특집논문 (Special Paper)
방송공학회논문지 제24권 제2호, 2019년 3월 (JBE Vol. 24, No. 2, March 2019)
https://doi.org/10.5909/JBE.2019.24.2.234
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

딥 러닝 기반의 SIFT 이미지 특징 추출

이 재 은[a)], 문 원 준[a)], 서 영 호[a)], 김 동 욱[a)‡]

SIFT Image Feature Extraction based on Deep Learning

Jae-Eun Lee[a)], Won-Jun Moon[a)], Young-Ho Seo[a)], and Dong-Wook Kim[a)‡]

요 약

본 논문에서는 일정 크기로 자른 영상의 가운데 픽셀이 SIFT 특징점인지를 판별함으로써 SIFT 특징점을 추출하는 딥 뉴럴 네트워크(Deep Neural Network)를 제안한다. 이 네트워크의 데이터 세트는 DIV2K 데이터 세트를 33×33 크기로 잘라서 구성하고, 흑백 영상으로 판별하는 SIFT와는 달리 RGB 영상을 사용한다. 그라운드 트루스(ground truth)는 옥타브(scale, octave)를 0, 시그마(sigma)는 1.6, 간격(intervals)은 3으로 설정하여 추출한 RobHess SIFT 특징들로 구성한다. VGG-16을 기반으로 컨볼루션 층을 13개에서 23개와 33개로 점점 깊은 네트워크를 구성하고, 영상의 스케일을 증가시키는 방법을 바꿔가며 실험을 수행한다. 출력 층의 활성화 함수로 시그모이드(sigmoid) 함수를 사용한 결과와 소프트맥스(softmax) 함수를 사용한 결과를 비교하여 분석한다. 실험결과 제안한 네트워크가 99% 이상의 추출 정확도를 가질 뿐 아니라 왜곡된 영상에 대해서도 높은 추출 반복성을 가진다는 것을 보인다.

Abstract

In this paper, we propose a deep neural network which extracts SIFT feature points by determining whether the center pixel of a cropped image is a SIFT feature point. The data set of this network consists of a DIV2K dataset cut into 33×33 size and uses RGB image unlike SIFT which uses black and white image. The ground truth consists of the RobHess SIFT features extracted by setting the octave (scale) to 0, the sigma to 1.6, and the intervals to 3. Based on the VGG-16, we construct an increasingly deep network of 13 to 23 and 33 convolution layers, and experiment with changing the method of increasing the image scale. The result of using the sigmoid function as the activation function of the output layer is compared with the result using the softmax function. Experimental results show that the proposed network not only has more than 99% extraction accuracy but also has high extraction repeatability for distorted images.

Keyword : SIFT Feature extraction, Deep learning, VGG, CNN(Convolutional Neural Network), Repeatability

a) 광운대학교 전자재료공학과(Department of Electronic Materials Engineering, Kwangwoon University)
‡ Corresponding Author : 김동욱(Dong-Wook Kim)
E-mail: dwkim@kw.ac.kr
Tel: +82-2-940-5167
ORCID: https://orcid.org/0000-0002-4668-743X

※ 이 논문의 연구 결과 중 일부는 “2018년 한국방송·미디어공학회 추계학술대회”에서 발표한 바 있음.
※ This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2016R1D1A1B03930691).
※ The present Research has been conducted by the Research Grant of Kwangwoon University in 2019.
· Manuscript received January 8, 2019; Revised March 6, 2019; Accepted March 8, 2019.

