논문 2024-19-23

깊이 추정 및 커널 필터링 기반 Visual-Inertial Odometry

(Visual-Inertial Odometry Based on Depth Estimation and Kernel Filtering Strategy)

송 지 민, 조 형 기, 이 상 준*
(Jimin Song, HyungGi Jo, Sang Jun Lee)

Abstract : Visual-inertial odometry (VIO) is a method that leverages sensor data from a camera and an inertial measurement unit (IMU) for state estimation. Whereas conventional VIO has limited capability to estimate scale of translation, the performance of recent approaches has been improved by utilizing depth maps obtained from RGB-D camera, especially in indoor environments. However, the depth map obtained from the RGB-D camera tends to rapidly lose accuracy as the distance increases, and therefore, it is required to develop alternative method to improve the VIO performance in wide environments. In this paper, we argue that leveraging depth map estimated from a deep neural network has benefits to state estimation. To improve the reliability of depth information utilized in VIO algorithm, we propose a kernel-based sampling strategy to filter out depth values with low confidence. The proposed method aims to improve the robustness and accuracy of VIO algorithms by selectively utilizing reliable values of estimated depth maps. Experiments were conducted on real-world custom dataset acquired from underground parking lot environments. Experimental results demonstrate that the proposed method is effective to improve the performance of VIO, exhibiting potential for the use of depth estimation network for state estimation.

Keywords : Autonomous driving, Deep learning, Visual-inertial odometry, Depth estimation, Filtering technique

Ⅰ. 서 론

로봇공학 분야의 주목할 만한 발전으로 로봇들이 다양한 실제 산업 시나리오에 점점 더 많이 적용되고 있다 [1, 2]. 예를 들어, 매니퓰레이터 로봇은 현대 사회의 공장 환경에서의 간단하고 반복적인 작업에 많이 적용되고 있다 [3]. 이와 같은 적용은 산업 효율성을 크게 향상시키는 데 상당한 기여를 하였으며, 이제 로봇 공학 분야 공동체는 다음 단계를 목표로 하고 있다. 최근 몇 년간, 모바일 매니퓰레이터 시스템 [4]과 자율주행 차량의 첨단 운전자 보조 시스템 (ADAS) [5]과 같은 차세대 응용 분야의 출현으로 해결해야 할 새로운 문제들이 제기되고 있다. 매니퓰레이터의 정교한 작업, 모바일 로봇의 자율 이동 그리고 차량의 장애물 회피 등은 기능 구현을 위해서 수많은 요소 기술들이 필요하다. 복잡한 시스템일수록 작은 오류가 큰 인명사고나 재산상의 피해를 입힐 수 있기 때문에 주변 환경과 시스템 자체 상태에 대한 더 정확한 인식 알고리즘을 필요로 한다.

사전에 알고 있는 환경에서 모바일 로봇의 위치를 추정하는 기술은 환경에 적합한 센서들로 시스템을 구성하여 구현할 수 있다. 미리 지정된 활동 영역에서 운영되는 실내 로봇 시스템은 위치 추정 및 보정을 위해 시각적 표지 [6]나 무선 주파수 식별 (RFID) 표지 [5]와 같은 외부적인 요소를 활용한다. 도로를 주행하는 차량과 같이 상대적으로 위치를 추정할 수 있는 물체가 없는 실외 시스템은 주로 전역 위치 시스템 (GPS) 기술을 사용하여 위치를 추정한다. 그러나 이러한 방법은 운영 환경에 위치 표지를 설치하거나 외부 시스템과 통신이 원활해야 하므로, 우주 탐사 [7]나 구조 작전 [8]과 같은 분야에서는 적용하기 어려울 수 있다. 이에 따라서 외부 요소 없이 모바일 로봇이 주변 환경을 자율적으로 인식하고 자신의 위치를 추정할 수 있는 동시적 위치 추정 및 맵핑 (SLAM) 기술의 수요가 지속해서 증가하고 있다. 본 논문에서는 일반적인 환경에서 카메라 기반 SLAM의 정확도를 향상시키기 위해서 딥뉴럴 네트워크를 통해 영상에서 화소 수준의 깊이 정보를 추정하는 것이 중요하다고 주장한다. 또한 추정한 깊이 맵에서 정확도가 낮을 수 있는 영역을 필터링하여 SLAM에 활용하는 방법을 제안한다.

최근 수년간 컴퓨터 비전 분야의 딥러닝 기술은 객체 감지, 객체 추적 및 의미적 분할과 같은 고전적인 작업뿐만 아니라 깊이 추정과 같은 보다 복잡한 작업에서도 많은 발전이 있었다. 깊이 추정 문제는 2차원 RGB 영상으로부터 3차원 정보인 화소 수준의 거리를 추정하는 어려운 역투영 문제이다. 하지만 높은 정확도로 거리를 측정할 수 있는 라이다 (LiDAR) 센서로 취득한 데이터를 활용하면 딥뉴럴 네트워크의 학습이 수월해질 수 있다. LiDAR 센서 데이터로부터 생성된 깊이 정보에 대한 ground truth를 학습에 사용

*Corresponding Author (sj.lee@jbnu.ac.kr)
Received: May 21, 2024, Revised: Jun. 12, 2024, Accepted: Jul. 1, 2024.
J. M. Song: Jeonbuk National University (Ph.D. Student)
H. G. Jo: Jeonbuk National University (Assist. Prof.)
S. J. Lee: Jeonbuk National University (Assist. Prof.)
※ 이 성과는 정부 (과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. RS-2024-00346415).

ⓒ IEMEK J. Embed. Sys. Appl. 2024 Aug. 19(4) 185-193
ISSN : 1975-5066
http://dx.doi.org/10.14372/IEMEK.2024.19.4.185

