JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 22, No. 11, pp. 1877-1883, Nov. 2021

Check for updates

ORB-SLAM의 재지역화 성능 개선을 위한 전문적인 시각적 어휘 활용

조 현 우[1] · 서 영 건[2] · 이 수 원[3*]
[1]한국전자통신연구원 홀로그래픽콘텐츠연구실
[2,3*]경상국립대학교 컴퓨터과학부, 경상국립대학교 기초과학연구소

Utilizing Specialized Visual Vocabulary to Improve Relocalization Performance of ORB-SLAM

Hyunwoo Cho[1] · Yeong Geon Seo[2] · Suwon Lee[3*]
[1]Electronics and Telecommunications Research Institute
[2,3*]School of Computer Science and The Research Institute of Natural Science, Gyeongsang National University

[요 약]

ORB-SLAM은 실시간 위치 추정 및 지도 작성이 가능하여 실시간성 응용에 널리 활용되고 있다. ORB-SLAM은 추적에 실패한 경우 재지역화를 시도하는데 어떤 시각적 어휘를 사용하느냐에 따라 재지역화 성능에 큰 영향을 미친다. 시각적 어휘는 ORB-SLAM이 실행되기 전에 준비되어야 하기 때문에 일반적인 이미지들에서 추출된 지역 특징들을 이용해 미리 학습을 시켜 놓는 것이 보통이다. 하지만 이렇게 학습된 시각적 어휘는 ORB-SLAM이 수행되는 환경을 고려하여 제작된 시각적 어휘가 아니다. ORB-SLAM이 수행되는 환경을 미리 알 수 있다면 해당 환경에 가장 알맞은 시각적 어휘를 사용하는 것이 가장 좋은 재지역화 성능으로 이어질 것이다. 본 논문은 전문적인 시각적 어휘를 제안하고, 실험을 통해 환경에 전문화된 시각적 어휘를 사용하는 것이 ORB-SLAM의 재지역화 성능을 개선할 수 있음을 보인다.

[Abstract]

ORB-SLAM is widely used in real-time applications since it enables real-time localization and mapping. ORB-SLAM tries to relocalize when tracking fails, and the relocalization performance varies greatly depending on which visual vocabulary is used. Since visual vocabulary has to be prepared before ORB-SLAM is executed, it is usually trained in advance using local features extracted from general images. However, the visual vocabulary was not created considering the environment in which ORB-SLAM is performed. If the environment in which ORB-SLAM is performed can be known in advance, using the visual vocabulary most appropriate for the environment will lead to the best relocalization performance. This paper proposes a specialized visual vocabulary and shows that using the specialized visual vocabulary for the environment can improve the relocalization performance of ORB-SLAM through experiments.

색인어 : 동시적 위치 추정 및 지도 작성, 시각적 어휘, 재지역화, 위치 추정, 지도 작성

Keyword : Simultaneous Localization and Mapping, Visual Vocabulary, Relocalization, Localization, Mapping

http://dx.doi.org/10.9728/dcs.2021.22.11.1877

Received 03 September 2021; Revised 17 September 2021
Accepted 19 October 2021

***Corresponding Author; Suwon Lee**

Tel: +82-55-772-1394
E-mail: leesuwon@gnu.ac.kr

Ⅰ. 서 론

동시적 위치 추정 및 지도 작성(simultaneous localization and mapping, SLAM)은 로봇이 임의의 공간을 이동하면서 현재 위치를 추정하는 동시에 주변의 지도를 작성하는 기술을 말한다. SLAM은 컴퓨터비전과 관련 센서 장비의 발전에 힘입어 그 어느 때보다 활발히 연구되고 있다. 특히, 카메라를 주요 센서로 사용하는 시각적 SLAM(visual SLAM)의 대표 알고리즘인 ORB-SLAM[1,2]은 고속으로 추출 및 매칭이 가능한 ORB[3]라는 이진 특징(binary features)을 사용하여 증강현실과 같이 실시간 SLAM을 필요로 하는 응용에 주로 활용되고 있다.

