

# 컨텍스트 이해를 통한 로봇의 작업을 위해 필요한 3D 작업공간 모델링

김은영\*, 이석한\*, 장대식\*, 한정현\*\*  
\*성균관대학교 정보통신공학부  
\*\*고려대학교 컴퓨터 과학과  
e-mail : samsam99@ece.skku.ac.kr

## 3D Workspace Modeling Based on Context Understanding for Robotic Manipulation

Eunyoung Kim\*, Sukhan Lee\*, Daesik Jang\*, JungHyun Han\*\*  
\*School of Information and communication engineering , SungKyunKwan University  
\*\*Dept. of Computer Science, Korea University

### 요 약

본 논문에서는 로봇이 작업을 계획하기 위해 필요한 3차원 작업 공간을 세 가지의 컨텍스트(context)들을 이해함으로써 빠르게 모델링하는 새로운 기법을 소개 하고 있다. 로봇이 사람과 비슷한 속도와 정확도로 작업 공간을 이해하고 모델링하는 것에 초점을 두고 있으며 이를 위해 작업 공간상의 특징적인 세 가지의 컨텍스트(작업공간의 간략화를 위한 전체 공간상의 평면특징, 데이터베이스에 미리 정의된 물체 그리고 로봇의 주어진 작업에 따라 다양한 상세함을 갖는 그 외의 장애물)를 정의하였고, 그것들을 빠르게 이해함으로써 어떻게 3차원 작업 공간을 형성하는지 설명하고 있다.

본 논문에서 3 차원 정보를 갖는 scale invariant feature transformation(SIFT)를 stereo-sis SIFT 로 간주했으며 이를 이용하여 위에서 언급한 컨텍스트들을 이해하였고 다양한 카메라의 위치로부터 얻어지는 여러 개의 장면들을 정합하였다.

또한, 실험을 통해 제안한 방법의 타당성도 검증하였다.

### 1. 서론

로봇 분야에서 환경 모델링은 지능형 서비스 로봇 또는 자체 이동 로봇이 이동이나 작업을 수행할 수 있는 환경 정보를 제공하는 것으로 예전부터 활발하게 연구되어 왔다. 대부분의 연구는 3 차원 range sensor 를 이용, 로봇의 주행이나 walk-through 를 위한 환경을 모델링하는 것에 초점을 두고 있다. [1][2][3][4][5] 그와 대조적으로 실시간으로 복잡하게 구성되어 있는 작업 공간을 모델링하는 것은 매우 어려운 일이었으며 거의 연구되어 온 바가 없었다.

본 논문에서는 스테레오 카메라로부터 얻어진 3 차원 점 정보를 이용하여 실시간으로 3 차원 작업환경을 구성하는 새로운 모델링 기법을 소개 하고 있다. 이를 위해 사람이 환경을 이해하기 위해 사용하는 기초적인 컨텍스트를 정의했으며 그 정보를 빠르게 이해함으로써 로봇이 작업하는데 필요한 3 차원 작업공간정

보를 만들어 주었다.

기본적인 컨텍스트는 세 가지이며 다음과 같이 정의되었다.

- 1) 공간상의 벽이나 바닥등 평면적 특성
- 2) 데이터베이스에 정의되어 있는 인식 및 조작해야 하는 물체
- 3) 주어진 작업에 따라 다양한 상세함을 갖는 기타 장애물

위에서 정의된 세가지 컨텍스트를 stereo-sis SIFT 를 이용하여 실시간으로 이해해 3 차원 작업 공간을 구성한다.

본 논문은 SIFT 기반 장면 정합(본문 2), 전체 공간 특징 추출(본문 3), 미리 정의된 다뤄야 할 물체 인식 및 위치 계산(본문 4), 장애물 표현 (본문 5) 그리고 실험 결과 (본문 6) 로 구성되어 있다.

2. SIFT 기반 장면 정합

스테레오 카메라로부터 얻어진 영상은 작업공간의 부분적인 정보만을 주게 된다. 그 부분적인 정보를 정합함으로써 하나의 전체적인 3 차원 작업공간을 만들어야 하는데 이 과정을 정합(registration)이라고 한다.

