

특집논문 (Special Paper)

방송공학회논문지 제28권 제4호, 2023년 7월 (JBE Vol.28, No.4, July 2023)

<https://doi.org/10.5909/JBE.2023.28.4.429>

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

특징 추출을 이용한 이미지 매칭 기술의 최근 동향 분석

박승기^{a)}, 채문주^{a)}, 조성인^{a)‡}

Analysis of Recent Trends in Image Matching Technology Based on Feature Extraction

Seunggi Park^{a)}, Moon Ju Chae^{a)}, and Sung In Cho^{a)‡}

요 약

이미지 매칭은 기본적인 컴퓨터비전 기술의 한 분야로, 두 개 이상의 이미지를 비교하여 유사한 특징을 찾아 매칭하는 것을 의미한다. 이미지 매칭은 이미지 검색, 이미지 스티칭, 이미지 차이 검출 등 컴퓨터 비전의 응용분야에서 사용되기에 중요성이 강조되고 있다. 그 중 특징 추출을 통한 이미지 매칭은 현재 이미지 매칭 분야에서 가장 널리 쓰이는 기법이며 과거의 규칙 기반 기법부터 최근 딥러닝 기술 활용 기법까지 다양한 기법들이 활발히 제안되어 왔다. 본 논문에서는 특징 추출을 통한 이미지 매칭 기법을 특징점 추출과 추출된 특징점을 사용한 매칭으로 단계를 나누어 소개한다. 또한 기존의 규칙 기반 기법부터 최근 딥러닝을 활용한 기법까지의 발전 과정을 비교 설명하고 실험결과를 통해 각 기법을 분석한다.

Abstract

Image matching is a field of basic computer vision technology meaning finding similar features by comparing two or more images. In addition, the importance of image matching is emphasized because it is used in various applications of various computer visions such as image search, image stitching, and image difference detection. Among them, image matching through feature extraction is the most widely used technique in the field of image matching. Recently, new techniques introduced with the development of deep learning technologies that from the past rule-based techniques. In this paper, image matching technology through feature extraction is introduced by dividing it into two processes: feature, descriptor extraction and matching using feature descriptors. We also describe the process from existing rule-based techniques to recent techniques using deep learning. Finally, we analyze the performance of each technique through experimental results.

Keyword : Image Matching, Local Feature Matching, Feature Extraction, Feature Descriptors, Deep Learning

a) 동국대학교 AI소프트웨어융합학부(Division of AISW convergence, Dongguk University)

‡ Corresponding Author : 조성인(Sung In Cho)

E-mail: csi2267@dongguk.edu

Tel: +82-2-2260-3336

ORCID: <https://orcid.org/0000-0003-4251-7131>

※ 본 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 한국연구재단 기초연구사업의 지원을 받아 수행된 연구임(No. RS-2023-00208763).

※ This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00208763).

· Manuscript July 17, 2023; Revised July 24, 2023 Accepted July 24, 2023.

Copyright © 2023 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.”

I. 서론

컴퓨터 비전 분야에서 중요한 연구 분야 중 하나인 이미지 매칭은 두 개 이상의 이미지를 비교하여 이미지 사이의 유사한 특징을 찾는 작업이다. 이미지 매칭을 통해 이미지 사이의 시각적 유사도를 판단하여 관련성이나 특정한 패턴을 파악할 수 있다. 이미지 매칭은 컴퓨터 비전 분야의 근본적인 문제이며 이미지 검색, 객체 인식, 이미지 검색, 이미지 스티칭 등 다양한 응용분야에서 기본적인 토대로 사용될 수 있다. 따라서 이미지 매칭의 중요성은 날이 강조되고 있다.

이미지 매칭 분야에서 특징 추출을 통한 이미지 매칭은 가장 널리 쓰이는 기법이며 그림 1의 예시로 확인할 수 있다. 그림 1에서 각 점은 추출된 특징이고 점을 이은 선은 이 특징점 사이의 매칭된 관계를 나타낸다. 특징 추출을 통한 이미지 매칭은 이미지의 특징점 찾기 및 특징점을 기술하는 디스크립터 생성, 생성된 특징점 매칭 2단계로 나타난다. 특징점 찾기 및 특징점 기술 과정에서 이미지에서 중요한 지점들을 찾아내고, 이를 이미지의 특징을 간결하게 표현하는 디스크립터로 기술한다. 특징 추출 이후에는 생성된 특징점들을 디스크립터를 통해 비교하고 유사도를 측정하여 매칭 결과를 도출한다.

기존에는 이미지 매칭 분야에 규칙 기반 기법들이 많이 사용되었지만 최근 딥러닝 기술의 비약적인 발전으로 새로운 기법들이 도입되며 많은 연구가 수행되고 있으며 뛰어난 성능을 보여준다. 본 논문에서는 특징 추출을 통한 이미지 매칭 기술을 특징점 추출 및 기술과 특징점 매칭으로 앞서 서술한 두 단계로 나누어 기존의 규칙 기반 기법부터

최근 딥러닝을 활용한 기법까지의 과정을 비교, 서술하고 실험을 통해 이미지 매칭 분야에서의 다양한 기법들의 장단점과 성능을 평가한다. 본 논문은 이를 통한 이미지 매칭 분야에서의 연구에 대한 체계적인 이해와 새로운 발전 가능성에 대한 기여를 목표로 한다.