ORB-SLAM은 추적 단계에서 카메라의 포즈를 적절히 예측하지 못하여 추적에 실패한 경우 기존에 습득한 정보를 바탕으로 현재의 대략적인 위치를 다시 추정하는 재지역화(relocalization) 단계를 거친다. 재지역화를 위해 입력 이미지와 가장 유사한 키프레임을 검색하여야 하는데 이때 시각적 어휘(visual vocabulary) 기반의 이미지 인식(image recognition)이 수행된다. 어떤 시각적 어휘를 사용하느냐에 따라 ORB-SLAM의 재지역화 성능이 달라지며, 재지역화 단계에서 수반되는 오류는 ORB-SLAM이 잘못된 지도를 만드는 결과로 이어질 수 있다.

시각적 어휘는 ORB-SLAM이 실행되기 전에 미리 준비되어야 하기 때문에 일반적인 이미지들에서 추출된 지역 특징들을 이용해 미리 학습을 시켜놓는 것이 보통이다. 하지만 이렇게 학습된 시각적 어휘는 ORB-SLAM이 수행되는 환경을 고려하여 제작된 시각적 어휘가 아니다. ORB-SLAM이 수행되는 환경을 미리 알 수 있다면 해당 환경에 가장 알맞은 시각적 어휘를 사용하는 것이 가장 좋은 재지역화 성능으로 이어질 것이다. ORB-SLAM의 수행 환경을 미리 아는 방법은 특정 응용이 사용되는 환경이 고정되어 있거나 ORB-SLAM을 수행하기 전 조도에 따른 실내/실외 인식, 조도에 따른 밤/낮 인식, GPS를 이용한 실내/실외 인식, 기타 외부 센서를 이용한 장소 정보 인식 등을 예로 들 수 있다.

본 논문은 일반화(generalization)와 전문화(specification)의의 정도에 따라 일반적인 시각적 어휘, 실내(indoor)/실외(outdoor) 시각적 어휘, 현장(onsite) 시각적 어휘 등의 3단계의 시각적 어휘를 제안하고, 실험을 통해 환경에 전문화된 시각적 어휘를 사용하는 것이 ORB-SLAM의 재지역화 성능을 개선할 수 있음을 보인다.

본 논문은 다음과 같이 구성된다. 2장에서는 ORB-SLAM에서의 시각적 어휘에 대해 살펴보고, 3장에서는 전문적인 시각적 어휘를 활용하는 방안에 대해 기술한다. 4장에서는 전문적인 시각적 어휘의 효과를 실험을 통해 분석한다. 마지막으로 5장에서 결론을 맺고 향후 연구 방향을 모색한다.

Ⅱ. ORB-SLAM과 시각적 어휘

2-1 시각적 어휘

지역 특징 기반의 이미지 인식의 문제점은 인식 대상의 수에 비례하여 저장 및 비교해야 할 특징의 수가 많아짐에 있다. 이는 1차적으로 메모리 문제를 야기하며, 2차적으로는 지역 특징 사이의 매칭 신뢰도를 급격히 하락시킨다. 이러한 메모리 문제와 매칭 신뢰도 문제를 한 번에 해결할 수 있는 방법이 시각적 어휘 기반의 이미지 인식[8]이다. 시각적 어휘는 다수의 시각 단어(visual words)로 구성된다. 다수의 지역 특징으로부터 미리 정의한 K개의 시각 단어를 학습해 두고, 새로운 지역 특징을 K 개의 시각 단어를 이용해 1 ~ K 사이의 숫자로 양자화(quantization)한다. 이는 다차원의 지역 특징 공간을 K 개의 공간으로 나누는 것으로 이해되며, 양자화된 지역 특징들을 저장해두고 같은 숫자끼리 매칭을 시도한다면 메모리 문제와 매칭 신뢰도 문제를 모두 해결하는 것이 가능하다. 하지만 일대다의 매칭이 수행되기 때문에 차후의 기하학적 검증(geometric verification) 등의 알고리즘을 통해 이상치(outlier)를 제거하는 것이 필수적이다. 뿐만 아니라 시각적 어휘 기반의 이미지 인식은 하나의 이미지를 하나의 벡터로 표현하는 것을 가능하게 한다. 양자화된 지역 특징들은 하나의 히스토그램(histogram)으로 표현이 가능하기 때문이다. 이를 통해 하나의 벡터로 표현된 이미지는 시점 변화 등에 강인하다는 지역 특징의 장점을 가짐과 동시에, 이미지 분류 등을 수행하는 분류기 등을 학습하는 목적으로 곧바로 쓰일 수 있다는 등의 장점을 가진다.

