

VILODE : A Real-Time Visual Loop Closure Detector Using Key Frames and Bag of Words

Hyesuk Kim[†] · Incheol Kim^{**}

ABSTRACT

In this paper, we propose an effective real-time visual loop closure detector, VILODE, which makes use of key frames and bag of visual words (BoW) based on SURF feature points. In order to determine whether the camera has re-visited one of the previously visited places, a loop closure detector has to compare an incoming new image with all previous images collected at every visited place. As the camera passes through new places or locations, the amount of images to be compared continues growing. For this reason, it is difficult for a visual loop closure detector to meet both real-time constraint and high detection accuracy. To address the problem, the proposed system adopts an effective key frame selection strategy which selects and compares only distinct meaningful ones from continuously incoming images during navigation, and so it can reduce greatly image comparisons for loop detection. Moreover, in order to improve detection accuracy and efficiency, the system represents each key frame image as a bag of visual words, and maintains indexes for them using DBoW database system. The experiments with TUM benchmark datasets demonstrates high performance of the proposed visual loop closure detector.

Keywords : Visual Loop Closure, Key Frames, SURF Feature Point, Bag of Visual Words, Consistency Checking

VILODE : 키 프레임 영상과 시각 단어들을 이용한 실시간 시각 루프 결합 탐지기

김혜숙[†] · 김인철^{**}

요약

본 논문에서는 키 프레임 영상과 SURF 특징점 기반의 시각 단어들을 이용한 효과적인 실시간 시각 루프 결합 탐지기 VILODE를 제안한다. 시각 루프 결합 탐지기는 과거에 지나온 위치들 중 하나를 다시 재방문하였는지를 판단하기 위해, 새로운 입력 영상을 이미 지나온 위치들에서 수집한 과거 영상들과 모두 비교해보아야 한다. 따라서 새로운 위치나 장소를 방문할수록 비교 대상 영상들이 계속해서 증가하기 때문에, 일반적으로 루프 결합 탐지기는 실시간 제약과 높은 탐지 정확도를 동시에 만족하기 어렵다. 이러한 문제점을 극복하기 위해, 본 시스템에서는 입력 영상들 중에서 의미 있는 것들만을 선택해 이들만을 비교하는 효과적인 키 프레임 선택 방법을 채택하였다. 따라서 루프 탐지에 필요한 영상 비교를 대폭 줄일 수 있다. 또한 본 시스템에서는 루프 결합 탐지의 정확도와 효율성을 높이기 위해, 키 프레임 영상들을 시각 단어들로 표현하고, DBoW 데이터베이스 시스템을 이용해 키 프레임 영상들에 대한 색인을 구성하였다. TUM 대학의 벤치마크 데이터들을 이용한 실험을 통해, 본 논문에서 제안한 시각 루프 결합 탐지기의 높은 성능을 확인할 수 있었다.

키워드 : 시각 루프 결합, 키 프레임, SURF 특징점, 시각 단어 집합, 일치성 검사

1. 서론

시각 루프 결합 탐지(visual loop closure detection) 기술은

카메라의 입력 영상들을 이용해 이동 경로 중 과거에 이미 지나온 위치나 장소 중 하나를 다시 재방문하였는지를 자동으로 판별해내는 기술이다[1]. 이러한 시각 루프 결합 탐지 기술은 잘 알려지지 않은 미지의 공간 안에서 움직이는 이동 물체의 위치 추적에 매우 중요한 핵심 기술로서, 영상 기반의 자율 주행(autonomous navigation), 장소 인식(place recognition), 위치 추적과 공간 지도 작성(Simultaneous Localization and Mapping, SLAM) 등 다양한 활용 분야를 가지고 있다[2, 3]. 하지만 정확한 시각 루프 결합 탐지를 위해서는 새로운 입력 영상을 이미 지나온 위치나 장소들에서 수집한 과거

※ 본 연구는 경기도의 경기도지역협력연구센터사업의 일환으로 수행하였음.
※ 이 논문은 2014년도 한국정보처리학회 추계학술발표대회에서 '키 프레임 영상을 이용한 실시간 시각 루프 결합 탐지기의 설계'의 제목으로 발표된 논문을 확장한 것이다.
† 준회원: 경기대학교 컴퓨터과학과 석사과정
‡ 종신회원: 경기대학교 컴퓨터과학과 교수
Manuscript Received: February 25, 2015
Accepted: March 8, 2015
* Corresponding Author: Incheol Kim(kic@kyonggi.ac.kr)

영상들과 모두 비교해보아야 한다. 따라서 새로운 위치나 장소를 방문할수록 비교 대상 영상들이 계속해서 증가하기 때문에, 일반적으로 루프 결합 탐지기는 실시간 제약과 높은 탐지 정확도를 동시에 만족하기는 어렵다.