II. 본론

본 장에서는 규칙 기반 기법과 딥러닝 활용 기법을 바탕으로 특징점 추출 알고리즘과 특징점 매칭 알고리즘을 나누어 설명하고 비교한다.

1. 규칙 기반 이미지 매칭 기법

규칙 기반의 이미지 매칭 기법은 이미지 간의 관련성을 규칙과 패턴을 기반으로 파악하는 기법이다. 이 방법은 이미지에 대한 특징점을 추출하기 위해 특정한 규칙이나 제약 조건을 정의하고, 이를 통해 매칭을 수행한다. 본 절에서는 규칙 기반의 특징점 추출과 특징점 매칭에 대해 설명한다.

1.1 규칙 기반 특징점 추출

이미지에서 일반적으로 가장 중요하고 구분하기 쉬운 특징점들은 코너와 엣지이다. 여기서 엣지란 한 방향 이상에서 픽셀 값의 변화가 급격하게 나타나는 점이며 코너란 두 방향 이상에서 변화가 급격하게 나타나는 점을 이야기한다. 이러한 변화가 일어나는 지점은 일반적으로 이미지 내부

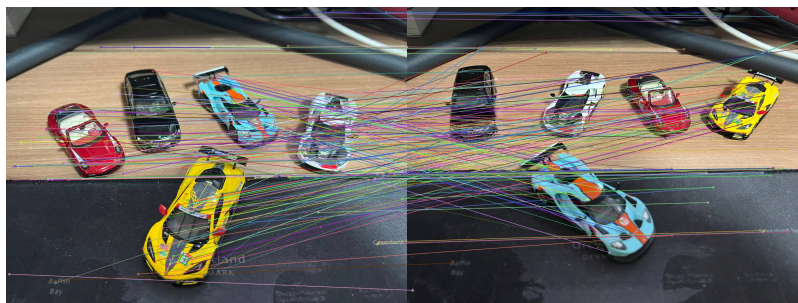


그림 1. 이미지 매칭의 예시
Fig. 1. Example of image matching

물체의 경계선이나 모서리와 관련되어 있기 때문에 많은 이미지 매칭 알고리즘에서 코너나 엣지를 특징점으로 사용한다.

Harris 코너 검출 알고리즘^[1]은 이미지에서 코너를 추출하는 알고리즘의 대표적인 예이다. Harris 코너 검출 알고리즘은 이미지의 각 픽셀에서 해당 픽셀을 중심으로 하는 작은 윈도우 영역을 추출한 뒤, 해당 윈도우 영역의 픽셀 값 변화를 기반으로 코너 여부를 판단한다. 변화의 크기가 작은 경우는 평탄한 영역이라고 가정하고, 큰 경우에는 코너 또는 경계선이라고 판단한다.

이후 등장한 Scale-Invariant Feature Transform (SIFT)^[2]는 20년간 이미지 매칭 분야에서 가장 널리 사용된 기법으로 이미지의 크기와 회전에 불변하는 특징을 추출하는 기법이다. SIFT는 입력 영상을 다양한 스케일로 변환하고 gaussian 블러 처리한 이미지 피라미드를 생성한 뒤, Difference of Gaussian (DOG)를 계산하여 겹치는 특징점을 Non-Maximum Suppression을 통해 제거하여 최대 최소의 극점을 가지는 특징점을 찾는다. 그림 2는 DOG를 계산하는 방법을 보여준다. 이후, 찾은 최댓값과 최솟값을 테일러 급수를 사용하여 보정시킨 뒤 특징점을 검출한다. Speeded Up Robust Features (SURF)^[3]는 SIFT 알고리즘을 기반으로 사용하지만 이미지 피라미드 형성방법의 변경과 적분영상을 사용하여 계산적인 효율성을 향상시킨 알고리

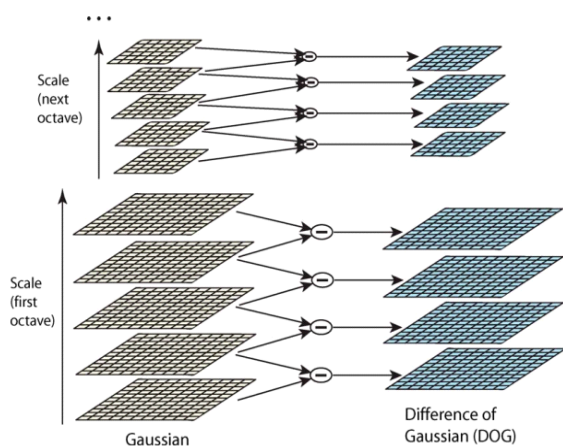


그림 2. Gaussian 블러 처리된 이미지 피라미드로 DOG를 계산하는 방법^[2]
 Fig. 2. Process of calculating DOG with Gaussian blurred image pyramid^[2]

즘이다.

