

딥러닝 기반 Keypoint/Descriptor 추출 및 Optical Flow 매칭을 통한 Visual SLAM 성능 향상 연구

이 영 현¹⁾ · 유 진 우^{*2)}

국민대학교 자동차공학전문대학원¹⁾ · 국민대학교 자동차IT융합학과²⁾

Improved Visual SLAM Framework via Deep Learning-based Keypoint/Descriptor and Optical Flow-based Matching

Yeonghyeon Lee¹⁾ · Jinwoo Yoo^{*2)}

¹⁾Graduate of Automotive Engineering, Kookmin University, Seoul 02707, Korea

²⁾Department of Automobile and IT Convergence, Kookmin University, Seoul 02707, Korea

(Received 14 April 2023 / Revised 26 May 2023 / Accepted 31 May 2023)

Abstract : In this paper, we are proposing a methodology that can replace handcrafted feature points and descriptors by using deep learning, and enhance image matching by using optical flow. To obtain feature points with robust repeatability and reliability, we will utilize R2D2, a deep learning network based on L2-Net. The inferred feature points and descriptors are only passed to SLAM if they satisfy the minimum requirements for repeatability and reliability. Moreover, in image areas where there is optical flow, matching is performed by relying on optical flow. Experimental results show that our methodology achieves lower path errors in terms of RMSE for the most part of the experimental data than the existing ORB Extractor in ORB SLAM.

Key words : SLAM(동시적 위치추정 및 지도 작성), Deep learning(딥러닝), Keypoint(특징점), Descriptor(기술자), Repeatability(반복성), Reliability(신뢰성), Optical flow(광류)

1. 서론

Visual SLAM(Simultaneous Localization And Mapping)은 카메라를 기반으로 이미지로부터 3차원 공간에서 측위 및 지도 작성을 동시에 수행하는 시스템이다. 카메라는 라이다보다 저렴하고 센서 배치에 유리하며 GPS와 HD(High Definition) 지도를 활용하기 어려운 환경에서도 사용할 수 있어서 실내 로봇, AR(Augmented Reality), VR(Virtual Reality), 자율 주행 등 다양한 분야에서 연구되고 있다.¹⁾ 이미지로부터 측위와 지도를 작성하기 위해서는 2개 이상에 이미지로부터 이미지 간에 기하학적 관계를 알아야 하는데, 이때 사용되는 대표적인 방식은 크게 Indirect methods와 Direct methods로 구분된다. Direct methods는 이미지 내에서 픽셀의 밝기 변화를 추적해서 카메라의 자세를 계산하는 방법이다. Indirect methods는

이미지에서 특징점을 추출하고 이러한 특징점들을 시간에 따라 변화하는 모습을 추적하며 지도를 작성한다. Direct methods는 가능한 모든 정보를 활용하기 때문에 이미지에 특징적인 텍스처가 부족하더라도 비교적 강한 성능을 보여준다. 그럼에도 불구하고 Direct methods는 두 프레임 사이에 밝기 일관성을 가정하기 때문에 조명에 변화가 큰 장면에서는 정확성이 좋지 않을 수 있다는 단점이 있다. 반면에 Indirect methods는 렌즈의 왜곡과 같은 기하학적 노이즈에 더 강하며 비교적 넓은 간격의 프레임을 매칭할 수 있고, 시점과 조명 변화에 강건하다는 장점이 있다.²⁾ 대표적인 Indirect methods SLAM에는 ORB SLAM, Parallel Tracking and Mapping(PTAM) 등이 있다. 특히 ORB SLAM은 현재까지도 전통적인 특징점 추출 방식에 슬랩에서 좋은 성능을 보여주고 있다.

*Corresponding author, E-mail: jwyo@kookmin.ac.kr

¹⁾This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited.

