 이용자 층(선원, 항만 종사자 등)을 고려했을 때 선박 및 항만 설비와 같은 환경에서 적용하기 어려운 부분이다. (3) 또한 해수에는 형태 및 크기가 다양한 입자들이 존재한다. 이러한 입자들은 초점을 흐리게 하여 노이즈를 발생 시키며, 플랑크톤을 판별하기 어렵게 만들 수 있다(Bell and Hopcroft[2008]; Bi *et al.*[2015]). 따라서 현재 이용되는 Net 방식의 선박평형수 샘플링 방법은 생물학적, 무생물학적(무기물 / 사체) 입자들이 매우 높은 밀도를 보이는 경우가 많아 차후 샘플링에 대한 방법론적 고민이 필요하다.

4. 요약 및 결론

영상 처리와 기계 학습을 접목하여 체폭 50 µm 이상의 생물을 분류함에 있어 본 연구에서는 상대적으로 크고, 모양이 단순하며, 형태적 변형이 적은 경우(자포동물: Chidaria, 모약동물: Chaetognatha, 극피동물: Echinodermata, 선형동물: Nematoda 등) 학습된 정보량에 비해 적응률이 우수한 것으로 나타났다. 반면에 수중에서 유영하는 방향에 따라 형태적으로 다른 양상을 보이거나(절지동물: Arthropoda, 연체동물: Mollusca), 몸체의 변형이 비교적 자유로운 분류군(척삭동물: Chordata, 환형동물: Annelida), 상대적으로 출현 빈도가 낮아서 해당 분류군에 대한 이미지 획득이 어려운 경우 생물의 특성에 따라 적응률이 떨어지는 것으로 확인되었다. 또한 분류군의 크기가 작아서 형태적 구분이 명확하지 않을 경우(섬모충류: Ciliophora, 윤형동물: Rotifera) 정보량에 비해 해독력이 떨어지는 것으로 나타났다. 학습된 정보량이 증가함에 따라 전체적인 시스템의 적응률이 90% 이상을 보였으나, 수중에서 생물을 분류하는 작업은 매우 어려운 작업으로 이미지를 이용한 분류 과정에서 적응률을 올리고, 오류를 줄이기 위해서는 생물의 다양한 변형성을 반영할 수 많은 이미지의 확보가 필요하고, 이미지의 선명도를 올리기 위한 하드웨어적(lens, camera, light source)인 개선이 뒤따라야 할 것으로 판단된다. 또한 차후 수중의 많은 입자에서 생존 개체를 검출하기 위해 형태적인 접근법과 더불어 수중의 입자간(생물 + 무생물) 이동 경향을 추적하고, 확인하기 위한 알고리즘(Kalman filter 등)에 대한 연구가 필요하다. 추가적으로 형광분석법 등을 고려한 통합적인 검출 시스템 개발을 통해 현장(항구, 선박 등)에서 즉시 결과 확인이 가능한 자동화 작업이 필요할 것으로 본다.

후 기

본 연구는 2018년도 산업통상자원부의 “항만국통제(PSC) 대응을 위한 선박평형수 통합 자가 검사 시스템 개발(No. 10067096)”과제의 지원을 받아 수행되었습니다.

References

- [1] Baek, S.H. and Shin, K.S., 2009, Applicability of fluorescein diacetate (FDA) and Calcein-AM to determine the viability of

