

딥러닝 기반 Depth Estimation 및 Segmentation 활용을 통한 Mono SLAM 알고리즘 성능 개선 연구

임 동 선¹⁾ · 유 진 우^{*2)}

국민대학교 자동차공학전문대학원¹⁾ · 국민대학교 자동차IT융합학과²⁾

Improved Mono-SLAM Algorithm Using Depth Estimation and Segmentation Based on Deep Learning

Dongsun Lim¹⁾ · Jinwoo Yoo^{*2)}

¹⁾Graduate School of Automotive Engineering, Kookmin University, Seoul 02707, Korea

²⁾Department of Automobile and IT Convergence, Kookmin University, Seoul 02707, Korea

(Received 19 April 2023 / Revised 26 May 2023 / Accepted 31 May 2023)

Abstract : In this paper, an improved Mono-SLAM(Simultaneous Localization and Mapping) algorithm is proposed to address the depth estimation of a monocular camera without using additional sensors. The proposed SLAM uses a depth estimation based on a deep-learning approach, and generates a refined depth map through semantic segmentation. The proposed algorithm is validated through the RGB-D mode of ORB SLAM2, and it demonstrated that localization accuracy can be improved, compared to using the Mono version of ORB SLAM2. Moreover, it can utilize RGB-D cameras in outdoor environments, though it may lead to performance degradation. To solve performance degradation in the depth estimation of the depth map at the boundary between distant objects due to insufficient ground information, a refined depth map that uses semantic segmentation is adopted, thus improving localization accuracy.

Key words : SLAM(동시적 위치추정 및 지도 작성), Deep learning(딥러닝), Depth estimation(거리 추정), Mono camera(단안 카메라), Disparity(시차) Semantic segmentation(의미론적 분할)

1. 서 론

자율주행의 안전성 확보를 위해서는 자차의 현재 위치와 주변 환경에 대한 정보를 획득할 수 있는 정밀한 측위 기술이 필요하다. 자차의 정밀한 위치 추정을 위해 정밀도지도와 랜드마크를 활용한 연구^{1,2)}도 진행되고 있으며 GPS 기반의 측위 기술은 주변에 고충 건물이 많거나, 터널 내부와 같은 주변 환경의 영향을 받아 위치 오차가 발생한다. 이를 해결하기 위해 SLAM(Simultaneous Localization and Mapping)에 대한 연구들이 진행되고 있다. SLAM은 동시적 위치추정 및 지도 작성으로 미지의 환경에 대한 지도를 작성할 수 있다. 따라서 로봇, 자율주행, 드론, 자율주차 등 다양한 분야에 활용될 수 있다.

차량의 위치추정 문제를 해결하기 위해 사용되는 SLAM은 주로 라이다, 카메라와 같은 센서들이 널리 사용되고 있다. 라이다는 레이저를 사용하여 주변 환경을 3차원 포인트 클라우드로 표현할 수 있으며 거리 정보가 정확하지만, 센서의 가격이 비싸다는 특징을 가지고 있다. 카메라를 주된 센서로 활용하는 Visual SLAM은 라이다 센서에 비해 저렴하고 영상 정보를 수집할 수 있어 많은 연구가 진행되고 있다.

Visual SLAM은 영상의 특징점(Keypoint) 검출 여부에 따라 Direct SLAM과 Indirect SLAM으로 분류할 수 있다. 대표적인 Direct SLAM인 Large-Scale Direct Monocular SLAM(LSD SLAM)³⁾ 특징 점을 추출하지 않고 이미지

* A part of this paper was presented at KSAE 2022 Spring Conference

* Corresponding author, E-mail: jwyoo@kookmin.ac.kr

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited.

픽셀값을 그대로 사용하여 회전량, 이동량만을 추정하기 때문에 계산량이 적다는 장점이 있으며, 특징점을 사용하는 방식과는 달리 영상의 픽셀값을 사용하기 때문에 Dense한 지도가 생성된다.

Indirect SLAM은 영상의 특징점을 뽑아서 디스크립터를 통해 매칭하는 과정을 거쳐 움직임을 추정하고 번들 조정을 통해 지도상의 3차원 포인트의 위치 및 카메라의 위치를 최적화 한다. 디스크립터는 많은 계산량을 필요로 하기 때문에 실시간으로 사용하기 어렵다. ORB(Oriented FAST and Rotated BRIEF) SLAM⁴⁾은 특징점 계산을 위해 FAST-9 알고리즘을 사용하였으며 이진화된 ORB 디스크립터를 통해 처리 속도를 향상시켰다. 또한 트래킹, 로컬 매핑, 루프 클로징의 세 개의 스크립트를 동시에 활용하는 멀티스레드를 통해 실시간성을 확보하였다.

