

신경망과 TFLITE를 이용한 한국수어 번역 알고리즘 구현

Implementation of Korean sign language on mobile using neural network

저자 (Authors)	정동화, 송준영, 이영훈, 이현승, 강소희, 임완수 Donghwa Jeong, JunYoung Song, YoungHun Lee, Hyeon seung Lee, Sohee Kang, Wansu Lim
출처 (Source)	한국통신학회 학술대회논문집 , 2019.11, 499-501(3 pages) Proceedings of Symposium of the Korean Institute of communications and Information Sciences , 2019.11, 499-501(3 pages)
발행처 (Publisher)	한국통신학회 Korea Institute Of Communication Sciences
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09277798
APA Style	정동화, 송준영, 이영훈, 이현승, 강소희, 임완수 (2019). 신경망과 TFLITE를 이용한 한국수어 번역 알고리즘 구현. 한국통신학회 학술대회논문집, 499-501
이용정보 (Accessed)	동신대학교 220.95.42.*** 2020/08/18 22:29 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

신경망과 TFLITE를 이용한 한국수어 번역 알고리즘 구현

정동화¹, 송준영¹, 이영훈¹, 이현승¹, 강소희², 임완수²

¹금오공과대학교 컴퓨터소프트웨어공학과

²금오공과대학교 IT융복합공학과

e-mail : wansu.lim@kumoh.ac.kr

Implementation of Korean sign language on mobile using neural network

Donghwa Jeong¹, JunYoung Song¹, YoungHun Lee¹,
Hyeon seung Lee¹, Sohee Kang², Wansu Lim²

Department of Computer Software Engineering¹

Department of IT Convergence Engineering²

Kumoh National Institute of Technology

요 약

본 논문은 기계학습을 이용한 수어번역 알고리즘을 모바일에 이식하는 방법을 제안한다. 수화는 동작을 통해 의사를 전달하므로 동영상상을 판별하는 방법이 필요하다. 이를 위해 Conv3D를 사용하였다. 또한 수화의 특성 상 비슷한 모양이 많으므로 인식률을 높이기 위해 Opticalflow를 계산하여 전처리를 시행하였다. 마지막으로 Android Application에 적용하기 위해 모델(ex. pb파일)을 lite 파일로 변환하는 과정을 거쳤다.

I. 서 론

청각 장애인의 수는 지금까지 꾸준히 증가하고 있다. 그리고 청각 장애인은 의사소통을 위해서 수화를 배우게 되며, 이에 따라 수화를 사용하는 사람의 수도 늘어나고 있다. 그러나 일반인들은 수화를 모르는 사람이 대부분이기 때문에, 수화를 알고 있는 사람과 그렇지 않은 사람 사이의 의사소통에 큰 어려움을 겪고 있다. 이러한 문제를 해결하기 위해 수화를 번역하기 위한 많은 노력들이 시도되어 왔고, 이는 머신러닝 기술의 발달과 함께 더욱 급속도로 진행 중이다.[1]

수화를 번역하는 데 있어서 가장 수월한 기기는 단연 스마트폰이라 할 수 있다. 준수한 카메라 성능을 갖추었으며, 우리 주변에서 매우 쉽게 접할 수 있는 기기이기 때문이다. 따라서 스마트폰을 이용한 수화 번역 어플리케이션은 이미 많이 존재한다. 그러나 그 정확도가 매우 떨어지거나, 서버단에서 수화를 번역하는 작업을 수행하기 때문에 인터넷 연결이 필수적이고, 이미지나 동영상을 서버에 전송하기 때문에 많은 데이터를 소모하고, 매우 느린 전송속도와 번역속도를 보여주는 경우가 대부분이다. 또한 서버를 유지하기 위한 노력과 비용도 추가로 요구된다.

이러한 문제를 해결하기 위하여 본 연구에서는, 수화번역 머신러닝 모델의 정확도를 높이기 위하여 Optical flow 와 3D Convolution 을 사용하게 되며, 머신러닝 모델 신경망을 서버가 아닌 모바일에 이식하여 개개인이 인터넷 연결에 구

애받지 않고, 언제 어디서나 사용할 수 있는 수화 번역 어플리케이션의 개발을 목표로 한다.

