

특집논문 (Special Paper)

방송공학회논문지 제26권 제1호, 2021년 1월 (JBE Vol. 26, No. 1, January 2021)

<https://doi.org/10.5909/JBE.2021.26.1.26>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

원격 탐사 영상 정합을 위한 딥러닝 기반 특징점 필터링

성 준 영^{a)}, 이 우 주^{b)}, 오 승 준^{a)†}

Deep Learning-based Keypoint Filtering for Remote Sensing Image Registration

Jun-Young Sung^{a)}, Woo-Ju Lee^{b)}, and Seung-Jun Oh^{a)†}

요 약

본 논문에서는 원격 탐사 영상에 대한 특징 기반 영상 정합 (Image Registration) 방법의 고속화를 위한 딥러닝 기반 특징점 필터링 방법인 DLKF (Deep Learning Keypoint Filtering)를 제안한다. 기존의 특징 기반 영상 정합 방법의 복잡도는 특징 매칭 (Feature Matching) 단계에서 발생한다. 이 복잡도를 줄이기 위하여 본 논문에서는 특징 매칭이 영상의 구조물에서 검출된 특징점으로 매칭되는 것을 확인하여 특징점 검출기에서 검출된 특징점 중에서 구조물에서 검출된 특징점만 필터링하는 방법을 제안한다. DLKF는 영상 정합을 위하여 필수적인 특징점을 잃지 않으면서 그 수를 줄이기 위하여 구조물의 경계와 인접한 특징점을 보존하고, 서브 샘플링 (Subsampling)된 영상을 사용한다. 또한 영상 분할 (Image Segmentation) 방법을 위해 패치 단위로 잘라낸 영상을 다시 합칠 때 생기는 영상 패치 경계의 잡음을 제거하기 위하여 영상 패치를 중복하여 잘라낸다. DLKF의 성능을 검증하기 위하여 아리랑 3호 위성 원격 탐사 영상을 사용하여 기존 특징점 검출 방법과 속도와 정확도를 비교하였다. SIFT 기반 정합 방법을 기준으로 SURF 기반 정합 방법은 특징점의 수를 약 18% 감소시키고 속도를 약 2.6배 향상시켰지만 정확도가 3.42에서 5.43으로 저하되었다. 제안하는 방법인 DLKF를 사용하였을 때 특징점의 수를 약 82% 감소시키고 속도를 약 20.5배 향상시키면서 정확도는 4.51로 저하되었다.

Abstract

In this paper, DLKF (Deep Learning Keypoint Filtering), the deep learning-based keypoint filtering method for the rapidization of the image registration method for remote sensing images is proposed. The complexity of the conventional feature-based image registration method arises during the feature matching step. To reduce this complexity, this paper proposes to filter only the keypoints detected in the artificial structure among the keypoints detected in the keypoint detector by ensuring that the feature matching is matched with the keypoints detected in the artificial structure of the image. For reducing the number of keypoints points as preserving essential keypoints, we preserve keypoints adjacent to the boundaries of the artificial structure, and use reduced images, and crop image patches overlapping to eliminate noise from the patch boundary as a result of the image segmentation method. the proposed method improves the speed and accuracy of registration. To verify the performance of DLKF, the speed and accuracy of the conventional keypoints extraction method were compared using the remote sensing image of KOMPSAT-3 satellite. Based on the SIFT-based registration method, which is commonly used in households, the SURF-based registration method, which improved the speed of the SIFT method, improved the speed by 2.6 times while reducing the number of keypoints by about 18%, but the accuracy decreased from 3.42 to 5.43. Became. However, when the proposed method, DLKF, was used, the number of keypoints was reduced by about 82%, improving the speed by about 20.5 times, while reducing the accuracy to 4.51.

Keyword: Computer Vision, Deep Learning, Image Registration, Remote Sensing Image, Image Segmentation

Copyright © 2021 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

1. 서론

영상 정합은 같은 장면의 감지 영상 (Sensed Image)과 기준 영상 (Reference Image)사이의 기하학적 관계를 추정하고 기하학적으로 일치시키기 위해 널리 사용되는 방법이며, 이러한 영상들은 다른 센서 (Sensor)나 다른 관점에 의해 획득될 수 있다^[1]. 영상 정합은 영상 융합, 변화 감지, 환경 감시 등 원격 탐사 영상 처리 분야에서 성능에 직접적인 영향을 미치는 중요한 과정이다.

원격 탐사 영상 정합 방법은 영역 기반 영상 정합 방법과 특징 기반 영상 정합 방법으로 나뉜다. 영역 기반 영상 정합 방법은 영상의 유사성을 이용하여 기하학적으로 변형시켜 정합된 영상을 얻는다^{[2][3]}. 특징 기반 영상 정합 방법은 점, 선, 면적 등의 유사점을 이용하여 기하학적으로 변형시킨다. 영역 기반 정합 방법은 구현이 쉽다는 장점이 있지만 밝기 변화, 조명 변화, 잡음 등 외부 변화에 민감하다는 단점이 있다. 특징 기반 정합 방법은 외부 변화에 상대적으로 강한 특성을 보여서 본 논문에서는 특징 기반 정합 방법을 기준 방법 (Baseline Method)으로 사용한다.

특징 기반 정합 방법은 특징점 검출, 특징 벡터 (Feature Vector) 표현, 특징 매칭, 좌표변환 식 계산의 네 단계로 나눌 수 있다. 첫 번째 단계에서는 코너 (Corner), 블롭 (Blob), T-형 코너 (T-junction) 등 영상에서 눈에 띄는 위치의 특징점을 선택한다. 특징점 검출기의 가장 중요한 특성은 반복성 (Repeatability)이다. 이는 다른 시야 조건에서 같은 물리적 특징점을 찾기 위한 검출기의 신뢰성을 나타낸다. 두 번째 단계에서는 모든 특징점의 근방을 특징 벡터로 표현하며 이를 서술자 (Descriptor)라 한다. 서술자는 고유해야 하며

동시에 잡음, 변위 탐지, 기하학적 및 광도학적 변형에 대해 강인해야 한다. 세 번째 단계에서 서술자는 서로 다른 영상 사이에서 매칭된다. 매칭은 마할라노비스 (Mahalanobis) 또는 유클리디안 (Euclidean) 거리 같은 벡터 사이의 거리 공식에 기초한다. 서술자의 차원은 걸리는 시간에 직접적인 영향을 미치기 때문에 빠른 특징 매칭에는 저차원이 바람직하다. 그러나 일반적으로 저차원 특징 벡터는 고차원 특징 벡터보다 정확도가 떨어진다. 마지막 단계에서는 매칭된 서술자 쌍을 이용하여 좌표변환 식을 계산한다. 본 논문에서는 잘못 매칭된 제어점 (Control Point)을 식별하여 좌표변환식 계산에서 제외하는 RANSAC (Random Sample Consensus) 알고리즘 (Algorithm)을 사용한다. 그 뒤에 어파인 (Affine) 매개변수를 구하여 기하학적으로 변형시켜 정합된 영상을 얻는다^[4].

영상 정합은 원격 탐사 영상 등 대형 영상에 대해 계산 복잡도가 크다는 문제점을 가진다. 원격 탐사 영상 정합의 경우 특징 매칭에서 모든 특징점에 대해 일대일 매칭 작업이 진행되기 때문에 검출기에서 검출한 특징점의 수가 많을수록 특징 매칭의 복잡도가 커진다. 표 1은 아리랑 3호 위성 원격 탐사 영상들에 대한 SIFT 기반 정합 방법에 대한 계산 시간의 평균을 보여준다. Total Time은 특징 기반 정합 방법 네 단계에 대한 전체 실행 시간을, Match Time은 특징 매칭 단계의 실행 시간을 나타낸다. 특징 기반 방법은 영상의 크기가 증가함에 따라 실행 시간이 기하급수적으로 증가하고 특징 매칭 단계에서 가장 많은 시간이 걸리는 것을 확인할 수 있다.