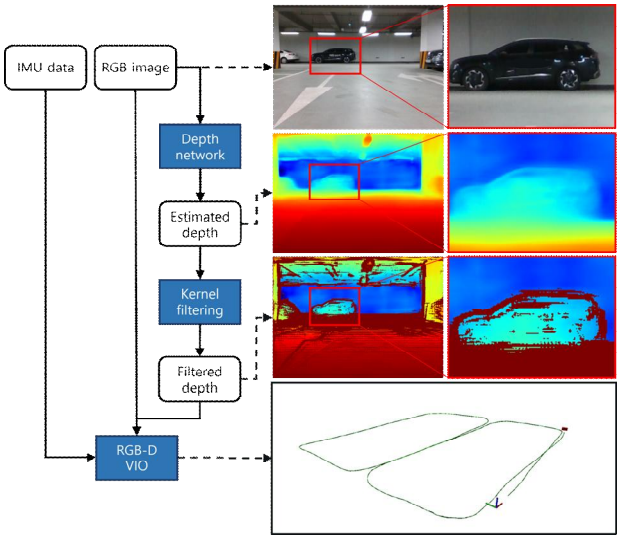

그림 1. 제안하는 VIO 파이프라인
Fig. 1. Overview of the proposed VIO pipeline

하는 지도 학습 방법은 학습 데이터와 유사한 환경에 대해서 추정된 깊이 정보의 정확도가 높다는 장점이 있다. 지속적인 연구를 통해서 네트워크를 통해 추정된 깊이 맵의 정확도가 향상되고 이에 따라 카메라 기반 SLAM 기술의 성능을 향상시키기 위해서 추정된 깊이 맵을 활용하는 연구도 증가되고 있다. 본 논문에서는 그림 1과 같은 새로운 RGB 기반 VIO 파이프라인을 제안한다. 제안하는 파이프라인은 깊이 추정이 어려운 물체 경계나 모바일 로봇이 배치된 평면에 평행한 평면에서의 추정 결과를 필터링하기 위한 커널 기반의 필터링 전략을 포함한다. 또한 이전 연구들에서 주로 활용된 사무실과 같은 한정된 실내 환경에 비해 확장된 환경인 지하 주차장에서 직접 수집한 실제 데이터셋을 사용하여 제안하는 방법의 효용성을 실증했다.

본 논문에서 제안하는 방법의 기여점을 정리하면 다음과 같다. 첫 번째는 직관적인 커널 기반의 필터링을 통해서 추가적인 학습 과정 없이 사전에 학습된 깊이 추정 네트워크의 성능을 향상시킬 수 있음을 보였다. 두 번째로는 실외 환경에서는 기존의 RGB-D 카메라의 깊이 맵을 활용하는 것 보다 잘 학습된 깊이 추정 네트워크의 결과를 활용하는 것이 VIO에 유리하다는 것을 실험적으로 보였다. 마지막으로 직접 취득한 데이터를 활용하여 제안하는 방법이 기존 방법에 비해서 실외 환경에서 VIO의 성능 향상에 효과적임을 정성적 및 정량적으로 보였다.

Ⅱ. 관련 연구

일반적으로 주행기록계 추정을 위해 관성 측정 장치(IMU)가 사용되며, 가속도 및 각가속도 데이터의 적분을 통해 이동 및 회전을 추정한다. 정확한 가속도 및 각가속도 데이터가 입력되면, 실제 값에 근사한 주행기록계를 추정할 수 있다. 그러나 실제 IMU의 측정 데이터에는 다양한 요인들로부터 발생하는 노이즈가 포함된다. 따라서 최근 제안되는 SLAM 알고리즘들은 IMU와 함께 추가적인 센서를 활용한다. 이에 따라서 SLAM 접근 방법은 추가적인 센서의 종류에 따라 LiDAR-inertial odometry (LIO)와 visual-inertial odometry (VIO)로 분류된다. GPS 신호가 사용 불가능한 실내 환경에서 LIO 알고리즘 [9]는 VIO 알고리즘의 성능을 평가하는 지표로 사용될 수 있을 만큼 정확도가 높다. 하지만 환경에 따른 위치 추정 정확도의 잠재적 변동성에도 불구하고, 시스템의 비용적인 이점을 감안하면 VIO 접근법의 성능을 개선하기 위해 지속적인 연구가 필요하다. VIO 알고리즘은 일반적으로 카메라와 IMU의 주행기록계 추정 과정을 독립적으로 사용하는 loosely-coupled sensor fusion 방법 [10, 11] 또는 두 센서의 추정값을 동시에 최적화하는 tightly-coupled sensor fusion 방법 [12, 13]으로 나누어 진다. 본 연구에서는 RGB-D 영상을 통합하는 tightly-coupled sensor fusion VIO 알고리즘 [14]를 기반으로 한 접근 방식을 제안한다. 우리의 목표는 깊이 추정 네트워크의 추정 결과를 효과적으로 활용하여 확장된 환경에서 VIO 알고리즘의 성능을 향상시키는 것이다.

단일 카메라 영상에서 거리를 추정하는 단안 깊이 추정은 스케일 모호성으로 인해 해결하기 어려운 문제이다. Eigen et al. [15]는 영상 내에서 거리 정보의 상대적인 차이를 고려할 수 있는 스케일 불변 손실함수를 제안했다. Lee et al. [16]은 DenseNet [17]의 인코더에서 깊이 정보를 효율적으로 추정하기 위해 설계된 디코더 모델을 소개했다. 이 접근 방식의 주목할 만한 측면 중 하나는 영상의 기하학적 특성을 포착하고 네트워크의 디코더 내에서 깊이 관련 정보를 계산하는 local planar guidance 기법이다. 우리는 깊이 추정 네트워크의 성능 향상시킬 수 있는 이 기술을 본 연구에서 활용했다. 깊이 추정 문제의 본질적인 어려움에도 불구하고, 다양한 응용 분야에서 유용성을 인정받아 적용되고 있다. Wang et al. [18]은 깊이 추정을 활용하여 의미적 세분화 성능을 향상시키는 학습 프레임워크를 제안했다. 또한 Dai et al. [19]는 깊이 추정 네트워크와 동적 물체 이동 추정을 통합하여 프레임 간 화소의 이동을 추정하는 optical flow 추정 성능을 향상시켰다. Wang et al. [20]은 깊이 맵을 3차원 포인트 클라우드 형태의 데이터로 변환하고, 이를 통해 3D 객체 감지 성능을 향상시켰다. 본 논문에서는 깊이 추정 네트워크의 추론 결과를 VIO 알고리즘에 활용하는 것이 추정 정확도를 향상시킬 수 있으며 특히 실내 환경보다 더 넓은 일반적인 공간에서 이점이 있다고 주장한다.

최근 몇 년간 RGB-D 카메라는 실내 환경에서 높은 성능을 보여주며 VIO 알고리즘에 활용되고 있다. 그러나 RGB와 RGB-D 카메라 간의 적용 범위 제한 및 비용 격차로 인해, 깊이 추정 네트워크를 활용하는 VIO 알고리즘에 대한 연구도 지속되고 있다. 깊이 추정 네트워크는 일반적으로 RGB 영상에서 상대적인 깊이를 추정하는 데 뛰어나지만, Wofk et al. [21]은 절대적인 깊이에 대한 스케일 및 시프트 정렬의 중요성을 강조하고 정확한 VIO를 위해 절대적인 깊이 정보를 추정하기 위한 새로운 프레임워크를 소개했다.

