논문 2021-16-23

# 3차원 합성곱 신경망 기반 향상된 스테레오 매칭 알고리즘
# (Enhanced Stereo Matching Algorithm based on 3-Dimensional Convolutional Neural Network)

왕 지 엔, 노 재 규*
(Jian Wang, Jackyou Noh)

Abstract : For stereo matching based on deep learning, the design of network structure is crucial to the calculation of matching cost, and the time-consuming problem of convolutional neural network in image processing also needs to be solved urgently. In this paper, a method of stereo matching using sparse loss volume in parallax dimension is proposed. A sparse 3D loss volume is constructed by using a wide step length translation of the right view feature map, which reduces the video memory and computing resources required by the 3D convolution module by several times. In order to improve the accuracy of the algorithm, the nonlinear up-sampling of the matching loss in the parallax dimension is carried out by using the method of multi-category output, and the training model is combined with two kinds of loss functions. Compared with the benchmark algorithm, the proposed algorithm not only improves the accuracy but also shortens the running time by about 30%.

Keywords : Stereo matching, 3D Convolutional Neural Network, Parallax dimension, Computation cost, Network structure

## Ⅰ. 서 론

스테레오 매칭은 컴퓨터 비전의 기초적인 연구로 3차원 재구성, 자율주행 및 로봇 내비게이션 등 다양한 분야에 활용된다. 전통적인 스테레오 매칭 알고리즘은 비용 계산과 시차 최적화에 관한 연구를 많이 진행하는데, 한편으로는 좋은 측정 함수를 설계하여 매칭 비용을 계산하고 다른 한편으로는 각 픽셀에 로컬 또는 글로벌을 사용하는 방법으로 시차 값을 배치한다 [1, 2]. 이러한 알고리즘은 모두 인공적으로 설계된 함수로 Pathological 영역 (예 : 무늬가 적은 영역)에 대해 정확한 결과를 얻지 못하는 경우가 많다.

딥러닝은 이러한 컴퓨터 비전 분야에서 강력한 그래픽 이해능력을 보여주고 있다. 특히 사전 정의된 범주에 가장 적합한 추정을 하는 객체 분류 (Classification), 범주 분류된 이미지에 해당하는 객체를 전체 이미지에서 탐색하는 객체 탐색 (Object detection) 및 픽셀 단위의 범주 분류 객체 이미지 탐색을 위한 의미론적 분할 (Semantic segmentation) 등에서 우수한 성능을 가지고 있으며, 딥러닝 기반의 스테레오 매칭 알고리즘에 대한 관심이 높아지고 있다 [3, 4]. 합성곱 신경망 (CNN, Convolutional Neural Network) 모델은 이미지 속에서 안정적인 특징을 추출할 수 있어 이미지 블록 사이의 유사성을 학습하기에 적합하다 [5, 6]. 매칭성이 모호한 영역에서 모델의 전반적인 최적화 능력을 더욱 높이기 위해 데이터 입력에서 추정 출력까지 신경망으로만 모든 모듈의 기능을 담당하여 추정을 수행하는 End-to-End 스테레오 매칭 방법은 시차 예측의 전체 과정을 CNN 모델에 통합하여 적용하기도 한다 [7, 8]. 그런데 이 알고리즘은 시차를 따르는 1차원적 알고리즘을 많이 적용해 시차 차원적 특징이 사라지게 된다. 스테레오 매칭에 3차원 합성곱 신경망 (3D CNN)을 도입해 3차원에서 전체적인 semantic 정보를 이해할 수 있도록 해 장면의 더 잘 이해할 수 있도록 했다 [9, 10].

스테레오 매칭 알고리즘에 3차원 합성곱 신경망을 도입할 경우에 매칭 과정에서 모델링 효과가 좋아지지만 매칭 비용과 계산량은 수십 배 증가하게 된다. 이러한 자원의 부담은 주로 3차원 합성곱 신경망에서 발생하게 된다. 이러한 자원 부담을 줄이기 위해 본 논문에서는 다음의 3가지 사항이 포함된 향상된 3차원 합성곱 신경망 스테레오 매칭 알고리즘을 제안하고자 한다.

1) 시차 차원에서의 희박 합성곱 모듈을 3차원 합성곱 신경망의 입력으로 구축하여 비디오 메모리 비용 및 계산량을 감소시킨다.

2) 단일 평행변환 길이에 3차원 합성곱 신경망의 출력 카테고리 수를 예측하고, 시차 차원에서 매칭하는 비용에 대해 출력 카테고리 수에 맞춰 세분화하여 샘플링한다.

3) 최대 확률 주변에서의 시차 회귀와 부분 픽셀의 교차 엔트로피 비용 (Loss)과 매끄러운 (smooth) L1-norm (평균절대오차) 비용 (Loss)을 결합하여 학습함으로써 모형이 시차 이미지를 더 정확하게 추정할 수 있게 하고 시차 범위를 확

*Corresponding Author (snucurl@kunsan.ac.kr)
Received: Aug. 20, 2021, Revised: Sep. 15, 2021, Accepted: Sep. 23, 2021
J. Wang: Kunsan National University (Ph.D. Course Student)
J.K. Noh: Kunsan National University (Prof.)
※ 본 논문은 2021년도 정부 (산업통상자원부)의 재원으로 한국에너지기술평가원의 지원을 받아 수행된 연구임 (20213030020120, 해상풍력발전 블레이드의 전주기 신뢰성 향상을 위한 생산품질 및 유지관리 기술 개발).

