오브젝트 정보를 이용한 다중 영역 기반 방사형 레이어 GCN의 개선 방안

장한별, 이칠우
전남대학교 전자컴퓨터공학과
e-mail : cerrant@gmail.com, leecw@jnu.ac.kr

The suggestion of improvement about multi region based radial layer GCN with object information

Han-byul Jang, Chil-woo Lee
Department of Electronics and Computer Engineering

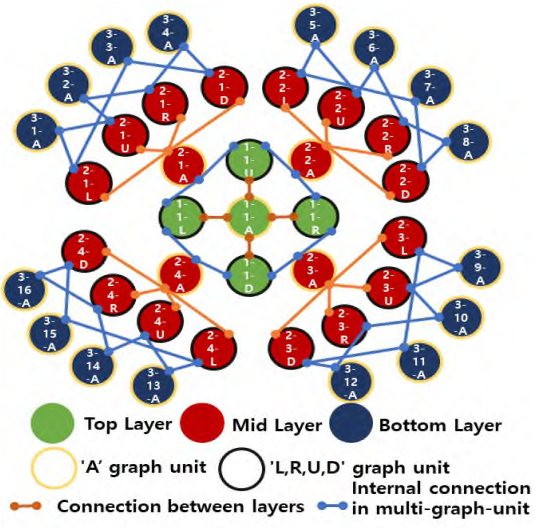
요 약

인간 행동 인식은 자동으로 위험 행동을 인지 할 수 있으므로, 지능형 감시 시스템에서 핵심적인 요소가 될 수 있다. 그러나 행동 인식 연구에서 일반적으로 사용되어온 스켈레톤 데이터는 감시 카메라를 통해 취득하기 어렵기 때문에, 선행 연구[1]에서는 오브젝트의 옵티컬 플로우와 이미지 그래디언트 정보에 기반한 다중 영역 기반 방사형 레이어 GCN(MRGCN) 알고리즘을 제안하였다. 그러나 MRGCN 알고리즘은 영상에서 사람 영역만을 검출하여 사용하므로 그 사람과 상호작용이 필요한 다른 물체(예를 들어 운동도구 등)가 등장하는 경우 행동들을 인식하지 못하는 단점이 있다. 이 논문에서는 이런 문제를 개선하기 위해서, Yolo 알고리즘을 사용하여 얻을 수 있는 오브젝트의 종류와 영역 정보를 MRGCN에 추가한 방법에 대해 기술 한다. 제안한 방법은 오브젝트의 영역 정보를 MRGCN 노드의 데이터 생성영역에 대해 상대적인 위치와 크기로 변환하여 기존의 데이터와 잘 융합될 수 있도록 변환하였고, 오브젝트 정보의 추가를 통해 기존의 이미지 그래디언트에 포함된 오브젝트 형태 정보를 보다 잘 표현할 수 있도록 확장하였다. 오브젝트 정보가 추가된 MRGCN 알고리즘은 도구를 사용하는 행동들에 대해서 보다 개선된 인식률을 얻을 수 있을 것으로 기대된다.

1. 서 론

인간 행동 인식은 영상 속 인물의 행동을 자동으로 판별하는 기술로 많은 자동화 시스템에서 중요하게 사용될 수 있다. 특히 지능형 감시 시스템에서 사용될 경우 위험 행동을 자동으로 감지하여 빠른 대처를 가능하게 하므로 핵심적인 역할을 할 수 있다. 그러나 현재의 행동 인식 연구는 감시 카메라를 통해서 정확한 취득이 불가능한 스켈레톤 데이터를 입력 데이터로 사용하는 경우가 일반적이기 때문에 지능형 감시 시스템에서 사용하기 부적합한 경우가 많았다.

