숏사이즈 유형 분류 확장을 위한 연구

이수연*, 한채윤*, 임양미**
**덕성여자대학교 IT미디어공학과
e-mail : yosimi@duksung.ac.kr

The Study for the Shot Size Types Classification Expansion

Soo-Yeon Lee*, Chae-Yun Han*, Yang-Mi Lim**
**Dept of IT Media Engineering, Duksung Women's University

요 약

최근 딥러닝의 발달로 정지 이미지에서의 객체 인식은 물론 동영상의 내용 흐름 분석이 가능해져 가고 있다. 본 연구는 영화나 드라마의 내용 분석을 위해 딥러닝을 위한 방대한 숏사이즈 유형별 학습데이터가 필요하다. 숏사이즈 유형에는 익스트림 클로즈업, 클로즈업, 미듐, 롱, 익스트림 롱으로 분류하였고, 익스트림 클로즈업과 익스트림 롱에 대한 숏이미지의 부족으로 데이터셋 확장을 본 연구에서 진행하였다. 데이터셋 확장은 YOLO에서 지원하는 라이브러리 사용과 이미지 합성을 시도하였다. 이와 같이 분류된 이미지 데이터는 영상들의 전체적인 숏사이즈 변화를 자동으로 파악할 수 있어, 향후 영화나 드라마에서 숏사이즈 변화 패턴 분석이 가능할 수 있다고 판단한다.

1. 서론

최근 딥러닝(deep learning) 분야의 연구가 본격화되면서 이미지보다 동영상 분석에 더 초점을 두고 있다. 이러한 현상은 사람의 인지가 정지된 이미지보다 움직이는 동영상에서 더 많은 정보를 이해하는데 쉽기 때문일 것이다.

딥러닝 기술은 대량의 데이터를 분석하여 패턴을 도출해 객체를 식별하도록 하는 기술로 데이터 확보, 데이터 분류, 분석 단계로 진행된다. 따라서 대량의 데이터 확보가 중요하다고 볼 수 있으며, 다양한 데이터를 확보하는 데에는 오랜 시간이 소요된다. 본 연구는 흥행 영화 숏구성의 패턴 분석을 위해서는 데이터 수 만장을 수집해야 한다. 따라서 흥행영화 분석 수행 이전에 본 연구는 수백만 장의 빅데이터 구축을 위해 영화 동영상 데이터의 숏 경계 추출 연구[1]를 통해 숏이미지 데이터를 수집하였다. 추출된 이미지 데이터들은 익스트림 클로즈업(extreme close-up), 클로즈업(close-up), 미들(middle), 풀(full), 익스트림 롱(extreme long)의 숏사이즈 유형별 분류를 하였다. 하지만 익스트림 클로즈업과 익스트림 롱의 데이터는 클로즈업과 미듐 숏사이즈 유형에 비해 상대적으로 적어 딥러닝 학습을 하기에는 역부족이었다.

본 연구에서는 숏사이즈 유형을 CNN(convolutional neural network)기반의 YOLO(you only look once)모델을 활용하여 익스트림 클로즈업 데이터를 확장과 이미지 합성을 활용하여 익스트림 롱 데이터를 확보하였다.

2. 관련연구

데이터셋을 확장하기 위한 매우 일반적인 방법은 이미지 반사, 자르기, 회전, 이동, 컬러팔레트 변경 등과 같은 방법을 취하며[2], 작은 데이터셋을 갖고 있는 ImageNet과 MNIST(mixed national institute of standards and technology)에서 데이터를 부풀리기 위해 주로 수행되어 사용되어왔다[3,4]. 하지만 과거에는 생성된 수많은 이미지의 가치성 판단이나 평가를 하지 못하고 사용하여 학습의 결과는 효율이 향상되었다고 할 수 없었다. 2014년에 발표된 Ian Goodfellow의 GAN(generative adversarial network)은 부풀려서 만들어진 이미지 데이터들의 진위여부 평가를 하여 학습의 효율을 더욱 향상시켰다[5]. 김과 남, 장은 사물의 자세-위치-행동 통합 인식 딥러닝 시스템을 통해 사람의 움직임 분석을 시도하였는데, 이 연구에서도 데이터의 부족으로 마스크를 이용한 이미지 합성을 활용하여 부족한 데이터를 확장하는 방법을 시도하였다[6].

