

특집논문 (Special Paper)

방송공학회논문지 제29권 제4호, 2024년 7월 (JBE Vol.29, No.4, July 2024)

<https://doi.org/10.5909/JBE.2024.29.4.399>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

다중 해상도 이미지를 활용한 자기 지도 학습 기반 단안 깊이 추정

류 건^{a)}, 김 원 준^{a)‡}

Self-supervised Monocular Depth Estimation with Multiple Resolution Images

Gun Ryu^{a)} and Wonjun Kim^{a)‡}

요 약

최근 자율 주행 분야 등에서 깊이 추정 기술에 대한 수요가 급증함에 따라 심층학습(Deep Learning)기반의 단안 깊이 추정 연구가 활발히 진행되고 있다. 라이다(LiDAR) 감지기를 기반으로 획득한 실측 정보를 정답 값(Ground Truth)으로 이용하여 신경망(Neural Network)을 학습하는 대부분의 기존 방법과는 달리, 본 논문에서는 날씨에 영향을 많이 받는 라이다 기반 실측 정보를 사용하지 않고 학습할 수 있는 자기 지도(Self-supervised) 학습 방법을 제안한다. 특히, 깊이 지도의 전체적인 기하학적 구조와 지역적 경계를 함께 고려할 수 있도록 다중 해상도 이미지 입력을 사용하였으며, 각 해상도에서 추출한 잠재적 특성을 혼합할 수 있는 신경망 구조를 제안하였다. 기준 데이터셋(Benchmark Dataset)에서의 실험 결과를 통해 제안하는 방법이 깊이 추정 성능을 효과적으로 개선할 수 있음을 보였다.

Abstract

Due to the rapidly increasing demand for accurate depth estimation and its applications, such as autonomous driving, self-supervised monocular depth estimation based on deep learning has been actively studied in recent years. Existing methods rely on the ground truth depth acquired through LiDAR sensors. In this paper, we propose a self-supervised learning method that enables training without LiDAR-based ground truth that is vulnerable to bad weather. Specifically, our method utilizes multiple-resolution image inputs to consider the overall geometric structure and regional boundaries of the depth map together, and the proposed neural network architecture merges the potential features extracted from each resolution. Experimental results on the benchmark dataset demonstrate that the proposed method can effectively improve the monocular depth estimation performance.

Keyword : Self-supervised learning, Monocular depth estimation, Deep neural network

a) 건국대학교 전기전자공학부(Department of Electrical and Electronics Engineering, Konkuk University)

‡ Corresponding Author : 김원준(Wonjun Kim)

E-mail: wonjkim@konkuk.ac.kr

Tel: +82-2-450-3396

ORCID: <https://orcid.org/0000-0001-5121-5931>

※ 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2023R1A2C1003699).

※ This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2023R1A2C1003699).

· Manuscript May 10, 2024; Revised June 10, 2024; Accepted June 10, 2024.

Copyright © 2024 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

I. 서론

자율 주행, 지능형 로봇 서비스 등 다양한 산업 분야에서 카메라로부터 목표 지점까지의 깊이 값을 알아내는 깊이 추정에 대한 수요가 지속적으로 증가하고 있다. 최근에는 심층학습(Deep Learning) 기술의 발전으로 신경망(Neural Network)을 활용한 단안 깊이 추정 연구가 활발히 진행되고 있다. 신경망을 이용한 단안 깊이 추정 방법은 라이다(LiDAR) 기반 실측에 비해 날씨 등의 외부 환경 변화에 강인하며, RGB 카메라만 사용하기 때문에 저렴한 비용으로 정밀하게 깊이 값을 추정할 수 있다는 장점이 있다.

초기 심층학습 기반 단안 깊이 추정 연구는 깊이 정답 값(Ground Truth)을 사용하는 지도(Supervised) 학습을 이용하였다. Eigen^[1] 등은 합성곱(Convolution) 기반의 압축기(Encoder)와 복원기(Decoder) 구조를 사용하여 대략적인 깊이 지도(Depth Map)를 생성한 후 다시 지역적 상세 정보를 추정하는 2단계 구조를 제안하였다. Xu^[2] 등과 Lui^[3] 등은 조건부 무작위장(Conditional Random Field)을 활용하여 다중 해상도 신경망으로부터 추출된 특성을 통합하여 깊이 추정의 정확도를 높였다. 또한, Cao^[4] 등은 깊이 정보의 속성을 기반으로 회귀(Regression) 문제를 분류(Classification) 문제로 재 정의하여 깊이 예측 성능을 개선하였다. 최근에는 트랜스포머(Transformer)^[5]를 활용한 다양한 방법들이 연구되고 있다. Bhat^[6] 등은 트랜스포머 구조를 통해 입력 이미지에 따라 깊이를 적응적으로 추정할 수 있는 깊이 눈금(Depth Bins)을 활용하는 방법을 제안하였다. Piccinelli^[7] 등도 트랜스포머 구조를 통해 군집화(Clustering)를 수행하여 깊이 패턴(Depth patterns)을 추출하는 방식으로 깊이 추정 성능을 향상시켰다. 그러나 라이다 기반 실측 정보를 정답 값으로 사용하는 지도 학습은 날씨 등 외부 요인에 의해 학습 정확도가 저하되는 문제점이 있다.