Ⅰ. 서 론

특징점 추출(feature extraction)은 객체 인식, 얼굴 인식, 영상 매칭과 같은 컴퓨터 비전의 기술의 하나로 연구되고 있다. 이때 영상의 형태나 크기, 카메라의 시점이나 조명 변화에 영향을 받지 않는, 변형에 강인한 점을 찾는 것이 중요하다. 특징점을 찾는 가장 초기 방법은 영상에서 코너점을 찾는 해리스 코너(Harris corner) 검출기이었다[1]. 코너점으로는 여러 방향에 대하여 값의 변화량을 비교하여 변화량이 큰 점을 이용한다. 이 방법은 영상의 스케일(scale) 변화에 강인하지 못하다는 약점이 있다. 이를 보완하는 방법으로는 Mikolajczyk의 해리스 라플라시안(Laplacian) 방법이 있다[2]. 이 방법은 여러 스케일에서 해리스 코너점을 찾은 뒤, 스케일 변화에 대해 극대인 점을 검출하는 방식이다. Shi와 Tomasi는 affine 변화까지 고려한 Shi& Tomasi 코너를 제안하였다[3]. 그러나 가장 많이 알려진 특징점 추출방법으로는 Lowe의 SIFT(Scale Invariant Feature Transform)이다[4]. SIFT는 원 영상에서뿐만 아니라 스케일 변화를 가한 영상에서 DoG(Difference of Gaussian)를 계산하여 변화량이 극대인 점을 찾는다. SIFT의 방대한 연산량 때문에 이를 해결하기 위한 시도로 여러 방법이 제안되었다. Bay의 SURF(Speed Up Robust Feature)[5]와 Rosten의 FAST(Features from Accelerated Segment Test)[6], 그리고 Mair의 AGAST가 있다[7]. 이 뿐만 아니라, 영상의 스티칭(stitching) 측면에서 스티칭 효율에 미치는 SIFT 특징점 추출 파라미터들을 분석하여 스티칭 효율저하를 최소로 하고 특징점 추출 연산시간을 줄일 수 있는 파라미터 값들을 제시한 연구도 발표되었다[8]. 또한 Rublee는 특징점 검출하고자 하는 알고리즘인 FAST와 특징점 표현자(feature descriptor) 생성 알고리즘인 BRIEF를 일부 수정하여 만든 ORB(Oriented FAST and Rotated BRIEF) 알고리즘을 제안하였다[9]. 특징점 추출 알고리즘이 오랜 기간 동안 연구되었음에도 불구하고 SIFT의 성능보다 확연하게 우세한 알고리즘이 아직 발표되지 않았다.

한편, 컴퓨터 비전 분야에서 최고수준의 성능을 내어 주목받고 있는 딥 러닝(deep learning)은 기계가 자동으로 데이터에서 중요한 패턴과 규칙을 학습하여 의사결정 또는 예측 등을 수행하는 기술을 말한다. GPU(Graphic Processing Unit)와 같은 하드웨어 발전으로 방대한 양의 연산을 짧은 시간에 수행할 수 있게 되었고, 빅 데이터(big data)로 분석 가능한 데이터의 양이 많아져 딥 러닝 연구는 활발히 진행되고 있다. 본 논문에서는 딥 러닝을 기반으로 SIFT 특징점을 추출하는 방법과 그 발전 가능성을 제안한다. 딥 러닝 모델은 VGG-16을 기반으로 네트워크 구조와 깊이를 변화시켜 그 결과를 비교, 분석하며 제안하는 네트워크가 높은 추출 정확도를 갖는다는 것을 실험을 통해 보인다. 또한 제안하는 네트워크의 특징점 추출 반복성(repeatability)을 보이기 위해 영상의 밝기와 흐림 정도를 변화하여 생성한 변형 영상에 대해 특징점을 추출하여 원 특징점과 비교함으로써 제안하는 네트워크가 높은 추출 반복성을 갖고 있음을 보인다.

본 논문은 다음과 같이 구성된다. Ⅱ장에서는 SIFT 알고리즘의 특징점 추출 방법을 간략히 설명한다. Ⅲ장에서는 SIFT 이미지 특징점 추출을 위한 딥 뉴럴 네트워크(Deep Neural Network, DNN)와 이 네트워크의 학습과 검증을 위한 데이터 세트를 구성한 방법을 설명한다. Ⅳ장에서는 네트워크 구조 변화에 따른 실험 결과를 비교·분석하고, 밝기 변화와 흐림 정도를 변화한 영상에 대한 추출 반복성 실험을 기존의 SIFT 알고리즘과 비교한다. 그 결과를 토대로 Ⅴ장에서 본 논문의 결론을 맺고 향후 연구 방향을 제시한다.

Ⅱ. SIFT

특징점을 추출하는 알고리즘 중 현재 가장 널리 사용되는 SIFT 알고리즘은 RGB 영상을 흑백 영상으로 변환한 후, 옥타브와 간격(intervals)에 따라 영상의 크기를 확대, 축소 및 블러링을 수행하여 scale space(영상의 크기를 변화시킨 여러 스케일의 영상들을 모아놓은 집합체)를 만든다. 블러링을 수행할 때 가우시안(Gaussian) 필터를 사용하는데, 분산(σ)을 파라미터로 블러링의 정도를 결정하며, σ는 옥타브에 의해 간접적으로 결정된다. σ가 클수록 영상을 강하게 블러링시키는데, 블러링이 강하게 된 영상은 스케일이 커지는 것으로 볼 수 있다. 따라서 옥타브를 스케일 벡터라고도 부른다. 식 (1)의 I는 영상, G는 옥타브에 따른 가우시안 필터(식 (2)), L은 결과 영상을 각각 뜻한다. 여기서

