

단안 카메라와 저정밀 GPS-IMU 신호를 융합한 맵매칭 방법

김용균¹, 구형일^{1*}, 강석완², 김준원², 김재관²
¹아주대학교 전자공학과, ²한화시스템

High accuracy map matching method using monocular cameras and low-end GPS-IMU systems

Yong-Gyun Kim¹, Hyung-Il Koo^{1*}, Seok-Won Kang², Joon-Won Kim², Jae-Gwan Kim²

¹Department of Electrical and Computer Engineering, Ajou University

²Hanwha Systems

요약 본 논문에서는 단안 카메라와 비교적 오차가 큰 GPS-IMU 센서를 이용하여 이동체의 정확한 포즈를 예측하는 고정밀 맵매칭 방법을 제안한다. 제안하는 방법은 카메라로부터 입력 받은 영상을 딥뉴럴 네트워크를 이용하여 의미상으로 분할한 결과와 시맨틱 지도 정보를 비교함으로써 달성된다. 카메라로부터 입력 받은 주행 영상은 시맨틱 분할 알고리즘을 통해서 두 개의 클래스로 분할되며, 시맨틱 지도 정보와 가능한 레이블 페어에 대해 설정된 가중치에 따라 비교 정렬함으로써 현재 이동체의 정확한 포즈를 예측할 수 있도록 한다. 이 과정에서 비교적 오차가 큰 GPS-IMU 센서의 신호는 해 공간의 범위를 효과적으로 줄여준다. 본 논문은 비교적 저렴한 센서를 이용하여 증강현실 및 자율주행 등에 필요한 고정밀 맵매칭이 가능함을 보여준다. 보정 전후의 차량 경로를 지도에 비교 도시하고 시맨틱 지도를 현재 입력 영상에 오버레이 하여 제안한 방법의 효과를 입증하였다. 또한 non-open-sky 환경과 같은 GPS-IMU 수신에 어려운 환경에서도 성능 개선이 있음을 확인하였다.

Abstract This paper presents a new method to estimate the pose of a moving object accurately using a monocular camera and a low-end GPS+IMU sensor system. For this goal, we adopted a deep neural network for the semantic segmentation of input images and compared the results with a semantic map of a neighborhood. In this map matching, we use weight tables to deal with label inconsistency effectively. Signals from a low-end GPS+IMU sensor system are used to limit search spaces and minimize the proposed function. For the evaluation, we added noise to the signals from a high-end GPS-IMU system. The results show that the pose can be recovered from the noisy signals. We also show that the proposed method is effective in handling non-open-sky situations.

Keywords : Augmented Reality, Deep Learning, Map Matching, Road Detection, Semantic Segmentation

1. 서론

GPS, IMU 등의 센서를 사용하여 측정된 신호를 바탕으로 현재 이동체(예, 차량, 로봇 등)의 정확한 포즈(위치, 방향)를 예측하는 방법의 개발은 자율 주행, AR 경로 안내 등의 응용을 위해 꼭 필요하다. 현재 차량의 종방향, 횡방향 위치와 바라보는 방향에 대한 정확한 정

보를 알고 있어야만 자율주행에 필요한 제어 신호를 생성할 수 있으며, 증강현실(Augmented Reality) 기반 네비게이션이나 원격 조정 같이 인간과의 인터랙션이 필요한 응용 분야에서도 이동체의 정확한 포즈 예측은 필수적이다. 일반적으로 이러한 고정밀 포즈 추정에는 센서의 정확도를 높임으로써 달성되었다. LiDAR 센서를 통해서 주변의 지형지물을 감지하고, 이를 획득된 3차원 지

본 논문은 한화시스템 연구과제로 수행되었음.

*Corresponding Author : Hyung-Il Koo(Ajou Univ.)

