딥 러닝 기반의 SIFT 이미지 특징 검출

*이재은 문원준 서영호 김동욱
광운대학교
*wodms6364@kw.ac.kr

SIFT Image Feature Detect based on Deep learning

*Lee, Jae-Eun Moon, Won-Jun Seo, Young-Ho Kim, Dong-Wook
Kwangwoon University

요약

본 논문에서는 옥타브(sacle vector, octave)를 0, 시그마(sigma)는 1.6, 간격(intervals)은 3으로 설정하여 검출한 RobHess SIFT 특징들로 데이터 셋을 만들어 딥 러닝 모델인 VGG-16을 기반으로 SIFT 이미지 특징을 검출하는 방법을 제안한다. DIV2K 데이터 셋을 33×33 크기로 잘라서 데이터 셋을 구성하였고, 흑백 영상으로 판별하는 SIFT와는 달리 RGB 영상을 사용하였다. 영상을 좌·우 반전, 밝기, 회전, 크기를 조절하여 원본 영상을 변형시켜 네트워크 학습 및 평가를 진행하였다. 네트워크는 영상의 가운데에 위치한 픽셀이 특징점인지 아닌지를 판별한다. 검증 데이터의 결과 98.207%의 정확도를 얻었다.

1. 서론

SIFT(Scale-Invariant Feature Transform) 알고리즘은 RGB 영상을 흑백 영상으로 변환 후, 옥타브(sacle vector, octave)에 따라 영상의 크기를 확대, 축소하여 scale space를 만든다. 그리고 시그마(sigma)와 간격(intervals)에 따른 가우시안 필터와 DoG(Difference of Gaussian)을 적용하여 DoG 피라미드를 만들어 인접한 픽셀과의 비교 및 테일러 급수를 진행해 극값들을 찾고 keypoint로 명시해준다. 그 다음, 낮은 콘트라스트를 가지는 keypoint와 엣지에 존재하는 keypoint를 제거하여 유효한 특징들을 검출한다[1]. SIFT 이후에 속도를 개선한 SURF, ORB 등 많은 특징 검출 알고리즘이 나왔지만 아직까지 SIFT 알고리즘을 확연하게 우세한 알고리즘을 찾기는 어렵다.

따라서 본 논문에서는, 요즘 뛰어난 결과를 내어 주목받고 있는 딥 러닝을 기반으로 특징을 검출하는 방법을 제안한다. 데이터 셋과 네트워크를 구성한 방법에 대해 설명한다. 학습시킨 네트워크의 가중치로 특징 점인 영상과 특징 점이 아닌 영상의 비율을 조정하고, 영상을 변형한 뒤 시험을 진행하여 결과를 비교·분석한다.

2. 제안하는 방법

풍경과 건물 영상이 비교적 많고 영상 크기가 너무 크지 않은 DIV2K 데이터 셋에서 저해상도 데이터를 사용하였다[2]. 훈련 데이터 셋 800장을 RobHess SIFT 알고리즘으로 특징을 추출하여 label을 구성하였다[3]. SIFT 옥타브는 0, 시그마는 1.6, intervals는 3으로 설정하여 원본 영상 크기에 대한 특징만 추출하였다. 그림 1은 사용한 하이퍼파라미터를 기반으로 검출한 SIFT 특징들을 나타낸다. 흑백 영상으로 특징들을 검출하는 SIFT 알고리즘과 달리 RGB 영상을 사용하였고, 간격을 한 픽셀로 두어 33×33씩 잘라내 사용한다. 훈련 데이터는 특징 점인 영상을 215,252개, 특징 점이 아닌 영상을 1,074,260으로 1:5로 비율을 두어 총 데이터 수는 1,291,512개이다. 변화에 강인한 특징점을 검출하기 위해 영상을 좌·우 반전시켜 학습을 진행한다. 검증 데이터도 훈련 데이터와 마찬가지로 특징 점인 영상과 특징 점이 아닌 영상에 대한 비율을 1:5로 하고 좌·우 반전시켜 학습의 진행정도를 확인한다. 그리고 영상의 밝기, 회전, 크기를 조절하여 원본 영상을 변형시켜 시험을 진행한다.

VGG-16은 필터의 크기를 작게 하는 대신 층을 깊게 하여 학습 효율을 증가시킨 네트워크이다. 네트워크 구조는 그림 2에 나타낸다. 입력은 RGB 정보를 포함한 33×33×3으로 진행한다. 네트워크는 3×3 필터 크기를 가진 13개의 컨볼루션 계층과 3개의 전결합층으로 구성한다. 컨볼루션을 2, 2, 3, 3, 3번 진행할 때마다 2×2 최대풀링이 진행되고 채널 수는 64, 128, 256, 512, 512로 점점 증가하며 2개의 전결합층 크기를 512로 유지하다가 특징점인지, 특징점이 아닌지를 분간하기 위해 마지막 출력 층 크기는 1×1×1로 설정한다[4]. 활성화 함수는 마지막 층에서만 sigmoid함수를 사용하고 그 외에는 모두 leaky ReLU 함수를 사용한다.