이 논문의 선행 연구[1]에서는 이러한 기존의 행동 인식 연구를 보완하기 위해 2차원 영상을 통해 쉽게 취득할 수 있는 옵티컬 플로우와 이미지 그래디언트 를 결합하여 입력 데이터로 사용한 GCN 기반의 새로운 행동 인식 알고리즘을 구현하였다. 선행 연구에서 발표된 다중 영역 기반 방사형 레이어 GCN(MRGCN:multi region based radial layer graph convolutional network)은 카메라 영상에서 사람의 몸 영역을 검출한 뒤, 그 안에 그래프 노드의 입력 데이터 취득을 위한 영역(RoM:region of motion)들을 체계적으로 배치하였다. 이때 그래프 노드들을 펼친 그림으로 그려 보면 그림 1과 같이 방사형으로 배치된 레이어 구조를 보이게 되므로 MRGCN이라는 이름을 붙였다. MRGCN은 취득된 옵티컬 플로우와 이미지 그래디언트 데이터를 32차원의 방향각에 따라 누적한 히스토그램으로 가공한 뒤 다시 6차원의 압축된 정보로 변환하여 보다 효율적인 신경망 학습을 구현하였으며, 입력 데이터의 특징에 따라 새롭게 설계한 그래프 연결을 통해 스켈레톤 데이터를 사용한 연구 이상의 행동 인식 결과를 얻을 수 있었다.

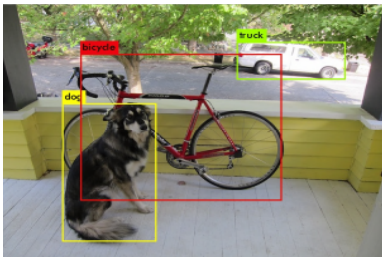


(그림 1) MRGCN 그래프 전개도(출처[1])

그러나 MRGCN 연구는 사람의 몸 영역에서 취득된 정보만을 사용하므로, 다른 오브젝트를 사용하는 행동은 잘 분석하기 어렵다는 단점을 갖고 있다. 사람이 하는 많은 행동 들이 도구 등 오브젝트를 사용하기 때문에, 오브젝트의 종류, 크기, 형태, 위치 등의 정보가 알고리즘에 사용된다면 인식률을 보다 개선할 수 있을 것이다. 따라서 이 논문은 MRGCN 알고리즘과 오브젝트 정보를 결합한 개선 방법에 대해 기술한다. 그림 2와 같이 영상에 등장하는 오브젝트는 Yolo 등을 이용하여 쉽게 검출 할 수 있으며, 오브젝트의 종류와 영역 박스 정보를 얻을 수 있다. 이 논문은 이러한 오브젝트 정보를 노드의 입력 데이터에 추가하는 방법을 제안한다.

2. 관련 연구

인간 행동 인식의 연구는 과거에는 패턴 인식을 이용하는 방법들이 사용되었으나, 현재는 딥러닝을 주로 사용하고 있다. 딥러닝은 대량의 데이터를 사용한 신경망 학습을 통해 기존의 방법으로 미처 찾아내지 못하는 특성을 분석할 수 있게 하기 때문에, 행동 인식과 같은 복잡한 문제를 다루기 적합하다. 특히 2018년 발표된 ST-GCN[2]은 그래프 구조의 데이터를 사용하여 학습하는 GCN(Graph convolutional network) 기법을 처음으로 행동 인식에 응용하여 획기적인 성과를 거두었다. GCN은 그래프의 연결 구조를 통해 데이터 안에 존재하는 맥락적 특징을 잘 드러낼 수 있으며, ST-GCN은 인체 형태의 그래프 연결을 이용해 행동의 발생 맥락을 학습 과정에서 잘 반영할 수 있었다. 그러나 스켈레톤 데이터를 사용한 ST-GCN은 데이터의 취득이 어려운 환경에서는 이용하기 어렵다는 단점이 있다. 이 연구의 선행 연구[1]는 이를 보완하기 위해 2차원 영상으로부터 추출할 수 있는 데이터를 대신 사용한 MRGCN 알고리즘을 구현하였다. 반면 [3]은 인체를 4개의 큰 부위로 나누어 처리하는 방법을 통해 스켈레톤 데이터의 오류를 개선한 알고리즘을 발표했다. [4]는 인체 형태의 그래프 연결 루트를 따라서만 학습이 진행되는 단점이 보완된 ASGCN(actional-structural GCN) 알고리즘을 통해 개선된 결과를 얻었으며, [5]는 그래프 학습을 최적화하는 GR-GCN(graph regression based GCN) 알고리즘을 제안하였다.

