

Joint Angle Feature 기반 딥러닝 LSTM 을 이용한 일상 휴먼 행동 인식

정단비, 오지현, Ismael Espinoza, 최용혁, 이강산, 김태성
경희대학교 전자정보융합공학과

dbingsu@khu.ac.kr, dhwlgjs3@khu.ac.kr, inespinoza24@khu.ac.kr, hope5090@khu.ac.kr,
kangsanl@khu.ac.kr, tskim@khu.ac.kr

Daily Human Activity Recognition Using Joint Angle Feature Based LSTM

Danbi Jung, Ji-Heon Oh, Ismael Espinoza, Yong-Hyeok Choi, Kangsan Lee, Tae-Seong Kim

Dept. of Electronics and Information Convergence Engineering, Kyung Hee Univ.

요 약

Human Activity Recognition (HAR) 기술은 Human Robot Interaction (HRI) 분야에서 핵심적인 요소 기술로 주목받고 있다. 본 논문에서는 일상 휴먼 행동의 텡스 영상에서 추출한 joint position feature 를 joint angle feature 로 변환하고, 딥러닝 LSTM 학습기를 학습하여 38 종의 일상행동에 대하여 92.89%의 인식률을 획득하였다.

I. 서 론

Human Activity Recognition (HAR) 기술은 인간의 동작 또는 행동을 인식하는 기술로, 최근 로봇의 HRI 분야에 활용 가능성이 높다. 기본적인 HAR 기술은 텡스 영상으로부터 joint position 정보를 추출하고, 딥러닝 학습기를 이용하여 이루어진다 [1]. 그러나 joint position 정보는 3 차원 공간에서 거리 및 인체 크기에 따라 변동한다. 본 연구에서는 거리와 인체 크기에 불변한 joint angle 정보에 따른 HAR 를 수행하였다

본 논문에서는 38 종의 일상 휴먼 행동의 joint position 정보를 joint angle 정보로 변환하고, 변환된 정보를 딥러닝 LSTM 모델을 학습시켜 행동 인식 정확도를 평가하였다.

II. 본론

Joint angle feature 기반 LSTM HAR 시스템의 전체 구성도는 그림 1 과 같다. 텡스 영상에서 추출한 joint position 정보를 joint angle 정보로 변환하고, 딥러닝 LSTM 학습기를 학습 후, 인식한 휴먼 행동 결과를 산출한다.

A. ETRI-Activity3D Dataset

본 연구에 활용된 데이터는 ETRI-Activity3D [2]으로 Azure Kinect Camera 기반 일상 휴먼 행동 3D RGB-D 영상 데이터베이스이다. 이 데이터베이스는 사람 몸의 25 개 joint 의 3D position (x, y, z), depth image 의 좌표 $D(u, v)$, 그리고 orientation 좌표 (l, m, n, w)로 구성되어 있다. 총 55 종의 일상 휴먼 행동으로 구성되었으며, 본 연구에서는 38 종 activity 의 3D joint position 데이터를

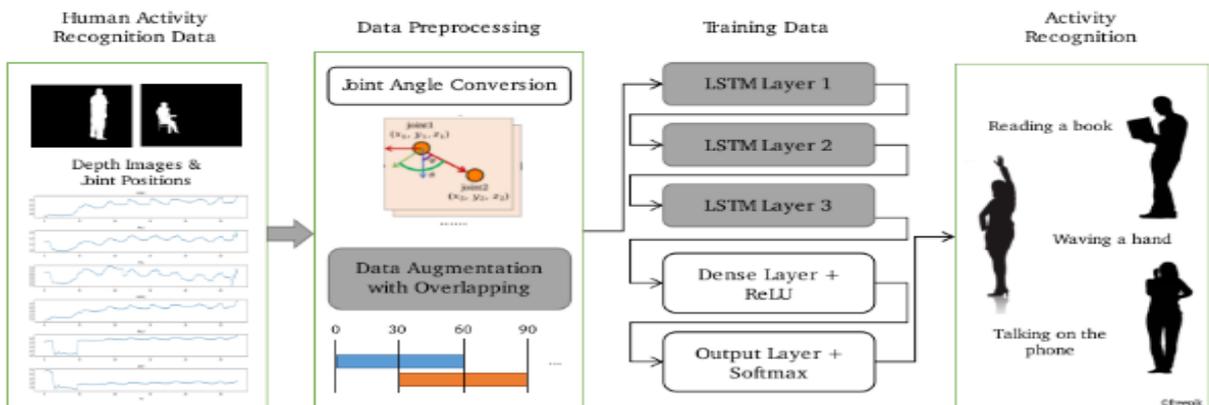


그림 1 Joint Angle Feature 기반 LSTM Human Activity Recognition 시스템

활용하였다. 38 종 일상 휴먼 행동은 표 1 과 같이 hand activities 19 종, upper-body activities 11 종, 그리고 whole-body activities 8 종으로 구성하였다.