이러한 문제점을 극복하기 위해, 본 시스템에서는 입력 영상들 중에서 의미 있는 것들만을 선택해 이들만을 비교하는 효과적인 키 프레임 선택 방법을 채택한다. 따라서 루프 탐지에 필요한 영상 비교를 대폭 줄일 수 있다. 또한 본 시스템에서는 루프 결합 탐지의 정확도와 효율성을 높이기 위해, 키 프레임 영상들을 SURF 특징점 기반의 시각 단어들(Bag of Visual Words, BoW)로 표현하고, DBoW 데이터베이스 시스템을 이용해 키 프레임 영상들에 대한 색인을 구성한다. 본 논문에서는 제안하는 실시간 시각 루프 결합 탐지기의 성능을 분석하기 위해, TUM(Technische Universität München) 대학의 벤치마크 데이터 집합[4]을 이용한 실험을 수행하고 그 결과를 소개한다.

2. 관련 연구

Sebastian의 연구[5]에서는 주로 컬러 영상(color image)을 루프 결합 탐지에 이용하던 기존의 연구들과는 달리, 깊이 영상(depth image)만을 이용하는 방법을 제안하였다. 이 연구에서는 컬러 영상 처리에 많이 사용되는 SIFT나 FAST와 같은 특징들을 깊이 영상들로부터 추출하고 이들을 루프 결합 탐지를 위한 영상 비교에 이용하였다. 한편, Kerl의 연구[6]에서는 컬러 영상과 깊이 영상을 모두 이용하는 루프 결합 탐지 방법을 제안하였다. 이 연구에서는 카메라 모션의 엔트로피(entropy)값을 기준으로 키 프레임들을 선별해내는 방법을 채택하였다. 그리고 루프 결합 탐지를 위한 두 영상 프레임 간의 비교에도 두 영상 사이의 카메라 모션 엔트로피값을 이용하였다. 하지만 Sebastian 및 Kerl의 방법 모두 영상들을 효율적으로 저장하고 검색하기 위한 별도의 색인 구조를 제공하지 않을 뿐 아니라, 원래 영상 자체를 비교하여 루프 결합을 탐지하기 때문에 광역 공간에 실시간으로 적용하기 어렵다는 공통적인 한계점이 있다. Galvez-Lopez의 연구[7]에서는 컬러 영상만을 이용하여 루프 결합을 탐지하는 방법을 제안하였다. 이 연구에서는 입력 영상들로부터 일정한 고정 시간 간격을 두고 키 프레임을 선택하는 방법을 채택하였다. 그리고 키 프레임들은 미리 정의해둔 시각 단어들(BoW)로 변환한 후, 루프 탐지를 위한 영상 비교에 이용하였으며, 키 프레임들을 효율적으로 저장하고 검색하기 위해 다양한 형태의 색인들을 이용하였다. 한편, Whelan의 연구[8]에서는 루프 결합 탐지를 위해 컬러 영상뿐만 아니라 깊이 영상도 함께 사용하는 방법을 제안하였다. 또한, 이 연구에서는 카메라 모션을 토대로 키 프레임을 선택하는 방법을 채택하였지만, 키 프레임 영상들의 저장과 검색을 위한 별도의 색인 구조는 제공하지 않았다.

3. 시각 루프 결합 탐지기 설계

이 절에서는 키 프레임을 이용한 실시간 시각 루프 결합 탐지기의 설계에 대해 설명한다. Fig. 1의 (a)는 루프 결합(loop closure)을 포함하고 있는 카메라 이동 경로의 한 예를 나타내고 있다. 이 그림에서 카메라의 이동 경로는 그림의 하단부에서 시작해서 시계 반대 방향으로 진행하여 다시 하단부로 돌아옴으로써, 작은 원으로 표시된 부분에서 루프 결합을 형성하고 있다. 본 논문에서는 카메라가 이동하는 동안 초당 약 30프레임의 컬러 영상이 카메라를 통해 입력된다고 가정한다.

