

지능로봇을 위한 3차원 환경인식 ☆

Recognition of 3D Environment for Intelligent Robots

장 대 식*
Daesik Jang

요 약

본 논문에서는 지능로봇에 활용될 수 있는 3차원 환경과 여기에 포함된 물체의 실시간 인식을 위한 새로운 접근 방법을 제안한다. 본 논문에서는 먼저 사람이 환경을 인식하고 상호작용하는 데 사용하는 3 가지 기본 원칙을 설정하고, 이 기본 원칙들을 이용하여 실시간 3차원 환경 및 물체 인식을 위한 통합된 방법을 제시한다. 이들 3가지 기본 원칙은 다음과 같다. 첫째, 전역적인 평면 특징들을 인식함으로써 작업환경의 기하적 구조에 대한 개략적 특성화를 고속으로 진행한다. 둘째, 작업환경 속에서 기존에 알려진 물체를 먼저 빠르게 인식하고 이를 데이터베이스 내에 저장되어 있는 물체의 모델로 교체한다. 셋째, 다중 해상도 Octree 표현 방법을 이용하여 기타 영역을 주어진 작업의 필요에 따라 적응적으로 실시간으로 모델링 한다. 본 논문에서는 3차원 SIFT 로 언급되는 3차원 좌표를 가지는 SIFT 특징들을 3차원 좌표정보와 함께 확장하여 사용함으로써 전역적 평면 특징의 빠른 추출, 고속의 물체 인식, 빠른 장면 정합 등의 기능에 활용하고 이와 동시에 스테레오 카메라로부터 입력되는 3차원 좌표의 잡음과 불완전성을 극복한다. 실험 결과에서는 본 논문에서 제안하는 지능형로봇의 조작 작업을 위한 실시간의 행위 중심의 3차원 환경모델링의 가능성을 보여준다.

Abstract

This paper presents a novel approach to real-time recognition of 3D environment and objects for intelligent robots. First, we establish the three fundamental principles that humans use for recognizing and interacting with the environment. These principles have led to the development of an integrated approach to real-time 3D recognition and modeling, as follows: 1) It starts with a rapid but approximate characterization of the geometric configuration of workspace by identifying global plane features. 2) It quickly recognizes known objects in environment and replaces them by their models in database based on 3D registration. 3) It models the geometric details on the fly adaptively to the need of the given task based on a multi-resolution octree representation. SIFT features with their 3D position data, referred to here as stereo-sis SIFT, are used extensively, together with point clouds, for fast extraction of global plane features, for fast recognition of objects, for fast registration of scenes, as well as for overcoming incomplete and noisy nature of point clouds. The experimental results show the feasibility of real-time and behavior-oriented 3D modeling of workspace for robotic manipulative tasks.

☞ Keyword : Intelligent Robot, Environment Recognition, SIFT, 3D Object Recognition, 지능로봇, 환경인식, SIFT, 3차원 물체인식

1. 서 론

1.1 연구 배경

본 논문에서는 스테레오 카메라를 이용하여 지능형로봇의 조작 작업을 위한 실시간의 3차원 환

경 인식 문제를 해결할 수 있는 방법을 제안한다. 이를 위하여 전역적 공간 특징들과 조작 목표 물체의 인식에기반한 새로운 실시간의 작업 공간 모델링 방법이 제안되었다.

환경의 인식 및 모델링 문제는 많은 연구 그룹들에서 적극적으로 연구되어 왔다[1][2][3][4]. Garcia와 Solana는 6 자유도 이동로봇의 궤적을 결정하고 동시에 작업환경의 3차원 모델을 구축하기 위한 스테레오 비전 기반의 알고리즘을 제안하였다[5]. Huber와 그 동료들은 3차원 공간상

* 정 회 원 : 국립군산대학교 컴퓨터정보과학과 교수
dsjang@kunsan.ac.kr

[2006/04/24 투고 - 2006/04/25 심사 - 2006/06/13 심사완료]

☆ 본 연구는 정보통신연구진흥원 선도기반기술개발사업의 지원을 받아 수행되었음

에서 한 점의 지역적 모양 정보를 계산할 수 있는 Spin-image 특징을 제안하고 이를 환경모델링에 적용하였다[6]. Liu와 Nüchter는 환경을 개략화하고 개선하며 3차원 좌표의 오차를 보완할 수 있는 환경 모델링 알고리즘을 제안하였다[7][8]. 그러나 이러한 알고리즘들은 주로 로봇의 조작 작업보다는 주행에 초점을 맞추고 있으며 조작을 위한 정확하고 빠른 작업 공간의 모델링에 관한 연구는 아직 미흡한 실정이다.

위에서 설명한 3차원 환경 인식 및 모델링 방법들은 지능로봇, 지능형 자동차와 같은 실시간 응용분야에 적합한 실시간성 보다는 보다 정확하고 섬세한 모델링에 집중되어 있다[9][10]. 그러나 실시간 응용 기술을 위해서는 상세도와 정확도에 대한 요구사항이 주어진 작업에 상대적이어야 하는 반면 실시간의 요구사항이 더욱 급하고 중요하다.

