

구조물 검출 네트워크 및 특징점 필터링을 이용한 원격 탐사 영상 정합

성준영 이우주 *오승준

광운대학교

jysung13mmlab@kw.ac.kr krosea@kw.ac.kr *sjoh@kw.ac.kr

Remote Sensing Image Registration using Structure Extraction and Keypoint Filtering

Jun-Young Sung Woo-Ju Lee *Seoung-Jun Oh

Kwangwon University

요 약

본 논문에서는 원격 탐사 영상 정합에서 정확도는 유지하면서 특징점 매칭 (Matching) 복잡도를 줄이기 위해 입력 영상을 전처리하는 구조물 검출 네트워크를 이용한 원격 탐사 영상 정합 방법을 제안한다. 영상 정합의 기존 방법은 입력 영상에서 특징점을 추출하고 설명자 (Descriptor)를 생성한다. 본 논문에서 제안하는 방법은 입력 영상에서 특징점 매칭에 영향을 미치는 구조물만 추출하여 새로운 영상을 만들어 특징점을 추출한다. 추출된 특징점은 필터링 (Filtering)을 거쳐 원본 영상에 매핑 (Mapping)되어 설명자를 생성하여 특징점 매칭 속도를 향상시킨다. 또한 구조물 검출 네트워크에서 학습 영상과 시험 영상의 특성의 차이로 생기는 성능 저하 문제를 개선하기 위해 히스토그램 매핑 기법을 이용한다. 아리랑 3 호가 획득한 원격 탐사 영상에 대한 실험을 통해 제안하는 방법은 정확도를 유지하면서 계산 시간을 SURF 보다 87.5%, SIFT 보다 92.6% 감소시킬 수 있다.

1. 서론

영상 정합은 동일한 장면의 감지 영상과 기준 영상 간의 기하학적 관계를 추정하여 기하학적으로 일치시키기 위해 널리 사용되는 방법으로 이 영상들은 서로 다른 센서나 다른 관점에 의해 촬영될 수 있다[1]. 영상 정합은 이미지 융합, 변화 탐지, 환경 모니터링 등 원격 탐사 영상 처리 분야에서 성능에 직접적인 영향을 미치는 중요한 과정이다.

원격 탐사 영상 정합을 위한 방법은 두 가지 방법으로 나뉜다. 첫 번째 방법은 영상의 유사성을 이용하여 최적의 기하학적 변환 파라미터를 구하는 방법인 지역 기반 영상 정합 방법이다[2], [3]. 두 번째 방법은 점, 선, 영역 등의 유사성을 이용하여 기하학적 변환 파라미터를 구하는 방법인 피처 (Feature) 기반 영상 정합 방법으로 SIFT[4], SURF[5] 등이 있다. 지역 기반 방법은 쉽게 구현할 수 있다는 장점이 있지만 밝기 변화나 조명 변화, 잡음 등과 같은 외부 변화에 민감하다는

단점이 있다. 피처 기반 방법은 상대적으로 외부 변화에 둔감한 특성을 보인다.

표 1. 영상 크기 별 SIFT 전체 속도 및 특징점 매칭 계산 속도 비교. 비율은 전체시간에서 매칭 시간의 비율을 의미한다.

크기	총 시간(초)	매칭 시간(초)	비율 (%)
500 × 500	0.67	0.51	76.11
1000 × 1000	21.73	21.06	96.89
2000 × 2000	315.18	312.28	99.07
4000 × 4000	3,473.50	3,462.34	99.67