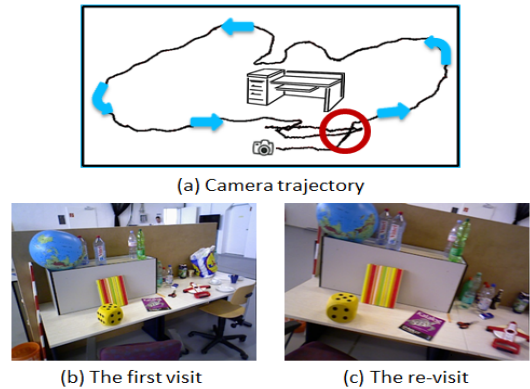


Fig. 1. Visual Loop Closure Detection

Fig. 1의 (b)는 카메라의 첫 방문 위치에서 입력한 영상을, (c)는 카메라가 해당 위치로 되돌아왔을 때 입력한 영상을 각각 나타낸다. 이 두 영상의 유사성을 토대로 카메라가 이전 위치로 되돌아왔다는 사실을 자동으로 판별해내는 것을 본 논문에서는 시각 루프 결합 탐지 문제로 가정한다. 따라서 본 논문에서 제안하는 시각 루프 결합 탐지기는 3차원 공간을 6자유도(DOF, Degree of Freedom)로 이동 중인 카메라의 입력 영상들을 이용하여, 그동안 지나온 이전 위치들 중 한 곳을 재방문하였는지를 실시간으로 판별해내는 시스템이다.

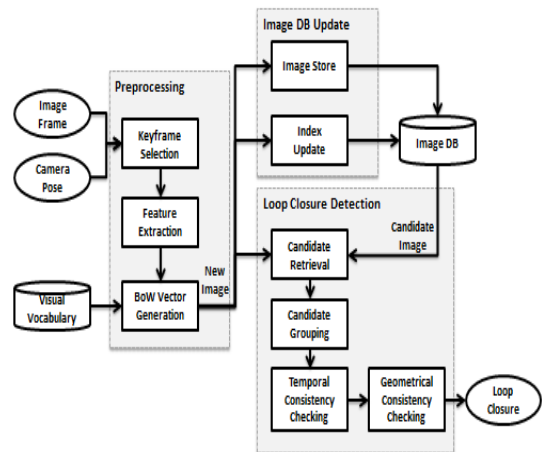


Fig. 2. System Architecture

Fig. 2는 본 논문에서 제안하는 시각 루프 결합 탐지기의 시스템 구조도를 나타낸다. 시각 루프 결합 탐지기는 크게 전처리부(Preprocessing), 이미지 데이터베이스 갱신부(Image DB Update), 그리고 루프 결합 탐지부(Loop Closure Detection)로 구성된다.

3.1 전처리

전처리(preprocessing) 과정은 키 프레임 선택(key frame selection) 단계, 특징 추출(feature extraction) 단계, BoW 벡터 생성(BoW vector generation) 단계들로 구성된다. 키 프레임 선택 단계에서는 루프 결합 탐지에 필요한 영상 프레임의 수를 효과적으로 줄이기 위해, 카메라 포즈(pose)의 변화량, 즉 주행 거리(odometry)를 기준으로 입력 영상들에서 키 프레임을 가려낸다. 따라서 이 단계에서는 마지막 키 프레임과 새로운 입력 영상 프레임 사이의 카메라 주행 거리가 일정 크기 이상이 되면, 이 영상 프레임을 새로운 키 프레임으로 선택한다.

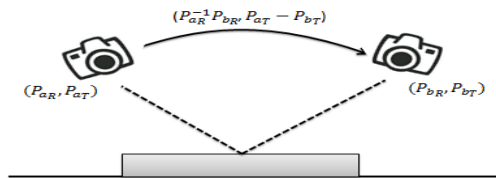


Fig. 3. Change of Camera Pose

Fig. 3은 3차원 공간에서 6자유도로 이동 중인 카메라의 포즈 변화를 나타내고 있다. 일반적으로 3차원 공간에서 움직이고 있는 카메라의 순간 포즈는 방향(orientation)과 위치(position)의 쌍으로 표현하며, 서로 다른 두 포즈 사이의 변화는 회전(rotation)에 의한 방향 변화와 병진(translation)에 의한 위치 변화로 나누어 표현할 수 있다. Fig. 3에서 (P_{a_r}, P_{a_t}) 와 (P_{b_r}, P_{b_t}) 는 이동 중인 카메라의 포즈들을 나타내며, $P_{a_r}^{-1} P_{b_r}$ 은 두 포즈 사이의 방향 변화를, $P_{a_t} - P_{b_t}$ 는 위치 변화를 각각 나타낸다. 이때 카메라 포즈의 변화량, 즉 주행 거리 m_{ab} 은 Equation (1)과 같이 방향 변화량과 위치 변화량의 합으로 계산할 수 있다.