ⓒ IEMEK J. Embed. Sys. Appl. 2021 Oct. 16(5) 179-186
ISSN : 1975-5066
http://dx.doi.org/10.14372/IEMEK.2021.16.5.179

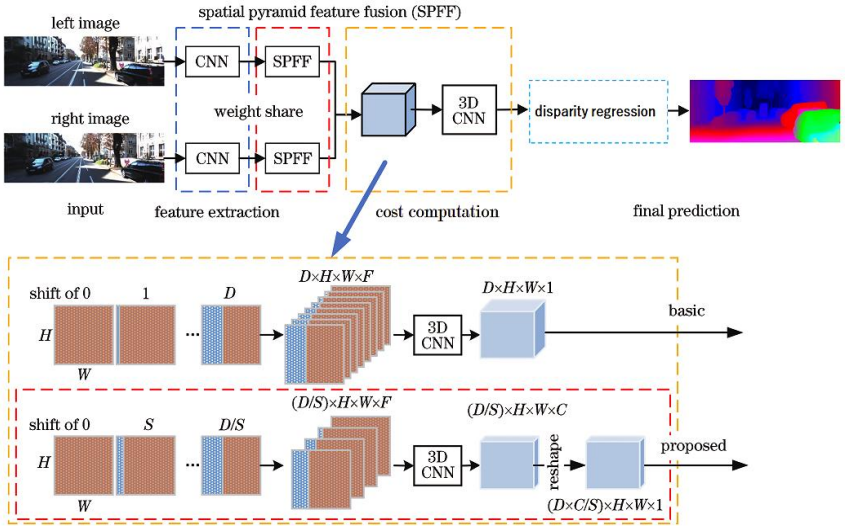


그림 1. 향상된 3차원 합성곱 신경망 스테레오 매칭 알고리즘 네트워크 구조 [5]
Fig. 1. The network structure of the proposed 3D CNN stereo matching algorithm [5]

장할 때 정밀도에 미치는 영향이 크지 않도록 한다.

## Ⅱ. 관련 연구

전통적인 스테레오 매칭 알고리즘은 일반적으로 매칭 비용 계산, 비용 집합, 초기 시차 계산, 시차 정제의 4단계를 포함한다. 이러한 스테레오 매칭 알고리즘에서의 합성곱 신경망은 블록 간 유사성을 계산하는 데 사용되는데 Zbontar 등은 Siamese 네트워크를 학습시켜 안정적인 이미지 특징을 추출하고 매칭 비용을 계산함으로써 전통적인 방법에 비해 성능이 크게 향상됨을 보여주었다 [6]. Luo 등은 블록 기반의 매칭 네트워크를 제시해 1s 미만으로 매우 정확한 시차를 계산할 수 있음을 보여주었다 [11].

End-to-End 학습은 알고리즘의 전체 최적화에서 더 나은 성능을 얻을 수 있는데, Mayer 등은 '인코딩-디코딩' 네트워크 구조를 적용하고 대규모 합성 데이터 집합을 만들어 시차의 End-to-End 학습을 수행하였다 [7]. 이 시차예측 네트워크를 바탕으로 Pang 등은 캐스케이드 네트워크를 통해 시차 정밀도를 높이고 있다 [8].

Liang 등은 합성곱 신경망과 베이지안 추론을 결합해 사전확률과 사후확률의 변함없는 특징을 학습해 시차예측과 조정을 하였다 [12]. Jie 등은 순환 신경망을 도입해 좌우 비전 이미지를 꾸준히 대조함으로써 시차 예측 결과를 개선하였다 [13].

End-to-End 학습법에 대해서는 Kendall 등이 3차원 합성곱 신경망 기반 End-to-End 학습 시차 예측을 제안하였다 [9]. 이 방법은 이미지의 기하학적 특성을 이용하여 깊이 있는 특징을 추출하고 이 구조에 합성곱 모듈을 적용하였다. 3차원 합성곱 신경망을 적용하여 시차 추정을 개선하는데 적용하였다. 그리고 부분 픽셀 정밀도의 시차 학습을 End-to-End로 구현할 수 있어 후처리나 정규화가 필요 없는 장점을 가지고 있다.

Yu 등은 명확한 비용 합계 submodule을 도입해 매칭 비용에 최적화하였다 [14]. Smolyanskiy 등은 좌우 비전 이미지의 기하학적 관계에 따라 semi-supervised 비용 함수를 구축하고, 희박 시차값으로 밀집한 오차 피드백 신호를 제공하였다 [15]. Chang 등은 평균 풀링 모듈을 적용해 특징 융합으로 특징추상력을 높이고 deeply-supervised 방식으로 매칭 비용 계산을 학습하였다 [10].

Zhang [16] 등은 Semi-global 집계 계층과 로컬 가이드 집계 계층의 Guided Aggregation Net (GA-Net)을 제안하면서 전통적인 방법과 현대적인 방법을 결합해 계산 효율을 높였다.

Huang [17] 등은 네트워크 구조가 복잡하고 소모성이 높은 문제에 대해 특징 추출 모듈을 간소화하고, 여분 층을 삭제하고 콘볼루션 커널 (convolution kernel)을 줄였다. 3D 콘볼루션 중 시차 차원을 줄이고, 3D 콘볼루션 출력에 다중 시차를 예측해 정밀도를 확보하면서 효율을 높였다.

기존 연구에서 주로 지도학습 (Supervised Learning) 스테레오 매칭 알고리즘이지만, 비지도학습 (Unsupervised Learning) 스테레오 매칭 알고리즘도 매우 주목할 만한 내용인데 이러한 알고리즘은 학습 과정에서 광범위하고 높은 품질 시차 데이터가 필요하지 않고 좌우 이미지의 기하학적 구속 관계에 따라 어떻게 시차를 예측하는지를 학습할 수 있어 학습 데이터 수집에 필요한 작업량을 크게 줄일 수 있다

## Ⅲ. 3차원 합성곱 신경망 기반 스테레오 매칭 알고리즘

스테레오로 매칭하는 것은 좌우 비전 이미지에서 대응점을 찾는 것이다. 출력이 밀집된 시차도를 출력하는 과정은 이를 End-to-End 도달한 값으로 되돌리는 작업으로 전환되며, 쌍안 이미지를 입력으로 예측한 시차도를 직접 출력한다. 기존의 3차원 합성곱 신경망 스테레오 매칭 알고리즘의 네트워크 구조는 그림 1의 상단과 같이 특징 추출 (Feature extraction), 공간 피라미드 특징 융합 (Spatial Pyramid Feature Fusion), 매칭 비용 계산 (Cost computation), 시차 회귀 (Disparity Regression)의 네 부분으로 이루어져 있다 [5]. 이러한 기존의 3차원 합성곱 신경망 스테레오 매칭의 프로세스는 다음과 같다.

1) 합성곱 신경망을 이용한 좌우 비전 이미지의 특징을 추출하고 공간 피라미드를 이용하여 특징을 융합한다.

2) 좌측 시각 특징과 평행변환 (Translation)하는 우측 시각 특징을 연결하여 시차 차원에서의 희박 합성곱 모듈을 구축하고, 3차원 합성곱 신경망으로 학습하여 매칭비용을 계산한다.

