Journal of Institute of Control, Robotics and Systems (2025) 31(2):153-159
https://doi.org/10.5302/J.ICROS.2025.24.0252 ISSN:1976-5622 eISSN:2233-4335

MCVL: 계절 및 광원 변화에 강인한 다방향 카메라 기반 영상 측위

MCVL: Multi-directional Camera-based Visual Localization Resistant to Seasonal and Illumination Changes

문 기 영, 김 학 일*
(Giyoung Mun[1] and Hakil Kim[1,*])
[1]Department of Electrical and Computer Engineering, Inha University

Abstract: The global navigation satellite system-based technology has inherent limitations due to its reliance on radio signals. In contrast, visual localization operates independently of radio communication, presenting a viable solution to overcome these limitations. However, it is susceptible to seasonal and illumination variations, highlighting the need for research to address these challenges. Therefore, this paper proposes the multi-directional camera-based visual localization, which is robust against seasonal and illumination changes. The proposed method combines image from multiple directions and extracts global deep learning features. Subsequently, local deep learning features are extracted to preserve the characteristics of each combined image, allowing for the identification of geographically similar images. This approach utilize multi-directional cameras, enabling resilient performance under various constraints. Moreover, it demonstrates an improvement of 7.56% in recall rate at 1-meter threshold compared to existing methods.

Keywords: image retrieval, visual localization, multi-directional image, illumination invariant

I. 서론

측위는 자율주행, 로보틱스 등 다양한 분야에서 활용되는 기술이다. 대부분의 측위 기술은 GNSS (Global Navigation Satellite System) 기반의 측위 기술을 사용한다. 이때 높은 수준의 정확도가 필요한 경우 RTK (Real Time Kinematic)를 결합하여 사용하기도 한다[1-2]. GNSS/RTK는 매우 정확한 위치 추정 결과를 제공하지만, 전파 통신에 장애가 발생하는 경우 매우 큰 오차를 보이기도 한다. 이런 특징을 극복하기 위해 기존과 다른 특징을 가진 기술로 서로의 한계를 보완할 수 있는 연구가 필요하다[3].

Visual localization (영상 측위)은 카메라를 사용하여 시스템의 위치를 파악하는 기술이다. 카메라 영상을 사용하기 때문에, 전파장애 상황에서도 독립적인 동작이 가능하다. 그래서 GNSS/RTK가 제대로 동작하지 못하는 환경에서도 시스템의 위치를 알 수 있다는 장점이 있다. 이러한 특징으로 GNSS/RTK의 한계를 보완할 수 있는 기술로 평가받으며[4], 다양한 연구들이 제안되고 있다. 하지만 영상 측위를 실제 시스템에 적용하기 위해서는 아직 많은 연구가 필요한 상황이다.

그림 1은 현재 영상 측위가 가진 한계를 보여주는 정량 평가 결과이다. Oxford Robotcar [14] 데이터셋을 이용하였으며, 데이터셋에 5월 19일 입력 데이터로 8월 28일 데이터를 사용한다. 그림 1에서 왼쪽은 입력 이미지(빨간 점)와 데이터셋(파란 점)이 촬영된 위치를 보여준다. 이 중 위치 추정 오차가 500m 이상일 경우 초록색 점으로 표시한다. 그리고 초록색 점의 정성 평가 결과가 오른쪽에 있다. 그림 1의 정성 평가 결과 영상 측위 기술은 역광과 비슷한 장면에서 정확도가 떨어지는 한계가 있음을 알 수 있다.

영상 측위가 역광에 취약한 이유는 역광이 발생하면 데이터셋에서 역광이 발생한 이미지를 찾기 때문이다. 이런 현상은 기술의 신뢰도와 안정성을 하락시키기 때문에 이를 보완하는 방법이 요구된다[3-4]. 역광 상황에도 정상적인 영상 측위를 하기 위해서는 역광이 발생하지 않은 다른 방향의 이미지를 사용해야 한다. 그래서 본 논문은 계절 및 광원 변화에 강인한 다방향 카메라 기반 영상 측위를 제안한다. 제안하는 방법의 목적은 다음과 같다.

- 전파에 종속적인 기존 측위 기술의 대안 제시
- 계절 및 광원에 강인한 실외 영상 측위 기술 제공
- 다방향 카메라를 사용한 새로운 영상 검색 기법 제안
- 이미지의 지역적 인코딩을 통한 다방향 이미지의 강건한 표현법 제안

*Corresponding Author
Manuscript received November 1, 2024; revised December 11, 2024; accepted December 26, 2024
문기영: 인하대학교 전기컴퓨터공학과 대학원생(gymoon@inha.edu, ORCID 0009-0002-0374-0576)
김학일: 인하대학교 스마트모빌리티공학과 교수(hikim@inha.ac.kr, ORCID 0000-0003-4232-3804)
※ 이 논문은 2024년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임 (P0017124, 2024년 산업혁신인재성장지원사업).