(x, y)는 영상에서의 화소위치, $*$는 컨볼루션 연산을 각각 뜻한다.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2)$$

가우시안 필터링 결과 영상에 DoG를 적용하여 DoG 피라미드를 형성하는데, 이 때 같은 옥타브 내에서 인접한 두 개의 영상간 DoG만을 생성한다. 이 과정은 그림 1에 나타내었다. 그 다음, 각 스케일의 DoG 영상 그룹에서 각 화소가 극대값 또는 극소값인지를 판별한다. 즉, 특정 DoG 영상 내의 한 DoG 화소에 대해, 주위의 8 화소, 한 단계 높고 낮은 스케일의 9 화소씩, 총 26 화소들과 비교하여 가장 크

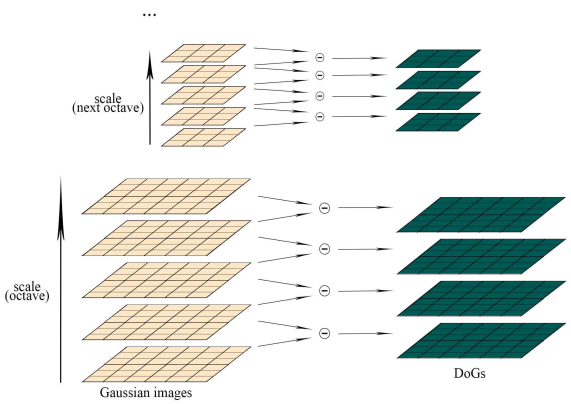


그림 1. SIFT DoG 피라미드 생성 과정
Fig. 1. The process to form a SIFT DoG pyramid

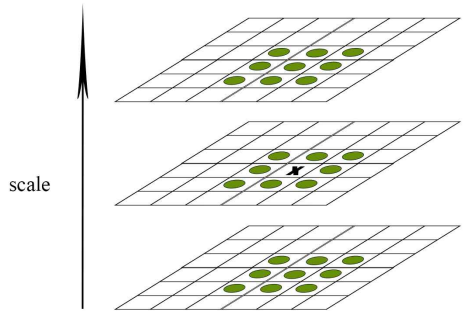


그림 2. SIFT 극점 추출 과정
Fig. 2. SIFT extrema extraction process

거나 가장 작은 값을 가지면 그 화소는 특징점 후보가 된다. 이 과정은 그림 2에 나타낸다. 마지막으로, 이 후보들 중 낮은 콘트라스트(contrast)를 갖거나 에지(edge)에 존재하는 후보들을 제거하여 최종 특징들을 결정한다.

SIFT는 영상에서 특정 화소가 주변 화소들에 비해 변화량이 많은 점들을 특징점으로 추출한다. 그러나 이 점들에서 변화량이 적거나 에지인 경우를 제외시킴으로써 특징 성질이 두드러지지 않거나 3차원 영상 등에서와 같이 시점변화로 사라질 수 있는 화소들은 특징점에서 제외시켰다. 따라서 SIFT 특징점은 다양한 영상처리 분야에서 가장 널리 사용되고 있으며, 특히 가상현실(Virtual Reality, VR) 영상생성을 위한 영상 스티칭(stitching)에 가장 많이 사용되고 있다.

그러나 SIFT는 컨볼루션 연산을 과다히 사용하기 때문에 연산량이 너무 많아 실시간 처리가 어렵고, 영상에 왜곡이 발생한 경우 추출 반복성이 다소 떨어진다는 단점을 갖고 있다. 반면, 딥 러닝을 기반으로 하는 DNN의 경우 훈련시간이 많이 걸리지만 훈련을 마친 DNN은 처리속도가 빠르기 때문에 실시간성에서 매우 유리하다. 뿐만 아니라 DNN은 영상처리 분야에서 뛰어난 성능을 보이고 있기 때문에 네트워크가 갖고 있는 가중치들을 충분히 활용한다면 추출 반복성에 있어서도 충분한 성능을 보일 것으로 기대할 수 있다. 이에 본 논문에서는 영상처리에 높은 성능을 보이는 컨볼루션-기반의 DNN의 성능을 개선하는 방안을 제안하여 높은 SIFT 특징점 추출률과 추출 반복성을 갖고 있음을 보이고자 한다.