본 논문에서는 이러한 실시간 응용에서의 요구사항들을 만족시키기 위한 새로운 3차원 환경인식 및 모델링 방법을 제안한다. 이를 위해 사람이 환경을 이해하기 위해 사용하는 근본적인 원칙들을 분석하여 이를 적용하고, 지역적 불변특징을 3차원 정보의 처리를 위한 핵심 특징으로 사용하여 실시간성을 증대시키면서 인식 및 모델링의 정확도와 잡음에 대한 강건성을 향상시킨다.

2. 본 논문에서 제안된 방법

본 논문에서는 3차원 환경의 효과적인 인식과 모델링을 위해 인간이 환경을 인지하고 이해할 때 사용하는 기본적인 원칙들을 관찰하고 이를 기본적인 가정으로 이용한다. 그리고 이러한 기본적인 원칙들을 이용하여 보다 빠르고 자연스럽게 3차원의 환경을 인식하고 모델링할 수 있는 방법을 제시한다. 이를 위해 본 논문에서는 인간이 환경을 인지할 때 사용하는 다음의 세 가지 기본 원칙을 가정하고 이를 적용한다.

- 1) 인간은 새로운 환경을 처음 접하였을 때 상세한 부분에 대한 정보를 인지하고 인식하기 이전에 먼저 벽이나 바닥, 천정과 같은 전반적이면서 개략적인 전역적인 특징들을 개략적이면서 빠르게 인식하고 이를 이용하여 상세한 부분들을 인식한다.
- 2) 인간은 환경 속에 학습되지 않은 새로운 물체가 있다면 이를 깊이 관찰하여 인식 및 모델링을 수행하여야 하지만, 이미 학습하여 알고 있는 물체가 있다면 그 물체를 오래 관찰하지 않고도 물체의 인식과 정합을 통해 빠르게 주변 환경과 물체를 결합하여 인지할 수 있다.
- 3) 인간은 관심의 대상이 되는 부분에 대해서는 자세하게 관찰하고 이를 인식 및 모델링하여 정보화하지만, 시야에 포함되는 비관심 영역에 대해서는 상세한 인식과 모델링 없이 개략적인 형태로 정보화하여 인지한다. 즉, 관심의 정도에 따라 환경을 인식하고 표현하는 상세도가 달라진다.

본 논문에서는 위에서 관찰한 세 가지의 기본 가정들을 통합함으로써 다양한 실시간의 응용 분야에서 보다 빠르고 자연스럽게 3차원의 환경을 인식하고 모델링할 수 있는 방법을 제안하는 것을 목표로 한다. 이를 위해 다음과 같은 구체적인 방법으로 위에서 설명한 원칙들을 적용한다.

- 1) 먼저 3차원 환경의 개략적이면서 빠른 특성화를 위해 본 논문에서는 3차원 공간 속에서 3차원의 평면들을 인식하여 사용한다. 빠르고 강건한 처리를 위해서 스테레오 카메라에서 획득된 3차원의 점들 뿐 아니라 3차원 좌표를 포함하는 3차원 SIFT 특징점들을 함께 사용하여 효율적으로 3차원 평면 특징들을 추출한다. 이렇게 추출된 3차원 평면 특징들은 물체정합을 위한 탐색공간을 줄여 속도를 향상할 뿐 아니라 장면 정합의 처리

속도를 향상시켜 실시간의 처리가 가능하도록 한다.

2) 빠른 물체의 인식과 3차원 정합을 위해 사전에 알려진 물체들의 3차원 특징들을 학습하여 물체 데이터베이스에 저장한다. 그리고 실시간에 물체를 인식하고 이들의 3차원 자세를 함께 정합하여 전역적 공간특징들과 통합함으로써 통합된 3차원 환경 모델을 신속하게 생성한다.

3) 관심의 대상이 되는 영역을 중심으로 하는 실시간의 주변 환경 모델링을 위해 전역적 공간특징이나 알려진 물체로 인식되지 않은 영역들은 다중 해상도를 가지는 Octree 표현을 이용하여 모델링한다. 이때 Octree로 표현되는 3차원 영역들은 관심의 대상이 되는 영역 주변에서는 높은 해상도를 가지며 그 외의 영역에서는 낮은 해상도를 가지도록 표현된다. 즉, Octree 셀의 크기가 관심의 대상이 되는 영역 주변에서는 매우 작게 표현되고 기타 영역에서는 크게 표현된다. 본 논문에서는 관심의 대상을 인식된 물체로 보고 인식된 물체의 주변에서 높은 해상도를 갖도록 Octree를 표현한다.

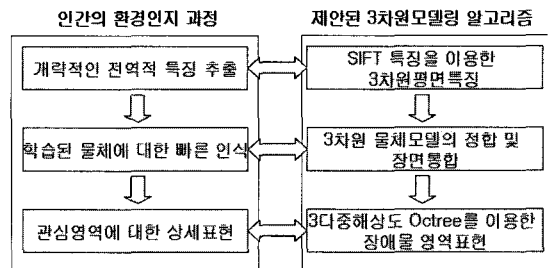
위의 세 단계에 걸쳐 전반적으로 3차원 공간상의 SIFT 특징들이 핵심 특징으로 사용되고 있다. 3차원 평면의 추출 시 뿐 아니라, 물체의 인식 및 정합, 장면 정합 등의 주요 단계에서 SIFT 특징점들을 사용함으로써 알고리즘의 속도를 향상시키고 획득된 3차원 좌표가 가지는 잡음과 불완전성을 보완한다. (그림 1)에서는 위에서 설명한 세 가지 기본 원칙에 기반하여 인식되고 모델링된 3차원 환경 모델의 예를 보여준다.

