

# 사전위치정보를 이용한 도심 영상의 의미론적 분할

## Semantic Segmentation of Urban Scenes Using Location Prior Information

왕 정 현<sup>1</sup>, 김 진 환<sup>+</sup>

Jeonghyeon Wang<sup>1</sup>, Jinwhan Kim<sup>+</sup>

**Abstract** This paper proposes a method to segment urban scenes semantically based on location prior information. Since major scene elements in urban environments such as roads, buildings, and vehicles are often located at specific locations, using the location prior information of these elements can improve the segmentation performance. The location priors are defined in special 2D coordinates, referred to as road-normal coordinates, which are perpendicular to the orientation of the road. With the help of depth information to each element, all the possible pixels in the image are projected into these coordinates and the learned prior information is applied to those pixels. The proposed location prior can be modeled by defining a unary potential of a conditional random field (CRF) as a sum of two sub-potentials: an appearance feature-based potential and a location potential. The proposed method was validated using publicly available KITTI dataset, which has urban images and corresponding 3D depth measurements.

**Keywords** Semantic Segmentation, Location Prior Map, Road-normal Coordinates

### 1. 서 론

최근 로봇 기술의 발전과 함께 자율 주행 차량에 대한 관심이 증대되고 있다. 자율 주행 차량은 차량에 장착된 각종 센서(예, 카메라 및 깊이 측정 센서)를 통해 주변 환경 정보를 계측하고, 계측 정보를 활용하여 상황에 맞는 적절한 주행을 하는 것을 목표로 한다. 도심 환경은 주변의 보행자 및 차량 등의 예측 불가능한 움직임으로 인해 자율 주행이 어려운 환경 중 하나로, 도심에서의 자율 주행을 위해서는 보다 높은 수준의 주변 환경 인식

결과를 요구한다.

의미론적 분할은 영상 내 최소 단위(즉, 화소 혹은 슈퍼픽셀)를 미리 정의된 레이블 클래스 집합 중 하나로 분류되도록 동일 의미를 가지는 영역으로 그룹화하는 문제이다. 최근 도심 영상의 분할 문제와 보행자 인식 및 차선 탐지 등의 관련 자율 주행 기술과의 융합을 통한 성능 개선 연구가 진행됨에 따라 학계에서 많은 관심을 끌고 있다. 조건부 무작위장(CRF: Conditional Random Field)<sup>[1]</sup>은 의미론적 분할 문제의 해결을 위한 확률 그래프 모델로, 확률 그래프의 노드와 엣지에 대한 포텐셜 함수를 모델링하기 용이하여 널리 쓰이고 있다.

기존의 CRF기반 의미론적 분할 연구에서는 취득한 영상 및 깊이 정보 등의 외형 특징(appearance feature)을 활용하여 영상을 레이블링하는 연구가 진행된 바 있다. 그러나, 도심 영상은 물체 간 가림(occlusion) 및 위장(camouflage) 효과가 빈번해 외형 특징만으로 영상을 레

Received : Apr. 13. 2017; Revised : Aug. 12. 2017; Accepted : Aug. 15. 2017

※This research was a part of the projects titled 'Development of Management Technology for HNS Accident' and 'Development of an autonomous ship-hull inspection system', funded by the Ministry of Oceans and Fisheries, Korea.

<sup>+</sup>Corresponding author: Department of Mechanical Engineering, and Robotics Program, KAIST, Daejeon, 34141, Korea (jinwhan@kaist.ac.kr)

<sup>1</sup>Robotics Program, KAIST, Daejeon, 34141, Korea (jh\_wang@kaist.ac.kr)

이블링하는 데 어려움이 있다. 상황 정보(contextual information)는 각종 센서로부터 취득한 외형 특징과는 다른 추가적인 정보로, 적절한 상황 정보의 모델링을 통해 의미론적 분할 성능을 개선시킬 수 있다는 점에서 많은 연구가 진행되고 있다<sup>[2]</sup>.

상황 정보는 다양한 형태로 의미론적 분할 문제에 적용될 수 있는데, 과분할(over-segmentation)을 통한 국소 영역의 그룹화<sup>[3]</sup>, 야외 건물 표면의 법선 벡터 추정을 통한 공간 정보 모델링<sup>[4]</sup>, 고차원 확률 그래프 모델링<sup>[5]</sup>, 상대위치선호도(relative location prior)<sup>[6]</sup> 및 영역 별 공간적 관계 모델링<sup>[7]</sup>을 포함한다.

Gould<sup>[6]</sup>는 상황 정보로서 레이블 가능한 모든 클래스 간 상대위치선호도를 학습하였으며, Ladicky<sup>[8]</sup>는 동시 발생 선호도를 학습하여 영상 내 어색한 레이블 쌍에 대한 제한 함수를 정의하였고, Diebold<sup>[9]</sup>는 [8]의 연구를 일반화하여 영상 내 대상 물체 간 화소 거리에 따른 동시 발생 선호도를 학습하였다. 그러나, 위 연구들은 실제 공간상의 대상 간 위치 관계를 고려하지 않았다는 점에서 가림 및 위장 효과가 존재하는 도심 영상을 분할할 때 한계점을 보인다.