Ⅲ. 제안하는 SIFT 특징점 추출 네트워크

본 논문에서 제안하는 특징점 추출 DNN(feature extraction DNN)은 입력 영상의 가운데 픽셀이 SIFT 특징점인지 아닌지를 판단한다. SIFT 특징점을 추출해 데이터 세트를 구성하는 방법을 먼저 설명하고, 그 다음에 DNN의 구조를 소개한다.

1. 데이터 세트

풍경과 건물이 비교적 많고 영상 크기가 너무 크지 않은

DIV2K 데이터 세트에서 저해상도 데이터를 사용하였다[10]. 훈련 데이터 세트 800장을 RobHess SIFT 알고리즘으로 특징점을 추출하여 label을 구성하였다[11]. SIFT 옥타브는 0, σ는 1.6, interval은 3으로 설정하여 원본 영상 크기에 대한 특징만 추출하였다. 그림 3은 설정한 하이퍼-파라미터(hyper-parameter)들을 기반으로 추출한 SIFT 특징들을 연두색 점으로 나타내었다. 흑백 영상으로 특징들을 추출하는 SIFT 알고리즘과 달리 본 논문에서는 RGB 영상을 사용하였고, 간격을 한 픽셀로 두어 33×33씩 잘라내어 사용하였다. 따라서 영상의 가장자리 16개의 픽셀들은 DNN에서 특징점 여부를 판별할 수가 없다. 훈련 데이터는 특징점인 영상을 215k개, 특징점이 아닌 영상을 2,152k개로 1:10의 비율을 두어 총 데이터 수는 2M개를 사용하였다. 변화에 강인한 특징점을 추출하기 위해 영상을 좌·우, 상·하 반전시켜 학습을 진행하였다. 검증(test) 데이터도 훈련 데이터와 마찬가지로 특징점인 영상과 특징점이 아닌 영상에 대한 비율을 1:10로 하고 좌·우, 상·하 반전시켜 학습의 진행 정도를 확인하였다. 그리고 영상의 밝기, 흐림 정도를 조절하여 원본 영상을 변형시켜 평가를 진행하였다.

그림 3. 옥타브를 0으로 설정하여 추출한 SIFT 특징들
Fig. 3. SIFT features extracted with octave set to 0

2. 네트워크 구조

2014년 이미지넷 챌린지(ImageNet challenge)에서 우수한 성능을 보여준 VGG-16을 기반으로 DNN를 구성하였다[12]. 5×5 필터로 컨볼루션 연산을 한 번 진행하는 것과 3×3 필터로 컨볼루션 연산을 두 번 진행하는 것을 비교하였을 때, 두 연산의 성능은 비슷하지만 후자의 연산량이 적다. VGG-16은 이 점에 착안하여 필터의 크기를 작게 하는 대신 층을 깊게 하여 학습 효율을 증가시킨 네트워크이다.

표 1에 본 논문에서 사용하는 모든 네트워크의 구성을 보이고 있다. 모든 네트워크의 입력 영상은 33×33 영상이며, RGB 정보를 모두 포함하여 3개의 채널(33×33×3)로 구성된다. 표 1의 A는 VGG-16의 기본 네트워크 구조를 나타내고 있다. 이 네트워크는 컨볼루션 계층(Conv)이 13개, 최대 풀링(max pooling)이 5개 그리고 전결합(fully connected, FC)층이 3개로 구성된다. 컨볼루션 계층은 모두 3×3 필터 크기를 갖고, 간격(stride)은 1×1로 하였다. 3×3 컨볼루션을 수행하기 때문에 각 영상의 네 변에 한 화소의 0-패딩(0 padding)을 진행하여 컨볼루션을 수행해도 영상의 크기는 유지되지만, 2×2 최대 풀링을 진행하면 영상의 높이, 너비가 반으로 감소한다. 컨볼루션을 2, 2, 3, 3, 3번 진행할 때마다 2×2 최대 풀링이 진행되며, 채널 수는 64, 128, 256, 512, 512로 점점 증가한다. 네트워크 출력부는 3개의 전결합층으로 되어있다. 2개의 전결합층 크기는 512로 유지되고 특징점 판단을 위해 마지막 출력층의 크기는 1×1×1로 설정한다. 마지막 층을 제외한 전결합층의 활성화(activation) 함수는 leaky ReLU 함수를 사용하고, 마지막 층에서만 0과 1사이 값으로 출력하는 시그모이드(sigmoid) 함수를 사용한다. 시그모이드 함수 출력이 특징점일 확률로 간주하여 0.5 이상이면 특징점, 0.5미만이면 특징점이 아니라고 판단한다.