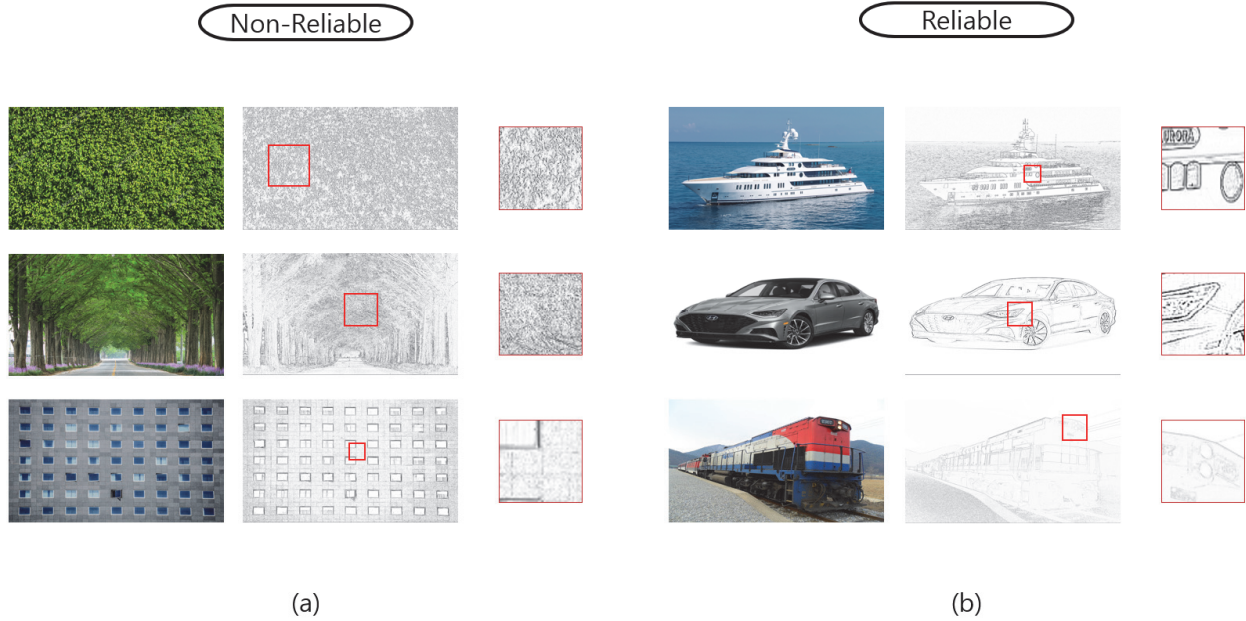


Fig. 1 The image on the (a) has a very strong edge line but low reliability. On the other hand, (b) has a very strong edge line and high reliability. Each patch image comes from the original image on the left

ORB(Oriented FAST and rotated BRIEF), FAST(Features from accelerated segment test), SIFT(Scale Invariant Feature Transform)³⁻⁵⁾와 같은 전통적인 특징 추출 알고리즘을 기반으로 하는 슬램은 빠른 속도와 높은 불변성을 보여주지만, 나뭇잎과 같이 유사한 텍스처(Fig. 1(a))가 반복되는 영역에서는 차별적인 특징점을 제대로 만드는 데 한계가 있다. 차별적이고 신뢰성 높은 특징점은 매칭의 정확성을 향상함으로써 모션 추정과 지도 작성에도 영향을 주기 때문에 매우 중요하다. Fig. 1(b) 우측에 있는 3개의 패치가 좌측 이미지 3장에서 각각 어디를 나타내는 지 정확하게 알 수가 있다. 이러한 영역에서의 특징점은 높은 신뢰성을 지니고 있을 것이다. 반면 Fig. 1(a)에 패치들은 신뢰성이 낮아 좌측에 원본 이미지에서 어디를 나타내는 지 정확히 알기가 어렵다.

최근에는 다양한 데이터로 학습된 딥러닝을 활용해서 조도나 회전 및 이동 변환에 강건한 특징점을 생성하는 모델들도 있다. 대표적으로 SuperPoint, GCNv2^{6,7)} 모델들은 CNN(Convolutional Neural Network)을 기반으로 특징점을 생성하는 딥러닝 모델이다. 이러한 딥러닝 기반의 특징점 생성 알고리즘으로 기존에 전통적인 알고리즘 기반의 슬램과 유사한 성능을 보여주는 사례도 있다.⁸⁾ 그러나 여전히 신뢰성을 고려하지 않은 특징점은 유사하고 반복적인 장면에서 좋지 못한 성능을 보여주었으며 몇몇 데이터에 대해서 꽤 큰 절대 경로 오차를 보여주었다. 본 논문에서는 신뢰성과 반복성을 동시에 고려

한 R2D2(Repeatable and Reliable Detector and Descriptor)⁹⁾를 활용하여 생성된 신뢰성 높은 특징점을 기반으로 특징점이 적거나 차별성이 부족한 장면에서 안정적인 SLAM을 수행하였으며, 동시에 모션 추정을 위한 매칭 정확도를 향상하기 위해 정제된 옵티컬 플로우를 활용한 방법론을 제안한다.

2. 관련 연구

컴퓨터 비전 분야는 다양한 환경과 조도, 시점에 대해서 반복적이고 신뢰성 있어야 하지만 그 동안에 전통적인 기법으로는 그 한계가 분명했다. 그러나 최근 딥러닝이 도입되면서 비약적인 발전이 일어나고 있다.¹⁰⁾ 이러한 수혜는 카메라 기반의 슬램 분야에서도 받고 있다. 실제로 딥러닝을 기반으로 하는 다양한 특징점 추출 모델들이 나오고 있으며 각 딥러닝 네트워크를 비교 평가하여 가장 적합한 모델을 선정하였다. 또한, 딥러닝을 통해 나온 특징점과 기술자는 시점과 환경에 따라 달라질 수 있는 불확정성이 있으므로 옵티컬 플로우를 통해 더 많은 매칭이 발생할 수 있도록 유도할 수 있어야 한다. 성공적인 매칭 유도를 위해 몇 개의 옵티컬 플로우 방법론을 비교 평가한 후 가장 적합한 방법론을 채택하였다.

2.1 ORB(Oriented FAST and Rotated BRIEF)

ORB³⁾는 FAST⁴⁾와 BRIEF(Binary Robust Independent Elementary Features)¹¹⁾를 기반으로 만들어진 특징점 추

출 및 기술자 생성 알고리즘이다. FAST는 중심 픽셀과 그 주변 원둘레 범위내에 픽셀들을 비교해서 임계값 이상에 변화를 감지한 경우 특징점을 추출하는 알고리즘으로 구현이 쉽고 빠르지만, 반복성이 높지 않다. ORB에서는 반지름 9 크기의 FAST-9 알고리즘을 사용하고 있다. 앞서 언급했듯 ORB에서는 BRIEF를 기반으로 기술자를 생성한다. BRIEF는 기술자를 이진화(Binarize)하여 기술자가 할당하는 용량의 크기를 줄였으며 시점과 조명에도 강인하다. ORB에서는 회전에 취약한 기존에 BRIEF를 회전 불변하게 만들기 위해서 특징점 주위의 그래디언트 방향을 계산하여, 회전 변환 행렬을 계산한 후 BRIEF 기술자를 회전 변환함으로써 회전이 발생하더라도 기술자를 생성하는 입장에서는 같은 기술자가 생성되게끔 유도한다. 이와 같은 단순한 방법은 특징점을 빠르고 많이 생성할 수 있지만 각 특징점의 신뢰성이 높은 편은 아니다. 특히 유사하고 반복적인 텍스처에서 특징점을 찾아내는 것은 단순히 두드러지는 명암으로 차별화하는 것이 어렵기 때문에 딥러닝을 통해 유사한 텍스처나 명암에 변화가 크지 않은 다양한 이미지를 학습함으로써 최적에 방법으로 반복성과 신뢰성이 높은 특징점을 찾아낼 수 있어야 한다. 본 논문에서는 앞서 언급한 반복성과 신뢰성 문제를 해결하기 위해 다음 장에서 소개할 딥러닝 방법론을 활용해서 ORB와 비교 평가할 것이다.