Features from Accelerated Segment Test (FAST)^[4]의 경우 실시간 특징점 추출을 고려하여 만들어진 알고리즘으로 주변 픽셀과 중심 픽셀의 밝기를 비교하여 중심 픽셀의 특징점 여부를 판단한다. 특징점은 주변 픽셀의 밝기 값과 중심 픽셀의 밝기 값과의 차이가 특정 임계값보다 크거나 작은 경우로 정의한다. 특징점 판단 이후 특징점으로 판단된 픽셀을 중심으로 16개의 이웃 픽셀을 검사하여 특징점의 신뢰도를 측정한다. 그림은 FAST의 16개 주변 픽셀의 위치를 보여준다.

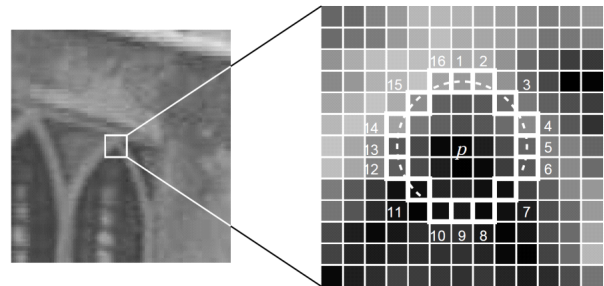


그림 3. 16개 이웃 픽셀의 위치^[4]
 Fig. 3. Location of 16 pixels of neighbor pixel^[4]

1.2 규칙 기반 특징점 매칭

특징점 검출은 과정 이후 찾은 특징점을 벡터로 저장하거나 표현하기 위해서 디스크립터를 생성하고, 이 디스크립터를 사용하여 특징점 매칭을 진행한다. SIFT를 기반^[2,3]으로 하는 이미지 매칭 기법은 디스크립터 측정 후, 측정된 디스크립터 사이의 유클리드 거리를 계산하여 가장 가까운 두 특징점을 매칭한다. SIFT의 디스크립터는 주변 영역을 16개의 구간으로 나누어 각 구간의 그래디언트를 집계하여 값을 생성하기 때문에 추출된 디스크립터의 크기는 128차원의 벡터이므로 연산량이 크다는 단점이 있다. 이를 보완하기 위해 Binary Robust Independent Elementary Features (BRIEF)^[5] 알고리즘이 고안됐다. BRIEF는 이미지의 디스크립터를 이진 특징 벡터로 기술하여 계산이 필요한 차원의 수를 크게 줄이는 알고리즘으로 SIFT 보다 빠른 연산이 가능하게 했다. Oriented FAST and Rotated BRIEF (ORB)^[6]는 특징점 추출 과정에 FAST 알고리즘을 사용하고 특징점 매칭 과정에 BRIEF 알고리즘을 사용하여 SIFT

보다 개선된 수행시간을 보였다. BRIEF는 간단하고 빠르게 계산될 수 있지만 특징점의 방향과 강도를 표현하는 능력은 SIFT보다 제한적이다. 또한 이러한 규칙 기반의 특징점 매칭은 결국 특징점의 디스크립터간의 관계를 파악하는 구조에서 그친다. 따라서 특징점을 기술하는 디스크립터가 표현하지 못하는 이미지 전체의 맥락은 파악할 수 없다는 단점을 가질 수밖에 없다.

2. 딥러닝 기반 이미지 매칭 기법

다양한 분야에서 딥러닝이 사용되고 발전함에 따라 이미지 매칭 분야 또한 딥러닝을 사용한 다양한 혁신적인 기법들이 등장하고 있다. 딥러닝은 규칙 기반의 특징점 매칭과 달리 많은 양의 데이터를 기반으로 학습하고, 높은 특징점 매칭 정확도와 일반화 성능을 제공한다. 본 절에서는 딥러닝을 사용한 특징점 추출 기법과 특징점 매칭 기법에 대하여 설명한다.

2.1 딥러닝 기반 특징점 추출

딥러닝 기법이 발달함에 따라 여러 가지 훈련 가능한 특징 추출 모델들이 등장하였다^[7,8]. 그 중 Learned Invariant Feature Transform (LIFT)^[7]는 Convolution Neural Network (CNN)을 사용하여 훈련이 가능하며 SIFT와 유사한 특징점 추출과 디스크립터 생성 기법을 제안했다. 그림 4는 LIFT의 동작 과정을 보여준다. LIFT는 다른 해상도에서 입력 이미지에 독립적으로 특징추출 CNN를 적용하여 스케일 공간에서 점수 피라미드를 얻는다. 그 후 SIFT^[2]와 유사한 Non-Maximum Suppression 방식을 사용하여 특징점 위치를 감지한 뒤 이를 훈련된 방향 추출 CNN 네트워크와 디스크립터 CNN을 사용하여 특징점의 디스크립터를 생성한다.