표 1에서 이전 층이 'ConvA-B'이고 현재 층이 'ConvC-D'일 때 현재 층에서 C×C×B 필터를 D개 사용하여 D 채널(각 필터가 하나의 채널 생성)의 동일한 크기의 영상을 생성한다. 'FC-X'층은 X개의 노드(node)를 갖고 이전 층의 모든 노드와 현재 층의 모든 노드가 가중치로 연결된 층을 의미한다.

표 1의 B와 C는 각각 A에 컨볼루션 계층을 10층, 20층 추가한 구조이다. D는 C에서 최대 풀링을 제거한 대신 6, 12, 19, 26, 33번째 컨볼루션층의 필터 간격을 2×2로 두어 최대 풀링의 역할인 영상의 스케일링을 수행하도록 하였다.

표 1. 제안한 DNN 구성
Table 1. Configurations of the proposed DNNs

A	B	C	D	E
16 weight layers	26 weight layers	36 weight layers	36 weight layers	36 weight layers
input (33×33 RGB image)				
conv3-64 conv3-64	conv3-64 conv3-64 conv3-64 conv3-64	conv3-64 conv3-64 conv3-64 conv3-64 conv3-64 conv3-64	conv3-64 conv3-64 conv3-64 conv3-64 conv3-64 conv3-64	conv3-64 conv3-64 conv3-64 conv3-64 conv3-64 conv3-64
maxpooling			stride : 2	
conv3-128 conv3-128	conv3-128 conv3-128 conv3-128 conv3-128	conv3-128 conv3-128 conv3-128 conv3-128 conv3-128 conv3-128	conv3-128 conv3-128 conv3-128 conv3-128 conv3-128 conv3-128	conv3-128 conv3-128 conv3-128 conv3-128 conv3-128 conv3-128
maxpooling			stride : 2	
conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256 conv3-256 conv3-256
maxpooling			stride : 2	
conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512 conv3-512 conv3-512
maxpooling			stride : 2	
conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512 conv3-512 conv3-512
maxpooling			stride : 2	
FC-512				
FC-512				
FC-512				
sigmoid				softmax

E는 A~D와 다르게 출력에 소프트맥스(softmax) 함수를 사용한 것이다. 즉, 출력에 시그모이드 함수를 사용하여 특징점일 확률을 계산한 A~D와 달리 E는 출력의 크기를 2로 두고 원-핫-인코딩(one-hot-encoding) 방식으로 01이면 특징점, 10이면 특징점이 아닌 것으로 판단하도록 하였다.

Ⅳ. 실험 결과

본 연구의 구현 환경은 파이썬(Python)이고, 사용한 PC는 Intel(R) Xeon(R) CPU E3-1275 v6 @3.80GHz, 64GB RAM을 갖고 있으며, 운영체계는 64-bit Windows, 그리고 GPU는 GTX 1080ti를 사용하였다. 이 PC로 DNN을 훈련

하는 데에는 최소 1주일에서 최대 2주일의 시간이 소요되었다. 훈련에서 미니배치(mini-batch)는 500으로 하였고, 300 에폭(epoch)까지 수행하였다.

표 2는 표 1의 DNN 구조에 따른 결과를 보이고 있다. DNN의 평균 훈련 데이터 정확도는 99.4%이었고, 가장 좋은 테스트 데이터 정확도는 99.109%이었다. C, D, E의 훈련 데이터 정확도는 훈련을 진행할수록 점점 증가하여 거의 100%에 도달하였지만, 테스트 데이터의 정확도는 일정 수준에서 더 이상 증가하지 않았다. A, B, C의 결과를 보면, 컨볼루션 층을 증가시킬수록 네트워크의 성능은 점점 좋아지는 것을 볼 수 있다. 특히, 테스트 데이터 정확도가 크게 증가한 것을 볼 수 있다. 그러나 시그모이드 대신 소프트맥스의 활성화 함수를 사용한 E의 검증 데이터 정확도는 오히려 감소하였다.