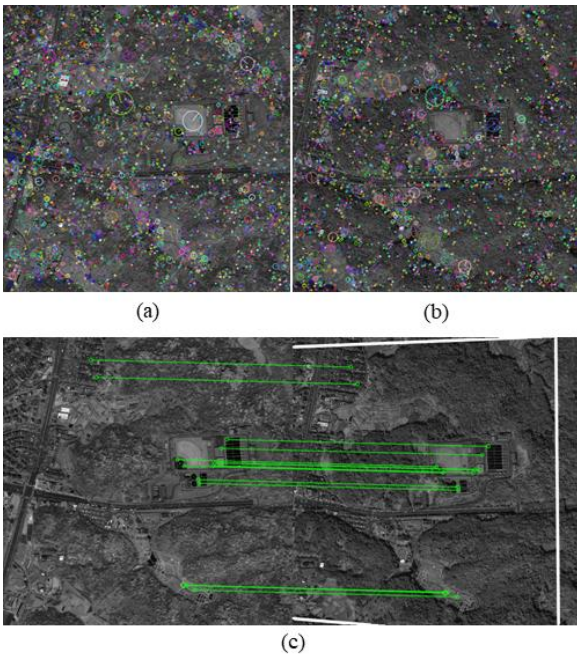


그림 1. SIFT 특징점 매칭 결과. (a) 입력 영상에서 추출된 특징점. (b) 기준 영상에서 추출된 특징점. (c) 매칭 단계에 사용된 특징점

표 1과 같이 영상 정합 방법에서 전체 시간 중 특징점 매칭 단계에서 가장 많은 시간이 걸린다. 또한 원격 탐사 영상 정합 방법에서 영상의 크기가 커질수록 특징점의 양이 기하급수적으로 많아진다. 그에 따라 특징점 매칭 단계에서의 계산 복잡도가 증가한다. 그림 1과 같이 원격 탐사 영상에서 획득되는 특징점 중 매칭에 사용되는 특징점은 대부분 영상 내 주요 구조물에서 추출된다. 따라서 구조물에서 추출되는 특징점만 남기고 배경에서 추출되는 특징점의 숫자를 줄인다면 영상 정합의 정확도는

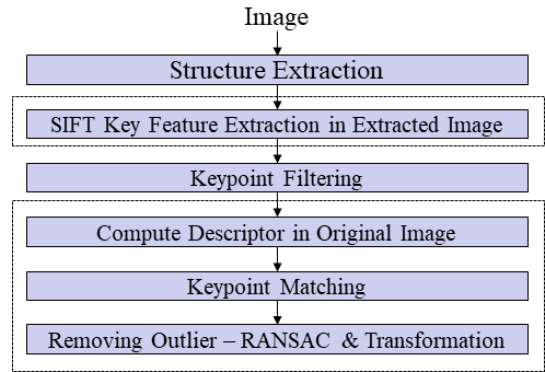


그림 2. 본 논문에서 제안하는 EF-SIFT 의 순서도. 점선으로 나타낸 부분은 기존 SIFT 방법이다.

유지하면서 계산 복잡도의 감소로 인한 속도 향상을 기대할 수 있다.

본 논문에서는 원격 탐사 영상 정합의 특징점 매칭 단계에서 매칭에 중요한 역할을 하는 주요 구조물인 건물과 도로를 제외한 배경 부분에서 추출되는 특징점을 제거하여 정확도는 유지하면서 속도를 개선하는 방법인 EF-SIFT 를 제안한다.

본 논문의 구성은 다음과 같다. 2 장에서는 EF-SIFT 의 구조를 설명하고, 3 장에서는 실험을 통해 EF-SIFT 와 SIFT 및 SURF 를 이용한 영상 정합 알고리즘과 성능 비교를 수행한다. 마지막으로 4 장에서는 본 연구의 결론을 맺고자 한다.

