

# 모바일 환경에서 ORB SLAM 3 기반 멀티유저 Global Map 생성 및 공유 알고리즘

성홍념<sup>0</sup>, 이철우<sup>\*</sup>

<sup>0</sup>전남대학교 소프트웨어공학과

<sup>\*</sup>전남대학교 컴퓨터정보통신공학과

<sup>0</sup>6590f1@naver.com, <sup>\*</sup>leecw@jnu.ac.kr

## 요약

복수의 모바일 단말들 사이에서 위치정보를 정밀하게 추정하기 여러 방법들이 사용되고, 그 방법 중 본 논문에서는 모바일 환경에 맞춰 Visual SLAM을 기반으로 한 Global Map작성 시스템을 제안한다. 하지만 여러 모바일 에이전트가 작성하는 Map을 합치기 위해서는 여러 가지 어려움이 존재한다. 본 논문에서는 이런 문제에 대해 설명하고 다수의 모바일 에이전트들이 생성한 맵을 실시간으로 융합하여 위치 정보를 동기화 할 수 있는 시스템을 제안한다.

## 1. 서론

복수의 단말(이하 에이전트 라고 함) 간의 상호 작용이 가능한 AR 시스템을 개발하기 위해서는 각각의 단말들이 물체를 정확히 인식하고 자신의 위치를 추정하면서 공동으로 사용이 가능한 글로벌 맵을 제작해야 한다. 그러나 서로 다른 특성을 지닌 카메라를 사용하여 장시간 위치 추정을 하고, 정확한 공동지도를 완성하는 것은 매우 어려운 일이다.

이 문제를 좀 더 구체적으로 나누어 고찰해보면 다음과 같다. 첫째 각각의 에이전트들이 취득한 영상으로부터 위치추정에 필요한 강인한 특징점을 어떻게 찾아낼 것인가에 대한 문제이다. 이 문제는 “강인한 특징점을 어떻게 찾을 것인가” 라는 단일 연구로 그 동안 많은 연구가 이루어져왔으나 모바일 디바이스, 즉 스마트폰을 사용할 경우 계산량을 줄이고 통신속도를 올려야하기 때문에 여러 가지 궁리가 필요하다. 두 번째는 각 카메라가 취득한 영상 속 특징점 간의 대응관계를 어떻게 구할 것인가의 문제이다. 고차원의 벡터 기술자(vector descriptor)를 사용하면 정확한 대응관계를 구하는데 도움이 되나 다수의 에이전트 간의 정보통합에는 많은 계산량이 필요하고, 내부 파라미터가 다른 카메라 사이에서는 같은 물체도 조금씩 다르게 왜곡되기 때문에 이에 대한 궁리가 필요하다. 세 번째

는 강인한 특징점들을 이용하여 각각의 에이전트들이 구한 로컬 맵을 만들고 모든 에이전트들이 사용이 가능한 공동의 맵을 어떻게 만들 것인가 하는 문제이다. 이 문제는 정확한 로컬 맵을 만들면 이를 바탕으로 정확한 글로벌 맵을 만들 수 있지만 글로벌 맵이 정확히 잘 만들어지면 에이전트들이 로컬맵을 만드는데 많은 정보를 제공할 수 있으므로 상호보완적이면서 점진적으로 이루어져야 하는 작업이다.

본 논문에서는 고속 처리를 위해 강인한 ORB-SLAM 3 알고리즘[1]을 기반으로 에이전트 사이에서 공유할 수 있는 특징점들을 추출하고 각 에이전트들의 특징점과 keyframe을 기록한 후 에이전트 간의 정보통합을 통해 공동 사용이 가능한 글로벌 맵 제작 기법에 대해 기술한다.

## 2. 시스템 구조

본 논문에서 제안한 시스템의 전체적인 구조는 그림 1과 같다. 이 구성은 ORB-SLAM3 에서 제안한 방법을 모바일 환경에서 다수의 에이전트가 사용할 수 있도록 개선되었다.