표 2. 제안한 DNN의 실험결과
Table 2. Experimental results for the proposed DNN

DNN	Train accuracy(%)	Test accuracy(%)
A	98.100	96.796
B	99.300	99.002
C	99.900	99.083
D	99.900	99.109
E	99.900	99.084

제안한 DNN의 또 다른 성능을 평가하기 위해 반복성 측정을 진행하였다. 반복성($r_{1,2}$) 측정은 식 (3)에서와 같이 원본 영상 I_1과 왜곡된 영상 I_2에서 추출된 특징점들 사이에 동일한 점의 개수 $C(I_1, I_2)$와 각 영상에서 추출된 특징점 개수 N_1, N_2의 최솟값($\min(N_1, N_2)$)의 비를 나타낸 수치이며, 이것은 왜곡에 대해 DNN이 얼마나 강인한지를 나타내는 척도이다. 따라서 왜곡된 영상에서 추출된 특징점들이 원본 영상에서 추출된 특징점들과 모두 동일하다면 반복성 1을 나타내며 이상적인 특징점 추출 방법이라고 할 수 있다[13].

$$r_{1,2} = \frac{C(I_1, I_2)}{\min(N_1, N_2)}, 0 \le r_{1,2} \le 1 \quad (3)$$

실험을 진행할 때 더 높은 신뢰도를 위해 검증 데이터의

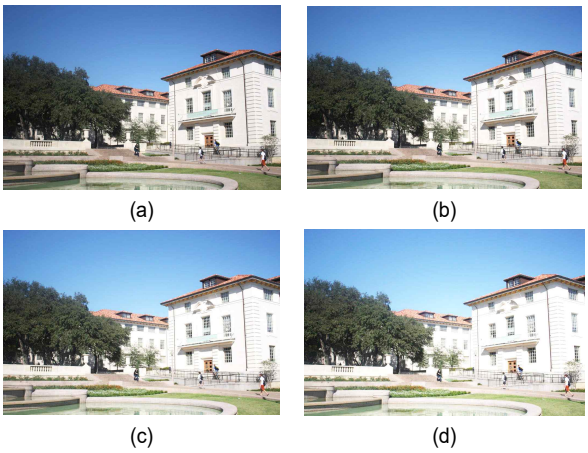


그림 4. 밝기 정도를 변화시킨 영상의 예; (a) +25, (b) +50, (c) +75, (d) +100
Fig. 4. Image examples with varying brightness level; (a) +25, (b) +50, (c) +75, (d) +100

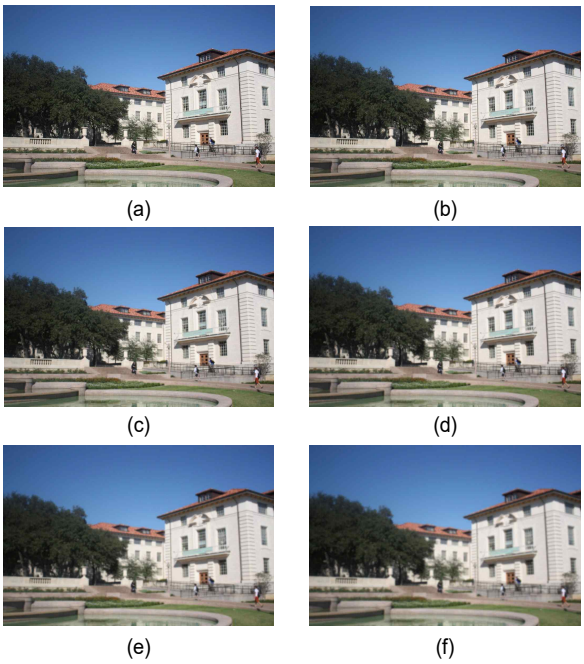
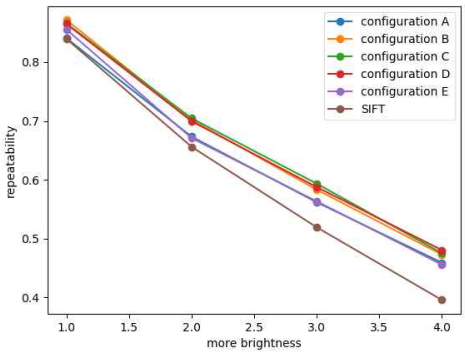


그림 5. 흐림 정도를 변화시킨 영상의 예; (a) 원본, (b) 반경 0.5, (c) 반경 1.0, (d) 반경 1.5, (e) 반경 2.0, (f) 반경 2.5
Fig. 5. Example Images with varying blur level; (a) original, (b) radius 0.5, (c) radius 1.0, (d) radius 1.5, (e) radius 2.0, (f) radius 2.5