2.2 R2D2(Repeatable and Reliable Detector and Descriptor)

R2D2⁹⁾ 모델은 원래 기술자를 생성하기 위해 만들어진 L2-Net¹²⁾을 기반으로 되어 있다. L2-Net은 7개의 3 x 3, 8 x 8 컨볼루션과 LRN(Local Response Normalization layer)으로 구성되어 있다. R2D2에서는 연산의 효율성을 위해 마지막 8 x 8 컨볼루션 레이어를 3개의 2 x 2 컨볼루션 레이어로 대체해서 동일한 정확도에 가중치 개수를 5배까지 감소시켰다. 또한, 모든 레이어에 대해서 Dilated 컨볼루션 구조를 취함으로써 모든 컨볼루션 레이어에서 입력 해상도를 유지하며 픽셀들의 지역적 정보가 비교적 덜 손상되어 전파된다. R2D2는 L2-Net과 다르게 기술자뿐만 아니라 반복적이고 신뢰성이 있는 특징점을 추출해야 한다. 6개의 3 x 3 컨볼루션과 2 x 2 컨볼루션 3개를 지나서 나온 출력 텐서를 제곱 연산하고 신뢰성과 반복성 점수 지도를 출력하는 크기 1의 컨볼루션 레이어를 지나가게 된다. 최종 출력물로는 모든 픽셀에 대해서 128개 채널로 구성된 기술자 텐서와 1개 채널로 구성된 신뢰성 점수 지도 그리고 반복성 점수 지도로 구성된다. 점수 지도에는 픽셀별로 반복성과 신뢰성에 대한 점수

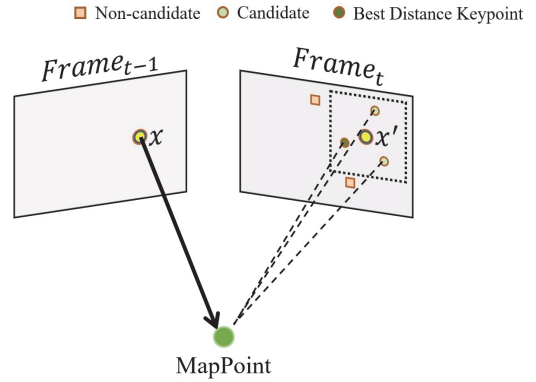


Fig. 2 To project onto the same map point from the current frame, select keypoints. The points within the bounding box are candidates that match the map point, and among them, the point with the closest distance becomes the best distance keypoint

값이 0에서 1 사이로 나타내어 있으며 임계값을 조절하여 적절한 특징점들을 추출할 수 있다. R2D2는 건축물 이미지 외에 가로수 길 같이 신뢰성 확보가 어려운 이미지도 함께 학습되어 있지만 신뢰성 및 반복성에 대한 점수 지도는 딥러닝을 통한 추론 결과이며 이는 불완전하다. 즉, 같은 위치에 대한 특징점이라도 기술자가 크게 달라질 수 있으며 이에 따라 단순히 기술자의 유클리드 안 유사도로 매칭하는 것은 SLAM에 치명적일 수 있다.

2.3 Optical Flow

연속되는 두 개 이상의 프레임이 존재할 때 프레임 내에서 변화가 있는 경우 두 프레임 간에 픽셀의 이동 방향과 거리를 알아낼 수 있다. 특징점은 이미지의 모든 위치에서 나타낼 수 있기 때문에 SLAM에서 사용하기 위해서는 조밀한 옵티컬 플로우를 만들어주는 Horn-Schunck¹³⁾와 같은 알고리즘이 필요하다. Horn-Schunck를 포함한 대부분에 조밀한 옵티컬 플로우 알고리즘들은 동일한 지점에 대한 밝기는 일정할 것이라는 밝기 항상성에 의존한다. 문제는 실제 외부 환경이 시간과 시점이 바뀔 때 따라 동일한 지점에서 밝기가 달라지기 때문에 플로우의 방향과 크기가 정확하지 않고 계속 변동적일 수 있다. TVL1¹⁴⁾은 Horn-Schunck와 유사한 전략으로 수행하는 밀집 옵티컬 플로우 알고리즘으로, Total Variation regularization과 L1 sparsity regularization을 통해 인접 픽셀값이 비슷하게 유지되도록 하여 경계선에서 픽셀값이 급격하게 변하지 않도록 하며 동시에 특정 특징이 두드러지게 나타나도록 유도한다. 이 두 가지 방식을 통해 이미지 변화량이 적을 때에도 정확한 대응 관계를 찾아내도록 하여 비교적 더 안정적인 밀집 옵티컬 플로우 생성이 가능하