(그림 2)에서는 인간의 환경인지과정과 이에 기반하여 본 논문에서 제안하는 3차원 환경인지과정의 관계를 설명한다. (그림 3)은 3차원 환경인지에 기반하여 환경모델링을 수행하기 위한 시스템의 구성도를 보여준다.

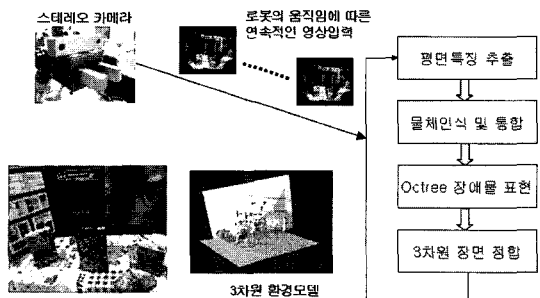


(a) 모델링할 실제 환경 (b) 인식된 3차원 모델

(그림 1) 생성된 3차원 환경모델



(그림 2) 인간의 환경인지과정에 기반한 모델링 과정



(그림 3) 환경모델링을 위한 시스템 구성도

본 논문은 다음과 같이 구성된다. 3장에서는 전역적 특징의 추출 방법에 대해 설명하고 4장에서는 조각의 대상이 되는 물체의 인식 방법을 설명한다. 그리고 5장에서는 기타 영역에 대한 다중 해상도의 Octree 표현 방법을 설명하고 6장에서는 3차원 SIFT 특징과 평면특징을 함께 이용한 3차원의 장면 정합 방법에 대해 설명한다. 7장에서 실험 결과를 보여주고 8장에서 결론을 설명한다.

3. 환경의 전역적 구조에 대한 빠른 특성화

본 논문에서 3차원 공간상의 평면들(평면 특징)을 전역적 공간 특징으로 사용된다. 제안된 평면 추출 알고리즘은 두 단계로 이루어진다. 첫 번째 단계에서는 RANSAC (Random Sampling Consensus) 알고리즘을 사용한다[11]. RANSAC 알고리즘은 하나의 모델을 실험 자료 집합에 적합시키고 평면들의 후보 집합을 생성한다. 이 평면 집합은 검증을 위해서 두 번째 단계로 입력된다. 그리고 두 번째 단계에서 검증된 후보들만이 평면으로 수용된다.

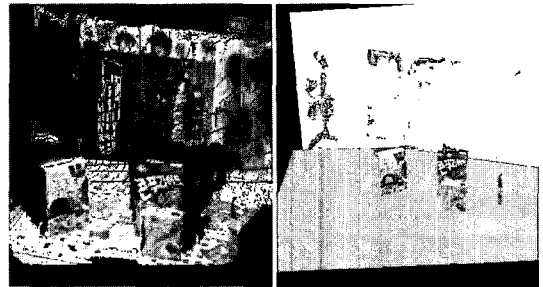
첫 번째 단계에서 RANSAC 알고리즘은 3차원 SIFT 특징들만을 처리한다. 세 개의 3차원 SIFT 특징 점들($P_1, P_2, P_3 \in R^3$) 이 임의로 선택되고 평면의 법선벡터 N 이 예측된다.

$$|N \cdot (X - P_i)| < t \quad (||N|| = 1) \quad (3.1)$$

여기서 t 는 사전에 정의된 임계값으로 본 논문에서는 실험을 위해 0.01m를 사용하였다.

제안된 평면 추출 알고리즘의 두 번째 단계에서는 식 (3.1)을 사용하여 장면 속의 모든 3차원 점들에 대해 위치 테스트를 수행한다. 만약 식 (3.1)을 만족하는 3차원 점들의 수가 임계치를 넘어가면 최적의 평면이 3차원 점들로 계산되고 이들이 작업공간의 전역적 공간 특징으로 선택된다. 만약 임계값 t 가 잘못 설정되면 평면이 안정적으로 추출되지 못하거나 평면이 아닌 영역이 평면으로 추출되는 오류가 발생할 수 있다.

새로운 카메라 위치로부터 새로운 장면이 획득되면 이전의 평면 집합이 갱신되고 새로운 점들이 추가된다. 새로운 3차원 점들로부터 추출된 새로운 SIFT 특징들이 현재의 평면들에 대해 테스트 되고 식 (1)을 만족하는 특징점들이 평면들을 갱신한다. 반면 현재의 평면들에 포함되지 않는 SIFT 특징점들은 두 단계의 평면 추출 과정을 통해 새로운 평면을 구성할 수 있다. 평면 추출 알



(a) 3차원 점의 집합 (b) 추출된 평면들

(그림 4) 추출된 평면을 포함하는 장면

고리즘의 시간 복잡도는 $O(N_p)$ 이고 여기서 N_p 는 장면 내의 3차원 점들의 개수이다.