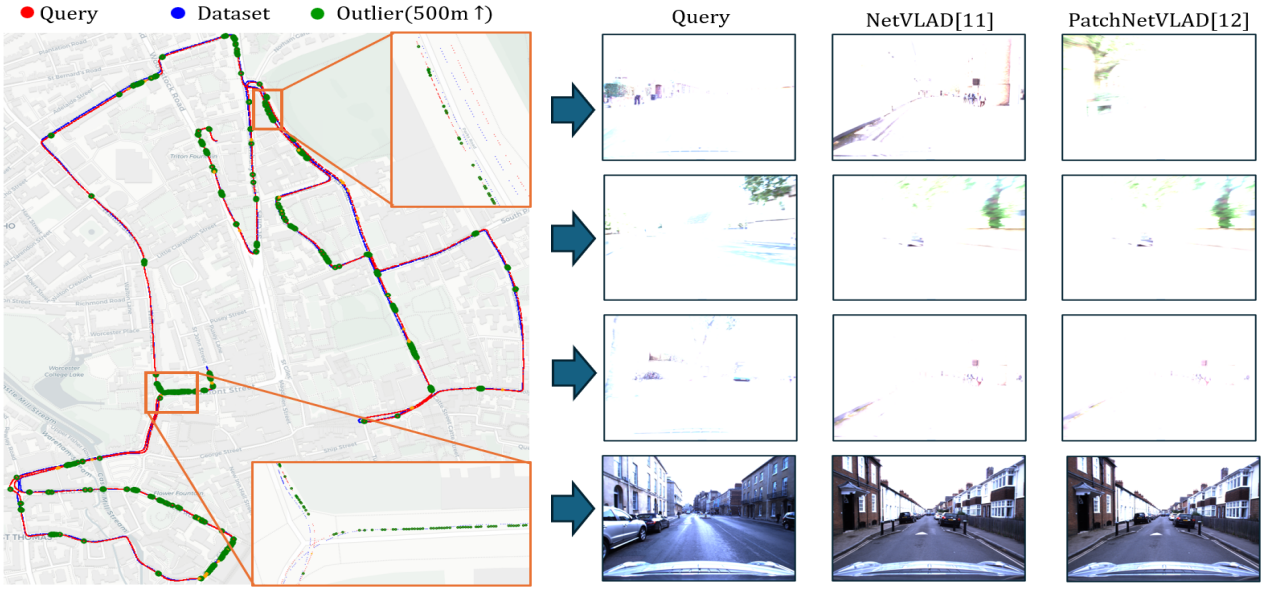

그림 1. 정성 평가 결과: 첫 번째부터 세 번째 행은 기존 방법이 역광으로 실패(500m+ 오차)한 경우. 마지막 행은 비슷한 장면에서 실패한 경우.

Fig. 1. Qualitative evaluation results: The first to third rows show cases where the existing method failed due to backlighting (500m+ error). The last row shows a case where the existing methods failed.

제안하는 방법은 위 목표를 위해 다양한 방향에서 촬영한 이미지들을 합친 후 전역 VLAD [11] 벡터로 인코딩한다. 이후 전체 데이터셋에서 후보군을 선정하여 슬라이딩 윈도우(sliding window) [13]를 활용한 지역 인코딩을 한다. 그리고 후보군 중 입력 이미지와 지리적으로 가까운 이미지를 다시 선택하여, 가장 유사도가 높은 이미지를 선택한다. 실험은 Oxford Robotcar [14]로 진행하고, RTK와 전, 후, 좌, 우 카메라가 서로 30ms 이내에 측정된 경우만 사용한다. 실험 결과 기존 1m recall rate에서 최대 7.56% 개선된 결과를 보인다.

II. 관련 연구

1. 영상 검색

영상 측위는 카메라 입력으로 들어온 영상의 위치를 추정하는 기술이다. 데이터셋의 형태에 따라 Structure-based (구조체 기반)와 영상 검색 기반으로 구분할 수 있다[5]. 구조체 기반 방식은 3차원 지도에서 위치를 추정한다. 3차원 지도는 LiDAR [6-7]나, SfM (Structure-from-Motion) [8]을 사용하여 제작한다. 하지만 LiDAR를 사용하는 경우 서로 다른 특성을 해결하기 어렵다는 한계가 있다[7]. 그리고 SfM을 사용하는 방식은 지도 제작에 노력이 필요하고, 넓은 지역에서 불리하다. 그래서 넓은 범위의 영상 측위에서는 구조체 기반 방식보다 영상 검색 기반 방식이 주로 사용된다[5].

영상 검색 기반 방식은 구조체 기반 방식보다 지도 구축이 편리하고, 넓은 장소에서의 위치 추정이 가능하다. 이전에 이미지와 위치가 같이 기록된 데이터셋을 만든 뒤 입력 이미지와 가장 유사한 이미지를 고른다. 그리고 입력 이미지와 데이터셋에서 찾은 이미지 사이의 상관관계를 통해 정밀한 위치를 추정한다. 대부분 SOTA (State-Of-The-Art) 수준의 영상 측위는 영상 검색 기반으로 되어있다[5]. 초기 영상 검색은 이미지 특징점을 사용[9-10]하지만, 최근엔 딥러닝을 사용한 연구들이 활발하다.