그림 1. 옥타브를 0으로 설정하여 추출한 SIFT 특징들
Fig. 1. SIFT features extracted with octave set to 0

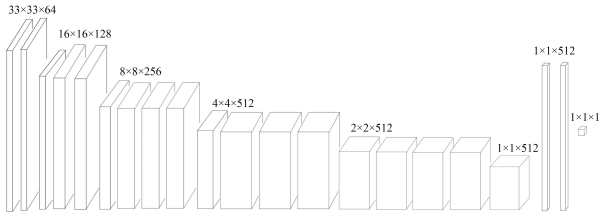

그림 2. VGG-16 네트워크 구조
Fig. 2. Architecture of VGG-16

3. 실험 결과

표 1. 영상의 특징 점인 것과 특징 점이 아닌 것의 비율을 조정하며 시험한 결과
Table 1. Test results of adjusting the ratio of feature point of image to non-feature point of image

rate	1:1	1:5	1:10	1:100	1:1000
accuracy	96.228%	98.01%	98.557%	98.901%	98.998%
feature	93.681%	93.681%	93.681%	93.681%	93.681%
non-feature	98.775%	98.876%	99.045%	98.953%	99.073%

표 2. 변형시킨 영상으로 시험한 결과
Table 2. Test results with modified images

	brightness	rotation	scale
accuracy	99.097%	98.063%	98.928%
feature	92.85%	59.15%	92.4%
non-feature	99.101%	98.124%	98.934%

네트워크의 훈련 데이터 정확도는 98.7%, 검증 데이터 정확도는 98.207%이다. 훈련 데이터 정확도는 에폭을 진행할수록 점점 증가하여 거의 100%에 도달하였지만, 시험을 진행할 때 더 높은 신뢰도를 위해 검증 데이터의 정확도가 높은 결과를 보여준 가중치로 시험을 진행하였다. 상황과 조건을 변경해가며 시험 데이터 셋을 구성하였으며, 그 결과는 표 1과 표 2에 나타낸다. 시험 결과는 전체 정확도, 특징 점인 것에 대한 정확도와 특징 점이 아닌 것에 대한 정확도로 나누어서 비교·분석한다. 모든 결과에서 특징 점인 것에 대한 정확도가 특징 점이 아닌 것에 대한 정확도에 비해 낮은 것을 볼 수 있다. 이는 훈련 데이터를 특징점인 데이터보다 더 많은 비율로 특징점이 아닌 데이터를 구성했기 때문이라고 사료된다.

표 1은 시험 데이터 셋에서 특징 점인 데이터는 모두 추출하였고, 특징 점이 아닌 데이터는 특징 점인 데이터의 1, 5, 10, 100, 1000배만큼 랜덤으로 추출하였다. 그 결과, 모든 경우에서 특징 점인 것에 대한 정확도는 똑같이 유지되었고, 특징 점이 아닌 것에 대한 비중을 크게 할수록 전체 정확도와 특징 점이 아닌 것에 대한 정확도가 증가하는 경향을 보였다.

표 2는 원본 영상을 밝기, 회전, 크기를 조절하여 변형시킨 데이터에 대한 시험 결과로, 변화에 강인한 정도를 보기 위해 진행하였다. 그 결과, 밝기 변화에 대해서는 좀 더 강인한 경향을 보였고, 회전 변화로 인해 특징 점인 것에 대한 네트워크의 판단력이 흐려진 것을 볼 수 있다.

4. 결론

본 논문에서는 훈련 데이터를 구성할 때, 특징 점인 데이터와 특징 점이 아닌 데이터에 대한 비율을 1:5로 두었고, 좌·우 반전에 대해서만 변형하여 학습시킨 네트워크의 결과를 나타내었다. 훈련 데이터에 특징점이 아닌 데이터의 비율을 높이거나, 밝기와 회전과 크기를 변형시킨 영상을 추가하여 학습을 진행한다면 정확도가 높아지고 변화에 강인한 네트워크를 만들 수 있을 것이라 기대된다.

본 논문은 딥 러닝을 사용하여 특징 검출을 할 수 있는 가능성을 높은 정확도로 증명하였고 SIFT 알고리즘보다 변화에 더 강인한 네트워크를 설계할 수 있을 것이라 기대한다.

감사의 글

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).

참고문헌

[1] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” International Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110, 2004.

[2] 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.

[3] R. Hess, “An Open-Source SIFT Library,” ACM Multimedia, pp. 1493-1496, 2010.

[4] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” In Proc. International Conference on Learning Representations(ICLR), 2015.