### 2.1 MapData

MapData는 Active Map(이하 Agent Map)과 Non Active Map(이하 Global Map)으로 이루어진다. 또한 각각의 Map들은 각 에이전트로부터 받아들이

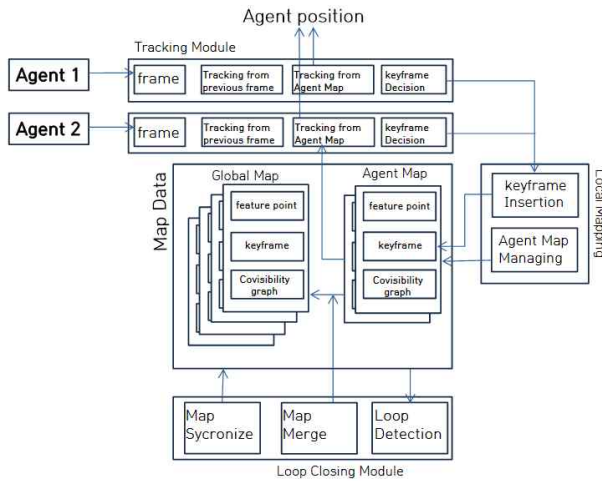


그림 1 시스템 구조도

영상에서 특징점, keyframe들과 그것들이 인식된 환경 안에서 측정된 3d좌표 그리고 특징점들의 대응관계정보를 추출하여 저장하고 있다. 본 논문에서 제안한 시스템에서는 MapData를 사용하여 각 에이전트의 Visual Odometry 계산과 함께 플레이어의 위치를 동기화한다.

### 2.1.1 Agent Map

Agent Map은 각각의 에이전트들이 비교적 짧은 시간 안에 SLAM을 진행하며 주변환경을 맵핑한 특징점과 key-frame으로 이루어진 Map이다. Agent Map은 자신의 위치를 추정하기 위한 Local Map이다. Active Map은 하나의 에이전트마다 하나씩 보유하고 각각의 Active Map은 유저가 공유되는 데이터를 통해 위치가 보정 및 동기화될 때 Global Map으로 전환되고, 유저는 보정된 위치를 기반으로 시작하는 새로운 Agent Map을 갖게된다.

### 2.1.2 Global Map

두 번째로 Global Map은 모든 에이전트가 공유하는 공동의 지도이다. 각각의 에이전트는 일정 주기마다 실행되는 loop detection 연산, 즉 이전에 왔던 곳인지 여부를 확인하고 Global Map을 통해 구한 현재 위치와 Agent Map을 통해 구한 현재 위치 사이의 오차를 계산하여 오차 범위가 일정 값 이하일 경우 새로운 Global Map으로 통합한다.

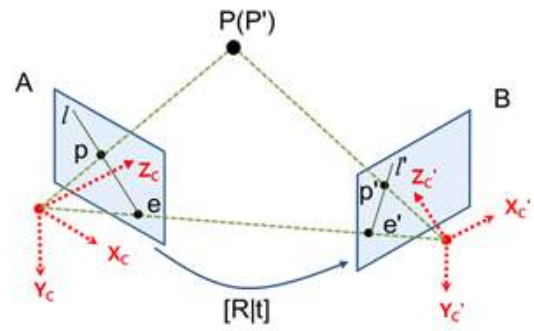


그림 2 에피폴라 지오메트리 제약 조건

## 2.2 Tracking

Tracking 모듈은 에이전트에서 입력한 현재 프레임과 직전 프레임의 특징점 그리고 Agent Map을 사용하여 Agent의 현재 위치를 추정한다. 이 경우, 그림 2와 같이 에피폴라 지오메트리 제약 조건을 사용하여 이전 프레임 또는 Active Map의 keyframe과 현재 프레임 사이의 매칭되는 특징점을 기반으로 위치를 추정한다. 또한 그림 1 시스템 구조도와 같이 현재 프레임과 추출된 특징점 등을 Local mapping 모듈로 전달하는 역할을 수행한다.

## 2.3 Local Mapping

Local Mapping 모듈은 Tracking 모듈로부터 입력을 받아 Agent Map을 관리하는 모듈이다. Local Mapping 모듈은 시스템 구성도와 같이 Agent Map에 키 프레임과 특징점을 추가하고, 중복된 키 프레임을 제거하며, 현재 프레임과 Agent Map에서 현재프레임과 관련있는 Keyframe들을 통해 Active Map을 보정한다

## 2.3 Loop closing

Loop closing 모듈은 Agent Map과 Global Map 사이의 공통구간을 감지하고 Global Map을 작성하는 역할을 한다. 그림 2와 같이 Agent Map과 Global Map 사이에서 공통된 특징점이나 keyframe의 대응관계가 검출되면, Global Map을 기반으로 Agent Map을 수정하여 Global Map에 Agent Map이 동기화되며 각 에이전트의 위치추정 정보인 Agent Map이 Global Map에 동기화되며 위치정보가 동기화 되게 된다.

