

특집논문 (Special Paper)

방송공학회논문지 제23권 제5호, 2018년 9월 (JBE Vol. 23, No. 5, September 2018)

<https://doi.org/10.5909/JBE.2018.23.5.598>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

스켈레톤 벡터 정보와 RNN 학습을 이용한 행동인식 알고리즘

김 미 경^{a)*}, 차 의 영^{b)}

Using Skeleton Vector Information and RNN Learning Behavior Recognition Algorithm

Mi-Kyung Kim^{a)*} and Eui-Young Cha^{b)}

요 약

행동 인식은 데이터를 통해 인간의 행동을 인식하는 기술로서 비디오 감시 시스템을 통한 위험 행동과 같은 어플리케이션에 활용되어 질 수 있다. 기존의 행동 인식 알고리즘은 2차원 카메라를 통한 영상이나 다중모드 센서, 멀티 뷰와 같은 장비를 이용한 방법을 사용하거나 3D 장비를 이용하여 이루어져 왔다. 2차원 데이터를 사용한 경우 3차원 공간의 행위 인식에서는 가려짐과 같은 현상으로 낮은 인식율을 보였고 다른 방법은 복잡한 장비의 구성이나 고가의 추가적인 장비로 인한 어려움이 많았다. 본 논문은 RGB와 Depth 정보만을 이용하여 추가적인 장비 없이 CCTV 영상만으로 인간의 행동을 인식하는 방법을 제안한다. 먼저 RGB 영상에서 스켈레톤 추출 알고리즘을 적용하여 관절과 신체부위의 포인트를 추출한다. 이를 식을 적용하여 변위 벡터와 관계 벡터를 포함한 벡터로 변형한 후 RNN 모델을 통하여 연속된 벡터 데이터를 학습한다. 학습된 모델을 다양한 데이터 세트에 적용하여 행동 인식 정확도를 확인한 결과 2차원 정보만으로 3차원 정보를 이용한 기존의 알고리즘과 유사한 성능을 입증할 수 있었다.

Abstract

Behavior awareness is a technology that recognizes human behavior through data and can be used in applications such as risk behavior through video surveillance systems. Conventional behavior recognition algorithms have been performed using the 2D camera image device or multi-mode sensor or multi-view or 3D equipment. When two-dimensional data was used, the recognition rate was low in the behavior recognition of the three-dimensional space, and other methods were difficult due to the complicated equipment configuration and the expensive additional equipment. In this paper, we propose a method of recognizing human behavior using only CCTV images without additional equipment using only RGB and depth information. First, the skeleton extraction algorithm is applied to extract points of joints and body parts. We apply the equations to transform the vector including the displacement vector and the relational vector, and study the continuous vector data through the RNN model. As a result of applying the learned model to various data sets and confirming the accuracy of the behavior recognition, the performance similar to that of the existing algorithm using the 3D information can be verified only by the 2D information.

Keyword : skeleton, feature vector, RNN, SELU, deep learning

I. 서론

오늘날 우리가 생활하고 있는 일상에서는 위험 요소들이 산재 되어 있어 비디오 감시 시스템을 통한 사회 안전 증대를 위해 많은 장소에 CCTV 감시 카메라를 구축하였지만 많은 양의 영상을 사람의 집중과 판단에 의존하는 기존의 시스템은 관리 소홀로 인한 취약점을 가진다^[1]. 이를 해결하기 위해서는 주어진 데이터를 분석하여 어떠한 행동을 하고 있는지 판단하는 행동 인식에 관한 연구가 필요하다. 올바른 행동인식을 위해서는 데이터로부터 시공간적 특징을 정확하고 빠르게 추출해야 하는데 장비적인 방법과 학습의 방법에서의 측면에서 분류할 수 있다.

장비적인 측면에서 보면 예전의 행동 인식의 방법으로는 2차원 카메라로 수집한 데이터를 활용하여 이루어 졌지만 인간의 행동은 3차원 공간에서 행해지기 때문에 겹쳐지는 부분으로 인해 낮은 인식율을 보였다. 이를 해결하기 위한 방법으로 가속도, 자이로, 고도 센서로 구성된 다중 센서를 이용한 방법^[23], 다중 카메라를 이용한 방법^[24]등 추가적인 장비가 필요한 방법이 있으며, 3D-컨볼루션을 이용한 방법^[2]은 많은 수의 파라미터 학습을 위해서 높은 사양을 지닌 하드웨어가 요구되며 depth map과 3D 스켈레톤^[3] 등을 적용한 방법은 이를 측정하기 위한 kinect와 같은 추가적인 장비가 필요한 단점을 지닌다.