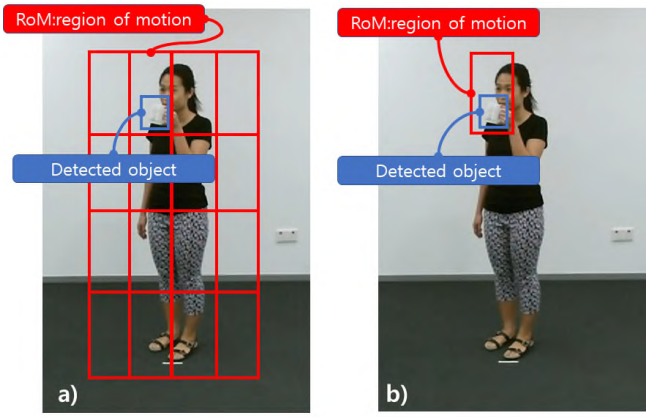


(그림 2) Yolo를 사용한 물체 인식
(출처:"https://pjreddie.com/darknet/yolo/")

3. 오브젝트 정보를 이용한 MRGCN 알고리즘의 개선 방안

그림 2와 같이 Yolo는 오브젝트를 인식하여 그 종류와 영역 정보를 출력한다. 이 장에서는 MRGCN의 학습에 행동과 연관된 오브젝트를 Yolo로 인식하고 그 종류, 영역, 형태 정보를 신경망과 결합 하는 방법을 기술한다. 그림 3-a와 같이 MRGCN은 다양한 데이터 취득 영역을 격자 형태로 사람의 몸 위에 배치한다. 하나의 데이터 취득 영역에서 한 그래프 노드의 입력 데이터를 생성하는데, 만약 그림 3-b와 같이 데이터 취득 영역과 오브젝트 영역이 겹친다면 그 노드는 오브젝트와 연관성을 갖고 있다고 생각할 수 있을 것이다.

따라서 겹쳐진 오브젝트의 정보를 노드의 입력 데이터에 추가함으로서 오브젝트 정보가 반영된 신경망 학습을 구현할 수 있다. 또한 MRGCN의 입력 데이터는 이미지 그래디언트 정보를 통해 데이터 취득 영역 안의 형태 정보를 사용하고 있기 때문에, 겹쳐진 오브젝트의 형태 정보도 신경망에 반영되고 있다고 볼 수 있다.



(그림 3) MRGCN의 RoM배치와 오브젝트(영상 출처:[6])
a) 최하단 레이어의 RoM 배치와 인식된 오브젝트(컵)
b) 오브젝트(컵)이 겹쳐진 RoM

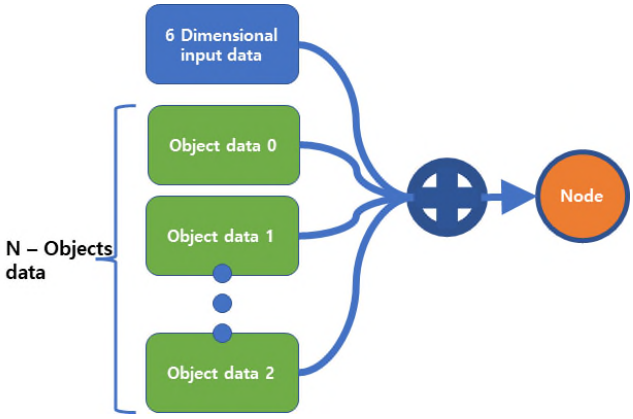
겹쳐진 오브젝트의 정보는 그 오브젝트의 종류와 영역이다. 이때 영역의 중심점 좌표와 가로, 세로 크기를 구한 후 다시 식(1), (2)를 사용하여 MRGCN 알고리즘에 융합할 수 있는 형태로 변환한다. 식(1), (2)에서 objectcx, objectcy는 오브젝트의 중심 위치 좌표를, RoMcx, RoMcy는 데이터 취득 영역(ROM)의 중심 위치 좌표를 의미한다. 또한 objectwidth와 objectheight는 오브젝트의 가로, 세로 길이를 의미하며, RoMwidth와 RoMheight는 데이터 취득 영역의 가로, 세로 길이를 의미한다. 즉 계산식을 통해 얻어진 objectdistx, objectdisty는 오브젝트와 데이터 취득 영역과의 거리를 의미한다. 또한 objectwidth2, objectheight2는 데이터 취득 영역의 크기에 대해 상대적 크기로 변환된 오브젝트의 영역의 가로, 세로 길이 정보이