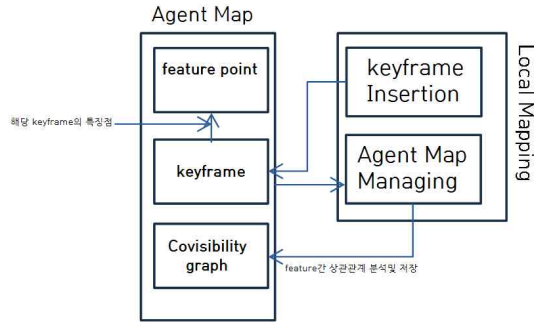


그림 3 local Mapping 구조도

### 3. 특징점 공유

서론에서 언급한 첫 번째와 두 번째 문제를 해결하기 위해서는 에이전트간의 간의 특징점을 공유하기 위해서는 모든 에이전트가 같은 특징점을 보았을 때 동일한 특징점이라고 인식할 수 있어야 한다. 하지만 시스템상의 에이전트들은 다른 내부 파라미터를 가진 카메라를 사용하기 때문에 특징점을 쉽게 공유하기 힘들다. 이런 문제를 위해 본 논문에서는 각 에이전트별로 camera calibration을 진행하여 카메라 내부 요인을 제거하고, 각 카메라별 내부 파라미터 차이에도 공유, 인식 될 수 있는 강인한 특징점을 사용하였다. 이러한 방식을 통해 처리를 하면 그림 4과 같이 서로 다른 에이전트의 카메라 사이에서도 특징점을 공유할 수 있다는 사실을 알 수 있다.

#### 3.1 특징점 공유

카메라로 촬영한 이미지는 실세계의 3차원 공간상의 점들을 2차원 평면에 투영하여 얻어진다. 이때, 3차원 공간상의 점들이 이미지에서 이미지 센서의 일직선상에 위치해야 하지만 실제로는 사용한 카메라의 렌즈나 이미지 센서와의 초점거리 등 카메라의 고유한 특성에 따라 달라진다. 이러한 점 때문에 각각의 에이전트들이 서버에 제공하는 프레임은 각각 조금씩 같은 공간을 보더라도 조금씩 다르게 인식될 수 있다. 따라서 본 논문에서는 각각의 에이전트마다 직선이 곡선으로 왜곡되는 Radial distortion을 보정하기 위해 수식 1과 같은 근사 방정식을 통해 왜곡계수를 구한다. 또한 이미지가 상이 땀히는 이미지 면과 평행하지 않게 왜곡된 Tangential Distortion을 수식 2과 같은 근사 방정식에서 왜곡계수를 구하여 사용한다. 결과적으로 수식 1, 수식 2를 통해 왜곡계수  $k_1, k_2, p_1, p_2, k_3$ 를



그림 4 ORB기술자를 이용한 특징점 매칭

구하여 에이전트마다 다른 왜곡계수를 적용하여 프레임을 보정하여 사용하였다. 이후 카메라의 초점거리와 광학 중심  $(f_x, f_y), (c_x, c_y)$ 를 구해 수식 3과 같은 카메라 메트릭스를 구하여 사용하였다.

#### 3.2 ORB(Oriented Fast and Rotated BRIEF)

ORB SLAM 3에서는 Oriented FAST and Rotated BRIEF features descriptor[2]를 사용한다. 이 방식은 특징점을 추출하기 위해 1차적으로 FAST알고리즘을 통해 특징점을 검출한 이후, 검출된 특징점들을 2차적으로 해리스 코너 검출 알고리즘에 통과시켜 강인한 특징점만들 추출하여 서로 다른 카메라 내부 파라미터를 가진 두 이미지 사이에서도 서로 공유, 인식 가능한 강인한 특징점들을 추출할 수 있다.