이러한 문제를 해결하기 위해, 정답 값을 사용하지 않는 자기 지도(Self-supervised) 학습 기반 깊이 추정 방법이 활발하게 연구되고 있다. Godard^[8] 등은 스테레오(Stereo) 카메라로 촬영된 이미지들을 통해 시차 지도(Disparity Map)를 추정하고 시차와 깊이 간의 기하학적 구조를 기반으로 깊이 지도를 생성하는 자기 지도 학습 방법을 제안하였다. 더 나아가, 단일 카메라에서 촬영된 영상을 기반으로 프레

임 간 카메라 위치의 변화를 추정하여 비디오를 활용한 자기 지도 학습 기반 단안 깊이 추정 방법 또한 제안하였다^[9]. 한편으로, Zhao^[10] 등은 합성곱 신경망 압축기에 트랜스포머 계층을 추가하여 전체적인 기하학적 구조 정보를 파악하고 이를 통해 효과적으로 깊이 추정 성능을 향상시켰다. Wang^[11] 등은 복원기에 트랜스포머 기반 모듈을 추가하여 전역적인 특성(Global Feature)과 지역적인 특성(Local Feature)의 관계를 나타낸 비용 용량(Cost Volume)으로부터 깊이 정보를 추출하는 방법을 제안하였다.

본 논문에서는 다중 해상도 입력을 통해 전체적인 기하학적 특성과 지역적 경계를 모두 고려할 수 있는 자기 지도 학습 기반 신경망 구조를 제안한다. 제안하는 방법은 서로 다른 해상도의 이미지로부터 각 해상도의 특징을 추출하고 혼합하여 깊이 추정 성능을 개선하였다. 기존 데이터셋(Benchmark Dataset)에 대한 성능 평가를 통해 기존 방법 대비 제안하는 방법이 자기 지도 학습 기반 단안 깊이 추정에 효과적임을 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 제안하는 다중 해상도 특성 추출 및 혼합 모듈에 대해 자세히 설명하며, 3장에서는 다양한 실험을 통해 제안하는 방법이 기존 자기 지도 학습 기반 깊이 추정 방법보다 효과적임을 검증한다. 마지막으로 4장에서는 본 논문의 결론을 서술한다.

II. 제안하는 방법

제안하는 방법은 다중 해상도 입력으로부터 각 해상도에 해당하는 깊이 지도를 생성한다. 또한, 혼합 신경망을 통해 서로 다른 해상도의 깊이 지도에서 깊이 추정에 유용한 특성을 추출하여 최종 깊이 지도를 생성한다. 본 장에서는 먼저 제안하는 방법의 신경망 구조에 대해 자세히 설명하고 제안하는 신경망의 훈련에 사용된 손실 함수에 관해 설명한다.

1. 다중 해상도 입력을 통한 깊이 지도 생성 및 깊이 지도 혼합

그림 1은 제안하는 신경망의 전체적인 구조를 보여준다.

제안하는 구조는 크게 두 개의 모듈로 이루어져 있다. 첫 번째 모듈은 다중 해상도 입력을 통해 깊이 지도를 생성하고, 두 번째 모듈에서는 두 프레임 간 이미지 특징 관계를 기반으로 카메라 위치를 추정한다. 본 논문에서는 $SQLDepth^{[1]}$ 모델을 이용하여 서로 다른 해상도의 깊이 지도를 생성한다. $SQLDepth^{[1]}$ 는 트랜스포머 기반 모듈을 통해 특성 지도로부터 전체적인 기하학적 구조와 지역적 특징을 포함한 비용 용량을 생성한다. 이후, 생성된 비용 용량을 통해 각 픽셀의 깊이에 대한 이산 확률을 추정하고, 이 확률들을 가중 합하여 깊이 지도를 생성한다. 본 논문에서는 사전 학습된 $SQLDepth^{[1]}$ 모델을 사용하며 훈련 중 가중치는 갱신하지 않는다. 서로 다른 해상도에서 학습된 특징의 차이는 그

림 2를 통해 확인할 수 있다. 그림 2 (b)와 (c)를 통해 제안하는 방법이 저해상도 이미지의 부족한 지역적 특성을 고해상도 이미지로부터 보완할 수 있음을 확인할 수 있다. 제안하는 방법은 서로 다른 특성이 반영된 깊이 지도에서 깊이 추정에 유용한 특성을 추출하여 혼합하는 깊이 지도 혼합 신경망을 기존 깊이 지도 생성 신경망에 추가한다. 깊이 지도 혼합 신경망의 입력은 고해상도 깊이 지도와 고해상도로 확대된 저해상도 깊이 지도의 연접(Concatenate)이다. 깊이 지도 혼합 신경망은 U자형 합성곱 신경망 구조(U-Net)를 활용하여 서로 다른 특성을 가진 다중 해상도 깊이 지도로부터 깊이 추정에 유용한 특성을 추출한다. 미리 정의된 깊이 최소 값과 최대 값을 사용한 정규화 기법은