Merrill et al. [22]는 VIO 알고리즘의 성능을 향상시키고 임베디드 시스템에서도 실시간 처리가 가능한 새로운 깊이 보충 방법을 소개했다. Almalioglu et al. [23]은 visual odometry, inertial odometry 및 깊이 추정을 위한 개별 네트워크를 동시에 학습하기 위한 적대적 프레임워크를 제안했다. Zuo et al. [24]는 VIO 프로세스 중 삼각측량을 통해 누적된 희박한 깊이 정보를 활용하여 깊이 네트워크의 추론 결과를 최적화하기 위한 방법을 제안했다. Sartipi et al. [25]는 삼각측량을 통해 생성된 희박한 깊이 정보를 보완하기 위해서 평면 분할 및 표면 법선 분할 네트워크를 활용하는 방법을 제안했다. 더 정확한 깊이 정보를 추정하는 것은 필수적이지만 본 논문에서는 데이터 혹은 네트워크 특성상 추정된 깊이 맵에서 부정확할 수 있는 영역을 필터링하기 위한 직관적인 필터링 방법을 제안하고 성능이 향상 될 수 있음을 실험적으로 보였다.

Ⅲ. 본 론

1. RGB 영상 기반 VIO 파이프라인

본 연구에서는 RGB-D 카메라를 활용하는 VIO 알고리즘인 VINS-RGBD [14]를 기반으로 RGB 기반 VIO 파이프라인을 구축했다. 카메라 영상으로부터 특징점 추출 및 추적과 IMU 데이터를 활용하여 두 개의 카메라 영상 사이의 위치 추정값을 보완 및 제한하는 pre-integration이 포함된 measurement processing 과정이 가장 먼저 수행된다. 그리고 카메라 영상과 IMU를 개별적으로 활용한 주행기록계 추정 결과를 통합하는 초기화 과정이 수행된다. 초기화 과정이 완료되면 그 이후에는 measurement processing 후에 local VIO 과정으로 이어진다. Local VIO에서는 일정 시간 동안의 입력만 활용하여 주행기록계를 추정한다. 그리고 특징점의 상관관계에 따라서 loop detection이 이루어지면 pose graph를 최적화하는 backend 과정이 수행되며 주행기록계 추정 결과의 정확도를 향상시킨다. 해당 과정에서 선행 연구와의 가장 큰 차이는 특징점의 거리 정보를 깊이 추정 네트워크의 추론 결과에 제안하는 필터링 기법을 적용한 깊이 맵에서 활용한다는 점이다.

2. 깊이 추정 네트워크

본 논문에서는 지도 학습 깊이 추정 네트워크 [16]을 활용하여 VIO의 성능을 향상시켰다. 그림 1에서 볼 수 있듯이, 깊이 추정 네트워크는 $H\times W$ 해상도의 RGB 영상을 입력으로 받고 동일한 해상도의 깊이 맵을 예측한다. 네트워크 아키텍처는 인코더-디코더 구조를 따른다. 인코더 백본으로는 공간 해상도가 $\frac{H}{8}\times\frac{W}{8}$인 특징맵을 추출하기 위해 DenseNet-161 [17]을 사용한다. 또한, 특징 추출을 위해 receptive field를 넓히기 위해 atrous spatial pyramid pooling 모듈 [26]을 활용한다. 이후, 추출된 특징맵은 local plan guidance layer를 활용하여 원래의 공간 해상도

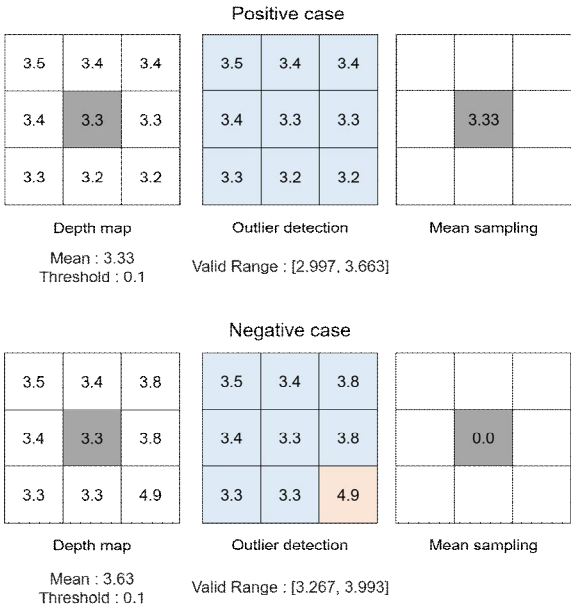

그림 2. 커널 기반의 필터링 예시
Fig. 2. Examples of kernel-based filtering

$H\times W$로 업샘플링된다. 각 디코더 블록의 출력은 채널 축을 따라 연결되고, 최종 깊이 추정을 위해 convolution layer에 입력된다. 이에 대한 출력에 데이터셋에 따라서 사전에 정의한 max depth를 곱하여 최종 깊이 추정 맵을 추정한다.

3. 커널 기반 필터링 기법

본 논문에서는 환경에 의한 변수나 딥뉴럴 네트워크 성능의 제약으로 인해 부정확하게 추정된 깊이 값을 제거하기 위해 그림 2와 같은 커널 기반 필터링 기법을 활용하는 방법을 제안한다. 제안하는 커널 필터링은 K로 표현되는 커널 크기와 T로 표현되는 이상치 탐지를 위한 임계값인 두 가지 휴리스틱 매개 변수의 활용이 필요하다. 이러한 매개 변수들은 깊이 추정 결과 중에 이상치를 식별하고 제거하는 데 중요한 역할을 한다. 본 연구에서는 각각 K를 3으로, T를 0.01로 설정하여 추정된 깊이 맵의 일부를 필터링했다. 커널 크기 K는 기준 화소 (i,j)에 대한 평균값 $\mu_{i,j}$을 계산할 때 활용된다.

$$\mu_{i,j}=\frac{1}{K^2}\sum_{x}\sum_{y}d_{x,y}, \tag{1}$$

여기에서 $d_{x,y}$은 화소 (x,y)에 대한 깊이 추정값을 의미한다. 기준 화소 (i,j)와 인접한 화소의 x와 y는 각각 $[i-\frac{K-1}{2},i+\frac{K-1}{2}]$와 $[j-\frac{K-1}{2},j+\frac{K-1}{2}]$의 수평 및 수직의 범위에 포함된다. 이상치 탐지를 위한 임계값 T를 활용한 필터링된 깊이 추정값 $\tilde{d}_{i,j}$은 다음과 같이 표현될 수 있다.

$$\tilde{d}_{i,j}=\begin{cases}\mu_{i,j} & \text{if } d_{x,y}\in[\mu_{i,j}(1-T),\mu_{i,j}(1+T)],\\ 0 & otherwise.\end{cases} \tag{2}$$

필터링된 깊이 추정값 $\tilde{d}_{i,j}$은 일정 수의 키프레임들에 대해서 최적화 기반으로 VIO를 추정하는 local VIO 과정에서 활용된다. 이 과정에서 카메라 영상으로부터 추출된 특징점의 깊이 값으로 활용되며 서로 다른 프레임에서 매칭된 특징점의 깊이 값들이 유사한 지 판단하는 깊이 검증 [27] 이후에 상수값으로 고정된다. 깊이 검증 또는 커널 필터링으로 깊이 값이 제거된 특징점의 경우 깊이 값을 삼각측량 [28]을 통해 추정하고 최적화를 통해서 변경될 수 있는 변수로 설정된다. 기존 방법 [14]의 경우 활용하는 RGB-D 카메라에 따라서 깊이 값의 유효 범위가 달라지고 약 20 m이내이지만, 제안하는 파이프라인에서는 깊이 추정 네트워크의 성능에 따라서 더 넓은 공간에서도 활용할 수 있다.