본 논문에서는 복잡한 작업 공간을 찍은 영상들을 빠르게 정합하기 위해 stereo-SIFT 를 사용하였다. 기존의 가장 널리 알려진 기법인 Iterative closed points(ICP) 는 점들 간의 간격을 바탕으로 대응되는 점을 찾아서 정합을 실행하는데 초기 위치를 알면서 outlier 가 적은 경우에만 적당하다. 잡음이 많고 초기 위치를 알지 못하는 경우 ICP 를 사용하기 힘들며 그것을 해결하기 위해 본 논문에서는 stereo-sis SIFT 를 이용하였다. 두 장면에서 대응되는 SIFT 들을 찾아 그 정보를 이용하였으며 그 결과 ICP 보다 적은 대응점을 이용하기 때문에 계산 시간을 줄 일수 있고 초기위치에 영향을 받지 않는다. 그림. 1 은 ICP 와 stereo-sis SIFT 기반 정합 방법에 대한 계산시간을 비교한 결과를 보여주며 그림. 2 는 stereo-sis SIFT 기반 정합 결과를 보여 주고 있다.

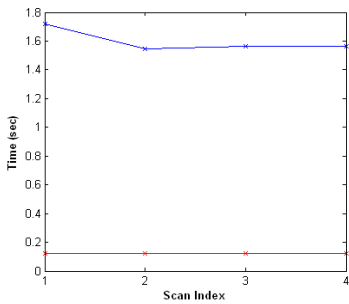


그림 1. ICP 와 Stereo-sis SIFT 정합 방법 계산 시간 비교(파란선:ICP, 빨강선: Stereo-sis SIFT)

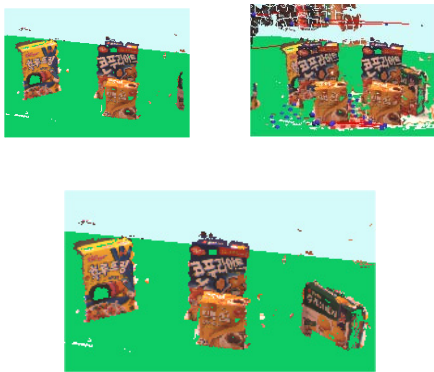


그림 2. 정합 결과

3. 전체 공간상의 평면 특징 추출

전체 공간상의 평면 특징을 추출하기 위한 방법은 크게 두 개의 단계로 구분된다. 첫 번째 단계는 일반적으로 사용되는 피팅 알고리즘인 RANSAC[6]을 이용하여 후보 평면 특징을 찾아 내는 것이다. 다음 단계는 검증 단계로 앞의 단계에서 구한 후보 평면들이 정말 전체 공간상의 평면을 의미하는 지를 검사하는 단계이다.

첫 번째 단계를 살펴보자. 계산시간의 향상을 위해 stereo-sis SIFT 를 RANSAC 에 적용하였다. 3 개의 SIFT ( $P_1, P_2, P_3 \in R^3$ )를 임의대로 선택한 후 평면의 노말  $N$  을 구한다. 주어진 평면 ( $N, P_1$ )을 가지고 전체 SIFT 점( $X$ )들에 대해 다음의 식(1)을 먼저 적용해 준다.

$$|N \cdot (X - P_1)| < \tau \quad (\|N\| = 1) \quad (1)$$

$\tau$ 는 사용자가 정의한 임계 거리다.

나아가 우리는 SIFT 점들의 방향성(orientation)도 고려한다. SIFT 는 영상에서 추출되며 최대 8 개의 이웃 점들을 갖게 된다. 이웃하는 점들을 이용하여 그림. 3 과 같이 해당 점의 방향을 구하게 된다. 이렇게 구한 방향이 식 (2)를 만족하는 지 검사한다.

$$M \cdot N > \sigma \quad (2)$$

$\sigma$  는 사용자가 정의한 임계치이며  $N$  은 평면의 노말이다. 위의 두 식(1),(2)을 만족하는 SIFT 점들의 수가 사용자가 정한 임계치보다 많으면 후보 평면 ( $N, P_1$ )가 추출된다.