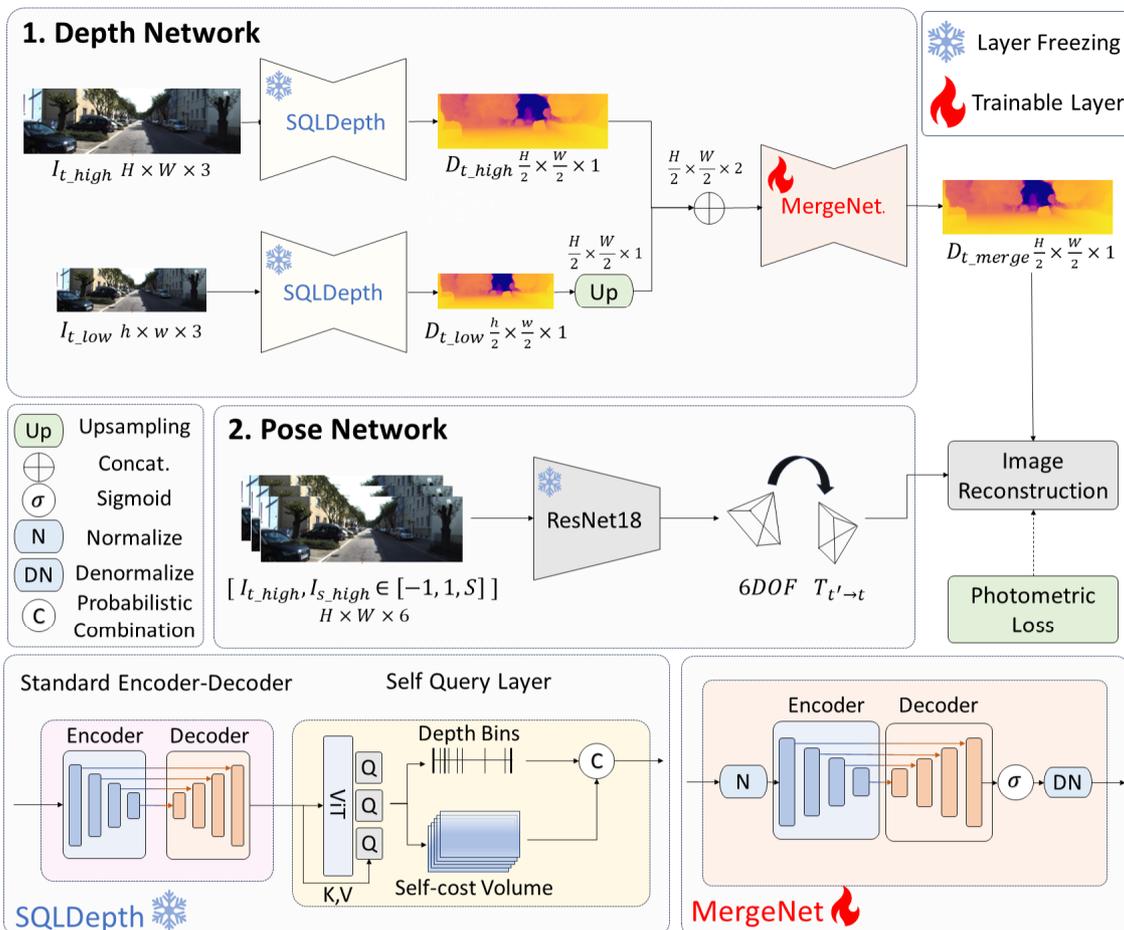


그림 1. 제안하는 방법의 신경망 상세 구조
 Fig. 1. Overview of the proposed method

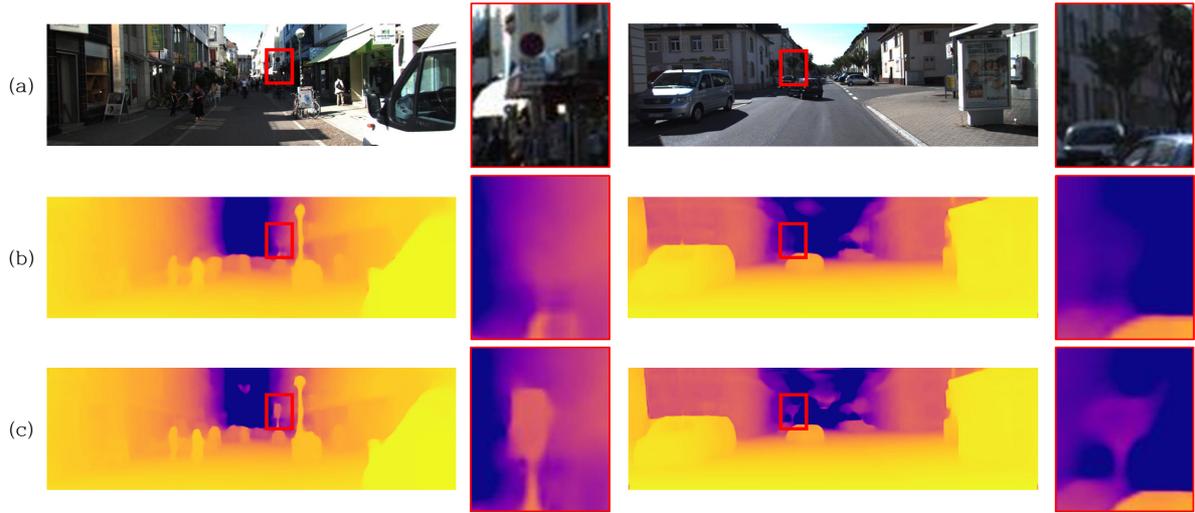


그림 2. 입력 이미지 해상도와 깊이 지도의 관계 (a): 입력 이미지 (b): 저해상도(640x192) 이미지 입력으로부터 생성된 깊이 지도 (c): 고해상도(1024x320) 이미지 입력으로부터 생성된 깊이 지도

Fig. 2. Relationship between the resolution of input images and depth maps (a): Input image (b): Depth maps of low-resolution(640x192) input (c): Depth maps of high-resolution(1024x320) input

신경망의 입력과 출력에 적용되며, 신경망의 학습 안정성을 높이고 출력된 특성 지도(Feature Map)가 깊이 값에 효과적으로 대응될 수 있도록 한다.

또한, 본 논문에서는 Zhou^[12] 등을 따라 자기 지도 학습 기반 단안 깊이 추정에서 널리 사용되는 카메라 위치 추정 신경망을 사용한다. 카메라 위치 추정 신경망은 연속하는 프레임을 입력으로 사용하여 카메라 위치 변화를 출력한다. 제안하는 방법은 기존 자기 지도 학습 기반 단안 깊이 추정 방법을 따라 생성된 깊이 지도와 카메라 위치 변화를 통해 식 1과 같이 주변 프레임 I_t 의 RGB 값으로부터 현재 프레임 $I_{t \rightarrow t'}$ 을 재구성한 이미지 $I_{t \rightarrow t}$ 를 생성한다.

$$I_{t \rightarrow t} = I_t \langle \text{proj}(D_{t_{\text{merge}}}, T_{t \rightarrow t'}, K) \rangle, \quad (1)$$