본 연구에서는 데이터 확장의 시간적 효율성을 위해 Keras에서 지원하는 ImageDataGenerator 함수를 사용하여 데이터 확장을 시도하였고, 획득한 데이터의 가치평가를 위한 정규화 과정을 추가하지 못하였지만, 새로운 이미지 생성방법이 아닌 기존 이미지 활용방안을 적용한 CNN을 활용을 통해 추출한 객체와 추출된 객체 이미지를 활용하여 전체 이미지의 관계를 추론하는 방식인 DNN을 사용하여 데이터 효용도를 높였다. [7,8].

3. 시스템 설계 및 구현

영화나 드라마의 숏 이미지는 영상 흐름의 의미가 있다. 예를 들어 클로즈업 숏은 인물 전체를 표현하거나 배경과 함께 상황을 보여주는 경우에 사용되고, 미들 숏은 인물 간의 대화 장면, 풀 숏은 인물전체를 표현하거나 배경과 함께 상황을 보여주는 경우에 사용되게 된다[9]. 본 연구에서는 숏 이미지의 영상의 의미 분석을 위해 숏사이즈 유형 데이터셋 구축을 하였다.

딥러닝을 위한 데이터셋 구축을 위해서는 ImageNet이나 MNIST, 기타 대학연구소에서 지원해주는 데이터를 사용을 해도 되나 본 연구의 목적에 맞는 학습을 위해서 사람을 중심으로 숏사이즈 class 유형 5개로 분류한 커스텀 모델을 만들었다. 데이터를 모으는 방법은 영화의 클라이맥스 부분의 2~3분 정도의 비디오데이터를 숏경계 추출과 숏사이즈 유형 분류 시스템을 구축하여 검출하였다[1, 10].

각각의 숏사이즈 유형별 class 5개에는 최소 1000장씩 있어야 하는데 영화 속에 포함된 익스트림 클로즈업과 익스트림 롱에 대한 데이터는 다른 숏사이즈 데이터에 비해 상대적으로 작았다. 따라서 익스트림 클로즈업은 API Keras에서 제공하는 Lambda Layer와 Concatenate를 사용하여 바운딩된 이미지 내의 객체만 잘라서 이미지를 확대하는 방법을 사용하였다. 익스트림 롱은 일반적인 합성 방식을 사용하여 바운딩된 이미지의 크기를 작게 하고 배경과 합성하였다. 그림 1은 데이터 확장 부분을 표시한 구조도이며, 트레이닝셋에 사용될 이미지들이 생성되는 곳을 점선 박스로 표기하였다.

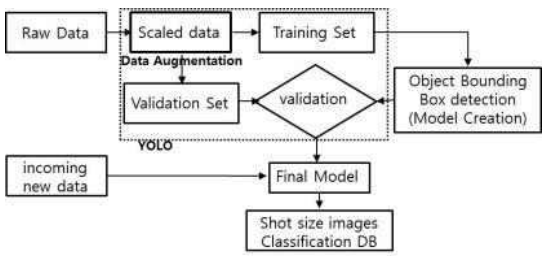

그림 1. 데이터셋 확장 시스템 구조도

숏경계검출 시스템[1,10]에서 검출된 숏이미지에서 객체를 탐색하여 바운딩 박스처리를 한 것이 그림 2와 3의 왼쪽 이미지이다. 숏사이즈 유형 판단은 전체 이미지의 크기와 바운딩 박스의 크기를 비교하여 클로즈업 숏임을 판별하였다. 클로즈업 숏 이미지에서 배우의 얼굴만 검출하기 위해 HAAR feature 기반의 cascade classifier를 사용하였다. 인식된 얼굴 중심으로 16:9의 이미지 크기를 보정하여 자르기를 시도하였다. 익스트림 롱의 경우는 배경으로 사용될 숏이미지를 구글에서 이미지를 파싱하여 획득한 것과 풀숏 이미지에 있는 객체를 검출하여 바운딩 박스를 중심으로 자른 후 합성하였다.