BoW (Bag of Words) [9]는 SIFT [15]나 SURF [16] 같은 지역 특징을 히스토그램으로 표현한 방식이다. 하지만 특징점의 위치나 구조적 관계를 사용하지 않아 공간 정보가 손실된다는 한계가 있다. 이후 제안된 FV (Fisher Vector) [17]는 BoW보다 특징점을 정교하게 표현하지만, 계산 복잡도를 간소화할 추가적인 연구가 필요하다. VLAD [10]는 이미지에서 추출한 지역 특징점이 클러스터 센터부터 얼마나 떨어져 있는지 벡터로 표현한다. FV보다 계산은 간단하고, BoW보다 정교한 표현이 가능하다.

NetBoW [18], NetFV [19], NetVLAD [11]는 기존 방법들을 신경망으로 학습한 결과를 보여준다. 하지만 딥러닝을 사용하여 전체 이미지를 축약적으로 표현하는 경우 이미지의 지역적인 특징을 표현하기 힘들다는 한계가 있다. Patch-NetVLAD [12]에서는 NetVLAD [11]를 활용한 지역적 표현과 매칭 방법을 제안한다. 그리고 계절 변화의 강인함을 보이기 위해 Oxford Robotcar Seasons V2 [14]를 이용하여 제안 방법의 우수성을 입증한다. 하지만 그림 1의 경우처럼 역광이 발생하는 상황을 보완할 방법이 필요하다.

2. 영상 검색 평가를 위한 데이터셋

영상 검색을 평가하기 위해서 같은 장소를 서로 다른 시간과 조건에 촬영한 데이터셋이 필요하다. Aachen Day-Night [20]은 낮과 밤의 성능을 평가할 수 있는 데이터셋을 제안한다. Baidu Mall, InLoc [21], GangnamStation는 실내에서 다양한 조건을 제공한다. 하지만 제안하는 방법을 평가하기 위해서는 다른 계절과 시간에 다방향 카메라를 사용한 데이터셋이 필요하다. Oxford Robotcar [14]는 다방향 카메라로 같은 경로를 1년 동안 100번 이상 기록한 데이터셋이다. 이 데이터셋을 이용하면 시간, 계절에 따른 종속성과 다방

향과 단방향 영상을 동시에 평가할 수 있다.

제안하는 방법을 평가하기 위해서는 실외에서 다방향 카메라를 사용하여 다양한 시간에 걸쳐 측정한 데이터가 필요하다. 본 논문에서 조사한 한 Oxford Robotcar [14]는 이 조건을 만족하는 거의 유일한 데이터셋이다. 같은 이유로 Patch-NetVLAD [12]에서도 동일한 데이터셋만 사용하여 계절 변화에 강인한 성능을 평가한 사례가 있다. 표 1은 관련 데이터셋을 정리한 자료이다. 그래서 본 연구는 Oxford Robotcar [14]를 사용하여 제안하는 방법을 평가한다.

III. MCVL: 다방향 카메라 영상 측위

제안하는 방법은 크게 이미지 병합, 전역 인코딩, 지역 인코딩, 유사도 평가로 구성된다. 제안하는 방법의 전체적인 구성은 그림 2의 시스템 개요도를 통해 확인할 수 있다. 우선 서로 싱크가 맞는 다양한 방향의 이미지를 병합한다. 그리고 전역 인코딩을 통해 딥러닝 벡터로 변환한 뒤, 데이터셋에서 후보군을 선택한다. 그리고 합쳐진 각 이미지의 특징을 살리기 위해 지역적인 인코딩을 하여 후보군 중 가장 지리적으로 가까운 이미지를 고른다.

이미지 병합에서는 네 방향의 이미지를 하나로 합치며 이를 식 (1)로 표현한다. 식 (1)에서 I_f, I_b, I_r, I_l는 순서대로 전방, 후방, 우측, 좌측 이미지를 의미한다. 그리고 I_{concat}은 합친 이미지 $\oplus$는 이미지 연결을 의미하는 수식이다. 이때 서로 다른 방향의 이미지를 동시에 사용하는 이유는 역광처럼 이미지 특징점이 전부 소실되는 상황에서도 다른 방향의 이미지를 사용하여 이미지 사이의 유사도를 찾기 위해서이다.

$$I_{concat} = I_f \oplus I_b \oplus I_r \oplus I_l \tag{1}$$

이미지를 병합한 이후엔 전역 인코딩한다. 전역 인코딩은 이미지를 NetVLAD [11] 벡터로 변환하는 과정으로 식 (2)로 표현한다. 식 (2)에서 $VLAD(\cdot)$는 이미지를 입력으로 받고, 출력으로 NetVLAD [11] 벡터를 반환하는 함수다. 원래 NetVLAD [11] 출력으로 나온 벡터 X_k는 4096차원이다. 하지만 제안하는 방법은 지역 인코딩을 추가로 하기에 PCA를 이용하여 512차원으로 축소한다. 512차원으로 축소하는 이유는 적은 메모리 사용으로 지역 인코딩을 통해 이미지의 지리적인 특징을 충분히 표현할 수 있기 때문이다.

$$X_k = VLAD(I_k) \tag{2}$$

입력 이미지를 $VLAD(\cdot)$ 함수로 벡터 변환한 결과를 $G=\{g_1,g_2,\cdots,g_{512}\}$라고 하고, 데이터셋의 인코딩 결과를 $D\in\{D_1,D_2,\cdots,D_n\}$라고 한다. 이때 데이터셋에 있는 각각의 이미지는 $D_n=\{d_1,d_2,\cdots,d_{512}\}$라고 표현한다. 영상 검색은 D 중에 G와 유사한 이미지를 고르며, 유사도는 L2 Norm을 통해 계산하고, D 중 G와 가장 유사한 N개의 이미지를 찾는 과정을 식 (3)으로 표현한다.

