

경량화된 행동 인식 분류 알고리즘을 위한 시계열 데이터 변환 및 특징 추출 시스템 제안

박재한, 이만희, 신수용

초연결 기반 협력형 무인자율이동체 연구센터, IT융복합공학과

국립금오공과대학교

rimitpark@kumoh.ac.kr, fordmore@kumoh.ac.kr, wdragon@kumoh.ac.kr

Time-Series Data Transformation and Feature Extraction System for a Lightweight Action Recognition Algorithm

Jae Han Park, Lee Man Hee, Soo Young Shin

Super Connected and Collaborative Unmanned Autonomous Mobility Research Center, Department of IT Convergence Engineering

Kumoh National Institute of Technology

요약

본 논문은 행동 인식 데이터 분류를 위한 CNN 기반 시간특성 분석 시스템을 제안한다. 제안된 시스템은 시계열 특성 분석을 위하여 시계열 데이터의 2차원 변환 알고리즘을 사용한다. 입력된 데이터는 2차원 이미지 데이터로 변환되어 CNN 입력으로 사용되며, 기존 RNN 혹은 Attention 알고리즘과 달리 고정된 연산량을 사용한다. 이런 변환 입력을 통하여 경량화된 행동 인식 데이터 분류를 시도한다.

I. 서론

최근 Teansformer 기반의 Attention 알고리즘을 사용한 다양한 비디오 분석 기법이 등장했다.[1][2] Transformer 기반 모델의 경우 입력 데이터를 시간 단위로 분리하여 Attention 알고리즘의 입력으로 사용하며, 모든 입력 요소 사이의 상관관계를 파악하기 때문에 매우 높은 분석 정확도를 보인다. 하지만 이런 특성은 입력 길이에 따라서 높은 연산량을 필요로 하는 문제가 있다. 비디오 데이터 기반의 행동 인식(Action Recognition) 분야 또한 많은 성능 향상이 이루어졌으나, 동일한 연산 효율 문제를 겪고 있다. 본 논문에서는 시계열 데이터를 GAF(Gramian Angular Field) 알고리즘을 사용하여 이미지화 하고, 경량화된 device에서도 사용 가능한 only-CNN(Convolution Neural Networks) 기반 행동 인식 데이터 분류 딥러닝 모델을 제안한다.

II. 본론

그림 1은 본 논문의 시스템에서 사용된 데이터 변환 예시를 나타낸다. 비디오에서는 공간 특성을 포함하는 프레임 2개를 추출하고, t 축에 대하여 x변화와 y 변화를 확인할 수 있는 시간 평면 2개로 변환한다. 변환된 시간 평면은 시간의 진행에 따라 변화하는 x값 혹은 y값을 가지는 1차원 데이터로 변형 가능하며, 1차원 데이터를 기반으로 GAF 알고리즘을 사용하여 2차원 이미지로 변환한다. 기존의 시계열 변화를 분석하기 위해서는 비디오를 프레임 단위로 분리하여 CNN을 사용하여 프레임의 특성을 추출하고 그 정보들은 Attention이나 RNN(Recurrent Neural Network) 계열의 모델 입력으로 사용하는 입력 길이에 따라 변화하는 구조이다. 본