II. 본론

1) 비디오의 특성

이미지가 특정한 순서로 존재하는 것이며 이로 인해 공간적 특징과 시간적 특징을 고려해야 한다. 따라서 수화 인식을 위해서 동영상을 판별하는 기술이 필요하다.

2) 동영상 판별을 위한 접근법

2-1) 비디오의 공간적 특성만을 사용(이미지 판별만을 사용) 이미지를 frame으로 나눈 후 적은 데이터 셋을 이용하여 효과를 극대화하기 위해 전이학습(transfer learning)을 이용한다. 광고나 스포츠와 같이 각 frame이 많이 다르면 유효한 방법이나 수화인식과 같이 각 frame이 비슷할 경우 정확도가 떨어져서 사용할 수 없다.

2-2) RNN을 사용하여 시간적 특성 사용

2-1절차를 수행하는데 프로그램 수정이 필요하다. CNN의 마지막 layer를 없애면 2048-vector 이미지 특징을 가진 layer를 사용 할 수 있는데 이 layer를 사용하여 나온 결과 값을 객체 파일로 저장한다. 이후 객체 파일을 사용하여 RNN에 교육시켜 모델을 얻어낼 수 있다.

2-3) Conv3d를 사용

기존의 이미지 판별을 위해 사용하던 Conv2D에서 다른 축을 하나 더 추가하여 이미지를 여러 개 쌓아서 input 할 수 있는 함수이다. 이 함수는 tensorflow 기본 라이브러리에서 제공되며 손쉽게 비디오를 판별 할 수 있다.

3) 손 추출

촬영을 진행할 때마다 배경 및 인물이 바뀌는 등 다양한 환경에서 원치않는 움직임 객체를 얻을 수 있게 된다. 그 결과 depth Sensor의 부재로 인한 손부분 외의 결과값 역시 분석하게 되는 문제가 발생한다. 따라서 관심 있는 부분인 손 부분만을 따로 추출하여 동영상 판별에 이용하는 것이 필요하다.

4) 손 추출을 위한 접근법

4-1) tensorflow object_detection 사용

tensorflow의 object_detection을 변형 시켜 손만을 인식할 수 있도록 한다. 데이터 셋은 The Egohands Dataset을 사용하였으며 ssdlitemobilenetv2에 전이학습(transfer learning)을 이용하여 학습시킨다. 이후 opencv의 canny edge detection algorithm을 이용하여 윤곽선을 추출하여 완성한다.

4-2) opencv Opticalflow 사용

Opticalflow란 두 개의 연속된 비디오 프레임 사이에 이미지 객체의 가시적인 동작 패턴을 말한다. 따라서 앞의 frame과 뒤의 frame을 비교하여 움직이는 물체를 판별 할 수 있다. 수화에서는 움직이는 부분이 손과 팔이므로 이 부분을 중점적으로 추출할 수 있기 때문에 효과적인 방법이라 할 수 있다.

5) Android Application의 제약 사항

모바일의 경우 PC보다 느리고 사용 가능한 함수 및 라이브러리에 있어 제한이 많다. 따라서 서버를 사용 여부에 따라 선택 가능한 접근법이 달라질 수 있고, 접근법에 따라 시간적 제약사항이 많이 다를 수 있다. 그러므로 서버 사용 여부 및 적합한 접근법을 이용하는 것이 중요하다.

5-1) 서버 사용

서버를 사용 할 경우 구현에 있어서 모든 접근법을 이용 가능하다. 클라이언트(Android Application)는 단순히 동영상 상을 서버로 전송하고 서버에서 손 추출 및 동영상 판별을 수행하여 결과값을 다시 전송해준다.

