

# 영상 기반 자율적인 Semantic Map 제작과 로봇 위치 지정

## Vision-based Autonomous Semantic Map Building and Robot Localization

임정훈\*, 정승도\*\*, 서일홍\*\*\*, 최병욱\*\*\*

Jounghoon Lim\*, Seungdo Jeong\*\*, Il Hong Suh\*\*\*, Byung-Uk Choi\*\*\*

**Abstract** - An autonomous semantic-map building method is proposed, with the robot localized in the semantic-map. Our semantic-map is organized by objects represented as SIFT features and vision-based relative localization is employed as a process model to implement extended Kalman filters. Thus, we expect that robust SLAM performance can be obtained even under poor conditions in which localization cannot be achieved by classical odometry-based SLAM

**Key Words** : Semantic Map, Vision-based Localization, SIFT Feature Model, Extend Kalman Filter

### 1. 서론

SLAM(Simultaneous localization and map-building)은 어떤 위치에서 주변 환경의 지도를 작성하고, 작성된 지도를 바탕으로 다시 움직인 로봇의 위치를 알아내는 반복을 통해 로봇의 위치와 주변 환경의 맵을 동시에 추정하는 기술이다.

SLAM에 관한 연구들은 상대 위치 추정 기술을 위해 로봇의 주행기록계(Odometry)정보를 이용해 왔다. 하지만, 로봇의 주행 시스템의 형태나 방식에 따라, 주행기록계를 위한 시스템 설계가 어렵고 복잡하며, 실제 정확도에 굉장히 민감한 성능을 보인다. 이와 같은 단점을 해결하기 위하여, 본 연구의 사전 연구로서, 비전 기반 상대 위치 추정 알고리즘을 제안한 바 있다[3]. 비전 기반 상대 위치 추정 알고리즘은 주행 시스템에 대한 복잡한 설계 없이 로봇에 장착된 카메라의 모션 정보를 이용하는 방법으로 비교적 정확한 위치 추정이 가능하다.

SLAM에 관한 연구에서 또 하나의 중요한 연구 과제가 주변 환경에 대한 지도를 제작하고, 이를 이용해 로봇의 위치를 추정하는 것이다. 주위 환경에 대한 지도는 로봇의 위치 추정에 사용 되는 표식으로 이루어져 있다. Davison은 이런 표식을 위해, Harris 코너 검출기와 스테레오 매칭을 통해 얻어진 3차원 코너점을 사용하였고, Se는 SIFT(Scale Invariant Feature Transform)를 통해 얻어진 특징점을 사용하였다[1]. 하지만, 단순한 좌표와 이미지 상의 특징만을 표식으로 사용하는 방법은 사람과 로봇 간, 의사 교환이나 delivery task에 대한 요청과 처리에 도움을 주기 어렵다.

본 연구에서는 인식을 통해 얻어진 물체의 특징을 표식으로 사용함으로써, 사람과 로봇 간 상호작용에 기반이 될 수 있는 semantic map을 제작하고자 한다. 또한, 영상을 이용한 상대 위치 추정 방식을 확장형 칼만 필터(Extended Kalman Filter)를 위한 프로세스 모델로 사용함으로써, 로봇의 주행기록계 정보를 사용하기 어려운 환경에서도 강인한 SLAM 성능을 제공하고자 한다.

### 2.9 SFM을 이용한 로봇 위치 지정 및 Semantic Map 제작

본 연구에서는 제안하는 시스템의 구성도는 그림 1 과 같다.

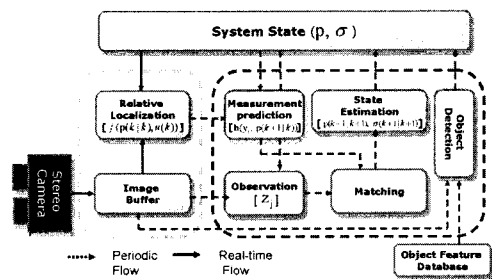


그림 1 시스템 구성도

제안하는 시스템은 스테레오 카메라를 이용한 영상 정보만을 센서 정보로 사용한다. 시스템은 크게 실시간으로 현재 로봇의 위치를 추정하는 부분(Realtime relative localization)과 SIFT를 이용한 물체 인식을 통해 표식을 생성하고 관찰하는 표식 인식 부분, 그리고, 확장형 칼만 필터를 이용해 두 측정 데이터를 융합하는 세 부분으로 이루어진다.