평면 특징으로 추출된 영역에 포함되는 3차원 점들은 추출된 평면으로 대체된다. (그림 4)는 추출된 평면들과 평면에 포함되지 않고 남아있는 3차원 점들을 나타낸다. 다음 과정에서 남아있는 점들의 일부는 조작의 대상이 되는 물체로 인식될 수 있다. 여기서 추출된 평면들은 전역적인 공간 특징으로 사용될 뿐 아니라, 평면 영역에서는 물체를 인식할 필요가 없기 때문에 물체 인식을 위한 탐색 공간을 줄여 전체적인 성능을 향상시키는 역할을 한다.

4. 작업환경과 물체의 통합

본 논문에서는 모든 조작의 대상이 되는 물체들은 물체 데이터베이스 내에 완벽한 3차원 모델을 가지고 있다고 가정한다. 그리고 이 데이터베이스는 각 물체에 대한 지역적 불변 특징들3차원 구조에 대한 충분한 정보를 가지고 있다고 가정한다. 조작의 대상이 되는 물체들을 인식하기 위해 전체 3차원 환경이 탐색 공간으로 정의 되고 만약 3차원 점들의 집합의 일부분이 목표 물체로 인식이 되면, 이 부분은 데이터베이스 내에 있는 목표 물체의 3차원 모델로 대체된다. 물체 인식 알고리즘은 다음과 같은 과정으로 수행된다.

- 1) 물체 인식을 위한 탐색 공간의 결정
- 2) SIFT 특징점들의 정합
- 3) 정합된 SIFT 특징점들의 검증
- 4) 인식된 목표 물체의 3차원 모델을 3차원 환경 모델과 통합

첫 번째 단계는 반드시 수행할 필요는 없지만 간단한 경험적 방법을 통하여 탐색 공간을 상당히 축소시킬 수 있다. 예를 들어, 모든 물체들이 책상이나 탁자와 같은 평면 위에 놓여 있다고 가정해 보자. 즉, 어떤 물체가 다른 물체의 위에 놓여 있거나 허공에 놓여있을 수는 없다. 이 가정을 사용하기 위해서 환경 상의 3차원 SIFT 특징점과 추출된 평면 사이의 거리가 계산된다. 만약 이 거리가 모든 평면에 대해서 목표물체의 높이보다 크다면, 이러한 SIFT 특징점들은 인식의 대상에서 제외된다. 여기서 제외되지 않고 남은 SIFT 특징점들이 물체 인식의 탐색 공간을 구성하게 된다.

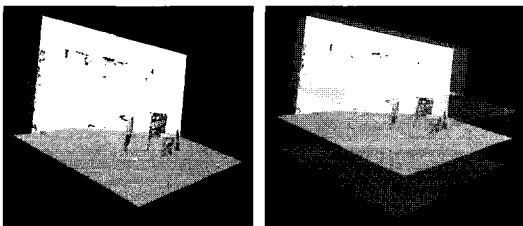
(그림 5)는 축소된 탐색 공간을 보여준다. (물체 인식 과정은 이러한 탐색 공간의 축소를 통하여 많은 잇점을 가지게 된다. 예를 들어 전체 공간에 대해 물체 인식 처리 시간이 375ms가 소요되는 경우 축소된 탐색 공간 내에서는 234ms 가 소요되었다.)

전통적으로 많은 응용 분야에서 다양한 불변 특징들이 사용되어 왔다. 최근에는 SIFT 특징점들이 2차원 영상에서의 물체 인식에 성공적으로

사용되었다[2]. 기존의 알고리즘들은 2차원 영상에서의 불변 특징에 기반을 두고 있다. 즉, 3차원 정보는 인식 과정에서 사용되지 않았다. 다른 연구에서 목표 물체의 3차원 모델 표면 상에 존재하는 각 SIFT 특징점이 전통적인 SIFT 정합 알고리즘을 이용하여 탐색공간의 SIFT 특징점들과 정합되었다[2]. 그러나 본 논문에서는 전통적인 SIFT 정합 알고리즘과 달리 3차원 위치정보를 가지는 SIFT 특징점들이 물체의 인식을 위한 주된 특징으로 사용되었다.

SIFT 정합 과정의 결과로 정합된 3차원 SIFT 특징점 쌍들의 집합이 만들어 진다. 이 SIFT 특징 쌍들은 잡음의 영향을 최소화하기 위해 3차원의 기하적 제약조건을 사용하여 재검증 절차를 거치게 된다. 본 논문에서는 이러한 3차원의 기하적 제약조건을 “3차원 삼각제약” 이라고 부른다. 만약 정합된 SIFT 특징 쌍의 개수 N 이 2보다 크다면, ${}_N C_3$ 개의 삼각형이 생성될 수 있다. 3차원 삼각제약 조건은 3개의 정합된 SIFT 특징 쌍들에 기반한 2개의 3차원 공간상의 삼각형들(물체모델에 존재하는 삼각형과 획득된 3차원 환경상에 존재하는 삼각형)이 일정한 오류 범위 내에서 서로 동치이어야 한다고 제약한다. 만약 SIFT 특징 쌍들이 정확하게 정합되었고 잡음이 포함되지 않았다면 이 점들이 주변 점들과 함께 이루는 삼각형들 또한 서로 동치이어야 한다는 점에서 이러한 제약 조건의 적용이 가능하다. 이러한 3차원 삼각제약을 만족하는 경우에만 삼각형을 이루는 각 SIFT 특징점들은 의미 있는 정합으로 검증된다. (그림 6)은 이 제약조건을 사용하지 않은 경우와 사용한 경우의 각각의 정합 결과를 비교하여 보여준다.