3) 원본 이미지 사이즈로 샘플링 합성곱 모듈을 사용하여 비용값을 계산한 다음 Softmax 함수로 시차 확률 분포로 바꾸고, 시차 회귀 함수를 통해 부분 픽셀의 예측 시차를 출력한다.

### 1. 네트워크 구조의 개선

왼쪽 시각의 특징과 평행변환하는 오른쪽 연결시각 특징으로 합성곱 모듈을 구축하는 과정은 그림 1의 하단에 나타내었다. 직사각형은 좌우 시각적 특징을 표현하고 겹침은 특징을 연결한다는 것을 의미한다. 일반적으로 우측 시각 특징 이미지를 평행변환시키는 거리 (S)를 1로 하고 최대 시차 (D) 범위 내에서 합성곱 모듈을 구성하면 그 데이터량이 특징 이미지의 D배로 변하며 단일 3차원 합성곱 층의 계산량은 단일 2차원 합성곱 층의 3D배 (합성곱 코어 크기는 3)가 된다.

이미지 특징에 대해 시차 차원 측면에서 작은 범위의 평행변환과 겹침을 진행하면 대량의 정보가 계속해서 남게 된다. 넓은 평행변환 거리 (S>1)를 적용하여 시차 차원의 희박 합성곱 모듈에 매칭하는 비용을 계산하면 평행변환이 1인 경우에 비해 3차원 합성곱 모듈의 비디오 메모리 비용과 계산량을 약 1/S로 낮출 수 있다. S = 1일 때와 비교하여 3차원 합성곱 모듈은 시차 차원에서는 원래의 1/S이 되고, 각 3차원 합성곱 층의 특징은 그대로이며 입출력도 원래의 1/S이기 때문에 필요한 비디오 메모리과 계산 자원은 모두 원래의 1/S이 된다.

평행변환 거리의 증가는 모형의 시차 차원에서의 세분화 능력을 약하게 하는데 이러한 영향을 감소시키기 위하여 각각의 매칭 비용에 대해 출력 카테고리 수 (C)의 변화를 통해 추정하였다. 매칭 비용을 비선형적으로 샘플링해 시차 확률 분포의 세분화 함수를 학습시켜 알고리즘 정밀도를 개선할 수 있는데 비용 출력층에만 가중치를 증가시키므로 알고리즘 효율에 미치는 영향은 적다.

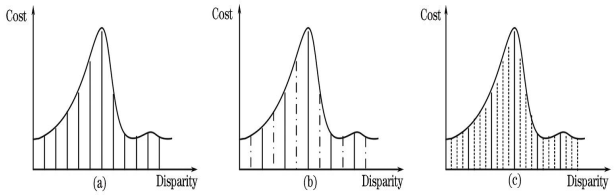


그림 2. 시차 차원 대응 샘플링 비용

Fig. 2. Visual description of the cost sampled in the parallax dimension(2(a) S=1, C=1; 2(b) S=2, C=1; 2(c)S=2, C=4)

S값과 C값을 변화시킴으로써 시차 차원에 매칭하는 비용에 해당하는 샘플링 수가 결정되는데 이를 그림 2에 나타내었다. 그림 2 (a)는 S=1, C=1인 경우로 각각의 시차값에 샘플링되는 비용이 대응하고 있으며 기준이 된다. 그림 2 (b)는 S=2, C=1인 경우로서 시차 축에서 넓은 간격으로 평행이동하는 경우로 점선 부분은 넓은 간격 평행이동에 따른 표현하지 못하는 샘플링을 나타내고 있다. 그림 2 (c)는 S=2, C=4인 경우로서 시차 축에서 넓은 간격으로 평행이동하되 여러 종류로 예측하는 경우이므로 많은 샘플링 매칭 비용 예측을 하고 있음을 알 수 있으며 비선형 샘플링이 이루어질 수 있음을 보여주고 있다.

설정된 S값이 클수록 효율은 높지만 정확도는 떨어진다. 그래서 정확도와 효율을 고려하여 적절한 S값을 선택해야 한다. 설정된 C값이 클수록 비선형상 시료채취 인자가 커지며 모델의 세분화 능력이 뛰어나지만 모델 학습의 난이도가 높아지기 때문에 모델의 수렴정확도와 속도에 영향을 준다.

### 2. 비용 함수의 개선

시차 회귀 모듈 예측 시차값과 매끄러운 L1-norm을 결합하여 모델링을 학습하였다 [18]. 시차 회귀 모듈은 Softmax 함수σ( • )를 사용하며, 매칭 비용의 C에 따라 시차 확률 분포를 계산하고, 시차 값을 가중 방식으로 부분 픽셀 추정을 하며, 그 표현식은 다음 식 (1)과 같다.

$$\widehat{d_A} = \sum_{n=0}^{N_d} d_n \sigma(-C_n). \tag{1}$$

$\hat{d}_A$은 시차값의 부분 픽셀 추정이며, dn은 가중치로 다음 식 (2)와 같다.

$$d_n = D_{\max} n / N_d. \tag{2}$$

$D_{max}$는 최대 시차이며 $N_d$는 시차 차원의 샘플링 수이고 n은 시차 차원의 인덱스를 의미한다.

시차 확률 분포가 단봉 (unimodal)이고 대칭일 때, 식 (1)은 비교적 좋은 부분 픽셀을 추정할 수 있다. 시차 확률 분포가 다봉 (multimodal)일 경우에는 부분 픽셀 추정값이 정확하지 못할 수도 있다. Kendall 등은 합성곱 신경망 학습이 출력값을 전처리하고 분포를 단봉성을 가지는 것으로 가정하였다 [9]. 그런데 이 전처리는 학습 단계에서 사용되지만

시험 단계에서도 모델 파라미터에 대한 재학습이나 조정이 필요할 수 있다. 이 문제를 해결하기 위하여 식 (3)과 같이 최대 확률의 시차값 인접 영역에서 가중 평균하였다.

$$\widehat{d_M}=\sum_{|d_n-\widehat{d_m}|\le\delta} d_n\sigma(-C_n). \tag{3}$$

$\hat{d}_M$ 은 시차 예측값이며 $d_m$은 다음 식 (4)와 같다.

$$d_m = D_{\max} m/N_d, \tag{4}$$

$$m=\underset{0\le n\le N_d}{\arg max}(-C_n). \tag{5}$$

본 연구에서는 식 (3)의 $\delta$=2로 설정하였다. 더 나은 시차 추정을 위해 부분 픽셀의 크로스 엔트로피 (Cross Entropy) 비용 $L^{CE}$와 매끄러운 $L_1$-norm 비용 $L^S$의 학습 모델에서의 비용함수는 다음 식 (6)과 같다.