표 1. SIFT 기반 정합 방법의 전체 계산 시간과 매칭 단계의 계산 시간
 Table 1. Calculation time of SIFT-based Registration method in overall and matching phase

Size	Total Time (s)	Match Time (s)	Time Rate (%)
1,000 × 1,000	21.73	21.06	96.89
2,000 × 2,000	315.18	312.28	99.07
4,000 × 4,000	3,473.50	3,462.34	99.67

기존 논문에서 다양한 영상 정합 고속화 방법이 제안되었다. SIFT (Scale Invariant Feature Transform)는 신뢰성 있는 영상 정합을 위한 안정적인 특징점을 제공하는 가장 널리 채택된 방법이다^[5]. 또한, [4]에서는 빠른 검출기와 서

a) 광운대학교 전자공학과(Department of Electronic Engineering, Kwangwoon University)

b) 광운대학교(Kwangwoon University)

‡ Corresponding Author : 오승준 (Seoung-jun Oh)

E-mail: sjoh@kw.ac.kr

Tel: +82-2-940-5102

ORCID: <https://orcid.org/0000-0002-5036-3761>

* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2020-2016-0-00288).

* This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2020-2016-0-00288) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation).

· Manuscript received November 23, 2020; Revised January 7, 2021; Accepted January 7, 2021.

술자 갖추어 SIFT보다 정확성을 유지하면서 속도는 개선된 SURF (Speed-Up Robust Features)가 제안되었다. [6]에서는 유사도 측정으로 상호 정보 (Mutual Information)를 이용한 영상 정합의 고속화 방법이 제안되었고 [7]에서는 영상 정합 방법의 고속화를 위하여 계층 병렬 SIFT (Layer Parallel SIFT)와 병렬 하드웨어 (Hardware) 설계가 제안되었다.

표 2는 아리랑 3호 시험 데이터 세트 (Test Data Set)를 위한 SIFT 및 SURF 기반 정합 방법의 실험 결과를 보여준다. Number of Keypoints는 두 영상에서 검출된 특징점의 합계이고 Matched Keypoints는 두 영상에서 매칭된 특징점 쌍이다. 표 2를 보면 SIFT 기반 정합 방법은 검출기에서 검출된 총 특징점의 0.05%만이 매칭된다. SURF 기반 정합 방법의 실험 결과는 SIFT 기반 정합 방법보다 속도가 항상 됨에도 불구하고 여전히 전체 특징점의 약 0.04%만이 매칭되고 정확도는 떨어지는 것을 보여준다. 이는 특징 매칭 단계에서 SURF 기반 정합 방법이 여전히 불필요한 특징점이 많다는 것을 보여준다.

표 2. SIFT 및 SURF 기반 정합 방법의 특징점 검출 및 특징 매칭 결과
Table 2. Keypoint detection and feature matching results of SIFT and SURF-based registration methods

Methods	Number of Keypoints	Matched Keypoints	Matching Rate (%)
SIFT	424,998	105	0.05
SURF	349,308	74	0.04

그림 1은 원격 탐사 영상인 감지 영상과 기준 영상의 SIFT 기반 정합 방법의 결과를 보여준다. 그림 1(a)와 (b)는 SIFT 검출기의 특징점 검출 결과이다. 그림 1(c)은 매칭된 특징점을 보여준다. 먼저 건물과 도로가 있는 영역을 구조물이라고 하고 나머지 영역을 배경이라고 했을 때, 전체 특징점 중 매칭되는 특징점이 구조물에서 검출되는 것을 보여준다. 따라서 구조물에서 검출되는 특징점만 보존하고 배경에서 검출되는 특징점을 제거하면 정합 방법의 정확성은 유지하면서 전체 특징점 개수의 감소로 인한 속도 향상을 기대할 수 있다. 특징점 필터링은 영상의 구조물을 제외한 나머지 영역인 배경을 딥러닝 기반 영상 분할 방법을 사용하여 제거한다. 그 후, 원본 영상과 배경이 제거된 영상을 비교하여 필터링을 수행한다. 영상 배경은 각 패치에 대해 개별적으로 제거되며 각 패치는 중첩되어 잘린다.

본 논문의 목표는 기존의 영상 정합 방법과 비교하였을 때 정확도 성능은 유지하면서 특징 매칭에 사용되는 특징점의 개수를 줄이는 특징점 필터링 방법을 개발하는 것이다. 아리랑 3호 데이터 세트에 대한 본 논문의 실험에서, DLKF는 전통적인 방법에 비해 정확도는 유지하면서 속도는 향상된다. 또한, 기존의 특징 기반 정합 방법에 적용될 수 있다.

이어지는 논문의 구조는 다음과 같다. 2장에서는 본 연구에 관련된 연구를 소개한다. 3장에서는 특징점 필터링을 포함한 특징 기반 영상 정합 구조를 제안하고, 4장에서는 수행한 실험을 설명하고 실험 결과를 분석한다. 마지막으로 5장에서 결론을 짓는다.

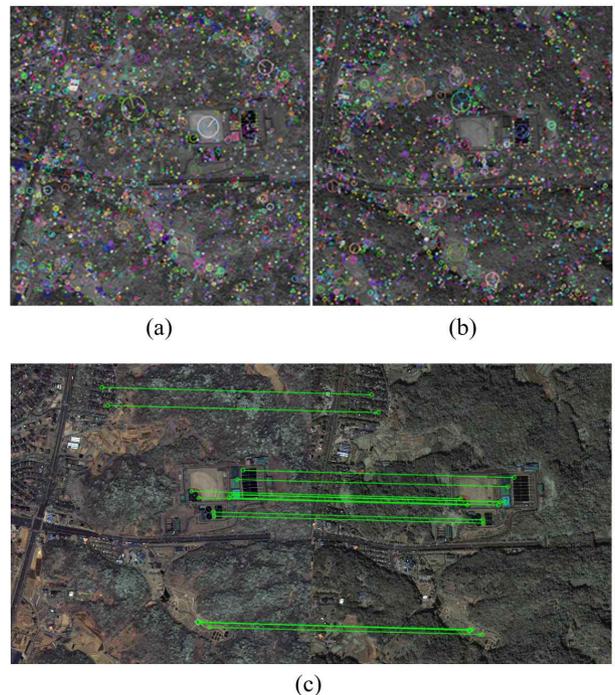


그림 1. 원격 탐사 영상에서 검출된 특징점과 SIFT 기반 정합 방법을 사용해 매칭된 결과

Fig. 1. The results of matching the features selected from the remote sensing image with the SIFT-based registration

II. 관련 이론

1. 딥러닝

컴퓨터 비전 (Computer Vision) 분야에서 딥러닝이 최근

에 주목을 받게 된 것은 객체 인식 (Object Recognition) 문제의 성공적인 적용에서 비롯된 면이 많다^[8]. 기계학습 관점에서, 객체 인식 문제는 분류 문제로 볼 수 있다. 일반적으로 분류 문제는 분류할 입력 데이터 세트의 특징을 보여주는 특징 벡터를 생성하고, 이 특징 벡터를 지지 벡터 모델 (Support Vector Model) 등의 분류기 (Classifier)에서 분류하는 방식으로 다루어진다. 이 분류기들은 주로 학습 데이터 세트 (Training Data Set)로부터 자동 구축되지만, 해당 분류 문제에 적합하며, 입력을 통해서 쉽게 생성할 수 있는 특징 벡터를 정의하는 것은 오랫동안 연구자의 역할이었다. 특징 벡터를 정의하는 과정도 학습 데이터 세트로부터 자동으로 구축하는 기술인 표현 학습 (Representation Learning) 연구가 되어 왔으며, 그 대표적인 예가 딥러닝이다. 보통 다단계 분류기 집합으로 다층신경망 (Multi-layer Neural Network)이 사용되는데, 특히 컴퓨터 비전에서는 다층신경망의 일종인 합성곱 신경망 (Convolutional Neural Network) 이 많이 사용된다.