여기서 $D_{t_{\text{merge}}}$ 는 혼합된 깊이 지도, $T_{t \rightarrow t'}$ 는 카메라 위치 변환 행렬, K 는 카메라 내부 파라미터를 나타내는 행렬이며, proj 는 생성된 깊이 지도 $D_{t_{\text{merge}}}$ 를 주변 프레임 t' 의 평면으로 투영(Projection)하는 연산을 의미하며, 그리고 $\langle \rangle$ 는 이미지 I_t 재구성을 위한 샘플링(Sampling)

연산을 의미한다.

2. 제안하는 신경망 훈련을 위한 손실 함수

본 논문에서는 MonoDepth2^[9]에서 제안된 손실함수를 사용하여 제안하는 신경망을 학습한다. 먼저, 측광 손실 함수(Photometric Loss)를 사용하여 재구성된 이미지 $I_{t \rightarrow t}$ 와 원본 이미지 I_t 간의 차이를 식 2와 같이 계산한다.

$$L_p = \min(\sum_{t'} pe(I_t, I_{t \rightarrow t})), \quad (2)$$

$$PE(I_a, I_b) = \frac{\alpha}{2}(1 - SSIM(I_a, I_b)) + (1 - \alpha)\|I_a - I_b\|_1, \quad (3)$$

여기서 측광 손실 함수 L_p 은 식 3과 같이 $SSIM$ ^[13]과 $L1$ 손실로 이루어져 있으며, α 는 0.85로 설정한다. 또한, 식 4와 같이 경계를 고려한 평활화 함수(Edge-aware Smooth Loss)를 통해 이웃한 픽셀의 색상 차이가 작은 영

역에서 깊이 값의 급격한 변화를 방지한다.

$$L_s = |\partial_x d_t^*| e^{-|\partial_x I_t|} + |\partial_y d_t^*| e^{-|\partial_y I_t|}, \quad d_t^* = d_t / \bar{d}_t, \quad (4)$$

여기서 I_t 는 현재 프레임이고, d_t 는 현재 프레임에서 예측된 깊이 지도, 그리고 \bar{d}_t 는 정규화 된 깊이 지도이다.

