

## 실시간 카메라 추적을 위한 3차원 공간에서의 Keypoint 선택 알고리즘

### Keypoint Selection Algorithm for Real-time Camera Tracking in 3D Environments

김기영, Kiyoung Kim\*, 우운택, Woontack Woo\*\*

**요약** 본 논문에서는 3차원 공간에서의 실시간 카메라 추적 방법과 효율적인 추적을 위한 Keypoint 선택 방법을 제안한다. 일반적으로 3차원 공간으로부터 카메라의 자세 (Pose)를 계산하기 위해서는, 3차원 공간상의 특징점 좌표 (coordinates) 와 영상 평면에 투영된 2차원 영상점이 요구된다. 실시간 카메라 추적은 3D-2D의 매칭 관계를 실시간으로 추적하는 과정이다. 본 논문에서는 오프라인에서 생성된 대용량의 3차원 정보가 주어졌을 경우, 실시간성을 보장하면서 강건하게 카메라의 자세를 회복하는 방법을 설명한다. 우선 오프라인 상에서 공간의 영상들을 획득하고, 이로부터 SIFT (Scale Invariant Feature Transform) 특징을 추출한다. 영상 기반 3차원 복원 기술을 통해 특징점들의 3차원 좌표와 매칭을 위한 SIFT 기술자 (descriptor)를 생성한다. 복원된 SIFT 특징점들을 실시간 카메라 추적에 그대로 사용하기에는 데이터의 양 및 높은 CPU 요구량 때문에 실시간성을 보장하기 힘들게 된다. 이를 극복하기 위해, 복원된 SIFT 특징점들을 3차원 공간상에서 균등하게 분할하고, 각 영역 당 보여 질 확률이 높은 특징점들만을 Keypoint로 선별한다. 이를 빠른 매칭을 위해 Octree 구조를 이용해 트리 레벨 별로 저장한다. 균등한 영역에서 선별·저장된 특징점들은 순차적으로 실시간 매칭에 사용되며, 가능성이 높은 정밀 영역을 정하는데 사용되고 최종적으로 확장된 매칭 과정을 수행한다. 제안된 방식을 80여장의 Keyframe으로 구성된 3차원 공간 적용한 결과, 640×480 영상에서 6~10ms이내에 카메라의 자세를 획득했다. 향후 제안된 알고리즘은 넓은 영역에서의 증강 현실 응용 및 로봇 내비게이션 등에 활용될 수 있을 것이다.

**Abstract** In this paper, we propose a real-time camera tracking method in 3D and a novel efficient keypoint selection algorithm for it. Generally, the 3D coordinates of features and their projected 2D points onto an image plane are required to compute a camera pose in a 3D environment. The real-time camera tracking is to find the matching relationship between 3D and 2D points. In this paper, we address a robust method which recovers camera poses in real-time when a large amount of 3D data are given as results from off-line processing. First of all, we obtain large image set of the environment and extract SIFT features from the image set. Then, we reconstruct the scene in 3D with SIFT descriptors. It is difficult to guarantee the real-time performance when we use all the 3D data because of large amount of data and CPU requirement. To overcome the problem mentioned above, we evenly divide all features in the 3D space. Then, we select the features that can be seen to many images and save them in an octree structure. The saved features are matched step by step from the higher level set. Finally, we carry out the fine matching. Using the proposed method, we could get 6~10ms to process one frame including the camera pose estimation with a 640×480 image. In the future, the proposed method will be exploited in Augmented Reality applications or robot navigation area.

**핵심어:** Camera Tracking, Keyframe Selection, Keypoint Selection, Augmented Reality

본 연구는 문화체육관광부 및 한국문화콘텐츠진흥원의 문화콘텐츠기술연구소육성사업의 연구결과로 수행되었음.

\*주저자 : 광주과학기술원 정보통신공학과 박사과정 e-mail: kkim@gist.ac.kr

\*\*교신저자 : 광주과학기술원 정보통신공학과 교수 e-mail: wwoo@gist.ac.kr

## 1. 서론

최근 영상기반 특징 인식 및 카메라 추적 기술은 마커리스 증강 현실 응용의 구현에 중요한 역할을 담당하고 있다. 본 논문에서는 3차원 공간을 효율적으로 인식하고, 카메라의 6DOF 자세를 실시간으로 추적하는 방법을 제안한다. 3차원 공간에서 인식 및 추적을 하기 위해서는 장면의 3차원 좌표 정보와 그에 상응하는 특징 정보가 요구된다 [1]. 이를 위해, 본 논문에서는 3차원 공간을 오프라인 상에서 복원하고, 복원된 공간에서 중요한 SIFT [2] 특징점을 균등한 공간상에서 관측 확률이 근거해 선별한다. 실제 카메라 추적 시에는 선별된 특징점을 이용해 매칭을 수행하고, 매칭은 카메라 공간상의 위치 정보를 이용해 추정된 카메라 FOV (Field of View) 영역 안에서 진행된다. 그러므로, 오프라인에서 넓은 영역을 복원해 특징 데이터가 증가함에 따라 카메라 추적 속도에 큰 영향을 미치지 않는다.