2. U-Net

U-Net^[9]은 생물의학 (Biomedical) 분야에서 영상 분할을 목적으로 제안된 단대단 (End-to-end) 방식의 완전 합성곱 신경망 기반 모델이다. 의료 연구에서는 자동화된 현미경 실험으로 몇 테라바이트 (Terabyte)에 달하는 영상이 생성되는데, 이에 대해 각 영상에 대해 시맨틱 분할 (Semantic Segmentation) 방법을 통해 셀 (Cell)과 셀 구성 요소를 분류했다. 결과적으로 U-Net의 구조를 활용하여 아주 적은 양의 학습 데이터만으로 데이터 증강 (Data Augmentation)을 활용하여 여러 생물의학 영상 분할 문제에서 우수한 성능을 보여주었다. U-Net은 고차원 형태의 영상을 저차원 형태의 영상으로 변경시켜주는 부호화기 (Encoder)와 부호화된 영상을 다시 원래 형태로 복원해주는 복호화기 (Decoder) 구조를 가지고 있다. 특히 기본적인 자동부호기 (Autoencoder) 구조에서 부호기에서 축약된 정보를 복호기에 연결해서 사용하는 특징을 갖는데, 이러한 과정으로 인해 기존 자동부호기 과정에서 활용하지 못했던 공간 정보를 얻을 수 있다는 장점이 있다.

3. 특징점 검출기

가장 널리 사용되는 검출기는 1988년에 제안된 해리스 코너 (Harris Corner) 검출기이다^[10]. 해리스 코너 검출기는 두 번째 모멘트 (Moment) 행렬의 고유값을 기초로 한다. 그러나 해리스 코너 검출기는 크기 변화에 강인하지 않다. [11]에서는 자동 크기 선택 개념을 소개했다. 이를 통해 영상에서 각각 고유한 특성 척도를 가진 특징점을 검출할 수 있다. 블롭과 같은 구조를 검출하기 위해 헤시안 (Hessian) 행렬의 결정인자 뿐만 아니라 라플라시안 (Laplacian)도 모두 실험했다. [12]에서는 반복성이 높은 크기 변화에 강인한 검출기를 만들었다. 해리스 측정값이나 헤시안 행렬의 결정 인자를 사용하여 위치를 선택하고, 라플라시안을 사용하여 크기를 선택했다. [13]에서는 속도에 초점을 맞추어 가우시안 차이 (DoG) 필터로 라플라시안 오브 가우시안 (LoG)을 근사화할 것을 제안했다. 크기 변화에 강인한 특징점 검출기들이 추가로 제안되었는데, 그 예로는 지역 내 엔트로피 (Entropy)를 최대화하는 [14]에서 제안된 두드러진 지역 검출기와 [15]에서 제안한 가장자리 기반 지역 검출기가 있다.

4. 특징점 서술자

가우스 (Gauss) 미분, 모멘트 불변, 복잡한 특징, 조정 가능한 필터 (Filter), 위상 기반 지역 특징, 특징점 인접 지역 내에서 소규모 특징의 분포를 나타내는 서술자와 같은 다양한 특징점 서술자가 제안되었다^{[16][17][18][19][20]}. [5]에서 제안한 SIFT는 다른 것들보다 성능이 뛰어난 것으로 나타났다. SIFT의 서술자는 특징점 주변의 지역 지향 그래디언트 히스토그램을 (Gradient histogram) 계산하고 128차원 벡터에 저장한다^[5]. [21]에서는 검출된 특징점 주위의 그래디언트 영상에 PCA (Principal Component Analysis)를 적용했다. PCA-SIFT는 SIFT의 2단계까지의 방법은 같다. SIFT에서 키포인트를 중심으로 128차원의 특징점을 서술하는데 반해, PCA-SIFT에서는 41×41 의 패치를 추출한다. 여기서 검출된 패치와 미리 계산된 고유벡터를 곱하여 36차원으로 축소된 특징점을 구할 수 있다. PCA-SIFT는 SIFT에 비해 차원수가 적어 특징 매칭 속도가 빠르다는 장점이 있

다. 하지만 [22]의 연구에서는 SIFT에 비해 정확도가 떨어지는 것이 입증되었다. [22]에서는 같은 수의 차원에서 훨씬 더 독특함이 입증된 GLOH (Gradient Location and Orientation Histogram)라는 SIFT의 변형을 제안했다. 그러나 GLOH는 데이터 압축에 PCA를 다시 사용하므로 계산 비용이 더 많이 든다. SIFT 서술자는 정확도가 높으면서도 실행 시간이 빠르기 때문에 가장 널리 사용되고 있다.

5. 고속화 알고리즘

SURF는 이차원 하르 웨이블릿 (Haar Wavelet)을 기반으로 하는 밝기의 기울기 정보를 이용하여 영상의 특징을 강인하게 표현할 수 있는 특징 검출 알고리즘이다^[4]. SURF는 특징의 회전이나 크기 변화에 강인하며 기존의 SIFT에서 가장 문제가 되었던 계산 속도를 개선하기 위한 3가지 방법을 적용하였다. 첫 번째는 적분 영상의 적용이다. 적분 영상이란 원점으로부터 각 픽셀 (Pixel) 위치까지의 사각형 영역의 모든 픽셀 값들의 합으로 만들어진 영상을 의미한다. 두 번째는 간단한 구조의 검출기와 서술자의 활용이다. 영상에서 SURF의 특징점은 64개의 원소를 가지는 벡터의 형태로 표현되고 개별 특징을 비교하는 작업은 많은 연산량을 필요로 한다. SURF의 검출기는 계산이 복잡해서 영상 전체 영역에 대하여 수행되지 않고 헤시안 검출기를 이용하여 특징이 잘 검출될 수 있는 영역에서만 연산이 이루어진다. 마지막은 간단한 특징 매칭이다. 앞서 헤시안 매트릭스에 의해 계산된 라플라시안의 부호를 비교하며 간단하게 매칭을 수행할 수가 있다. 연산량이 증가하지 않으면서 성능 감소가 없기 때문에 속도를 효율적으로 향상시킬 수 있다. [6]에서 상호 정보의 추정을 통한 영상 정합을 위해 두 가지 접근방식을 사용한다. 첫 번째는 MLMI (Maximum Likelihood Mutual Information) 기반 정합 방법이고 두 번째는 히스토그램 (Histogram) 기반 정합 방법이다. 상호 정보 추정에서 MLMI 기반 정합 방법이 히스토그램 기반 정합 방법 보다 훨씬 적은 계산 시간을 필요로 한다. [6]에서는 영상 정합의 순환을 단계별로 수행하고 그에 따라 최대 상호 정보를 찾을 수 있다. LPSIFT (Layer Parallel SIFT)는 프레임 수준 계산으로 인한 긴 지연시간을 피하기 위해 상자 커널 (Box Kernel)로 재구성된 계층 병렬 방법을 사용하

여 반복된 가우스 블러 연산을 대체했다^[7]. 상자 커널의 계산은 서브 커널 (Sub Kernel)의 합계를 재사용하는 통합 영상 접근법으로 더욱 단순화되었다. 하드웨어 설계의 경우에는 부분적으로 일시적인 결과만 저장하는 즉석 특징 검출 (On-the-fly Feature Extraction)을 사용했다. 또한 정밀도가 사이클 (Precision Equivalent Cycles)을 갖추어 특징점의 지역화 (Localization) 단계에서의 비용을 줄였다.