두 번째 단계는 3 차원 점 정보를 이용하여 앞에서 구한 후보 평면들을 검증하는 단계이다. 3 차원 점 정보에 식(1)를 적용하여 만족하는 점의 수가 임계치보다 많다면 그 평면은 최종적인 전체 공간상의 평면 특징이 된다. 이 때 식(1)을 만족하는 점들은 이 평면을 구성하는 점으로 간주하고 그 점을 이용, 평면의 최적 피팅을 수행한다. 그 후, 해당 점들은 평면으로 대체된다.

카메라가 이동을 해서 새로운 장면을 얻게 되면 기존에 있던 평면 특징은 확대되면서 업데이트된다.

위의 과정을 통해 알아낸 평면 특징들은 무한한 크기를 가지고 있다. 그러나 작업공간을 모델링 후 작업 계획 등을 하기 위해서는 정확한 평면의 크기를 알아야 한다. 우리는 평면을 사각형 형태라고 가정하고 식 (1)을 통해 평면에 해당된다고 판단된 모든 점을 평면에 투영한 후 2 차원 oriented bounding box(OBB)를 구해서 크기를 정한다.

그림 4 는 평면 추출 결과를 보여 주고 있다.

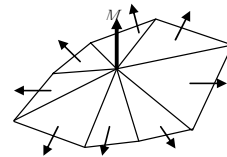


그림 3. SIFT 점의 방향

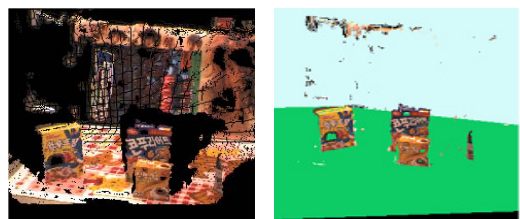


그림 4. 평면 특징 추출

#### 4. 물체 인식 및 위치 판단

물체 인식을 위해서 본 논문에서는 로봇이 조작해야 할 물체는 완전한 솔리드 형태이고 많은 텍스처 정보를 가지고 있으며 grasping 을 위해 국지적 접근 가능한 방향 정보를 가지고 있다고 가정하였다. 이 정보는 데이터 베이스에 저장되어 있으며 저장된 물체 즉 작업 해야 할 물체를 카메라로부터 얻은 장면에서 찾아서 그 물체의 위치를 계산, 미리 데이터 베이스에 정의되어 있던 물체를 환경에 가져다 놓는다.

본 논문에서 제안한 방법은 stereo-SIFT 를 이용한다. 이미 정의된 물체의 SIFT 와 카메라로부터 얻어진 장면에서의 SIFT 사이의 대응되는 정보를 이용하여 우리는 물체를 인식하고 위치도 정할 수 있다. 그러나 물체와 장면의 모든 SIFT 를 이용하여 대응점을 찾게 되면 오랜 시간이 필요하다. 그러므로 우리는 앞에서 구한 평면 정보를 이용, 인식 해야 할 물체가 있을 가능성이 높은 곳을 지정, 그곳에서 대응점을 찾게 함으로써 계산 시간을 줄여주었다. 이때 우리는 물체들이 책상이나 바닥 등에 놓여 있다고 가정하였다. 평면 특징으로부터 일정 거리 이상 떨어진 SIFT 정보는 물체 인식 계산에 불필요하다고 판단, 제거해 준다. 그림. 5 에서 평면을 이용, 어떻게 searching space 가 정해지는지 보여주고 있다.

비록 SIFT 가 다른 특징점 들에 비해 강건하다고 하나 간혹 잘못된 대응점이 생성될 수 있다. 그 경우 물체 인식을 해서 위치를 정할 경우 물체가 비뚤어질 수 있다. 그 문제를 해결하기 위해 우리는 삼각형 제한(triangle constraint)를 두었다. 대응되는 SIFT 를 삼각화하여 장면의 삼각형 정보와 물체의 삼각형 정보가 일치하는지를 검사, 일치하는 경우 제대로 대응되었다고 본다. 그림. 6 는 삼각형 제한을 이용한 모습을 보여준다.