$$m_{ab} = \|r(P_{a_r}^{-1} P_{b_r})\|_2 + \|P_{a_t} - P_{b_t}\|_2 \quad (1)$$

본 논문에서는 카메라의 포즈 간 주행거리가 일정한 임계값(threshold) 이상을 넘어설 때마다, 현재 입력 영상을 새로운 키 프레임으로 선택하는 주행거리 기반의 키 프레임 선택 방법을 이용한다.

특징 추출 단계에서는 키 프레임으로 선택된 컬러 영상을 회색 음영 영상(gray-scale image)으로 변환한 뒤, 이 영상으로부터 SURF 특징점(feature point)과 서술자(descriptor)를 추출해낸다. SURF 특징점 추출은 고속 헤이시안(fast Hessian) 방법을 이용한다. 고속 헤이시안 방법을 사용하기 위해서는 입력 영상으로부터 적분 영상을 구해야 한다.

$$I(x,y) = \sum_{i=0}^x \sum_{j=0}^y I(i,j) \quad (2)$$

Equation (2)는 (x,y) 번째 픽셀의 적분 영상값 $I(x,y)$ 을 구하는 계산식을 나타낸다. Equation (2)에서 $I(i,j)$ 는 입력 영상의 (i,j) 번째 픽셀의 화소값을 나타낸다. Equation (2)에 따라 적분 영상이 구해지면, 헤이시안 행렬식을 적용하여 SURF 특징점을 추출한다. Fig. 4의 (a)는 키 프레임 영상의 한 예를 나타내고, (b)는 회색 음영 영상으로부터 추출된 SURF 특징점들을 나타내고 있다.



(a) Key frame image



(b) SURF feature points

Fig. 4. Key Frame Image and SURF Feature Points

BoW 벡터 생성 단계에서는 입력 영상에서 추출한 SURF 특징 벡터들을 미리 정의해놓은 시각 어휘집(visual vocabulary)을 이용해 시각 단어들(BoW)로 변환한다. 한편, 시각 어휘집은 실시간 루프 결합 탐지 작업 이전에 동일 공간에서 미리 수집해놓은 훈련 영상 집합으로부터 SURF 특징 벡터들을 추출하고, 이들 중에서 영상을 대신 표현할 수 있는 대표 시각 단어를 선정하는 오프라인 작업을 통해 정의한다. 그리고 시각 어휘집에는 선정된 대표 단어들과 더불어 TF-IDF(Term Frequency-Inverse Document Frequency) 방식으로 계산한 해당 단어의 가중치(weight)들도 함께 저장한다.

3.2 이미지 데이터베이스 갱신

카메라의 이동 경로를 따라 수집되는 각 키 프레임들은 BoW 벡터 형태로 변환한 후, DBoW 데이터베이스 시스템에 저장한다. 효율적인 루프 결합 탐지를 지원하기 위해, DBoW 데이터베이스 시스템은 Fig. 5에 나타난 것과 같이 대용량의 영상 집합에 대한 어휘 트리(vocabulary tree), 정방향 색인(direct index), 그리고 역방향 색인(inverse index) 등을 구성한다.

어휘 트리는 오프라인 단계에서 추출한 SURF 특징 벡터들의 집합에 대해 몇 단계에 걸쳐 반복적으로 k-중심 군집화(k-median clustering)를 수행하여 특징 벡터들을 계층적으로 분할한 뒤, 군집들 간의 계층 관계를 트리 형태로 표

현한 것이다. 어휘 트리의 각 노드는 해당 군집을 대표하는 중심 벡터를 저장하고 있으며, 어휘 트리의 단말 노드에는 입력 영상을 표현하는 데 사용할 단어들 저장된다. 이러한 어휘 트리는 향후 입력 영상들에서 등장할 다양한 종류의 SURF 특징 벡터들을 미리 선정해놓은 유한개의 단어들로 대체함으로써, 이들을 이산화하는 목적으로 사용된다.