III. 실험 결과 및 분석

본 논문에서는 제안하는 방법의 성능 평가를 위해 KITTI 데이터셋^[14]을 사용하였다. KITTI 데이터셋^[14]은 다양한 교통 시나리오가 포함된 도로 주행 영상 데이터셋이며 10Hz ~ 100Hz 프레임의 6시간 분량 영상으로 구성되어 있다. Zhou^[12] 등을 따라, 71634 프레임이 학습을 위해 사용되었고, 4424 프레임이 검증을 위해 사용되었다. 또한, 697 프레임은 Eigen^[1] 등을 따라 성능 평가에 사용되었다. 제안하는 방법은 PyTorch^[15] 프레임워크를 기반으로 구현되었다. 본 논문에서는 신경망 가중치를 최적화하기 위한 알고리즘으로 Adam^[16]을 사용하였고, β_1 과 β_2 값은 각각 0.9과 0.999로 설정하였다. 학습 속도(Learning Rate)는 1×10^{-4} 로 설정하고, 15 에포크(Epoch) 지점에서 10^{-1} 만큼 감소시켰으며, 총 25 에포크 동안 학습을 진행하였다. 학습 시에는 640×192 픽셀 크기와 1024×320 픽셀 크기의 이미지를 모두 사용하였고, 데이터 증강을 위해 이미지의 밝기,

대비, 채도 및 색조를 변경하였다. 학습과 성능 평가에는 Intel(R) Core(TM) i7-6850K CPU @ 3.60GHz CPU와 NVIDIA GeForce RTX 3090 GPU 2대가 사용되었다.

제안하는 방법의 성능 개선 효과를 검증하기 위해, 기존 단안 깊이 추정에서 주로 사용하는 평가 지표로 제안하는 방법의 성능을 측정하였다. 단안 깊이 추정 성능 평가 지표는 깊이를 정규화 하여 예측 깊이와 실제 깊이 간의 오류를 계산하는 절대 상대 오차(Absolute Relative Error)와 제곱 상대 오차(Square Relative Error), 정규화하지 않고 오류를 계산하는 평균 제곱근 오차(RMSE)와 로그 평균 제곱근 오차(LogRMSE) 그리고 예측 깊이가 특정 임계값 안에 속하는지 평가하는 임계값 미만의 정확도(Accuracy Under Threshold)로 이루어져있다. 대표적 자기 지도 학습 기반 단안 깊이 추정 방법인 MonoDepth2^[9], ManyDepth^[17], MonoViT^[10], 그리고 SQLDepth^[11]와 성능을 비교하여 표 1에 결과를 나타내었다. 표 1에서 제안하는 방법의 성능 개선 효과를 확인할 수 있다.

제안하는 다중 해상도 입력 및 혼합 모듈이 전체적인 기하학적 구조와 지역적 경계를 함께 고려하는지 검증하기 위해 SQLDepth^[11]와 비교한 정성적인 단안 깊이 추정 결과를 그림 3에 나타내었다. 또한, 제안하는 방법의 정성적인 단안 깊이 추정 결과를 대표적 자기 지도 학습 기반 단안 깊이 추정 방법들과 함께 그림 4에 나타내었다. 그림 3과 4를 통해, 제안하는 방법을 적용했을 때 신경망이 지역적인 특성과 전역적인 특성을 모두 고려하여 깊이 지도를 생성하는 것을 확인할 수 있다.

표 1. KITTI 데이터셋^[14]에서의 정량적 평가 비교

Table 1. Performance comparison on the KITTI^[14] eigen benchmark

Method	Resolution	Depth Error				Depth Accuracy		
		Lower is better				Higher is better		
		AbsRel	SqRel	RMSE	Log RMSE	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
MonoDepth2 ^[9]	640x192	0.106	0.861	4.699	0.185	0.887	0.960	0.982
ManyDepth ^[17]	640x192	0.090	0.713	4.261	0.170	0.914	0.966	0.983
MonoViT ^[10]	640x192	0.099	0.708	4.372	0.175	0.900	0.967	0.984
SQLDepth ^[11]	640x192	0.088	0.697	4.175	0.167	0.919	0.969	0.984
MonoDepth2 ^[9]	1024x320	0.106	0.806	4.630	0.193	0.876	0.958	0.980
ManyDepth ^[17]	1024x320	0.087	0.685	4.142	0.167	0.920	0.968	0.983
MonoViT ^[10]	1024x320	0.093	0.671	4.202	0.169	0.912	0.969	0.985
SQLDepth ^[11]	1024x320	0.082	0.607	3.914	0.160	0.928	0.972	0.985
SQLDepth ^[11] + Ours	Both	0.082	0.594	3.853	0.159	0.928	0.972	0.985