Ⅳ. 실 험

1. 실환경 데이터셋

본 논문의 실험에서는 지하 주차장 환경에 중점을 두었다. 데이터셋 취득에 활용된 환경의 바닥은 천장의 조명이 반사되는 재질이고, 피사체는 주로 차량 및 오토바이 또는 사람이 있다. 이러한 특징은 실제 시나리오에 포함되어야 하는 필수적인 요소들이다. 지하 주차장 환경은 기존의 RGB-D 카메라를 활용해도 양질의 깊이 맵을 얻을 수 있는 사무실과 같은 전형적인 실내 환경에 비해 더 큰 공간을 제공한다. 또한 자동 주차 및 실내 배달과 같은 실외 도로 주행 시나리오와는 다른 응용기술에 활용할 여지가 있다. 데이터 수집을 위해 그림 3과 같은 모바일 로봇 구성을 활용했다. 센서 데이터 수집 및 jackal unmanned ground vehicle (UGV) 제어를 위해서 jetson AGX Xavier를 사용했다. 센서 구성에는 수직 및 수평 시야각이 각각 90° 와 360° 인 ouster OS-0 32채널의 LiDAR가 포함되어 있다. 또한, 우리는 RGB-D 카메라, 특히 수직 및 수평 시야각이 각각 42° 와 69° 인 realsense D435i를 모바일 로봇의 전면과 후면에 배치했다. 깊이 추정 및 VIO를 위해서 640×480 픽셀의 RGB-D 영상을 30Hz의 프레임 속도로 독립적으로 취득했다.

깊이 추정 네트워크의 지도 학습 및 성능 평가를 위해 LiDAR와 카메라 간의 오프라인 캘리브레이션을 통해 얻은

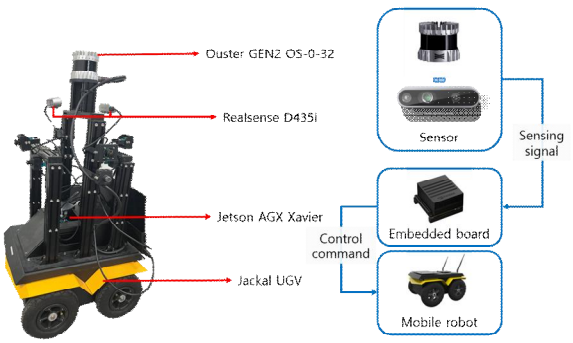

그림 3. 데이터 취득을 위한 모바일 로봇 구성
Fig. 3. Configuration of mobile robot for acquiring dataset

외부 파라미터를 활용하여 ground truth 깊이 맵을 생성했다. VIO의 평가를 위해 각각 225 m, 225 m, 122 m, 44 m, 44 m 및 26 m의 거리를 포함하는 6가지 별도의 주행 시나리오에서 데이터를 수집했다. 지하 주차장 환경에서는 VIO에 대한 ground truth 주행기록계는 LiDAR 기반의 SLAM 기술인 Faster-LIO [9]를 활용하여 생성했다. 이러한 데이터셋은 다양한 환경 조건 및 주행 시나리오에서 우리의 제안된 방법의 성능을 평가하기 위한 포괄적인 기반을 제공한다.

2. 실험 구성

깊이 추정 및 VIO 실험은 AMD EPYC 7313P 16코어, 64GB DDR4 RAM, 그리고 NVIDIA GeForce RTX 4090을 포함한 하드웨어 환경에서 진행되었다. 우리는 깊이 추정을 위해서 기존의 지도 학습 기반 네트워크와 학습 파이프라인을 활용했다 [16]. 우리는 NYUv2에서 사전 학습된 깊이 추정 모델의 매개변수를 0.00001의 학습률로 500쌍의 RGB 영상과 LiDAR 기반의 ground truth 깊이 맵을 사용하여 미세 조정을 했다. 안정적인 VIO를 위해, 오프라인 캘리브레이션을 통해 카메라와 IMU 사이의 외부 파라미터를 얻었으며, VIO 프로세스 중에 외부 파라미터에 대한 추가적인 최적화는 수행되지 않았다. 640×480 해상도의 영상에 대해, 모든 VIO 실험에서 추적하는 특징점의 최대 수와 두 특징점 사이의 최소 거리는 각각 180 개과 20 화소로 고정되었다. 또 다른 VIO에 대한 휴리스틱 파라미터인 keyframe parallax와 ceres solver [29]의 최대 시간 및 선형 최소 제곱 문제를 해결하는 최대 반복 횟수는 각각 10 화소, 0.04 ms 및 8 회로 고정되었다.

3. 평가 지표

본 논문에서는 깊이 추정의 정량적 평가를 위해 5개의 오차 평가 지표와 3개의 정확도 평가 지표를 사용한다. 오차 평가 지표에는 절대 상대 오차 $AbsRel$, 제곱 상대 오차 $SqRel$, 평균 제곱근 오차 $RMSE$, 로그 제곱근 평균 오차 $RMSElog$ 및 스케일 불변 로그 오차 $SIlog$가 포함된다. 정확도 평가 지표에는 $\max(\hat{d}/d, d/\hat{d})$로 계산된 상대 오차에 대해서 threshold $\delta \in [1.25, 1.25^2, 12.5^3]$ 이내에 포함되는 픽셀 비율로 계산되며, 여기서 $\hat{d}$와 d는 네트워크에 의해서 예측된 깊이 값과 ground truth 깊이 값을 의미한다. 오차 평가 지표와 정확도 평가 지표는 각각 낮은 값과 높은 값에서 더 나은 성능을 나타내며, 더 자세한 식은 이전 연구 [15]에서 확인할 수 있다. VIO 성능을 평가하기 위해 상대 위치 오차 (RPE)의 이동 및 회전 오차와 절대 궤적 오차 (ATE)의 RMSE를 포함한 3 개의 평가 지표를 활용했다 [30].

