다중 객체의 행동 인식을 위한 심층신경망

김승현*, 김도연**
*순천대학교 컴퓨터과학과
**순천대학교 컴퓨터공학과
skim@sunchon.ac.kr

A Deep Neural Network for Activity Recognition of Multi-object

Seunghyun Kim*, Do-Yeon Kim**
*Dept of Computer Science, Sunchon National University
**Dept of Computer Engineering, Sunchon National University

요 약

행동 인식을 위한 기존의 심층신경망은 행동 패턴 모델링과 행동 인식 성능 향상에 큰 기여를 하였다. 그러나 이 신경망은 영상 전체를 하나의 행동 인식 대상으로 보기 때문에 다중 객체의 개별적인 행동 인식에는 한계가 있다. 이에 본 논문에서는 R-CNN과 LSTM을 융합한 RC-LSTM 심층신경망을 통해 다중 객체의 행동 인식을 위한 방법을 제안한다.

1. 서론

컴퓨터를 이용한 행동 인식은 미리 정의된 행동 패턴 모델과 새로 입력된 행동 패턴의 비교를 통해 수행된다. 이러한 행동 인식 과정은 행동의 형태적 변화에 관계없이 강건한 인식 능력을 갖춰야하기 때문에 고수준으로 일반화된 패턴 모델을 구축할 필요가 있다. 최근 재조명된 심층학습(deep learning)은 학습 데이터를 통해 데이터 전반의 특징 학습 및 가중치 조절이 스스로 수행되기 때문에 패턴 모델링을 위한 효과적인 방법으로써 각광받고 있다.

CNN(convolutional neural networks)[1]은 심층학습의 대표적인 신경망 중 하나로써, 컨볼루션 연산을 수행하는 전처리 레이어를 통해 영상의 주요 특징 요소를 반복적으로 추상화한다. R-CNN(regions with CNN features)[2]은 CNN을 응용한 것으로, 영상에 나타난 다중 객체의 검출과 인식을 위해 사용된다. 객체가 존재할 가능성이 높은 영역들을 컴퓨터 비전 기법을 통해 선정하고, 이후 CNN을 통해 각 영역에 대한 판별과 인식을 수행하는 방식이다[3]. 한편, 기존 신경망들을 융합하여 새로운 목적의 신경망을 설계한 연구들도 있다.

D. Jeffrey et al.[4]은 LRCN(long-term recurrent convolutional network) 모델을 통한 행동 인식 방법을 제안하였다. 이 모델은 기존의 CNN과 LSTM(long short-term memory network)을 융합한 모델로써, 시계열(time-series) 기반의 비디오 영상을 통해 행동 인식을 위한 학습과 분류를 수행한다. LSTM은 기억 능력을 지닌 신경망으로 각 프레임에 나타난 순간적인 행동 특징들의 연계성을 학습할 수 있다.

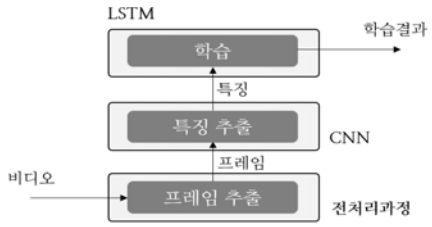
하지만 LRCN 모델은 다중 객체의 동시 다발적 행동에 대한 인식이 어렵다는 한계가 있다. 영상에 나타난 다중 객체의 개별적인 행동을 독립적인 행동으로 인식하지 못하고, 영상 전체를 하나의 행동 인식 대상으로 보기 때문이다. 따라서 영상에서 다중 객체의 뒤섞인 행동 특징 분포가 나타나게 될 경우, 전혀 엉뚱한 분류 결과를 출력하게 된다.

본 논문에서는 R-CNN과 LSTM을 융합한 RC-LSTM 심층신경망을 통해 각 객체의 행동 패턴을 인식할 수 있는 방법을 제안한다. RC-LSTM은 기존 LRCN을 이용한 행동 인식 방법을 확장한 것으로 R-CNN을 통해 관심영역(ROI, region of interest)을 설정하고, 관심영역 내의 다중 객체에서 표현되는 행동 특징들을 LSTM으로 독립적으로 인식하는 것을 최종 목적으로 한다.

2. 학습 데이터 구성 및 학습 과정

학습 데이터는 행동에 따라 분류된 비디오 데이터를 통해 구성된다. 신경망 학습을 위한 학습 데이터의 획득은 기존의 공개된 데이터베이스를 이용하거나, 직접 촬영하는 방법을 통해 이루어질 수 있다. 각 비디오 영상은 2~10초 사이의 짧은 촬영시간, 피사체와의 거리, 배경 변화, 카메라 고정 여부 등을 고려해 구성한다. 이는 학습에 투입되는 시간의 절약이나 효율성을 높일 수 있기 때문이다.