Mono 카메라를 사용하는 대부분의 Visual SLAM 시스템의 단점은 다음과 같다. 첫 번째, 주변 환경이 정적인 상황을 가정하여 카메라 위치 추정과 주변 환경에 대한 맵을 작성하기 때문에 주변 환경에 동적 물체가 있다면 정확한 위치 추정과 맵 작성에 어려움이 생긴다. 두 번째, 카메라를 통해 획득한 영상 정보는 정확한 3차원 공간 정보를 유추하기 어렵다. 정확한 거리 정보를 알 수 없기 때문에 지도의 스케일을 결정할 수 없는 문제가 발생한다. 앞선 문제를 해결하기 위해 SLAM과 딥러닝을 융합하는 연구가 활발히 진행되고 있는데 주된 내용은 Semantic segmentation을 활용하여 주변 환경의 동적 물체를 판별하고 정적 물체에서 추출된 특징점을 사용하는 방식이다. 대표적으로 DynaSLAM,⁵⁾ SLAMANTIC,⁶⁾ DS-SLAM⁷⁾ 등이 있다. 또한 CNN SLAM⁸⁾은 거리 정보의 문제를 해결하기 위해 LSD SLAM에 Depth estimation을 적용하는 연구를 진행하였다. 하지만 이 알고리즘은 실내 환경에서만 검증을 진행한 한계점이 있다.

본 논문에서는 Mono 카메라의 단점인 거리 추정 문제를 해결하기 위해 딥러닝 기반 Depth estimation을 활용하여 Depth map을 획득하고, Segmentation 정보를 추가해 원거리 물체를 제거한 Refined depth map을 획득하여 ORB SLAM2 RGB-D에 적용하는 연구를 진행하였으며, 타당성 검증을 위해 공개된 실외 환경 오픈 데이터 세트와 비교 검증을 시행하였다. 그 결과, 스테레오 카메라, RGB-D 카메라 없이 Mono 카메라로 실외 환경에서의 SLAM 위치추정 성능을 향상할 수 있음을 보였다.

2. 관련 연구

2.1 Depth Estimation 네트워크

Depth estimation(거리 추정)은 카메라와 영상 픽셀에

해당하는 거리를 추정하는 일이다. 영상의 거리 정보는 카메라 센서 특성상 3차원 공간 정보를 2D 이미지 평면에 투영시켜 이미지를 획득하기 때문에 거리 정보의 소실이 발생한다. 기존의 연구에서는 인간이 거리를 추정하는 방식과 같이 두 개 이상의 카메라 센서를 사용하는 스테레오 카메라를 사용하여 거리를 추정하거나 적외선 센서를 사용하는 RGB-D 카메라를 통해 거리 정보 문제를 해결할 수 있지만 야외 환경에서 잘 작동하지 않는 문제점이 가지고 있다. 딥러닝이 발전하면서 이미지의 픽셀의 거리를 추정하는 Depth estimation 알고리즘에 관한 연구도 활발히 이루어지고 있다. Eigen 등⁹⁾은 다른 비전문야에서 상당한 성공을 거둔 Convolutional Neural Network(CNN)을 통해 Single image의 Depth map prediction이 가능함을 보여주었다. Ground truth가 존재하는 데이터를 사용하는 지도학습 방식으로 거리 추정 문제를 공식화하였으며 Coarse 네트워크와 Fine 네트워크를 연결하여 Coarse 네트워크에서는 이미지의 전체적으로 보고 입력 영상의 1/4의 해상도의 대략적인 깊이 영상을 추정하여 Fully Convolutional Network(FCN)로 구성된 Fine 네트워크에서 깊이 정보를 보정하는 방식을 제안하였다.

이미지의 깊이 정보는 실내 데이터 세트의 경우에서도 이미지마다 크게 다르다. 예를 들어, 가구 같은 물체 이미지의 깊이 정보는 작은 값에 집중되지만 복도 이미지의 깊이 정보는 큰 값에 집중될 수 있다.

Bhat 등¹⁰⁾은 앞서 언급된 이미지마다 거리 분포가 크게 다르다는 점에 주목하여 전역 정보를 사용하여 깊이의 범위를 추정하는 트랜스포머 기반의 아키텍처 블록을 추가하여 매끄러운 깊이 맵을 생성하여 전체 깊이 추정 성능을 향상시켰다.

2.2 Semantic Segmentation 네트워크

Semantic segmentation은 딥러닝 기반 객체 검출(Object detection)과는 달리 모든 픽셀에 대해 각각의 픽셀이 어떤 물체인지 구분하여 의미 정보를 부여하는 작업이다. 객체의 분류를 통하여 Visual SLAM에서 동적 물체를 구별하는 등 다양하게 사용될 수 있다.

FCN¹¹⁾은 기존의 이미지 분류에서 우수한 성능을 보인 CNN 기반 VGG16¹²⁾ 모델을 Semantic segmentation을 수행하기 위해 변형시켰으며 일반적인 CNN 기반 이미지 분류 모델은 네트워크 뒷단에 Fully connected layer를 붙여 이미지 분류 작업을 수행하는데 Fully connected layer를 거치고 나면 위치 정보가 사라지기 때문에 Segmentation에 적합하지 않다. 위 문제를 해결하기 위해 Fully connected layer를 Convolution layer로 바꾸어 위치 정보의 손실을