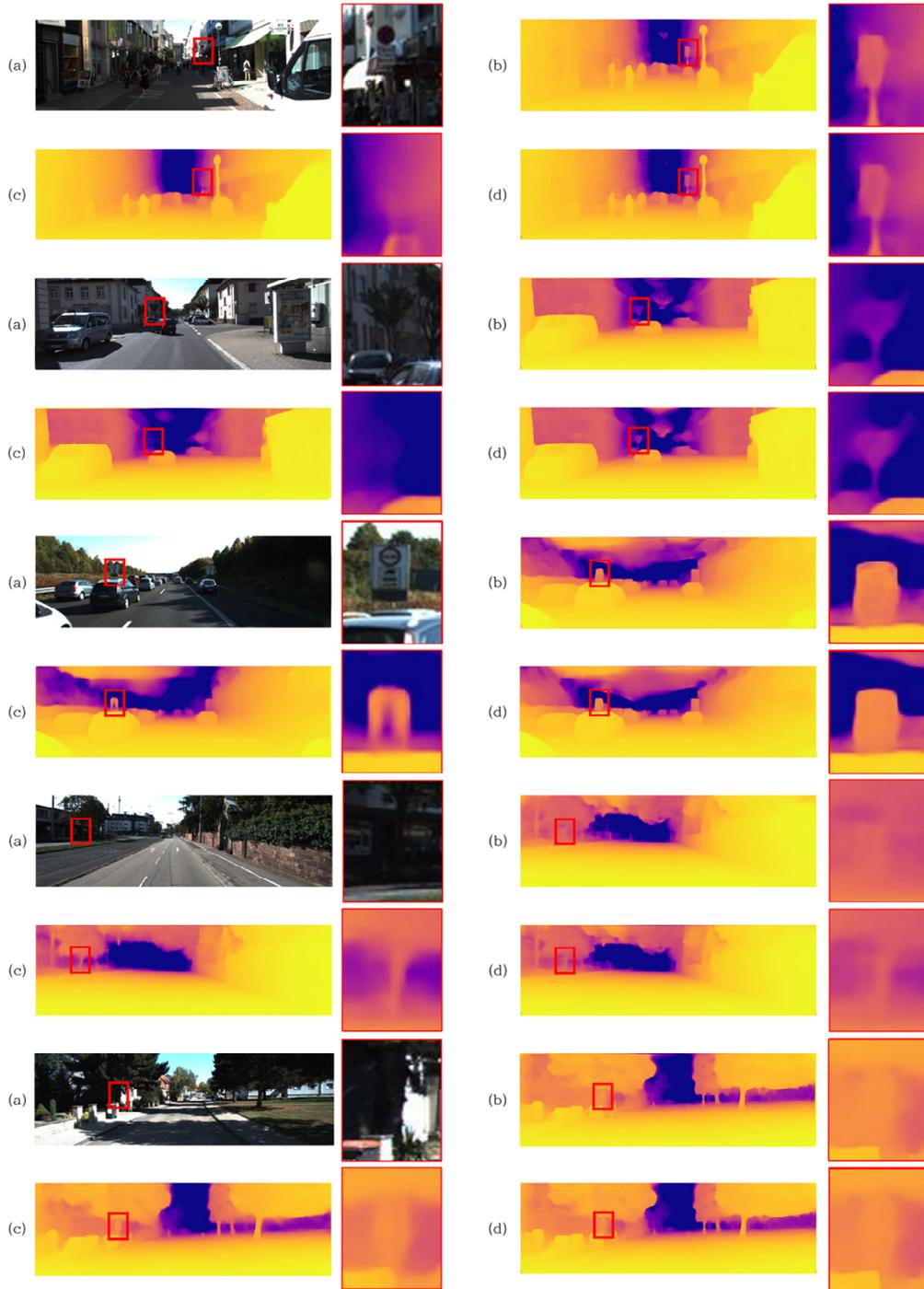


그림 3. KITTI 데이터셋^[14]에 대한 단안 깊이 추정 결과, (a): 입력 이미지 (b): 고해상도(1024x320) 이미지 입력으로부터 생성된 SQLDepth^[11]의 깊이 지도 (c): 저해상도(640x192) 이미지 입력으로부터 생성된 SQLDepth^[11]의 깊이 지도 (d): 제안하는 방법으로부터 생성된 깊이 지도
 Fig. 3. Results of monocular depth estimation on the KITTI dataset^[14] (a): Input image (b): Depth maps of high-resolution(1024x320) inputs by SQLDepth^[11] (c): Depth maps of low-resolution(640x192) inputs by SQLDepth^[11] (d): Depth maps by the proposed method