구성된 학습 데이터를 신경망에 학습시키는 과정은 기존 신경망의 학습 과정과 비슷하다. (그림 1)은 이 과정을 나타낸 것이다.



(그림 1) 제안된 신경망의 학습과정

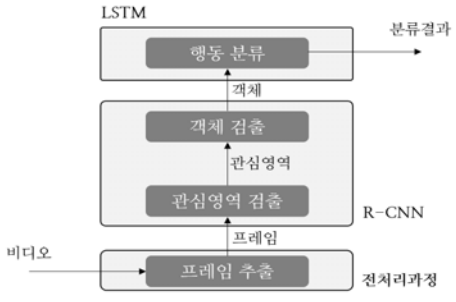
행동 인식을 위한 학습은 지도 학습(supervised learn -ing)을 통해 진행되며, 각 비디오 영상은 행동 부류에 맞는 라벨을 통해 표지된다(labelled).

제시한 학습 데이터는 단일 객체를 대상으로 한다. 여기에는 두 가지 이유가 있다. 첫 째, 다중 객체의 행동 인식 단계는 먼저 R-CNN이 영상에 나타난 각 객체를 독립적으로 검출해주고, 이후 LSTM에서 이들의 행동을 분류하는 순서를 거친다. 객체가 독립적으로 검출된 그 시점부터 각 객체의 행동은 단일 객체의 행동으로 볼 수 있다. 둘 째, 단일 객체의 행동만으로 구성된 학습 데이터는 지도 학습을 위한 행동 클래스 표지가 쉽기 때문이다. 다수의 객체가 각기 다른 행동을 행하는 학습 데이터는 어떤 행동 클래스로 표지해야 할지 결정하는 것이 쉽지 않다.

3. 다중 객체의 행동 인식 과정

다중 객체 행동 인식을 위해 영상에 나타난 객체들의 검출 과정이 선행되어야 한다. 객체 검출은 컴퓨터 비전 기법을 이용한 관심영역 검출을 통해 수행된다. 관심영역 검출 기법은 접근 방법에 따라 상향 또는 하향식 방법으로 나뉘며, 상향식 방법은 다시 통계적 접근 또는 스펙트럼 접근 방법으로 나뉜다[5].

관심영역은 관찰 대상 객체가 존재할 것으로 예상 되는 영역을 의미한다. LSTM이 시계열 행동 패턴에 대한 충분한 학습이 완료되었다고 가정했을 때, 다중 객체의 행동 인식을 위한 과정은 (그림 2)와 같다.



(그림 2) 제안된 신경망의 다중 객체 행동인식 과정

그림에서 비디오 영상은 먼저 프레임 추출을 위한 전처리 과정을 거치게 된다. 이 과정에서 다중의 RGB 프레임과 옵티컬플로우(optical flow) 이미지가 출력되며, 이미지 내부의 다중 객체 검출을 위해 R-CNN의 입력으로 들어간다. R-CNN은 입력된 프레임으로부터 객체 검출을 수행하기 위해 관심영역 검출을 위한 연산을 수행한다. 검출된 관심영역들은 객체 검출을 위한 연산의 입력으로 들어가는데, 이 과정에서 CNN이 사용된다. CNN의 출력결과는 해당 객체가 행동 인식 대상 인지 아닌지에 대한 판별 결과이며, 전자의 경우 본격적인 행동 인식을 위한 LSTM의 입력으로 들어가게 된다.

4. 결론 및 향후 연구과제

비디오 영상에서 다중 객체에 대한 행동 인식을 수행하려면 영상에 나타난 객체들의 검출이 선행되어야 한다. 본 논문에서는 R-CNN과 LSTM을 융합한 RC-LSTM 심층신경망을 통해 관심영역 검출을 기반으로 객체들을 검출하고, 다중 객체의 동시 다발적 행동 인식을 위한 방법에 대하여 제안하였다. 향후 연구과제로 논문에서 제안한 신경망을 실제로 구현하고 임의의 테스트 영상을 통해 다중 객체의 행동 인식을 잘 수행할 수 있는지 시험하고자 한다.

사사의 글

본 연구는 원자력안전위원회의 재원으로 한국원자력안전재단의 지원을 받아 수행한 원자력안전연구개발사업의 연구결과입니다. (No. 1403025)

참고문헌

[1] Karpathy, Andrej, et al. "Large-scale video classification with convolutional neural networks." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2014.

[2] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[3] 김인중. "시각 인식을 위한 딥러닝 기술의 최근 발전 동향." 정보과학회지, 33.9 (2015.9): 15-20.

[4] D. Jeffrey et al, “Long-term Recurrent Convolutional Networks for Visual Recognition and Description,” in Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston-Massachusetts: US, pp.2625-2634, June. 2015.

[5] 김원준, 김창익. "관심영역 검출 기술과 적용사례." 전자공학회지, 39.2 (2012.2): 20-28.