Journal of the Korea Academia-Industrial cooperation Society
Vol. 24, No. 12 pp. 360-365, 2023

https://doi.org/10.5762/KAIS.2023.24.12.360
ISSN 1975-4701 / eISSN 2288-4688

스테레오 비전에서의 가변 블록 사이즈를 활용한 심층 학습 기반 깊이 추정 방법 및 비교

김창환, 이덕우*
계명대학교 컴퓨터공학과

Comparison of Depth Estimation Based on Variable Block Size with Deep Learning Model in Stereo Vision

Changhwan Kim, Deokwoo Lee*
Department of Computer Engineering, Keimyung University

요 약 본 논문은 컴퓨터 비전 및 그래픽 분야에서 중요한 주제인 영상 기반 깊이 추정 기법을 분석하는 방법을 제시한다. 특히, 디스패리티 추정을 위한 블록사이즈의 변화를 동반한 심층학습 모델 중 하나인 2DtoDepth 모델에 기반한 스테레오 비전 시스템을 이용한 깊이 추정에 중점을 두고 있으며, 컬러 영상에 비해 비교적 정보가 적은 그레이스케일 영상을 사용한다. Stereo Block Matching, Semi-Global Block Matching, 2DtoDepth 세 가지 기법을 이용하여 깊이를 추정하고 성능을 비교하였다. SBM과 SGBM은 두 개의 이미지를 비교하여 깊이를 추정하는 방식이고, 2DtoDepth는 딥러닝 기반 학습 모델로 깊이를 추정하는 방식으로 단일 이미지로 깊이를 추정하는 방식이다. 결과적으로 본 논문에서는 2DtoDepth 모델로 추출한 깊이맵과 SBM, SGBM 방법으로 추출한 깊이맵을 비교한 결과를 제시한다.

Abstract This paper presents a method for analyzing image-based depth estimation techniques, which is an important topic in the field of computer vision and graphics. In particular, this paper chiefly focuses on depth estimation using a stereo vision system based on a deep learning model with variable block size for disparity estimation. Grayscale images were used for simple processing with less information compared to color images. Depth was estimated using three techniques: stereo block matching (SBM), semi-global block matching (SGBM), and the 2DtoDepth method. The performance of the methods was compared. The SBM and SGBM are methods of estimating depth by comparing two images, and the 2DtoDepth method estimates depth with a deep learning-based trained model with a single image. As a result, this paper presents the results of comparison between the depth map extracted by the 2DtoDepth model and the depth map extracted by the SBM and SGBM methods.

Keywords : Computer Vision, Image Processing, Depth Map, Disparity Map, Depth Estimation

1. 서론

스테레오 비전은 컴퓨터 비전 분야에서 핵심 연구 주제 중 하나로서, 두 가지 혹은 그 이상의 카메라에서 얻은 이미지를 사용하여 실제 세계에서의 깊이 정보를 도출하는 기술이다. 이는 로봇 공학[1], 자율 주행, 3D 모델링[2] 등 다양한 응용 분야에서 중요한 역할을 수행한다. 스테레오 비전에서 가장 핵심적인 알고리즘은 스테

본 결과물은 교육부 및 한국연구재단의 재원으로 첨단분야 혁신융합대학 사업비와 2023년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구 (No.2019R1G1A1100172)임.
*Corresponding Author : Deokwoo Lee(Keimyung Univ.)
email: dwoolee@kmu.ac.kr
Received September 18, 2023 Revised October 11, 2023
Accepted December 8, 2023 Published December 31, 2023

레오 매칭(Stereo Matching)으로, 두 이미지 사이의 유사한 부분을 매칭하여 깊이 정보를 추출한다. 이러한 과정에서 여러 가지 복잡한 문제점들이 발생하는데, 이를 극복하기 위해 여러 가지 스테레오 매칭 알고리즘들이 개발되었다[3].