Tel: +82-31-219-2479 email: hikoo@ajou.ac.kr

Received March 16, 2018

Revised April 5, 2018

Accepted April 6, 2018

Published April 30, 2018

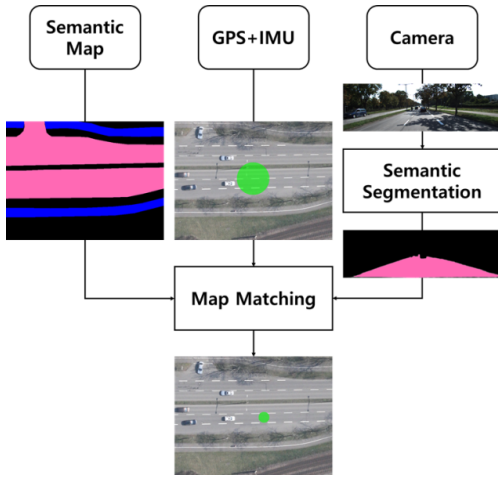


Fig. 1. Block diagram of proposed method. The method yields high-accuracy pose estimates using sensor fusion.

도 정보와 비교함으로써 맵매칭을 수행하거나, 고정밀 GPS-IMU 센서를 이용하여 센서 레벨에서 현 이동체의 포즈를 높은 정확도로 추정하는 방식을 취한 것이다 [1,2]. 하지만 이와 같이 고정밀 센서에 의존하는 방식은 높은 가격 때문에 일반적인 응용에 적용하기 어렵다는 문제점이 있다. 그 외에도 GPS 를 기반으로 하는 방법은 수신 상황에 따라 (Open-sky 환경이 아닌 경우) 통제할 수 없는 방식으로 오차가 증가할 수 있으며, LiDAR 센서를 이용하는 방식은 3차원 지도의 유지, 보수가 필요하다는 단점이 있다.

본 논문에서는 최근 다양한 영상 인식 문제에 성공적으로 적용되고 있는 딥러닝 기술을 이용하여 전방 영상에서 시맨틱 정보를 얻고 이를 비교적 오차가 큰 GPS-IMU 센서의 신호와 융합하여 지도에서의 이동체의 위치 추정 정확도를 높이는 방법을 제안한다. 구체적으로 제안하는 방법의 블록 다이어그램은 Fig. 1에 소개되어 있다.

그림에서 볼 수 있듯이 제안하는 방법은 비교적 저렴한 GPS-IMU 센서를 사용하며 물체의 대략적인 포즈를 알 수 있지만 그 정확도는 상대적으로 높지 않다. 이 상황에서 CNN (convolutional neural network)을 사용하는 시맨틱 분할(semantic segmentation) 알고리즘을 카메라 영상에 적용하여 시맨틱 마스크를 얻고 [3], 이 마스크와 시맨틱 지도 정보를 비교하여 정확한 이동물의 포즈를 예측한다. 제안하는 방법의 효과를 입증하기 위

해 알고리즘 적용 전/후의 차량 궤적을 지도에서 비교하였으며 GPS 정보의 오차가 큰 상황에서 제안하는 방법이 차량의 포즈의 예측 정확도를 높임을 확인하였다.

2. 기하학적 관계

본 논문은 현 이동체의 정확한 포즈를 유클리드 그룹인 SE(3)의 원소로 표현하고, 각 시간(t)에 이 값을 비교적 오차가 큰 GPS-IMU 센서가 제공하는 신호와 전방 카메라를 통해 취득된 영상으로부터 추정하는 방법을 제안한다. 이를 위해 우선 포즈의 매개화 방법 및 기하학적 관계를 다룬다.

2.1 포즈의 매개화

GPS-IMU 센서는 이동체에 고정되어 설치되어 있어 센서의 포즈를 예측하는 것으로 이동체의 포즈를 추정 가능하다. 우리가 구하고자 하는 시간 t 에서의 GPS-IMU 센서의 포즈 $P_S(t)$ 를 SE(3)의 원소로 표현한다.

$$P_S(t) = \begin{bmatrix} R(t) & d(t) \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

여기서 $R(t)$ 는 SO(3)의 원소로 시간 t 에서 센서의 3차원 회전 자세를 나타낸다. 본 논문에서는 RPY 각으로 $R(t)$ 를 매개화한다. 구체적으로 RPY 각은 롤(roll, $\varphi(t)$), 피치(pitch, $\theta(t)$), 요(yaw, $\psi(t)$)로 이루어져 있으며 $R(t)$ 는

$$R(t) = \begin{bmatrix} \cos \psi(t) & -\sin \psi(t) & 0 \\ \sin \psi(t) & \cos \psi(t) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 \cos \varphi(t) - \sin \varphi(t) \\ -\sin \theta(t) & 0 \cos \theta(t) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 \cos \varphi(t) - \sin \varphi(t) \\ 0 \sin \varphi(t) & \cos \varphi(t) \end{bmatrix} \quad (2)$$

으로 주어진다. 한편, $d(t) = [x(t) \ y(t) \ z(t)]^T$ 는 월드 좌표계 (직교 좌표계)기준으로 현재 GPS-IMU 센서 중심의 좌표를 나타낸다.