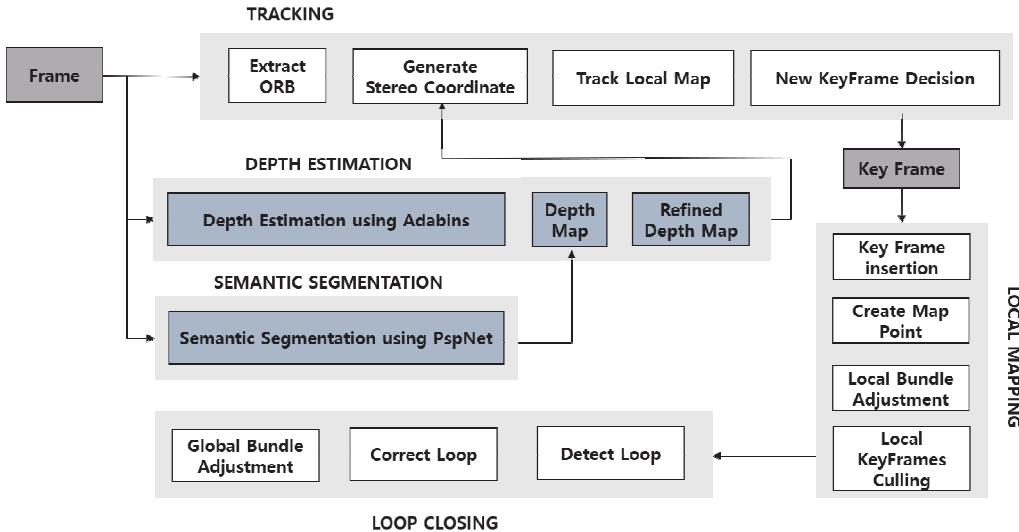


Fig. 1 Proposed SLAM system architecture

방지하여 핵심마다 Segmentation을 가능하게 하였다.

PspNet¹³⁾은 기존의 FCN의 문제점인 주변 상황과 맞지 않는 분류(Mismatched relationship), 헷갈릴 수 있는 분류(Confusion categories), 눈에 잘 띄지 않는 물체의 분류(Inconspicuous classes)을 주변상황(Global information)의 도입으로 형상과 주변 상황을 모두 고려하여 mIOU를 향상시켰다.

3. Proposed Mono SLAM

본 장에서는 Mono 카메라의 단점인 거리 정보의 부정확함을 해결하기 위해 딥러닝 기반의 Depth estimation과 Semantic segmentation을 사용하는 것을 제안한다. Fig. 1은 제안된 SLAM 시스템의 아키텍처를 나타내며, Depth estimation을 통해 거리 정보를 담은 깊이 맵을 추론하고, Semantic segmentation을 통해 구별된 영역으로 Depth map을 보정하여 Refined depth map을 생성하는 방식을 사용하였다.

3.1 Depth Estimation 블록

딥러닝 네트워크를 활용하여 Mono 카메라 이미지 거리를 추정한다. 거리 추정 네트워크는 이미지 속 물체와의 거리가 얼마나지를 추정하는 딥러닝 네트워크이다. 본 논문에서는 Depth estimation network로 Adabins를 사용하였으며 실외 데이터 세트인 KITTI Eigen split을 통해 훈련되었다. 해당 데이터세트는 Ground truth를 획득하기 위해 라이다를 사용하며, 거리 정보를 담고 있는 포인트 클라우드를 카메라 이미지 평면에 사영하여 Ground truth를 획득하였다.



Fig. 2 Comparison of camera image and semi-dense depth map. (a) camera image, (b) ground truth

Fig. 2의 (b)는 라이다 데이터를 이미지 평면에 사영한 결과를 나타낸 것이며 먼 거리의 물체는 어두운색을 가진다. 본 실험에 사용한 Adabins network는 이미지의 모든 핵심에 대해 Depth 정보를 추론하여 Dense depth map을 생성한다. 생성된 Dense depth map은 16bit image로서 핵심값을 0 ~ 65,535까지의 값을 가질 수 있다. 해당 Dense depth map은 거리 정보를 담고 있는 행렬로, 256의 핵심마다 1 m의 거리를 나타낸다. 즉 핵심값 512는 2 m의 거리를 나타낸다. 또한, 핵심값 0은 해당 핵심에 라이다가 없음을 의미한다. Fig. 3은 원본 이미지와 Depth estimation 결과를 나타낸 그림이다. 카메라로부터 거리가 가까운 물체는 밝은색을 나타내며, 멀어질수록 어두

운 보라색으로 추정된다. 학습에 사용되는 데이터는 라이다 센서의 특성상 데이터가 희소한 것이 특징이며, 수직 Field of View(FOV)가 좁고 제한된 센서 탐지 범위를 가지고 있어 거리가 먼 물체에 대한 정보가 부족하다. 따라서 Ground truth 정보가 부족하여 학습에 사용되는 데이터가 제한적이고 그 결과 Depth estimation 결과인 Depth map에 대해 거리를 추정할 수 없는 하늘 영역에 대한 결과가 부정확하며 가까운 거리로 인식하는 것을 확인하였다.