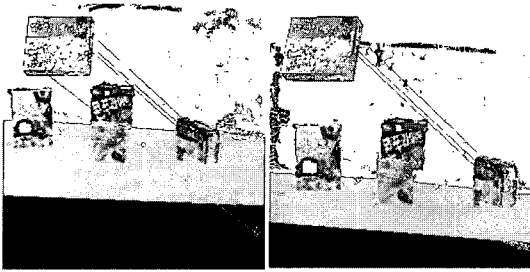
본 논문에서 제안하는 물체 인식 알고리즘의 시간 복잡도는 $O(M_s N_s)$ 이고, 여기서 M_s 는 물체 모델에 존재하는 SIFT 점들의 개수이고 N_s 는 감소된 탐색 공간 내에 존재하는 SIFT 점들의 개수이다.



(a) 추출된 평면들

(b) 탐색 공간

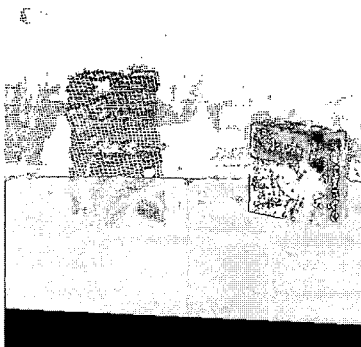
(그림 5) 목표 물체에 대한 축소된 탐색 공간



(a) SIFT 정합 (b) 제약조건 적용
(그림 6) 물체 인식 결과

만약 정합된 3차원 SIFT 특징 쌍의 개수가 3보다 작으면 목표 물체의 인식은 실패한 것으로 판단하고, 그렇지 않으면 3차원의 완전한 물체 모델의 자세는 Hom의 알고리즘을 이용하여 정합된 환경 속의 SIFT 특징의 위치로 변환된다[3]. 일단 3차원 물체 모델이 3차원 환경에 결합되면 그 물체에 포함되는 3차원 점들은 작업 공간에서 모두 제거되어 3차원 모델로 대체되게 된다. (그림 7)은 이러한 예를 보여준다.

그림에서 보는 바와 같이 여기서 한 가지 주목할 만한 사실은, 3차원의 환경이 서로 다른 유형과 수준의 표현 방법들을 통해 통합되어 모델링된다는 것이다. 즉, 결합된 완전한 3차원 물체 모델은 가장 상세한 수준의 정보를 표현하고 있는 반면 전역적 공간 특징인 평면 특징은 가장 단순화된 환경 정보를 표현하고 있다.

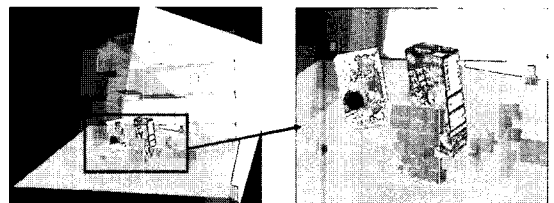


(그림 7) 목표 물체의 3차원 모델과 작업 공간의 통합

5. Octree를 이용한 다중 해상도 환경 표현

이제 추출된 평면들에 포함되지 않고 또한 인식된 목표 물체에도 포함되지 않는 3차원 점들의 집합은 인식이 불가능한 기타 영역으로 분류되어 인식의 대상은 아니지만 공간의 일부를 차지하는 불명확한 영역으로 표현된다. 본 논문에서는 이들 기타 영역들을 Octree를 이용하여 계층적으로 표현한다. 여기서 Octree는 3차원 공간의 접근 분석, 충돌 감지, 동작 계획 등의 분야에서 주로 사용되어지는 모델링 기술이다. Octree 표현 방법은 3차원의 공간을 분할할 수 있는 간단하고 빠른 보편적인 공간 분할 방법이기 때문에 선택되었으며 다른 방법들에 비해 상대적으로 효율적인 알고리즘으로 알려져 있다. Octree 표현 방법에서 전체 장면은 초기에 하나의 3차원 육면체 셀로 만들어 지고 이 셀이 8등분되어 8개의 같은 크기를 갖는 셀들로 재귀적으로 분할된다. 일반적으로 이 재귀적 분할은 최대 재귀 수준에 이르기까지 계속되거나 어떤 셀이 미리 정해진 숫자의 점들 보다 작은 점들을 가질 때까지 계속된다. 제안된 방법에서는 Octree의 구축을 위해 다중 해상도 모델링 방법이 사용된다.

본 논문에서는 조작의 대상이 되는 물체를 중심으로 고해상도의 공간표현을 모델링한다. 즉 목표 물체로부터 먼 거리에 있는 Octree 셀들은 개략적인 해상도의 표현만으로도 충분한 반면 목표 물체에 가까이 있는 Octree 셀들은 보다 상세한 해상도로 표현할 필요가 있다. (그림 8)은 이러한 다중해상도 모델링에 기반한 Octree 표현의 예를 보여준다.