학습 방법 측면에서 보면 예전의 기계학습(Machine Learning)방법에서는 추출할 특징들을 연구자가 직접 설계해야 하기 때문에 많은 시간이 필요하며 효율적인 특징 추출을 위해서는 어느 정도 전문적 지식이 필요하다, 이런 단점을 보완하기 위해 데이터로부터 직접 유용한 특징들을 학습하는 딥러닝 방식을 이용한 행동 인식에 관한 연구가 최근 활발히 진행되어 지고 있으며 본 논문은 딥러닝의 알고리

즘을 이용하여 학습의 성능을 높이고자 한다.

데이터를 통한 행동 인식을 정확하게 할 수 있으면 지능형 감시 시스템이나 로봇 비전, 게임 컨트롤 등 다양한 분야에서 사용 가능하고 지능형 감시 시스템의 행동 인식 성능을 높임으로써 사회 안전망 구축에 기여할 수 있다.

본 논문은 장비적인 측면에서는 추가적인 장비 구축이나, 고가의 장비가 필요한 단점을 해결하여 2차원의 RGB 영상을 사용하였으며 학습에서는 딥러닝을 통한 행동 인식의 방법을 제안한다. 영상은 2차원의 동영상을 사용하며 스켈레톤 추출 알고리즘을 딥러닝을 이용하여 학습하여 관절과 신체부위의 포인트를 추출한다. 추출된 포인트들 간의 변위 벡터와 관계 벡터를 포함한 벡터로 변형한다. 이후 RNN 모델을 통하여 연속된 벡터 데이터를 학습하여 다양한 데이터 세트에 적용시켜 행동 인식 정확도를 확인한다. 이러한 방법을 이용하여 성능은 더 좋거나 버금가는 알고리즘을 제안하였으며 기존 발표된 알고리즘들을 구현한데 의의가 있다.

본 논문의 2장은 논문에서 사용한 행동 인식 알고리즘에 관한 설명을 하고 3장은 행동 데이터 세트에 알고리즘 적용 후 실험 결과에 관해 기술한다. 마지막으로 4장에서 결론을 기술한다.

II. 실험에서 적용한 행동인식 알고리즘

1. 스켈레톤 추출 알고리즘

스켈레톤을 추출하기 위하여 Eldar et al.(2017)의 ArtTrack^[5]을 사용하여 영상에서 사람의 각 부위와 관절의 포인트를 추출하였다. 알고리즘은 이미지에서 관절의 위치를 포함한 여러 사람을 탐지하고 포즈를 추정하는 것을 목표로 한다. 장면에서 사람의 수를 먼저 추론하고 신체부위를 식별하는 방법으로, 사람들을 먼저 감지하고 이어서 몸자세를 추정하는 방법과 대조적이다. 알고리즘은 CNN기반 부분 탐지기로 생성된 신체 부분 가설 세트의 분할 및 레이블링 공식을 제안 하였다. 여러 사람일 경우, 단일 인물의 경우 모두 ArtTrack을 이용한 인물의 탐지 및 포즈 추정할 수 있는 코드는 GitHub에서 다운로드 할 수 있으며 이를

a) 부산대학교 전자공학과(Dept. of Electronics Engineering, Pusan National University)

b) 부산대학교 정보컴퓨터공학과(Dept. of Computer Science and Engineering, Pusan National University)

† Corresponding Author : 김미경(Mi-kyung Kim)

E-mail: ddosun@ousan.ac.kr

Tel: +82-51-510-2878

ORCID: <https://orcid.org/0000-0002-3884-1812>

※ 이 논문의 연구결과 중 일부는 "IPTU 2018"에서 발표한 바 있음.

※ Manuscript received April 23, 2018; Revised August 24, 2018; Accepted August 31, 2018.

이용하여 관절 포인트 추출하여 그림 1과 같은 결과를 얻었다.

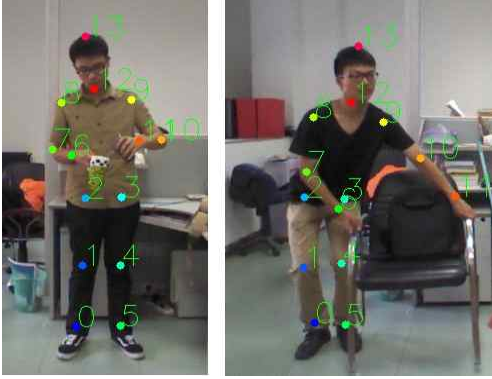


그림 1. 사람의 포인트 추출 예시 그림
Fig 1. Person's point extraction example

2. 특징 벡터 추출

사람마다 같은 행동임에도 키와 신체 부위별 크기에 따른 차이에 의해 추출된 스켈레톤 특징은 사람마다 다른 값을 가진다. 각 행동의 공통된 값을 얻기 위해 Diago et al(2017)이 제안한 방법^[6]을 적용하여 관절과 부위별 포인트를 관계 벡터와 변위 벡터값으로 변형한다. 스켈레톤 추출 알고리즘의 포인트 수와 Diago et al의 사람의 포인트 수가 다르므로 이를 반영하여 관계 벡터와 변위 벡터 값으로 변환하였다.