3.2 Semantic Segmentation 블록

본 연구에서는 위에서 언급한 Estimated depth map의 성능을 개선하기 위하여 Semantic segmentation을 사용하-

는 방식을 제안한다. Semantic segmentation은 이미지 속 모든 픽셀의 Class를 예측하는 작업으로 본 연구에서는 Class를 Car, Person, Road, Traffic sign, Wall, Building, Sky 등으로 분류하게 된다. Fig. 3의 (c)는 Segmentation 결과이며 각 픽셀의 Class 별로 다른 색상을 띠게 된다. 하늘 색으로 표기된 영역은 하늘로 인식된 영역이며 Depth estimation 결과 하늘 영역에 대한 값이 부정확한 점과 상당히 먼 거리는 점을 고려하여 SLAM의 위치추정 성능에 도움이 되지 않는다고 가정한다. 앞선 가정을 기반으로 해당 영역의 Depth 값을 0으로 수정하여 해당 영역에는 거리 정보가 존재하지 않음을 나타내는 방식으로 Refined depth map을 생성하였다. Fig. 3의 (d)는 Refined depth map을 나타낸 그림이며 (b)와 비교하였을 때 SLAM 위치추정 성능에 도움이 되도록 하였다.

3.3 Proposed SLAM System

이번 절에서는 본 연구에서 제안하는 SLAM 시스템에 대해 설명한다. 기존의 ORB SLAM2 RGB-D 버전을 사용하였으며 로컬 매핑, 루프 클로징 스레드는 기존의 ORB SLAM2와 동일하게 사용하였다. 해당 시스템의 기존 입력 데이터는 카메라와 깊이 센서가 결합된 시스템에서 획득한 이미지와 Depth map이다. 앞선 절에서 언급한 방식대로 본 연구에서는 Mono 카메라를 통해 획득한 이미지만을 사용하여 Depth estimation과 Segmentation을 통해 Refined depth map을 획득하고 이를 입력 데이터로 사용한다. RGB-D 버전은 입력 이미지에서 특징점(Keypoint)을 찾고 이를 기반으로 ORB 디스크립터를 생성한다. 각각의 특징점의 이미지 좌표계 (u_L, v_L) 에 식 (1)과 같이 거리 정보 d 를 사용하여 가상의 Stereo 좌표계 (u_R, v_R) 를 생성한다.

$$u_R = u_L - \frac{f_x b}{d} \quad (1)$$

f_x 은 수평 Focal length를 나타내며, b 는 RGB-D 카메라의 카메라와 거리 측정 센서 사이의 베이스라인을 의미한다. 기존의 ORB SLAM2 RGB-D는 데이터 획득에 사용한 실제 센서의 파라미터를 사용하지만, 본 연구에서는 Depth map을 딥러닝을 통해 획득하기 때문에 베이스라인이 존재하지 않는다. 따라서 검증에 사용할 KITTI 데이터 세트를 고려하여 베이스라인을 0.54 m로 설정하였다. 이후 카메라가 이동할 때마다 특징점들 사이의 대응 관계를 찾아 이를 기반으로 카메라의 위치와 자세를 추정한다. 또한, 번들 조정을 통해 모든 키프레임의 위치와 Map points의 오차를 최소화하고, Bag of words를 사

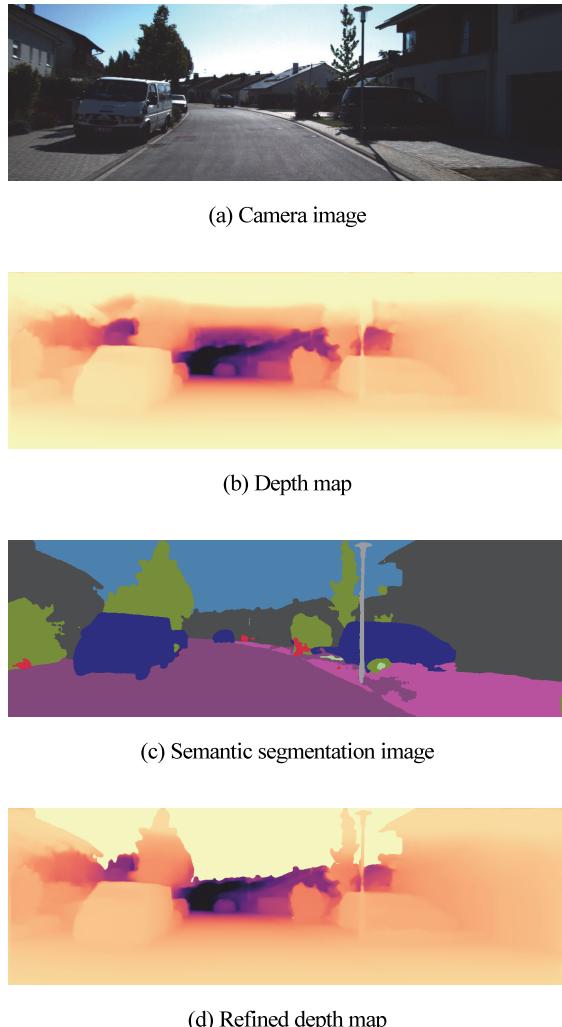


Fig. 3 Results of depth estimation and segmentation.