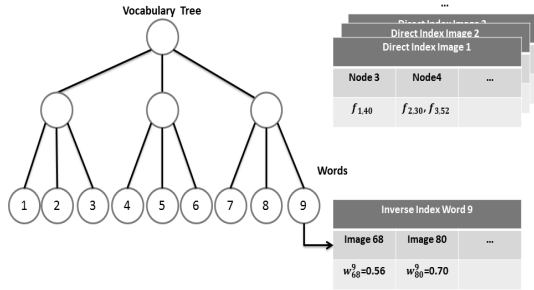


Fig. 5. DBoW Image Database System

어휘 트리에 따라 BoW 벡터로 변환된 두 영상의 유사도는 Equation (3)과 같이 계산되며, 이 유사도 계산은 3.3절에서 시각 루프 결합 탐지를 위한 후보들을 결정하는 데 이용된다.

$$s(v_1, v_2) = 1 - \frac{1}{2} \left| \frac{v_1}{|v_1|} - \frac{v_2}{|v_2|} \right| \quad (3)$$

한편, 역방향 색인에는 각 단어별로 해당 단어를 포함하고 있는 영상들에 접근할 수 있는 정보를 보관하고, 정방향 색인에는 각 영상별로 해당 영상에 포함된 시각 단어들에 접근할 수 있는 정보를 담고 있다. 새로운 키 프레임 영상이 DBoW 이미지 데이터베이스 시스템에 저장될 때마다, 이것을 반영하여 역방향 색인과 정방향 색인도 함께 갱신된다.

3.3 시각 루프 결합 탐지

카메라를 통해 새로운 키 프레임 영상이 입력될 때마다 수행되는 실시간 시각 루프 결합 탐지 과정은 후보 검색(candidate retrieval) 단계, 후보 그룹화(candidate grouping) 단계, 시간 일치성 검사(temporal consistency checking) 단계, 공간 일치성 검사(geometrical consistency checking) 단계 등 크게 4단계로 구성된다. 후보 검색 단계에서는 새로운 입력 영상과 루프 결합을 이룰 수 있는 후보 영상들을 DBoW 데이터베이스 시스템을 통해 검색한다. 새로운 입력 영상과 검색된 후보 영상이 각각 BoW 벡터 v_t 와 v_{t_j} 로 주어지면, Equation (4)를 이용하여 각 후보 영상의 정규화된 유사도 점수를 계산하고 미리 정해놓은 기준보다 낮은 유사도를 가진 영상들은 후보 집합에서 제거한다.

$$\eta(v_t, v_{t_j}) = \frac{s(v_t, v_{t_j})}{s(v_t, v_{t-\Delta t})} \quad (4)$$

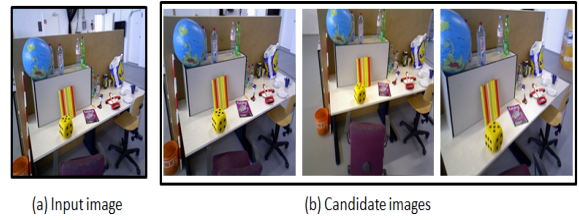


Fig. 6. Candidate Image Retrieval

Fig. 6의 (a)는 신규 입력 영상의 예, (b)는 이미지 데이터베이스에서 검색된 루프 결합 후보 영상들의 예를 각각 나타낸다. 한편, 후보 그룹화 단계에는 매우 가까운 인접 시간대에서 수집된 후보 영상들 n_1, \dots, m_i 을 찾아 그룹 V_{T_i} 단위로 묶는 작업을 수행한다. 가능한 규모가 큰 후보 그룹들을 생성하기 위해, 후보 그룹 V_{T_i} 는 신규 입력 영상 v_t 과의 유사도를 Equation (5)와 같이 평가한다.