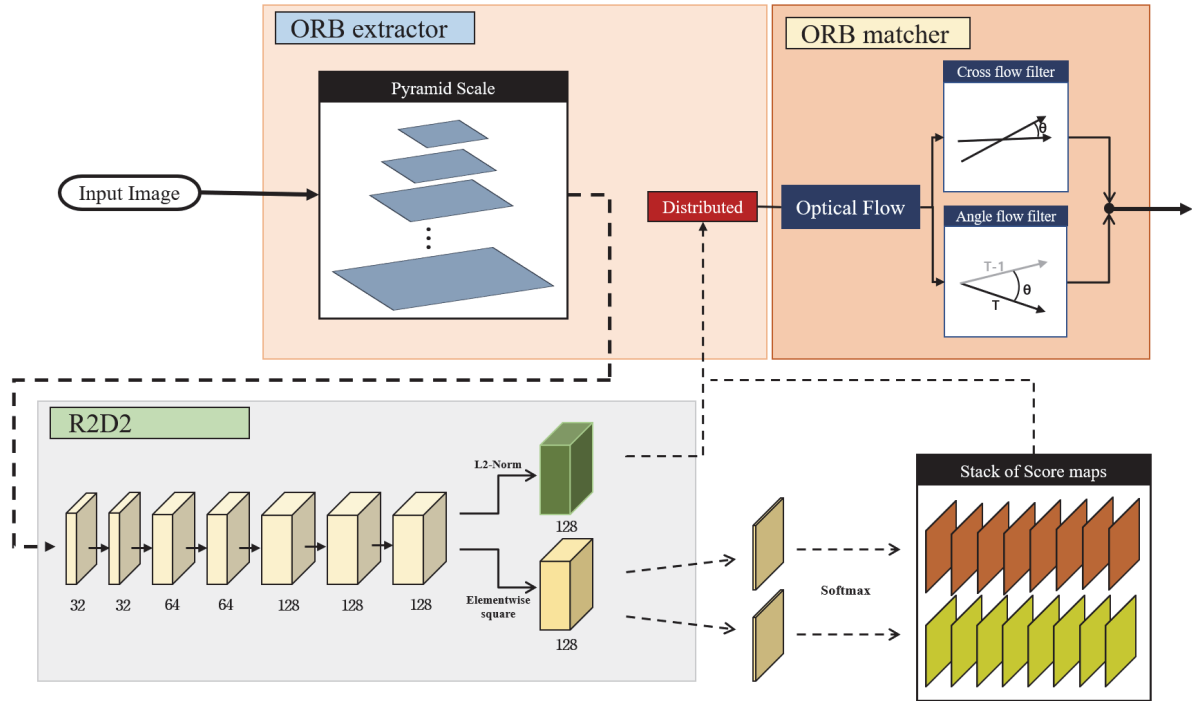


Fig. 3 Overview of our process system. You can see jointly R2D2 and ORBSLAM in ORB extractor. And you can see Optical Flow module for increase matching in ORB matcher

로 TVL1을 매칭 보조를 위한 옵티컬 플로우로 채택하였다. 또한, 밝기 향상성에 의존하는 한계를 극복하기 위해 옵티컬 플로우에서 신뢰할 수 있는 플로우만 선별하는 작업을 수행한다.

3. Proposed SLAM

본 연구는 전통적인 특징점 기반 SLAM 중 성능이 우수한 ORB SLAM¹⁵⁾을 기반으로 진행되었다.

3.1 Structure of Proposed Method

위에서 언급한 R2D2 모델은 ORB SLAM과 다른 언어를 기반으로 동작하기 때문에 두 개의 패키지를 결합하기 위해 언어에 상관없이 데이터 통신을 가능하게 해주는 ROS(Robot Operating System)를 사용하였다. ROS를 사용하기 위해서 노드별로 토픽과 데이터를 받을 수 있게 ROS화 하는 작업을 진행했다. 특징점을 생성하는 ORB extractor 단계에서 이미지가 도착하면 4에서 8개의 Pyramid scale로 다운 샘플링을 취한 후 이미지들을 차례로 딥러닝 모델 내에 전달하게 된다. 전달된 각 이미지는 9개의 레이어 층을 지나서 128개의 채널로 이루어진 기술자와 각 픽셀에 대한 신뢰성 및 반복성에 대한 점수를 0에서 1로 나타낸 Score map으로 만들어서 다시 ORB SLAM으로 전달하게 된다.

이후에는 기존에 ORB SLAM 시스템과 동일하게 동작하며 Keypoint 분배를 위해 사용되는 Response는 반복성 Score map의 점숫값에 의존한다. 전체 파이프라인에 대한 순서도가 Fig. 3에 나타내어 있다.

3.2 Optical Flow Task

모션에 움직임을 추적하거나 새로운 맵 포인트를 생성하기 위해서는 현재 이미지로부터 이전 이미지 나 기존에 맵 포인트와 매칭이 이루어져야 한다. 도출된 특징점들은 매칭 후보군에 기술자와 유사도를 계산해서 정해진 임계값보다 가까운 후보가 매칭 쌍으로 선정된다. 그러나 실제 환경에서는 같은 특징점에 기술자가 시점이나 조도에 따라 달라질 수도 있다. 만약에 매칭의 정확성을 위해서 임계값을 너무 낮게 하면 매칭 쌍들의 정확성은 높아지겠지만 영상이 진행됨에 따라 이전에 매칭되었던 쌍의 매칭이 안 되거나 잘못된 특징점과 매칭이 이루어질 수 있고 이미지가 단순한 경우 매칭 개수가 부족해서 추적에 실패할 수도 있다. 반대로 임계값을 높게 하면 잘못된 매칭 쌍이 많이 생성되어 누적 오차가 심해질 수 있다. 그러므로 본 논문에서는 기존에 유사도 측정을 통한 매칭뿐만 아니라 옵티컬 플로우를 사용해서 같은 영역을 나타내는 특징점인데도 불구하고 기술자가 달라져서 매칭이 이루어지지 않는 문제를 해결하였다.