$$\begin{aligned} \| G - D_n \|_2 &= \sqrt{\sum_{k=1}^{512} (g_k - d_{n,k})} \\ D_c &= \arg\min_{D_c \in D, |D_c| = N} \| G - D_n \|_2 \end{aligned} \tag{3}$$

D_c를 구한 뒤엔 지역 인코딩을 통해 G와 유사도를 다시 비교한다. 지역 인코딩은 슬라이딩 윈도우[13]로 이미지의 지역적인 부분을 인코딩한다. 이렇게 전역적으로 인코딩한 이미지를 다시 지역적으로 인코딩하는 이유는 합쳐진 각 이미지에서 지역적 특징을 살리기 위해서다. 지역 인코딩은 이미지를 p 등분으로 나눈 뒤 각 지역을 식 (2)를 통해 인코딩한다. 그리고 지역 인코딩으로 나온 결과를 병합하여 하나의 벡터로 만든다.

이때 입력 이미지와 후보군 D_c에서 추출한 지역 벡터는 식 (4)를 통해 표현된다. $L\in\{L_1,L_2,\cdots,L_p\}$로 표현하고, $L_p=\{l_1,l_2,\cdots,l_{512}\}$로 구성되어 있다. 그리고 후보군 이미지의 지역 인코딩도 $D_L\in\{D_{L,1},D_{L,2},\cdots,D_{L,N}\}$로 표현하고, $D_{L,N}$은 지역 VLAD 벡터 집합을 의미한다. 그리고 $D_{L,N}=\{d_{l,1},d_{l,2},\cdots,d_{l,p}\}$로 구성되어 있다. $d_{l,p}$는 지역 VLAD 벡터를 의미하고 $d_{l,p}=\{d'_1,d'_2,d'_{512}\}$로 표현한다.

$$\begin{aligned} L &= \{VLAD(G)\}_{i=1}^{p} \\ D_L &= \{VLAD(D_i)\}_{i=1}^{p} \end{aligned} \tag{4}$$

식 (5)는 지역 인코딩을 통해 후보군 D_c 중 입력 이미지와 가까운 이미지를 다시 구하는 과정을 표현한다. 지역 인코딩을 통해 다시 계산되는 D'_c은 유사도에 따라 순서대로 정렬된다. 이 과정을 통해 다방향 이미지의 영상 측위를 수행하며, 이 과정을 알고리즘 1로 정리한다. 알고리즘 1은 제안하는 방법의 의사코드를 정리한 것이다.

$$D'_c = \{D_{L,i} | i \in 1, 2, \cdots, k\} \tag{5}$$

$$\text{where } \| L - D_{L,1} \| \leq \| L - D_{L,2} \| \leq \cdots \leq \| L - D_{L,k} \|$$

표 1. 영상 검색 평가를 위한 데이터셋 요약: Oxford Robotcar [14]는 다방향 영상 검색을 평가하기 적합한 데이터셋.

Table 1. Dataset summary for image retrieval evaluation: Oxford Robotcar [14] is a dataset suitable for evaluating multi-directional image retrieval.

Dataset	Camera	Method	Season	Day and Night	Multi-directional	Environment
Aachen Day-Night [20]	Phone	Handheld	X	O	X	Outdoor
Baidu Mall [22]	DSLR, Phone	Handheld	X	X	X	Indoor
InLoc [21]	Phone, RGB-D	Handheld	X	X	X	Indoor
Gangnam Station B2 [23]	Industrial Camera, Phone	Robot	X	X	O	Indoor
Oxford Robotcar [14]	Industrial camera	Vehicle	O	O	O	Outdoor