- (a) camera image, (b) depth map, (c) segmentation image, (d) refined depth map

Table 1 RMSE comparisons of absolute trajectory for both ORB SLAM Mono and Proposed SLAM [m]

Sequence	ORB SLAM Mono RMSE [m]	Initial proposed SLAM (Depth)		Proposed SLAM (Depth + Segmentation)	
		RMSE [m]	Improvement [%]	RMSE [m]	Improvement [%]
KITTI 00	6.68	2.90	56.587	2.67	60.03
KITTI 02	21.75	5.82	73.241	5.65	74.023
KITTI 06	14.68	1.44	90.191	1.28	91.281
KITTI 07	3.36	0.90	73.214	0.89	73.512
KITTI 08	46.58	10.49	77.48	9.60	79.39
KITTI 09	7.62	1.81	76.247	1.49	80.446
KITTI 10	8.68	2.86	67.051	2.59	70.161

용하여 이전에 방문한 장소인지를 확인하여 루프 클로징을 수행해 지도 오차를 보정하게 된다.

4. Experiment Result

4.1 실험 환경

본 연구는 CPU: AMD Ryzen 7 3800XT/ GPU: RTX 3070/ RAM: 32GB 환경에서 실험되었다. SLAM의 위치 추정 성능 향상을 확인하기 위해 ORB SLAM2 RGB-D를 통해 검증하였으며, 실외 환경에서의 SLAM 위치추정 성능을 확인하기 위해 실외 데이터 세트인 KITTI Odometry 데이터 세트¹⁴⁾를 활용하였으며 KITTI는 도심환경과 고속도로 주행 환경에서 획득한 데이터 세트로 평가를 위해 KITTI Sequence 00 ~ 10까지 GPS 기반의 실측 경로를 제공한다. 또한, KITTI Odometry 데이터 세트는 많은 회전과 상대적으로 빠른 차량의 속도로 인해 Mono 카메라 기반 SLAM은 많은 어려움을 가진다. 본 실험에서는 식 (2)¹⁵⁾와 같이 해당 GPS 데이터와 제안된 SLAM을 통해 추정된 카메라 경로간의 위치 오차를 Absolute Trajectory Error(ATE) RMSE를 통해 비교하여 SLAM 위치추정 성능을 평가하였다. ATE는 Ground truth와 계산된 경로를 정렬한 후, 두 경로 사이의 절대적인 거리를 오차로 계산하는 것으로 주로 SLAM의 위치추정 성능 평가를 위해 사용된다. T_i 는 i 시간대에서의 위치 오차를 의미한다.

$$ATE_{RMSE} = \left(\frac{1}{n} \sum_{i=1}^n \| T_i \| \right)^{\frac{1}{2}} \quad (2)$$

Depth estimation network로 Adabins, Semantic segmentation network로 PspNet을 사용하였다. 각각의 네트워크는 KITTI Eigen split 데이터 세트와 Cityscape 데이터 세트¹⁶⁾로 사전 훈련된 가중치를 활용하였다. Table 2는 Adabins 네트워크가 훈련된 옵션을 나타낸다.

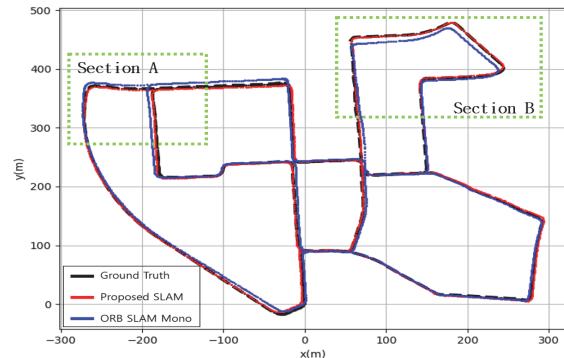
Table 2 Training options of Adabins network

Parameter name	
Optimizer	AdamW
Weight decay	10^{-2}
Max learning rate	3.5×10^{-4}
Max epochs	25
Batch size	16

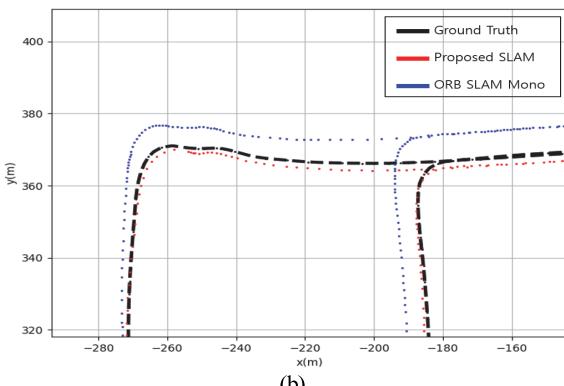
4.2 실험 결과

실험 비교군으로 ORB SLAM2 Mono를 선정하였으며 Refined depth map의 성능을 검증하기 위하여 Depth map을 적용한 경우와 Refined depth map을 적용한 경우로 나누어 진행하였다.