SuperPoint^[9]는 LIFT 이후 등장한 특징추출 모델이다. 이미지를 입력 받아 CNN을 통해 LIFT처럼 특징점 검출과 디스크립터 생성을 동시에 수행하지만, LIFT와 달리 이미지 크기의 제한이 없으며, 훈련 과정에서 레이블이 있는 데이터로 훈련을 진행한 뒤 레이블이 없는 이미지 쌍을 데이

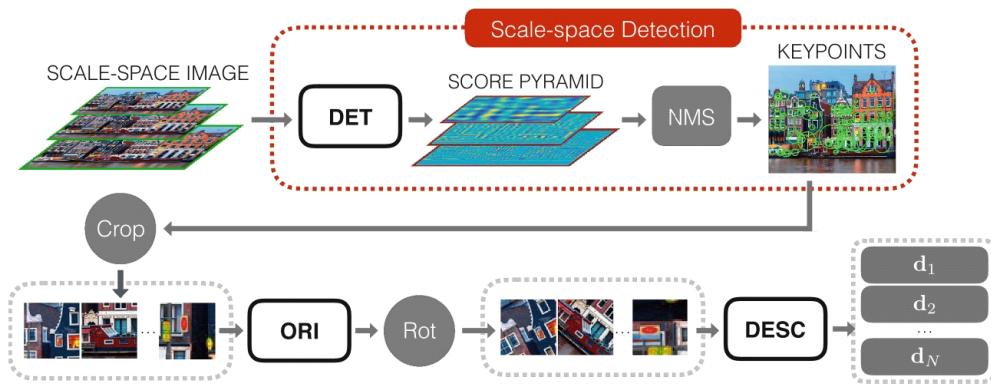


그림 4. Learned Invariant Feature Transform (LIFT)의 동작과정^[7]
 Fig. 4. Pipeline of Learned Invariant Feature Transform (LIFT)^[7]

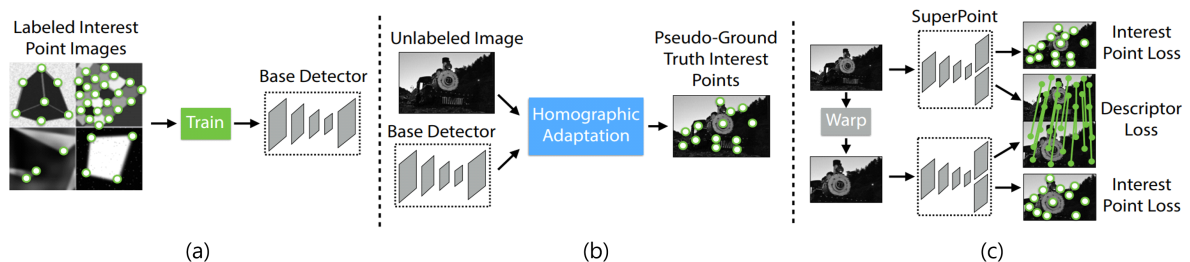


그림 5. SuperPoint의 자체 감독 훈련 개요 (a) 관심 지점 사전 학습 (b) 관심 지점 자체 레이블 지정 (c) 합동 훈련^[9]
 Fig. 5. Self-supervised training overview of SuperPoint (a) interest point pre-training (b) interest point self-labeling (c) joint training^[9]

터로 사용하여 자체 감독 훈련을 수행한다. 기존 딥러닝을 사용한 특징추출 모델은 훈련하기 위해 지도된 결과 값을 가진 훈련 데이터를 사용해야 했지만 SuperPoint는 자체 감독 훈련을 적용하여 이를 극복했다. 그림 5는 SuperPoint의 자체 감독 훈련 개요를 보여준다. SuperPoint는 레이블이 없는 데이터에 대해 기본적인 도형 이미지를 통해 훈련된 기존 특징추출 모델과 Homographic 적용 절차를 적용하여 레이블을 자동으로 지정한다. 이후 생성된 레이블을 통해 CNN를 훈련한다.

2.2 딥러닝을 사용한 특징점 매칭

딥러닝을 사용하여 특징점을 추출하는 방법뿐만 아니라 딥러닝을 사용하여 특징점을 매칭하는 방법 또한 연구되고 있다. 기존의 연구들은 두 이미지의 특징점의 디스크립터를 통해서만 관계를 파악하는 한계가 있었기 때문에 딥러닝을 사용하여 이러한 한계를 극복하려 했다. 특히 특징점 매칭 시 특징점의 디스크립터만을 고려하는 것이 아닌 특징점 사이의 관계도 고려하여 매칭 하는 연구가 진행되고 있다.

SuperGlue^[10]는 SuperPoint의 후속 연구로 Graph Neural Networks (GNN)을 사용하여 두 이미지 사이 특징점들 간 관계뿐만 아니라 같은 이미지에 있는 특징점의 관계 또한 고려하여 특징점 매칭을 수행한다. SuperGlue는 이미 추출된 특징점과 디스크립터를 입력데이터로 받아 매칭을 수행하며 추출된 특징점에 노드로 같은 이미지에 존재하는 특징점간의 엣지를 생성한 그래프인 self 그래프와 매칭 쌍 이미지에 존재하는 특징점간 엣지를 생성한 cross 그래프

를 두 개 생성하여 GNN을 적용한다. 이를 통해 SuperGlue는 본 이미지에서의 특징점의 맥락과 상대 이미지의 특징점과의 맥락을 전부 반영하여 이미지 매칭을 수행할 수 있다. 그림 6은 SuperGlue의 동작 과정을 보여준다.