본 논문에서는 스테레오 블록 매칭(Stereo Block Matching, SBM), 세미 글로벌 블록 매칭(Semi Global Block Matching, SGBM), 그리고 2DtoDepth 모델을 활용하여 깊이 추정을 진행한다. 블록 매칭은 가장 기본적인 스테레오 매칭 알고리즘 중 하나로서, 이미지의 한 블록을 다른 이미지의 같은 위치에 있는 블록과 비교하는 방식을 사용한다. 하지만 이 알고리즘은 텍스처가 없거나 반복적인 패턴, 깊이 변화가 심한 영역에서 문제를 일으키는 등 여러 가지 한계를 가지고 있다[4,5]. 이에 비해 SGBM은 블록 매칭의 한계를 해결하기 위한 방법 중 하나로, 전역 최적화(Global Optimization)와 지역 최적화(Local Optimization)를 절충(Trade-Off)한 방식을 채택한다. 이 방법은 복잡도와 정확도 사이에서 균형을 잡아냄으로써, 블록 매칭이 가진 문제점들을 효과적으로 개선할 수 있다[6].

이러한 스테레오 매칭 알고리즘들에 이어, 2DtoDepth 모델을 통해 깊이 추출을 진행한다. 2DtoDepth 모델은 2D 이미지에서 3D 깊이 정보를 추출하는 심층 학습 기반의 방법으로, 이는 스테레오 비전에 대한 접근법을 제공한다[7]. 심층 학습 기반은 기존 알고리즘 방식에서 발생하는 한계를 넘어서며, 깊이 추정에 있어서 강력한 성능을 발휘한다. 이 모델은 다양한 현실 환경에서 깊이를 예측하고, 이를 통해 3D 환경을 더욱 정확하게 이해하는데 도움이 된다.

본 연구에서는 이 세 가지 기술을 기반으로 한 결과를 상호 비교하고 분석한다. 이런 과정을 통해, 각 기술의 성능과 특징을 분명히 이해하고, 그들 사이의 상호 작용을 비교한다. 이를 통해 각 기술의 장단점을 명확히 제시하며, 특히 어떤 환경에서나 조건에서 특정 기술이 우세할 수 있음을 보여준다. 각 기술들을 어떻게 활용하면 서로 상호 작용할 수 있으며, 각각의 성능과 한계가 어떤 부분에서 달리 하는지에 대해 심층적으로 이해하는 것이다. 해당 결과를 통해 스테레오 비전에 대한 이해를 더욱 확장하고 세밀화 하는데 기여할 수 있을 것이다. SBM과 SGBM, 2DtoDepth 모델을 통합적으로 비교하고 분석함으로써 스테레오 비전 기술에 대한 깊이 정보 추정의 비교 결과를 제공하여, 이에 기반한 최적화 방안을 탐색해 나가는 데에 중점을 둔다.

2장에서는 SBM, SGBM, 2DtoDepth 모델로 추출한 깊이 맵 추출 방법에 대해 설명한다. 3장에서는 깊이 맵 사이의 육안 및 이미지 크기 비교에 대해 설명하고, 4장에서 본 논문의 끝을 맺는다.

2. 본론

SBM은 이미지 정합을 위해 한 쌍의 이미지를 활용하여 깊이 정보를 구하는 알고리즘이다[8]. 좌우 이미지 사이에서 일정한 크기의 윈도우로 픽셀 주변 영역을 비교한다. 이후 이 픽셀들의 유사성을 측정하여, 비용이 최소화되는 위치를 정합점으로 결정한다. 이 과정을 모든 픽셀에 반복 적용하여 깊이 추정 계산이 진행되고, 깊이 맵이 결과물로서 도출된다. 두 영상을 깊이 맵 추출을 위해 활용한 입력 영상의 예시를 Fig. 1에 나타내었다.