4. 깊이 추정 성능 분석

표 1은 유효한 깊이 범위에 따른 성능 변화를 보여준다. Realsense D435i가 제공하는 깊이 맵은 제조사에서 권장하는 거리 범위인 0.3 m에서 3.0 m에서 깊이 추정 네트워크의 추론 결과 및 제안하는 필터링 기법을 적용한 깊이 맵과 정

표 1. 깊이 맵에 대한 정량적 성능 비교
Table 1. Comparison of quantitative performance of depth maps

Depth range	Method	Ratio	Error metric ↓					Accuracy metric ↑		
			AbsRel	*SqRel*	*RMSE*	*RMSElog*	*SIlog*	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
3.0 m	Realsense D435i	0.200	0.030	**0.003**	0.086	0.035	**0.206**	**0.999**	**1.000**	**1.000**
	Depth network	0.220	0.026	**0.003**	0.072	0.031	0.240	**0.999**	**1.000**	**1.000**
	Ours	0.137	**0.025**	**0.003**	**0.070**	**0.030**	0.225	**0.999**	**1.000**	**1.000**
10.0 m	Realsense D435i	0.653	0.107	0.197	0.977	0.157	1.491	0.897	0.972	0.988
	Depth network	0.651	0.050	0.043	0.517	0.075	0.709	0.978	**0.997**	**0.999**
	Ours	0.354	**0.047**	**0.038**	**0.481**	**0.069**	**0.654**	**0.982**	**0.997**	**0.999**
20.0 m	Realsense D435i	0.917	0.140	0.453	1.797	0.199	1.937	0.839	0.952	0.980
	Depth network	0.916	0.063	0.099	0.982	0.097	0.946	0.961	0.994	**0.999**
	Ours	0.512	**0.056**	**0.068**	**0.825**	**0.082**	**0.799**	**0.975**	**0.997**	**0.999**
35.0 m	Realsense D435i	0.990	0.144	0.557	2.324	0.206	2.027	0.825	0.949	0.980
	Depth network	1.000	0.069	0.172	1.442	0.109	1.073	0.954	0.991	0.997
	Ours	0.559	**0.059**	**0.106**	**1.149**	**0.088**	**0.860**	**0.972**	**0.996**	**0.999**

표 2. 커널 필터링 기법의 영향 분석을 위한 절제 연구
Table 2. The ablation study for analyzing the effect of the kernel filtering strategy

Kernel	Threshold	Case 1	Case 2	Case 3	Average
RMSE of ATE [m]					
Without filtering		2.2144	1.0160	0.8355	1.4488
3×3	0.1	2.1584	1.0592	0.8135	1.4391
	0.05	2.1552	1.0131	0.8023	1.4173
	0.01	2.2043	0.8987	0.7603	1.3827
5×5	0.1	2.1792	1.0047	0.8623	1.4363
	0.05	2.1516	0.9245	**0.7616**	1.3724
	0.01	2.4732	1.3102	1.3546	1.7771
7×7	0.1	2.1819	1.0195	0.8349	1.4373
	0.05	**2.0931**	**0.8963**	0.8288	**1.3526**
	0.01	2.7641	1.4077	1.2241	1.9020
Translation error of RPE [m]					
Without filtering		0.0418	0.0349	0.0345	0.0375
3×3	0.1	0.0410	0.0371	**0.0328**	0.0377
	0.05	0.0416	0.0362	0.0334	0.0377
	0.01	0.0421	**0.0339**	0.0349	**0.0373**
5×5	0.1	0.0407	0.0358	0.0349	0.0375
	0.05	0.0410	0.0365	0.0340	0.0377
	0.01	0.0516	0.0456	0.0488	0.0486
7×7	0.1	0.0414	0.0366	0.0333	0.0378
	0.05	**0.0401**	0.0426	0.0344	0.0399
	0.01	0.0660	0.0532	0.0530	0.0582
Rotation error of RPE [°]					
Without filtering		0.1408	0.1680	0.1900	0.1620
3×3	0.1	0.1404	0.1727	0.1788	0.1613
	0.05	0.1411	0.1728	0.1861	0.1632
	0.01	0.1376	0.1503	0.1841	0.1525
5×5	0.1	0.1401	0.1709	0.1935	0.1636
	0.05	0.1391	0.1754	0.1860	0.1634
	0.01	**0.1370**	0.1465	0.1819	**0.1503**
7×7	0.1	0.1400	0.1657	0.1929	0.1614
	0.05	0.1388	0.1684	0.1826	0.1598
	0.01	0.1415	**0.1443**	**0.1787**	0.1505

량적 평가 지표에서 유사한 성능을 보인다. 그러나 성능 평가 범위가 이상적인 범위를 벗어나면 깊이 추정 네트워크의 추론 결과에 비해서 모든 측정 항목에서 유의한 성능 저하가 관찰된다. 이는 RGB-D 카메라의 이상적인 범위를 넘어가는 깊이 정보를 VIO에 사용하는 것은 적합하지 않다는 것을 의미한다. 이에 따라서 이상적인 범위 내에 특징이 많지 않은 개방된 공간에서 VIO는 RGB-D 카메라를 활용하는 것 보다 깊이 추정 네트워크를 활용하는 것이 유리할 수 있다고 할 수 있다.

우리가 제안하는 필터링 기법을 깊이 네트워크에서 생성된 추정된 깊이 지도에 적용하는 경우, 모든 측정 항목에서 필터링 기법을 적용하지 않은 경우보다 우수한 성능을 보인다. 깊이 추정의 정량적인 성능 계산에는 ground truth 깊이 맵에 존재하는 화소에 대해서만 계산하며, 필터링을 적용하게 될 경우 성능 평가에서 제외되어 결과적으로 깊이추정값의 오차가 큰 영역에 대해서 필터링 할 수 있음을 의미한다. 깊이 맵에서 유효한 깊이 값의 비율이 감소하더라도, 신뢰할 수 있는 깊이 값의 사용은 VIO에서 정확도를 향상시킬 수 있다. 이는 표 2에서 제시된 실험 결과에서 확인할 수 있다. 그림 4에서 RGB-D 카메라의 깊이 맵은 특히 붉은 색으로 표현된 먼 거리에서 추정이 불가능하며 일부가 추정되더라도 일관성이 부족하다. 네트워크를 통해서 추정된 깊이 맵은 전체적으로 입력 영상에 대응되는 합리적인 추정 결과를 보인다. 또한 이에 우리가 제안하는 필터링을 적용한 결과는 가장 우측 열에서 볼 수 있으며, 오차가 클 수 있는 객체의 경계선 및 먼 거리에 있는 지면 등을 성공적으로 제거한 것을 확인할 수 있다.

