인터랙티브 미디어 창작을 위한 Semantic Segmentation 모델을 활용한 동영상 배경 변환 연구

송복득, 최홍규, 김성훈
한국전자통신연구원

bdsong@etri.re.kr, hk-choi@etri.re.kr, steve-kim@etri.re.kr

A Study on the Transformation of Video Background Using Semantic Segmentation Model for Interactive Media Creation

Bok Deuk Song, HongKyw Choi, Sung-Hoon Kim

Intelligent Convergence Research Laboratory
Electronics and Telecommunications Research Institute(ETRI).

요 약

본 논문은 인터랙티브 미디어 창작 서비스 플랫폼에서 사용자의 시나리오에 따라 영상을 창작하기 위하여 위하여 Semantic Segmentation 모델을 활용한 동영상 배경 변환 연구에 관한 내용이다. 영상 내 객체 인식은 Semantic Segmentation 모델을 이용하여 사용자가 추출하고자 하는 객체를 인식하여 인식된 객체외 영상을 배경으로 간주하고 합성하고자 하는 타켓 이미지를 선택하여 객체외 영역을 배경 영상으로 변환 할 수 있다. 이러한 연구는 사용자 시나리오에 맞게 영상을 새로 촬영할 필요 없이 기존 촬영된 영상을 활용하여 배경이 변환된 새로운 영상을 생성할 수 있는 장점을 제공한다.

Ⅰ. 서 론

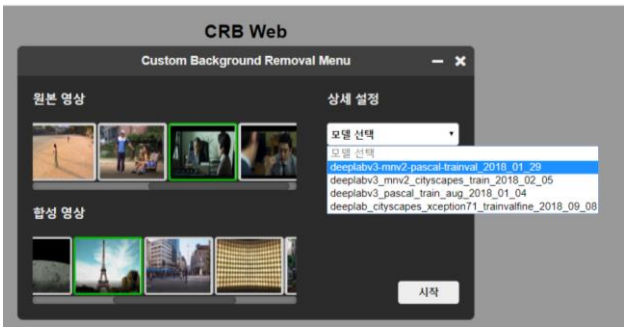
전세계적으로 인공 지능을 기반으로 영상 내 객체 인식, 추적, 변환등 영상 인식 기술에 대한 연구가 활발히 진행 되고 있다. 이러한 기술은 자율 주행, 의료, 디지털 미디어 산업등에 많은 영향을 주고 있으며 특히 AR, VR, MR 등 현실 세계와 가상 세계의 객체 인식 및 변환등에 적용되고 있는 추세이다. 본 논문에서는 Semantic Segmentation 모델을 활용하여 각 모델에서 인식하는 객체외의 영상 영역을 배경으로 간주하여 추출된 객체에 사용자가 변환하고자 하는 배경 영상을 선택하여 영상 합성 과정을 수행한 후 사용자 시나리오에 맞는 새로운 영상을 생성하는 동영상 배경 변환 연구를 수행하였다.

Ⅱ. 본론

본 논문에서 객체 인식을 위해 사용한 Semantic Segmentation 모델은 최근 구글에서 발표한 DeepLab V3+ 이다. DeepLab V3+ 는 내부에 여러 CNN 모델들의 조합으로 구성되며, encoding 과 decoding 프로세스를 통해 학습 속도를 개선한 구조이다[1]. 따라서 DeepLab V3+ 에서 사용할 CNN 모델들을 선택하여 구성할 수 있고, 정확도를 위한 XCEPTION 과 빠른 속도의 MobileNet 을 backbone 으로 사용해 배경 객체 인식을 위한 semantic segmentation 모델을 구성했다[2].

본 논문에서는 온라인에서 제공되는 많은 학습용 데이터셋 중에 유명한 Pascal Coco 와 Cityscape 데이터셋을 기본적으로 사용하였다[3][4].