(a) (b)

Fig. 1. A pair of images for depth map extraction
(a) Input Image 1 (b) Input Image 2

SBM을 활용하여 생성된 객체는 깊이 맵 생성을 위한 다양한 속성들을 지니고 있으며, 이 속성들의 값에 따라 깊이 맵의 품질이 결정된다. 깊이 맵에 주요하게 영향을 미치는 블록 크기 조정을 통해 깊이 맵의 정확도가 결정되고, 정확도와 처리 시간 사이의 절충이 발생한다. 이는 블록의 크기에 따라 처리를 위한 정보의 양과 그에 따른 정확도 향상 및 비교적 많은 시간이 필요로 하기 때문이다. 정확도는 깊이 맵 생성에 있어서, 좌우 이미지 사이의 유사성에 맞게 깊이 맵을 그릴 수 있는 중요한 요소이다. 예를 들어, 블록의 크기를 19로 설정한 경우와, 11로 설정한 경우에는 Fig. 1의 한 쌍의 영상으로부터 생성되는 깊이 맵의 차이는 Fig. 2와 같다.

SBM으로 생성된 깊이 맵은 잡음이 많고, 불연속적인 경계가 잘 보존되지 않는다는 약점을 지니고 있다. 이러한 SBM의 약점을 개선한 방법이 SGBM이다. SGBM은 SBM의 기존 방식에 기하학적 제약 조건을 추가하여 유사성 평가를 진행한다[9]. 인근 픽셀 간의

관계를 고려하여 불필요한 깊이 값을 제거하며, 경계 부분에서의 정확한 결과 도출을 위해 서브 픽셀 정확도 향상을 통해 결과 영상에 반영하는 과정을 거친다. SGBM을 통해 도출된 깊이 맵은 Fig. 2에서 확인할 수 있다.

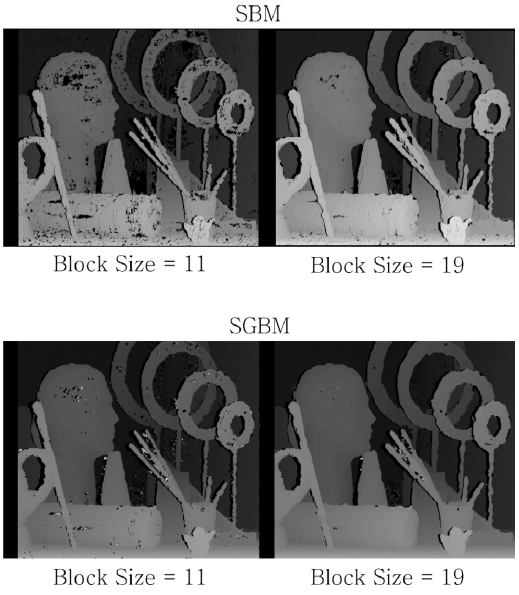

Fig. 2. Depth map difference according to block size

깊이 추출을 위한 전통적인 방식은 SBM과 SGBM처럼 물체를 바라보는 다른 각도의 카메라 두 개로부터 촬영된 영상 두 장을 활용한다. 그러나 최근 심층 신경망 모델을 활용한 심층 학습 기반의 깊이 정보 추정에 대한 연구가 활발히 진행되어 왔다[10]. 그 중 단일 영상만을 활용하여 깊이 정보를 구하는 2DtoDepth 모델을 SBM과 SGBM과 같은 전통적인 방식과 비교해 보고자 한다. 2DtoDepth 모델은 Frozen People 챌린지에서 촬영된 사진들을 데이터셋으로 학습하여, 한 장의 영상만으로도 데이터셋을 통해 학습된 정보를 통해 깊이 맵을 추출할 수 있다. 기존의 깊이 맵들은 가까울수록 밝은 색을 띄었다는 점과 달리, 2DtoDepth 모델을 통해 추출된 깊이 맵은 가까운 물체를 밝게 표현한다. 이러한 모델을 특성을 활용하여, 한 쌍의 영상들 중 첫 번째 영상을 입력 영상으로 사용하였다. 이를 통해 추출한 깊이 맵의 예시는 Fig. 1의 영상을 활용하여 추출하였고, Fig. 3를 통해 확인할 수 있다.