지역 특징을 양자화하기 위해서는 시각적 어휘의 학습이 선행되어야 한다. 시각적 어휘를 학습하기 위해서는 다수의 지역 특징들이 필요하며, 이는 새로운 이미지들을 사용할 수도 있고, 표현하고자 하는 이미지를 사용할 수도 있다. 표현하고자 하는 이미지에서 추출된 지역 특징을 사용한다면 해당 이미지들의 영역(domain)에 최적화된 시각적 어휘를 학습할 수 있다. 하지만 이렇게 학습된 시각 단어를 이용해 새로운 영역의 이미지를 표현한다면 그 변별력이 많이 떨어질 것이다. 이는 일반화와 전문화의 문제에 해당한다.

일반적으로 시각적 어휘를 학습하기 위해서는 비지도 학습(unsupervised learning)의 대표 주자인 군집화(clustering) 기법이 사용된다. 다수의 지역 특징들을 미리 정의한 K개의 군집(cluster)으로 나누고, 최종 군집들의 대표(reference) 특징들이 시각 단어가 된다. 유클리디안(Euclidean) 거리 기반의 k-평균(means) 군집화가 가장 일반적이며 이진 특징으로 구성된 시각적 어휘를 학습하기 위해 k-다수(majority) 군집화[9] 등이 제안되었다.

한편, 큰 규모의 시각적 어휘는 정확도뿐만 아니라 웹 스케일에서 발생하는 메모리와 검색 속도의 문제 또한 일부 해결해 준다. 큰 규모의 시각적 어휘를 이용해 이미지를 표현하여

벡터를 생성하면 대부분의 값이 0이 된다. 일반적으로 시각 단어의 개수가 추출된 특징의 개수보다 훨씬 많기 때문이다. 이런 희소성(sparseness)을 이용한다면 역색인(inverted index) 검색 기법을 적용할 수 있다. 모든 이미지를 벡터로 저장하는 것이 아니라, 각 시각 단어를 가진 이미지의 목록을 저장하고, 거꾸로 검색을 한다면 메모리 사용량과 검색 속도를 획기적으로 줄일 수 있다. 보통 큰 규모의 시각적 어휘는 학습이 비현실적으로 오래 걸리기 때문에 계층적 군집화(hierarchical clustering)[10]나 근사(approximate) k-평균[11] 등이 이용되고 있다.

2-2 ORB-SLAM

카메라를 주요 센서로 사용하는 시각적 SLAM은 다른 센서들에 비해 상대적으로 값싼 카메라만 갖추면 동작한다는 장점을 가진다. 시각적 SLAM은 보통 3단계의 파이프라인으로 구성된다. 추적(tracking) 단계에서는 현재 카메라 프레임에서 지역 특징(local features)을 추출한 후 모션 기반 무기조정(bundle adjustment)[4,5]을 수행한다. 프레임 중 일부는 키프레임(keyframe)으로 결정된다. 지역 지도 작성(local mapping) 단계에서는 지역 지도(local map)를 만들고 최적화한다. 키프레임의 지역 특징을 기 구축된 지도 포인트들과 대조하여 새로운 지도 포인트를 생성하거나 불필요한 지도 포인트들을 걸러낸다. 되돌아옴(loop closing) 단계에서는 같은 장소에 되돌아온 경우를 판단하여 그래프 최적화를 통해 누적된 오류를 보정한다.