본 연구에서는 기존 연구의 한계점을 반영하여 상황 정보로서 실제 공간상의 거리를 반영한 사전위치지도를 제시하고 그 유용성을 보이고자 한다. Fig. 1의 도심 풍경에 대한 영상의 예에서 볼 수 있듯이, 도심에 존재하는 건물, 나무 등의 고정 구조물은 일반적으로 도로변을 따라 위치하고, 자동차 및 보행자 등의 동적 물체는 도로 위에 위치한다. 이처럼 주요 대상들이 공간 내 규칙적인 방식으로 배치된 도심 환경은 각 대상 간 사전위치 정보를 모델링하기 용이한 형태이다.

먼저, 사전위치정보의 적용을 위해 도로법선좌표계(road-normal coordinates)라는 특수한 형태의 평면좌표계를 정의한다. 도로법선좌표계는 도로의 방향과 수직인 공간에서 정의되는 평면좌표계로, Fig. 2에서 이에 대한 개념을 보인다. 도로법선좌표계를 정의하면 대상 간 3차원 위치 관계를 모델링하는 데 발생할 수 있는 방향 모호성을 해결해줌으로써 위치 관계를 보다 간단하고 직관적으로 표현할 수 있다. 또한, 도로법선좌표계를 정의하기 위해 의미론적 분할을 수행하기 전 매개변

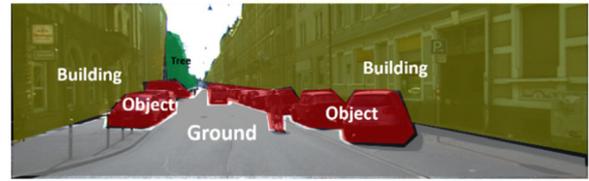


Fig. 1. An example of urban scene segmentation result

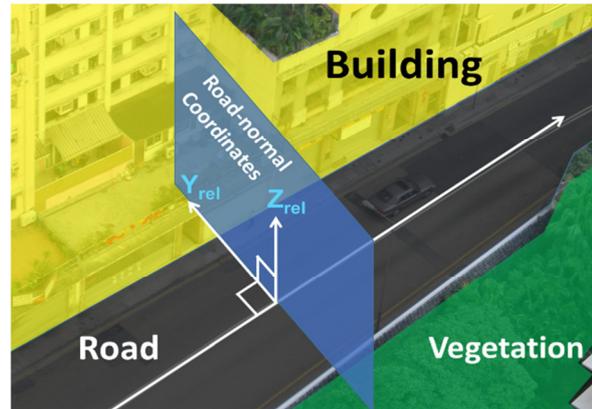


Fig. 2. Definition of road-normal coordinates

수화한 도로 모델을 우선 추정하여 도로의 방향 및 폭을 찾는다.

본 논문의 나머지 부분 중, 이어지는 2장에서는 CRF 기반 의미론적 분할 모델을 전반적으로 소개하고, 3장에서 제안하는 사전위치지도의 정의 및 학습 결과를 자세히 기술한다. 제안된 방법의 유용성을 검증하기 위한 실험 및 검증 결과를 4장에서 제시하며, 5장에서 본 연구의 결론에 대해 정리한다.

## 2. 문제 정의

### 2.1 CRF 기반 의미론적 분할 모델

본 절에서는 의미론적 분할 문제의 해결을 위해 CRF 모델을 도입하고 모델의 각 항에 대해 정의한다. 먼저,  $N$ 개의 슈퍼픽셀로 구성된 영상 및 깊이 정보  $x = \{x_i\}_{i=1}^N$ 와 상응하는  $N$ 개의 레이블 변수  $y = \{y_i\}_{i=1}^N$ 를 정의할 수 있고, 레이블 변수는  $M$ 개의 가능한 분류 범주 중 하나의 값을 갖는다. CRF 기반 의미론적 분할 문제는 입력(즉, 영상 및 깊이 정보)이 주어졌을 때 다음의 에너지 함수를

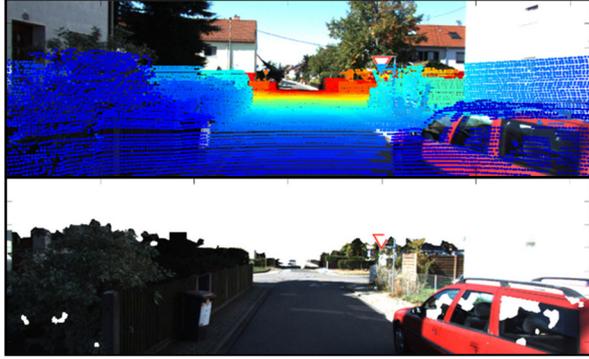


Fig. 3. The 3D point cloud re-projected on the image (Top), and the corresponding area of the image (Bottom)

최소화하는 N개의 레이블 집합을 찾는 문제로 귀결된다.

$$E(y, x) = \sum_i \phi(y_i, x) + \lambda \sum_{i,j} \psi(y_i, y_j, x) \quad (1)$$

식 (1)에서  $\phi(\cdot)$  및  $\psi(\cdot)$ 는 각각 단항 포텐셜(unary potential) 및 쌍별 포텐셜(pairwise potential)로, 단항 포텐셜은 입력 정보를 통해 분류된 i번째 슈퍼픽셀의 레이블 클래스 별 적합도이며, 쌍별 포텐셜은 인접한 레이블 간 유사도를 나타낸다. 여기서, 제안하는 모델은 센서로부터 계측된 외형 특징을 이용해 계산한 레이블 적합도  $\phi_{app}(\cdot)$ 와 사전위치정보를 나타내는  $\phi_{loc}(\cdot)$ 의 합을 단항 포텐셜로 정의함으로써 구성할 수 있고, 이를 아래 수식에 보인다.