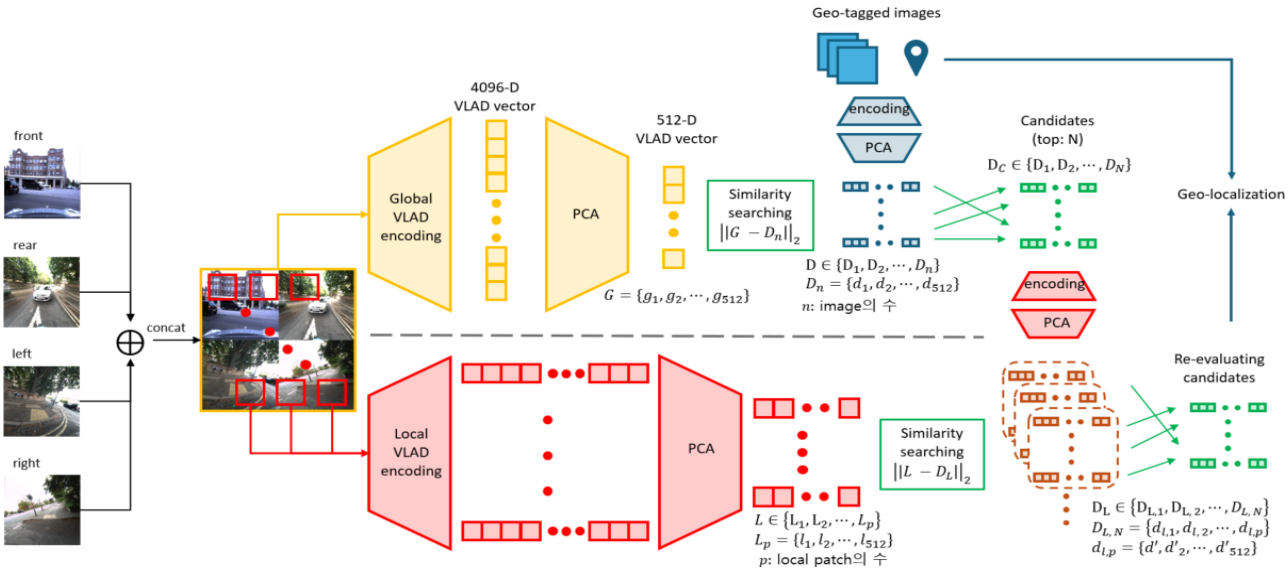

그림 2. 시스템 개요도: 크게 이미지 병합, 전역 인코딩, 지역 인코딩, 유사도 평가로 구성.

Fig. 2. System overview diagram. It consists of image concatenating, global encoding, local encoding, and similarity evaluation.

IV. 실험 및 결과

제안하는 방법은 Oxford Robotcar [14]에서 서로 다른 계절의 데이터를 선택하여 실험한다. 평가 척도는 Recall rate와 오차에 따른 이상치의 비율, 위치 오차와 회전 오차를 사용한다. Recall rate로 평가하는 이유는 전체 데이터셋에서 정답 비율이 희소해서 Precision 등의 지표를 사용할 경우 모델을 과소평가할 수 있기 때문이다. 모든 센서는 30(ms) 이내에 측정된 경우를 사용하고, ground truth는 동시에 기록된 RTK를 사용한다. 표 2는 실험에 사용한 데이터의 수를 동기화 전과 후로 비교한 결과이다.

평가는 위치 오차(translation error)와 회전 오차(rotation error)를 기준으로 평가한다. 재현율(Recall)은 n(m or °)으로 표현한다. n(m or °)은 위치 혹은 회전만 고려했을 때 오차가 n(m) 이내에 속하는 비율을 의미한다. n(m), n(°)의 경우 위치 오차 n(m), 회전 오차 n(°) 이내를 동시에 만족하는 경우를 의미한다. 그리고 위치 오차의 경우 위도와 경도를 (m) 단위로 변환하기 위해, [4]에서 사용한 방법을 사용하여 오차를 (m)로 계산한다. 비교군은 NetVLAD [11]와 PatchNetVLAD [12]를 사용한다. 표 3은 재현율을 기준으로 실험한 결과이며, 단위는 (%)다.

평가 결과 제안하는 방법이 모든 구간에서 기존 방법보다 우수한 성능을 보인다. 이런 결과가 나온 이유는 기존 방법

표 2. 데이터 동기화 결과, 4대의 카메라와 RTK가 30 [ms] 이내에 속하는 경우만 사용.

Table 2. Data synchronization results, only cases where the four cameras and RTK fall within 30 [ms] are used.

	Unsynced	Synced
Front Image	30k	10k
Left/Right/Rear Image	20k	10k
RTK	20k	10k

알고리즘 1. 제안하는 방법의 의사코드.

Algorithm 1. Pseudo-code of proposed method.

```
# Step 1: Image Concatenation and Resizing
Input: Image set (front, left, right, rear)
Concatenated_Image = Concatenate_Images(Image set)
Resized_Image = Resize_Image(Concatenated_Image,
    Target_Size)
# Step 2: VLAD Encoding
VLAD_Vector = Encode_VLAD_Vector(Resized_Image)
# Step 3: Similarity Calculation with Dataset
Similar_Images_List = Empty_List()
For each Image in Dataset:
    Dataset_Image_VLAD =
      Encode_VLAD_Vector(Current_Image)
    Similarity_Score = Calculate_Similarity(VLAD_Vector,
      Dataset_Image_VLAD)
    Append (Current_Image, Similarity_Score) to
      Similar_Images_List
End For
# Step 4: Ranking Using Local VLAD Encoding
Ranked_Images_List = Empty_List()
For each Image in Top_N_Images:
    Local_VLAD_Vector =
      Encode_Local_VLAD(Current_Image)
    Local_Similarity_Score =
      Calculate_Local_Similarity(VLAD_Vector,
      Local_VLAD_Vector)
    Append (Current_Image, Local_Similarity_Score) to
    Ranked_Images_List
End For
# Sort Final Ranked Images
Ranked_Images_List =
    Sort_By_Similarity(Ranked_Images_List)
# Step 5: Return Results
Output = Ranked_Images_List
```

표 3. 재현율을 기준으로 한 평가 결과, (m or °)는 단독 평가, (m), (°)는 둘 다 만족하는 경우.

Table 3. Evaluation based on recall, (m or °) for individual assessment, (m), (°) when both are satisfied.