$$H(v_t, V_{T_i}) = \sum_{j=n_i}^{m_i} \eta(v_t, v_{t_j}) \quad (5)$$

후보 그룹화 단계에 이어, 시간 일치성 검사 단계가 수행된다. 루프 결합을 찾기 위한 검색에서 시간 일치성은 질의 영상들(query images)의 시간대가 비슷하면 이들과 매치되는 후보 그룹들의 시간대도 서로 같거나 매우 가까워야 한다는 것을 의미한다. 따라서 입력 영상의 BoW 벡터 v_t 와 매치되는 후보 그룹 V_{T_i} 를 $\langle v_t, V_{T_i} \rangle$ 로 표시하면, 직전 k번의 입력 영상의 BoW 벡터들과 이들의 후보 그룹들 $\langle v_{t-\Delta t}, V_{T_1} \rangle, \dots, \langle v_{t-k\Delta t}, V_{T_k} \rangle$ 은 시간적으로 매우 인접해야 한다. 시간 일치성 검사 단계에서는 현재 입력 영상의 BoW 벡터 v_t 와 매치되는 모든 후보 그룹 V_{T_i} 들에 대해 이와 같은 시간 일치성 검사를 수행하고, 이 검사를 통과하는 각 그룹 V_{T_i} 에 대해서는 현재 입력 영상의 BoW 벡터 v_t 와 유사도가 가장 높은 후보 영상의 BoW 벡터 v_{t_j} 을 하나씩만 후보 집합에 유지하고 나머지는 삭제한다.

공간 일치성 검사 단계에서는 후보 영상이 공간 기하학적 관점에서 질의 영상과 충분히 일치하는지를 검사한다. 본 논문에서는 RANSAC(RANdom Sample Consensus) 특징 정합(feature matching) 방법을 적용하여 후보 영상이 최소 12개의 SURF 특징점들에서 질의 영상과 정합될 수 있는지 검사한다. 그리고 이 과정에서 SURF 특징 벡터들, 즉 단어 간의 계층적 유사도를 잘 표현하고 있는 어휘 트리를 활용함으로써 보다 효율적인 정합 작업을 진행한다.

4. 구현 및 실험

앞서 설명한 설계를 토대로 실시간 시각 루프 결합 탐지

기를 구현하고, 성능 평가 실험을 수행하였다. 시각 루프 결합 탐지기는 Windows 7 64bit, Intel Core @ 3.20GHz 플랫폼에서 visual studio 2010과 C++언어를 이용하여 구현하였다. 실험을 위한 데이터는 TUM 대학에서 제공하는 벤치마크 데이터 집합을 사용하였다. Table 1은 실험에 사용한 TUM 벤치마크 데이터의 종류와 수집 시간(s), 그리고 이동 거리(m) 등을 나타낸다. 사용된 데이터의 종류는 총 4가지로 각각 fr1/room, fr3/long office, fr2/pioneer_slam, 그리고 fr2/pioneer_slam2이다. 이들은 사무실, 복도 등 각기 다른 장소에서 다양한 카메라 움직임을 통해 수집한 데이터들이며, 모두 하나 이상의 루프 결합을 포함하고 있다.

Table 1. Test Datasets

Dataset	Duration (s)	Distance (m)
fr1/room	48.90 s	15.989 m
fr3/long office	87.09 s	21.455 m
fr2/pioneer_slam	155.72 s	40.380 m
fr2/pioneer_slam2	115.63 s	21.7553 m

본 논문에서는 시각 루프 결합 탐지기의 성능을 분석해보기 위해, 각 실험 데이터별로 다양한 키 프레임 선택 방법들을 적용하여 키 프레임당 평균 루프 결합 탐지 소요 시간(execution time)과 루프 결합 탐지 정확도(precision)를 측정해보았다. 루프 결합 탐지 정확도는 시스템이 탐지해낸 루프 결합점들의 실제 진위 여부를 사람이 판별하는 방법으로 측정하였다. 또, 키 프레임 선택을 위한 임계값은 카메라 주행 거리 기반의 선택 방법(Camera Odometry)은 주행 거리 임계값을 각각 0.04와 0.07, 0.1로, 고정 간격의 선택 방법(Fixed Interval)은 프레임 간격을 각각 0, 3, 5, 10으로 설정하여 실험을 수행하였다. 한편, 중복 특징 기반의 키 프레임 선택 방법(Feature Overlaps)은 FAST 특징 추출 방식과 RANSAC 특징 정합 방식을 적용하여 중복 특징들을 계산하였고, 최소 중복 특징 수를 각각 30, 50, 70개로 설정하여 실험을 수행하였다.