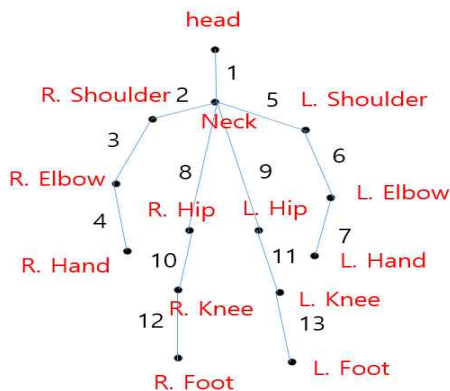


그림 2. 사람의 신체부위별 그림
Fig 2. Picture of human body part

표 1. 변위 벡터를 구성하는 하위그룹

Table 1. Subgroups constituting the displacement vector

feature	Subgroups by body
f1	head, left hand, right hand, left foot, right foot
f2	neck, left elbow, right elbow, left knee, right knee
f3	left shoulder, right shoulder, left hip, right hip

표 2. 관계 벡터를 구성하는 하위그룹

Table 2. Subgroups that make up the relationship vector

feature	Subgroups by body	relationship branch
f4	head, right hand, left hand	neck
f5	neck, right hand, right foot	left heap
f6	head, left hand, left foot	right heap
f7	right hand, left hand	head

$$v_i^s = \frac{p_i^{s+1} - p_i^s}{\Delta T} \mid 1 < s < \tau \quad (1)$$

$$f1(x, y) = v_i^s(x, y) / D(1, 4, 7, 12, 13) \quad (2)$$

$$f2(x, y) = v_i^s(x, y) / D(1, 3, 6, 10, 11) \quad (3)$$

$$f3(x, y) = v_i^s(x, y) / D(2, 5, 10, 11) \quad (4)$$

식 (1)에서 s 는 영상의 프레임 번호를 의미하며 τ 는 영상의 전체 프레임 수이다. p_i^s 는 스켈레톤 포인트의 i 번째 (x, y) 좌표를 의미한다. 변위 벡터의 경우 식 1을 적용하여 구한 나중의 프레임에서 $f1, f2, f3$ 의 부위별 위치의 ($x1, y1$) 값과 먼저 프레임에서 $f1, f2, f3$ 부위별 위치 ($x0, y0$) 값의 차를 구하고($x1-x0, y1-y0$) 사람의 신체 부위별 크기 차이를 고려하여 값을 구한다. $D(1, 4, 7, 12, 13)$ 의 경우 그림 2에서 1(머리와 목 연결 선), 4(오른쪽 손과 오른쪽 팔꿈치 연결선), 7(왼쪽 손과 왼쪽 팔꿈치 연결선), 12(오른쪽 무릎과 오른쪽 발 연결선), 13(왼쪽 무릎과 왼쪽 발 연결선)의 거리의 합을 의미한다.

$$w^{s_i, k} = p_i^s - p_k^s \mid i \neq k \quad (5)$$

$$f4(x, y) = (p_i^s - p_k^s) / D(1, 3, 4, 6, 7) \quad (6)$$

$$f5(x, y) = (p_i^s - p_k^s) / D(3, 4, 10, 12) \quad (7)$$

$$f6(x, y) = (p_i^s - p_k^s) / D(6, 7, 11, 13) \quad (8)$$

$$f7(x, y) = (p_i^s - p_k^s) / D(3, 4, 6, 7) \quad (9)$$

식 (5)에서 p_i^s 와 p_k^s 는 같은 스켈레톤 내 신체 별 하위 그룹과 관계 지점의 (x, y) 좌표를 의미한다. f1 5개 포인트, 5개 f2 포인트, 4개 f3 포인트(변위 벡터), f4 3개 포인트, f5 3개 포인트, f6 3개 포인트, f7 2개 포인트(관계 벡터), 총 25개의 포인트 (x, y) 좌표에서 수식 (1),(2)를 적용하여 얻은 50개 벡터값을 저장한다.