시각적 SLAM의 대표 알고리즘인 ORB-SLAM[1,2]은 ORB라는 이진 특징을 지역 특징(local features)으로 사용한다. 이진 특징은 기존의 경사도(gradient) 기반 특징의 무거운 연산량을 해결하기 위한 방법으로 단순히 두 픽셀 쌍들의 밝기 값 비교를 통해 특징을 추출한다. ORB는 FAST 키포인트[6]와 BRIEF 기술자[7]에 방향 변화에 불변한 성질을 추가한 이진 특징이다. FAST는 특정 반지름에 놓인 모든 이웃 픽셀들 보다 어둡거나 밝은 픽셀을 키포인트(keypoint)로 검출한다. BRIEF는 키포인트를 중심으로 한 이미지 조각(image patch)에 대해 미리 정해놓은 일련의 픽셀 쌍의 밝기 값 비교를 통해 0과 1로 이루어진 비트 스트링을 반환하는 이진 기술자(binary descriptor)이다. 이진 특징은 경사도 기반 특징에 비해 다양한 변화에 다소 약하지만 최대 10배 이상의 빠른 연산량을 자랑한다. 뿐만 아니라 이진 특징은 최종 특징 벡터가 0과 1로만 이루어지기 때문에 경사도 기반 특징보다 훨씬 적은 메모리로 표현이 가능하며, 해밍(Hamming) 거리의 장점을 이용할 수 있어서 특징들 사이의 유사도를 빠르게 계산할 수 있는 부가적인 장점을 지닌다. ORB를 지역 특징으로 사용하는 ORB-SLAM은 지역 특징의 고속 추출과 고속 매칭이 가능하여 증강현실과 같이 실시간 SLAM을 필요로 하는 응용에 주로 활용되고 있다.

2-3 ORB-SLAM에서의 시각적 어휘

ORB-SLAM의 추적 단계에서 카메라의 포즈를 적절히 예측해내지 못한 경우를 추적에 실패했다고 한다. 카메라가 빠르게 움직이거나 충분한 수의 지역 특징을 추출하지 못하는 지점을 바라보았을 때 자주 발생한다. ORB-SLAM은 추적에 실패한 경우, 기존에 습득한 정보를 바탕으로 현재의 대략적인 위치를 다시 추정하는 재지역화 단계를 거친다. 재지역화 단계에서 발생하는 오류는 ORB-SLAM이 잘못된 지도를 만드는 결과로 이어질 수 있다. 결국 재지역화를 실패하면 실시간 시스템에서 장기적으로 구동되는 SLAM 시스템을 신뢰할 수 없게 된다. 재지역화 단계에서는 현재 프레임과 가장 닮은 키프레임을 검색하는 일이 가장 중요하며 선행되어야 한다. ORB-SLAM은 키프레임을 검색하기 위해 시각적 어휘 기반의 이미지 인식을 수행한다. 특별히 ORB-SLAM은 DBoW[12]라는 시각적 어휘를 사용하고 있는데 DBoW는 장소 인식(place recognition)을 목적으로 학습되고 공개된 시각적 어휘이다. 이렇듯 ORB-SLAM에서의 시각적 어휘는 추적에 실패했을 때 재지역화를 위한 목적뿐만 아니라 되돌아옴 단계에서 같은 장소에 되돌아온 경우를 판단하는 데에도 활용된다.

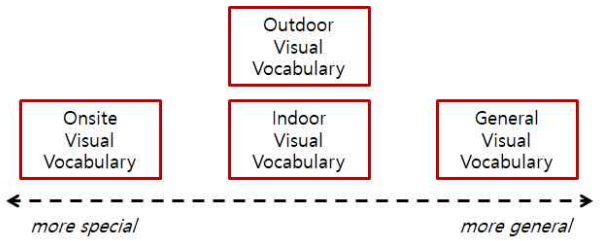

그림 1. 일반화와 전문화의 정도에 따라 다양하게 학습된 시각적 어휘들

Fig. 1. Various trained visual vocabularies according to the degree of generalization and specialization

Ⅲ. 전문적인 시각적 어휘 활용

ORB-SLAM에 탑재되는 시각적 어휘는 ORB-SLAM의 지역화 및 지도 작성의 전체 성능에 영향을 미치며 특히 재지역화 성능에 지대한 영향을 미친다. ORB-SLAM을 이용한 응용 어플리케이션을 개발할 때 대부분의 경우에는 의심할 여지없이 DBoW와 같이 누군가 미리 학습하여 배포해 놓은 시각적 어휘를 탑재하여 사용한다. 이러한 시각적 어휘는 이미지 인식이나 장소 인식 태스크를 통해 검증 과정을 거친 것이 대부분이긴 하지만 특정 응용 어플리케이션이 구동되는 환경에 따른 재지역화 성능을 고려하여 제작한 시각적 어휘가 아니다. 특정 응용 어플리케이션이 수행되는 환경에 대한 정보를 미리 알 수 있다면 해당 환경에 가장 알맞은 시각적 어휘를 학습하여 사용하는 것이 가장 좋은 ORB-SALM의 재지역화 성능으로 이어질 것이다.