$$L=L^{CE}+wL^S. \tag{6}$$

여기서 w는 $L^S$의 가중치로, 두 가지 비용함수의 중요성에 대한 균형에 사용되며, 본 연구에서는 0.1을 사용하였다. $L^{CE}$와 $L^S$의 구체적인 형태는 다음 식 (7) 및 (8)과 같다.

$$L^{CE}=\frac{1}{N}\sum_{i=0}^{N}\sum_{n=0}^{N_d}-Q(d^{gt},d_n)\ln[\sigma(-C_n)], \tag{7}$$

$$L^S=\frac{1}{N}\sum_{i=0}^{N} f_{SL}(|d^{gt}-d_A|). \tag{8}$$

여기서 N은 시차 태그 값을 갖는 픽셀 수이며, i는 픽셀 인덱스이다. Q함수는 다음 식 (9)와 같이 정의되며 목표 확률 분포를 나타낸다.

$$Q(d^{gt},d_n)=\exp(-|d_n-d^{gt}|/b). \tag{9}$$

이는 시차 태그 값 $d^{gt}$를 중심으로 divergence가 b인 라플라스 분포로서 본 연구에서는 b=2를 사용하였다. $f_{SL}$은 식 (10)과 같은 함수이다.

$$f_{SL}(x)=\begin{cases}0.5x^2, & |x|<1\\ |x|-0.5, & otherwise\end{cases}. \tag{10}$$

## Ⅳ. 실 험

본 연구에서 제안하고 있는 개선된 네트워크 구조 및 비용함수를 사용한 3차원 합성곱 신경망을 적용한 스테레오 매칭 알고리즘의 성능을 평가하기 위하여 실험을 수행하였다. SceneFlow 데이터 집합, KITTI2015 데이터 집합, KITTI2012 데이터 집합에서 평가지표 $E_{EP}$와 $E_{D1}$을 사용하여 알고리즘을 평가한다. $E_{EP}$ (EP:end-point)는 예측 시차와 참값의 차이의 절대값이며, $E_{D1}$ (D1은 디지털 시스템 디스플레이 형식 표준)은 각 그룹 이미지의 평가 영역에 대한 오류 픽셀의 백분율을 나타내며, $E_{EP}$가 3픽셀보다 작거나 $E_{EP}$가 참값의 5%보다 작을 경우에는 정확한 픽셀로 간주하고, 그렇지 않으면 오류 픽셀로 간주한다.

### 1. 실험사항

개선된 알고리즘은 PyTorch를 사용하여 학습과 시험을 하였으며 Nvidia 2080ti 그래픽카드를 사용하였다. 배치경사하강법 (small batch random gradient descent)을 사용하여 학습하며, 한 번에 반복 샘플 크기는 2, 4 ,8을 사용했으며 경사하강법으로 업데이트한 후 반복 횟수는 8, 4, 2를 적용하였으며 업데이트의 샘플 크기를 16으로 확대하였다. 모델의 학습은 모두 Adam optimizer를 사용하고 지연율 파라미터 (0.9, 0.999)를 적용하였으며 전처리된 이미지 크기는 256픽셀*512픽셀이며 최대 시차는 192픽셀로 하였다 [19]. 사용할 데이터 집합은 다음과 같다.

1) SceneFlow 데이터 집합: 학습 이미지 (SF-train)와 시험 이미지 (SF-test)모두 사용하였으며 이미지 픽셀 크기는 540픽셀*960픽셀로 세밀한 시차 그래프의 참값 (ground truth)을 제공한다. 실험에서 계산 비용과 평가 지표를 계산할 때 시차가 192픽셀보다 큰 이미지는 제외했다.

2) KITTI2015 데이터 집합 및 KITTI2012 데이터 집합: 모두 다른 날씨 조건에서 실제 블록의 장면을 기록하는 데이터 집합이며, KITTI2015 데이터 집합은 200쌍의 학습 이미지 (K15-train)과 200쌍의 시험 이미지 (K15-test)를 가지고 있으며, KITTI2012 데이터 집합은 194쌍의 학습 이미지 (K12-train)와 195쌍의 시험 이미지 (K12-train)를 포함하고 있다. 이미지 픽셀 크기는 375픽셀*1242픽셀로 희박한 레이저 데이터를 시차 이미지의 참값으로 사용한다. 설정에 따른 결과 분석을 위해 학습 이미지의 앞 160쌍을 학습용 집합 (K-train), 나머지는 벨리데이션 집합 (K-val)으로 활용하였다. 모델의 일반화 능력을 높이기 위해 학습 데이터의 색상과 공간 변환을 강화하였다. 색상 증강은 색조 증가, 대비율 증가, 밝기 증가, 랜덤 그레이스케일이 포함된다.

하이퍼 파라미터에 대한 분석은 3그룹으로 나누어 앞에 2개의 그룹은 S값과 C값의 역할을 분석하기 위해 SF-train 학습 집합만 사용하고 3번째 그룹은 비용함수 비교 분석을 위해 SF-train과 K-train 학습 집합을 사용하였다.

### 2. S값이 알고리즘 성능에 미치는 영향

S값을 분석함과 동시에 기준 알고리즘과 직접 비교하기 위해 본 실험은 문헌 [14]의 모형설정 (C=1)에 기초하여 S의 범위를 [1, 8]로 설정하였으며 성능 비교 결과는 표 1과 같다. 표 1에서 GPU/GB는 256픽셀*256픽셀의 이미지가 학습 단계에서 차지하는 GPU의 비디오 메모리 용량을 의미하고, $E_{EP}$와 ED1 모두 학습이 완료됐을 때 벨리데이션 집합에서 테스트한 결과이며, $t_{Run}$은 20개의 랜덤 샘플 (각 샘플의 이미지 크기는 540픽셀*960픽셀)에서 테스트한 평균값이다. S값이 커지면서 알고리즘 오차가 커졌고 계산 부담은 낮아진 것으로 보인다. 문헌 [14]에 비해 동일한 모형이 설치된