본 논문에서는 물체의 SIFT 특징 모델(SFM)을 통해 인식된 물체의 SIFT 특징점을 지도상의 표식으로 사용한다. 이로

### 저자 소개

- \*한양대학교 정보통신대학원 석사과정
- \*\*한양대학교 전자통신컴퓨터공학과 박사과정
- \*\*\*한양대학교 정보통신대학원 교수, 공학박사

써, 기존 연구와 달리 표식과 의미의 연결하는 추가적인 과정 (anchoring) 없이, 표식 등록과 동시에 표식에 물체의 의미를 부여하여 semantic map을 제작하고자 한다.

또한 SLAM을 위한 상대 위치 추정방식으로 로봇의 주행 기록계 정보가 아닌, 기존 연구를 통해 제안한 비전 기반 상대 위치 추정 기법 사용하고자 한다. 확장형 칼만 필터 기반 SLAM에 비전 기반 상대 위치 추정기법을 적용하기 위해 예측 모델과 측정값 예측 모델을 제안하고, 복잡한 주행 시스템 설계 없이 적용 가능한 SLAM 시스템을 실현하고자 한다.

### 2.1 시스템 상태 벡터와 공분산 행렬

로봇의 위치와 표식들의 불확실성 및 상호 간에 관계를 규정하기 위해, Davison이 제안한 시스템 상태 벡터(State Vector)와 공분산(Covariance) 행렬을 사용하였다[1].

$$p = \begin{pmatrix} x_v \\ y_1 \\ y_2 \\ \vdots \end{pmatrix} \sigma^2 = \begin{pmatrix} \sigma^2_{x_v, x_v} & \sigma^2_{x_v, y_1} & \sigma^2_{x_v, y_2} & \dots \\ \sigma^2_{y_1, x_v} & \sigma^2_{y_1, y_1} & \sigma^2_{y_1, y_2} & \dots \\ \sigma^2_{y_2, x_v} & \sigma^2_{y_2, y_1} & \sigma^2_{y_2, y_2} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (1)$$

본 논문에서 로봇의 상태는 평면에 위치한 것으로 가정하고, 위치와 방향은  $x_v = (x \ z \ \theta)^T$ 으로 나타냈다. 각 표식에 대한 상태 벡터는  $y_i$  과 같고, 물체의 SIFT 특징점을 이용해 인식한 물체의 특징점들의 절대 좌표를 나타냈다[2].

### 2.2 예측 모델

본 논문에서 예측 모델(process model)은 기존 연구를 통해 소개한 Pyramid Lucas-Kanade Optical Flow와 LM 알고리즘을 사용하는 상대 위치 추정을 통해 획득된 변위  $\Delta x$ ,  $\Delta z$ 와  $\Delta \theta$ 를 사용한다.

$$x_v(k+1|k) = f(x_v(k|k), u(k)) = \begin{bmatrix} x \\ z \\ \theta \end{bmatrix} + \begin{bmatrix} \Delta x_r \\ \Delta z_r \\ \Delta \theta \end{bmatrix}, \quad (2)$$

where  $\Delta x_r, \Delta z_r$  is

$$\begin{bmatrix} \Delta x_r \\ \Delta z_r \end{bmatrix} = \begin{bmatrix} \cos(\theta + \Delta\theta) & -\sin(\theta + \Delta\theta) \\ \sin(\theta + \Delta\theta) & \cos(\theta + \Delta\theta) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix}$$

$x_v(k|k)$ 와  $u(k)$ 는 각각 기존의 로봇의 위치와 로봇의 움직임을 나타내는 변위이다. 기존 위치의 카메라 좌표계를 기준으로 추정된 변위  $\Delta x$ ,  $\Delta z$ 를 지도상의 기준 좌표계로 변환한 변위가  $\Delta x_r, \Delta z_r$ 이다.

$$y_i(k+1|k) = y_i(k|k), \forall i \quad (3)$$

$y_i(k+1|k)$ 는 예측된  $i$ 번째 표식의 현재 위치를 나타내고, 식 (3)에서와 같이 움직이지 않는 표식을 가정한다.

$$\begin{aligned} \sigma^2_{x_v, x_v}(k+1|k) &= \nabla_{x_v} f \sigma^2_{x_v, x_v}(k|k) \nabla_{x_v} f^T + \nabla_u f \sigma^2_u(k|k) \nabla_u f^T \\ \sigma^2_{y_i, x_v}(k+1|k) &= \nabla_{x_v} f \sigma^2_{y_i, x_v}(k|k) \\ \sigma^2_{y_i, y_i}(k+1|k) &= \sigma^2_{y_i, y_i}(k|k) \end{aligned} \quad (4)$$

시스템에 대한 공분산은 식 (4)과 같이 예측되며,  $\nabla_{x_v} f$ 와  $\nabla_u f$ 는 각각 로봇 상태 벡터와 변위에 대한 예측함수  $f$ 의 Jacobian이다.

$$\sigma^2_u = \begin{bmatrix} C_x |\Delta x| + U_x & 0 & 0 \\ 0 & C_z |\Delta z| + U_z & 0 \\ 0 & 0 & C_\theta |\Delta \theta| + U_\theta \end{bmatrix} \quad (5)$$

$\sigma^2_u$ 은 카메라 모션 정보 추정 과정에서 발생하는 프로세스 노이즈이고,  $C_x, C_z$ 와  $C_\theta$ 는 각 변위에 따라 영향을 주는 여러 상수이고,  $U_x, U_z$  그리고  $U_\theta$ 는 변위와 상관없이 발생하는 비 시스템 에러를 나타내는 상수이다.