5-2) 서버 없이 Android Phone에서 처리

Android Phone에서 모두 처리 할 경우 PC기반으로 만들어진 프로그램을 같은 동작을 할 수 있도록 적합하게 옮기는 것이 필요하다. 먼저 동영상 판별 접근법을 비교하면 2-2 RNN을 이용하는 경우 CNN 의 마지막 layer를 제거한 모델에 각 frame을 넣고 그 결과값을 객체 파일로 저

장 후 RNN모델에 넣어야 하는 작업이 필요하다. 그에 반해 2-3 Conv3d의 경우 tensorflow에서 제공하는 함수이기 때문에 복잡한 과정 없이 이용 할 수 있다. 또한 파일 I/O의 경우 cost가 높으므로 2-2 접근법 대신 2-3 Conv3d를 채택하였다.

다음으로 손 추출 접근법을 비교하면 4-1의 경우 object_detection 하는 것에 많은 cost가 소요되며 더 큰 문제점으로 기계학습을 통한 손의 인식이기 때문에 정확히 판단 할 수 없을 때도 있다. 그에 반해 4-2 Opticalflow의 경우 움직이는 객체를 추적하는 것이기 때문에 손뿐만 아니라 팔의 움직임도 함께 포함시킬 수 있어 인식률을 더욱 높일 수 있다. 따라서 4-2 opencv Opticalflow를 채택하였다. 이후 python으로 구현된 로직을 Android java언어로 변환한다.

6) tensorflow lite 사용을 위한 pb모델 변환(5-2연장)

모델을 Android Application에서 사용하기 위해서 tensorflow mobile 또는 lite를 사용하는 것이 필요하다. tensorflow mobile의 경우 PC에서 사용하는 pb파일을 그대로 사용할 수 있는 장점이 있지만 속도가 느리고 지금 시점에서 폐기 상태에 있기 때문에 사용하지 않는 것이 좋다. 따라서 tensorflow lite를 사용하여 구현한다. tensorflow lite의 경우 PC에서 사용하던 모델(ex. pb파일)을 lite파일로 변환시켜 사용해야 한다. 변환 시 아직 lite에서 제공하지 않는 함수가 있는데 여기서는 Conv3d 함수를 제공하지 않으므로 실험적기능인 select_ops를 사용하여 변환하였다.

7) 결과



<그림 1> 녹화



<그림 2> Opticalflow 계산 및 추론



<그림 3>결과 출력

III. 결론

본 논문에서는 안드로이드 스마트폰 카메라를 이용하여 수어를 번역하는 알고리즘을 모바일에 이식하였다. Depth Sensor의 부재로 인한 객체의 인식의 문제점을 해결하기 위하여 Conv3d와 Optical flow를 활용하여 전처리를 계산하는 알고리즘을 구축하였으며 tensorflow lite 사용을 위한 모델 변환을 통해 모바일에 이식함으로써 알고리즘을 모바일에서도 구동이 가능하도록 하였다.

ACKNOWLEDGMENT

본 연구는 한국연구재단의 지원을 받아 수행된 연구임(No. 2017R1CB5016837)

이 연구는 금오공과대학교 학술연구비로 지원되었음 (2019-104-155)

참 고 문 헌

- [1] 강상연, 조현, 황원준. (2018). 딥러닝 기반 동영상 장면 분류 기법. 한국통신학회 추계종합학술대회논문집, 311-312.
- [2] G. Anantha Rao, "Selfie Continuous Sign Language Recognition with Neural Network Classifier" 2017, (https://link.springer.com/chapter/10.1007/978-981-10-4280-5_4).
- [3] Tensorflow Object Detection (Github) (https://github.com/tensorflow/models/tree/master/research/object_detection)
- [4] Real-time Hand-Detection (Github) (<https://github.com/victordibia/handtracking>)
- [5] Converting pb file To tflite file (https://www.tensorflow.org/lite/guide/ops_select)
- [6] Sign Language Recognition (Github) (<https://github.com/FrederikSchorr/sign-language>)
- [7] Sign Language Dataset (<http://aicompanion.or.kr/kor/tech/data.php>)