Fig. 3. Depth map extracted through 2DtoDepth

3. 실험

세 가지 방식으로 추출한 깊이 맵 모두 흑백 단차를 활용하여 물체 간의 깊이 차이를 식별할 수 있다. 입력 영상으로 활용되었던 한 쌍의 영상과 추출된 깊이 맵들 사이의 육안적 비교를 진행하여, 각 방식의 깊이 맵 추출 성능에 대한 분석이 가능하다. 방식 별 비교를 Fig. 4에 나타내었다.

SBM으로 추출된 깊이 맵은 입력 영상들과 비교하였을 때에, 입력 영상들 속에 존재하는 사물들을 모두 표현하고 있음을 확인할 수 있다. 마네킹의 머리나 주전자, 원 모양의 장식물들에서 잡음을 확인할 수 있으나, 깊이를 인식함에 있어서 어려움을 발생시키는 요소로 분별되지 않는다. SBM으로 추출한 깊이 맵은 전반적으로 가까운 물체를 보다 더 밝게 표현함을 확인할 수 있다.

SGBM으로 추출된 깊이 맵 또한 SBM 방식과 같은 맥락의 결과를 지니고 있다. 입력 영상에서 존재하는 모든 물체들에 대한 깊이를 제공하고 있으며, 흑백 단차를 활용하여 그 깊이를 관찰할 수 있다. SBM과 다른 특징으로는 물체 위에서 발생한 잡음이 상대적으로 적은 것을 확인할 수 있으며, 가까운 물체 또한 비교적 어둡게 표현함을 확인할 수 있다.

2DtoDepth로 추출된 깊이 맵은 앞선 2가지 방식 대비 흐릿한 물체 표현과 물체의 일부분을 인식하지 못하는 점을 확인할 수 있다. Fig. 4의 Art 영상으로 추출한 깊이 맵들은 원본 영상이 연속적으로 이어지는 깊이를 지니고 있음에도 맵에서는 깊이를 2단에 걸쳐 표현하고 있다. 다른 물체들에 대한 세부적인 깊이 묘사가 이루어지지 않고 있다는 특징 또한 확인할 수 있다. 이는 일반