III. DLKF (Deep Learning Keypoint Filtering)

본 논문에서 제안하는 DLKF를 포함한 특징 기반 영상 정합 방법의 전체 구조는 그림 2와 같다. 전체 구조는 SIFT 기반 정합 방법에 DLKF를 추가한 형태이다. 먼저 입력 영상이 들어오면 서브 샘플링을 진행한다. 서브 샘플링된 영상은 특징점 검출기에서 특징점을 검출하고 동시에 구조물 분할 방법에서 구조물만 분할된 영상을 검출한다. 이때 구조물 분할 방법의 결과인 이진 영상 (Binary Image)에 대해 영상 팽창 (Dilation)을 적용하고 원본 영상과 요소 곱을 통하여 변환한다. 그 뒤 합성된 영상과 원본 영상은 검출된 특징점의 정보를 바탕으로 서술자를 계산한다. 두 영상에서 생성된 같은 위치의 서술자를 비교하여 특징점을 필터링한다. 마지막으로 필터링된 정보를 바탕으로 소스 영상을 기하학적으로 변형시켜 정합된 영상을 얻는다.

1. 구조물 검출 및 특징점 필터링

구조물에서 검출된 특징점과 배경에서 검출된 특징점을 구분하기 위하여 그림 3과 같이 구조물과 배경을 구분하는 구조물 분할 방법을 추가하였다. 구조물 분할 방법은 영상 분할 분야에서 널리 사용되고 있는 U-Net을 사용하였다^[9]. 입력 영상은 256×256 크기의 패치 단위로 변환이 이루어진 후 다시 원본영상의 크기로 합쳐진다. 특징점 필터링 방법은 다음과 같은 순서로 진행된다. 먼저 구조물 분할 방법을 통해서 이진 영상으로 변환된 영상은 다시 원본 영상과의 요소 곱을 통하여 배경이 제거된 영상으로 변환된다. 그 후, 배경이 제거된 영상은 SIFT의 검출기에서 검출된 원본 영상의 특징점에 대한 정보를 기반으로 서술자를

생성한다. 생성된 서술자는 그림 4와 같이 원본 영상에서 같은 위치에서 생성된 서술자와 일치 여부를 비교한다. 두

영상에서 생성된 서술자가 일치한다면 원본 영상에서 그 위치의 특징점은 보존되고 일치하지 않으면 제거된다.

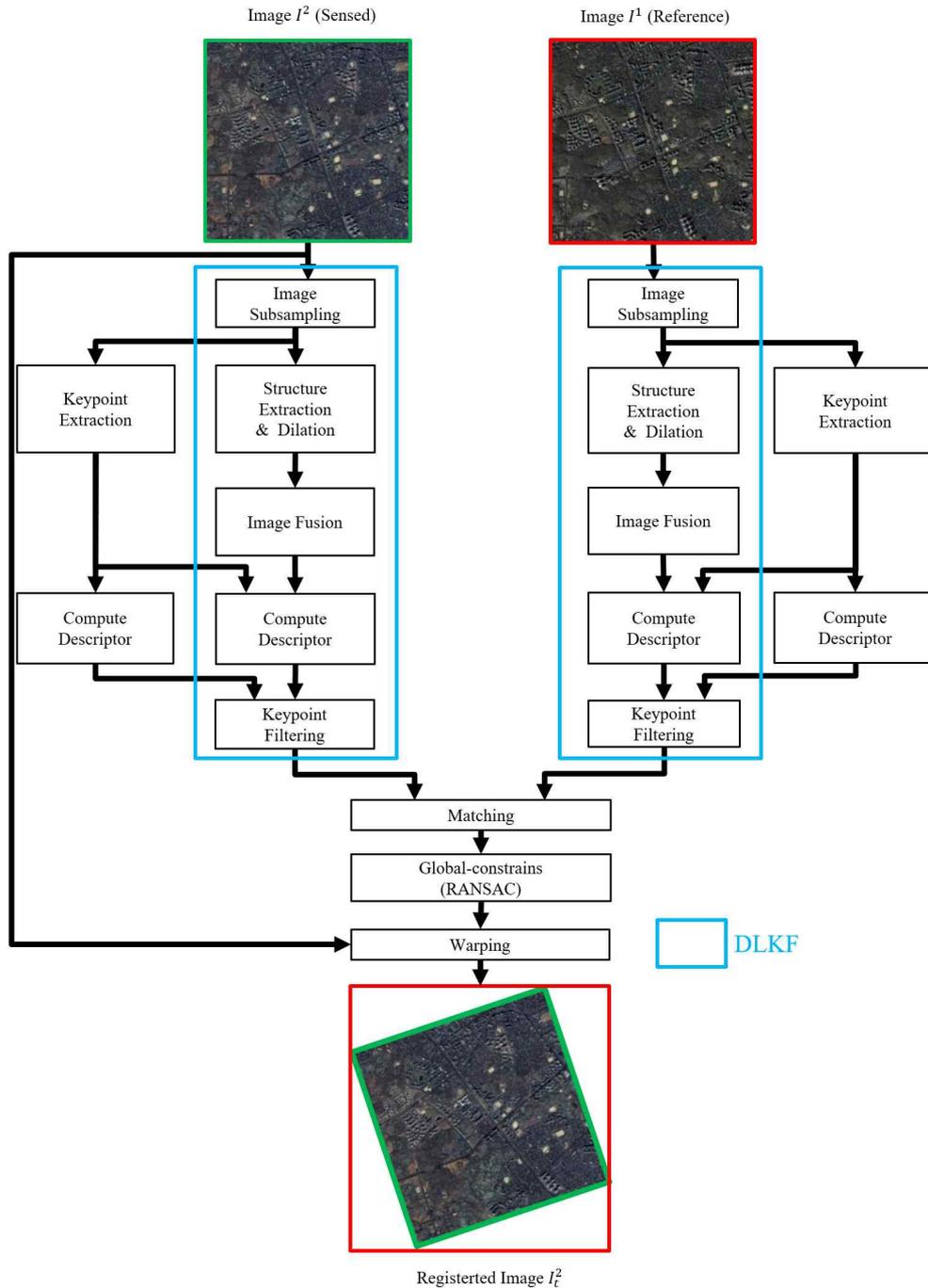


그림 2. DLKF를 포함한 특징 기반 영상 정합 방법의 전체 구조
 Fig. 2. The architecture of the feature-based image Registration method, including DLKF

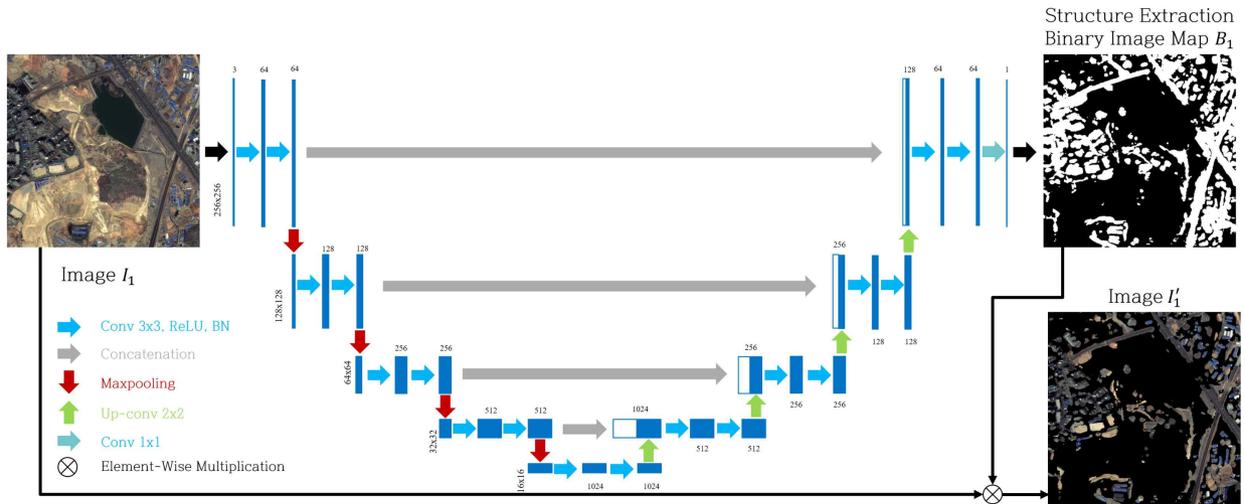


그림 3. U-Net 기반 구조물 분할 방법의 구조
 Fig. 3. The architecture of U-Net based structure segmentation method