Recall (%)	NetVLAD [11]	Patch - NetVLAD[12]	**MCVL (Proposed)**
1m	10.42	15.67	**17.98**
2.5m	29.46	42.37	**52.02**
5m	50.99	63.00	**76.92**
7.5m	59.47	70.11	**88.99**
10m	62.30	72.07	**90.46**
1°	38.57	42.08	**46.87**
2.5°	57.26	64.92	**72.42**
5°	66.96	74.17	**86.35**
7.5°	71.56	79.42	**91.02**
10°	73.56	81.00	**93.19**
1m, 1°	5.59	8.22	**7.64**
2.5m 2.5°	23.67	34.22	**39.75**
5m, 5°	46.71	57.64	**71.39**
7.5m, 7.5°	56.90	68.12	**86.21**
10m, 10°	60.73	70.99	**88.96**

들은 역광이 발생할 때, 데이터셋에서 역광 이미지를 고르기 때문이다. 하지만 제안하는 방법은 다른 방향의 이미지를 같이 사용하여 역광에도 정상적인 이미지를 선택한다. 그림 1은 사용한 데이터셋에서 역광이 발생할 때 기존 방법과 제안 방법의 차이를 정성적으로 보여준다. 제안 방법의 실험 결과가 제안 연구들[11-12]과 큰 차이가 나는 이유는 역광이 자주 발생한 데이터를 입력으로 선정했기 때문이고, 그림 1에서 확인할 수 있다.

표 4는 실험 결과 중 위치 오차와 회전 오차를 최소, 최대, 중간, 평균값으로 정리한 것이다. 위치 오차는 (m) 단위이며, 회전 오차는 (°)이다. 그리고 위치 오차의 최댓값이 중간값과 큰 차이가 나면서 결과의 경향을 파악하기 힘들다. 그래서 각 방법의 원활한 비교를 위해 $+1\sigma$의 값을 사용하며, $+1\sigma$는 전체 오류 중 84.2%에 해당한다. 표 4에서 회전 오차의 최솟값에서 *이탤릭체*로 표현된 부분은 소수점 두 자리로는 평가할 수 없어, 네 자리로 표현한 것이다. 실험 결과 제안 방법이 대부분 조건에서 좋은 성능을 보이며, 이는 다방향 영상 검색이 기존 방법보다 계절, 역광 등의 외부 요인에 강하다는 것을 보여준다.

표 4에서 위치 오차 최댓값의 $+1\sigma$를 보면, 제안하는 방법과 차이가 크게 난다. 이런 결과는 제안하는 방법이 기존 방법보다 이상 치(outlier)의 비율이 적다는 것을 의미한다. 표 5는 이상 치의 비율을 정리한 것으로 각각 위치 오차와 회전 오차의 비율을 보여준다. 5의 결과를 통해 오차의 빈도를 확인할 수 있다. 그 결과 영상 검색은 매우 적은 수의 큰 오차가 전체 시스템의 성능을 하락시키는 특징을 갖고 있다. 그리고 제안하는 방법이 기존 방법보다 우수한 이유는 기존 방법에선 추정할 수 없는 이미지에서도 위치를 추정하기 때문이다.

그림 3은 표 3을 기준으로 재현율을 도식화한 것이다. 제안하는 방법의 우수성과 기존 방법의 경향을 파악할 수 있다. 왼쪽 그래프는 위치 오차, 가운데는 회전 오차, 오른쪽은 둘 다를 동시에 만족하는 경우를 의미한다. 그림 3의 결과 위치 오차와 회전 오차가 서로 비슷한 경향을 보인다. 이는 비슷한 시야를 검색하기 위해서는 위치와 회전을 동시에 만족해야 하기 때문이다.

표 4. 위치 오차와 회전 오차의 최소, 최대, 중간, 평균값, 이탤릭체는 평가를 위해, 네 자리 소수점 사용.

Table 4. Minimum, maximum, median, and mean values of position and heading errors, four decimal places are used for evaluation.

	Translation Error (m)				Rotation Error (°)			
	Min	Max($+1\sigma$)	Median	Mean	Min	Max($+1\sigma$)	Median	Mean
NetVLAD[11]	0.03	356.86	124.53	150.53	6.24	40.52	12.17	18.36
PatchNetVLAD[12]	0.02	88.31	4.30	110.40	***0.0002***	17.38	0.52	13.45
Proposed	**0.01**	**5.67**	**2.89**	**28.20**	*0.0004*	**4.69**	**0.18**	**4.69**

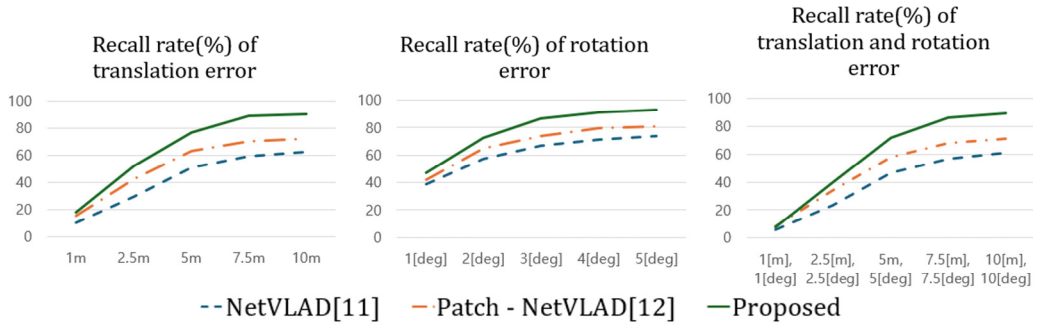

그림 3. 표 3의 기준에 따른 재현율 그래프: 왼쪽은 위치 오차만 평가한 결과, 가운데는 회전 오차만 평가한 결과, 오른쪽은 위치와 회전을 동시에 평가한 결과.