$$\phi(y_i, x) = \phi_{app}(y_i, x) + \phi_{loc}(y_i, x) \quad (2)$$

## 2.2 에너지 함수의 정의

본 절에서는 식 (1), (2)에서 제시한 항들에 대해 구체적으로 정의한다. 에너지 함수의 정의는 Cadena<sup>[10]</sup>의 연구 결과를 따르는데, 이는 영상을 분할 하는 데 서로 다른 탐지 범위를 가지는 센서들의 정보를 융합하기 위해서는 탐지 범위가 중첩되는 영역과 중첩되지 않는 영역의 포텐셜 함수가 독립적으로 정의된다는 내용을 포함한다. 따라서 본 연구에서는 카메라 및 라이다로부터 취득한 계측 정보의 중첩 영역 및 미 중첩 영역에 대해 각각 정의한다. 중첩 영역은 영상을 기준으로 라이다로

Table 1. List of feature descriptors

Source	Features		
	Observation	Dim.	Description
Image	RGB	6	Mean and std. dev.
	LAB	6	Mean and std. dev.
	SIFT BoW	25	SIFT descriptor w/ GMM
3D	$v_i$	1	Superpixel vertical location
	$p_i$	3	Point cloud position
	$(\mu_{\Delta d_i}, \sigma_{\Delta d_i})$	2	Mean and std. dev. of depth
	$1 - \text{mean}(\ n_i n_N\ )$	1	Neighboring planarity
	$\frac{3\sigma_3}{\sigma_1 + \sigma_2 + \sigma_3}$	1	Superpixel planarity
	$ n_i \cdot \vec{z} $	1	Superpixel vertical orientation

$n_N$  : Normal vector of neighboring superpixels

$n_i$  : Normal vector of i<sup>th</sup> superpixel

$\sigma$  : Singular values of the superpixel's point cloud such that  $\sigma_1 > \sigma_2 > \sigma_3$

$\vec{z}$  : Unit vector of z axis in the global coordinates

GMM: Gaussian mixture model

부터 얻은 깊이 정보를 영상에 투영시켰을 때 해당되는 영역이고, 미 중첩 영역은 영상에서 깊이 정보가 존재하지 않는 영역을 나타낸다. Fig. 3에 설명한 각 영역의 예를 보인다.

### 2.2.1 단항 포텐셜

본 연구에서 단항 포텐셜은 식 (2)와 같이 두 포텐셜의 합으로 표현할 수 있고, 각 포텐셜을 아래 수식에 보인다.

$$\phi_{app}(y_i, x) = \begin{cases} -\log p(y_i|x_{im}), & \text{if } i \in im \setminus 3D \\ -\log p(y_i|x_{im,3D}), & \text{if } i \in im \cap 3D \end{cases} \quad (3)$$

$$\phi_{loc}(y_i, x) = M(T_{3D}^{RN} \cdot T_{im}^{3D} \cdot p_i) \quad (4)$$

$\phi_{app}(\cdot)$ 은 영상을 기준으로 계측 센서들 간 중첩 탐지 영역 및 미 중첩 영역에 대해 각각 정의되고, 이는 입력 (즉, 영상 혹은 영상 및 깊이 정보)이 주어졌을 때 레이블  $p(y_i|x)$ 의 확률에 음의 로그(negative log)를 취한 형태로 정의할 수 있다. 또한, 위 조건부확률은 각 영역에서

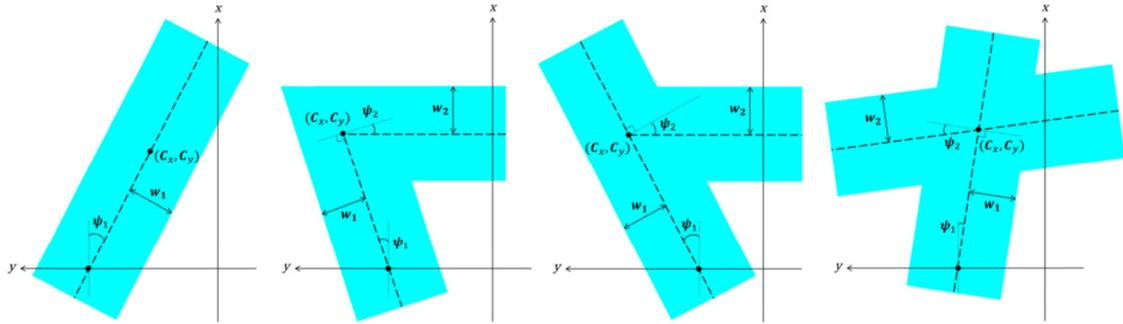


Fig. 4. Road layout models: (from left) Straight, Curve, T-junction, and Crossroad