일반적인 시각적 어휘를 사용한다면 모든 경우에 준수한 성능을 보이지만 전문적인 시각적 어휘를 사용한다면 특수한 경우에 더 좋은 성능을 발휘할 수 있을 것이다. 가장 쉬운 예로 실내와 실외를 구분하거나 낮과 밤을 구분하여 시각적 어휘를 학습할 수 있다. 그림 1은 이러한 일반화와 전문화에 따라 선택할 수 있는 시각적 어휘의 범위를 보여준다. 가장 오른쪽은 DBoW와 같이 환경의 구분 없이 수집된 이미지를 이용해 학습된 시각적 어휘를 선택한 경우이다. 특별히 특정 응용 어플리케이션의 환경을 제약하지 않거나 재지역화의 성능을 크게 신경 쓰지 않거나 새롭게 시각적 어휘를 학습할 수 없는 경우에 선택한다. 가운데 위치한 시각적 어휘는 실내/실외 혹은 밤/낮 등으로 구분하여 특정 어플리케이션이 수행되는 환경에 따라 다르게 선택한 경우이다. 특정 응용 어플리케이션이 실내에서만 수행된다는 보장이 있다면 실내에서만 촬영된 이미지들을 수집하여 전문적인 시각적 어휘를 학습할 수 있다. 가장 왼쪽은 가장 전문적인 시각적 어휘이며 특정 응용 어플리케이션이 특정 환경에서만 수행된다는 보장이 있다면 해당 환경에서만 촬영된 이미지들을 수집하여 가장 전문적인 시각적 어휘를 학습할 수 있다.

그림 2. 데이터베이스의 샘플 이미지들
Fig. 2. Sample images in the database

Ⅳ. 실험 및 분석

4-1 평가 데이터베이스 제작

특정 시각적 어휘의 사용에 따른 ORB-SLAM의 재지역화 성능을 평가하기 위해서는 일부러 추적에 실패하는 상황을 연출해야한다. ORB-SLAM은 추적에 실패했을 때 재지역화를 시도하기 때문이다. 이를 위해 카메라를 천천히 움직이다가 갑자기 빠르게 화면을 전환하여 추적을 놓친 상황을 만들고, 천천히 제자리로 돌아가는 방식으로 영상을 촬영하였다. 시각적 어휘의 일반화와 전문화 정도에 따른 재지역화 성능을 평가할 수 있도록 실내와 실외를 구분하여 촬영하였다. 그림 2는 이렇게 제작된 데이터베이스의 샘플 이미지를 보여준다.

4-2 평가 기준 마련

ORB-SLAM은 현재 상태에 따라 초기화 전(NOT_INIT), 추적성공(OK), 추적실패(LOST) 등 3개의 상태를 가진다. 영상의 전체 프레임에서 추적시도 대비 추적성공 비율을 이용해 시각적 어휘의 재지역화 성능을 평가한다. 특정 시각적 어휘가 재지역화 성능이 좋다면 추적실패 상태에서 추적성공 상태로 바뀌는 시간이 짧으며 이는 추적시도 대비 추적성공 비율이 높은 결과로 이어지기 때문이다.

표 1. DBoW 학습에 쓰인 이미지 데이터베이스
Table 1. Image database used for training DBoW

Dataset	Description	Image size
New College [13]	Outdoor, dynamic	512×384
Bicocca 2009-02-25b [14]	Indoor, static	640×480
Ford Campus 2 [15]	Urban, slightly dynamic	600×1600
Malaga 2009 Parking 6L [16]	Outdoor, slightly dynamic	1024×768
City Centre [17]	Urban, dynamic	640×480

4-3 시각적 어휘 준비

실험을 위해 시각적 어휘의 일반화와 전문화 단계를 3단계로 구성하였다. ORB-SLAM의 원 저자가 사용하고 있는 공개된 시각적 어휘인 DBoW를 일반적인 시각적 어휘로 정의하였다. 표 1은 DBoW 학습에 쓰인 이미지 데이터를 보여준다.