Fig. 3. Recall rate graph based on the criteria in Table 3. The left side shows the evaluation results for position error only, the middle shows the evaluation results for heading error only, and the right side shows the evaluation results for both position and heading errors combined.

표 5. 이상 치(outlier) 비율.

Table 5. Outlier ratio.

ratio (%)	NetVLAD [11]	Patch - NetVLAD[12]	**MCVL (Proposed)**
10m +	41.31	27.06	**9.17**
50m +	24.99	16.86	**2.67**
100m +	20.76	14.61	**2.05**
500m +	11.99	8.17	**1.32**
1000m +	2.22	1.53	**0.34**
5 °	37.18	28.33	**14.43**
10 °	27.51	20.00	**7.03**
30 °	17.44	12.50	**4.29**
50 °	13.31	9.29	**3.59**
100 °	6.16	4.63	**0.95**

그림 4. 비슷한 장면에서 실패 사례.

Fig 4. Failure cases in similar scenes.

V. 결론 및 한계

본 연구는 계절과 광원에도 변함 없는 영상 측위를 위해 다방향 영상 검색과 측위 방법을 제안한다. 제안하는 방법은 다양한 방향에서 취득한 이미지를 병합한 뒤, 전역 및 지역 인코딩한다. 이 결과로 나온 벡터와 데이터셋 이미지 사이의 유사도를 기준으로 영상 검색을 수행한다. 다양한 방향에 이미지를 동시에 사용하는 이유는 역광에도 강인한 특징을 갖기 위해서이다. 지역적으로 다시 인코딩하는 이유는 합쳐진 이미지의 특징을 사용하기 위해서다.

실험은 Oxford Robotcar [14]에서 역광이 심한 날을 선택한다. 실험 결과 제안하는 방법은 기존 방법보다 우수한 성능을 보인다. 하지만 비슷한 장면이 반복되는 장소에서 성능이 하락하는 한계도 보인다. 그리고 실험 데이터를 통해 위치, 회전 오차의 빈도와 경향을 통해 영상 검색이 가진 특징을 확인한다. 영상 검색은 적은 수의 이상 치에 의해 전체 성능이 하락해 보이는 경향이 있다. 이런 특징을 통해 영상 검색은 드물게 발생하는 장애를 개선하는 것만으로, 상당한 성능 향상 효과가 있는 것을 알 수 있다.

하지만 제안하는 방법은 회전 오차의 최솟값에서 기존 방법보다 부족한 성능을 보이기도 한다. 그리고 비슷한 장면들이 반복되는 경우 정확도가 떨어지는 경향을 보이기도 한다. 그림 4는 제안하는 방법이 500(m) 이상 위치 오차를 발생한 경우를 정성적으로 보여준다. 이를 보완하기 위해서는 비슷한 장면에서도 이미지의 지리적 특징을 구분할 수 있는 표현 방법이 필요하다. 이를 위해 새로운 학습 데이터셋과 딥러닝 모델 연구가 필요함을 알 수 있다.

REFERENCES

[1] T. Li, H. Zhang, Z. Gao, Q. Chen, and X. Niu, "High-accuracy positioning in urban environments using single-frequency multi-GNSS RTK/MEMS-IMU integration," *Remote Sensing*, vol. 10, no. 2, p. 205, 2018. doi: https://doi.org/10.3390/rs10020205

[2] M.-G. Petovello, M.-E. Cannon, G. Lachapelle, J. Wang, C. K. H. Wilson, O. S. Salychev, and V.-V. Voronov, "Development and testing of a real-time GPS/INS reference system for autonomous automobile navigation," *Proceedings of the 14th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2001)*, pp. 2634-2641. Salt Lake City, UT, Sept. 2001. doi:https://www.ion.org/publications/abstract.cfm?articleID=1941

[3] S.-J. Park, S.-T. Kim, Y.-M. Kim, J.-H. Lee, J.-W. Song, and E.-J. Kim, "GNSS-DR/INS integrated navigation algorithm using GNSS Speed information fusion for overcoming GNSS denial environment," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 29, no. 1, pp. 72-79, 2023. doi: https://doi.org/10.5302/j.icros.2023.22.0180

[4] G. Mun, and H. Kim, "Kalman Filter based Image Re-selection: Handling Collinear and High Co-visibility Situations in Image Retrieval based Visual Localization," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 30, no. 9, pp. 954-959, 2024. doi: https://doi.org/10.5302/j.icros.2024.24.0119

[5] M. Humenberger, Y. Cabon, N. Pion, P. Weinzaepfel, D. Lee, N. Guérin, T. Satter, and G. Csurka, "Investigating the role of image retrieval for visual localization: An exhaustive benchmark," *International Journal of Computer Vision*, vol. 130, no. 7, pp. 1811-1836, 2022. doi: https://doi.org/10.1007/s11263-022-01615-7