정의된 특징 기술자(feature descriptor)로 학습된 임의의 숲(Random forest)<sup>[11]</sup> 기반 분류기 결과로 모델링할 수 있다. Table 1에 학습에 사용된 특징 기술자를 나타내었다. 중첩 영역에 대해서는 총 46개의 모든 기술자를 사용하여 분류기를 학습하였으며, 미 중첩 영역은 영상을 통해서만 획득 가능한 38개의 기술자로 학습하였다. 사전위치정보에 대한 포텐셜  $\phi_{loc}(\cdot)$ 은 영상 내 깊이 정보가 유효한 부분에 대해서만 정의되며,  $M(\cdot)$ 은 미리 학습된 사전위치지도를 나타낸다. 사전위치지도는 도심 공간 내 정의된 특수한 좌표계인 도로법선좌표계 내에 정의되는데, 사전위치정보에 대한 포텐셜을 다음의 일련의 과정을 통해 계산할 수 있다. 먼저, 각 superpixel을 대표하는 공간상의 좌표는 해당 superpixel로 투영된 point cloud값의 평균으로 정의되며, 이를 식 (4)에서  $p_i$ 로 표현한다. 평균좌표값  $p_i$ 는 좌표계 변환 행렬을 통해 도로법선좌표계 내 좌표( $T_{3D}^{RN} \cdot T_{im}^{3D} \cdot p_i$ )로 변환될 수 있다. 사전위치정보에 대한 포텐셜은 도로법선좌표계 내 해당 좌표에서의 사전위치지도 값으로 정의된다. 사전위치지도에 대한 자세한 내용은 다음 장에서 기술한다.

### 2.2.2 쌍별 포텐셜

본 연구에서 쌍별 포텐셜은 일종의 스무딩(smoothing) 효과로, 이웃한 슈퍼픽셀들이 서로 다른 종류로 레이블링되는 경우 페널티를 부과하는 방식으로 모델링할 수 있다. 이는 Contrast-sensitive Potts 모델의 형태를 따르며, 다음과 같이 표현할 수 있다.

$$\psi(y_i, x) = \psi_{im}(y_i, y_j, x) + \psi_{3d}(y_i, y_j, x) \quad (5)$$

$$\psi_{im}(y_i, y_j, x) = \begin{cases} \exp(-\|I_i - I_j\|_2), & \text{if } y_i \neq y_j \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

$$\psi_{3d}(y_i, y_j, x) = \begin{cases} \exp(-\|p_i - p_j\|_2), & \text{if } y_i \neq y_j \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

여기서,  $I_i$ 는  $i$ 번째 슈퍼픽셀의 CIELab색공간 값이고,  $p_i$ 는  $i$ 번째 슈퍼픽셀의 공간상의 좌표를 나타낸다. 슈퍼픽셀의 이웃  $i$ 와  $j$ 의 관계에 대한 정의는 Wang<sup>[12]</sup>의 연구 결과를 따른다.

## 3. 사전위치지도의 학습 및 정의

본 장에서는 제안하는 사전위치지도에 대해 정의한다. 사전위치지도는 영상 및 깊이 정보로부터 재구성한 도심의 3차원 레이아웃 내 정의된 특수한 좌표계인 도로법선좌표계에 정의된다. 도로법선좌표계는 도로의 방향에 수직인 2차원 평면좌표계로, 해당좌표계를 정의하기 위해서는 도로 모델의 매개변수 추정을 통해 도로의 방향을 결정해야 한다. 이 일련의 내용들은 1) 도로 모델의 추정; 2) 도로법선좌표계의 정의; 3) 사전위치지도의 학습 순으로 기술된다.

### 3.1 도로 모델의 추정

먼저, 매개변수화한 도로 모델  $R = \{k, \mathbf{c}, \psi, \mathbf{w}\}$ 을 정의한다.  $k$ 는 도로 토폴로지로, 본 연구에서는 직선 도로, 커브길, T자형 도로 및 교차로의 총 4가지 도로 모델을 정의하였다.  $\mathbf{c}, \psi, \mathbf{w}$ 는 각각 도로의 중심점, 도로의 방향 및 도로의 폭을 나타내며, 정의한 4개의 모델에서 매개변수들의 의미를 Fig. 4에 나타낸다. 도로 모델의 추정에

는 외형 정보를 특징 기술자로 하여 학습된 임의 숲 기반 분류기의 결과를 사용하고, 각 슈퍼픽셀 별 분류 결과  $\hat{y}_{app\_i}$ 를 다음과 같이 정의한다.

$$\hat{y}_{app\_i} = \operatorname{argmin}_{y_i} \phi_{app}(y_i, x) \quad (8)$$

도로 매개변수 추정 문제는 모든 슈퍼픽셀 별 외형 정보 기반 분류 결과가 주어졌을 때, 다음의 확률 모델을 최대화하는 매개변수집합  $\hat{R}$ 을 찾는 문제로 정의할 수 있다.

$$\hat{R} = \operatorname{argmax}_R p(R | \hat{y}_{app}) \quad (9)$$

where  $p(R | \hat{y}_{app}) \propto p(R) p(\hat{y}_{app} | R)$

식 (9)의 조건부확률은 베이즈 정리(Bayes' theorem)를 이용해 사전 확률  $p(R)$ 과 우도 함수(likelihood function)  $p(\hat{y}_{app} | R)$ 의 곱으로 표현할 수 있다. 여기서 도로 모델의 매개변수에 대한 사전확률은 매개변수 별 정의역 내 균등분포를 따른다 가정한다. 우도함수는 다음의 볼츠만 분포(Boltzmann distribution)로 모델링할 수 있다.