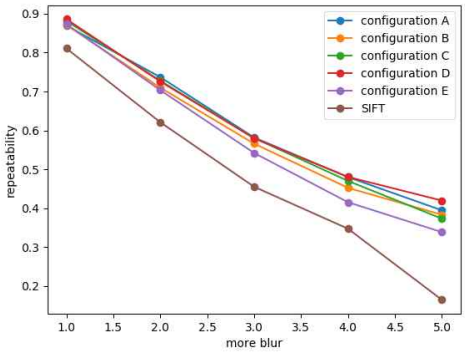
정확도가 높은 결과를 보여준 가중치로 진행하였다. 검증 데이터 세트는 밝기 정도와 흐림 정도를 변경해가며 구성

하였으며, 그림 4와 그림 5에 예를 보이고 있다. 그림 5(a)는 그림 4와 그림 5의 모든 영상의 원본 영상이며, 그림 4는 밝기 정도를 4단계로 적용한 영상이고, 그림 5(b)~(f)는 흐림 정도를 5단계로 적용한 영상이다.

이와 같이 왜곡된 영상들에 대한 특징점 추출 실험을 진행하였으며, 결과를 원 SIFT(알고리즘적 추출) 실험결과와 그림 6에서 비교하였다. 그림 6(a)는 밝기 변화에 따른 반복성으로, 모든 변형영상에서 제안한 DNN이 원 알고리즘보다 높은 추출 반복성을 보였다. 밝기 1, 2단계에서는 원 SIFT 알고리즘과 제안한 DNN의 반복성 차이가 크지 않지만, 밝기 변화가 심해질수록 차이가 커지는 것을 볼 수 있다. 그림 6(b)는 흐림 정도 변화에 따른 반복성인데, 이 실험에서는 약한 흐림 정도부터 원 알고리즘에 비해 크게 높은 추출 반복성을 보였다. 또한 흐림 정도를 크게 적용할수록 그 차이도 커지는 것을 확인할 수 있다.



(a)



(b)

그림 6. 왜곡된 영상의 특징점 반복성 측정결과; (a) 밝기 변화, (b) 흐림 정도 변화

Fig. 6. Results of the feature point repeatability measure for the distorted images; (a) change in brightness, (b) change in blur

표 2의 추출 실험 결과와 그림 6의 추출 반복성 실험 결과에서 보듯이, 제안한 DNN은 원 알고리즘에서 추출한 특징점의 99% 이상을 추출할 수 있고, 추출 반복성 실험에서도 기존의 SIFT 알고리즘보다 밝기와 흐림 정도 변화 모두에 훨씬 더 강인하다는 것을 확인하였다.

V. 결 론

본 논문에서는 딥 러닝을 사용한 SIFT 특징점 추출 DNN을 제안하였다. DNN은 VGG DNN(VGG-16)을 기반으로 네트워크의 깊이(컨볼루션 계층 수)를 증가하고, 풀링 방법과 활성화 함수를 달리하여 총 5개의 DNN을 구성하였다. 실험결과 제안한 모든 DNN이 99%가 넘는 추출 정확도를 보였으며, 밝기와 흐림 정도를 변화하여 만든 왜곡영상에서 원 영상의 특징점을 추출하는 반복성이 기존 SIFT 알고리즘보다 높음을 보였다. 따라서 제안한 DNN은 SIFT 특징점 추출에 있어 추출률 자체의 성능저하가 거의 없는 높은 추출 반복성으로 악의적/비악의적 영상왜곡이 발생한 경우 원 알고리즘보다 더 효과적으로 사용할 수 있을 것으로 사료된다.

실험결과에 의하면 고려한 5개의 DNN은 네트워크의 깊이가 깊을수록 추출률 성능이 향상되는 것으로 나타났으나 추출 반복성에 있어서는 큰 차이를 보이지 않았다. 풀링 방법으로는 최대풀링보다 간격을 2×2로 한 경우가 추출률과 추출 반복성 모두에서 약간 높은 성능을 보였다. 활성화 함수로 시그모이드 함수를 사용한 경우가 소프트맥스를 사용한 경우보다 추출률과 추출 반복성 모두에서 더 좋은 성능을 보였다. 따라서 네트워크 깊이를 깊게, 그리고 2×2 간격으로 스케일링을 수행하고, 시그모이드 함수를 활성화 함수로 사용하는 것이 가장 좋은 성능을 보였다.