SuperGlue는 추출된 특징점과 디스크립터를 입력으로 받는 기법이므로 추출된 특징점간의 관계를 고려하여 매칭을 수행하지만 특징점이 검출되지 않은 구역이나 이미지의 전체적인 맥락은 고려하지 못한다. 또한 다른 위치의 특징점이 같은 디스크립터를 가진다면 둘 사이의 차이를 구별하는 위치 정보를 고려하지 못한다. 이를 해결하기 위해 제안된 Detector-Free Local Feature Matching with Transformers (LoFTR)^[11]은 SuperGlue나 SIFT같은 기존 방법처럼 추출된 특징점을 매칭하는 것이 아니라 이미지에 CNN을 통해 추출된 feature map을 사용한 뒤 관계성을 판단하고 추후 coarse-to-fine module을 통해 특징점을 지정한다. 그림 7은 LoFTR의 동작 과정을 보여준다. LoFTR은 추출된 feature map을 같은 이미지의 특징 사이의 관계를 나타내는 self attention layer와 매칭 쌍 이미지의 특징 사이의 관계를 나타내는 cross attention layer를 생성한 뒤 디스크립터 없이 바로 이미지 매칭을 수행한다. 더불어 매칭된 Feature map에서 Coarse-to-Fine Module을 제안하여 상세히 특징점의 위치를 추정한다. LoFTR은 SuperGlue와 같이 self attention layer, cross attention layer를 통해 본 이미지 내부의 관계와 매칭 대상 이미지와의 관계를 고려하지만 GNN이 아닌 transformer^[12]를 사용하여 성능을 발전시키고 추출된 특징점이 아니라 feature map을 사용하여 특징의 위치적 정보를 고려할 수 있는 기법이다. 또한 LoFTR의 높은

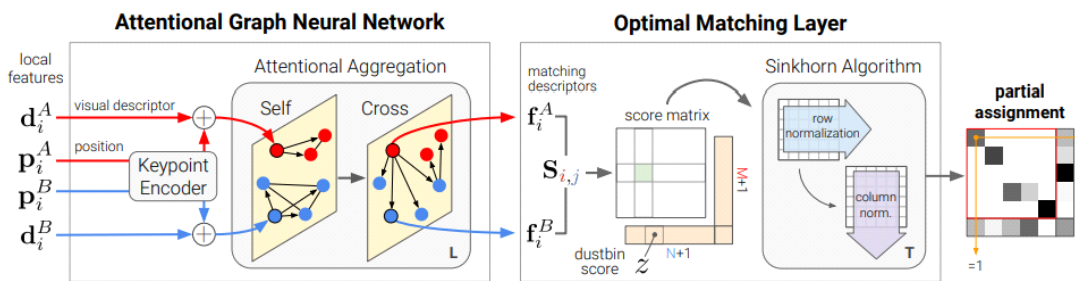


그림 6. SuperGlue의 동작과정^[10]
 Fig. 6. Pipeline of SuperGlue^[10]

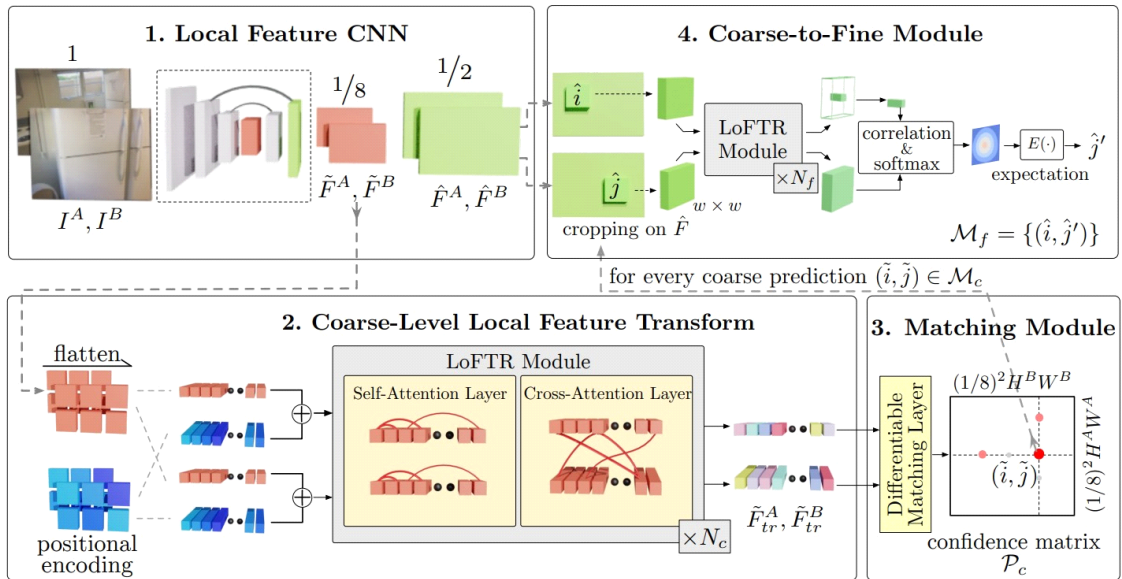


그림 7. LoFTR의 동작과정^[11]
 Fig. 7. Pipeline of LoFTR^[11]

성능 때문에 이후에도 LoFTR을 토대로 사용한 다양한 이미지 매칭 기법들이 등장하였다^[13,14].