$$p(\hat{y}_{app} | R) \propto \exp(\rho \cdot f(R)) \quad (10)$$

여기서  $\rho$ 는 점유격자지도(occupancy grid map)의 개념에서 착안한 변수로, 각 슈퍼픽셀이 공간상에 점유된 상태인지를 외형 정보 기반 분류 결과를 통해 판단한다. 본 연구에서는 도심 영상 분할 문제의 해결을 위해 땅(GND), 동적 물체(OBJ), 건물(BLDG), 숲(VEG)의 총 4종류의 클래스를 사용하였다. 깊이 정보를 포함하는 슈퍼픽셀은 하늘(SKY) 클래스를 정답(ground truth)으로 가질 수 없으므로 이를 제외한 4종류의 클래스를 정의한다.  $\rho$ 값이 1인 경우 점유상태(occupied)라 정의하며, -1인 경우 비어있는 상태(free)라 정의한다. 만약 슈퍼픽셀의 분류 결과  $\hat{y}_{app\_i}$ 가 건물(BLDG) 혹은 숲(VEG)인 경우 해당 영역을 지나갈 수 없으므로 점유 상태라 판단할 수 있다. 반면, 분류 결과가 땅(GND)인 경우 지나갈 수 있으므로 비어있는 상태라 판단할 수 있다. 또한, 분류

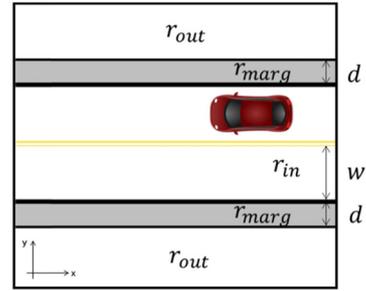


Fig. 5. Road layout potential

결과가 자동차 및 보행자를 포함하는 동적 물체(OBJ)인 경우  $\rho$ 값을 0으로 정의하였다. OBJ 클래스의 분류 결과는 다른 클래스 대비 상대적으로 불확실하여 잘못된 정보를 줄 가능성이 존재하므로 0의 중립 값을 갖게 설정하였다. 앞에서 기술한 내용을 식 (11)에 보인다.

$$\rho_i = \begin{cases} 1(\text{occupied}), & \text{if } \hat{y}_{app\_i} \in \{\text{BLDG}, \text{VEG}\} \\ 0(\text{not used}), & \text{if } \hat{y}_{app\_i} \in \{\text{OBJ}\} \\ -1(\text{free}), & \text{if } \hat{y}_{app\_i} \in \{\text{GND}\} \end{cases} \quad (11)$$

Fig. 5에 도로 모델의 영역 구분의 정의 및 영역 별 가중치를 나타낸다. 도로 모델은 크게 도로 안쪽 부분, 안쪽의 경계로부터  $d$ 만큼 떨어진 영역까지의 경계 영역 및 바깥쪽 부분으로 구분할 수 있다. 안쪽 부분의 거리값은 추정된 도로의 폭으로 정의되고, 경계 영역의 폭은 임의로 조정 가능한 매개변수이며, 경계 영역 외 모든 영역에 대해 바깥쪽 영역으로 정의된다. 식 (10)에서  $f(R) = [r_{in}, r_{marg}, r_{out}]^T$ 은 추정된 도로 모델의 영역 별 가중치로, 실제 깊이 측정 정보의 특징을 감안하여 임의로 설정하였다.

구하고자 하는 최적의 도로 모델 매개변수 집합은 식 (10)의  $\rho$ 와  $f(R)$ 의 곱을 최대화하는 매개변수로 결정되는데, 이는 도심 영상 내 GND 클래스로 분류된 슈퍼픽셀들이 추정된 도로 영역의 내부에 존재하고 BLDG 및 VEG 클래스를 갖는 슈퍼픽셀들이 도로의 외부에 존재하도록 도로의 방향, 폭 및 토폴로지가 결정될 때 최적의 매개변수 집합  $\hat{R}$ 을 갖는다. 최적의 매개변수의 집합은 Geiger<sup>[13]</sup>의 방법을 참고하여 마르코프 연쇄 몬테 카를

로(Markov Chain Monte Carlo) 방법 중 메트로폴리스-헤이스팅스(MH: Metropolis-Hastings) 샘플링 알고리즘을 통해 추정된다.

### 3.2 도로법선좌표계의 정의

앞 절에서 추정한 도로 모델을 통해 도로법선좌표계를 정의할 수 있다. 도로법선좌표계는 추정한 도로 영역의 중심을 가로지르는 방향을 한 축으로 가정했을 때, 이에 수직인 평면좌표계로 정의할 수 있다. 직선 도로를 제외한 모든 종류의 도로 모델은 2개의 직선 도로의 조합으로 표현할 수 있으므로, 하나의 도로 모델 내 2개의 도로법선좌표계가 정의될 수 있다. 단, 사전위치정보 포텐셜  $\phi_{loc}(\cdot)$ 의 계산을 위해 각 슈퍼픽셀에 해당하는 공간상의 점을 도로법선좌표계에 투영시킬 경우, 이중 계산(double counting) 문제가 발생할 수 있다. 따라서 각 좌표계로 투영된 점 중 도로 폭 방향 거리가 가까운 쪽의 투영 결과만 유효하다 가정한다.