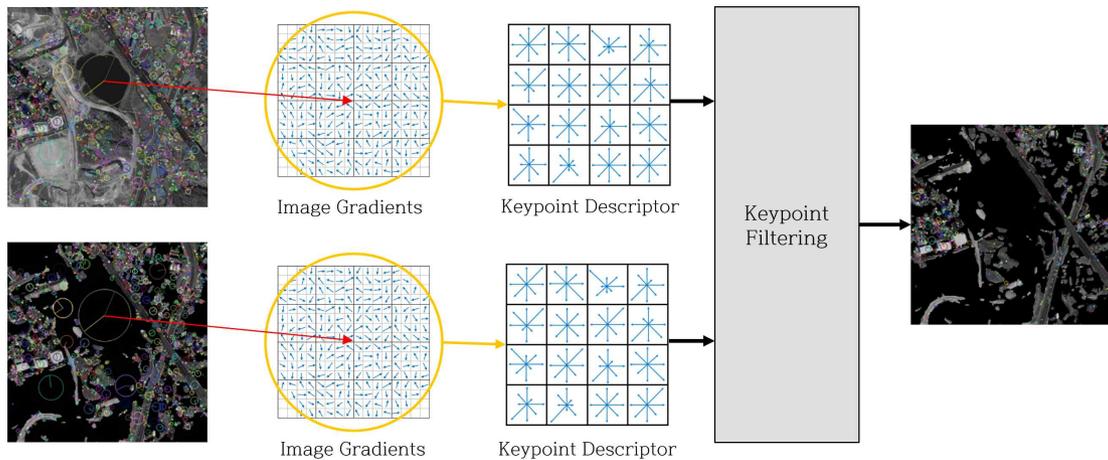


그림 4. 특징점 필터링 과정 예시
 Fig. 4. Example of the keypoint filtering process

2. 구조물 경계 확장

특징점 필터링 과정에서 구조물의 내부 경계와 가까운 위치에서 검출되는 특징점은 서술자를 구하는 과정에서 그 범위가 구조물의 범위를 벗어나 영상이 제거된 영역까지 포함된다. 이 특징점은 구조물에서 검출되는 특징점이지만 서술자를 비교하는 과정에서 원본 영상의 같은 특징점의

서술자와 일치하지 않는 문제가 발생한다. 그림 5의 (a)는 원본 영상의 특징점 검출기의 결과를 보여주고 (b)는 같은 영상에 특징점 필터링을 적용한 결과를 보여준다. 초록색 상자 영역은 원래 특징점이 있던 위치로 구조물에서 뽑히는 특징점이지만 서술자의 크기 범위를 나타내는 빨간색 상자 영역에 배경이 제거된 영역이 포함되어 필터링 과정 이후 제거되는 모습을 볼 수 있다.

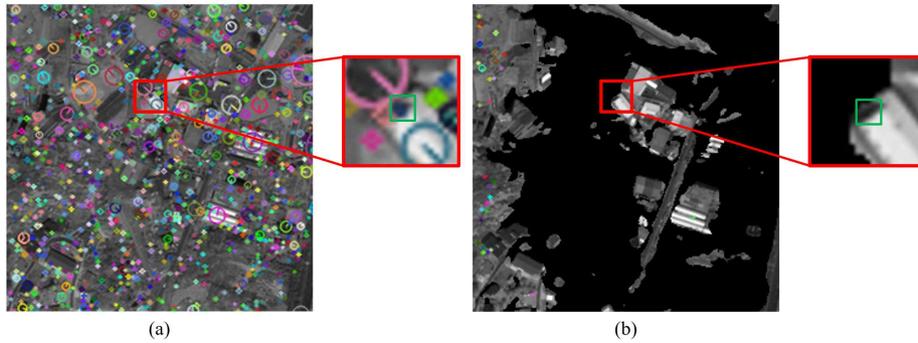


그림 5. 구조물의 경계와 가까운 특징점에 대한 특징점 필터링 결과. (a) 입력 영상에 대한 SIFT 기반 특징점 검출기의 결과, (b) 입력 영상에 특징점 필터링을 적용한 결과

Fig. 5. The keypoint filtering results for close keypoint to boundaries of artificial structure. (a) SIFT-based feature point detector results for the input image, (b) The result of applying keypoint filtering to the input image

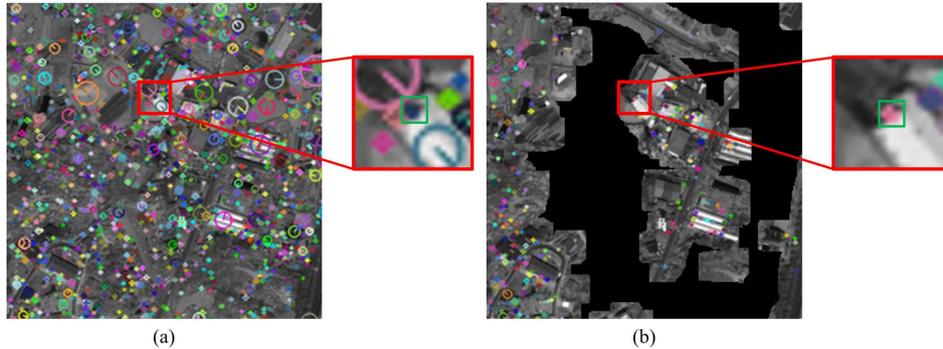


그림 6. 구조물의 경계와 가까운 특징점에 대한 영상 팽창을 포함한 특징점 필터링 결과. (a) 입력 영상에 대한 SIFT 기반 특징점 검출기의 결과, (b) 입력 영상에 영상 팽창을 포함한 특징점 필터링을 적용한 결과

Fig. 6. The keypoint filtering including image expansion results for close keypoint to boundaries of artificial structure. (a) SIFT-based keypoint detector results for the input image, (b) Result of applying keypoint filtering including image dilation to the input image

본 논문에서는 구조물의 경계에서 보존되는 영상의 영역을 확장하는 방법을 추가하여 이 문제를 해결하고자 한다. 확장 방법으로는 영상 팽창을 적용한다. 특징점이 영상의 경계에서 뽑히는 경우까지 고려하여 특징점 서술자의 절반 크기만큼의 크기의 영상 팽창을 진행한다. 또한 영상 팽창을 적용할 경우 확대될 잡음을 고려하여 침식 (Erosion)을 먼저 적용한 뒤 팽창을 적용하였다. SIFT의 경우 서술자의 크기가 16×16 이기 때문에 소수점 오차를 고려하여 5×5 크기만큼의 침식을 적용하고 23×23 크기만큼의 팽창을 적용한다. 영상 팽창을 적용하면 그림 6의 (b)와 같이 구조물의 가장자리에서 뽑히는 특징점도 특징점을 기준으로 서술자의 크기만큼의 영역에서 영상이 제거된 영역을 포함하지 않는 것을 확인할 수 있다.