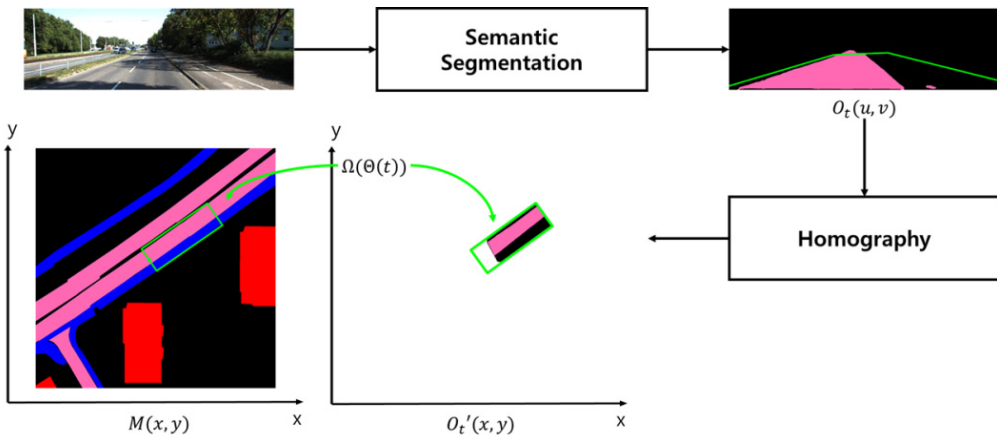


Fig. 2. Proposed registration method

2.2 GPS-IMU 센서 신호

GPS-IMU 센서는 위치를 경도(longitude), 위도(latitude), 고도(altitude) 형태로 제공하며 3차원 회전 자세는 RPY 각으로 주어진다. 우리는 이 값을 직교 좌표계인 UTM 좌표계 (universal transverse mercator coordinate system)로 변환하여 사용한다.

GPS-IMU 신호에서 롤과 피치는 비교적 정확한 값을 제공하고 있으며 고도 변화는 크지 않다고 가정 한다면 우리가 찾고자 하는 포즈는 요의 오차 $\eta(t)$ 와 위치 오차 $(\alpha(t), \beta(t))$ 로 아래와 같이 표현 가능하다.

$$\varphi(t) = \varphi_{se}(t) \quad (3)$$

$$\theta(t) = \theta_{se}(t) \quad (4)$$

$$\psi(t) = \psi_{se}(t) + \eta(t) \quad (5)$$

$$x(t) = x_{se}(t) + \alpha(t) \quad (6)$$

$$y(t) = y_{se}(t) + \beta(t) \quad (7)$$

$$z(t) = h \quad (8)$$

여기서 $\varphi_{se}(t)$, $\theta_{se}(t)$, $\psi_{se}(t)$, $x_{se}(t)$, $y_{se}(t)$ 는 관성 센서로부터 얻은 신호값이다. 따라서 GPS-IMU 신호를 이용하여 $P_S(t) \in SE(3)$ 를 예측하는 문제는 $\Theta(t) = [\alpha(t), \beta(t), \eta(t)]$ 를 예측하는 문제로 볼 수 있다.

2.3 카메라와의 기하학적 관계

이동체에 카메라와 항법센서가 고정 설치되어 있으며

로 카메라의 포즈 $P_C(t) \in SE(3)$ 는 $P_S(t)$ 와 강제운동 관계로 표현할 수 있으며 이 관계식은 고정되어 있다. 따라서 고정된 변환 $T_{CS} \in SE(3)$ 에 대해서 $P_C(t) = T_{CS}P_S(t)$ 로 표현할 수 있으며 초기 캘리브레이션을 통해서 T_{CS} 를 측정할 수 있다.