### 3.3 사전위치지도의 학습 및 적용

사전위치지도는 각 슈퍼픽셀의 레이블 정답 및 상응하는 깊이 정보를 사용하여 학습되는데, 레이블링된 깊이 정보를 앞 절에서 정의한 도로법선좌표계로 투영시킨다. 각 영상 당 추정된 도로의 폭이 상이하므로, 투영된 깊이 정보의 도로법선좌표계 내 폭 방향 축 성분에 대해 도로 폭으로 나누어 정규화(normalize)한다. 투영된 깊이 정보를 높이 값에 대해 이산화 한 후, 각 레이블 별 같은 높이 값을 갖는 깊이 집합에 대해 커널밀도추정(kernel density estimation)을 통해 사전위치지도를 생성한다. Fig. 6에 깊이 정보의 도로법선좌표계로의 투영 결과 예를 보인다.

일반적으로 깊이 정보는 도심에 존재하는 대상의 바깥면에서만 반사되어 계측되므로, 커널밀도함수를 통해 추정된 사전위치지도는 바깥면 부근(예, 건물의 외벽)에 높은 사전확률값을 갖는 경향이 있다. 이러한 깊이 정보의 불투과성을 고려하기 위해 커널밀도추정 결과 취득한 확률분포와 이 분포의 누적분포와의 합을 사전위치지도로 정의한다. Fig. 7에 학습한 레이블 클래스 별 사

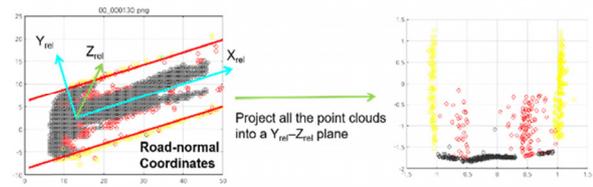


Fig. 6. An example of projecting point clouds into road-normal coordinates. Red circles represent “OBJ” class, yellow circles are “BLDG” class, and Black circles are “GND” class

전위치지도를 보인다.

각 클래스 별 사전위치지도는 탐지될 대상의 크기를 고려하여 서로 다른 높이 범위 내 생성되었다. 사용된 데이터 셋의 깊이 정보가 바닥으로부터 약 1.7 m 위에서 계측된 관계로<sup>[14]</sup>, 장착된 센서의 위치를 기준으로 바닥인 -1.7 m부터 사전위치지도가 생성되었다. Fig. 7 내 검정색 평면은 도로의 외부와 내부의 경계를 나타내며, 도로의 좌우에서 도심 내 대상의 위치 특성은 동일하므로 도로 중앙선을 기준으로 한 쪽 차선에 대해서만 표현하였다. 사전위치지도의 학습 결과 건물(BLDG) 및 숲(VEG)에 대해서는 대체로 도로의 바깥에서 높은 사전위치확률값을 갖는 것을 알 수 있고, 동적 물체(OBJ)의 경우 도로 내에서 높은 값을 갖는 것을 알 수 있다. 위 결과는 도로 위에는 자동차 및 보행자가 존재하고 도로를 따라 나무 및 건물들이 존재하는 도심의 일반적인 환경이 반영되었다는 점에서 의미가 있다.

## 4. 실험적 검증 및 결과

본 연구에서는 제안하는 알고리즘의 성능 검증을 위해 KITTI 도심 데이터 셋<sup>[14]</sup>을 사용하였다. KITTI 데이터 셋은 독일 도심 환경에 대한 스테레오 영상 및 라이다로부터 계측된 동기화된 깊이 정보를 포함한다. 이론적으로 깊이 정보는 스테레오 영상의 쌍을 통해서도 추정할 수 있으나, 라이다를 통해 계측한 깊이 정보 대비 정확도가 떨어진다. 본 연구에서는 사전위치지도의 활용에 있어 보다 정확한 깊이 정보의 사용을 위해 라이다로부터 계측된 깊이 정보를 사용한다. KITTI 도심 데이터 중, 선택된 140개의 영상 정보와 상응하는 깊이 정보

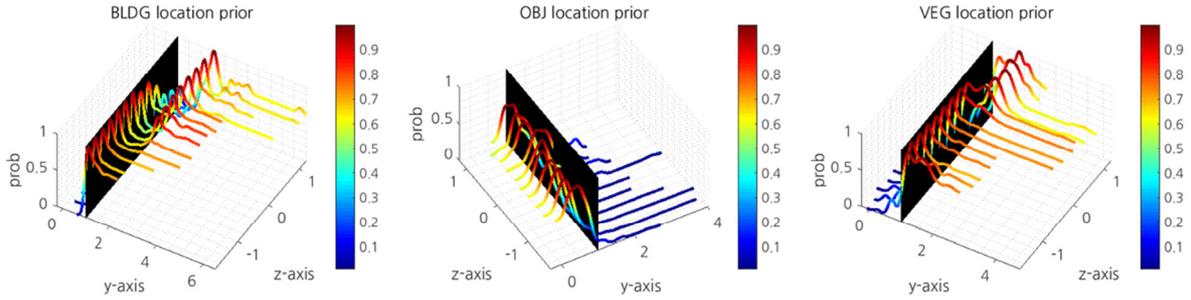


Fig. 7. Trained location prior maps in road-normal coordinates. Colorbars represent the probability of the location prior in each class