2.3절에서 언급했듯이 옵티컬 플로우 알고리즘으로 TVLI을 선정하였다. 정확성을 위해서 TVLI을 선택했지만, 속도의 향상을 위해 이미지를 다운 샘플링하였으며 다음에 설명할 2개의 필터링 작업까지 끝낸 후 원래 크기로 업 샘플링한다. 보통 업 샘플링을 할 때는 주변에 값을 사용하는데 실제로 주변 픽셀들과의 플로우 차이가 크지 않으므로 허용할 수 있는 수준이다. 다운 샘플링을 많이 하면 속도가 빨라지고 적게 하면 정확성이 높아질 것이며 실험을 통해 속도와 정확성을 모두 고려하여 2배수로 선정하였다.

연속적인 매칭을 유지하기 위해서는 플로우가 지속적으로 생성되어야 하나 실제로 단일 이미지에서는 플로우가 생성되지 않는 영역을 자주 발견할 수 있다. 플로우의 주기적인 부재는 맵 포인트를 생성하거나 모션을 추적할 때 연속적인 매칭이 불가능하게 해서 최악에 경우 모션 추적이 실패할 수 있다. 그래서 본 논문에서는 단일 이미지에 플로우가 생성되지 않는 영역을 줄이기 위해 3배수로 다운 샘플링된 이미지의 플로우를 생성하여 기존에 플로우 맵에 추가함으로써 주기적인 부재를 감소시킨다.

3.3 Rotation and Cross Flow Filtering

밀집 옵티컬 플로우는 텍스처 상태에 따라 모든 픽셀에서 완벽하게 플로우를 생성할 수 없다. 픽셀과 그 주변이 유사한 흐름인 것을 생각하면 약간에 허용 오차는 괜찮지만, 과한 오차는 오히려 잘못된 매칭을 만들 수 있으므로 다음과 같은 2가지 과정을 통해 신뢰성 있는 옵티컬 플로우만을 남겨둔다. 첫 번째로 회전에 변화가 큰 경우 제거한다. 일반적으로 연속적인 동영상은 이미지와 다음 이미지의 변화가 매우 작으면서 연속적이다. 그러므로 이전 이미지와 현재 이미지에서 플로우 각도는 큰 차이를 나타낼 수 없다. 만약 동일한 지점에서 이전 플로우 각도와 현재 플로우 각도를 확인하여 차이가 큰 경우 플로우는 제거된다. 또한, 직진하는 상황에서는 플로우가 전 방향으로 작은 크기의 값을 가진다. 이전 이미지에 플로우를 기준으로 삼기 때문에 플로우에 크기가 너무 작은 경우 신뢰할 수 없는 값으로 생각하고 제거된다. 다음으로 플로우끼리 교차가 된 경우 제거된다. 두 개의 플로우가 교차가 되려면 큰 회전이나 움직임이 발생해야 하지만 일반적인 자동차의 움직임을 고려했을 때 무시해도 될 정도에 Roll과 Pitch가 발생하며 연속적인 영상에서는 Yaw에 변화도 크지 않으므로 플로우가 교차하였다는 건 옵티컬 플로우의 방향이 잘못되었다는 것을 알 수 있다. 옵티컬 플로우는 3.2절에서 언급한 2가지 케이스에서 사용되며 그중 하나는 모션을 추적할 때 사용된

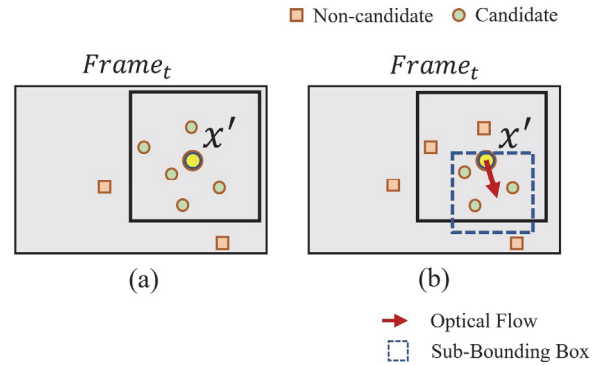


Fig. 4 Two different picture for Keypoint projection to map point. The left picture is original ORBSLAM. And right side is proposed our method

다. 모션 추적은 새로운 이미지가 들어올 때마다 이루어지며 이전 이미지에 존재하는 맵 포인트와 투영을 통한 매칭을 진행할 때 이전 이미지에 해당하는 키 포인트 위치에서 주변을 탐색한다(Fig. 4(a)). 다른 하나는 현재 프레임으로부터 새로운 맵 포인트를 만들 때 이미지 사이에 기하학적 관계를 구하기 위해서 사용되며 앞서 언급한 것과 동일한 과정으로 동작한다. 정제된 플로우가 생성된 후 플로우가 존재하는 위치에서는 플로우에 의존해서 비교적 여유 있는 임계값으로 매칭이 이루어지며 플로우가 제거된 위치인 경우 낮은 임계값을 사용하여 매칭된다. 이때 플로우를 통해 추정된 위치에서 기존에 바운딩 박스보다 작은 Sub-bounding box가 생성되며 기존에 바운딩 박스와 새롭게 생성된 바운딩박스의 교차 영역을 가장 신뢰할 수 있는 지역으로 한정해서 유사도를 확인한다(Fig. 4(b)). 각각에 대한 임계값은 실험을 통해 결정되었다.