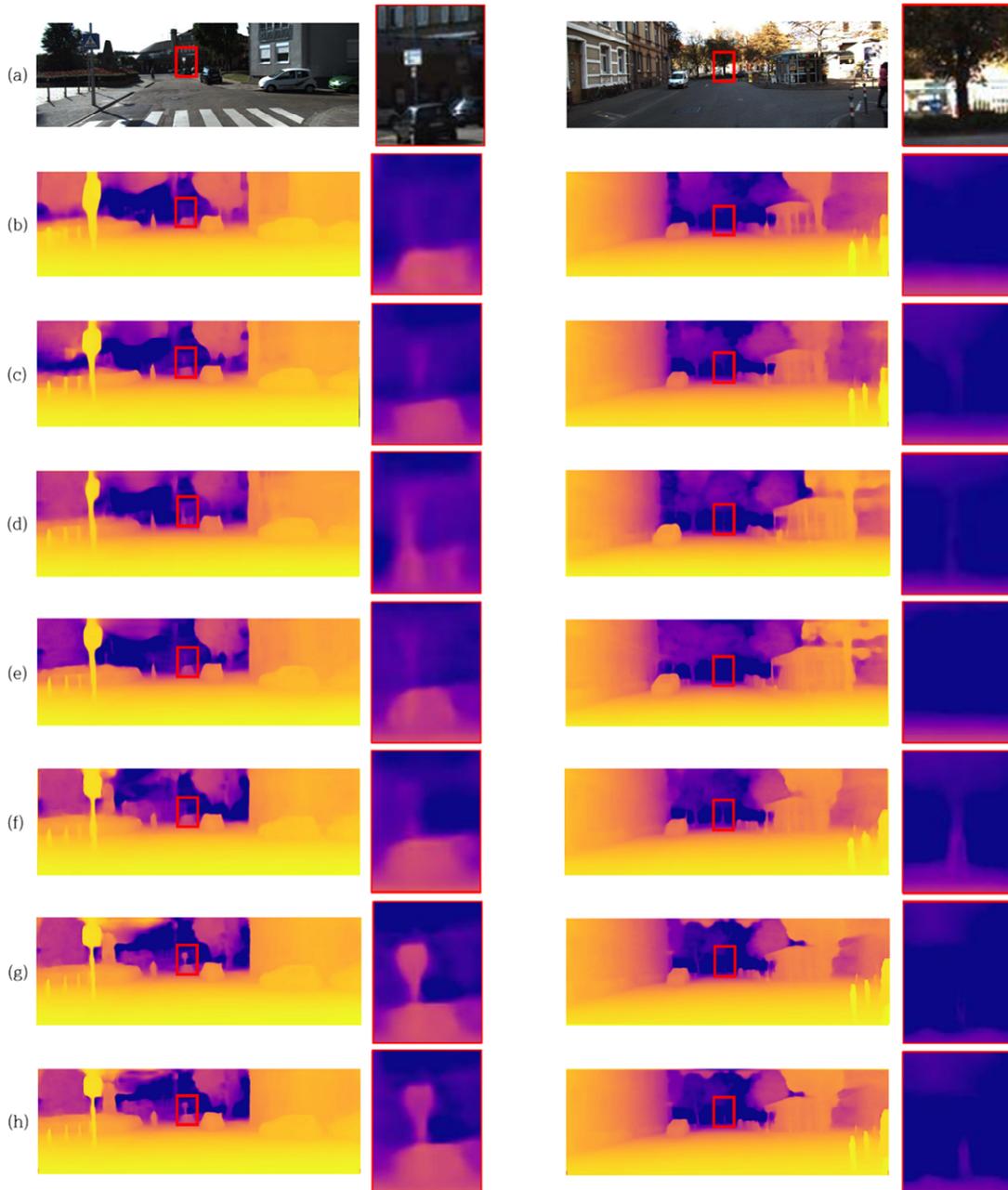


그림 4. KITTI 데이터셋^[14]에 대한 단안 깊이 추정 결과 (a): 입력 이미지 (b): 저해상도(640x192) 이미지 입력으로부터 생성된 MonoDepth2^[9]의 깊이 지도 (c): 고해상도(1024x320) 이미지 입력으로부터 생성된 MonoDepth2^[9]의 깊이 지도 (d): 저해상도 이미지 입력으로부터 생성된 ManyDepth^[17]의 깊이 지도 (e): 고해상도 이미지 입력으로부터 생성된 ManyDepth^[17]의 깊이 지도 (f): 저해상도 이미지 입력으로부터 생성된 SQLDepth^[11]의 깊이 지도 (g): 고해상도 이미지 입력으로부터 생성된 SQLDepth^[11]의 깊이 지도 (h): 제안하는 방법으로부터 생성된 깊이 지도

Fig. 4. Results of monocular depth estimation on the KITTI dataset^[14] (a): Input image (b): Depth maps of low-resolution(640x192) inputs by MonoDepth2^[9] (c): Depth maps of high-resolution(1024x320) inputs by MonoDepth2^[9] (d): Depth maps of low-resolution inputs by ManyDepth^[17] (e): Depth maps of high-resolution inputs by ManyDepth^[17] (f): Depth maps of low-resolution inputs by SQLDepth^[11] (g): Depth maps of high-resolution inputs by SQLDepth^[11] (h): Depth maps by the proposed method

IV. 결론

본 논문에서는 자기 지도 학습 기반 단안 깊이 추정에서 다중 해상도 입력과 깊이 지도 혼합 모듈을 제안하였다. 제안하는 방법은 다중 해상도의 입력으로부터 깊이 추정에 유용한 서로 다른 특성을 추출하고 혼합 모듈로 지역적 및 지역적 특성을 효과적으로 결합한다. 이를 통해, 제안하는 방법은 전체적인 기하학적 구조와 지역적 경계를 함께 고려한 깊이 지도를 생성할 수 있다. 다양한 실험 결과를 통해 제안하는 방법이 자기 지도 학습 기반 단안 깊이 추정 성능을 효과적으로 향상시킬 수 있음을 확인하였다.