5. VIO 성능 분석

기존의 VIO 알고리즘인 VINS-Mono [27]와 VINS-RGBD [14], 그리고 우리가 제안하는 RGB VIO 접근 방식 간의 정량적 및 정성적 성능 비교 분석을 수행했다. case 1, 2, 3

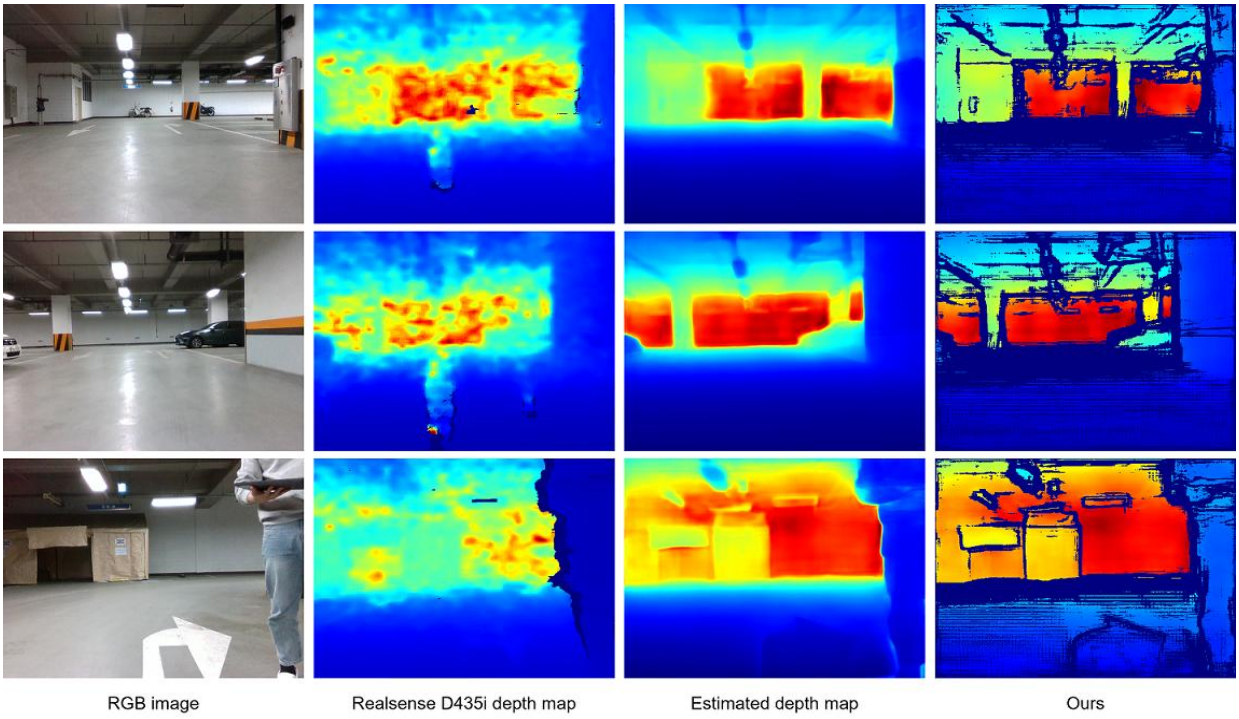

그림 4. 깊이 맵에 대한 정성적 성능 비교
Fig. 4. Comparison of the quality of depth maps

표 3. 기존 VIO 알고리즘과의 정량적 성능 비교
Table 3. Quantitative comparison with existing VIO algorithms

Method	Case 1	Case 2	Case 3	Average
RMSE of ATE [m]				
VINS-Mono [27]	7.6614	2.0864	2.7397	4.4185
VINS-RGBD [14]	5.8164	7.8056	0.9733	5.5654
Ours	**2.2043**	**0.8987**	**0.7603**	**1.3827**
Translation error of RPE [m]				
VINS-Mono [27]	0.1518	0.0539	0.0563	0.0929
VINS-RGBD [14]	0.1217	0.5044	0.0753	0.2623
Ours	**0.0421**	**0.0339**	**0.0349**	**0.0373**
Rotation error of RPE [°]				
VINS-Mono [27]	0.1379	**0.1442**	0.1844	**0.1503**
VINS-RGBD [14]	0.3601	0.4537	0.3722	0.3995
Ours	**0.1376**	0.1503	**0.1841**	0.1525

에서는 지하 주차장에서의 일반적인 주행 상황에서 성능을 평가하고자 했다. 표 3에서 볼 수 있듯이 제안하는 방법의 ATE의 RMSE는 VINS-Mono [27] 및 VINS-RGBD [14]와 비교하여 약 4배에서 5배 낮았다. 우리는 시간 경과에 따른 누적 오류가 아닌 순간적인 오류를 제공하는 RPE 관점에서도 성능을 평가했다. 모든 주행 경로에서 우리의 방법은 이동에 대한 오차 평가 지표에 대해서 가장 높은 성능을 보였다. VINS-Mono [27]은 우리의 방법과 비교하여 약간 더 나은 회전 오류 성능을 보였으나, 평균적인 편차가 0.0023°로 미미했다. case 4, 5, 6에서는 각각 연속적인 회전, 90도 각도의 급격한 회전 및 반복되는 예각 및 둔각 회전과 같은 다양한 궤적 조건에서 성능을 평가하고자 했다. 그림 5에서 보여진 모든 경우의 추론 결과 시각화에서 VINS-Mono [27] 및 VINS-RGBD [14]에서 발생 큰 추정 오류를 확인할 수 있다. 특히, VINS-Mono [27] 와 VINS-RGBD [14]는 case 5와 case 6 그리고 case 2에서 추정 결과에 누적된 오차가 허용 가능한 임계값을 초과하여 VIO 시스템의 재시작하기도 했다. 우리가 제안하는 방법을 사용하여 VIO를 수행했을 때, 결과적으로 ground truth에 더 가까운 성능을 보였다. 따라서 RGB-D 카메라의 깊이 맵에 의존하는 대신 깊이 추정 네트워크에서 추정된 깊이 맵을 활용하는 것이 지하 주차장과 같은 환경에서 더 효과적이라고 할 수 있다.

6. 절제 연구

우리는 본 논문에서 제안된 커널 필터링 방법에서 사용되는 휴리스틱 파라미터인 커널 크기 K와 임계값 T의 영향을 분석하기 위해 절제 연구를 수행했다. 이 실험에는 필터링 기법을 적용하지 않은 깊이 네트워크에 의해 추정된 깊이 맵의 모든 값이 VIO에 활용되었다. 표 2에서 볼 수 있듯이, 커널 크기 K가 3, 5, 7일 때 각각 임계값 T를 0.01, 0.05, 0.05으로 설정하게 되면 필터링 기법을 적용하지 않을 때에 비해서 ATE의 RMSE 기준으로 약 93.36%에서 95.44% 정도 오차율을 낮출 수 있었다. 이는 깊이 추정 네트워크의 추정 결과 중에서 정확도가 낮을 수 있는 영역을 필터링하는 것이 확실하게 성능을 향상 시킬 수 있음을 보여준다.