4. Experiment

검증을 위해 KITTI 00에서 10번까지의 데이터를 사용하였으며, KITTI¹⁶⁾ 데이터에서 제공하는 Ground truth 값과 키 프레임 경로 간에 절대 경로 오차를 계산하여 Root Mean Squared Error(RMSE) 형태로 다른 모델들의 오차와 비교 평가한다.

4.1 Experimental Setting

본 연구의 실험 환경으로는 AMD Ryzen 7 3700X / RTX 3090 / RAM:128gb에서 진행되었다. R2D2에 대한 학습 데이터로는 Random web images, Aachen Day-Night이 사용되었으며 텍스처의 차별성이 적은 영역에서 Reliability 향상을 위해 검증에 사용되지 않으면서 유사한 텍스처로 반복되는 KITTI 12, 14, 17 데이터가 소량 선

별되어 사용되었다.

네트워크 훈련을 위해 Optimizer = ADAM, Learning Rate = 0.0001, MaxEpochs = 300으로 학습을 진행하였다. 이때 Repeatability와 키 포인트 개수에 영향을 주는 N 패치 크기는 실험을 통해 48로 결정하였다. 키 포인트 추출은 다양한 배경과 조도에서도 항상 발생해야 하므로 가장 Loss가 낮은 모델 대신 50회 학습된 모델을 검증용으로 사용하였다.

4.2 Experimental Results

실험 비교를 위해 기존에 ORB SLAM과 Deng 등⁸⁾에서 제안한 GCNv2-ORB SLAM 및 SuperPoint-ORB SLAM을 비교군으로 나타내었다. Table 1에 우리가 제안하는 알고리즘 R2D2 with optical flow와 앞선 비교군에 결과를 나타내었다. 각 방법론은 Keypoint 인지와 Matching을 제외하고는 ORB SLAM과 동일하게 작동되므로 Table 1에는 알고리즘별로 Keypoint 추출 FPS(Frame Per Second)와 Matching FPS가 작성되었다.

GCNv2에 경우 매우 효율적이고 경량 네트워크로 딥러닝임에도 빠른 속도를 보여주었다. 그러나 모든 KITTI 데이터에 대해서 SLAM을 수행하지 못해서 사용이 어렵다는 문제가 있다. 반면 SuperPoint를 활용한 모델에서는 이미지에서 특징점을 추출하는 과정이 가장 느리면서 동시에 기존 ORB SLAM에서 수행할 수 있었던 몇몇 데이터에서 SLAM 수행에 실패하는 결과를 볼 수 있다. R2D2 with optical flow의 경우 Optical flow 필터를 위해 모든 이미지를 탐색하는 과정에서 많은 시간이 소요되어 Matching FPS가 낮게 나오고 있지만 대부분에 데이터에 대해서 가장 낮은 RMSE를 보여주었으며 ORB와의 비교에서는 KITTI 03을 제외한 모든 데이터에 대해서 낮은 RMSE를 보여준다. 특히 KITTI 01번에 경우 ORB에서도 실패했으나 R2D2 with optical flow에서는 좋은 결과를 볼 수 있었다. 그러나 몇몇 데이터에 대해서는 SuperPoint가 좋은 결과를 보여주었는데, SuperPoint는 추출된 특징점 맵을 일정한 그리드로 잘라서 그리드별로 다른 임계값을 적용해서 특징점을 추출하기 때문에

Table 1 The table presents the RMSE results of ORB and Deep learning-based SLAM methods on the KITTI 0-10 dataset. The bold font represents the best-performing method, while the asterisk(*) indicates cases where the deep learning-based SLAM method outperforms ORB SLAM. The proposed R2D2 with optical flow method showed better performance than ORB SLAM on all data except KITTI 03, and exhibited stable performance without failing to perform SLAM on any of the data. While the SuperPoint-based ORB SLAM showed the best performance on some of the data, it frequently failed to initialize. On the other hand, GCNv2 failed to initialize on all data, resulting in overall instability or poor performance

KITTI	RMSE (m)			
	ORB SLAM	GCNv2 + ORB SLAM	SuperPoint + ORB SLAM	R2D2 with optical flow + ORB SLAM (Proposed)
00	6.93	Failure	Failure	5.01
01	Failure			16.02
02	27.02			
03	1.05		1.04	2.24
04	1.44		0.35	*1.39
05	6.04		3.73	*4.96
06	14.22		14.27	12.37
07	2.43		3.02	1.30
08	48.74		*39.63	30.86
09	60.45		Failure	24.46
10	9.53		5.31	*7.39
Average RMSE (m) (03-08, 10)	11.92	-	9.62	8.64
Success Rate	0.91	0.00	0.63	1.00
Keypoint (FPS)	1777.10	164.00	3.71	8.25
Matching (FPS)	2357.5	3926.5	2357.9	1.9394

모든 영역에 균일한 특징점을 생성함으로써 좋은 결과를 나타낸 것으로 판단된다. 실제로 KITTI 03, 04, 05, 10은 이미지의 좌, 우가 전혀 다른 텍스처 구조로 되어 있는 경우가 많아서 동일한 임계값일 때 불균형적인 특징점 추출이 이루어지는 것을 확인할 수 있다. 이렇듯 제안하는 SLAM은 일부 KITTI 데이터에서 SuperPoint 기반 SLAM에 비해 RMSE가 높았지만, Table 1에서 알 수 있듯이 3개 모델에 대해서 공통적으로 Failure가 발생하지 않은 데이터에 대한 Average RMSE 수치는 Proposed SLAM이 가장 낮은 것을 확인할 수 있었다. 또한, 전체 데이터들을 대상으로 SLAM 동작 성공 데이터의 비율을 나타내는 Success rate도 Proposed SLAM이 가장 높다는 사실을 통해 여러 환경에 대한 일반화적 성능도 가장 우