(그림 8) 기타 영역에 대한 다중 해상도 Octree 표현

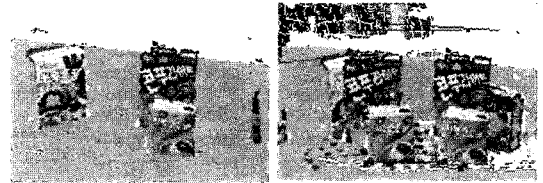
6. 3차원 SIFT 특징과 평면 특징을 이용한 장면의 정합

본 장에서는 움직이는 스테레오 카메라로부터 여러 각도에서 획득된 다중 시점의 3차원 점의 집합들을 효율적으로 정합하여 하나의 통합되고 일관된 3차원 공간을 형성하기 위한 방법을 설명한다. 카메라가 움직일 때 한 순간에 획득된 3차원 점들의 집합은 전체 환경에 대한 부분적인 영역의 정보만을 표현할 수 있다. 특히 스테레오 카메라의 경우 사람이 볼 수 있는 시야각보다 훨씬 작은 시야각을 가지고 있어 환경을 효과적으로 표현하기 위해서는 다양한 각도에서의 3차원 장면들을 획득하고 이들을 통합하는 과정이 필요하다.

따라서 여러 시점에서 획득된 3차원 점들의 집합들은 서로 연결되어 하나의 확장된 3차원 공간을 형성하여 사용되어야 한다. 이러한 과정을 3차원 장면 정합이라 표현한다.

본 논문에서는 기존의 3차원 장면 정합 알고리즘으로 효과적으로 수행할 수 없었던 복잡한 환경에서의 실시간의 강건한 3차원 장면 정합방법을 제안한다. 본 논문에서는 장면 정합을 위해 3차원 점 대신 3차원 공간상의 SIFT 특징점들을 사용한다. 그리고 개략적인 전역적 공간특징을 표현하는 평면 특징들을 사용하여 유효한 SIFT 특징점들의 개수를 줄여 효율적으로 작업을 수행한다. 제안된 장면 정합 알고리즘은 다음과 같은 단계로 설명될 수 있다.

- 1) 3차원 SIFT 특징 집합 S_{t-1} 과 S_t 가 계산된다. 여기서 S_{t-1} 은 $t-1$ 시점에서 획득된 3차원 특징점들의 집합이고 S_t 는 t 시점에서 획득된 3차원 특징점들의 집합이다.
- 2) 3차원 SIFT 특징 집합 S_{t-1} 에서 평면 특징 위에 놓여있는 특징점들만을 선택하여 사용함으로써 새로운 감소된 특징 집합 S'_{t-1} 를 생성한다.
- 3) 감소된 특징 집합 S'_{t-1} 와 t 시점에서의 특



(a) 한 시점의 3차원 장면 (b) 두 장면에서 정합된 특징 쌍들



(c) 정합이 완료된 3차원 장면

(그림 9) 3차원 SIFT 특징과 평면 특징을 이용한 3차원 장면 정합

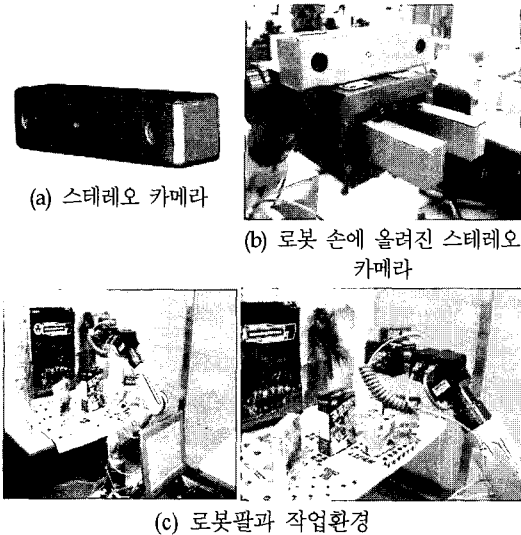
징 집합 S_t 를 정합하여 정합된 특징 쌍들의 집합인 $M_{t-1,t}$ 를 구한다.

- 4) 정합된 특징 쌍들의 집합 $M_{t-1,t}$ 을 기반으로 Horn의 변환 알고리즘에 적용하여 서로 다른 좌표계를 가지는 점들의 집합을 연결할 수 있는 변환 행렬을 계산한다[3].
- 5) 계산된 변환 행렬을 이용하여 t 시점의 모든 3차원 점들의 좌표를 $t-1$ 시점의 3차원 좌표계로 변환하여 이동시켜 줌으로써 정합을 완료한다.

제안된 알고리즘의 시간 복잡도는 $O(N_t \cdot N_{t-1})$ 이다. 여기서 N_{t-1} 는 S_{t-1} 에서의 SIFT 특징점들의 개수이고 N_t 는 S_t 에서의 개수이다. (그림 9)는 제안된 방법을 사용하여 수행되는 장면 정합의 과정과 결과를 보여준다.