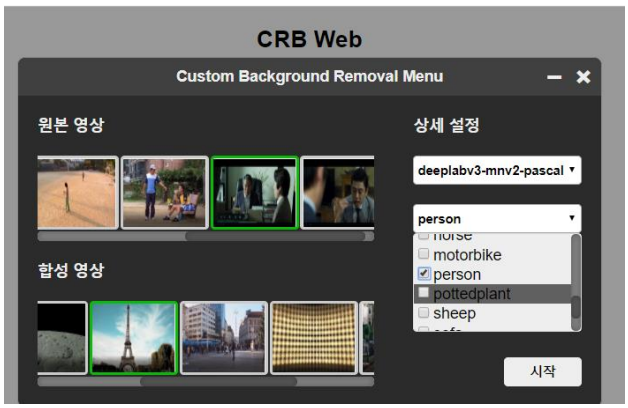
본 논문에서는 실제 객체 인식이 필요한 영상에서 데이터들을 추출하고 가공하여 학습 데이터셋을 구축하여 보다 정확한 인식이 가능하게 하는 학습 데이터 구축을 위하여 사람을 제외한 일반 객체들(산, 하늘, 차, 건물 등)은 기존 Cityscape, Coco 데이터 셋의 데이터를 그대로 사용하거나 조금 변형하여 사용했다. 주로 사람을 인식 대상 객체로 지정하는 경우가 많고, 테스트로 사용한 영상에서 사람 위주의 영상이었기 때문에 사람의 경우 영상에서 직접 사람 이미지를 추출하고, image augmentation 을 통해 데이터셋을 확장하여 새로운 학습 데이터셋을 구축했다. 그림 1 과 같이 다양한 딥러닝 기반의 영상 학습 모델을 선택하여 영상 내 객체 인식을 수행한다.



<그림 1> 배경 객체 인식용 딥러닝 모델 종류 선택 화면

그림 2 는 객체 인식을 위하여 영상 학습 모델에 따라 인식하고 하는 객체를 선택하는 화면으로 예를 들어 영상에서 사람을 인식하려면 person 을 선택한다. 영상에서 사람만을 인식하기 때문에 인식된 사람

객체외는 배경으로 인식한다. 입력 비디오 프레임과 배경이 설정되면 get_segmented 함수에서 배경 제거가 진행된다. 먼저, 입력 비디오 프레임의 입력 값으로 Deeplabv3 모델의 prediction 결과인 입력 이미지에 대한 segmentation map 을 입력 받는다. segmentation map 에서 원하는 클래스에 해당되지 않는 모든 값들을 0 으로 할당한다. 그리고 segmentation map 은 흑백 masking 처리를 통해 변환된다. 객체에 해당되는 pixel 들은 RGBA (255, 255, 255, 255) 값을 갖고, class(객체)에 해당되지 않는 pixel 들은 RGBA (0, 0, 0, 0) 값을 갖는다.



<그림 2> 인식하고자 하는 객체 종류 선택 화면

객체를 인식하고 인식된 개체외의 영역을 배경이므로 다른 영상을 배경 영상으로 합성하는 과정을 수행한다. 배경 변환 영상의 길이가 객체를 추출할 원본 영상의 길이와 같도록 하기 위해서 배경 영상에서 프레임을 조정하여 각 프레임에 해당하는 프레임을 배경 영상으로부터 합성하는 과정을 수행한다 그림 3 은 본 논문의 연구 결과로써 원본 영상에서 Semantic Segmentation 모델은 을 적용하여 사람 객체를 인식하고 사람외의 영역을 배경으로 간주하고 배경에 타겟 이미지로 합성한 결과 영상을 보여준다.

<그림 3> 원본 영상(상)에서 동영상 배경 변환 결과(하)

Ⅲ. 결론

본 논문에서는 사용자의 시나리오에 따라 영상 내 배경 객체 변환을 위하여 영상 학습 모델을 활용하여 객체를 인식하고 인식된 객체외 영역을 배경으로 간주하여 변환하고자 하는 타겟이미지를 배경으로 합성하는 동영상 배경 변환 연구를 수행하였다. 이는 인터랙티브 미디어 창작시 기존 영상을 활용하여 새로운 영상을 생성할 수 있으며 기존 동영상 배경 변환과는 다르게 영상에서 인식된 객체외의 영역을 배경으로 간주하는 차이점을 제공한다. 향후 다양한 영상 학습 모델에서 제공하는 객체 인식 기능을 활용하여 동영상 배경 변환 기능을 사용할 수 있을 것으로 예상되며 추가적으로 개발되는 영상 학습 모델도 적용할 수 있을 것으로 예상된다.

ACKNOWLEDGMENT

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [20AR1110, Development of programmable interactive media creation service platform based on open scenario].

참 고 문 헌

[1] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam,"Encoder-decoder with atrous separable convolution for semantic image segmentation", In ECCV, 2018.

[2]Franc¸ois Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions" In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[3] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick., "Microsoft COCO: Common objects in context", In ECCV, 2014

[4] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler,R. Benenson, U. Franke, S. Roth, and B. Schiele. "The cityscapes dataset for semantic urban scene understanding ", In CVPR, 2016