표 1. S값에 따른 제안 방법의 성능 평가 (C=1)
Table. 1. Performance evaluation of proposed method with different S (C=1)

| Method | S | $E_{EP}$(pixel) | $E_{D1}$(%) | $t_{Run}$(s) | GPU(GB) |
|---|---|---|---|---|---|
| PSMNeT[10] | 1 | 1.09 | - | - | - |
| Proposed | 1 | 1.02 | 3.41 | 0.75 | 2.16 |
| | 2 | 1.12 | 3.89 | 0.47 | 1.51 |
| | 3 | 1.17 | 4.34 | 0.37 | 1.32 |
| | 4 | 1.22 | 4.81 | 0.30 | 1.20 |
| | 5 | 1.17 | 5.36 | 0.25 | 1.11 |
| | 6 | 1.30 | 5.62 | 0.25 | 1.08 |
| | 7 | 1.40 | 6.05 | 0.23 | 1.01 |
| | 8 | 1.42 | 6.23 | 0.22 | 1.01 |

표 2. C값에 따른 제안 방법의 성능 평가
Table. 2. Performance evaluation of proposed method with different C

| S | C | $E_{EP}$(pixel) | $E_{D1}$(%) | $t_{Run}$(s) | GPU(GB) |
|---|---|---|---|---|---|
| 1 | 1 | 1.02 | 3.41 | 0.75 | 2.16 |
| 2 | 1 | 1.12 | 3.89 | 0.47 | 1.51 |
| | 2 | 1.07 | 3.71 | 0.48 | 1.68 |
| | 3 | 1.06 | 3.63 | 0.52 | 1.85 |
| | 4 | 1.08 | 3.81 | 0.53 | 2.02 |
| | 5 | 1.08 | 3.79 | 0.54 | 2.20 |
| | 6 | 1.07 | 3.69 | 0.56 | 2.37 |
| 3 | 1 | 1.17 | 4.34 | 0.37 | 1.32 |
| | 2 | 1.13 | 4.06 | 0.37 | 1.42 |
| | 3 | 1.12 | 3.99 | 0.40 | 1.55 |
| | 4 | 1.14 | 4.06 | 0.41 | 1.66 |
| | 5 | 1.14 | 4.03 | 0.42 | 1.78 |
| | 6 | 1.07 | 3.89 | 0.43 | 1.89 |

경우, 본 실험에서 $E_{EP}$ 오차가 약 6% 감소했는데, 이는 학습 주기의 연장과 학습율의 적절한 조정에 따른 것이다. S=1에 비해 S=2, 3의 경우 EEP가 약 10%, 15% 증가하고 $E_{D1}$은 0.58%, 0.93% 증가한다. $t_{Run}$은 약 37%, 50% 낮아진다. 문헌 [14]에 비해 S=2, 3의 경우 오차가 약 1%, 7% 증가하는 데 그쳐 오차 증가가 적고 계산시간은 현저히 단축되었다.

### 3. C값이 알고리즘 성능에 미치는 영향

S=2, 3일 때 알고리즘 성능은 현저히 떨어지지는 않고, 계산 효율은 현저히 향상되었다. 따라서 S=2, 3으로 설정할 때 C의 범위를 [1, 6]으로 적용하여 실험하였다. 알고리즘의 성능 대비 결과는 표 2와 같다. S값이 고정되어 있을 때 알고리즘의 계산 부담은 C값의 증가에 따라 약간 증가하였음을 알 수 있으며, C값의 범위가 [1, 3]일 때는 C값의 증가에 따라 오차가 줄어들며, C값의 범위가 [4, 6]일 때는 이러한 변화 추세를 유지할 수 없는데 이는 큰 C값으로 인해 모델의 학습이 어려워지기 때문이다. S=1, C=1에 비해 S=2, 3, C=3일 때 $E_{EP}$는 약 4% , 10% 증가했고, $E_{D1}$은 0.22%, 0.58% 증가했다. $t_{Run}$은 약 31%, 47% 낮아졌으며, 문헌 [14]에 비해 S=2, C=3일 때 EEP는 약 4% 낮아지고, $t_{Run}$은 약 31% 낮아졌다. 알고리즘 정밀도와 효율은 모두 향상됐다.

표 3. 다른 설정에서의 제안 방법의 성능 평가
Table. 3. Performance evaluation of proposed method with different settings

| Setting | | | | Max disparity | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | 192 | | | | 384 | |
| S | C | Loss | D | $E_{EP}$ | $E_{D1}$ | $t_{Run}$ | GPU | $E_{EP}$ | $E_{D1}$ |
| 1 | 1 | L1 | Tri | 1.02 | 3.41 | 0.75 | 2.16 | 1.33 | 3.64 |
| 2 | 3 | L1 | Tri | 1.05 | 3.63 | 0.51 | 1.85 | 1.38 | 3.90 |
| 2 | 3 | CE | Bi | 1.04 | 2.71 | 1.45 | 1.50 | 1.29 | 2.92 |
| 2 | 3 | CE+L1 | Bi | 1.04 | 2.69 | 0.45 | 1.56 | 1.28 | 2.87 |
| 3 | 3 | L1 | Tri | 1.11 | 3.99 | 0.39 | 1.55 | 1.49 | 4.30 |
| 3 | 3 | CE | Bi | 1.13 | 2.73 | 0.36 | 1.30 | 1.39 | 3.40 |
| 3 | 3 | CE+L1 | Bi | 1.12 | 2.75 | 0.36 | 1.28 | 1.37 | 3.00 |

표4. K-val에 따른 제안 방법의 성능 평가
Table. 4. Performance evaluation of proposed method with different settings on K-val

| Setting | K15-val | | K12-val | |
|---|---|---|---|---|
| | $E_{EP}$(pixel) | $E_{D1}$(%) | $E_{EP}$(pixel) | $E_{D1}$(%) |
| $S_1$ | 0.74 | 2.33 | 0.62 | 2.05 |
| $S_2$ | 0.75 | 2.02 | 0.63 | 1.78 |
| $S_3$ | 0.81 | 2.33 | 0.70 | 1.98 |

S=3, C=3일 때 $E_{EP}$는 약 2% 증가했고, tRun은 약 47% 낮아졌다. 비용은 적게 들지만 효율은 현저히 향상되었다.