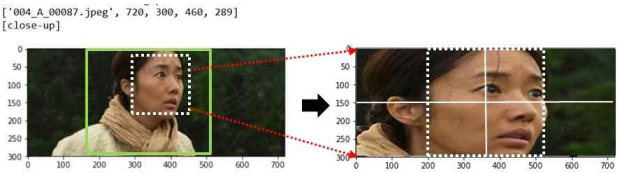

그림 2. 클로즈업 숏이미지에서 익스트림 클로즈업 숏이미지로 확대하기

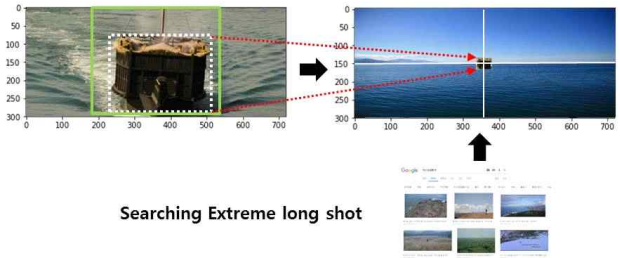

그림 3. 익스트림 롱 숏이미지 수집 방법

4. 결론

본 연구는 영화의 영상 데이터로부터 숏사이즈를 분류를 위해 부족한 데이터셋을 확장하는 방안의 연구를 진행하였다. 하지만 숏사이즈 분류는 최소 익스트림 클로즈업, 클로즈업, 미들, 풀, 롱, 익스트림 롱으로 분류해야 본 연구의 궁극적 목적인 숏사이즈 분류를 통한 클라이맥스 패턴 연구가 가능하다. 뿐만 아니라 숏앵글이나 subject/object 등 여러 가지를 구현해야 할 필요성이 있다. 영화에서 표현하는 연출 방식이 다양하여 적합한 데이터 수집 연구가 계속되어야 한다. 향후 숏사이즈 검출 모델의 정확도를 더욱 올리기 위한 연구를 추가적으로 모색해서 클라이맥스 패턴 분석이나 자동 카메라 워킹 편집에 응용되도록 진행할 예정이다.

참고문헌

[1] S.R. Park and Y.M. Lim, “The Implementing a Color, Edge, Optical Flow based on Mixed Algorithm for Shot Boundary Improvement,” Journal of Korea Multimedia Society, Vol. 21, No. 8, pp. 829-836, August 2018.

[2] H. S. Baird, Document image analysis. chapter Document Image Defect Models, pp. 315 - 325. IEEE Computer Society Press, Los Alamitos, CA, USA, 1995.

[3] Deep mnist for experts. https://www.tensorflow.org/get_started/mnist/pros. (accessed Dec. 24, 2018)

[4] The mnist database of handwritten digits. http://yann.lecun.com/exdb/mnist. (accessed Dec. 24, 2018)

[5] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber. “Deep Big Simple Neural Nets Excel on Handwritten Digit

Recognition," the Journal of Neural Computation, Vol. 22, No. 12, December 2010
[6] J.S. Kim, C.J. Nan and B.T. Zhang, "Deep Learning-based Video Analysis Techniques," Journal of Communications of the Korea Information Science Society, Vol. 33, No. 9, pp. 21-31, September 2015.
[7] M.O. Huy, K.M. Kim and B.T. Jang, "Deep Learning based Video Story Learning Technology," Journal of Korea Multimedia Society, Vol. 20, No. 3, pp. 23-40, 2016.
[8] R. B. Girshick, J. Donahue, T. Darrell, and J.Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 580-587, Columbus, Ohio, USA, June 2014.
[9] Y.M. Lim, "The Climax Expression Analysis Based on the Shot-list Data of Movies," Journal of The Korean Society Of Broad Engineers, Vol. 21, No. 6, pp. 965-976, November 2016.
[10] S.R. Park, J.E. Eom and Y.M. Lim, "The System Design and Implementation for Detecting the Types of Shot Size," Proceeding of Conference Korea Multimedia Society, Vol. 21, No. 1, pp. 968-996, Seoul, Korea, May 2018.

이 논문은 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임. (No.NRF-2017R1D1A1B03028804)