### 4. Global Map의 작성

ORB-SLAM3 시스템에서 카메라의 위치를 계산하기 위해서는 이전 프레임과 특징점 대응관계와 위치가 저장된 map data를 생성하여 활용한다. 특히 ORB SLAM에서는 맵 데이터를 크게 short-term data association, Mid-term data association, Long-term association 3가지 data association을 사용하여 위치를 추정한다. 본 논문에서는 다수의 모바일 에이전트의 효과적이고 정확한 위치추정 및 동기화를 위해 여러가지 에이전트 간 Map Fusion 방법을 도입하여 시스템을 구성하였다.

$$\begin{aligned} x_{corrected} &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ y_{corrected} &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \end{aligned}$$

수식 2 Radial distortion

$$\begin{aligned}x_{corrected} &= x + [2p_1xy + p_2(r^2 + 2x^2)] \\y_{corrected} &= y + [p_1(r^2 + 2y^2) + 2p_2xy]\end{aligned}$$

수식 3 Tangential Distortion

$$cameramatrix = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

수식 3 카메라 매트릭스

#### 4.1 Short-term data Association

short-term data association은 기존 visual Odometry와 같이 현재 프레임 특징점과 현재 프레임과 근접한 프레임 사이에서 데이터를 결합하여 계산하는 방식으로 지속적인 추정 드리프트가 발생한다.

#### 4.1 Mid-term data Association

mid-term data association은 현재 프레임의 특징점과 현재 프레임과 시간 차이가 얼마 나지 않는 맵 데이터를 결합하여 계산하는 방식이다. 일정 시간 시간 안에서 작성된 Active Map(Local Map)을 통해 공간을 인식하여 short-term 방식보다 더 넓은 범위에서 특징점을 검색하여 누적된 drift를 어느정도 해소해 줄 수 있다.

#### 4.3 Long-term data Association

long-term data association은 현재 위치가 시간에 관계 없이 이전에 방문했던 위치인지 확인연산으로 에이전트의 움직임중 loop를 감지하고, 이전에 방문하여 작성했던 맵정보를 이용하여 현재 위치를 relocalization을 하는데 사용된다. 이를 위해서 이전에 Mid-term data association에 사용되었지만 시간이 흘러 Non-Active Map으로 전환된 글로벌 맵과의 비교를 통해 연산이 진행된다.

#### 4.4 Map Fusion

다수의 모바일 에이전트의 위치정보를 추정하고 동기화하기 위해서 본 논문에서는 각 에이전트가 수집한 정보 short,mid, 그리고 long-term data로 변환하여 활용하여 position estimation을 진행하고 Global Map(Non active map)과 각 에이전트 별 Local Map(active map)을 구성한다. 하지만 모든 에이전트는 서로 다른 방향과 위치에서 촬영한 프레임을 서버에 제공한다. 그렇기 때문에 하나의 에이전트만

존재 할 때와 달리 각기 다른 에이전트로부터 입력된 프레임 사이에는 연속성이 존재하지 않아 Short, Mid-term data association을 진행 할 수 없다. 그렇기 때문에 본 논문에서는 각 에이전트마다 short-term association, Mid-term association 연산을 통해 Visual Odometry를 계산을 하고, 강인한 특징, 키프레임을 통해 Active Map을 생성하게 된다. 이를 통해 생성된 각각의 Active Map은 합쳐져 Non Active Map으로 통합된다. 그 다음 Long term data association 연산을 통해 이전에 방문했던 장소의 Non Active Map을 인식하여 상대적인 위치정보를 보정한다.

#### 4.4.1 Local Map Making

기존 ORB-SLAM3에서는 short-term association연산을 통해 Visual Odometry를 계산하고 Mid-term association 연산을 통해 현재 에이전트가 인식하고 drift 보정에 사용하는 Active Map을 구성하게 된다. 본 논문에서는 다중 에이전트에 대해 위치추정 시스템을 제공하기 위해 각 에이전트 마다 Agent Map을 하나씩 할당하여 각자의 로컬 Map을 생성하며 위치를 추정하게 된다.