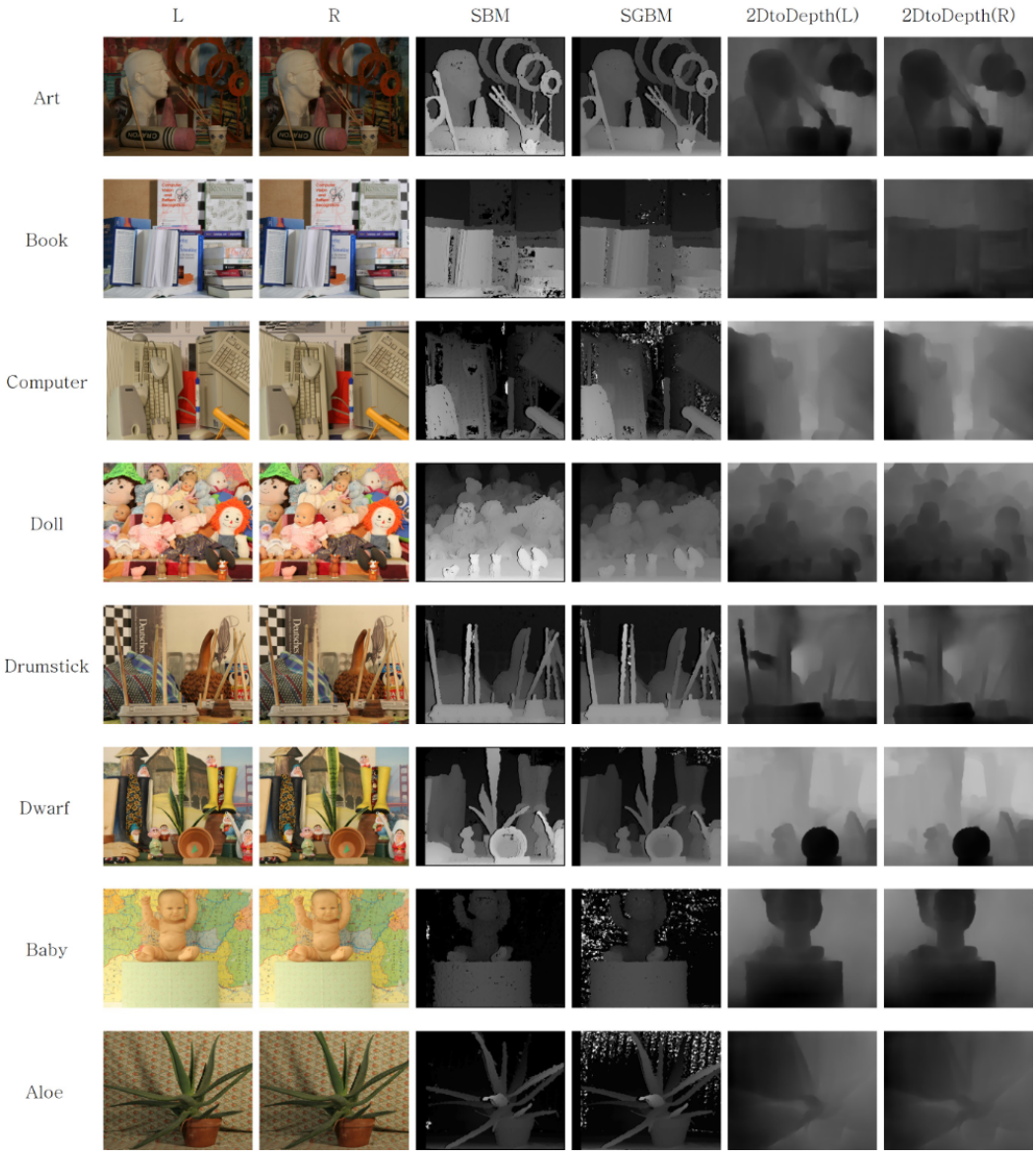

Fig. 4. Image comparison between SBM, SGBM, 2DtoDepth

적으로 2장의 영상을 활용하여 깊이 맵을 추출하는 방식과 달리, 특정 데이터셋을 학습한 모델에 1장의 영상만으로 깊이를 추출하기 때문에 발생하는 특징들이다. 하지만 심층 학습 모델을 활용하면 1장의 영상만으로도 깊이를 추출할 수 있으며, 이는 깊이 맵 추출을 위한 영상 데이터 처리량이 SBM과 SGBM에 비해 급격하게 낮다고 평가할 수 있다. 깊이 맵 추출을 위해 사용된 2장의 영상들의 데이터 크기는 다음과 같다.

Fig. 5의 IMG_1은 1,787,928 바이트이며, IMG_2는 1,793,030 바이트 크기의 데이터를 가지고 있다. 2DtoDepth는 IMG_1 영상 1개를 입력 영상으로 활용하였으며, IMG_2를 깊이 맵 추출에 사용하지 않고 있다. 두 영상의 총합 크기인 3,580,958 byte에 비해 2DtoDepth는 영상 하나의 크기인 1,787,928 byte를 사용하기에, SBM과 SGBM 대비 처리 데이터가 약 50% 가량 감소한다.