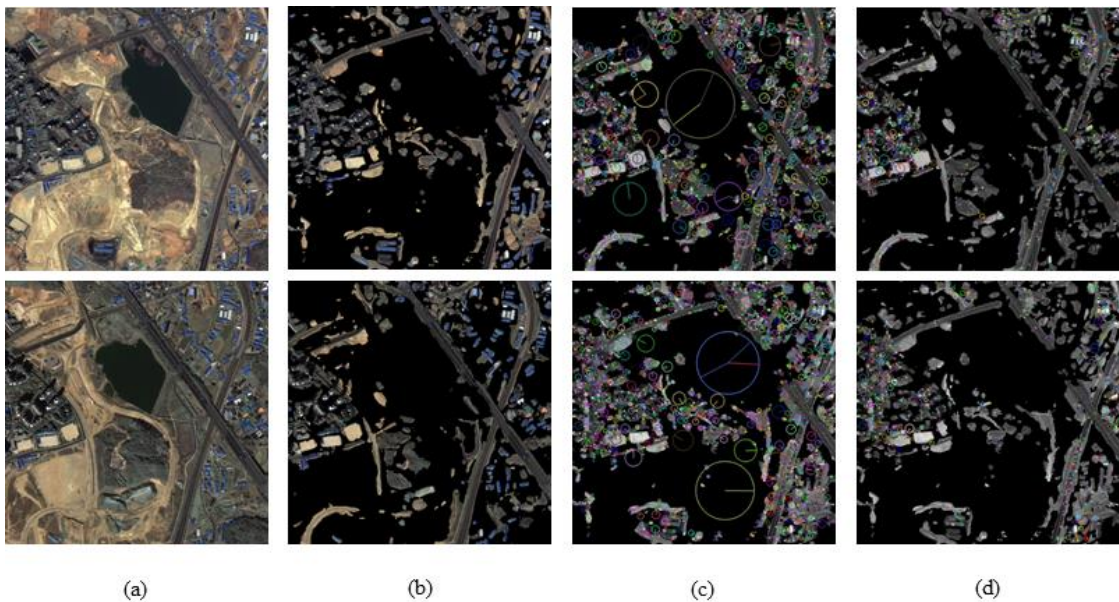


그림 3. 제안하는 특징점 필터링 과정. (a) 원본 영상, (b) 배경이 제거된 영상, (c) 배경이 제거된 영상에서 뽑힌 특징점. (d) 특징점 필터링이 적용된 영상

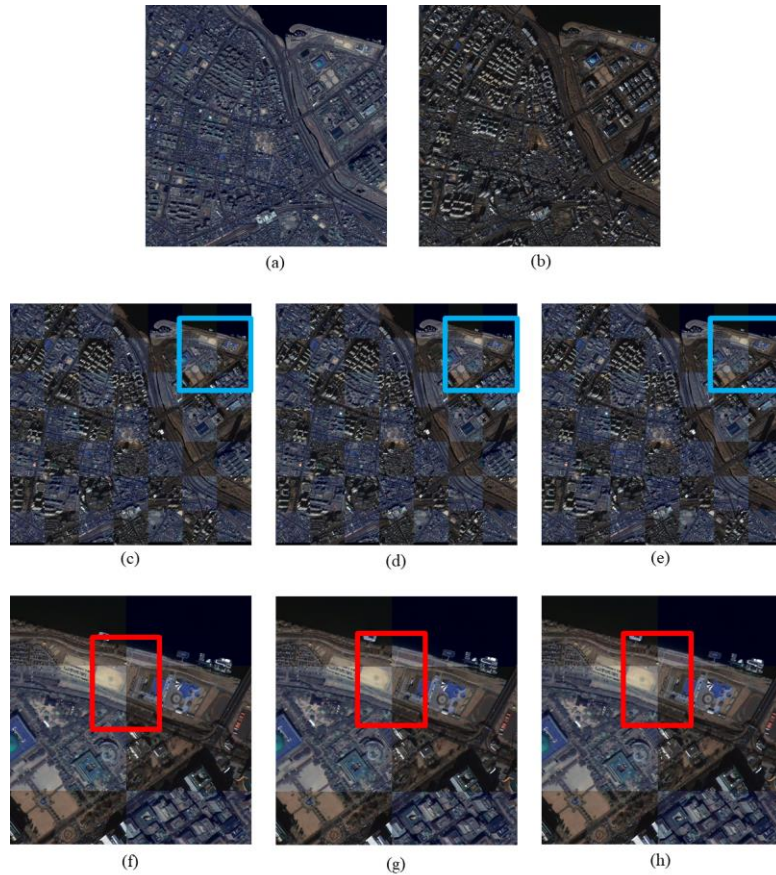


그림 4. 아리랑 3 호에서 취득한 원격 탐사 영상에 대한 SIFT 와 EF-SIFT 결과 영상. (a) 감지 영상, (b) 기준 영상, (c) SIFT 정합 결과 영상과 기준 영상을 모자이크 한 결과, (d) SURF 정합 결과 영상과 기준 영상을 모자이크 한 결과, (e) EF-SIFT 정합 결과 영상과 기준 영상을 모자이크 한 결과, (f) SIFT 정합 결과 영상의 파란 영역을 확대한 영상, (g) SURF 정합 결과 영상의 파란 영역을 확대한 영상, (h) EF-SIFT 정합 결과 영상의 파란 영역을 확대한 영상