다. 또한 오브젝트의 종류인 objecclass 코드를 더하여 하나의 그래프 노드와 연관된 오브젝트 i의 정보는 다음 식 3과 같이 표현한다. 그래프 노드의 데이터 취득 영역은 여러개의 오브젝트와 겹쳐질 수 있으므로, 식 3의 $objectinfo_i$를 오브젝트의 크기가 큰 순서대로 N개 만큼 기존의 입력 데이터에 추가한다. 즉 오브젝트 정보가 추가된 입력 데이터의 차원은 (6+5*N)의 크기가 된다.

$$objectdistx = \frac{\sqrt{(objectcx - RoMcx)^2}}{RoMwidth}, \quad objectdisty = \frac{\sqrt{(objecty - RoMy)^2}}{RoMheight} \qquad (식\ 1)$$

$$objectwidth2 = \frac{objectwidth}{RoMwidth}, \quad objectheight2 = \frac{objectheight}{RoMheight} \qquad (식\ 2)$$

$$objectinfo_i = (objectdistx_i, objectdisty_i, objectwidth_i, objectheight_i, objectclass_i) \qquad (식\ 3)$$



(그림 4) 그래프 노드에서의 오브젝트 정보 추가

4. 결론과 향후 과제

사람의 행동은 도구 등의 오브젝트를 사용하는 경우가 많기 때문에, 오브젝트 정보를 사용하면 보다 개선된 행동 인식을 구현 할 수 있다. 이 논문은 선행 연구에서 발표된 MRGCN 알고리즘에 오브젝트 정보를 추가한 개선 방안을 기술하였다. 기존의 MRGCN은 오브젝트 정보를 사용하지 않기 때문에 도구를 사용하는 행동을 잘 구분하지 못하였으나, 제안된 방법을 사용하면 도구를 사용하는 행동들을 보다 잘 처리 할 수 있을 것으로 기대된다.

추가된 오브젝트 정보들은 MRGCN 노드의 데이터 취득 영역에 대한 오브젝트의 상대적 위치 크기 정보와 오브젝트의 종류이다. MRGCN의 노드 데이터는 이미지 그래디언트를 통해 데이터 취득 영역의 형태 정보를 갖고 있으므로, 겹쳐진 오브젝트의 형태 정보도 학습에 사용되고 있다고 할 수 있다.

논문의 개선 방안은 아직 검증 실험이 수행되기 전 상태이다. 실험을 위해 Yolo 알고리즘을 이용할 예정이므로, 개선 방안의 정확한 성능 검증을 위해서 먼저 Yolo의 인식 가능 오브젝트 들과 성능을 검증할 필요가 있다. 또한 오브젝트가 등장하는 새로운 데이터 셋을 생성하여, 제안된 알고리즘의 인식률을 평가하는 과정이 남아있다.

감사의 글

본 연구는 문화체육관광부 및 한국컨텐츠진흥원의 2021년도 문화기술연구개발 지원사업으로 수행되었음. (R2020060002)

참 고 문 헌

[1] Han-Byul Jang, Chil-Woo Lee, “A human action recognition based on MRGCN using overlapped data acquisition regions”, SMA 2021: The 10th International Conference on Smart Media and Applications. 2021.

[2] Yan, Sijie, Yuanjun Xiong, and Dahua Lin. "Spatial temporal graph convolutional networks for skeleton-based action recognition." Thirty-second AAAI conference on artificial intelligence. 2018.

[3] Thakkar, Kalpit, and P. J. Narayanan. "Part-based graph convolutional network for action recognition." arXiv preprint arXiv:1809.04983 (2018).

[4] Li, Maosen, et al. "Actional-structural graph convolutional networks for skeleton-based action recognition." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[5] Gao, Xiang, et al. "Optimized skeleton-based action recognition via sparsified graph regression." Proceedings of the 27th ACM International Conference on Multimedia. 2019.

[6] Shahroudy, Amir, et al. "Ntu rgb+ d: A large scale dataset for 3d human activity analysis." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.