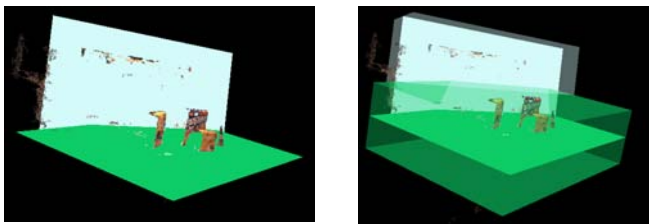


그림 5. 평면특징을 이용한 searching space 제한



그림 6. 삼각형 제한을 안 한 경우와 한 경우 비교: 삼각형 제한을 사용하면 잘못된 대응점은 제거된다.



그림. 7 물체 인식 및 위치 계산 결과

물체 인식 할 때 대응점의 수가 3 개 이하면 인식은 실패한다. 만약 3 개 이상이라면 그 물체는 인식된 것이고 Horn[7]의 쿼터니온 기반 방법을 이용하여 DB 에 정의된 물체를 3D 작업 공간에 추가시켜 주었다. (그림 7)

#### 5. 작업 중심 기타 장애물 표현

카메라로부터 얻은 3 차원 점 정보를 이용하여 우리는 물체를 인식하고 평면 특징을 추출하여 3 차원 작업환경을 구했다. 그러나 작업 계획 등을 위해 평면이나 물체로 표현되지 않은 나머지 장애물들에 대한 표현이 필요하다.

또한, 우리의 목적은 복잡한 환경에서 로봇이 작업할 수 있도록 3 차원 정보를 제공하는 것이기 때문에 이를 위해서는 작업할 물체 주변의 장애물 정보에 대해서는 자세히 표현해줘야 한다. 그렇기 때문에 우리는 Octree 와 OBB 표현 형태를 이용하였다.

Octree 는 전체 공간을 포함여부를 따져가면서 공간을 8 개씩 분할, 형성되며 형성된 정보는 같은 크기의 상자로 나타내진다. OBB 는 점의 분포 계산 통해 점의 분포에 가장 잘 맞는 상자를 의미한다.

먼저, Octree 를 이용하여 전체 공간상의 장애물을 거칠게 표현한다. 그 후, 물체가 인식이 되면 물체 주변의 Octree 를 좀더 상세하게 표현하기 위해 OBB 를 이용한다. 이때 OBB 는 다양한 상세함을 갖게 된다. 즉, OBB 가 생성되고 그 OBB 가 항상 점들로 잘 차여 있는지를 검사한다. 만약 OBB 내의 점이 잘 차여 있으면 그 OBB 는 더 이상 분할될 필요가 없다. 그러나 만약 OBB 가 점들로 제대로 차여 있지 않다면 차질 때까지 계속 분할 된다.

그림. 8 은 다양한 상세함의 OBB 들을 보여 주고 있으며 그림에서 보듯이 평평한 부분은 적은 수의 OBB 로 표현되고 구부러진 부분은 많은 수의 OBB 를 표현, 다양한 상세함을 가지면서 물체를 상세하게 표현해준다.

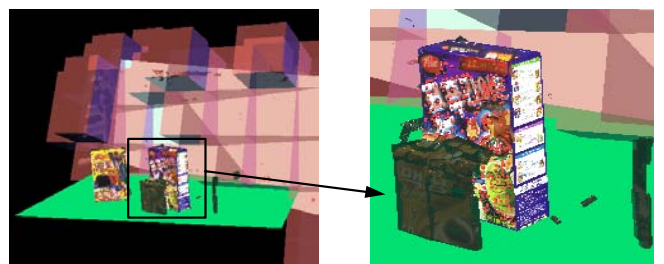


그림 8. Octree 와 다양한 상세함을 가진 OBB

6. 실험 결과

본 실험에서는 스테레오 카메라가 텍스처가 없는 환경에서 3 차원 정보를 주지 못하기 때문에 작업환경이 텍스처가 많은 환경이라고 가정하였다. 그림 11의 스테레오 카메라(왼쪽)를 사용하였으며 카메라는 로봇 팔 위에 장착되어 있다. 로봇 팔이 움직이면서 계속적으로 장면을 찍어내게 된다. 그림 11의 가운데 그림은 실제 작업환경 및 로봇 팔을 보여 주고 있으며 오른쪽 그림은 데이터베이스에 미리 정의된 물체를 보여 주고 있다. 물체에 보여지는 빨간 점은 stereo-sis SIFT 를 의미한다.