수한 것을 증명할 수 있었고, 이를 통해 제안하는 SLAM framework의 범용적인 성능 향상을 확인할 수 있었다. 추가로 데이터마다 RMSE 편차가 크게 나타나는 경향을 모든 SLAM에서 동일하게 확인할 수 있었다. 이는 데이터별로 강건한 특징점을 균일하게 만들기 어려운 장면에 분량으로 달라진다. 특히 이러한 장면에서 큰 회전이 자주 발생하거나 매칭 오류가 지속해서 발생하는 경우 스케일 드리프트가 크게 발생하게 된다. 일례로 KITTI 08 경우 초반에 강건한 특징점 추출이 어려운 장면에서 큰 회전이 자주 발생하여 스케일이 바뀌는 모습을 확인할 수 있다(Fig. 5(c)). Bold 채로 된 글씨는 가장 낮은 RMSE를 나타내며 *문구는 ORB보다 낮은 RMSE를 갖는 경우를 나타낸다.

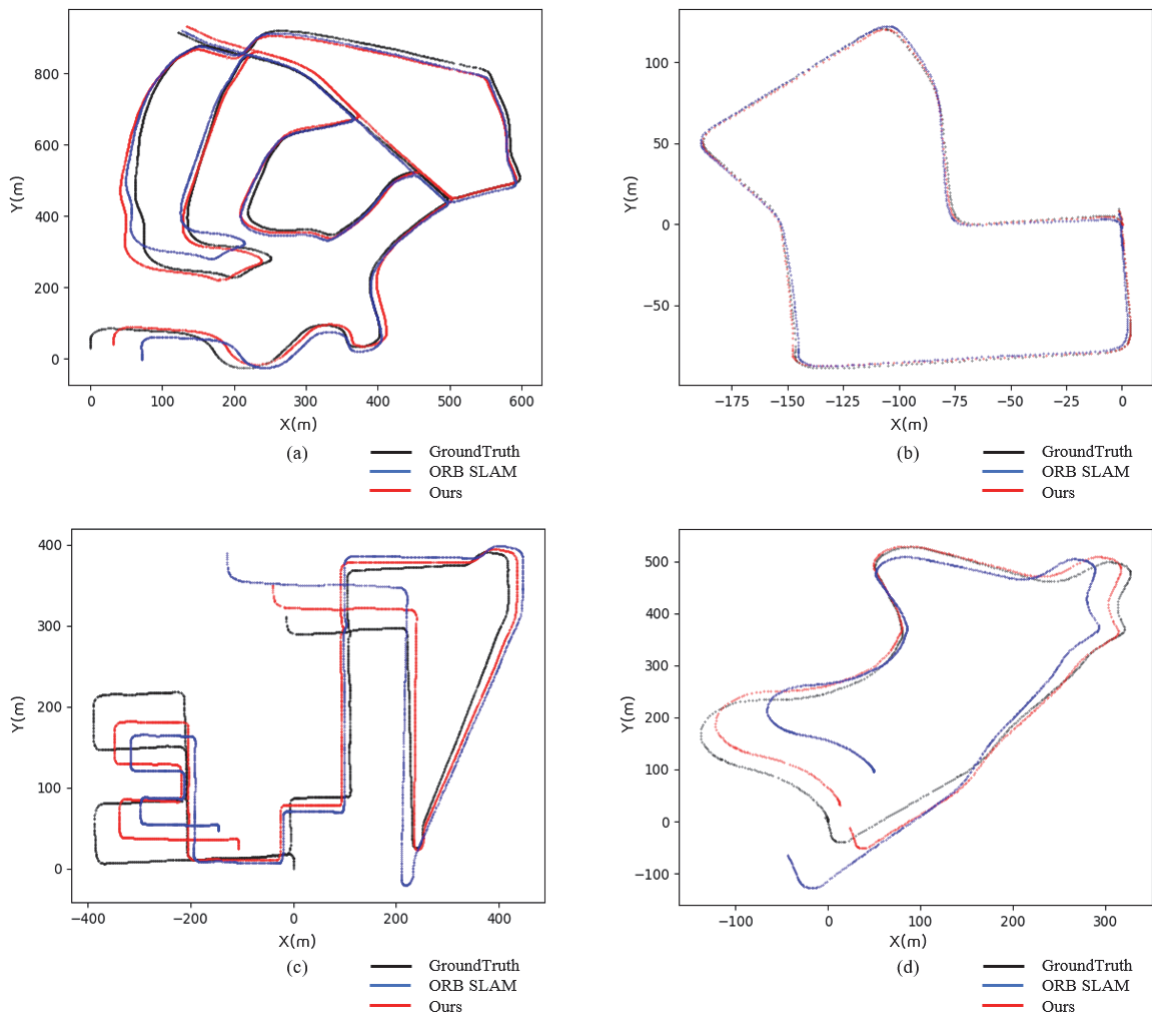


Fig. 5 Image comparing the path graphs of KITTI datasets that show large RMSE differences compared to ORB SLAM. The path graph results for KITTI 02, 07, 08, and 09 are shown in order from (a) to (d). The blue line represents the results of ORB SLAM, while the red line represents the results obtained using the method proposed in this paper. The black line represents the ground truth path

KITTI 02에서는 도로에 많은 부분이 나뭇잎이 무성한 나무들과 풀, 시멘트벽으로만 구성된 이미지가 자주 출현한다. 긴 직선 도로가 지속되다가 회전이 이루어지는 구간에서 경로 추정에 정확도가 크게 차이 남을 알 수 있다(Fig. 5(a)). Fig. 5(c)와 Fig. 5(d)에 경우에도 유사한 배경으로 길게 직선 연결된 도로가 많으며 이러한 구간에서 경로가 크게 벗어났음을 알 수 있다. 특히 Fig. 5(d)에서 초기 구간은 나뭇잎으로만 구성된 도로로 정확한 매칭이 어려운 환경이다. 실제로 ORB SLAM보다 본 논문에서 제안한 방법론이 Ground truth 경로에 매우 가깝다는 것을 알 수 있다.

