

딥 러닝을 이용한 화면 전환 검출

이재은 · 서영호 · 김동욱

광운대학교 전자재료공학과

Deep Learning-based Scene Change Detection

Jae-eun Lee · Young-Ho Seo · Dong-wook Kim

Kwangwoon University, Department of Electronic Materials Engineering

E-mail : jelee@kw.ac.kr / yhseo@kw.ac.kr / dwkim@kw.ac.kr

요 약

본 논문에서는 딥 러닝을 이용해 화면 전환을 검출하는 방식을 제안한다. 특징점을 추출할 때는 딥 뉴럴 네트워크를 사용하였고 추출한 특징점을 SIFT(Scale Invariant Features Transform) 기술자를 이용해 128차원 벡터를 생성한다. 이를 기반으로 각 픽셀마다 매칭 여부를 판단하여 25% 미만일 경우 화면 전환이라고 판단한다.

ABSTRACT

In this paper, we propose a method to detect the scene change using deep learning. To extract feature points, we use a deep neural network and express extracted feature points as 128 dimensional vectors using SIFT descriptor. If it is less than 25%, it is determined that the scene is changed.

키워드

Deep Learning, Scene Change, SIFT

I. 서 론

최근 온라인에서 영상에 대한 무분별한 공급에 따른 저작권 문제도 많이 발생하고 있다. 따라서 이를 방지하기 위한 기술들이 연구되고 있다. 이를 위해선 가장 기본적으로 화면 전환을 검출하는 기술이 필요하다.

따라서, 본 논문에서는 요즘 뛰어난 결과로 주목 받고 있는 딥 러닝을 기반으로 화면 전환을 검출하는 방식을 제안한다.

II. 제안하는 방법

본 논문에서 제안하는 화면 전환 검출 방식은 그림 1에 나타내었다. 본 논문에서는 특징점 추출 DNN(feature extraction DNN)과 SIFT(Scale Invariant Features Transform) 기술자(Descriptor)를 사용하였다.

먼저, 화면 전환을 판단할 두 영상을 각각 DNN을 이용하여 특징점을 검출한다. 검출한 특징점들을 SIFT 기술자로 128차원의 벡터를 생성한다. 그 다음 두 영상의 특징점 세트를 비교하여 매칭점을 찾아 그 비율에 따라 화면 전환인지를 판단한다.

1. 특징점 검출 네트워크

데이터 세트는 DIV2K 저해상도 데이터 800장을 사용하였다[1]. 옥타브는 0, σ 는 1.6, interval은 3으로 설정하여 RobHess SIFT 알고리즘으로 특징점을 추출하여 학습을 위한 정답을 구성하였다[2]. 입력 데이터는 33×33으로, 간격을 한 픽셀로 두었다. 훈련 데이터는 특징점인 영상과 특징점이 아닌 영상의 비율을 1:10으로 두어 각각 215k개, 2,152k개로 총 데이터 수는 2M개를 사용하였다.

DNN 구조는 VGG-16을 기반으로 구성된다[3]. 그림 2에 본 논문에서 사용하는 네트워크의 구조

를 나타내었다. 이 네트워크는 컨볼루션 계층(Conv)이 23개, 최대 풀링(max pooling)이 5개 그리고 전결합(fully connected, FC)층이 3개로 구성된다. 마지막 층을 제외한 전결합층의 활성화(activation) 함수는 leaky ReLU 함수를 사용하였고, 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.

2. SIFT Descriptor and Matching

SIFT 기술자를 구현하기 위해 openCV 함수를 이용한다. DNN으로 추출한 특징점을 기준으로 주위의 4x4 픽셀들을 8차원의 방향을 판단하여 128차원의 벡터를 생성한다.

두 영상을 매칭시킬 때 특징점 세트의 128차원 벡터에 대해 유클리디안 거리를 구하여 최소인 점을 매칭점이라고 한다. 이때 최소인 점과 두번째로 최소인 점의 유클리디안 거리의 비율이 0.8보다 크면 매칭점일 확률이 급격히 커지기 때문에 매칭을 시키지 않는다. 두 영상에 대해 매칭을 진행한 후, 특징점에 대한 매칭점의 비율이 25%미만일 경우 화면 전환이라고 판단한다.

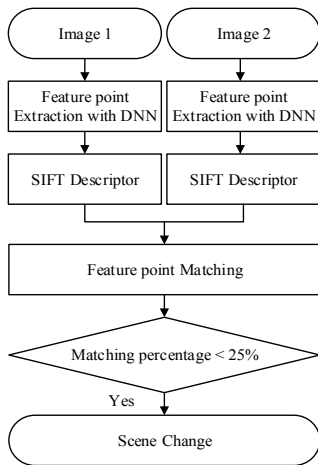


그림 1. 제안하는 화면 전환 검출기

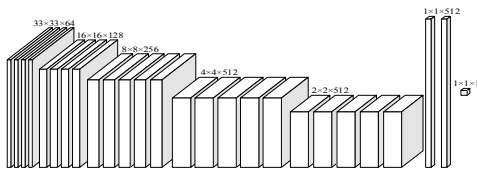


그림 2. 특징점 추출 DNN 구조

III. 실험 결과

실험 데이터는 30fps를 가진 80초짜리 동영상은 10프레임마다 추출하였다. 총 240장으로 화면 전환은 26번이 발생한다. 제안하는 딥 러닝을 이용한 화면 전환 검출기는 정확도 80.7%를 가진다. 그림 3은 화면 검출 결과들을 보여준다. 그림 3(a)와 (b)는 검출된 영상이고 그림 3(c)와 (d)는 미검출된 영상이다.

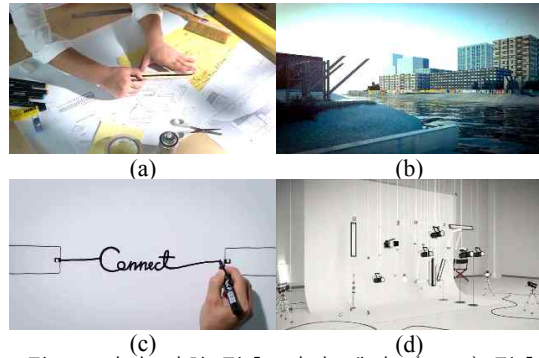


그림 3. 화면 전환 검출 결과 예시; (a, b) 검출, (c, d) 미검출

IV. 결론

본 논문에서는 딥 러닝을 이용한 화면 전환 검출기를 제안하였다. 특징점 추출 DNN은 VGG-16을 기반으로 컨볼루션 계층 수를 증가하여 DNN을 구성하였다. 추출한 특징점을 SIFT descriptor를 사용해 128차원의 기술자를 생성하여 이를 기반으로 매칭점을 추출하였다. 특징점 추출뿐만 아니라 descriptor도 딥 러닝을 사용한다면 더 높은 정확도가 나올 것이라 기대된다.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2018R1D1A1B07043220).

References

- [1] 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.
- [2] R. Hess, "An Open-Source SIFT Library," ACM Multimedia, pp.1493-1496, 2010.
- [3] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," In Proc. International Conference on Learning Representations (ICLR), 2015.