### 2.3 SFM을 이용한 표식 인식과 측정값 예측 모델

표식에 대한 인식은 물체에 대해 구성된 SFM을 이용하여 이루어지고, 인식된 표식의 3차원 좌표에 대한 측정값은 보정 (Calibration)된 스테레오 카메라를 이용하여 얻고, 이에 대한 측정값 예측 모델은 식(6)과 같다.

$$\begin{aligned} h_i(k+1) &= h(y_i, x(k+1|k)) = \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} \\ &= [R][y_i - x'_v], \end{aligned} \quad (11)$$

$$\text{where } x'_v = (x \ 0 \ z)^T, R = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}$$

$y_i$ 는 로봇의 시작 위치를 기준으로 하는 기준 좌표계에 대한 표식의 3차원 절대 좌표이고, 이를 카메라를 기준으로 하는 3차원 좌표로 변화시키는 함수가  $h_i$ 이다.  $x'_v$ 는 기준 좌표계에 대한 로봇의 평면 좌표를 나타내고, 회전  $R$ 은 기준 좌표계와 카메라 좌표계 간의 회전 변환을 나타내는 행렬로, 로봇의 각도 성분  $\theta$ 를 변수로 하는 행렬식으로 나타낼 수 있다.

### 2.4 시스템 상태 벡터 및 공분산 추정

$$\begin{aligned} K &= \sigma^2 \nabla_{x_v} h_i^T \cdot \sigma^{-2}_{IN} \\ &= \begin{pmatrix} \sigma^2_{x_v, x_v} \\ \sigma^2_{y_i, x_v} \\ \vdots \end{pmatrix} \frac{\partial h_i^T}{\partial x_v} \sigma^{-2}_{IN} + \begin{pmatrix} \sigma^2_{x_v, y_i} \\ \sigma^2_{y_i, y_i} \\ \vdots \end{pmatrix} \frac{\partial h_i^T}{\partial y_i} \sigma^{-2}_{IN} \end{aligned} \quad (12)$$

$$\begin{aligned} p(k+1|k+1) &= p(k+1|k) + K v_y(k+1) \\ \sigma^2(k+1|k+1) &= \sigma^2(k+1|k) - K \cdot \sigma^2_{IN}(k+1) \cdot K^T \end{aligned} \quad (14)$$

식 (7)을 이용하여 각각 표식에 대한 칼만 이득(Kalman Gain)을 계산한다.  $\sigma^2_{IN}$ 은 각 표식에 대한 이노베이션 공분산을 의미한다. 각각 표식에 대해 계산된 칼만 이득과 이노베이션 행렬을 식 (8)과 같이 시스템 상태 벡터에 적용하는 과정을 연속적으로 수행하여 갱신을 완료한다.

또한, 시스템 공분산 행렬에 대해서도, 상태벡터 갱신과정과 유사한 과정으로 식 (9)을 이용하여 연속적으로 수행한다.

### 2.5 표식 생성 및 삭제

인식된 물체에서 새로운 3D SIFT 특징점이 발견되었을 때, 측정값 예측 모델의  $y_n(x_v, h_n)$ 을 이용하여 절대 좌표를 계산하여 시스템 상태 벡터에 추가하고, Davison이 제안한 방식을 사용하여 공분산을 초기화한다[1].

표식에 대한 삭제 시에는, 공분산 행렬에서  $y_i$ 항을 포함하는 각 행과 열을 모두 삭제한다.

### 3. 실험 결과

#### 3.1 영상 기반 상대 위치 추정 성능 실험

상대 위치 추정에 대한 성능을 평가하기 위해 직선거리 이동 시 추정된 각도와 위치 성분을 측정하였다.

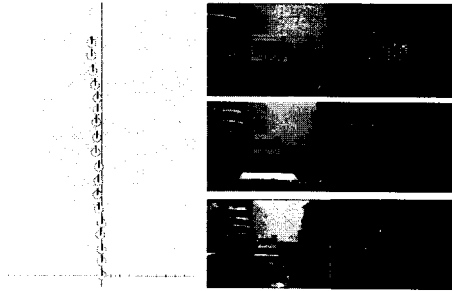


그림 2 상대위치 추정 성능 실험

그림 2 은 카메라를 900 millimeter를 직선으로 진행하고, 이에 대한 실제 경로와 추정된 경로를 나타낸다. 시작 위치를 기준으로 상대위치 추정 오류는 표 1과 같다.