5. 결론

본 논문에서는 기존에 Handcrafted 알고리즘의 한계를 넘고자 최근 컴퓨터 비전에서 적극 수용하고 있는 딥러닝을 활용했다. 반복적이고 신뢰성이 좋은 특징점과 기술자를 SLAM에 도입하여 특징점이 적거나 차별성이 부족한 장면에서도 SLAM 시스템을 성공시켰다. 또한, 모션 추적을 위한 투영 단계에서 이전 프레임에 동일한 특징점과 매칭을 하고자 할 때, 단순히 일정한 범위내에서 거뒀값만을 가지고 추적하였을 때는 유사한 텍스처가 반복되는 영역에서 잘못된 매칭이 발생되었다. 매칭 오차를 최소화하기 위해 Rotation과 Cross 필터를 사용해서 정제된 유틸리티 플로우를 통해 더 정확한 탐색 범위로 매칭 정확도를 높였다. 실제로 KITTI 데이터를 사용해서 R2D2와 TVL1 유틸리티 플로우 기반에 방법론이 기존에 ORB SLAM보다 성능이 더 좋음을 확인할 수 있으며 기존에 다른 Deep learning 기반에 SLAM과 비교해서도 성능이 우수함을 확인할 수 있었다.

제안된 시스템은 특징점을 비교적 큰 네트워크를 통해서 다양한 환경과 신뢰성을 확보할 수 있도록 학습하고 유틸리티 플로우를 통해 매칭을 보조하였다. 네트워크에 대해서 Knowledge distillation 방법론 등을 통해 경량화가 적극적으로 이루어지고 유틸리티 플로우 기법이 발전함에 따라 해당 시스템에 성능과 실시간성은 더욱 확보될 것이다.

후 기

이 논문은 2023년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임(P0020536, 2023년 산업혁신인재성장지원사업). 또한, 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(과제번호: NRF-2021R1F1A1062153).

References

- 1) Y. Lee and J. Yoo, "Improve Performance of Visual SLAM Using Keypoint and Descriptor based on Deep Learning," KSAE Fall Conference Proceedings, p.2092, 2022.
- 2) G. Younes, D. Asmar and E. Shammass, "A Survey on Non-filter-based Monocular Visual SLAM Systems," arXiv preprint arXiv:1607.00470, 2016.
- 3) E. Rublee, V. Rabaud, K. Konolige and G. Bradski, "ORB: An Efficient Alternative to SIFT or SURF," 2011 International Conference on Computer Vision, pp.2564-2571, 2011.
- 4) E. Rosten and T. Drummond, "Machine Learning for High-speed Corner Detection," Computer Vision—ECCV 2006: 9th European Conference on Computer Vision, pp.430-443, 2006.
- 5) D. G. Lowe, "Distinctive Image Features from Scaleinvariant Keypoints," International Journal of Computer Vision, pp.91-110, 2004.
- 6) D. DeTone, T. Malisiewicz and A. Rabinovich, "Superpoint: Self-supervised Interest Point Detection and Description," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp.224-236, 2018.
- 7) J. Tang, L. Ericson, J. Folkesson and P. Jensfelt, "GCNv2: Efficient Correspondence Prediction for Real-time SLAM," IEEE Robotics and Automation Letters, Vol.4, No.4, pp.3505-3512, 2019.
- 8) C. Deng, K. Qiu, R. Xiong and C. Zhou, "Comparative Study of Deep Learning Based Features in SLAM," 4th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS), pp.250-254, 2019.
- 9) J. Revaud, P. Weinzaepfel, C. D. Souza, N. Pion, G. Csurka, Y. Cabon and M. Humenberger, "R2D2: Repeatable and Reliable Detector and Descriptor," Proceedings of the 33rd International Conference on Neural Information Processing Systems, pp.12414-12424, 2019.
- 10) S. Kim, Y. Kim, H. Jeon, D. Kum and K. Lee, "Autonomous Driving Technology Trend and Future Outlook: Powered by Artificial Intelligence," Transactions of KSAE, Vol.30, No.10, pp.819-830, 2022.
- 11) M. Calonder, V. Lepetit, C. Strecha and P. Fua, "BRIEF: Binary Robust Independent Elementary Features," Computer Vision—ECCV 2010: 11th European Conference on Computer Vision, pp.778-792, 2010.
- 12) Y. Tian, B. Fan and F. Wu, "L2-Net: Deep Learning of Discriminative Patch Descriptor in Euclidean

- Space,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.661-669, 2017.
- 13) B. K. P. Horn and B. G. Schunck, “Determining Optical Flow,” Artificial Intelligence, Vol.17, Nos.1-3, pp.185-203, 1981.
 - 14) A. Chambolle and T. Pock, “A First-order Primal-dual Algorithm for Convex Problems with Applications to Imaging,” Journal of Mathematical Imaging and Vision, Vol.40, pp.120-145, 2011.
 - 15) 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, 2017.
 - 16) A Geiger, P Lenz, C Stiller and R. Urtasun, “Vision Meets Robotics: The KITTI Dataset,” The International Journal of Robotics Research, Vol.32, No.11, pp.1231-1237, 2013.