3. 서브 샘플링 및 패치 중복시켜 잘라내기

영상 정합을 위하여 필수적인 특징점을 잃지 않으면서 그 개수를 줄이기 위하여 추가로 서브 샘플링을 사용한다. 표 3은 원본 영상과 쌍입방 보간법 (Bicubic Interpolation)을 사용하여 서브 샘플링된 영상을 SIFT 기반 정합 방법을 사용하여 비교 실험한 결과이다. 성능 비교에 대한 지표는 [23]에 소개된 측정 지표를 사용하였다. 가장 보편적으로 사용되는 SIFT 기반 정합 방법을 기준으로 하여 비교하였을 때 서브 샘플링된 영상을 사용하는 방법은 특징점의 수를 약 67% 감소시키고 속도를 약 9.3배 향상시키면서 정확도는 3.42에서 3.64로 정확도가 비슷하게 유지되었다.

표 3. 원격 탐사 영상에 대한 SIFT 기반 정합 방법 및 서브 샘플링이 추가된 방법의 결과 비교

Table 3. Comparison of results of SIFT-based registration method for remote sensing images and method with subsampling added

Methods	N_{red}	RMS_{all}	RMS_{LOO}	P_{quad}	BPP	S_{kew}	S_{cat}	ϕ	Number of Keypoints	Calculation Time (s)
SIFT	105	11.85	11.88	1.00	0.98	0.17	0.83	3.42	424,998	2,266.55
SIFT + Subsampling	97	12.68	12.77	1.00	0.92	0.18	0.88	3.64	136,069	244.85

그림 7은 원격 탐사 영상에 대한 구조물 분할 결과를 보여준다. 영상의 빨간 상자 영역을 확대해보면 패치의 경계에 잡음이 생기는 것을 확인할 수 있다. 기존의 U-Net은 그림 8의 (a)와 같이 영상을 영역이 중복되지 않는 패치 단위로 잘라서 영상 분할을 진행하였기 때문에 영상을 다시 하나로 합칠 때 패치 경계 부분에 잡음이 발생하는 문제가

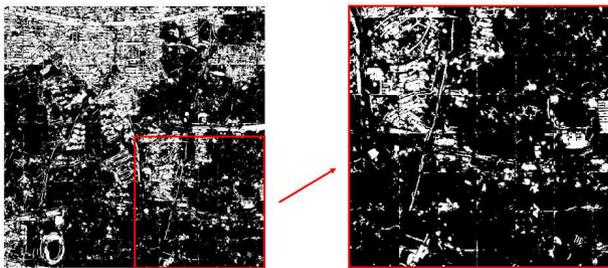


그림 7. 구조물 분할 결과
Fig. 7. The results of structure segmentation

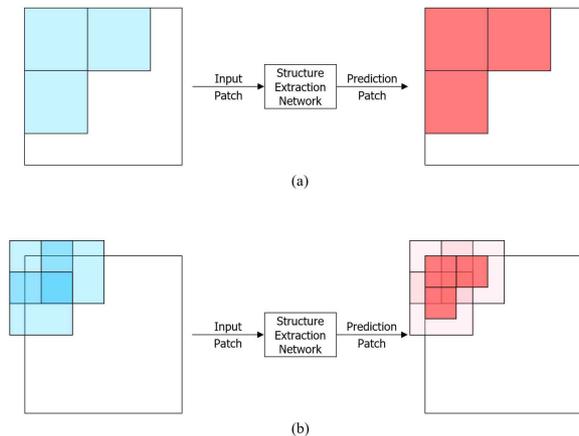


그림 8. 패치 잘라내기 방법 비교 (a) 패치를 중복시키지 않고 잘라낸 방법의 예시, (b) 패치를 중복시켜 잘라낸 방법의 예시
Fig. 8. Comparison of patch crop methods. (a) Example of how to crop patches without overlapping, (b) Example of how to crop overlapping patches

존재한다. 본 논문에서는 이 부분을 개선하기 위해 그림 8의 (b)과 같이 영상의 패치를 자를 때 각각의 패치가 50%씩 중복되도록 패치를 잘라낸다. 잘라낸 영상 패치는 구조물 분할 방법을 통과하여 전체 영역 중에서 중심 부분을 차지하는 영역 25%만을 사용하여 다시 하나의 영상으로 합쳐진다. 그림 9는 기존의 방법과 패치 중복시켜 잘라내기 방법의 실험결과를 보여준다. 그림 9의 (a)에서 패치 경계에서 보이는 잡음이 패치 중복시켜 잘라내기 방법을 추가한 분할 결과인 그림 9의 (b)에서는 없어진 것을 확인할 수 있다.

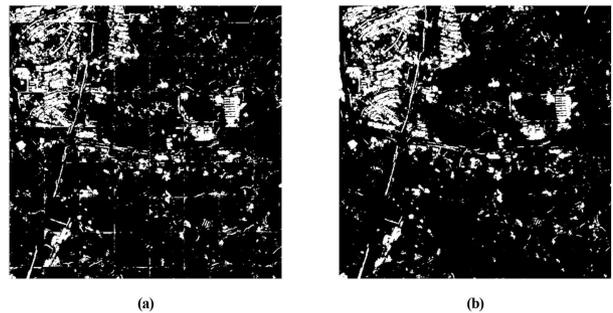


그림 9. 잘라내기 방법에 따른 구조물 분할 결과 비교 (a) 패치를 중복시키지 않고 잘라낸 방법의 분할결과, (b) 패치를 중복시켜 잘라낸 방법의 분할결과
Fig. 9. Comparison of structure segmentation results according to the crop method. (a) The segmentation result of how to crop patches without overlapping, (b) The segmentation result of how to crop overlapping patches

IV. 실험 결과

본 논문에서는 SIFT 및 SURF 기반 정합 방법과 성능 비교를 진행한다. 구현 및 실험은 PyCharm 인터페이스에서 이루어졌고, GPU는 Nvidia GeForce 1080ti를 사용하였다. 시험 데이터 세트는 2012년 5월 18일에 발송된 다목적실용

위성인 아리랑 3호 (KOMPSAT-3)에서 취득된 서울 및 여의도 지역의 동일한 지역을 다른 시간대에 촬영한 위성 탐사 영상을 사용하였다. 구조물 분할 방법의 학습 데이터 세트는 공개된 영상 분할 데이터 세트인 Massachusetts Roads 데이터 세트와 Massachusetts Buildings 데이터 세트를 사용하였다. 원격 탐사 영상인 시험 데이터 세트와 항공 영상인 학습 데이터 세트의 특성의 차이를 보완하기 위해 히스토그램 사상 (Mapping)을 통한 데이터 증강 (Data Augmentation)을 진행하였다.

본 논문의 실험에서는 [23]에 소개된 측정 지표를 사용하여 제어점 수 (N_{red}), 모든 제어점을 기반으로 픽셀 크기로 정규화된 루트 평균 제곱 오차 (Root Mean Square Error, RMS_{all}), 리브-원-아웃 (Leave-one-out) 방법을 기반으로 제어점의 잔차로 계산된 루트 평균 제곱 오차 (RMS_{LOO}), 사분면에 걸친 잔차 분포의 통계적 평가 (P_{quad}), 표준 값이 1.0보다 큰 불량점의 비율 ($BPP(1.0)$), 잔차 산점도 (Residual Plot)의 선호 축 존재에 대한 통계적 평가 (S_{kew}), 영상 전체에 걸친 제어점 분포의 선도에 대한 통계적 평가 (S_{cut}), 위의 7개의 지표의 가중치 합 (ϕ)를 포함하여 영상 정합의 정확도 성능을 평가한다. N_{red} 를 제외하고 모든 지표는 값이 작을수록 성능이 좋다. 또한 정합 결과에 대한 모자이크 영상은 영상 정합의 정성적 평가에 사용된다.