III. 실험결과

본 장에서는 앞서 소개한 특징점 추출과 특징점 매칭을 합친 최신 기법들의 이미지 매칭 성능을 비교한다. 성능 지표의 경우 기존 논문에서 제공된 정확도 및 지표를 참고하였다.

1. 이미지 매칭 성능 비교 지표

이미지 특성 검출 성능 비교를 위하여 homographic 추정 방법을 사용했다. Homographic 행렬이란 한 평면에서 다른 평면으로의 변환을 나타내는 행렬이다. Homographic 행렬이 발견되면 원본 이미지를 표적 이미지에 투영하는 데 사용할 수 있다. 이후 Area Under the Curve (AUC)를 활용하여 성능 평가를 진행하는데, AUC는 표적 이미지의 특징점의 특정 거리 임계값 내에 있는 원본 이미지의 특징점의 비율을 계산하여 계산된다. AUC가 높을수록 homographic



그림 8. HPatches 데이터 셋의 예시^[15]
 Fig. 8. Examples of HPatches dataset^[15]

행렬이 원본 이미지의 특징점을 표적 이미지에 더 잘 투영한다는 것을 나타낸다. 두 이미지의 이미지 매칭을 수행하고 매칭된 특징점들의 위치를 통해 homographic 행렬을 계

산하여 AUC를 나타낸다.

일반적으로 이미지 매칭 기법을 평가하기 위한 데이터 셋은 HPatches^[15], ScanNet^[16], YFCC100M^[17]이 있다. 실험은 이중 실험은 HPatches를 사용하여 진행했다. 그림 8은 HPatches 데이터 셋의 예시를 보여준다. HPatches는 한 사물을 각기 다른 각도와 조명에서 찍은 사진의 모음이다. 116개의 장면이 696개의 이미지로 구성되어 있고 57개의 장면은 조명의 큰 변화를 가지며 59개의 장면은 시점의 큰 변화를 가진다.

2. 이미지 매칭 성능 비교

표 1은 특징점 추출 방법에 따른 homography 추정 실험 결과를 보여준다. ϵ 는 임계값으로 매칭된 특징점의 투영 후 오차범위를 의미한다. 실험을 위해 SuperPoint^[9], LIFT^[7],

표 1. 특징점 추출 방법에 따른 HPatches의 homography 추정 결과^[9]
 Table 1. Result of HPatches homography estimation according to feature extraction method^[9]

Homography estimation (AUC)			
	$\epsilon = 1$	$\epsilon = 2$	$\epsilon = 3$
SuperPoint	.310	.684	.829
LIFT	.284	.598	.717
SIFT	.424	.676	.759
ORB	.150	.395	.538

SIFT^[2], ORB^[6]를 평가했다. 모든 이미지는 480x640 해상도에서 검출된 특징점 1000개를 사용하였으며 특징점 매칭 방법은 생성된 디스크립터의 가장 가까운 유클리드 거리를 가진 기준으로 계산했다. ϵ 가 1일 때를 제외하고 딥러닝 기반의 기법들이 더 좋은 성능을 제공하는 것을 확인할 수 있다.

표 2. 특징점 매칭에 따른 HPatches의 homography 추정 결과^[11]
 Table 2. Result of HPatches homography estimation according to feature matching method^[11]

Homography estimation (AUC)			
	$\epsilon = 3$	$\epsilon = 5$	$\epsilon = 10$
SP+SuperGlue	.539	.683	.817
LoFTR	.659	.756	.846

표 2는 특징점 매칭 방법에 따른 homography 추정 실험 결과를 보여준다. 실험에서는 SuperPoint와 SuperGlue를 같이 사용한 모델과 LoFTR을 평가했다. 실험결과에서 볼 수 있듯이 모든 조건에서 LoFTR이 SuperGlue보다 더 뛰어난 성능을 제공한다.