#### 4.3 Global Map Making

본 논문의 목표인 동기화를 하기 위해선 Local Map Making을 통해 생성된 각 에이전트들의 Agent Map을 동기화해야 한다. 기존의 ORB SLAM3에서는 Agent Map을 생성하는 중 이전에 생성되어 저장되어있는 Global Map과 겹치는 부분에 오게 되면 해당 정보를 바탕으로 Agent Map을 보정하게 된다. 본 논문에서는 이런 과정을 진행하기 위해 어떤 Global Map 정보를 우선적으로 사용할 것 인지 기준을 작성하였다. 첫 번째 기준은 시간이다. 단안시 기반 visual slam은 시간이 지나면서 누적 오차가 발생한다. 이를 최소화 하기 위해 본 논문에서는 Long-term data association을 진행할 때 Non Active Map에서 검색된 여러 feature들 중에서 가장 오래된 feature를 사용하여 보정 및 위치정보를 추정하였다. 두 번째 기준은 가장 높은 신뢰도를 주기위해 해상도가 높은 에이전트로부터 생성된 Non Active Map 기준으로 loop closing등의 연산을 함으로 높은 신뢰도의 에이전트의 정보를 바탕으로 공유되는 Map을 구성하였다.

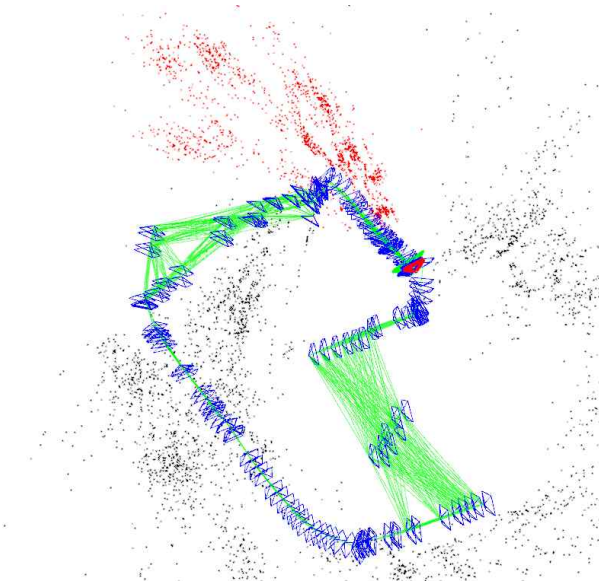


그림 5 단일 에이전트 입력을 통해 구성한 Global Map

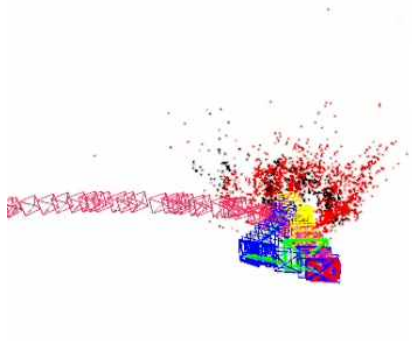


그림 6 다수 에이전트 입력을 통해 구성한 Global Map

## 5. 실험

기존의 ORB SLAM 3의 경우 모바일 에이전트 입력을 통해 실행 할 경우 그림 5와 같이 진행을 하는 도중 loop closing을 진행 할 때 하나의 카메라 입력으로 정확한 결과를 도출하지 못하고 Active Map을 제대로 보정하지 못하고 맵이 망가지는 모습을 볼 수 있다. 하지만 여러 입력을 통해 Global Map을 구성하여 맵을 작성하면 그림 6과 같이 Loop Closing에서 keyframe간 상관관계를 더욱 정교하게 추정하여 맵이 망가지지 않고 연속적인 움직임을 자세하게 인식할 수 있어 더 정밀한 Global Map을 작성하여 SLAM 시스템의 정확도가 향상된 것을 확인 할 수 있다.

## 6. 결론

본 논문은 ORB SLAM3를 기반으로 다수의 모바일 에이전트를 대상으로 Global Map을 제공할 수 있는 시스템을 구상하여 실험하였다. 실험 결과 다수의 에이전트를 대상으로 각각의 Local Map을 본 논문에서 제안한 기준에 따라 융합하여 Global 맵을 구성할 수 있다는 것을 알 수 있었다. 또한 기존 ORB SLAM 3와 다수의 에이전트를 사용하기 때문에 여러 방면에서 공간을 인식할 수 있어 각각의 에이전트가 서로 features를 공유하는 과정에서 기존의 ORB SLAM 3보다 더 정밀하게 위치를 추정 할 수 있었다.

## 참고문헌

- [1] Carlos Campos, Richard Elvira, Juan J. Gomez Rodriguez, Jose M.M. Montiel and Juan D. Tardos, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM," IEEE Transactions and Robotics, 2021
- [2] Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary Bradski, "ORB: an efficient alternative to SIFT or SURF", International Conference on Computer Vision, pp. 2564-2571, 2011