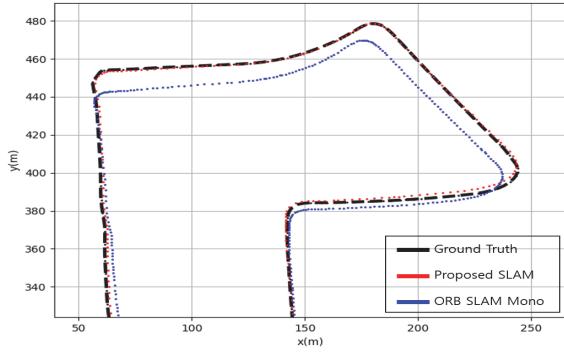
Table 1은 제안된 알고리즘을 활용하여 실험을 진행한 결과를 나타낸다. 이 중 ORB SLAM Mono는 기존 ORB SLAM2 Monocular에 대한 결과를 나타낸다. Initial proposed SLAM은 Depth estimation만을 사용한 결과를 나타내며, Proposed SLAM은 Depth map에 Segmentation을 적용하여 하늘에 속하는 영역에 대한 거리 정보를 삭제한 Refined depth map을 적용한 결과를 나타낸다. 실험 결과, ORB SLAM2 Mono와 비교하였을 때 Depth map을 추가한 경우 평균적으로 73.43 % ATE RMSE가 감소하였으며 Refined depth map을 추가한 경우는 75.55 % ATE RMSE가 감소하였다. 가장 높은 성능 향상을 보인 것은 KITTI Sequence 06으로 Depth map과 Refined depth map을 적용했을 경우 각각 90.191 %, 91.281 % 에러가 감소한 것을 확인하였다. 이외 다양한 Sequence에서도 위치 추정 성능이 향상된 것을 확인할 수 있다. 이는 추가 센서 없이 Mono 카메라와 딥러닝을 사용하여 SLAM의 위치추정 성능을 개선할 수 있다는 것을 의미한다.



(a)



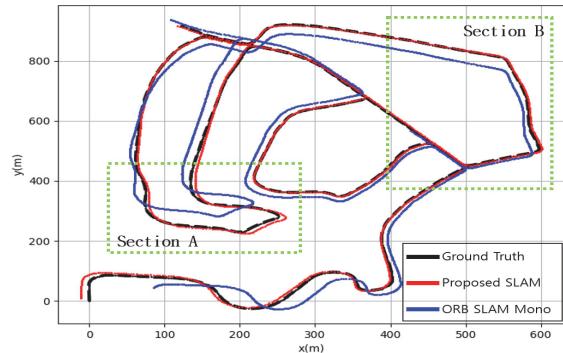
(b)



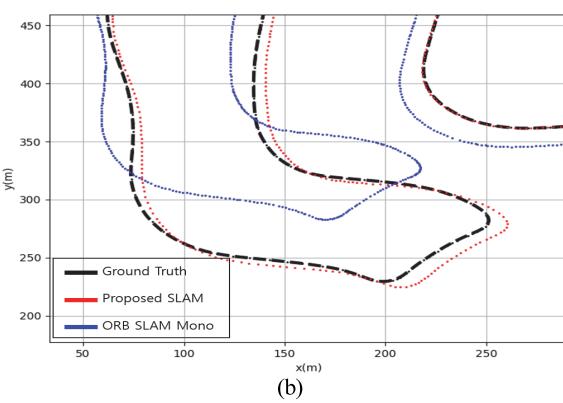
(c)

Fig. 4 In KITTI odometry sequence 00, a comparison of ground truth (black), proposed SLAM (red) and ORB SLAM Mono (blue). (b) the Trajectory of section A, (c) the trajectory of section B

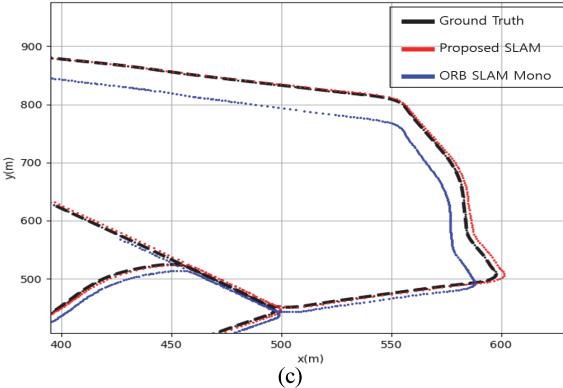
Fig. 4는 KITTI Sequence 00에 대한 경로를 시각화 한 것이다. 실제 경로는 검은 선으로 표시되었으며 빨간 선으로 표시된 경로는 Proposed SLAM, 파란 선은 기존의 ORB SLAM2 Mono를 나타낸다. Fig. 4(b)는 Section A를 확대한 결과이며, Fig. 4(c)는 Section B를 확대한 결과를 나타낸다.