3. 이동체의 포즈 예측

앞 장에서 관성센서의 좌표계 $P_S(t)$, 카메라의 좌표계 $P_C(t)$ 를 $\Theta(t)$ 로 매개화 하였다. 이 결과를 바탕으로 영상과 시맨틱 지도를 비교하여 $\Theta(t)$ 를 구하는 방법을 설명하도록 한다.

3.1 영상 분석

영상을 분석하기 위해서는 각 픽셀이 어떤 클래스에 속하는지 판단하는 시맨틱 분할(semantic segmentation) 알고리즘을 적용한다[3]. 이 알고리즘은 영상 정보를 VGG16 모델을 이용하여 인코딩하고 이 결과를 FCN (fully convolutional network)를 이용하여 디코딩하는 방식으로 동작하며 영상의 각 픽셀을 2개의 클래스(도로 픽셀, 도로 이외 픽셀)로 표현하고 있다. 즉, 시간 t 에 들어온 영상 입력을 영상 분할 신경망에 적용하여 우리는 각 입력 영상에 대해서 Fig. 2와 같은 영상 분석 결과 $O_t(u, v)$ 를 얻을 수 있다.

3.2 시맨틱 지도와 영상 분석 결과의 관계

현재 카메라 포즈 $P_C(t)$ 와 카메라 캘리브레이션 매트릭스 K (3x4 행렬)가 주어졌다면 월드 좌표계의 한 점 $[X, Y, Z]^T$ 이 이미지 좌표계의 점 $[u, v]^T$ 에 어떻게 대응되는 지 표현할 수 있다. 즉,

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \simeq K(P_C(t))^{-1} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = K(T_{CS}P_S(t))^{-1} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (9)$$

으로 주어진다. Fig. 2에서 볼 수 있듯이 우리가 도로를 평면이라고 가정하면 $O_t(u, v)$ 를 x-y 평면으로 옮기는 변환은 $\theta(t)$ 를 변수로 가지는 호모그래피로 표현된다. 이 호모그래피를 통해 $O_t(u, v)$ 가 변환된 결과를 $O_t'(x, y)$ 라고 한다.

3.3 시맨틱 지도

시맨틱 지도 $M(x, y)$ 는 월드 좌표계를 기준으로 정의된 지리적 위치 정보를 가지고 있다. 시맨틱 지도는 도로, 주차 공간, 배경, 건물, 인도에 대한 레이블을 제공한다. Fig. 2에서 볼 수 있듯이 도로는 분홍색으로 나타나고 인도는 파란색, 건물은 붉은색, 배경은 검은색으로 나타난다. 그림에서는 볼 수 없지만 주차 공간은 노란색으로 나타난다. 시맨틱 지도 정보는 영상 분석 결과와 차이를 측정하기 위해 도로, 주차 공간 그리고 나머지 정보들을 하나로 묶어 배경으로 사용한다.

3.4 지도 정렬

월드 좌표계를 기준으로 정의된 영상 분석 결과 $O_t'(x, y)$ 와 시맨틱 지도 $M(x, y)$ 가 차이를 측정할 수 있도록 목적 함수를 정의하며 이 함수를 최적화하여 $\theta(t)$ 를 예측한다. Fig. 2에서 볼 수 있듯이 시야각과 가리워짐 문제로 $O_t'(x, y)$ 는 전체 지도에서 일부만 정의되며 시맨틱 지도 $M(x, y)$ 와는 일부 영역에서만 비교를 수행한다. 구체적으로 현재 $\theta(t)$ 값을 이용하여 두 시맨틱 지도 비교 범위 $\Omega(\theta(t))$ 를 정의한다. 이 영역은 이동체의 좌우 일정 범위와 전방 일부를 포함하며 직사각형 형태로 면적은 일정하다. 최종적으로 제안하는 목적함수는

$$E(\theta(t)) = \int_{x, y \in \Omega(\theta(t))} d(M(x, y), O_t'(x, y)) dx dy \quad (10)$$

으로 주어진다. 여기서 $d(M(x, y), O_t'(x, y))$ 는 $M(x, y)$ 가 가질 수 있는 세 가지 경우와 $O_t'(x, y)$ 가 가질 수 있는 두 가지 경우에 대해서 가중치를 의미하며 Table 1에 정리되어 있다.