2. 제안 방법

본 논문에서 제안하는 EF-SIFT 의 전체 구조는 그림 2와 같이 구조물 추출 네트워크와 SIFT 가 결합한 구조이다. 구조물 추출 네트워크는 다양한 영상 분할 모델 가운데 U-Net[6]을 사용한다. 추출된 구조물의 정보는 이진 맵 (Map)으로 출력되어 원본 영상에 매핑된다. 딥 러닝 기반 구조물 추출 네트워크는 시험 데이터 세트인 위성 영상과 학습 데이터 세트인 항공 영상의 특성의 차이 때문에 같은 특성의 데이터 세트를 사용하였을 때보다 성능이 떨어진다. 따라서 그 차이를 보완하기 위하여 항공 영상 데이터의 히스토그램 매핑을 이용하여 데이터 세트를 늘려 학습한다.

특징점 추출 과정에서 구조물을 추출하며 생긴 경계선에 특징점이 생성된다. 경계선에서 생기는 특징점은 매칭에는 영향을 미치지 않지만 계산 복잡도를 증가시키는 문제가 있다. 따라서

경계선에 생긴 특징점을 필터링하기 위해 배경이 제거된 영상에서 특징점을 취득하여 배경이 제거된 영상과 원본 영상 두 영상 모두에 설명자를 생성한다. 경계선에서 추출된 특징점에서 설명자가 생성되면 설명자의 영상 패치 (Patch)는 배경으로 제거된 영역이 존재한다. 따라서 원본 영상의 같은 위치의 패치와 비교했을 때 유사도가 떨어진다. 구조물에서 추출된 특징점은 설명자와 패치가 원본 영상과 유사하다. 설명자의 유사도 판단 척도는 L2 Norm 을 사용한다. 설명자의 유사도의 임계값은 실험적으로 2.5 로 정하였다. 필터링에 대한 결과는 그림 3과 같다. 그림 3의 (a)는 원본 영상, (b)는 배경이 제거된 영상이다. (c)는 (b)에서 뽑힌 특징점이고 (d)는 (c)에 필터링이 적용된 뒤의 특징점을 보여준다. 필터링 방법은 (c)의 특징점을 (d)와 같이 67.7% 줄여준다. 이후 영상 정합에 대한 변환 매트릭스는 필터링된 설명자에 의해 매칭된 특징점을 이용하여 RANSAC 을 통해 계산된다.

3. 실험 결과

본 논문의 구현 및 실험은 PyCharm 인터페이스에서 이루어졌고, GPU는 Nvidia GeForce 1080ti를 사용하였다. 구조물 추출 네트워크의 학습 데이터 세트는 공개된 영상 분할 데이터 세트인 Massachusetts Roads 데이터 세트와 Massachusetts Buildings 데이터 세트를 사용하였다. EF-SIFT의 성능을 측정하기 위해, 아리랑 3 호에서 획득된 원격 탐사 영상을 이용했다. 시험 결과 영상들 중 그 예를 그림 4에서 보인다.

그림 4의 (a)와 (b)는 2014년 3월과 2015년 10월에 아리랑 3 호에서 획득된 여의도 지역에 대해 4000×4000 크기로 자른 한 쌍의 영상을 보여준다. 그림 4의 (c)-(e)는 이 한 쌍의 영상을 SIFT, SURF, EF-SIFT 방법을 통해 하나의 체크보드 영상으로 나타낸다. 그림 4의 (f)-(h)는 각각 (c)-(e)의 파란 영역이다. 빨간 영역을 비교했을 때, SURF의 결과 영상에서 끊어져 있는 도로 부분이 EF-SIFT의 결과 영상에서 SIFT의 결과 영상과 유사하게 매끄럽게 이어져 있는 것을 확인할 수 있다.