그림 10은 이미지 매칭 방법에 따른 실험결과를 보여준다. 실험에 사용된 이미지는 같은 장소에서 피사체 위치에 변화를 두어 촬영한 이미지이다. 실험에서는 네 가지의 다른 기법을 적용하여 이미지 매칭을 실행하였다. 규칙 기반

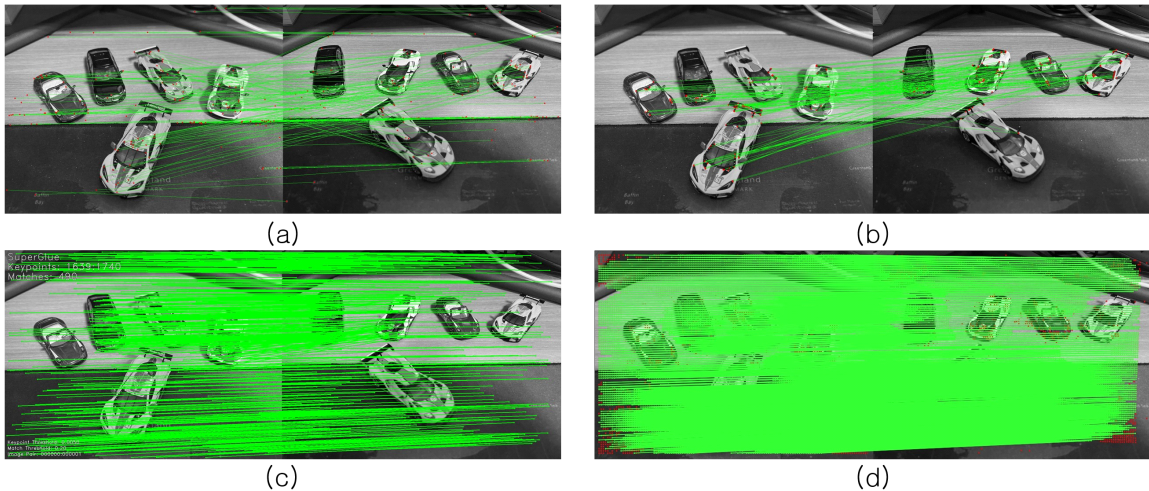


그림 9. 이미지 매칭 기법에 따른 결과 (a) SIFT^[2] (b) ORB^[6] (c) SP + SuperGlue^[9,10] (d) LoFTR^[11]
 Fig. 9. Results according to image matching method (a) SIFT^[2] (b) ORB^[6] (c) SP + SuperGlue^[9,10] (d) LoFTR^[11]

의 기법으로 SIFT를 디스크립터의 유클리디언 거리를 통해 매칭한 기법, ORB를 사용하였고 딥러닝을 사용한 기법으로 SuperPoint와 SuperGlue를 같이 사용한 모델과 LoFTR을 사용했다.

실험결과에서 SIFT와 ORB는 이미지의 추출된 특징점의 위치는 코너나 엣지가 있는 부분으로 한정되어 있으며 특히 ORB의 경우 SIFT에 비해 상대적으로 흐릿한 영역의 특징점을 추출하는 능력이 부족하다는 것을 확인할 수 있다. 또한 딥러닝을 사용한 기법이 그렇지 않은 기법보다 많은 특징점을 매칭 시키는 것을 실험 결과에서 확인할 수 있다. 특히, LoFTR의 경우 코너나 엣지가 아닌 평탄한 부분에서도 특징점을 매칭 시킬 수 있는 성능을 보여주고 있으며 특징점 간 공간적인 정보를 더 잘 유지하는 것을 확인할 수 있다.

표 1, 표 2에서의 결과를 통해 각도와 조명을 다르게 찍은 이미지인 HPpatches 데이터 셋에서 딥러닝 기반의 기법들이 더 좋은 성능을 보여주었다. 하지만 그림 10에서의 결과를 비교해 보면 두 이미지 사이에서 이동하는 물체의 경우 딥러닝 기반의 매칭된 특징점들이 물체의 이동을 제대로 표현하지 못하는 경향이 있으며 오히려 규칙 기반의 특징점 매칭 방법들이 그 관계를 더 효과적으로 표현하고 있다. 이는 이미지 내부의 특징점들의 관계와 맥락을 고려하는 딥러닝 기반의 모델이 이미지 내부의 맥락을 변화시키는 물체 이동을 매칭 결과에 반영하는 데 어려움을 가진다고 해석할 수 있다. 또한 규칙 기반의 기법들이 특징점들 사이의 관계를 직접적으로 활용하기 때문에 물체 이동과 같은 지역적인 변화를 포착하는 데 있어서는 더 유용한 역할을 수행할 수 있다.

IV. 결론 및 향후 연구 방향

본 논문에서는 특징 추출을 통한 이미지 매칭 기법에 대해서 특징 점 추출과 특징 점 매칭의 두 단계로 나누어 각 절차를 설명하고 널리 알려진 이미지 매칭 기법에 대해 살펴보았다. 이후 실험을 통해 각 기법의 성능을 평가하고 이미지 매칭의 정성적 결과를 논의하였다. 특징 추출을 통한 homography 추정에서는 SuperPoint가 가장 우수한 성적을

보여주었으며. 특징점 매칭 기법 중에서는 LoFTR이 가장 우수한 매칭 성능을 보여주었다. 또한 딥러닝을 사용한 기법들이 규칙기반의 기법보다 더 많은 특징점을 정확히 매칭 시켰으며 LoFTR의 경우 특징점 간의 공간적 정보를 보존하여 매칭이 어려운 영역에서도 우수한 성능을 보여주었다.