Table 2. Computation Time(ms/frame)

Dataset		fr1/room	fr3/long office	fr2/pioneer slam	fr2/pioneer slam2
Camera Odometry	0.04	680.0	808.6	157.5	252.1
	0.07	617.0	665.1	146.3	224.1
	0.1	551.8	717.0	<u>129.6</u>	<u>187.8</u>
Fixed Interval	0	682.1	814.5	547.8	419.6
	3	617.3	663.9	411.6	317.0
	5	572.6	724.2	404.1	302.7
	10	281.9	702.8	362.4	230.1
Feature Overlaps	70	326.9	300.2	802.7	682.3
	50	310.2	282.6	660.1	627.6
	30	<u>220.8</u>	<u>206.7</u>	720.8	586.0

Table 2는 실험의 결과로 얻어진 키 프레임당 평균 루프 결합 탐지 소요 시간(ms/frame)을 나타내고, Table 3은 루프 결합 탐지 정확도(%)를 나타낸다. 어떤 키 프레임 선택 방법을 적용하든 프레임 선택 간격이나 거리가 늘어남에 따라, Table 2의 키 프레임당 루프 결합 탐지 소요 시간은 감소한 반면, Table 3의 루프 결합 탐지 정확도는 향상된 결과를 보여준다. 이러한 실험 결과는 키 프레임을 이용한 루프 결합 탐지 방법이 소요 시간과 정확도 면에서 모두 긍정적인 효과를 가진다는 것을 의미한다. 한편, Table 2의 밑줄 친 부분들은 소요 시간이 가장 짧은 경우들을, Table 3의 밑줄 친 부분들은 정확도가 가장 높은 99% 이상의 경우들을 각각 나타낸다.

이들을 중심으로 살펴보면, fr2/pioneer slam와 fr2/pioneer slam2 같은 광역 공간 실험 데이터들에서는 주행 거리 임계값을 0.1로 설정한 카메라 주행 거리 기반의 키 프레임 선택 방법(Camera Odometry)이 소요 시간과 정확도 면에서 가장 우수한 결과를 보였다. 하지만 fr1/room과 fr3/long office와 같이 비교적 좁은 실험 공간 데이터들에서는 중복 특징 기반의 선택 방법(Feature Overlaps)이 소요 시간 면에서 가장 우수했고, 카메라 주행 거리 기반의 선택 방법(Camera Odometry)은 정확도 면에서 가장 우수한 결과를 보였다. 이와 같은 실험 결과를 통해, 본 논문에서 제안한 주행 거리 기반의 키 프레임 선택 방법(Camera Odometry)이 주로 광역 공간들을 중심으로 루프 결합 탐지 소요 시간 대비 루프 결합 정확도가 가장 높음을 확인할 수 있었다.

Table 3. Detection Accuracy (%)

Dataset		fr1/room	fr3/long office	fr2/pioneer slam	fr2/pioneer slam2
Camera Odometry	0.04	82.7	76.8	87.5	92.1
	0.07	98.5	92.0	98.6	97.5
	0.1	<u>100.0</u>	<u>99.0</u>	<u>100.0</u>	<u>100.0</u>
Fixed Interval	0	65.7	56.8	67.6	72.8
	3	75.0	62.7	83.4	81.5
	5	88.6	97.6	98.9	99.5
	10	93.5	<u>99.5</u>	<u>100.0</u>	<u>100.0</u>
Feature Overlaps	70	70.8	62.2	86.2	82.6
	50	92.7	72.3	94.5	94.3
	30	<u>99.6</u>	89.2	97.2	95.5

Fig. 7은 시각 루프 결합 탐지기의 성능을 정성적으로 분석해보기 위해, 실험 데이터별로 시각 루프 결합 탐지 결과를 카메라 이동 궤적 위에 표시해보았다. Fig. 7의 (a)는 fr1/room 데이터의 루프 결합 탐지 결과를 나타내고, Fig. 7의 (b)는 fr3/long office 데이터의 루프 결합 탐지 결과를 나타낸다. 그림에서 붉은색 선 부분은 루프 결합이 탐지된 위치들을 나타낸다. Fig. 7의 (a)와 (b)를 통해, 두 실험 데이터 집합 모두 루프 결합을 오판한 경우도 일부 포함하고

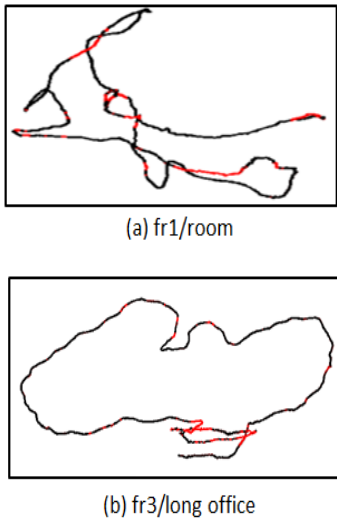


Fig. 7. Detected Loop Closures

있으나, 대부분 실제 루프 결합점들과 일치하는 탐지 결과들을 얻었음을 확인할 수 있다.