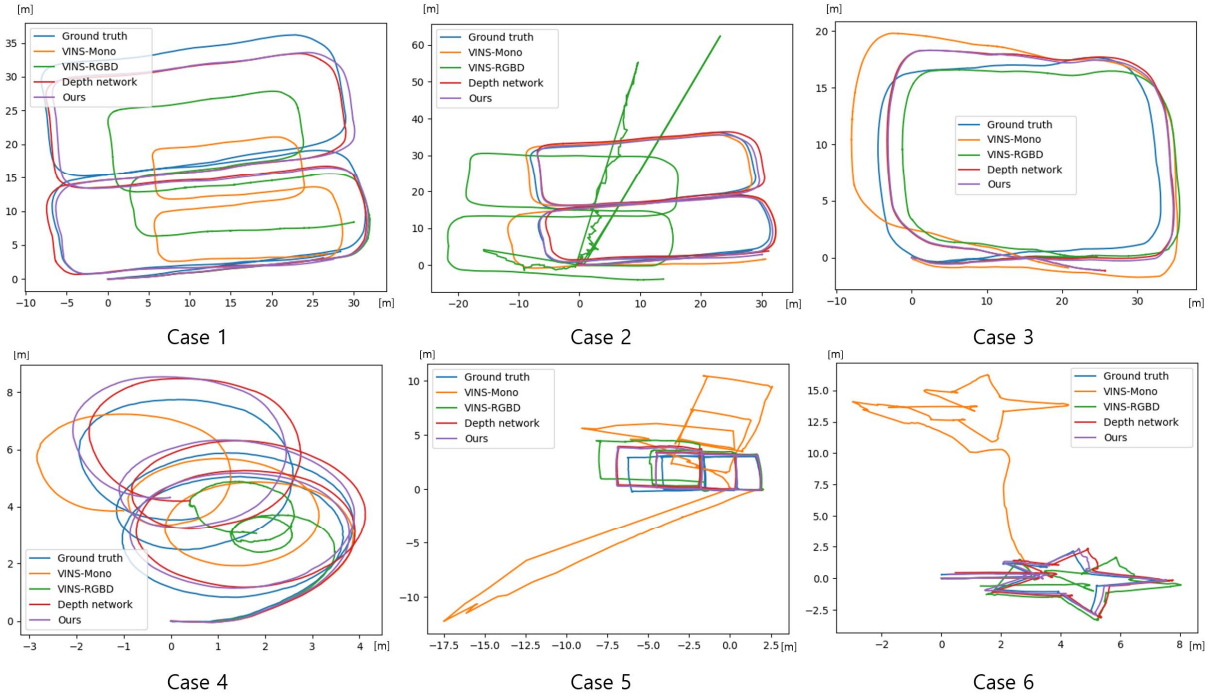

그림 5. VIO 결과에 대한 기존 방법과 제안하는 방법의 정성적 성능 비교
Fig. 5. Comparison of the VIO results based on the previous and proposed method

하지만 커널 크기 K를 5와 7로 설정하고 임계값 T를 0.01로 설정한 경우, 필터링 기법을 적용하지 않은 경우 보다 성능이 낮았다. 이는 커널 크기 K가 커지게 됨에 따라서 비교해야 할 화소가 증가하지만, 이때 임계값 T를 낮게 설정하게 되면 상당수의 화소를 필터링하게 되어 성능 하락이 발생할 수 있음을 의미한다.

V. 결 론

본 논문에서는 깊이 추정 및 필터링 기법이 포함된 RGB 기반 VIO 접근법을 소개했다. 이 방법은 깊이 추정 네트워크의 추론 결과를 활용한다. 실환경 데이터셋에서의 실험 및 다양한 관점에서의 분석을 통해서 깊이 추정 및 VIO에 대한 성능 향상에 커널 기반 필터링이 유효함을 보였다. 우리가 제안한 필터링 방법은 딥러닝 기반 깊이 추정이 특히 객체 경계나 모바일 로봇의 시각에서 먼 지역과 같은 곳에서 도전을 겪는다는 기본적인 가정에 따라 동작한다. 관측된 성능 향상은 RGB-D 카메라의 이상적인 깊이 범위를 벗어난 보다 넓은 공간에서 VIO 실험을 수행한 데에서 부분적으로 기인할 수 있다. 기존 VIO 연구는 주로 실내 환경에서 RGB-D 카메라를 활용해 왔지만, 본 연구는 더 큰 공간에서 깊이 정보를 획득하기 위한 대안적 방법에 대한 필요성을 강조한다. 본 연구 결과는 다양한 환경 및 상황에서 VIO 성능을 향상시키기 위해서 새로운 방향을 제시함에 있어 도움이 되길 기대한다. 또한 추후에는 객체의 다양성과 도로노면의 구배 등과 같이 지하 주차장의 일반적인 특징들을 포함할 수 있는 데이터셋을 구축하게 되면 실제 산업 적용을 앞당길 수 있을 것이라 생각된다.

References

[1] R. Bloss. "Sensor Innovations Helping Unmanned Vehicles Rapidly Growing Smarter, Smaller, more Autonomous and more Powerful for Navigation, Mapping, and Target Sensing," Sensor Review, Vol. 35, No. 1, pp. 6-9, 2015.

[2] J. Wang, A. Chortos, "Control Strategies for Soft Robot Systems," Advanced Intelligent Systems, Vol. 4, No. 5, 2022.

[3] V. S. D. M. Sahu, P. Samal, C. K. Panigrahi, "Modelling, and Control Techniques of Robotic Manipulators: A Review," Materials Today: Proceedings, Vol. 56, No. 5, pp. 2758-2766, 2022.

[4] P. Štibinger, G. Broughton, F. Majer, Z. Rozsypálek, A. Wang, K. Jindal, A. Zhou, D. Thakur, G. Loianno, T. Krajník, M. Saska, "Mobile Manipulator for Autonomous Localization, Grasping and Precise Placement of Construction Material in a Semi-structured Environment," IEEE Robotics and Automation Letters, Vol. 6, No. 2, pp. 2595-2602, 2021.

[5] J. Nidamanuri, C. Nibhanupudi, R. Assfalg, H. Venkataraman, "A Progressive Review: Emerging Technologies for ADAS Driven Solutions," IEEE Transactions on Intelligent Vehicles, Vol. 7, No. 2, pp. 326-341, 2021.