Table 1. Weight table for possible pairs

$M(x, y) \backslash O_t'(x, y)$	road	parking	background
road	0.0	0.5	0.9
background	1.0	0.5	0.4

시맨틱 지도의 클래스 중 주차 공간은 영상 분석 결과와 대응되는 좌표에서 도로로도 분류되고 배경으로도 분류될 수 있으므로 0.5라는 값을 할당했다. 영상 분석 결과의 오검출과 미검출을 고려하여 시맨틱 지도의 클래스가 배경이고 영상 분석 결과에서는 도로인 경우에는 0.9, 배경인 경우에는 0.4로 할당했다. 최적화는 GPS+IMU의 오차를 고려하여 가능한 모든 범위에서 그리디 서치를 진행한다. ($\alpha^2(t) + \beta^2(t) \leq 36, |\eta(t)| \leq 3^\circ$)

4. 실험

실험을 위해서 KITTI tracking benchmark[4]에서 제공하는 주행 영상과 GPS-IMU 정보를 사용하였으며 지도 정보는 Air-Grounded-KITTI [5]에서 제공하는 정보를 사용하였다. KITTI 데이터 셋 취득에 고정밀 GPS+IMU 센서가 사용되었으므로 이 값을 참값으로 보고 추가적인 잡음을 가해서 비교적 저가의 센서로부터 얻게 될 데이터를 생성했다. 구체적으로 위치 정밀도가 2m 이내, 방향(yaw) 정밀도가 1도 이내가 되도록 하였다.

Fig. 3은 10 프레임 간격으로 주행 차량의 위치를 연결하여 이동 경로를 보여주고 있다. Fig. 3-(a)는 고정밀 센서(KITTI 데이터)를 이용하여 취득한 경로이며 Fig. 3-(b)는 잡음이 추가된 경로이다. 제안하는 방법을 Fig. 3-(b)에 적용하며 얻은 결과는 Fig. 3-(c)에 표현되어 있다. 그림에서 확인할 수 있듯이 제안하는 방법은 GPS+IMU

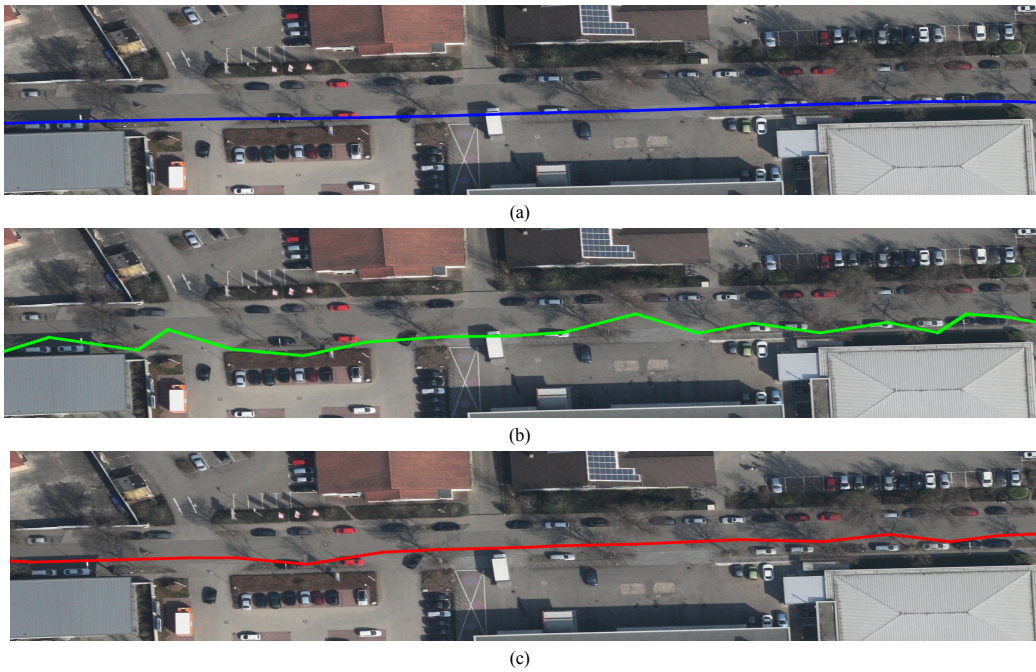


Fig. 3. (a) GPS+IMU(blue) (b) GPS+IMU+Noise(green) (c) our sensor fusion method(red).