그러나, 제안하는 방법을 적용하기 위해서는 다중 해상도 입력에 대해 사전 훈련된 깊이 추정 모듈이 필요하다. 이에 따라, 앞으로 사전 훈련된 모듈을 사용하지 않고, 중단 간 훈련으로 다중 해상도 이미지로부터 유용한 특성을 추출할 수 있는 모델 설계가 요구된다.

참고 문헌 (References)

- [1] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," in Proc. Adv. Neural Inf. Process. Syst., Dec. 2014, pp. 2366-2374.
doi: <https://doi.org/10.48550/arXiv.1406.2283>
- [2] D. Xu, E. Ricci, W. Ouyang, X. Wang, and N. Sebe, "Multi-scale continuous CRFs as sequential deep networks for monocular depth estimation," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 161-169.
doi: <https://doi.org/10.1109/cvpr.2017.25>
- [3] F. Liu, C. Shen, and G. Lin, "Deep convolution neural fields for depth estimation from a single image," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 5162-5170
doi: <https://doi.org/10.1109/cvpr.2015.7299152>
- [4] Y. Cao, T. Zhao, K. Xian, C. Shen, Z. Cao, and S. Xu, "Monocular depth estimation with augmented ordinal depth relationships," IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 8, pp. 2674-2682, Aug. 2020.
doi: <https://doi.org/10.1109/tip.2018.2877944>
- [5] A. Dosovitskiy et al., "An image is worth 16x16 words: transformers for image recognition at scale," in Proc. Int. Conf. Learn. Represent., May 2021, pp. 1-12.
doi: <https://doi.org/10.48550/arXiv.2010.11929>
- [6] S. F. Bhat, I. Alhashim, and P. Wonka, "AdaBins: depth estimation using adaptive bins," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2021, pp. 4009-4018.
doi: <https://doi.org/10.1109/cvpr46437.2021.00400>
- [7] L. Piccinelli, C. Sakaridis, and F. Yu, "iDisc: Internal discretization for monocular depth estimation," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2023, pp. 21477-21487.
doi: <https://doi.org/10.1109/cvpr52729.2023.02057>
- [8] C. Godard, O. M. Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2017, pp. 6602-6611.
doi: <https://doi.org/10.1109/cvpr.2017.699>
- [9] C. Godard, O. M. Aodha, M. Firman, and G. Brostow, "Digging into self-supervised monocular depth estimation," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Nov. 2019, pp. 3827-3837.
doi: <https://doi.org/10.1109/iccv.2019.00393>
- [10] Zhao, Chaoqiang, et al., "Monovit: Self-supervised monocular depth estimation with a vision transformer," in Proc. IEEE Int. Conf. 3D Vis., Sep. 2022, pp. 668-678.
doi: <https://doi.org/10.1109/3dv57658.2022.00077>
- [11] Y. Wang, Y. Liang, H. Xu, S. Jiao, and H. Yu, "SQLdepth: Generalizable self-supervised fine-structured monocular depth estimation," in Proc. AAAI, 2024, pp. 5713-5721.
doi: <https://doi.org/10.1609/aaai.v38i6.28383>
- [12] T. Zhou, M. Brown, N. Snavely, and D. Lowe, "Unsupervised learning of depth and ego-motion from video," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 1851-1858.
doi: <https://doi.org/10.1109/cvpr.2017.700>
- [13] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE transactions on image processing, vol. 13, no. 4, pp. 600-612, 2004.
doi: <https://doi.org/10.1109/tip.2003.819861>
- [14] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," Int. J. Robot. Res., vol. 32, no. 11, pp. 1231-1237, Aug. 2013.
doi: <https://doi.org/10.1177/0278364913491297>
- [15] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. Devito, Lin, A. Desmaison, L. Antiga, A. Lerer, "Automatic differentiation in pytorch," in Proc. Neural Inf. Process. Syst., Dec. 2017, pp. 1-4.
doi: <https://openreview.net/pdf?id=BJJsrMfCZ>
- [16] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in Proc. Int. Conf. Learn. Representation, May 2015, pp. 1-15.
doi: <https://doi.org/10.48550/arXiv.1412.6980>
- [17] J. Watson, O. Mac Aodha, V. Prisacariu, G. Brostow, and M. Firman, "The temporal opportunist: Self-supervised multi-frame monocular depth" in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2021, pp. 1164-1174.
doi: <https://doi.org/10.1109/cvpr46437.2021.00122>

저 자 소 개



류 건

- 2019년 3월 ~ 현재 : 건국대학교 전기전자공학부 학사 과정
- ORCID : <https://orcid.org/0009-0003-0992-9270>
- 주관심분야 : 컴퓨터 비전, 영상처리, 심층학습



김 원 준

- 2012년 8월 : 한국과학기술원(KAIST) 박사
- 2012년 9월 ~ 2016년 2월 : 삼성종합기술원 전문 연구원
- 2016년 3월 ~ 2020년 2월 : 건국대학교 전기전자공학부 조교수
- 2020년 3월 ~ 현재 : 건국대학교 전기전자공학부 부교수
- ORCID : <https://orcid.org/0000-0001-5121-5931>
- 주관심분야 : 컴퓨터 비전, 영상처리, 기계학습