전문적인 시각적 어휘를 학습하기 위해 표 1의 이미지들 중에서 실내와 실외에 속한 이미지들을 따로 구분하여 새로운 시각적 어휘를 학습하고, 각각 실내 시각적 어휘, 실외 시각적 어휘로 정의하였다. 학습 도중 메모리가 부족한 경우가 발생해 가상 메모리를 디스크에서 크게 할당해 해결하였다.

가장 전문적인 시각적 어휘는 ORB-SLAM이 구동되는 현장에서 수집된 이미지들을 이용해 학습된 시각적 어휘이다. 4-1절에서 제작한 평가 데이터베이스의 영상들만을 이용해 시각적 어휘를 학습하였다. 즉, 각 영상마다 하나의 시각적 어휘를 학습하고, 현장 시각적 어휘로 정의하였다.

4-4 결과 및 분석

표 2. 시각적 어휘에 따른 추적 성공률
Table 2. Tracking success rate according to visual vocabulary

	General visual vocabulary	Indoor visual vocabulary	Outdoor visual vocabulary	Onsite visual vocabulary
Indoor1	0.824	0.829	0.818	0.83
Indoor2	0.857	0.860	0.851	0.862
Indoor3	0.896	0.899	0.89	0.902
Outdoor1	0.804	0.799	0.814	0.817
Outdoor2	0.811	0.809	0.815	0.819
Outdoor3	0.839	0.833	0.844	0.846
Total	0.838	0.838	0.838	0.846

표 3. 과적합된 현장 시각적 어휘
Table 3. Overfitted onsite visual vocabulary

Test \ Train	Indoor1	Indoor2	Indoor3	Outdoor1	Outdoor2	Outdoor3
Indoor1	0.83	0.823	0.82	0.808	0.81	0.812
Indoor2	0.853	0.862	0.853	0.843	0.846	0.847
Indoor3	0.892	0.895	0.902	0.881	0.884	0.882
Outdoor1	0.798	0.796	0.79	0.817	0.804	0.802
Outdoor2	0.805	0.80	0.81	0.808	0.819	0.81
Outdoor3	0.834	0.834	0.832	0.836	0.832	0.846

실험 결과를 표2에 정리하였다. 모든 영상에서 빠르게 화면을 전환하여 다른 곳을 바라보는 상황, 즉 추적에 실패한 상황에 대한 프레임은 2~3초 정도로 짧게 제작되었기 때문에 일반적인 시각적 어휘에서도 높은 추적 성공률을 보이고 있다.

실내 시각적 어휘는 일반적인 시각적 어휘에 비해 실내 영상에 대해서는 4.2% 정도 높은 성능을 보이지만 실외 영상에 대해서는 오히려 5.2% 정도 낮은 성능을 보이고 있다. 비슷하게 실외 시각적 어휘는 일반적인 시각적 어휘에 비해 실외 영상에 대해서는 7.7% 정도 높은 성능을 보이지만 실내 영상에 대해서는 오히려 7% 정도 낮은 성능을 보이고 있다. 실내 영상에 대해 실내 시각적 어휘를 사용하고, 실외 영상에 대해 실외 시각적 어휘를 사용한다면 일반적인 시각적 어휘에 비해 5.9%의 성능 향상을 얻을 수 있다.

현장 시각적 어휘는 일반적인 시각적 어휘보다 8.9% 높은 성능으로 모든 영상에 대해 가장 좋은 성능을 보이고 있다. 이는 실내 영상에 대해 실내 시각적 어휘를 사용하고 실외 영상에 대해 실외 시각적 어휘를 사용한 경우보다도 3% 정도 높은 성능이다. 현장 시각적 어휘가 가장 좋은 성능을 보이는 것은 학습 데이터와 테스트 데이터가 같기 때문이다. 이는 일종의 모델의 과적합(overfitting)으로 볼 수 있다. 일반적으로는 일반화를 높이기 위해 모델의 과적합을 피해야하는데 본 논문은 ORB-SLAM이 수행되는 환경이 고정되어 있거나 미리 알 수 있다는 가정 하에 모델의 과적합을 이용하였다. 표 3은 이러한 현장 시각적 어휘의 과적합을 보여준다. 학습 데이터와 테스트 데이터가 같을 때는 가장 좋은 성능을 보이지만 다를 때는 오히려 일반적인 시각적 어휘보다 좋지 않은 성능을 보이고 있다. 실내에 속하는 현장 시각적 어휘를 실외 영상에 적용하거나 실외에 속하는 현장 시각적 어휘를 실내 영상에 적용한 경우가 가장 좋지 않은 성능을 보이고 있다.