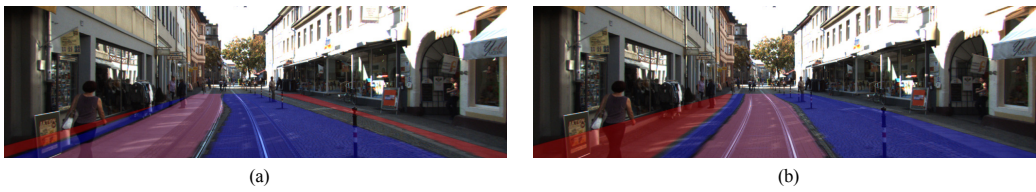


Fig. 4. The semantic map projected into the image (Non-open-sky situations). Pink is road, blue is sidewalk, red is building. (a) GPS+IMU (b) our sensor fusion method

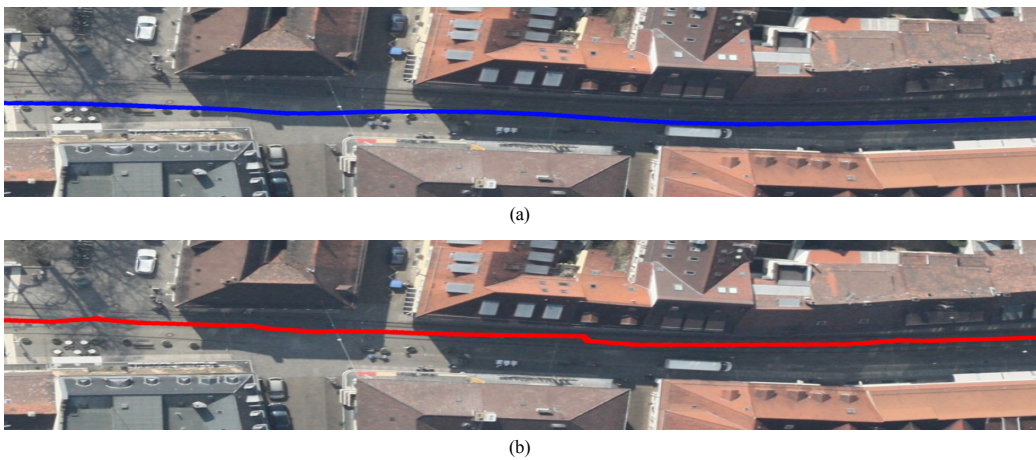


Fig. 5. Trajectories of moving vehicle (Non-open-sky situations)
(a) GPS+IMU(blue) (b) our sensor fusion method(red)

센서의 오차를 효과적으로 줄이는 것을 알 수 있다. 특히, Fig. 3-(a)에 확인할 수 있듯이 고정밀 센서를 사용하더라도 횡방향 오차가 존재하여 도로의 경계에 주행 경로가 표시되는 것이 관찰되는데 이는 Fig. 3-(c)에서 상당부분 완화되었다.

고정밀 센서를 사용하더라도 Fig. 4와 Fig. 5처럼 Open-sky 환경이 아닌 경우는 오차가 커질 수 있다. Fig. 4-(a)는 고정밀 센서로부터 얻는 포즈를 바탕으로 지도 정보(시멘틱 맵 $M(x,y)$)를 영상에 도시한 결과이다. 그림에서 확인할 수 있듯이 도로 경계가 불일치하는 것을 확인할 수 있다. 이와 같은 상황에 제안하는 알고리즘 적용하여 카메라의 포즈를 예측하면 Fig. 4-(b)와 같이 도로 경계를 올바르게 정렬되는 것을 알 수 있다. 이는 제안하는 방법이 GPS+IMU 센서의 오차를 줄여 AR 관련 사용자에게 좀 더 나은 경험을 제공할 수 있음을 보여준다. Fig. 5-(a) 와 Fig. 5-(b) 는 제안하는 방법을 적용하기 전후의 경로이며 제안하는 방법이 오차를 보정함을 확인할 수 있다.