표 4는 기준 방법인 SIFT 기반 정합 방법을 실험한 결과와 본 논문에서 제안하는 방법들을 하나씩 추가했을 때의

실험 결과들이다. 기준 방법에 성능 향상의 요인 4가지를 순차적으로 추가해서 사용함에 따라 성능이 올라감을 확인할 수 있다. SIFT + Ours 1은 구조물 분할 방법과 특징점 필터링이 추가된 정합 실험 결과이다. 3장에서 언급한 구조물에서 검출된 특징점 중에서 서술자를 생성할 때 배경영역을 포함하여 필터링되는 문제점으로 인해 성능이 크게 저하된다. SIFT + Ours 2는 SIFT + Ours 1에 영상 팽창을 추가한 정합 실험 결과이다. 기준 방법에 비해 정확도는 5.16으로 저하되지만, 속도가 약 4.2배 개선된다. SIFT + Ours 3는 SIFT + Ours 2에 서브 샘플링을 추가하여 정합 실험을 진행한 결과이다. 이전 방법에 비해 정확도는 5.90으로 저하되지만, 속도가 약 5.3배 개선된다. 마지막으로 SIFT + DLKF는 패치를 중복시켜 잘라낸 방법까지 모두 추가한 제안 방법으로 정합 실험을 진행한 결과이다. 이전 방법에 비해 속도는 약 0.9배 저하되지만, 정확도가 4.51로 향상된다.

표 5는 아리랑 3호에서 취득된 원격 탐사 영상 10개 데이터 세트로 7개의 지표에 대하여 성능 평가를 진행한 결과이다. 가장 보편적으로 사용되는 SIFT 기반 정합 방법을 기준으로 하여 비교하였을 때 SIFT 기반 정합 방법 대비 SURF 기반 정합 방법은 정확도가 떨어지지만, 속도 측면에서 개선이 있었다. DLKF를 포함한 SIFT 기반 정합 방법은 SURF 기반 정합 방법보다 정확도 측면에서 성능 저하가 덜하면서 속도 측면에서 SIFT 및 SURF 기반 정합 방법보

표 4. 제안하는 방법의 추가에 따른 기준 방법과의 성능 비교.

Table 4. Performance comparison with the baseline method according to the addition of the proposed method.

Methods	N_{red}	RMS_{all}	RMS_{LOC}	P_{quad}	BPP	S_{kew}	S_{cut}	ϕ	Calculation Time (s)
SIFT	105	11.85	11.88	1.00	0.98	0.17	0.83	3.42	2,266.55
SIFT + Ours 1	9	382,546.97	433,120.55	1.00	0.98	0.52	0.96	104,066.21	34.86
SIFT + Ours 2	63	18.76	19.03	1.00	0.79	0.22	0.86	5.16	540.61
SIFT + Ours 3	73	21.65	21.80	1.00	0.94	0.23	0.90	5.90	102.20
SIFT + DLKF	73	16.02	16.26	1.00	0.91	0.22	0.98	4.51	110.66

표 5. DLKF를 포함한 SIFT 기반 정합 방법과 기존 방법들의 성능 비교.

Table 5. Performance Comparison between SIFT-based registration methods with DLKF and the conventional methods.

Methods	N_{red}	RMS_{all}	RMS_{LOC}	P_{quad}	BPP	S_{kew}	S_{cut}	ϕ	Number of Keypoints	Calculation Time (s)	Speed Improvement Ratio
SIFT	105	11.85	11.88	1.00	0.98	0.17	0.83	3.42	424,998	2,266.55	100%
SURF	74	19.83	19.91	1.00	0.95	0.16	0.90	5.43	349,308	870.75	260%
SIFT + DLKF	73	16.02	16.26	1.00	0.91	0.22	0.98	4.51	76,532	110.66	2050%

다 개선된 것을 확인할 수 있다. SURF 기반 정합 방법은 특징점의 수를 약 18% 감소시키면서 속도를 약 2.6배 향상시켰지만 정확도가 3.42에서 5.43으로 저하됐다. 그러나 DLKF를 포함한 SIFT 기반 정합 방법을 사용하였을 때 특징점의 수를 약 82% 감소시키고 속도를 약 20.5배 향상시키면서 정확도는 4.51로 저하됐다.

DLKF를 포함한 SIFT 기반 정합 방법으로 시험한 결과 영상 중 그 예를 그림 10에서 보인다. 그림 10의 (a)와 (b)는 2014년 3월과 2015년 10월에 여의도 지역에서 취득된 영상에 대해 $4,000 \times 4,000$ 크기로 자른 한 쌍의 영상을 보여준다. 그림 10의 (c)-(e)는 이 한 쌍의 영상을 SIFT 및 SURF 기반 정합 방법 및 DLKF를 통해 하나의 체크 보드 (Check Board) 영상으로 나타낸다. 초록 상자 영역을 확대한 부분의 빨간 상자 영역을 비교했을 때, SURF 기반 정합 방법의 결과 영상에서 어긋나 있는 건물 부분이 DLKF를 포함한 SIFT 기반 정합 방법의 결과 영상에서는 SIFT 기반 정합

방법의 결과 영상과 유사한 것을 확인할 수 있다.

V. 결 론

본 논문에서는 원격 탐사 영상에 대한 특징 기반 영상 정합 방법의 고속화를 위한 딥러닝 기반 특징점 필터링 방법인 DLKF를 제안했다. 기존의 특징 기반 영상 정합 방법의 복잡도는 특징 매칭 단계에서 발생한다. 이 복잡도를 줄이기 위하여 본 논문에서는 특징 매칭이 영상의 구조물에서 검출된 특징점으로 매칭되는 것을 확인하여 특징점 검출기에서 검출된 특징점 중에서 구조물에서 검출된 특징점만 필터링하는 방법을 제안하였다. DLKF는 영상 정합을 위하여 필수적인 특징점을 잃지 않으면서 그 수를 줄이기 위하여 구조물의 경계와 인접한 특징점의 보존하고, 서브 샘플링된 영상을 사용하며, 구조물 분할 방법의 결과에서

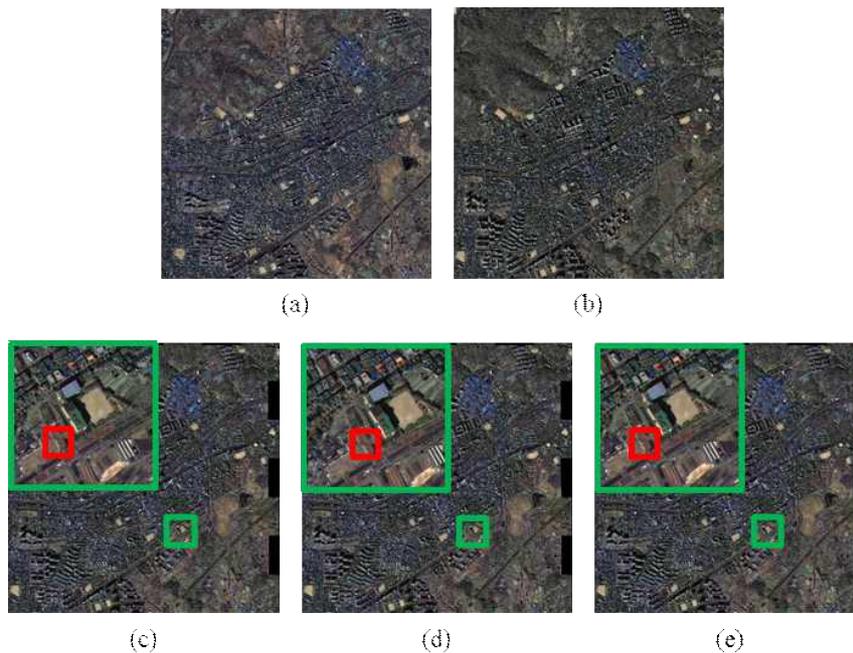


그림 10. 아리랑 3호 위성 원격 탐사 영상에 대한 기존 방법들과 DLKF를 포함한 SIFT 기반 정합 방법의 결과. (a) 감지 영상, (b) 기준 영상, (c) SIFT 기반 정합 방법의 결과 영상과 기준 영상을 모자이크 한 결과, (d) SURF 기반 정합 방법의 결과 영상과 기준 영상을 모자이크 한 결과, (e) DLKF를 포함한 SIFT 기반 정합 방법의 결과와 기준 영상을 모자이크 한 결과