그림 3은 시각적 어휘에 따른 실내 영상에 대한 ORB-SLAM의 추적 상태 변화를 보여준다. 모든 시각적 어휘에서 동일하게 약 20번째 프레임에서 초기화를 끝낸 뒤 추적성공 상태를 유지하다가 약 350번째 프레임에서 추적실패 상태로 변환된다. 이때 카메라가 빠르게 전환되어 추적을 실패한 것이다. 이후 매 프레임에 대해 재지역화를 시도하게 되며, 재지역화를 성공한 경우 다시 추적성공 상태로 전환된다. 그림에서 보듯이 현장 시각적 어휘, 실내 시각적 어휘, 일반적인 시각적 어휘 순으로 재지역화를 성공하였다.

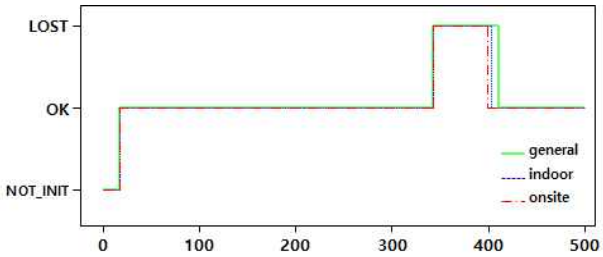

그림 3. 시각적 어휘에 따른 실내 영상에 대한 추적 상태 변화
Fig. 3. Tracking status change for an indoor video according to visual vocabulary

본 논문에서 사용한 영상들은 ORB-SLAM이 추적에 실패하는 상황을 연출했다가 잠시 후 다시 재지역화를 성공하도록 제작되었지만 일반적인 영상에서는 ORB-SLAM이 한번 추적을 놓친 이후에는 재지역화 성능에 따라서 이후 프레임들에서는 재지역화에 진입하지 못하는 상황도 발생한다. 또한 실시간 시스템에서는 ORB-SLAM의 재지역화의 성능 차이로 장기간으로 생성한 지도를 신뢰할 수 없게 되거나 증강현실과 같은 응용에서는 사용자의 몰입감을 해치는 요소로 작용한다.

Ⅴ. 결 론

본 논문에서는 ORB-SLAM의 재지역화 성능을 개선하기 위한 방법으로 전문적인 시각적 어휘 사용을 제안하였다. 일반화와 전문화의 정도에 따라 일반적인 시각적 어휘, 실내/실외 시각적 어휘, 현장 시각적 어휘 등의 3단계의 시각적 어휘를 학습하였다. 시각적 어휘의 전문화 정도에 따른 ORB-SLAM의 재지역화 성능을 비교하기 위해 평가 데이터베이스를 제작하고 평가 기준을 마련하였다. 실험 결과 ORB-SLAM의 재지역화 성능은 현장 시각적 어휘, 실내/실외 시각적 어휘, 일반적인 시각적 어휘 순으로 나타났으며, 이를 통해 환경에 전문화된 시각적 어휘를 사용하는 것이 ORB-SLAM의 재지역화 성능을 개선할 수 있다는 결론을 내릴 수 있었다.

감사의 글

본 연구 논문은 한국전자통신연구원 및 정보통신기획평가원의 출연금으로 수행하고 있는 한국전자통신연구원 실·가상환경 해석 기반 적응형 인터랙션 기술 개발 (2011-2021-0-00230) 위탁연구과제의 연구결과입니다.