5. 결론

본 논문에서는 딥러닝 기반의 영상 분석 기술과 GPS-IMU 센서, 지도 정보를 사용한 고정밀 맵매칭 방법을 제안했다. 단안 카메라로부터 받은 차량의 전방 영상을 분석하여 도로와 배경에 해당하는 픽셀로 분류하였고, 영상 분석 결과와 지도 정보를 비교하여 정확한 포즈를 찾았다. 실험을 통해 제안하는 방법은 GPS-IMU 오차가 비교적 큰 상황에서 영상 정보를 사용하여 이를 효과적으로 보정할 수 있으며 다양한 분야에 적용될 수 있음을 확인하였다.

References

- [1] M. Limmer, J. Forster, D. Baudach, F. Schüle, R. Schweiger and H. P. A. Lensch, "Robust Deep-Learning-Based Road-Prediction for Augmented Reality Navigation Systems at Night," *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1888-1895, 2016.
DOI: <https://doi.org/10.1109/ITSC.2016.7795862>
- [2] M. Szczot, M. Serfling, O. Löhlein, F. Schüle, M. Konrad and K. Dietmayer, "Global positioning using a digital map and an imaging radar sensor," *2010 IEEE*

Intelligent Vehicles Symposium, pp. 406-411, 2010.
DOI: <https://doi.org/10.1109/IVS.2010.5548043>

- [3] M. Teichmann, M. Weber, J. Zoellner, R. Cipolla and R. Urtasun, "MultiNet: Real-time Joint Semantic Reasoning for Autonomous Driving," arXiv preprint arXiv:1612.07695, 2016.
- [4] A. Geiger, P. Lenz and R. Urtasun, "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite," *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3354-3361, 2012.
DOI: <https://doi.org/10.1109/CVPR.2012.6248074>
- [5] G. Mátyus, S. Wang, S. Fidler and R. Urtasun, "HD Maps: Fine-grained Road Segmentation by Parsing Ground and Aerial Images," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3611-3619, 2016.
DOI: <https://doi.org/10.1109/CVPR.2016.393>

김 용 균(Yong-Gyun Kim)

[정회원]



- 2014년 2월 : 아주대학교 전자공학과 (공학사)
- 2016년 2월 : 아주대학교 전자공학과 (공학석사)
- 2017년 3월 ~ 현재 : 아주대학교 전자공학과 (공학박사과정)

<관심분야>
영상처리, 머신러닝

구 형 일(Hyung-Il Koo)

[정회원]



- 2004년 2월 : 서울대학교 전기컴퓨터 공학부 (공학석사)
- 2010년 2월 : 서울대학교 전기컴퓨터 공학부 (공학박사)
- 2010년 3월 ~ 2012년 8월: 쉐컴 리서치 코리아 선임 연구원
- 2012년 3월 ~ 현재 : 아주대학교 전자공학과 교수

<관심분야>
컴퓨터비전, 머신러닝

강 석 원(Seok-Won Kang)

[정회원]



- 1993년 2월 : 국민대학교 공과대학 전자공학과 (공학석사)
- 1993년 2월 ~ 2002년 12월 : 대우통신(주)
- 2003년 2월 ~ 현재 : 한화시스템(주) 수석연구원

<관심분야>
무인체계, 정보통신

김 준 원(Joon-Won Kim)

[정회원]



- 2000년 2월 : 연세대학교 공과대학 기계공학과 (공학사)
- 2000년 2월 ~ 현재 : 한화시스템(주) 수석연구원

<관심분야>
무인로봇 원격통제, 증강현실

김 재 관(Jae-Gwan Kim)

[정회원]



- 2002년 8월 : 한국과학기술원 통신공학 (공학석사)
- 2008년 8월 : 한국과학기술원 통신공학 (공학박사)
- 2009년 5월 ~ 2011년 3월 : 국방과학연구소 선임연구원
- 2011년 4월 ~ 현재 : 한화시스템(주) 지상시스템팀 수석연구원

<관심분야>
무인체계, 지휘통제