딥 러닝이 가장 많이 적용되고 있는 분야가 영상처리 분야이다. 일반적으로는 네트워크의 크기와 깊이가 증가할수록 좋은 성능을 보인다고 알려져 있으나, 최근 연구에서는 응용 분야에 따라 네트워크 크기에 대한 포화가 발생한다고 보고되고 있다. 네트워크의 크기나 깊이는 연산량과 직결되므로 응용분야에 따라서 시간적인 문제

가 발생할 수도 있다. 따라서 향후 연구에서는 영상처리에 있어서 어떤 응용분야에서는 어느 정도의 네트워크를 사용하는 것이 가장 합리적인지를 판별하는 연구를 진행하고자 한다.

참 고 문 헌 (References)

[1] C. Harris, M. Stephens, “A combined corner and edge detector,” Proceedings of the Alvey Vision Conference, pp.147-151, 1988.
[2] K. Mikolajczyk, C. Schmid, “Indexing based on scale invariant interest points,” ICCV, Vol.1, pp. 525-531, 2001.
[3] J. Shi, C. Tomasi, “Good features to track,” 9th IEEE Conference on Computer Vision and Pattern Recognition, Springer, Heidelberg, 1994.
[4] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” International Journal of Computer Vision, Vol.60, No.2, pp.91-110, 2004.
[5] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features,” In European Conference on Computer Vision, Vol.1, No.2, May 2006.
[6] E. Rosten, T. Drummond, “Machine learning for high-speed corner detection,” Proc. 9th European Conference on Computer Vision (ECCV'06), May 2006.
[7] E. Mair, G. Hager, D. Burschka, M. Suppa, and G. Hirzinger, “Adaptive and generic corner detection based on the accelerated segment test,” Computer Vision-ECCV 2010, Vol.2, No.2, pp.183-196, 2010.
[8] M. WonJun, S. Youngho, and K. Dongwook, “Parameter Analysis for Time Reduction in Extracting SIFT Keypoints in the Aspect of Image Stitching,” Journal of Broadcast Engineering, Vol.23, No.4, pp.559-573, July 2018.
[9] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: an efficient alternative to SIFT or SURF,” In Proc. of the IEEE Intl. Conf. on Computer Vision (ICCV), Vol.13, 2011.
[10] E. Agustsson, R. Timofte, “NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study,” In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2017.
[11] R. Hess, “An Open-Source SIFT Library,” ACM Multimedia, pp.1493-1496, 2010.
[12] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” In Proc. International Conference on Learning Representations (ICLR), 2015.
[13] K. Mikolajczyk, C. Schmid, “Scale and affine invariant interest point detectors,” IJCV, Vol.1, No.60, pp.63-86, 2004.

저 자 소 개

이 재 은

- 2019년 2월 : 광운대학교 전자재료공학과 졸업(공학사)
- 2019년 3월 ~ 현재 : 광운대학교 전자재료공학과 (공학석사)
- ORCID : https://orcid.org/0000-0001-9760-4801
- 주관심분야 : 영상 처리, 딥 러닝, 뉴로모픽 시스템

문 원 준

- 2018년 2월 : 광운대학교 전자재료공학과 졸업(공학사)
- 2018년 3월 ~ 현재 : 광운대학교 전자재료공학과 (공학석사)
- ORCID : https://orcid.org/0000-0002-9620-9524
- 주관심분야 : Virtual Reality, 워터마킹, 2D 영상 신호처리, 딥 러닝

저 자 소 개

서 영 호

- 1999년 2월 : 광운대학교 전자재료공학과 졸업(공학사)
- 2001년 2월 : 광운대학교 일반대학원 졸업(공학석사)
- 2004년 8월 : 광운대학교 일반대학원 졸업(공학박사)
- 2005년 9월 ~ 2008년 2월 : 한성대학교 조교수
- 2008년 3월 ~ 현재 : 광운대학교 전자재료공학과 정교수
- ORCID : http://orcid.org/0000-0003-1046-395X
- 주관심분야: 실감미디어, 2D/3D 영상 신호처리, 디지털 홀로그램, SoC 설계

김 동 욱

- 1983년 2월 : 한양대학교 전자공학과 졸업(공학사)
- 1985년 2월 : 한양대학교 공학석사
- 1991년 9월 : Georgia 공과대학 전기공학과(공학박사)
- 1992년 3월 ~ 현재 : 광운대학교 전자재료공학과 정교수
- ORCID : http://orcid.org/0000-0002-4668-743X
- 주관심분야 : 3D 영상처리, 디지털 홀로그램, 디지털 VLSI Testability, VLSI CAD, DSP설계, Wireless Communication