이 실험은 Pentium 4 2.8GHz PC 에서 수행되었으며 5 개의 장면을 모두 정합하여 세가지 컨텍스트를 이용하여 3 차원 작업환경을 계산하는데 평균적으로 3.72 초가 걸렸다. 이 결과는 전의 모델링 방법보다 훨씬 빠른 결과로 실시간에 로봇 작업을 수행 가능하도록 3 차원 작업 환경을 로봇에게 전달 할 수 있는 가능성을 보여 주는 것이다. 표 1 은 각 단계의 평균 계산 속도를 보여 주고 있다.

그림 13 은 복잡한 환경을 모델링 했을 때의 최종 결과를 보여준다.

7. 결론

본 논문은 다양한 컨텍스트를 정의, 빠르게 이해함으로써 작업공간을 모델링하는 알고리즘을 보여 주고 있다. 제안된 방법은 복잡하게 구성된 환경에서 실험을 했으며 그 결과 실시간으로 로봇이 작업할 수 있도록 3 차원 작업 환경을 구성할 수 있다는 것을 보여 주었다.

향후에는 데이터베이스에 정의되는 물체를 더 다양화하고 SIFT 이외의 3 차원 invariant feature 및 제안한 3 차원 작업 공간을 이용하여 로봇이 작업을 수행하는 것과의 결합을 할 것이다.



그림 11. 실험 환경. 스테레오 카메라, 실험 환경, 데이터베이스에 저장되어 있는 물체들

표 1. 성능 평가

	Avg. time (ms)
SIFT feature calculation	178
Planar feature extraction	21
Object recognition (3 object models)	342
Obstacle representation (Octree + multi-resolution OBBs)	75
Scene registration	183
<b>Total</b>	<b>799</b>

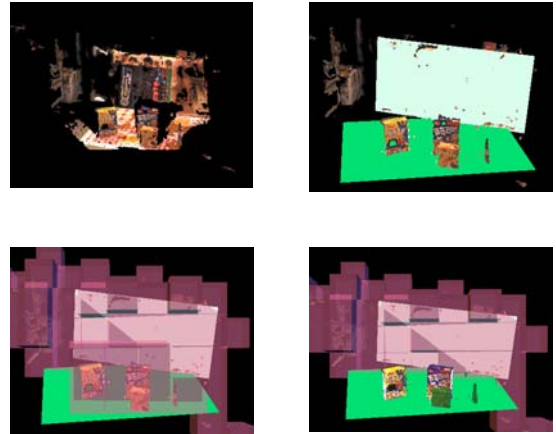


그림 13. 제안된 방법을 이용해서 모델링된 작업환경

ACKNOWLEDGMENT

이 논문은 과학 기술부에서 추진하는 21 세기 프론티어 과제 중에 하나인 지능형 로봇 개발 프로그램으로부터 지원을 받았다.

참고문헌

- [1] M. A. Garcia and A. Solana, "3D Simultaneous localization and modeling from stereo vision," International Conference on Robotics & Automation (ICRA'04), New Orleans, LA, April 2004
- [2] A. Nüchter, H. Curmann, and J. Hertzberg, "Automatic model refinement for 3D reconstruction with mobile robots," 3-D Digital Imaging and Modeling, 2003.
- [3] D. Huber, O. Carmichael and M. Hebert, "3-D Map reconstruction from range data," In proceedings of the IEEE Conference on Robotics and Automation, 2000.
- [4] Y. Liu, R. Emery, D. Chakrabarti, W. Burgard, and S. Thrun, "Using EM to learn 3D models of indoor environments with mobile robots," In proceedings of the 18<sup>th</sup> Conference on Machine Learning, Williams College, July 2001.
- [5] D. Hahnel, W. Burgard and S. Thrun, "Learning compact 3D models of indoors and outdoor environments with a mobile robot," Journal Robotics and Autonomous systems, vol. 44, 2003.
- [6] M.A. Fischler and R.C. Bolles, "Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography," Commun. Assoc. Comp. Mach., vol. 24, pp. 381 – 395, 1981.
- [7] P. Besl and N. McKay, "A method for registration of 3D shapes," IEEE Transactions on Pattern Analysis and Machine Intelligent, vol. 14, pp. 239-256, 1992