### 4. 시차 회귀 함수가 알고리즘 성능에 미치는 영향

S=1, C=1보다 S=2, 3, C=3일 때 오차는 약 4%, 10% 증가하지만 효율은 약 31%, 47% 향상됐다. 따라서 S=2, 3, C=3을 설정하고 두 종류의 비용함수가 알고리즘에 미치는 영향을 실험하였으며 성능 대비 결과는 표 3과 같다. 표 3에서 L1은 식 (8)의 매끄러운 L1-norm 비용으로 모델 학습하였으며, CE는 식 (7)의 크로스 엔트로피 비용으로 모델 학습하였고, CE+L1은 식 (6)으로 모델 학습하였다. Tri/Bi는 3/2개 차원에서 매칭 비용을 선형적으로 샘플링하였다. 학습 모델의 매끄러운 L1-norm 비용보다 크로스 엔트로피 비용이 평가지표인 ED1의 개선에서 두드러지고 시차를 확장했을 때 알고리즘 정밀도의 낙폭이 더 적다는 것을 알 수 있다. 파라미터 {1,1,L1,Tri}과 {2,3,CE+L1,Bi}로 설정했을 때 $E_{EP}$는 약 2%, 10% 증가했다. $E_{D1}$은 0.72%, 0.66% 낮아진다. $t_{Run}$은 약 40%, 52% 낮아진다. 이상점 (Outlier) 픽셀수를 줄일 뿐 아니라 효율을 크게 높였다.

표 3의 세 그룹 모델 S1={1,1,L1,Tri}, S2={2,3,CE+L1,Bi}, S3={3,3,CE+L1,Bi}를 설정하고, K-train에서 미세 조정하여 K-val에서 테스트한 성능 비교 결과는 표 4과 같다. 표 4에서는 K15-val와 K12-val은 각각 K-val 중 두 개의 서브세트에 대응한다. 알고리즘 운영의 효율성을 높일 뿐 아니라 $E_{D1}$을 평가 지표로 삼을 때 유리하다는 것을 알 수 있다.

### 5. 기존 딥러닝 기반 알고리즘과의 성능 비교

K15-test와 K12-test 시험 이미지를 사용하여 개선된 알고리즘 (표 4의 설정 파라미터 S2)에 대해 시차를 추정하였

표 5. K12-테스트에 대한 다른 방법과의 성능 평가
Table 5. Performance evaluation of different methods on K12-test

| Method | $\gamma_2$(%) | | $\gamma_3$(%) | | $\gamma_5$(%) | | Mean $E_{EP}$ | |
|---|---|---|---|---|---|---|---|---|
| | Noc | All | Noc | All | Noc | All | Noc | All |
| DispNetC[6] | 7.38 | 8.11 | 4.11 | 4.65 | 2.05 | 2.39 | 0.9 | 1.0 |
| MC-CNN[7] | 3.90 | 5.45 | 2.43 | 3.63 | 1.64 | 2.39 | 0.7 | 0.9 |
| iResNet[8] | 2.69 | 3.34 | 1.71 | 2.16 | 1.06 | 1.32 | 0.5 | 0.6 |
| GC-net[9] | 2.71 | 3.46 | 1.77 | 2.30 | 1.12 | 1.46 | 0.6 | 0.7 |
| PSMNet[10] | 2.44 | 3.01 | 1.49 | 1.89 | 0.90 | 1.15 | 0.5 | 0.6 |
| Proposed | 2.28 | 2.95 | 1.33 | 1.85 | 0.86 | 1.16 | 0.6 | 0.6 |

다. 기존 딥러닝 기반 알고리즘과 개선된 제안 알고리즘의 성능 비교는 표 5와 표 6에 나타내었다. 이 성능 비교의 visual results는 그림 3와 그림 4에 나타내었다. 'All'은 평가 시 모든 픽셀을 포함하고, 'Noc'은 비차폐 영역 내 픽셀만 고려한다. 표 6에서는 $E_{D1-bg}$, $E_{D1-fg}$와 $E_{D1-all}$은 배경영역, 전경영역, 모든 영역에서 계산된 평가지표 $E_{EP}$를 나타낸다.

표 6 . K15-테스트에 대한 다른 방법과의 성능 평가
Table 6. Performance evaluation of different methods on K15-test

| Method | $E_{D1}$ All (%) | | | $E_{D1}$ Noc (%) | | | (s) |
|---|---|---|---|---|---|---|---|
| | bg | fg | all | bg | fg | all | $t_{Run}$ |
| DispNetC[6] | 4.32 | 4.41 | 4.34 | 4.11 | 3.72 | 4.05 | 0.06 |
| MC-CNN[7] | 2.89 | 8.88 | 3.89 | 2.48 | 7.64 | 3.33 | 67.00 |
| iResNet[8] | 2.55 | 3.40 | 2.44 | 2.07 | 2.76 | 2.19 | 0.12 |
| GC-net[9] | 2.21 | 6.16 | 2.87 | 2.02 | 5.58 | 2.61 | 0.90 |
| PSMNet[10] | 1.86 | 4.62 | 2.32 | 1.71 | 4.31 | 2.14 | 0.41 |
| Proposed | 1.67 | 3.94 | 2.03 | 1.46 | 3.47 | 1.79 | 0.34 |

표 5에서 $\gamma_n$는 평가 구역에서 error 픽셀 백분율이다. 오차맵은 0 (검은색)과 >=5 (흰색) 픽셀 사이에서 선형적으로 스케일링된다. 기존 딥러닝 기반 알고리즘에 비해 정확도와

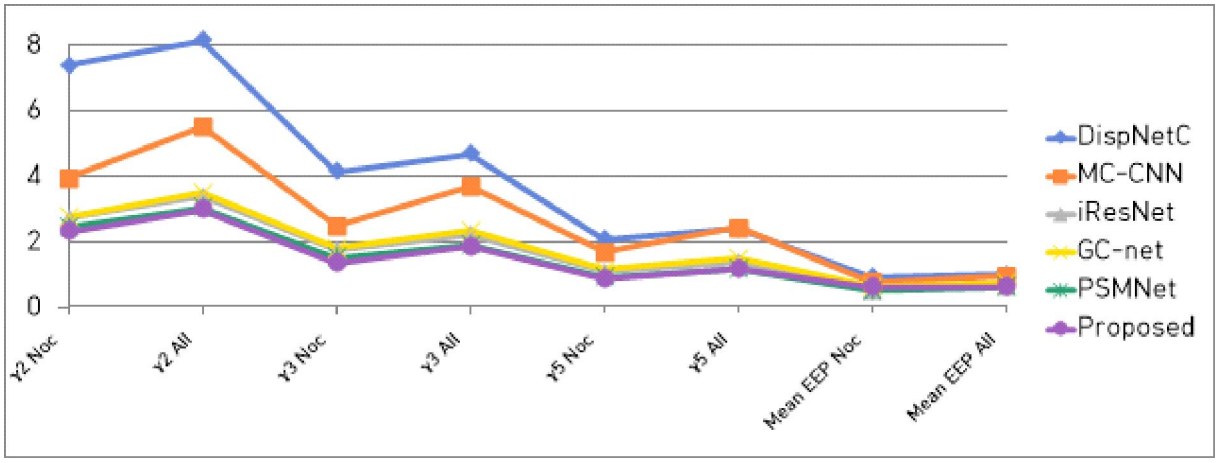


그림 3. K12-테스트에 대한 다른 방법과의 성능 평가
Fig. 3. Performance evaluation of different methods on K12-test