(a)



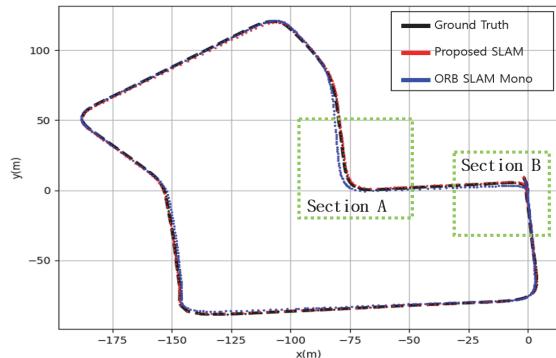
(b)



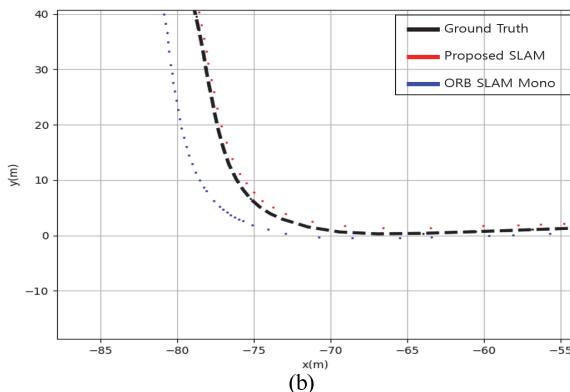
(c)

Fig. 5 In KITTI odometry sequence 02, a comparison of ground truth (black), proposed SLAM (red) and ORB SLAM Mono (blue). (b) the Trajectory of section A, (c) the trajectory of section B

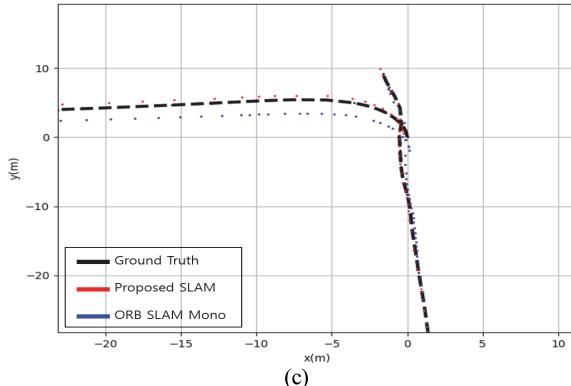
Fig. 5는 KITTI Sequence 02에 대한 경로를 보여준다. Fig 5(b)는 Section A를 확대한 결과이며, Fig 5(c)는 Section B를 확대한 결과를 나타낸다. 해당 시나리오에서는 초기 맵을 생성할 때 기존의 Mono버전은 오래 걸리지만 Refined depth map을 추가한 경우에 더 빠르게 맵을 생성하는 것을 확인할 수 있다. 이는 초기 맵 생성 작



(a)



(b)

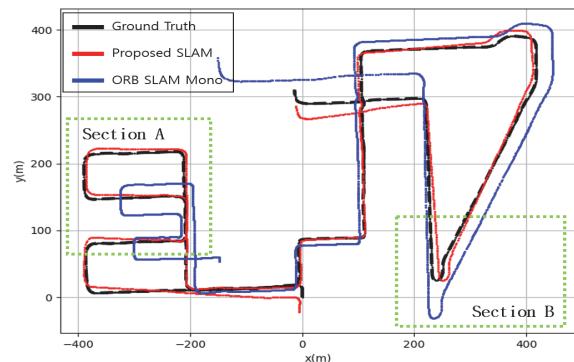


(c)

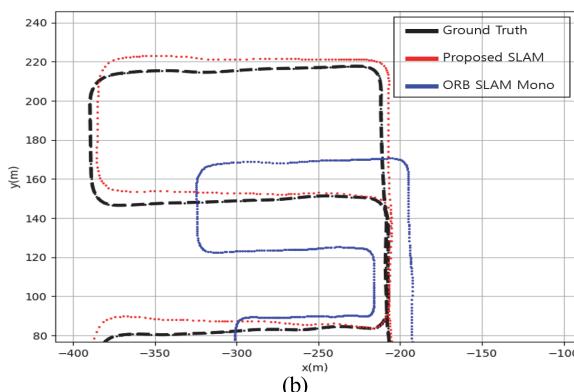
Fig. 6 In KITTI odometry sequence 07, a comparison of ground truth (black), proposed SLAM (red) and ORB SLAM Mono (blue). (b) the Trajectory of section A, (c) the trajectory of section B

업에 거리 정보 추가하였기 때문으로 보인다.

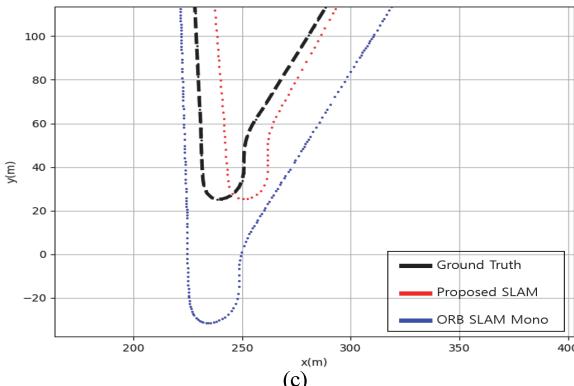
Fig. 7은 KITTI Sequence 08에 대한 경로이다. Fig. 7(b)는 Section A를 확대한 결과이며, Fig. 7(c)는 Section B를 확대한 결과를 나타낸다. Mono 카메라에서는 전체적인 이동량과 스케일에 대한 정보가 부족하여 오차가 크지만, 제안된 SLAM은 거리 정보의 추가로 인하여 개선된 모습을 보여준다.