Fig. 10. The registration results of the conventional methods and SIFT-based registration methods with DLKF for images of KOMPSAT-3. (a) Sensed image, (b) Reference image, (c) The mosaic image of reference images and result of SIFT-based registration methods, (d) The mosaic image of reference images and result of SURF-based registration methods, (e) The mosaic image of reference images and result of SIFT-based registration methods with DLKF

생기는 영상 패치 경계의 잡음을 제거하기 위하여 영상 패치를 중복하여 잘라냄으로써 정합 속도를 향상시킨다. 또한, 알고리즘의 학습 단계에서는 학습 영상과 시험 영상의 특성의 차이로 인해 생기는 성능 저하문제를 개선하기 위해 히스토그램 사상 기법을 사용하여 데이터 증강을 진행하였다. 제안하는 방법의 성능을 검증하기 위하여 아리랑 3호 위성 원격 탐사 영상을 사용하여 기존 특징점 검출 방법과 속도와 정확도를 비교하였다. 가장 보편적으로 사용되는 SIFT 기반 정합 방법을 기준으로 하여 비교하였을 때 SIFT 방법의 속도를 개선시킨 SURF 기반 정합 방법은 특징점의 수를 약 18% 감소시키고 속도를 약 2.6배 향상시켰지만 정확도가 3.42에서 5.43으로 저하되었다. 그러나 제안하는 방법인 DLKF를 사용하였을 때 특징점의 수를 약 82% 감소시키고 속도를 약 20.5배 향상시키면서 정확도는 4.51로 저하되었다.

참 고 문 헌 (References)

- [1] B. Zitova, J. Flusser, "Image registration methods: a survey" *Image and Vision Computing*, Vol.21, No.11, pp.977-1000, June 2003.
- [2] S. Suri, P. Reinartz, "Mutual-information-based registration of terra-sar-x and ikonos imagery in urban areas" *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48, No.2, pp.939-949, November 2009.
- [3] J.P. Kern, M.S. Pattichis, "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] H. Bay, A. Ess, T. Tuytelaars, L.V. Gool, "Speeded-up robust features (surf)" *Computer Vision and Image Understanding*, Vol.110, No3, pp.346-359, June 2008.
- [5] D. Lowe, "Distinctive image features from scale-invariant keypoints" *International Journal of Computer Vision*, Vol.60, No.2, pp.91-110, January 2004.
- [6] M.I. Patel, V.K. Thakar, "Speed Improvement in Image Registration using Maximum Likelihood based Mutual Information" *International Conference on Advanced Computing and Communication Systems*, Coimbatore, India, pp.1-3, 2015.
- [7] L.C. Chiu, T.S. Chang, J.Y. Chen, N.Y.C. Chang, "Fast SIFT Design for Real-Time Visual Feature Extraction" *IEEE Transactions on Image Processing*, Vol. 22, No. 8, pp.3158-3167, August 2013.
- [8] Y. LeCun, Y. Bengion and G. Hinton, "Deep learning," *Nature*, Vol. 521, pp.436-444, May 2015.
- [9] 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.
- [10] C. Harris, M. Stephens, "A combined corner and edge detector" *Alvey Vision Conference*, Manchester, UK, pp. 147-151, 1988.
- [11] T. Lindeberg, "Feature detection with automatic scale selection", *International Journal of Computer Vision*, Vol. 30, No.2, pp.79-116, November 1998.
- [12] K. Mikolajczyk, C. Schmid, "Indexing based on scale invariant interest points" *International Conference on Computer Vision*, Vancouver, Canada, pp.525-531, 2001.
- [13] D. Lowe, "Object recognition from local scale-invariant features" *International Conference on Computer Vision*, Kerkyra, Greece, pp.1150-1157, 1999.
- [14] T. Kadir, M. Brady, "Scale, saliency and image description" *International Journal of Computer Vision*, Vol. 45, No.2, pp.83-105, November 2001.
- [15] F. Jurie, C. Schmid, "Scale-invariant shape features for recognition of object categories" *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, pp.90-96, 2004.
- [16] L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, M.A. Viergever, "General intensity transformations and differential invariants" *Journal of Mathematical Imaging and Vision*, Vol.4, No.2, pp.171-187, May 1994.
- [17] F. Mindru, T. Tuytelaars, L. Van Gool, T. Moons, "Moment invariants for recognition under changing viewpoint and illumination" *Computer Vision and Image Understanding*, Vol.94 No.1 - 3, pp.3-27, April 2004.
- [18] A. Baumberg, "Reliable feature matching across widely separated views" *IEEE Conference on Computer Vision and Pattern Recognition*, Hilton Head Island, SC, USA, pp.774-781, 2000.
- [19] W.T. Freeman, E.H. Adelson, "The design and use of steerable filters" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.13, No.9, pp.891-906, September 1991.
- [20] G. Carneiro, A.D. Jepson, "Multi-scale phase-based local features" *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Madison, WI, USA, pp.736-743, 2003.
- [21] Y. Ke, R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors" *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, pp.506-513, 2004.
- [22] K. Mikolajczyk, C. Schmid, "A performance evaluation of local descriptors" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.27, No.10, pp.1615-1630, October 2003.
- [23] H. Goncalves, J.A. Goncalves, L. Corte-Real, "Measures for an objective evaluation of the geometric correction process quality" *IEEE Geoscience and Remote Sensing Letters*, Vol.6, No.2, pp.292-296, April 2009.

저 자 소 개



성 준 영

- 2019년 2월 : 광운대학교 전자공학과 학사
- 2019년 3월 ~ 현재 : 광운대학교 전자공학과 석사과정
- ORCID : <https://orcid.org/0000-0002-6493-8511>
- 주관심분야 : 영상신호처리, 컴퓨터비전, 고해상도 위성영상 처리



이 우 주

- 2018년 2월 : 광운대학교 전자공학과 학사
- 2020년 8월 : 광운대학교 전자공학과 석사
- 2020년 9월 ~ 현재 : 광운대학교 졸업 후 연구원
- ORCID : <https://orcid.org/0000-0002-3388-3765>
- 주관심분야 : 영상신호처리, 컴퓨터비전, 고해상도 위성영상 처리



오 승 준

- 1980년 2월 : 서울대학교 전자공학과 학사
- 1982년 2월 : 서울대학교 전자공학과 석사
- 1988년 5월 : 미국 Syracuse University 전기/컴퓨터공학과 박사
- 1982년 3월 ~ 1992년 8월 : 한국전자통신연구원 멀티미디어연구실 실장
- 1986년 7월 ~ 1986년 8월 : NSF Supercomputer Center 초청 학생연구원
- 1987년 5월 ~ 1988년 5월 : Northeast Parallel Architecture Center 학생연구원
- 1992년 3월 ~ 1992년 8월 : 충남대학교 컴퓨터공학부 겸임교수
- 1992년 9월 ~ 현재 : 광운대학교 전자공학과 교수
- 2002년 3월 ~ 2017년 12월 : SC29-Korea 전문위원회 대표위원
- ORCID : <https://orcid.org/0000-0002-5036-3761>
- 주관심분야 : 비디오데이터처리, 영상압축, 컴퓨터비전