표 1. 38 종 Daily Human Activities

| Category | Daily Human Activity |
|------------------------------|---|
| Hand Activities (19 종) | 수저 또는 포크로 음식 집어먹기, 물 또는 음료를 컵에 따르기, 약 먹기, 채소 다듬기, 과일 깎기, 이빨 닦기, 화장품 바르기, 립스틱 바르기, 머리 빗기, 리모컨으로 TV 컨트롤하기, 책 읽기, 글 쓰기, 스마트폰 조작하기, 컴퓨터 키보드 치기, 담배 피기, 박수 치기, 이리 오라고 손짓하기, 손가락으로 가리키기, 손을 좌우로 흔들기 |
| Upper-body Activities (11 종) | 세수하기, 수건으로 얼굴/머리 닦기, 머리 드라이기로 말리기, 설거지하기, 식탁을 행주로 닦기, 물건을 찾기 위해 두리번거리기, 신문 보기, 전화 걸거나 받기, 두 손으로 얼굴 비비기, 목 돌리기 운동 하기, 어깨 셀프 안마 하기 |
| Whole-body Activities (8 종) | 상의 입기, 진공청소기 사용하기, 걸레로 엮드려서 바닥 닦기, 창문이나 가구 등 닦기, 빨래 널기, 맨손체조 하기, 쓰러지기, 서 있다가 눕기 |

B. Data Preprocessing

Human activity 의 joint position feature 에서 인접한 joint 사이의 벡터를 계산하여 14 개의 Joint vector 를 추출하였다. Joint angle feature 는 추출된 joint 벡터의 두 회전각 (ϕ)와 (θ)로 구성되었다. ϕ 와 θ 를 구하는 수식은 다음과 같다[3].

$$\phi = \arccos\left(\frac{z_2 - z_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}}\right) \quad (1)$$

$$\theta = \arctan\left(\frac{x_2 - x_1}{y_2 - y_1}\right) \quad (2)$$

본 논문에서는 각 행동의 epoch data 를 2 초, 60 frames 로 정의한다. 38 종 행동에 대해 2 초 단위로 자르고 50%의 overlapping 후, one-hot encoding 방식으로 data augmentation 을 수행하였다.

C. LSTM 모델

본 연구에서 LSTM 모델은 3 개의 LSTM layer 와 2 개의 dense layer 로 구성된다. Dense layer 의 activation function 은 ReLU 를 사용하고, output layer 에서 Softmax activation function 을 통해 38 종 human activity 다중 분류를 수행한다. LSTM 모델은 training data 와 test data 는 8:2 의 비율로 나누어 LSTM 모델에 학습과 테스트를 진행하였다. 인식률은 joint position 대 joint angle feature 기반 정확도를 비교 평가하였다.

D. 연구결과

일상 행동 인식 모델의 결과는 표 2 와 같다. 세 그룹의 휴먼 행동에 대하여, joint position 특징에 기반한 평균 인식률은 89.04%인 반면, joint angle 특징에 기반한 평균 인식 정확도는 92.89%로 성능이 향상됨을 확인하였다.

표 2. Joint Position Feature 와 Joint Angle Feature 를 사용했을 때 Activity 에 따른 각각의 평균 Accuracy

| Daily Human Activities(종) | Accuracy | |
|----------------------------|----------------|-------------|
| | Joint Position | Joint Angle |
| Hand Activities (19) | 87.05% | 94.43% |
| Upper-body Activities (11) | 88.4% | 93.07% |
| Whole-body Activities (8) | 91.67% | 91.17% |
| Average (38) | 89.04% | 92.89% |

III. 결론

본 논문에서는 38 종의 일상 행동을 평균 인식률 92.89%로 인식하고, joint angle feature 의 효용성을 검증하였다. 특히 joint angle feature 는 parameter 개수를 joint position feature 의 75 개에서 28 개로 약 63% 감소시켜, 효율적인 학습을 가능하게 하였다.

ACKNOWLEDGMENTS

본 연구는 한국연구재단 중견연구사업 (NRF-2023R1A2C100358511), 산업통상자원부 사회적약자 자립지원로봇기술개발사업 (RS-2023-00232141), 과학기술정보통신부 및 정보통신기획평가원 대학 ICT 연구센터사업 (IITP-2024-RS-2024-00438239)의 연구결과로 수행되었음.

참고 문헌

- [1] Md. Milon Islam, et al., "Multi-level feature fusion for multimodal human activity recognition in Internet of Healthcare Things," Elsevier Science Publishers B. V., pp. 17-31., June 2023.
- [2] Jang, Jinhyeok, et al., "ETRI-activity3D: A large-scale RGB-D dataset for robots to recognize daily activities of the elderly," 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS.), pp 10990-10997., March 2020.
- [3] Fan, Bingfei, et al. "Effects of IMU sensor-to-segment misalignment and orientation error on 3-D knee joint angle estimation," 2021 IEEE Sensors Journal., pp 2543-2552, Feb 2022.