Fig. 8. Examples of urban scene labeling results on the KITTI dataset. This figure shows input images (first column), ground truth (second column), [12] (third column), and the results by the proposed method (last column)

를 임의의 숲 기반 분류기 및 사전위치지도의 학습에 사용하였고, 37개의 영상을 알고리즘의 검증에 사용하였다. 또한, 도심 영상의 의미론적 분할을 위해 땅(GND), 동적 물체(OBJ), 건물(BLDG), 숲(VEG)의 4종류의 클래스를 사용하였고, 도로 레이아웃의 영역 별 가중치  $[r_{in}, r_{margin}, r_{out}]^T$  및 경계의 폭  $d$ 를 각각  $[-1, 7, 5]$ , 1로 설정하였다. 식 (1)의 에너지 함수를 최소로 하는 레이블 집합을 찾기 위해 Alpha-expansion 알고리즘<sup>[15]</sup>을 사용하였고, 음의 로그우도(negative log-likelihood)함수를 최소화하는 매개변수  $\lambda$ 를 학습하였다.

Table 2에 의미론적 분할 성능 결과를 보이고, Fig. 8에 레이블 정답, Wang<sup>[12]</sup>의 방법 및 제안하는 방법을 통한 분할 결과의 예를 보인다. 각 알고리즘은 MATLAB 코드로 작성되어 Intel i7-2600 @ 3.40Ghz x 8 프로세서가 탑재된 PC에서 실행되었다. Wang의 연구는 기존의 조건부무작위장 기반 의미론적 분할 방법론을 따름과 동시에 깊이 정보를 통해 영상 내 레이블에 대해 공간 일관성(spatial coherence)의 개념을 적용시켜 분할 성능을 개

선시킨 연구이다. 이는 사전위치정보의 적용을 위해 영상 및 깊이 정보의 융합을 전제로 하는 본 연구의 문제 정의와 일치한다. 따라서 본 논문에서는 Wang의 방법과 영상 분할 성능을 비교하였고, 각 방법론 별 재현율(recall)을 척도로 성능을 비교하였다. Table 2에서 Global은 전체 슈퍼픽셀 개수 대비 재현율이며, Class는 레이블 별 재현율의 평균, GND, OBJ, BLDG 및 VEG는 각 클래스 별 재현율을 나타낸다. 여기서, 모든 재현율 계산 결과는 라이다로 측정한 깊이 정보를 영상에 투영시켰을 때 중첩되는 탐지 영역에 한해 계산하였다. 이는 두 센서 간 서로 다른 탐지 영역에 대한 처리를 포함하는 것으로, 평균 55% 정도의 영역이 중첩 탐지 영역에 해당하는 것으로 확인하였다. 제안하는 방법은 Wang의 연구와 비교했을 때 전체 슈퍼픽셀 개수 대비 재현율(Global)과 레이블 별 재현율의 평균(Class)에서 각각 7.3% 및 7.8% 향상된 성능을 보인다. 각 레이블 별 재현율에 있어서도 대체로 성능이 향상되었으나, OBJ 클래스에 대해서는 재현율이 소폭 감소하였음을 알 수 있다. 이는 기존

Table 2. Performance evaluation on the KITTI dataset

Methods	Performance (KITTI dataset)						Computational cost (ratio)
	Global	Class	GND	OBJ	BLDG	VEG	
Wang <sup>[12]</sup>	80.0	78.6	95.9	<b>93.5</b>	44.1	86.7	1
Proposed	<b>87.3</b>	<b>86.4</b>	<b>96.2</b>	80.0	<b>78.0</b>	<b>91.3</b>	1.57

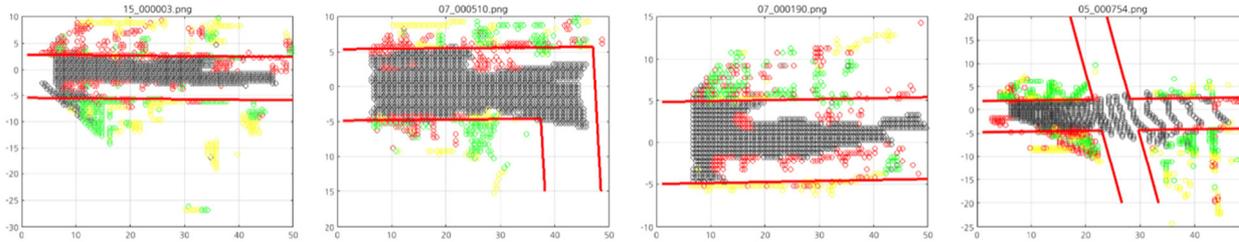


Fig. 9. Estimated road layouts: (from left) Straight, Left curve, Straight, Crossroad, Red, green, yellow, and black-colored circle in this figure correspond to 'OBJ', 'VEG', 'BLDG', and 'GND'. Red line represents the boundary of the estimated road layout

Wang의 방법이 OBJ 클래스와 BLDG 클래스를 구분하는 데 어려움이 있어, OBJ 클래스에 과적합된 결과를 도출하였기 때문에 해석할 수 있다. 따라서 BLDG 클래스의 재현율 값의 대폭 상승은 OBJ 클래스로 과적합되는 경향이 개선된 결과로 볼 수 있다. 마지막으로 Wang의 방법과 제안하는 접근법의 상대적인 계산량 차이를 비교하면 다음과 같다. 제안하는 기법은 사전위치정보의 활용을 위해 추가적인 계산을 필요로 하나 이는 Wang의 알고리즘의 계산량 대비 57%의 수준으로, 분류 성능의 개선 결과를 감안했을 때 용인할 수 있는 수준이다.