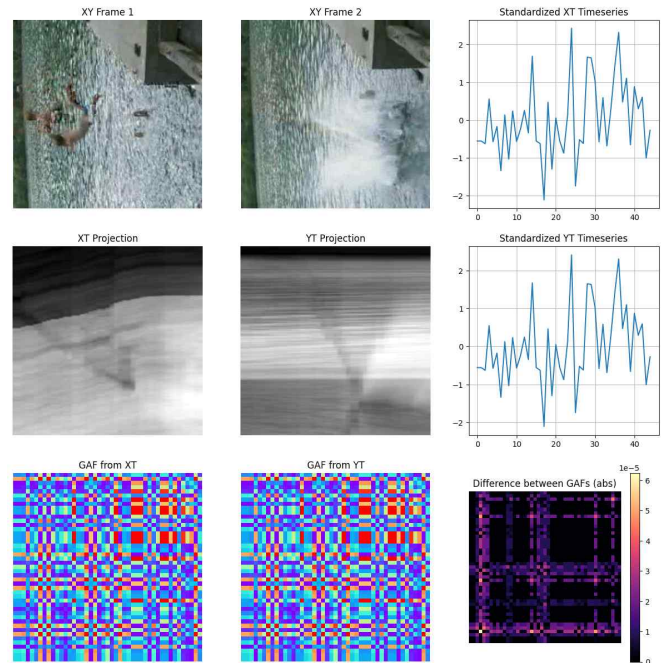


그림 1 비디오 데이터세트 HMDB51 다이빙 비디오 변환 예시 시스템에서는 비디오의 시간축에 따라 변화하는 x, y 데이터를 GAF 알고리즘의 입력으로 사용하여 2차원 이미지로 재구성한다. 이렇게 생성된 이미지는 데이터의 시간적 특성을 포함하고 있어, 공간적 특징 분석에 특화된 CNN 모델의 입력으로 사용 가능하다.[3][4][5] 또한, 시계열 데이터를

이미지로 변환하였기 때문에 기존의 다양한 이미지 증강(Augmentation) 기법을 그대로 적용할 수 있다는 장점이 있다.

이렇게 변환된 GAF 이미지(시간적 특징)는 원본 비디오 프레임(공간적 특징)에서 추출된 정보와 함께 CNN 모델의 입력으로 사용된다. CNN은 GAF 이미지의 고유한 텍스처와 패턴을 분석하여 원본 좌표 데이터의 동적 특성을 학습한다. 예를 들어, 주기적인 움직임은 GAF 이미지에서 반복적인 패턴으로 나타나며, 빠르고 급격한 움직임은 고주파(high-frequency) 텍스처로 표현된다. 모델은 이러한 시각적 단서들을 통해 움직임의 속도, 주기성, 방향 전환과 같은 복잡한 시간 정보를 포착한다. 이렇게 추출된 시간적 특징 벡터는 공간적 특징 벡터와 결합 되며, 모델은 이 통합된 정보를 바탕으로 비디오의 클래스를 최종적으로 판별한다.

III. 결론

본 논문에서는 Attention 기반 모델의 높은 연산량 문제로 인해 행동 인식 시스템을 저사양 기기에 구현하기 어려운 점을 해결하고자 하였다. 이를 위해 비디오의 시간 정보를 담고 있는 1차원 데이터를 GAF 알고리즘을 사용하여 2차원 이미지로 재구성하고, 대표 프레임이 내포한 공간 정보와 함께 CNN 기반 알고리즘으로 처리하는 경량화된 행동 인식 모델을 제안하였다.

제안된 시스템은 비디오의 모든 프레임을 처리하지 않고도 GAF 이미지를 통해 움직임, 속도, 주기성과 같은 동적 특징을 효과적으로 표현하는 방법을 시도하였다. 이를 통해 기존 CNN 기반 이미지 분석 기술의 장점을 사용하여 전체적인 모델 경량화를 달성하며, 모바일, 임베디드 시스템 등의 경량 환경에서 비디오 분석 및 분류를 시도하였다.

하지만, 본 연구는 단일 객체나 특정 지점의 좌표를 추적하는 것을 전제로 하며, x, y의 변화량이 압축된 1차원 데이터를 사용하므로 여러 객체가 복잡하게 상호작용하는 영상에서는 특징 추출에 한계가 있을 수 있다. 현재는 다중 객체의 움직임과 x,y의 변화량을 압축하지 않은 시계열 데이터를 GAF나 다른 이미지 인코딩 기법으로 시각화하여, 더 복잡하고 섬세한 행동을 인식하는 모델로 확장하는 연구를 진행 중이다.

ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 ICT혁신인재 4.0 사업의 연구결과로 수행되었음" (IITP-2025-RS-2022-00156394, 50%)

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 연구결과로 수행되었음" (IITP-2025-RS-2024-00437190, 50%)

참 고 문 헌

[1] Han, Kai, et al. "A survey on vision transformer." IEEE transactions on pattern analysis and machine intelligence 45.1 (2022): 87-110.

[2] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[3] He, Kaiming, et al. "Identity mappings in deep residual networks." Computer Vision - ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11 - 14, 2016, Proceedings, Part IV 14. Springer International Publishing, 2016.

[4] Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[5] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.