본 논문의 실험에서는 [7]에 소개된 측정지표를 사용하여 영상 정합 성능을 평가한다. EF-SIFT의 성능을 SIFT 및 SURF와 비교한다. 표 2는 아리랑 3 호에서 획득된 원격 탐사 영상 데이터 세트로 7개의 지표에 대하여 성능 평가를 진행한 결과이다. EF-SIFT의 시험 영상에 대한 실험 결과는 정확도 측면에서 Phi 값을 기준으로 SURF의 결과 값보다 0.51 낮고, SIFT의 결과 값보다 0.25 높다. 또한, 속도 측면에서 SURF보다 87.5%, SIFT보다 92.6%의 시간 감소를 보이는 것을 확인할 수 있었다.

4. 결론

본 논문에서는 원격 탐사 영상 정합에서 정확도는 유지하면서 특징점 매칭 복잡도를 줄이기 위해 입력 영상을 전처리하는 구조물 검출 네트워크를 이용한 원격 탐사 영상 정합 방법인 EF-SIFT를 제안하였다. EF-SIFT는 입력 영상에서 특징점 매칭에 영향을 미치는 구조물만 추출하여 새로운 영상을 만들어 특징점을 추출했다. 추출된 특징점은 필터링을 거쳐 원본 영상에 매핑되어 설명자를 생성하여 특징점 매칭 속도를 향상시켰다. 또한 구조물 검출 네트워크에서 학습 영상과 시험 영상의 특성의 차이로 생기는 성능 저하 문제를 개선하기 위해 히스토그램 매핑 기법을 이용했다. 아리랑 3 호가 획득한 원격 탐사 영상에 대한 실험을 통해 EF-SIFT가 기존 방법보다 정확도를 유지하면서 계산 시간을 SURF보다 87.5%, SIFT보다 92.6% 감소시켰다.

표 2. 제안한 방법과 기존 방법들의 성능 비교

Metric	SIFT	SURF	EF_SIFT
RMS	7.68	10.62	8.25
RMSLoo	7.69	10.77	9.02
Pquad	1.00	1.00	1.00
BPP	0.94	0.96	0.76
Skew	0.19	0.18	0.19
Scat	0.89	0.87	0.90
Phi	2.38	3.14	2.63
Num of Keypoints	392,493	346,444	86,904
Inference Time (s)	1,546.17	915.93	114.82

감사의 글

“본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음” (IITP-2020-2016-0-00288)

참고문헌

- [1] Zitova, B., Flusser, J. “Image registration methods: a survey”. *Image and Vision Computing*. Vol.21, No.11, pp 977-1000. June 2003.
- [2] Suri, s., Reinartz, p. “Mutual-information-based registration of terrasars-x and ikonos imagery in urban areas” *IEEE Transactions on Geoscience and Remote Sensing*. Volume: 48, No.2, pp939-949. November 2009
- [3] Kern, J.P., Pattichis, M.S. “Robust multispectral image registration using mutual-information models”. *IEEE Transactions on Geoscience and Remote Sensing*. Vol.45, No.5, pp 1494-1505. April 2007.
- [4] Lowe, D.G. “Distinctive image features from scale-invariant keypoints”. *International Journal of Computer Vision*. Vol.60, No.2, pp 91-110. January 2004.
- [5] Bay, H., Ess, A., Tuytelaars, T., Gool, L.V. “Speeded-up robust features (surf)”. *Computer Vision and Image Understanding*. Vol.110, No3, pp 346-359. June 2008.
- [6] O. Ronneberger, P. Fischer, and T. Brox. “U-Net: Convolutional networks for biomedical image segmentation”. *Medical Image Computing and Computer-Assisted Intervention*. Vol.9351. pp 234-241. November 2015.

- [7] Goncalves, H., Goncalves, J.A., Corte-Real, L.
“Measures for an objective evaluation of the geometric
correction process quality”. *IEEE Geoscience and
Remote Sensing Letters*. Vol.6, No2, pp 292-296. April
2009