7. 실험결과

본 논문에서는 제안된 알고리즘을 실험하고 검증하기 위해 실내 환경에서 서비스를 제공하는 지능로봇이 있고 이 로봇이 탁자 위에 놓여있는



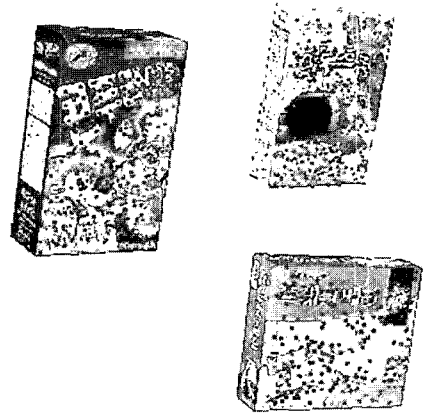
(그림 10) 지능로봇의 조작을 위한 작업환경



(a) 3차원 점들의 집합 (b) 대응되는 컬러영상
(그림 11) 획득된 3차원 환경 정보와 2차원 정보

학습된 물체들을 조작하는 상황을 가정한다. 그리고 조작 대상이 되는 물체들은 탁자와 같은 평면 위에 놓여 있다고 가정한다. 실험에 사용한 스테레오 카메라는 실시간으로 3차원의 좌표 정보와 이에 대응되는 2차원의 컬러 화소 정보를 함께 획득할 수 있다. (그림 10)은 본 실험에 사용된 스테레오 카메라와 실험환경를 보여 준다.

작업환경에 텍스처 정보가 충분이 존재하여 획득된 거리 정보가 충분한 수의 3차원 특징점들을 포함한다고 가정한다. 이렇게 텍스처가 충분한 환경에서는 의미있는 SIFT 특징점들 또한 많이 추출될 수 있다. 거리 정보는 최고 30 frames/sec 의

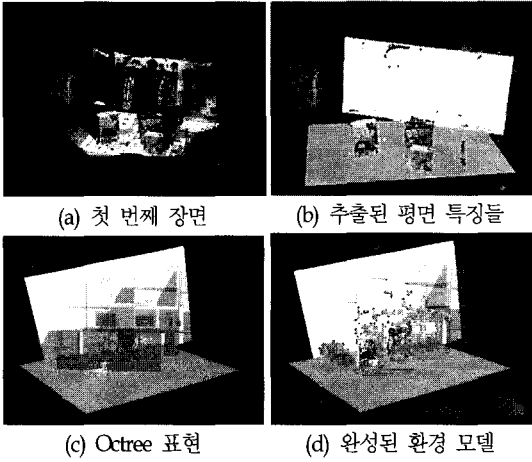


(그림 12) 데이터베이스 내의 물체 모델들

속도로 획득될 수 있다. (그림 11)은 획득된 3차원 점들의 집합과 이에 대응되는 2차원의 영상을 보여준다. 여기서 3차원 점들에는 2차원 컬러 화소의 정보를 함께 도시하였다.

모든 조작의 대상이 되는 목표 물체들은 물체 데이터베이스 내에서 3차원의 완전한 모델을 가진다고 가정된다. 이들 3원 물체 모델들은 물체 표면의 SIFT 특징점들이 미리 계산되어 함께 저장되어 있다. (그림 12)는 본 논문에서 사용된 데이터베이스 내의 물체 모델들의 예를 보여준다. (그림 13)은 복잡한 작업환경에 대한 모델링의 예를 보여준다.

본 실험은 최신 PC (Pentium 4, CPU 2.8 GHz)에서 구현 되었으며, 표 1은 각 단계의 알고리즘들이 수행되는 시간을 측정한 결과를 보여 준다. 표에서 보는 바와 같이 두 개의 연속된 장면들로 인식과 모델링을 수행하는데 평균 0.779 초의 시간이 소요되며 비교적 빠른 시간 안에 3차원의 환경을 인식하고 모델링 할 수 있어 실시간의 응용분야에 적용이 가능하다는 것을 알 수 있다. 또한 제안된 접근 방법이 기존의 다른 방법들 보다 훨씬 빠른 시간 안에 3차원의 환경을 모델링할 수 있다는 것을 알 수 있다[2][3][9].

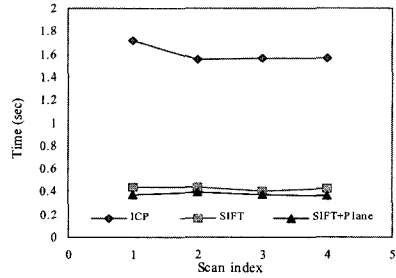


(그림 13) 5개의 연속된 장면을 통해 구축된 환경 모델

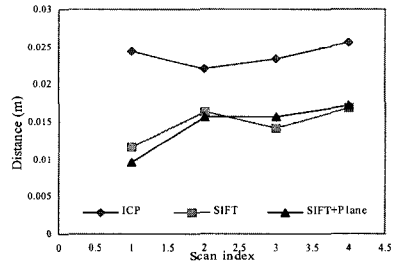
(표 1) 시스템 성능의 측정

	평균 처리 시간 (ms)
SIFT 특징 계산	178
평면 특징 추출	21
물체 인식 및 정합 (3 개의 물체)	342
기타 영역 모델링	75
3차원 장면 정합	183
전체 계산시간	799

성능의 평가를 위해 제안된 3차원 장면 정합 알고리즘을 기존의 ICP 기반 접근방법[3]과 비교하여 실험하였다. (그림 14)의 실험 결과에서 보여주듯이 제안된 3차원 장면 정합 알고리즘이 기존의 ICP 기반 방법보다 더 빠르면서 정확하다는 것을 알 수 있다. (그림 14)의 (a)의 계산시간은 SIFT 특징 계산과 SIFT 정합 및 좌표계 변환을 모두 포함한 시간을 의미한다. (그림 14)의 (b)에서는 SIFT 기반 방법을 추출된 평면 특징을 사용하는 경우와 사용하지 않는 경우로 나누어 비교하였으며, SIFT 특징점과 평면 특징을 함께 사용하였을 때 오류가 가장 적게 나타나는 것을 알 수 있다. 기존의 ICP 기반 방법들은 임의의 3차원 점들을 기반으로 정합을 수행하므로 오류가