Fig. 9에는 도심 영상의 도로 모델 추정 결과를 보인다. 검증에 사용된 도심 영상 내 T-junction 모델을 ground truth로 갖는 영상이 존재하지 않아 Straight, Curve, Crossroad를 정답으로 갖는 영상에 대해서만 추정한 결과를 보인다. 상대적으로 GND 클래스는 많은 개수가 예측되는 반면 OBJ, VEG, BLDG 클래스에 대해서는 적은 개수의 공간상의 점이 예측된다. 도로 레이아웃의 영역 별 가중치를 설정할 때, 학습에 사용된 도심 영상 내 클래스 별 데이터 비율을 활용하여 VEG 및 BLDG 클래스는 적은 수가 예측되어도 큰 효과를 발휘하게 많은 가중치를 부여하는 방식으로 값을 설정하면 도로 모델의 매개변수를 빠르게 찾을 수 있다. MH 알고리즘의 반복 횟수는 2000번으로 설정하였고, 대체로 도로 모델

을 잘 추정함을 Fig. 9를 통해 확인할 수 있다.

## 5. 결론

본 연구에서는 사전위치정보를 활용하여 의미론적 분할 문제를 해결하였다. 도심 환경의 규칙적인 특성을 반영하여 사전위치정보를 학습하였고, 사전위치정보의 활용을 위해 다양한 유형의 도로 모델을 추정하였다. 추정된 도로 모델은 도로 방향에 수직인 도로법선좌표계를 설정하는 데 사용되었으며, 해당좌표계에 사전위치정보가 정의되었다. 제안하는 방법의 유용성을 검증하기 위해 KITTI 데이터 셋을 사용하였으며, 기존 알고리즘 대비 성능이 개선됨을 확인하였다.

## References

- [1] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *International Conference on Machine Learning (ICML)*, Williamstown, USA, pp. 282-289, 2001.
- [2] S. Divvala, D. Hoiem, J. Hays, A.A. Efros, and M. Hebert, "An empirical study of context in object detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, USA, pp.

1271-1278, 2009.

[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 34, no. 11, pp. 2274-2282, November, 2012.

[4] D. Hoiem, A.A. Efros, and M. Hebert, "Geometric context from a single image," in *IEEE International Conference on Computer Vision (ICCV)*, Beijing, China, pp. 654-661, 2005.

[5] P. Kohli, and P.H. Torr, "Robust higher order potentials for enforcing label consistency," *International Journal of Computer Vision (IJCV)*, vol. 82, no. 3, pp. 302-324, May, 2009.

[6] S. Gould, J. Rodgers, D. Cohen, G. Elidan, and D. Koller, "Multi-class segmentation with relative location prior," *International Journal of Computer Vision (IJCV)*, vol. 80, no. 3, pp. 300-316, December, 2008.

[7] S. Gould, R. Fulton, and D. Koller, "Decomposing a scene into geometric and semantically consistent regions," in *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan, pp. 1-8, 2009.

[8] L. Ladicky, C. Russell, P. Kohli, and P.H. Torr, "Graph cut based inference with co-occurrence statistics," in *European Conference on Computer Vision (ECCV)*, Crete, Greece, pp. 239-253, 2010.

[9] J. Bergbauer, C. Nieuwenhuis, M. Souiai, and D. Cremers, "Proximity priors for variational semantic segmentation and recognition," in *IEEE International Conference on Computer Vision Workshops (ICCVW)*, Sydney, Australia, pp. 15-21, 2013.

[10] C. Cadena and J. Kosecka, "Semantic segmentation with heterogeneous sensor coverages," in *IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China, pp. 2639-2645, 2014.

[11] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, October, 2001.

[12] J. Wang and J. Kim, "Semantic segmentation of urban scenes with enhanced spatial contexts," in *IEEE International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, Xian, China, pp. 263-266, 2016.

[13] A. Geiger, M. Lauer, C. Wojek, C. Stiller, and R. Urtasun, "3d traffic scene understanding from movable platforms," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 36, no. 5, pp. 1012-1025, May, 2014.

[14] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *International Journal of Robotics Research (IJRR)*, vol. 32, no. 11, pp. 1231-1237, September, 2013.

[15] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 23, no. 11, pp. 1222-1239, November, 2001.



**왕 정 현**

2012 한양대학교 미디어통신공학과 학사  
 2014 KAIST 로봇공학학제전공 석사  
 2014~현재 KAIST 로봇공학학제전공 박사  
 과정

관심분야: 영상처리, 의미론적 분할, 센서 융합



**김 진 환**

1993 서울대학교 조선해양공학 학사  
 1995 서울대학교 조선해양공학 석사  
 2007 Stanford University 항공우주공학 박사  
 1995~2000 선박해양플랜트연구소  
 2007~2010 Optimal Synthesis Inc. Research  
 Scientist  
 2010~현재 KAIST 기계항공공학부 부교수

관심분야: 이동로봇 및 항법제어