표 1 상대 위치 추정에 대한 오류 측정

	PLK-optical flow and LM Algorithm		
	maximum	minimum	avg.
x	8.785	0.122	3.310
z	5.153	0.880	3.254
$\theta$	0.969	0.013	0.366

#### 3.2 물체 표식 기반 SLAM 성능 평가

##### 3.2.1. 확장형 칼만필터 위치 추정을 통한 보정 결과

본 논문에서 제안한 물체 인식 통한 표식 등록 과정은 실시간 실험이 어렵다. 따라서 본 논문에서는 물체 표식과 확장형 칼만 필터를 이용한 위치 지정 및 semantic map 작성을 비동기적으로 처리하였다.



그림 3 물체의 SFM

표식을 위해 사용되는 물체는 20도 간격으로 영상을 찍고, 물체 SIFT 특징점을 획득한다. 획득된 물체의 SIFT 특징점은 DB에 저장되고, 물체 인식을 위해 사용된다.

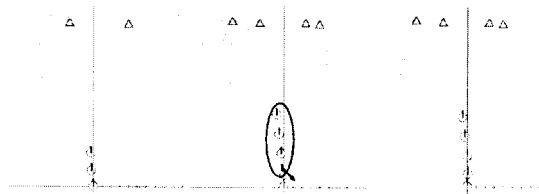


그림 4 확장형 칼만 필터를 이용한 보정 결과

실시간 상대 위치 추정을 수행 후, 3번째 프레임으로부터 물체 인식하고, 특징점을 등록하였다. 등록된 표식과 이전에 추정된 상대 위치 정보를 확장형 칼만 필터로 융합하여 그림 4와 같이 로봇의 위치를 다시 추정하여 수정하였다.

#### 3.2.2 Semantic Map 제작

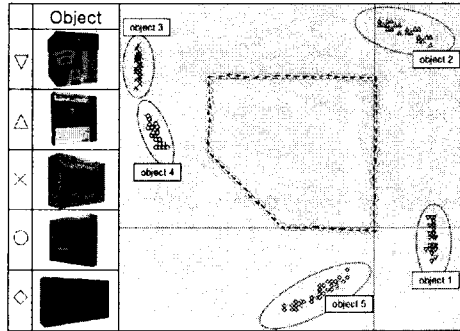


그림 5 Semantic Map 제작 결과

알려진 5개의 물체가 배치된 제한된 공간에서 제안하는 방법을 이용하여 지도를 제작하는 과정을 실험하였다. 각각 5개의 물체에 대한 영상과 제작된 지도는 그림 5와 같다. 지도에서 각 물체는 각각의 물체에 대한 표식으로 표시된다.

### 4. 결론

본 연구에서는 SIFT를 이용한 물체 인식을 통해 얻어진 물체의 특징을 표식으로 사용함으로써, 사람과 로봇 간 상호 작용에 기반이 될 수 있는 semantic map을 제작하였다. 또한 확장형 칼만 필터 기반 SLAM에서 영상을 이용한 상대 위치 추정 방식을 사용하기 위해 예측 모델과 측정값 예측 모델을 제안하여, 복잡한 주행 시스템 설계 없이 적용 가능한 SLAM을 실험하였다.

물체에 대한 영역 정보는 기존의 SLAM과 달리 주변 환경에 대한 지도 제작과 동시에 이루어진다. 이는 로봇에게 있어 새로운 능력을 부여할 수 있다. 첫째, 제안하는 방법을 사용하는 로봇은 알려지지 않은 환경에서 delivery tasks를 수행하면서, 동시에 다음 작업에 사용될지도 모를 물체에 대한 지도상의 배치 정보를 얻을 수 있다. 둘째, semantic map은 로봇 주위 환경에 대한 윤곽리지를 구축하는데 사용되어, path planning이나 task planning에 응용될 수 있다.

### 참 고 문 헌

- [1] A. J. Davison and W. Murray, "Simultaneous Localization and Map-Building Using Active Vision," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp.865-880, July, 2002
- [2] D. G. Lowe, "Distinctive Image Feature from Scale-Invariant Keypoints," *International Journal of Computer Vision*, pp.91-110, February, 2004
- [3] S. Jeong, J. Chung, S. Lee, I.H. Suh, B. Choi, "Design of A Simultaneous Mobile Robot Localization and Spatial Context Recognition System", *LNAI 3683 (KES2005)*, pp.945-952. 2005