이미지 매칭은 두 이미지 안에서 특징점 간 관계, 전체적인 이미지에서 특징점의 맥락 등을 고려해야 하는 복잡하고 까다로운 작업이다. 딥러닝 기술의 발전과 함께 기존 기법들의 단점을 해결하고 다양한 이미지의 특징을 고려하는 새로운 기법들이 등장하고 있다. 그러나 논문에 제시된 기존의 기법들은 각 프로젝트의 목적과 매칭 이미지의 특성에 따라 제약사항이 발생할 수 있다. 따라서 이러한 목적과 특성을 고려하여 기존 기법의 단점을 분석하고 극복할 수 있는 새로운 이미지 매칭 방법에 대한 연구가 진행되어야 한다.

참고 문헌 (References)

- [1] C. Harris and M.J. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147 - 152, 1988. doi: <https://doi.org/10.5244/c.2.23>
- [2] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91-110, 2004. doi: <https://doi.org/10.1023/b:visi.0000029664.99615.94>
- [3] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features." *Lecture notes in computer science* 3951, 20 404-417, 2006. doi: https://doi.org/10.1007/11744023_32
- [4] Rosten, Edward, and Tom Drummond. "Machine learning for high-speed corner detection." *Computer Vision - ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9*. Springer Berlin Heidelberg, 2006. doi: https://doi.org/10.1007/11744023_34
- [5] Calonder, Michael, et al. "Brief: Binary robust independent elementary features." *Computer Vision - ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11*. Springer Berlin Heidelberg, 2010. doi: https://doi.org/10.1007/978-3-642-15561-1_56
- [6] Rublee, Ethan, et al. "ORB: An efficient alternative to SIFT or SURF." *2011 International conference on computer vision*. Ieee, 2011. doi: <https://doi.org/10.1109/iccv.2011.6126544>
- [7] Yi, Kwang Moo, et al. "Lift: Learned invariant feature transform." *Computer Vision - ECCV 2016: 14th European Conference,*

- Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14. Springer International Publishing, 2016.
doi: https://doi.org/10.1007/978-3-319-46466-4_28
- [8] Verdie, Yannick, et al. "Tilde: A temporally invariant learned detector." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
doi: <https://doi.org/10.1109/cvpr.2015.7299165>
- [9] DeTone, Daniel, Tomasz Malisiewicz, and Andrew Rabinovich. "Superpoint: Self-supervised interest point detection and description." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018.
doi: <https://doi.org/10.1109/cvprw.2018.00060>
- [10] Sarlin, Paul-Edouard, et al. "Superglue: Learning feature matching with graph neural networks." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
doi: <https://doi.org/10.1109/cvpr42600.2020.00499>
- [11] Sun, Jiaming, et al. "LoFTR: Detector-free local feature matching with transformers." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.
doi: <https://doi.org/10.1109/CVPR46437.2021.00881>
- [12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. NeurIPS, 2017.
doi: <https://doi.org/10.48550/arXiv.1706.03762>
- [13] SHEN, Zehong, et al. Semi-Dense Feature Matching With Transformers and its Applications in Multiple-View Geometry. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
doi: <https://doi.org/10.1109/TPAMI.2022.3223530>
- [14] Tian, Liang. "Matching Images from Different Viewpoints with Deep Learning Based on LoFTR and MAGSAC++." Proceedings of the 2023 5th International Conference on Image Processing and Machine Vision. 2023.
doi: <https://doi.org/10.1145/3582177.3582181>
- [15] Balntas, Vassileios, et al. "HPatches: A benchmark and evaluation of handcrafted and learned local descriptors." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
doi: <https://doi.org/10.1109/cvpr.2017.410>
- [16] Dai, Angela, et al. "Scannet: Richly-annotated 3d reconstructions of indoor scenes." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
doi: <https://doi.org/10.1109/cvpr.2017.261>
- [17] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. YFCC100M: The new data in multimedia research. Communications of the ACM, 59(2):64 - 73, 2016.
doi: <https://doi.org/10.1145/2812802>

저 자 소 개



박 승 기

- 2023년 : 동국대학교 멀티미디어공학과 학사
- 2023년 ~ 현재 : 동국대학교 멀티미디어공학과 석사
- ORCID : <https://orcid.org/0009-0003-3215-2260>
- 주관심분야 : 준 지도 학습, 항공사진을 이용한 변화감지



채 문 주

- 2022년 : 동국대학교 멀티미디어공학과 학사
- 2022년 ~ 현재 : 동국대학교 멀티미디어공학과 석사
- ORCID : <https://orcid.org/0009-0006-7195-1454>
- 주관심분야 : 3D데이터에서의 휴먼 제스처 인식

저 자 소 개



조 성 인

- 2010년 : 서강대학교 전자공학과 학사
- 2015년 : 포항공과대학교 전자전기공학부 박사
- 2017년 : LG 디스플레이 선임연구원
- 2019년 : 대구대학교 전자전기공학부 조교수
- 2019년~ 현재 : 동국대학교 AI소프트웨어융합학부 부교수
- ORCID : <https://orcid.org/0000-0003-4251-7131>
- 주관심분야 : 영상처리, 컴퓨터 비전, 딥러닝