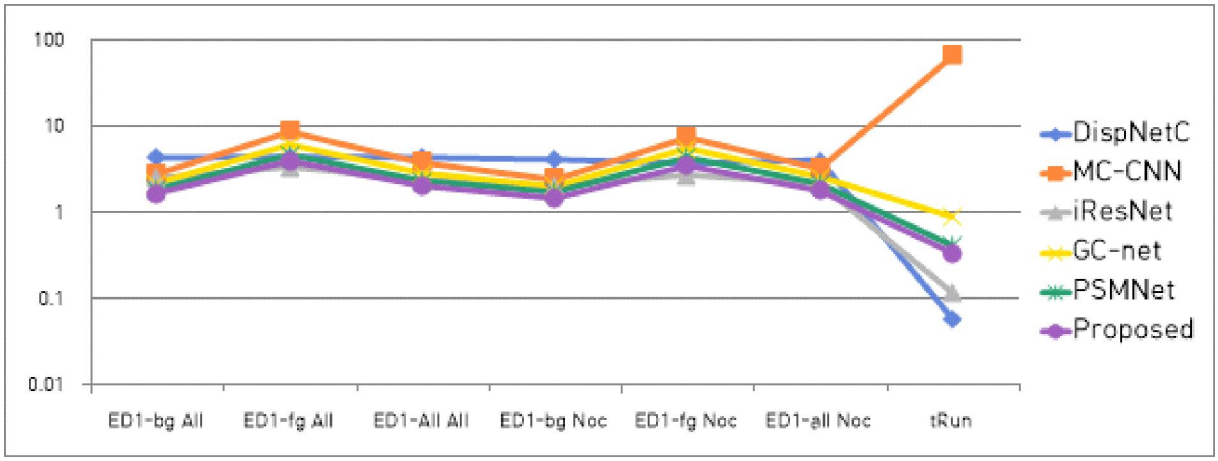


그림 4. K15-테스트에 대한 다른 방법과의 성능 평가
Fig. 4. Performance evaluation of different methods on K15-test

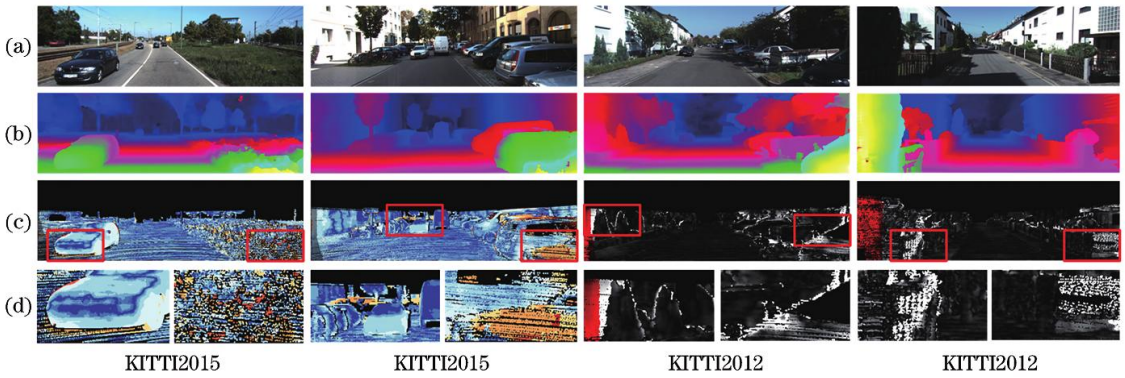


그림 5. 제안된 방법에 의해 예측된 시차 (a) 왼쪽 이미지, (b) 변위 지도, (c) 오류 지도, (d) 로컬 세부 정보
Fig. 5. disparity predicted by proposed method (a) left image; (b) disparity map; (c) error map; (d) local details

계산 효율에서 강점을 가지고 있으며, K15-test와 K12-test의 비차폐 영역에서 가장 높은 정밀도를 가짐을 알 수 있다.

제안하는 개선된 알고리즘의 주요 문제점을 분석하기 위한 K15-test와 K12-test에서 각각 오차가 큰 2개의 이미지 그룹을 선택하여 평가하였다. 결과는 그림 5과 같다. 오차맵(error map)은 추정하는 시차도와 기준값 사이의 절대오차이다. 왼쪽 2개 그룹은 K15-test ($E_{D1-all}$은 3.80%와 3.55%),

## V. 결 론

스테레오 매칭에 3차원 합성곱 신경망을 도입해 3차원에서 전체적인 semantic 정보를 이해할 수 있도록 해 장면의 더 잘 이해할 수 있도록 적용하는 경우에 모델링 효과가 좋아지지만 매칭 비용과 계산량은 증가하게 되는데 이러한 자원의 부담을 줄이기 위해 본 논문에서는 1) 시차 차원에서의 희박 합성곱 모듈을 3차원 합성곱 신경망의 입력으로 구축하여 비디오 메모리 비용 및 계산량을 감소시키고, 2) 평행변환 길이에 3차원 합성곱 신경망의 출력 카테고리 수를 예측하여 시차 축에서 매칭하는 비용에 대해 출력 카테고리 수에 맞춰 세분화하여 샘플링하고, 3) 최대 확률 주변에서의 시차 회귀와 부분 픽셀의 교차 엔트로피 비용과 매끄러운) L1-norm 비용을 결합하여 학습함으로써 모형이 시차 이미지를 더 정확하게 추정할 수 있게 하고 시차 범위를 확장할 때 정밀도에 미치는 영향이 크지 않도록 하는 3가지 특징을 가지는 향상된 3차원 합성곱 신경망 스테레오 매칭 알고리즘을 제안하였다.