- marine plankton, *Ocean Polar Res.*, 31(4), 349-357.
- [2] Baek, S.H. and Shin, K.S., 2015, A staining method to determine marine microplanktonic organism viability and investigate the efficacy of a ships ballast water treatment system, *J. Korea Acad. Industr. Coop. Soc.*, 16(6), 4328-4334.
- [3] Battle, J. (Ed.), 2009, Silent invasion: The spread of marine invasive species via ships' ballast eater, WWF for a Living Planet.
- [4] Bell, J.L. and Hopcroft, R.R., 2008, Assessment of Zoolmage as a tool for the classification of zooplankton. *J. Plankton Res.*, 30(12), 1351-1367.
- [5] Benfield, M.C., Grosjean, P., Culverhouse, P.F., Irigoien, X., Sieracki, M.E., Lopez-Urrutia, A., Dam, H.G., Hu, Q., Davis, C.S., Hansen, A., Pilsikaln, C.H., Riseman, E.M., Schultz, H., Utgoff, P.E. and Gorsky, G., 2007, RAPID: research on automated plankton identification. *Oceanography*, 20(2), 172-187.
- [6] Bi, H., Cook, S., Yu, H., Benfield, M.C. and Houde, E.D., 2012, Deployment of an imaging system to investigate fine-scale spatial distribution of early life stages of the ctenophore *Mnemiopsis leidyi* in Chesapeake Bay. *J. Plankton Res.*, 35(2), 270-280.
- [7] Bi, H., Guo, Z., Benfield, M.C., Fan, C., Ford, M., Shahrestani, S., Sieracki, J.M., 2015, A semi-automated image analysis procedure for in situ plankton imaging systems. *PloS one*, 10(5), e0127121.
- [8] Bickel, S.L., Tang, K.W. and Grossart, H.P., 2009, Use of aniline blue to distinguish live and dead crustacean zooplankton composition in freshwaters, *Freshwater boil.*, 54(5), 971-981.
- [9] Culverhouse, P.F., Williams, R., Reguera, B., Herry, V. and González-Gil, S., 2003, Do experts make mistakes? A comparison of human and machine identification of dinoflagellates, *Mar. Ecol. Prog. Ser.*, 247, 17-25.
- [10] Drake, L.A., Ruiz, G.M., Galil, B.S., Mullady, T.L., Friedmann, D.O. and Dobbs, F.C., 2002, Microbial ecology of ballast water during a transoceanic voyage and the effects of open-ocean exchange, *Mar. Ecol. Prog. Ser.*, 233, 13-20.
- [11] Elliott, D.T. and Tang, K.W., 2009, Simple staining method for differentiating live and dead marine zooplankton in field samples, *Limnol. Oceanogr. Methods*, 7(8), 585-594.
- [12] Gollasch, S., Lenz, J., Dammer, M., Andres, H.G., 2000, Survival of tropical ballast water organisms during a cruise from the Indian Ocean to the North Sea, *J. Plankton. Res.*, 22(5), 923-937.
- [13] Gollasch, S. and Leppäkoski, E., 2007, Risk assessment and management scenarios for ballast water mediated species introductions into the Baltic Sea, *Aquat. Invasions*, 2(4), 313-340.
- [14] Gorsky, G., Ohman, M.D., Picheral, M., Gasparini, S., Stemmann, L., Romagnan, J.B., Cawood, A., Pesant, S., Garcia-Comas, C., Prejger, F., 2010, Digital zooplankton image analysis using the ZooScan integrated system, *J. Plankton Res.*, 32(3), 285-303.
- [15] Hu, Q. and Davis, C., 2005, Automatic plankton image recognition with co-occurrence matrices and support vector machine, *Mar. Ecol. Prog. Ser.*, 295, 21-31.
- [16] Hyun, B., Shin, K., Chung, H., Choi, S.Y., Jang, M.C., Lee, W.J. and Choi, K.H., 2014, Application of Neutral Red Staining Method to Distinguishing Live and Dead Marine Plankton for the Investigation of Efficacy of Ships Ballast Water Treatment System, *The Sea*, 19(4), 223-231.
- [17] International Maritime Organization, 2004, International Convention for the Control and Management of Ships' Ballast Water and Sediments, 2004.
- [18] Leow, L.K., Chew, L.L., Chong, V.C., Dhillon, S.K., 2015, Automated identification of copepods using digital image processing and artificial neural network, *BMC bioinformatics*, 16(18), S4.
- [19] Matuszewski, D.J., 2014, Computer vision for continuous plankton monitoring, Doctoral dissertation, PhD Thesis, University of Sao Paulo, Sao Paulo.
- [20] Rodenacker, K., Hense, B., Jütting, U., and Gais, P., 2006, Automatic analysis of aqueous specimens for phytoplankton structure recognition and population estimation, *Microsc. Res. Tech.*, 69(9), 708-720.
- [21] Schulze, K., Tillich, U.M., Dandekar, T. and Frohme, M., 2013, PlanctoVision-an automated analysis system for the identification of phytoplankton, *BMC bioinformatics*, 14(1), 115.
- [22] Sieracki, M.E., Benfield, M., Hanson, A., Davis, C., Pilsikaln, C.H., Checkley, D., Sosik, H.M., Ashjian, C., Culverhouse, P., Cowen, R., Lopes, R., Balch, W., Irigoien, X., 2010, Optical plankton imaging and analysis systems for ocean observation, *Proceedings of ocean Obs*, 9, 21-25.
- [23] Tang, X., Stewart, W. K., Huang, H., Gallager, S. M., Davis, C. S., Vincent, L. and Marra, M., 1998, Automatic plankton image recognition, *Artificial intelligence review*, 12(1-3), 177-199.
- [24] Wright, D.A., Gensemer, R.W., Mitchelmore, C.L., Stubblefield, W.A., van Genderen, E., Dawson, R., Orano-Dawson, C.E., Berr, J.S., Mueller, R.A. and Cooper, W.J., 2010, Shipboard trials of an ozone-based ballast water treatment system, *Mar. Pollut. Bull.*, 60(9), 1571-1583.

Received 6 November 2018

Revised 16 March 2019

Accepted 17 April 2019