## 2. 제안 방법

특징 인식 및 추적의 전체적인 과정은 그림 1과 같다. 프레임워크는 오프라인 모드와 온라인 모드로 구성된다. 오프라인 모드에서는 영상 기반 모델링을 수행하고, 추적 데이터를 선별해 저장한다. 온라인 모듈은 카메라의 자세를 실시간으로 추출하기 위해, 두 쓰레드 기반의 멀티코어 프로그래밍 접근 방법을 사용한다. SIFT 인식 과정은 백그라운드에서 실행되며, 인식된 Keypoint는 affine warping 기반의 추적 알고리즘으로 추적되어 안정적인 결과를 제공한다.

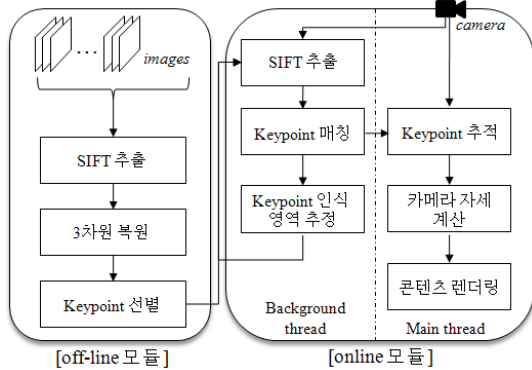


그림 1. 3D 환경 인식 및 추적 프레임워크 개요

### 2.1 Keypoint 선별

3차원 복원 후 생성된 모든 Keypoint를 추적에 사용할 경우, 매칭 특징점들의 증가로 광범위한 영역에서의 성능이 급격히 감소한다. 본 논문에서는 이를 해결하기 위해, 3차원으로 복원된 특징점들을 octree로 구성한다. 트리 레벨 (1, ..., m)에 따라, 3차원 공간을  $N_1, \dots, N_m$  크기로 분할한다. 분할된 영역에서 관찰 빈도수가 가장 높은 특징점들만을 선별해 저장한다. 레벨 1에서 추출된 Keypoint 세트를  $L_{k,1}$ 라 하자. m은 최종적으로 복원된 특징점 수에 의해 결정되

며, 5를 넘지 않는다. 추후, 빠른 특징점 매칭 과정을 위해 각 영역에 속한 특징점들은  $k-dTree$ 를 이용해 저장한다.

### 2.2 Keypoint 매칭 및 인식 영역 선별

매칭은 2.1에서 추출된  $L_{k,1}$ 부터 높은 트리 레벨로 차례대로 수행한다. 트리 매칭은 특정 조건을 만족할 경우, 인식될 확률이 높은 공간으로 판별한다. 즉, 특정 레벨에 속한 영역 중 매칭 과정에서 관련 없는 공간의 특징점들을 대부분 제거할 수 있다. 매칭 확률이 높다고 판별된 영역은 미리 저장된  $k-dTree$ 를 이용해 카메라 영상으로부터 추출된 특징점과 비교하여 정밀한 매칭 결과를 얻는다.

## 3. 구현 및 결과

실험에서는 CPU 2.66GHz (Core 2 Duo)의 일반 노트북을 활용했다. 그림은 일반 장면을 3차원으로 복원된 결과를 보인 것이다.  $640 \times 480$  지원의 웹캠을 사용했으며, 총 1500의 영상에서 80장의 키프레임을 선별해 복원을 수행했다. 한 프레임 당 카메라 추적은 6~10ms안에 수행이 되었으며, SIFT 매칭은 GPU 기반의 알고리즘을 이용해 프레임 당 33ms안에 수행이 되었다.

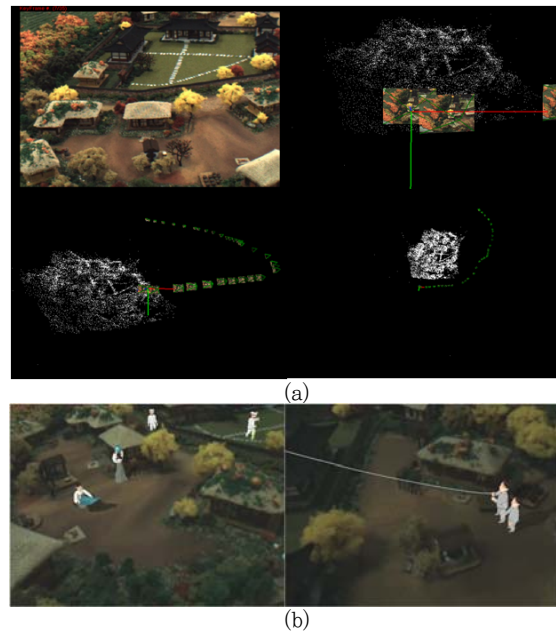


그림 2. 복원된 3차원 공간 및 증강 현실 적용 결과 (a) 영상 기반 복원 결과 (b) 3차원 공간에 가상 캐릭터 증강

## 참고문헌

- [1] Z. Dong, G. Zhang, J. Jia, and H. Bao, "Keyframe-Based Real-Time Camera Tracking", IEEE International Conference on Computer Vision (ICCV), 2009.
- [2] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, 60, 2, pp. 91-110, 2004.