(a)



(b)



(c)

Fig. 7 In KITTI odometry sequence 08, a comparison of ground truth (black), proposed SLAM (red) and ORB SLAM Mono (blue). (b) the Trajectory of section A, (c) the trajectory of section B

5. 결 론

본 논문에서는 Mono 카메라의 단점인 거리 추정 문제를 해결하기 위해 추가 센서 없이 Mono 카메라의 이미지를 통해 딥러닝 기반 Depth estimation을 활용하여 추정된 Depth map과 Semantic segmentation을 통해 Refined depth map을 생성하는 알고리즘을 개발하였다.¹⁷⁾ ORB

SLAM2 RGB-D를 통해 검증 하였으며 ORB SLAM2 Mono 대비 위치 추정 성능이 향상한 것을 확인할 수 있었다. Mono 카메라는 정확한 거리정보를 획득할 수 없기 때문에 딥러닝을 통한 Depth estimation으로 해당 문제를 해결하여 SLAM의 위치 추정 성능이 향상된 것으로 보인다. 실외 환경에서 RGB-D 카메라의 성능이 떨어지기 때문에 해당 문제를 딥러닝을 통해 대체할 수 있다는 것이 장점이다. 또한 Ground truth 정보가 충분하지 않아 Depth map의 성능이 거리가 먼 물체의 경계 부분에서 성능이 하락하는 문제를 해결하기 위해 Semantic segmentation을 사용하여 Refined depth map을 생성하였을 때 위치추정 성능이 추가 개선된 것을 고려해봤을 때 추후 Depth estimation 네트워크의 거리 추정 정확도가 향상되면 SLAM 위치추정 성능 향상을 기대할 수 있다.

후 기

이 논문은 2023년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임(P0020536, 2023년 산업혁신인재성장지원사업). 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(과제번호: NRF-2021R1F1A1062153).

References

- 1) Y. Na, S. Kim, Y. Kim, J. Park, J. Jeong, K. Jo, S. Lee, S Cho, M. Sunwoo and J. Oh, "HD Map Usability Verification for Autonomous Car," Transactions of KSAE, Vol.28, No.11, pp.797-808, 2020.
- 2) E. Jang, J. Suhr and H. Jung, "Lane-end-point Detection and Automated Performance Evaluation for Vehicle Localization on the Highway," Transactions of KSAE, Vol.27, No.1, pp.47-54, 2019.
- 3) J. Engel, T. Schöps and D. Cremers, "LSD-SLAM: Large-scale Direct Monocular SLAM," Computer Vision-ECCV 2014, pp.834-849, 2014.
- 4) R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An Open-source SLAM System for Monocular, Stereo, and RGB-D Cameras," IEEE Transactions on Robotics, Vol.33, No.5, pp.1255-1262, 2017.
- 5) B. Bescos, J. M. Fácil, J. Civera and J. Neira, "DynaSLAM: Tracking, Mapping, and Inpainting in Dynamic Scenes," IEEE Robotics and Automation Letters, Vol.3, No.4, pp.4076-4083, 2018.
- 6) M. Schörghuber, D. Steininger, Y. Cabon, M. Humenberger and M. Gelautz, "Slamantic - Leveraging Semantics to Improve VSLAM in Dynamic Environments," IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp.3759-3768, 2019.
- 7) C. Yu, Z. Liu, X. Liu, F. Xie, Q. Wei and Q. Fei, "DS-SLAM: A Semantic Visual VSLAM towards Dynamic Environments," IEEE/RSJ International Conference on Intelligent Robots and System (IROS), pp.1168-1174, 2018.
- 8) K. Tateno, F. Tombari, I. Laina and N. Navab, "CNN-SLAM: Real-time Dense Monocular SLAM with Learned Depth Prediction," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp.6243-6252, 2017.
- 9) D. Eigen, C. Puhrsch and R. Fergus, "Depth Map Prediction from a Single Image using a Multi-scale Deep Network," Neural Information Processing Systems(NIPS), 2014.
- 10) S. F. Bhat, I. Alhashim and P. Wonka, "AdaBins: Depth Estimation Using Adaptive Bins," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), pp.4009-4018, 2021.
- 11) J. Long, E. Shelhamer and T. Darrel, "Fully Convolutional Networks for Semantic Segmentation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp.3431-3440, 2015.
- 12) K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-scale Image Recognition," arXiv preprint arXiv:1409.1556, 2014.
- 13) H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, "Pyramid Scene Parsing Network," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp.2881-2890, 2017.
- 14) A. Geiger, P. Lenz and R. Urtasun, "Are we Ready for Autonomous Driving? The KITTI Vision Benchmark Suite," IEEE Conference on Computer Vision and Pattern Recognition, pp.3354-3361, 2012.
- 15) A. Kasar, "Benchmarking and Comparing Popular Visual SLAM Algorithms," arXiv preprint arXiv: 1811.09895, 2018.
- 16) M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp.3213-3223, 2016.

- 17) D. Lim, J. Kim and J. Yoo, "Improving Performance of Mono SLAM Using Depth Estimation Based on Deep Learning," KSAE Spring Conference Proceedings, p.1041, 2022.