실험을 통하여 제안된 알고리즘이 K15-test와 K12-test에서는 기존의 딥러닝 기반 방법과 비교하여 전체적으로 정밀한 성능은 최적화되었고, 특히 비교 대상 방법에 비해 알고리즘 정확도가 높아졌을 뿐 아니라 계산 시간도 대폭 단축됨을 확인할 수 있었다. 다만 알고리즘에 대한 절대적인 계산 비용이 높아 여전히 실시간 요구사항은 만족시킬 수는 없는 한계를 가지고 있다. 향후 정밀도가 높고 계산 효율이 높아 실시간성을 만족시킬 수 있는 3차원 합성곱 신경망 기반 스테레오 매칭 알고리즘 개발을 위해 네트워크 구조에 대한 추가 연구가 필요하다.

## References

[1] V.D. Nguyen, D.D. Nguyen, S.J. Lee, J,W. Jeon,, “Local Density Encoding for Robust Stereo Matching,” Transactions on Circuits and Systems for Video Technology, Vol. 24, No. 12, pp. 2049-2062, 2014.

[2] R.A. Hamzah, H. Ibrahim, A.H.A. Hassan, “Stereo Matching Algorithm Based on per Pixel Difference Adjustment, Iterative Guided Filter and Graph Segmentation,” Journal of Visual Communication and Image Representation, pp. 145-160, 2017.

[3] S.Q. Ren, K.M. He, R. Girshick, J. Sun, “Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks,” Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, No. 6, pp. 1137-1149, 2015.

[4] E. Shelhamer, J. Long, T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, No. 4, pp. 640-651, 2017.

[5] J.S. Xiao, H. Tian, W.T. Zou, “Stereo Matching Based on Convolutional Neural Network,” Acta Optica Sinica, 2018.

[6] J. Zbontar, Y. LeCun, “Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches,” Journal of Machine Learning Research, Vol. 17, No. 1, pp. 2287-2318, 2016.

[7] N. Mayer, E. Ilg, P. Hausser, “A Large Dataset to Train Convolutional Network for Disparity, Optical Flow, and Scene Flow Estimation,” Conference on Computer Vision and Pattern Recognition, Las Vegas,

NV, USA, New York, pp. 4040-4048, 2016.
[8] J.H. Pang, W.X. Sun, J.S. Ren, "Cascade Residual Learning: A Two-stage Convolutional Neural Network for Stereo Matching," International Conference on Computer Vision Workshops, Venice, Italy. New York, pp. 878-886, 2017.
[9] A. Kendall, H. Martirosyan, S. Dasgupta, P. Henry, R. Kennedy, A. Bachrach, A. Bry, "End-to-end Learning of Geometry and Context for Deep Stereo Regression", International Conference on Computer Vision, Venice, Italy. New York, pp. 66-75, 2017.
[10] J.R. Chang, Y.S. Chen, "Pyramid Stereo Matching Network," Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA. New York, pp. 5410-5418, 2018.
[11] W.J. Luo, A.G. Schwing, R. Urtasun, "Efficient Deep Learning for Stereo Matching," Conference on Computer Vision and Pattern Recognition, Las Vegas, UT, USA. New York, pp. 5695-5703, 2016.
[12] Z.F. Liang, Y.L. Feng, Y.L. Guo, H. Liu, W. Chen, L. Qiao, L. Zhou, J. Zhang, "Learning for Disparity Estimation Through Feature Constancy," Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA. New York, pp. 2811-2820, 2018.
[13] Z.Q. Jie, P.F. Wang, Y.G Ling, B. Zhao, Y. Wei, J. Feng, W. Liu, "Left-right Comparative Recurrent Model for Stereo Matching," Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA. New York, pp. 3838-3846, 2018.
[14] L.D. Yu, Y.C. Wang, Y.W. Wu, Y.D. Jia, "Deep Stereo Matching with Explicit Cost Aggregation Sub-architecture," http://cn.arxiv.org/abs/1801.04065, 2018.
[15] N. Smolyanskiy, A. Kamenev, S. Birchfield, "On the Importance of Stereo for Accurate Depth Estimation: an Efficient Semi-supervised Deep Neural Network approach," Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA. New York, pp. 1120-1128, 2018.
[16] F.H. Zhang, V Prisacariu, R. Yang, P.H.S. Torr, "GA-Net: Guided Aggregation Net for End-To-End Stereo Matching," IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach, CA, USA, pp. 185-194, 2019.
[17] J.H. Huang, R.F. Zhang, Y.H. Liu, "Research on An Optimized Deep Learning Stereo Matching Algorithm," Laser & Optoelectronics Progress, Vol. 58, No. 24, pp. 37-55, 2021.
[18] R. Girshick, "Fast R-CNN," International Conference on Computer Vision, Santiago, Chile. New York, pp. 1440-1448, 2015.
[19] D.P. Kingma, J. Ba, "Adam: A Method for Stochastic Optimization," http://cn.arxiv.org/abs/1412.6980, 2017.

**Jian Wang (왕 지 엔)**

2014 Department of Naval Architecture and Ocean Engineering from Kunsan National University (B.S)
2017 Department of Naval Architecture and Ocean Engineering from Kunsan National University (M.S)

2018~Department of Naval Architecture and Ocean Engineering from Kunsan National University (Ph.D)
Career:
2017~2018 Ulsan Department of Naval Architecture and Ocean Engineering from Ludong University, China, Research assistant

Field of Interests: Stereo Matching, Deep Learning.

Email: wjza989@hotmail.com

**Jackyou Noh (노 재 규)**

1996 Naval Architecture & Ocean Engineering from Seoul National University, Seoul, Republic of Korea (B.S.)
1998 Naval Architecture & Ocean Engineering from Seoul National University, Seoul, Republic of Korea (M.S.)

2009 Naval Architecture & Ocean Engineering from Seoul National University, Seoul, Republic of Korea (Ph.D.)
2010~Naval Architecture & Ocean Engineering, in Kunsan National University (Prof.)
Career:
2014~Korean J. of Computational Design and Engineering, Associate Editor
2014~IT Convergence of The Korean Society of Mechanical Engineers, Member of board of directors
2021~The Korean Society of Mechanical Engineers, Member of board of directors
Field of Interests: Systems Control, and Deep Learning
Email: snucurl@kunsan.ac.kr