Image 1	Image 2

Input Image	Data Size(Byte)
Image 1	1,787,928
Image 2	1,793,030

Fig. 5. Data size comparison of input images

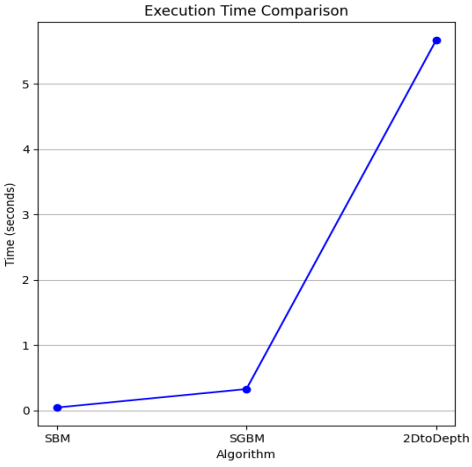

Fig. 6. Comparison results of execution time

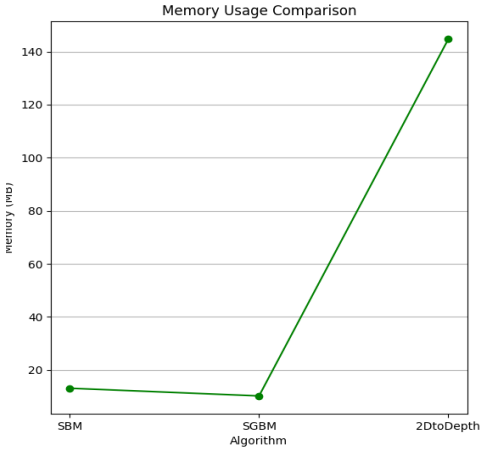

Fig. 7. Comparison results of memory usage comparison

SBM과 SGBM, 2DtoDepth는 깊이 맵 추출에 필요한 시간이 약 0.043초, 0.326초, 5.665초이며, 사용하는 메모리의 양은 약 13.121 MB, 10.227 MB, 144.645 MB를 소요한다. 전통적인 깊이 맵 추출 방식인 SBM과 SGBM에 비해, 2DtoDepth는 추출 시간과 메모리 사용량 두 부분 모두에서 저조한 성능을 보인다. 하지만 깊이 맵 추출을 위해 1장의 영상만을 필요로 한다는 점과, 학습된 데이터셋을 활용하여 깊이 맵을 추출하는 방식이기에 추가적인 훈련을 통해 추출 성능을 향상시킬 수 있다는 특징이 있다. 또한 대략적인 물체 식별 및 이를 통한 깊이 확인이 가능하다.

결과적으로, SGBM은 SBM보다 높은 성능의 깊이 맵을 추출하였으며, 노이즈가 상대적으로 적고 깊이의 식별이 더 정확하게 이루어졌다. 그러나 SBM과 SGBM 모두 두 장의 이미지를 처리해야 하므로 데이터 처리량이 크다는 단점이 있다. 또한, 깊이를 정확하게 계산하려면 두 이미지 간의 정확한 정합성이 요구되며, 이는 비교적 복잡한 전처리 과정을 요구한다.

4. 결론

본 연구를 통해, 전통적인 방법과 모델을 사용한 깊이 맵(depth map) 추출의 성능을 비교하였다. 특히, Stereo Block Matching과 Semi Global Block Matching을 사용한 전통적인 방법과 2DtoDepth라는 심층 학습 모델을 사용한 방법 간의 차이를 살펴보았다. 결과적으로 심층 학습 모델은 일부 상황에서만 기존 방식에 비해 향상된 결과물을 확인할 수 있었다. 이는 심층 학습 모델의 성능이 항상 우수하다는 일반적인 인식과는 다소 다른 결과이다. 이는 대중들의 심층 학습 기반 기술 발전 기대에 대해 중요한 시사점을 제공한다. 심층 학습은 많은 가능성을 제공하지만, 그것이 모든 상황에서 가장 적합한 해결책이라는 것을 보장하지는 않는다. 따라서 그 특성과 한계를 충분히 이해하고 적절한 상황에서만 사용하는 것이 중요하다. 이는 단순 성능만을 기준으로 기술을 선택하는 것이 아니라, 특정 상황과 요구 사항을 고려하여 유연하게 결정되어야 함을 시사한다.

본 논문에서는 심층 학습 기반 방식과 전통적인 방식 사이의 깊이 맵 품질 차이를 확인할 수 있다. 이를 통해 향후 연구에서는 두 가지 방식을 적절히 응용한 깊이 맵 추출 기법을 개발하여, 다양한 상황에서 우수한 품질의 깊이 맵을 추출할 수 있는 강인한 알고리즘을 만들 수 있을 것으로 기대한다.