참고문헌

[1] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, Vol. 31, No. 5, pp. 1147-1163, Oct 2015. https://doi.org/10.1109/TRO.2015.2463671

[2] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, Vol. 33, No. 5, pp. 1255-1262, Oct 2017. https://doi.org/10.1109/TRO.2017.2705103

[3] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proceeding of the International Conference on Computer Vision*, Barcelona, Spain, pp. 2564-2571, 2011. https://doi.org/10.1109/ICCV.2011.6126544

[4] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment a modern synthesis," in *Proceeding of the International Workshop on Vision Algorithms*, Corfu, Greece, pp. 298-372, 1999. https://doi.org/10.1007/3-540-44480-7_21

[5] R.Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[6] E. Rosten and T. Drummond, "Machine Learning for High-Speed Corner Detection," in *Proceeding of European Conference on Computer Vision*, Graz, Austria, pp. 430-443, 2006. https://doi.org/10.1007/11744023_34

[7] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary Robust Independent Elementary Features," in *Proceeding of European Conference on Computer Vision*, Crete, Greece, pp. 778-792, 2010. https://doi.org/10.1007/978-3-642-15561-1_56

[8] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proceeding of European Conference on Computer Vision*, Nice, France, pp. 1470-1470, 2003. https://doi.org/10.1109/ICCV.2003.1238663

[9] S. Lee, "Binary Visual Word Generation Techniques for A Fast Image Search," *Journal of KIISE*, Vol. 44, No. 12, pp. 1313-1318, Dec 2017. https://doi.org/10.5626/JOK.2017.44.12.1313

[10] D. Nister and H. Stewenius, "Scalable Recognition with a Vocabulary Tree," in *Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, NY, pp. 2161-2168, 2006. https://doi.org/10.1109/CVPR.2006.264

[11] J. Philbin, J, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, Minn, pp. 1-8, 2007. https://doi.org/10.1109/CVPR.2007.383172

[12] D. Gálvez-López, D and J. D. Tardós, "Bags of binary words for fast place recognition in image sequences," *IEEE Transactions on Robotics*, Vol. 28, No. 5, pp. 1188-1197, May 2012. https://doi.org/10.1109/TRO.2012.2197158

[13] M. Smith, I. Baldwin, W. Churchill, R. Paul, and P. Newman, "The new college vision and laser data set," *The International Journal of Robotics Research*, Vol. 28, No. 5, pp. 595–599, May 2009. https://doi.org/10.1177/0278364909103911

[14] A. Bonarini, W. Burgard, G. Fontana, M. Matteucci, D. G. Sorrenti, and J. D. Tardos, "Rawseeds: Robotics advancement through web-publishing of sensorial and elaborated extensive data sets," in *Proceeding of International Conference on Intelligent Robots and Systems*, Beijing, China, 2006.

[15] G. Pandey, J. R. McBride, and R. M. Eustice, "Ford campus vision and lidar data set," *The International Journal of Robotics Research*, Vol. 30, No. 13, pp. 1543–1552, Nov 2011. https://doi.org/10.1177/0278364911400640

[16] J.-L. Blanco, F.-A. Moreno, and J. Gonzalez, "A collection of outdoor robotic datasets with centimeter-accuracy ground truth," *Autonomous Robots*, Vol. 27, No. 4, pp. 327–351, Nov 2009. https://doi.org/10.1007/s10514-009-9138-7

[17] M. Cummins and P. Newman, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *The International Journal of Robotics Research*, Vol. 27, No. 6, pp. 647–665, June 2008. https://doi.org/10.1177/0278364908090961

조현우(Hyunwoo Cho)

2012년 : 한국과학기술원 대학원 (공학석사)

2012년～현　재: 한국전자통신연구원
※관심분야 : 증강현실(Augmented Reality), 컴퓨터비전(Computer Vision) 등

서영건(Yeong Geon Seo)

1987년 : 경상대학교 전산과 학사
1997년 : 숭싱대학교 전산과 박사

1989년～1992년: 삼보컴퓨터
1997년～현재: 경상국립대학교 컴퓨터과학부 교수
2014년～현재: 경상국립대학교 대학원 문화융복합학과 교수
2011년～현재: 경상국립대학교 공학연구원 멤버
※관심분야 : 의료 영상 처리, 머신 러닝, SLAM, 영상 인식,컴퓨터 네트워크, 컴퓨팅 사고

이수원(Suwon Lee)

2012년 : 한국과학기술원 (공학석사)
2017년 : 한국과학기술원 (공학박사)

2018년～현　재: 경상국립대학교 컴퓨터과학부 조교수
※관심분야 : 증강현실(Augmented Reality), 컴퓨터비전(Computer Vision) 등