5. 결 론

본 논문에서는 키 프레임 영상과 SURF 특징점 기반의 시각 단어들을 이용한 효과적인 실시간 시각 루프 결합 탐지기를 제안하였다. 본 시스템에서는 비교 연산을 줄이고 실시간성을 만족하기 위해, 입력 영상들 중에서 루프 결합 탐지에 중요한 키 프레임들만을 효과적으로 선택하는 방법을 채택하였다. 또한 본 시스템에서는 루프 결합 탐지의 정확도와 효율성을 높이기 위해, 키 프레임 영상들을 오프라인 단계에서 미리 선정해둔 시각 단어들로 표현하고, DBoW 이미지 데이터베이스 시스템을 이용해 후보 영상들을 검색하도록 설계하였다. 실험을 통해 본 논문에서 제안한 시각 루프 결합 탐지기의 높은 성능을 확인할 수 있었다. 계획하고 있는 향후 연구로는 루프 결합 탐지를 위해 컬러 영상뿐만 아니라 깊이 영상도 함께 효과적으로 이용할 수 있도록 시스템을 확장하는 것이다. 즉, 깊이 영상으로부터 점 군(point cloud) 데이터를 얻은 다음, 이 점 군 데이터의 정합을 통해 보다 정확한 공간 기하학적 일치성 검사를 수행함으로써, 루프 결합 탐지 정확도를 더 향상시킬 수 있을 것으로 기대한다.

References

[1] Y. Liu, H. Zhang, "Visual Loop Closure Detection with Compact Image Descriptor," Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.1051-1056, 2012.

[2] A. Angeli, S. Doncieux, J.A. Meyer, and D. Filliat, "Real-Time Visual Loop-Closure Detection," Proceedings of IEEE International Conference on Robotics and Automation, pp. 1842-1847, 2008.

[3] K. L. Ho, P. Newman, "Detecting Loop Closure with Scene Sequences", *International Journal of Computer Vision*, Vol.74, No.3, pp.261-286, 2007.

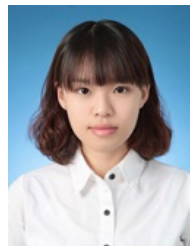
[4] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A Benchmark for the Evaluation of RGB-D SLAM Systems," Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012.

[5] S. A. Sebastian, A. Kloss, and A. Zell, "Loop Closure Detection using Depth Images," Proceedings of the European Conference on Mobile Robots, pp.100-106, 2013.

[6] C. Kerl, J. Sturm, and D. Cremers, "Dense Visual SLAM for RGB-D Cameras," Proceedings of IEEE/RSJ International Conference on Intelligence Robots and Systems, pp.2100-2106, 2013.

[7] D. Galvez-Lopez, J. D. Tardos, "Bags of Binary Words for Fast Place Recognition in Image Sequences," *IEEE Transaction on Robotics*, Vol.28, No.5, pp.1188-1197, 2012.

[8] T. Whelan, M. Kaess, J. J. Leonard, and J. McDonald, "Deformation-based Loop Closure for Large Scale Dense RGB-D SLAM," Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.548-555, 2013.



김혜숙

e-mail : chia@kyonggi.ac.kr

2013년 경기대학교 컴퓨터과학과(학사)

2013년~현 재 경기대학교 컴퓨터과학과 석사과정

관심분야 : 인공지능, 컴퓨터비전, 지능로봇



김인철

e-mail : kic@kyonggi.ac.kr

1985년 서울대학교 수학과(학사)

1987년 서울대학교 전산학과(이학석사)

1995년 서울대학교 전산학과(이학박사)

1996년~현 재 경기대학교 컴퓨터과학과 교수

관심분야 : 인공지능, 기계학습, 지능시스템