(a) 수행시간의 비교



(b) 거리 오류의 비교

(그림 14) 3차원 장면 정합 알고리즘의 평가

포함될 가능성이 큰 반면 SIFT 기반 방법은 오류가 적은 강한 특징값들만을 사용하므로 오류의 가능성을 줄일 수 있다. 특히 평면 특징이라는 전역적인 특징을 사용하여 잡음의 발생가능성을 최소화할 수 있다. 이와 같이 본 논문에서 제안한 SIFT 특징점과 평면에 기반한 3차원 장면 정합은 기존의 방법들보다 복잡한 환경 속에서 더 정확하고 강건하게 정합을 수행할 수 있다고 할 수 있다.

8. 결론 및 향후과제

본 논문에서는 다양한 실시간의 응용 분야에서 활용될 수 있는 빠르고 강건한 3차원 환경의 인식 및 모델링 방법을 제안하였다. 특히 인간이 환경을 인지하는 3가지 기본 원칙을 관찰하고 이를 적용함으로써 보다 빠르고 자연스럽게 시스템이 3차원의 환경을 인식할 수 있는 방법을 개발하였

다. 즉, 3차원 환경에서 먼저 개략적인 전역적 공간 특징을 추출하고, 이를 기반으로 조각의 대상이 되는 목표 물체들을 인식한다. 그리고 기타 영역들에 대해서는 관심영역을 중심으로 하는 다중 해상도의 Octree를 이용하여 효율적으로 3차원 환경을 표현하였다. 특히 이러한 세 단계의 기본 원칙을 수행함에 있어 각각의 단계에서 도출되는 정보들이 독립적으로 표현되는 것이 아니라, 서로 유기적으로 작용하면서 시스템의 효율성과 정확성을 위해 활용되고 있다. 예를 들어 전역적 공간 특징으로 추출된 평면 특징 들은 3차원 장면의 정합과정에서 활용되어 정확도와 속도를 향상시키는데 사용되고 또한 목표물체의 인식 단계에서도 탐색 영역을 축소하여 오류를 줄이고 인식 속도를 향상시키는데 활용된다. 그리고 인식된 물체의 영역은 관심 영역으로 지정되어 기타 영역을 Octree로 표현할 때 관심 영역에 대해서만 고해상도의 표현을 함으로써 전체적인 효율을 높이는 효과를 주고 있다. 실험 결과를 통해 본 논문에서 제안하는 방법이 충분히 빠르고 강건하게 실시간 응용들을 위한 3차원 환경의 인식과 모델링을 수행할 수 있음을 보여 주었고, 제안된 방법이 기존에 상세한 그래픽적인 정확도에만 치우치던 3차원 모델링 방법들에 비해 훨씬 빠르고 효과적으로 3차원 환경을 인식하고 모델링 할 수 있음을 증명하였다.

추후 연구에서는 물체 데이터베이스의 확장, 3차원 불변 특징의 추가, 그리고 다양한 응용에의 적용 등에 대해 지속적으로 연구할 계획이다.

참 고 문 헌

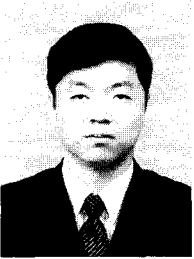
- [1] M. A. Garcia and A. Solana, "3D Simultaneous localization and modeling from stereo vision," International Conference on Robotics & Automation (ICRA'04), New Orleans, LA, April 2004
- [2] A. Nüchter, H. Curmann, and J. Hertzberg, "Automatic model refinement for 3D reconstruction with mobile robots," 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 18th Conference on Machine Learning, Williams College, July 2001.
- [5] S. Se, D. Lowe and J. Little, "Vision-based mapping with backward correction," 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2002
- [6] 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
- [7] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP," The 3rd International Conference on 3D Digital Imaging and Modeling, 2001
- [8] D. Lowe. "Object recognition from local scale invariant features," In proceedings of the Seventh International Conference on Computer Vision (ICCV'99), pages 1150-1157, Kerkyra, Greece, September 1999.
- [9] 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.
- [10] Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," In proceedings of the IEEE Conference on

Robotics and Automation, 1991.

[11] 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.

● 저자 소개 ●



장 대 식

1994년 숭실대학교 전자계산학과 졸업(학사)
1996년 숭실대학교 대학원 전자계산학과 졸업(석사)
1999년 숭실대학교 대학원 전자계산학과 졸업(박사)
2000~2002년 줄라이네트(주) 연구소 책임연구원
2002~2004년 (주)인컴아이엔씨 연구소 책임연구원
2004~2005년 성균관대학교 연구교수
2005~현재 국립군산대학교 컴퓨터정보과학과 교수
관심분야 : 컴퓨터비전, 로봇비전, 패턴인식, 지능형로봇, etc.
E-mail : dsjang@kunsan.ac.kr