References

[1] J.-H. Ko, "Robot Vision System based on 3D Depth map and Object Recognition," Journal of the Institute of Electronics and Information Engineers, vol. 57, no. 3, pp. 101-105, Mar. 2020.
DOI: https://doi.org/10.5573/ieie.2020.57.3.101

[2] R. Ait-Jellal and A. Zell, "A fast dense stereo matching algorithm with an application to 3D occupancy mapping using quadrocopters," 2015 International Conference on Advanced Robotics (ICAR), Jul. 2015.
DOI: https://doi.org/10.1109/icar.2015.7251515

[3] J. Hyun, I. Lee, and B. Moon, "A Performance Evaluation of Post-Processing Algorithms for Disparity Refinement of Stereo Vision," The Journal of Korean Institute of Information Technology, vol. 15, no. 2, pp. 99-107, Feb. 2017.
DOI: https://doi.org/10.14801/jkiit.2017.15.2.99

[4] J. Kim, C. Park, J. O Kim, and D. Lee, "OCCLUSION HANDLED BLOCK-BASED STEREO MATCHING WITH IMAGE SEGMENTATION," Computer Science Information Technology (CSIT), Mar. 2019.
DOI: https://doi.org/10.5121/csit.2019.90303

[5] B.-W. Cheon and N.-H. Kim, "Digital Filter Algorithm based on Complex Weight and Pixel Matching in AWGN Environment," 2021 IEEE 16th Conference on Industrial Electronics and Applications (ICIEA), Aug. 2021.
DOI: https://doi.org/10.1109/iciea51954.2021.9516313

[6] W.-J. Jeong, M. Lee, and S.-Y. Park, "Iterative Improvement of Stereo Disparity Image through Minimum Cost Voting of SGM Aggregation Matching Cost Volume," Journal of the Institute of Electronics and Information Engineers, vol. 58, no. 7, pp. 63-74, Jul. 2021.
DOI: https://doi.org/10.5573/ieie.2021.58.7.63

[7] Z. Li et al., "Learning the Depths of Moving People by Watching Frozen People," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2019.
DOI: https://doi.org/10.1109/cvpr.2019.00465

[8] J.-S. Kim, C.-H. Park, D.-W. Lee, "Block-Based Stereo Matching Using Image Segmentation," The Journal of Korean Institute of Communications and Information Sciences, 44(7), 1402-1410, Jul. 2019.
DOI: https://doi.org/10.7840/kics.2019.44.7.1402

[9] Xiang, L., Tang, L., Gai, J., and Wang, L, "Measuring stem diameter of sorghum plants in the field using a high-throughput stereo vision system," Transactions of the ASABE, 64(6), 2021.
DOI: https://doi.org/10.13031/trans.14156

[10] Kim, H., Lim, H., Jee, M., Lee, Y., Yoon, M., & Kim, C, "High-Precision Depth Map Estimation from Missing Viewpoints for 360-Degree Digital Holography," Applied Sciences, 12(19), 9432. Sep. 2022.
DOI: https://doi.org/10.3390/app12199432

김 창 환(Chang-Hwan Kim) [준회원]

- 2018년 3월 ~ 현재 : 계명대학교 공과대학 컴퓨터공학과 학사과정

〈관심분야〉
컴퓨터비전, 영상처리, 신호처리

이 덕 우(Deokwoo Lee) [정회원]

- 2007년 2월 : 경북대학교 전자전기컴퓨터학부 졸업 (공학사)
- 2012년 12월 : North Carolina State University, NC, USA, 전기컴퓨터공학부 석사, 박사
- 2013년 1월 ~ 4월 : Washington Univ. in St. Louis, USA, 박사후 연구원
- 2013년 9월 ~ 2016년 8월 : 삼성전자 무선사업부 책임 연구원
- 2016년 9월 ~ 2018년 2월 : 영산대학교 가상현실콘텐츠 전공 조교수
- 2018년 3월 ~ 현재 : 계명대학교 공과대학 컴퓨터공학과 조교수

〈관심분야〉
컴퓨터비전, 영상처리, 신호처리