[6] R. Muñoz-Salinas, R. Medina-Carnicer, "UcoSLAM: Simultaneous Localization and Mapping by Fusion of Keypoints and Squared Planar Markers," Pattern Recognition, Vol. 101, pp. 107193, 2020.

[7] S. Hong, A. Bangunharcana, J. Park, M. Choi, H. Shin, "Visual SLAM-based Robotic Mapping Method for Planetary Construction," Sensors, Vol. 21, No. 22, pp. 7715, 2021.

[8] W. Deng, K. Huang, X. Chen, Z. Zhou, C. Shi, R. Guo, H. Zhang, "Semantic Rgb-d Slam for Rescue Robot Navigation," IEEE Access, Vol. 8, pp. 221320-221329, 2020.

[9] C. Bai, T. Xiao, Y. Chen, H. Wang, F. Zhang, X. Gao, "Faster-LIO: Lightweight Tightly Coupled LiDAR-inertial Odometry Using Parallel Sparse Incremental Voxels," IEEE Robotics and Automation Letters, Vol. 7, No. 2, pp. 4861-4868, 2022.

[10] S. Lynen, M. W. Achtelik, S. Weiss, M. Chli, R. Siegwart, "A Robust and Modular Multi-sensor Fusion Approach Applied to Mav Navigation," IEEE/RSJ International Conference on Intelligent Robots and Systems, 2013.

[11] S. Weiss, M. W. Achtelik, S. Lynen, M. Chli, R. Siegwart, "Real-time Onboard Visual-inertial State Estimation and Self-calibration of MAVs in Unknown Environments," IEEE International Conference on Robotics and Automation, 2012.

[12] M. Bloesch, S. Omari, M. Hutter, R Siegwart, "Robust Visual Inertial Odometry Using a Direct EKF-based Approach," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015.

[13] Z. Yang, S. Shen, "Monocular Visual - inertial State Estimation with Online Initialization and Camera - IMU Extrinsic Calibration," IEEE Transactions on Automation Science and Engineering, Vol. 14, No. 1, pp. 39-51, 2016.

[14] Z. Shan, R. Li, S. Schwertfeger, "RGBD-inertial Trajectory Estimation and Mapping for Ground Robots," Sensors, Vol. 19, No. 10, pp. 2251, 2019.

[15] D. Eigen, C. Puhrsch, R. Fergus, "Depth Map Prediction from a Single Image Using a Multi-scale Deep Network," Advances in Neural Information Processing Systems, 2014.

[16] J. H. Lee, M. Han, D. W. Ko, I. H. Suh, "From Big to Small: Multi-scale Local Planar Guidance for Monocular Depth Estimation," arXiv preprint, 2019.

[17] G. Huang, Z. Liu, L. V. D. M, K. Q. Weinberger, "Densely Connected Convolutional Networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[18] Q. Wang, D. Dai, L. Hoyer, L. V. Gool, O. Fink, "Domain Adaptive Semantic Segmentation with Self-supervised Depth Estimation," Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

[19] Q. Dai, V. Patil, S. Hecker, D. Dai, L. V. Gool, K. Schindler, "Self-supervised Object Motion and Depth Estimation from Video," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020.

[20] Y. Wang, W. Chao, D. Garg, B. Hariharan, M. Campbell, K. Q. Weinberger, "Pseudo-lidar from Visual Depth Estimation: Bridging the Gap in 3d Object Detection for Autonomous Sriving," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.

[21] D. Wofk, R. Ranftl, M. Müller, V. Koltun, "Monocular Visual-inertial Depth Estimation," IEEE International Conference on Robotics and Automation (ICRA), 2023.

[22] N. Merrill, P. Geneva, G. Huang, "Robust Monocular Visual-inertial Depth Completion for Embedded Systems," IEEE International Conference on Robotics and Automation (ICRA), 2021.

[23] Y. Almalioglu, M. Turan, M. R. U. Saputra, P. P. D. Gusmão, A. Markham, N. Trigoni, "SelfVIO: Self-supervised Deep Monocular Visual - Inertial Odometry and Depth Estimation," Neural Networks, Vol. 150, pp. 119-136, 2022.

[24] X. Zuo, N. Merrill, W. Li, Y. Liu, M. Pollefeys, G. Huang, "CodeVIO: Visual-inertial Odometry with Learned Optimizable Dense Depth," IEEE International Conference on Robotics and Automation (ICRA), 2021.

[25] K. Sartipi, T. Do, T. Ke, K. Vuong, S. I. Roumeliotis, "Deep Depth Estimation from Visual-inertial Slam," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020.

[26] L. Chen, G. Papandreou, I. Kokkinos. K. Murphy, A. L. Yuille, "Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected Crfs," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 40, No. 4, pp. 834-848, 2017.

[27] T. Qin, P. Li, S. Shen, "Vins-mono: A Robust and Versatile Monocular Visual-inertial State Estimator," IEEE Transactions on Robotics, Vol. 34, No. 4, pp. 1004-1020, 2018.

[28] J. Civera, A. J. Davison, J. M. M. Montiel, "Inverse Depth Parametrization for Monocular SLAM," IEEE Transactions on Robotics, Vol. 24, No. 5, pp. 932-945, 2008.

[29] S. Agarwal, K. Mierle, "Ceres Solver: Tutorial & Reference," Google Inc, 2012.

[30] J. Sturm, N. Engelhard, F. Endres, W. Burgard, D. Cremers, "A Benchmark for the Evaluation of RGB-D SLAM Systems," IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012.

Jimin Song (송 지 민)

2022 Department of Electronic Engineering from Jeonbuk National University (B.S.)
2024 Department of Electronic Engineering from Jeonbuk National University (M.S.)
2024~Department of Electronic Engineering from Jeonbuk National University (Ph.D.)

Field of Interests: Artificial intelligence, Computer vision, Deep learning, Robotics
Email: jimin_song@jbnu.ac.kr

HyungGi Jo (조 형 기)

2012 Electrical and Electronic Engineering from Yonsei University (B.S.)
2020 Electrical and Electronic Engineering from Yonsei University (Ph.D.)
2021~Div. of Electronic Engineering from Jeonbuk National University (Assist Prof.)

Field of Interests: SLAM, Robot Perception, Autonomous Navigation, Spatial AI
Email: hygijo@jbnu.ac.kr

Sang Jun Lee (이 상 준)

2011 Electrical Engineering from POSTECH (B.S.)
2018 Electrical Engineering from POSTECH (Ph.D.)

Career:
2018~2020 Samsung Advanced Institute of Technology (Senior Researcher)
2020~Jeonbuk National University (Assist Prof.)

Field of Interesst: Artificial intelligence, Computer vision, Deep learning, Robotics
Email: sj.lee@jbnu.ac.kr