[6] M. Feng, s. Hu, M-H. Ang Jr, and G-H Lee, "2d3d-matchnet: Learning to match keypoints across 2d image and 3d point cloud," *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 4790-4796, 2019. doi: https://doi.org/10.1109/icra.2019.8794415

[7] J. Li and G-H Lee, "DeepI2P: Image-to-point cloud registration via deep classification," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15960-15969, 2021. doi: https://doi.org/10.1109/cvpr46437.2021.01570

[8] A. Fisher, R. Cannizzaro, M. Cocharane, C. Nagahawatte, and J. L. Palmer, "ColMap: A memory-efficient occupancy grid mapping framework," *Robotics and Autonomous Systems*, vol, 142, 103755, 2021. doi: https://doi.org/10.1016/j.robot.2021.103755

[9] J. Sivic and A. Zisserman, "Video Google: a text retrieval approach to object matching in videos," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1470-1477, 2003.

doi: https://doi.org/10.1109/iccv.2003.1238663

[10] R. Arandjelovic and A. Zisserman, "All about VLAD," *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 1578-1585, 2013.
doi: https://doi.org/10.1109/cvpr.2013.207

[11] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN architecture for weakly supervised place recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5297-5307, 2016.
doi: https://doi.org/10.1109/cvpr.2016.572

[12] S. Hausler, S. Garg, M. Xu, M. Milford, and T. Fischer, "Patch-netvlad: Multi-scale fusion of locally-global descriptors for place recognition," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14141-14152, 2021.
doi: https://doi.org/10.1109/cvpr46437.2021.01392

[13] J. Xie, K. Hu, G. Li, and Y. Guo, "CNN-based driving maneuver classification using multi-sliding window fusion," *Expert Systems with Application*, vol. 169, 114442, 2021.
doi: https://doi.org/10.1016/j.eswa.2020.114442

[14] W.Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The oxford robotcar dataset," *The International Journal of Robotics Reasearch,* vol. 36, no. 1, pp. 3-15, 2017.
doi: https://doi.org/10.1177/0278364916679498

[15] D. G. Lowe, "Oject recognition from local scale-invariant features," *Proceedings of the IEEE International Conference on Computer Vision,* vol. 2, pp. 1150-1157, 1999.
doi: https://doi.org/10.1109/iccv.1999.790410

[16] H. Bay, A. Ess, T. Tuytelaara, and L. V. Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding,* vol. 110, no. 3, pp. 346-359, 2008.
doi: https://doi.org/10.1016/j.cviu.2007.09.014

[17] J. sanchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: Theory and practice," *International Journal of Computer Vision*, vol. 105, pp. 222-245, 2013.
https://doi.org/10.1007/s11263-013-0636-x

[18] E.J. Ong, S. S. Husain, M. Bober-Irizar, and M. Bober, ""Deep architectures and ensembles for semantic video classification," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 29, no. 12, pp. 3568-3582, 2018.
doi: https://doi.org/10.1109/tcsvt.2018.2881842

[19] F. Perronnin, and D. Larlus, "Fisher vectors meet neural networks: A hybrid classification architecture," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3743-3752, 2015.
doi: https://doi.org/10.1109/cvpr.2015.7298998

[20] A. Benbihi, C. Pradalier, and O. Chum, "Object-guided day-night visual localization in urban scenes," *IEEE International Conference on Pattern Recognition*, pp. 3786-3793, 2022.
doi: https://doi.org/10.1109/icpr56361.2022.9955638

[21] H. Taira, M. Okutomi, T. Sattler, M. Cimpoi, M. Pollefeys, J. Sivic, T. Pajdla, and A. Torii, "InLoc: Indoor visual localization with dense matching and view synthesis," *IEEE/CVF International Conference on Pattern Recognition*, pp. 7199-7209, 2018.
doi: https://doi.org/10.1109/cvpr.2018.00752

[22] X. Sun, Y. Xie, P. Luo, and L. Wang, "A dataset for benchmarking image-based localization," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7436-7444, 2017.
doi: https://doi.org/10.1109/cvpr.2017.598

[23] D. Lee, S. Ryu, S. Yeon, Y. Lee, D. Kim, C. Han, Y. Cabon, P. Weinzaepfel, N. Guerin, G. Csurka, and M. Humenberger, "Large-sacle localization datasets in crowded indoor spaces," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3327-3236, 2021.
doi: https://doi.org/10.1109/cvpr46437.2021.00324

문 기 영

2022년 동양미래대학교 자동화공학과(공학사). 2023년~현재 인하대학교 전기컴퓨터공학과 석사과정. 관심분야는 SLAM, Visual Localization, 영상 처리.

김 학 일

1983년 서울대학교 제어계측공학과(공학사). 1985년 Purdue대학교 전기컴퓨터공학과(공학석사). 1990년 Purdue대학교 전기컴퓨터공학과(공학박사). 1990년~현재 인하대학교 정보통신공학과 및 스마트모빌리티공학과 교수. 관심분야는 자율주행, 컴퓨터비전, 패턴인식, 바이오영상처리.