3. RNN구조

각 프레임마다 사람의 스켈레톤에서 프레임 당 50개의 벡터로 변환하고 연속된 프레임을 모아서 데이터를 만들었다. 데이터를 training data와 validation data, test data는 4:1:1로 나누어 RNN(recurrent Neural Network)의 학습에 사용하였다. RNN 구조는 2층의 BiGRU(Bidirectional gated recurrent unit)^[7]와 학습 성능의 향상을 위해 Batch Normalization^[8], 50%의 Dropout^[9]을 적용하였다. GRU의 경우 별도의 파라미터를 씌워서 입력 frame의 수만큼, 많은 층의 back propagation시 loss에 대한 weight의 변화 값에 따른 gradient vanishing, exploding 현상을 방지함으로써 학습이 잘 이루어지게 한다. 그리고 Batch Normalization의 경우 입력데이터를 평균이 0, 분산이 1인 데이터 분포로 바꾸어 균일한 분포의 데이터로 바꾸어 loss 함수에서 global minimum으로의 수렴을 도와 학습이 잘 이루어지도록 한다. 식

10,11의 Batch Normalization은 $\gamma=1.0$, $\beta=0$, $\epsilon=10^{-5}$ 를 적용하여 초기화하였다. 그리고 optimizer로 Adam^[10]을 사용하였으며 식 12,13,14는 Adam식이다.

$$\hat{x}_i \leftarrow \frac{x_i - \mu_B}{\sqrt{\delta_B^2 + \epsilon}} \quad (10)$$

$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv BN_{\gamma, \beta}(x_i) \quad (11)$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (12)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_2^t} \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (13)$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t y_i \leftarrow \gamma \hat{x}_i + \beta \equiv BN_{\gamma, \beta}(x_i) \quad (14)$$

Adam은 학습률을 개선하기 위한 RMSprop과 gradient를 계산하는 방법을 개선하기 위한 momentum 알고리즘을 결합한 형태이다. Adam 알고리즘에서 gradient는 momentum과 이전 gradient의 이동평균으로, 학습률은 RMSprop과 마찬가지로 gradient 제곱의 이동평균에 반비례하도록 설정된다. 이 알고리즘은 hyperparameter의 선택에 덜 민감하다고 알려져 있다. Adam은 RMSprop과 momentum의 장점을 모두 지니므로 일반적인 deep learning의 학습 알고리즘에 주로 사용되고 있다.

학습 성능의 향상을 위해 활성화 함수로 ReLU(rectified

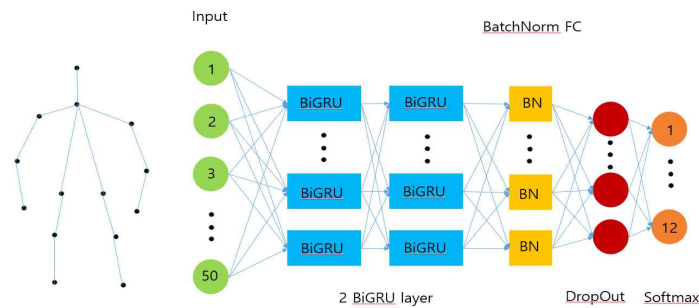


그림 3. 논문에서 사용한 RNN 구조

Fig 3. RNN structure used in the paper

linear unit)^[11]와 SELU(scaled exponential linear unit)^[12]를 적용하고 두 함수 간 성능을 비교하였다. SELU를 적용하여 학습 시 입력데이터를 정규화하는 특성을 보이므로 Batch Normalization 층을 제외하고 학습을 진행하였다. 학습 결과 SYSU 3D Human-Object Interaction Set의 경우 ReLU를 적용 시 행동 인식 정확도는 63.4%였고 SELU를 적용 시 69.3%로 성능이 향상됨을 확인하였다.

III. 실험결과 및 분석

1. Data Set 종류

1.1 Northwestern-UCLA Multiview Action 3D Dataset

위 데이터 세트는 UCLA의 JingWang와 Xicohan Nie가 만들었으며 한 손으로 물체 들기, 두 손 으로 물체 들기, 쓰레기 버리기 등 10개의 행동에 대하여 10명의 영상으로 구성되어 있다. 각 영상은 서로 다른 시점에서 촬영되었다. 데이터 세트는 RGB 영상과 깊이 정보, 스켈레톤 정보를 포함하나 본 논문에서는 RGB 영상만을 사용하여 실험을 진행하였다.

1.2. MSR-Daily Activity 3D Dataset

위 데이터 세트는 Microsoft Research Redmond에서 Wanqing Li가 작성했으며, 마시기, 먹기, 책 읽기 등 16가지의 행동으로 구성되어 있다. 데이터 세트는 각 행동을 키넥트를 사용하여 얻은 RGB 영상과 깊이 정보, 스켈레톤 정보를 포함하나 본 논문에서는 RGB 영상만을 사용하여 실험을 진행하였다.

1.3. SYSU 3D Human-Object Interaction Set

위 데이터 세트^[4]는 중산대학에서 만든 데이터셋으로 40명이 각각 12가지 행동에 대하여 자유롭게 행동하고 그 영상을 저장하였다. 행동은 핸드폰을 사용하여 전화를 받는 동작, 가방을 메는 동작 등을 포함한다. 총 480(12 x 40)개의 1.9초에서 21초 사이의 다른 프레임 길이 영상으로 구성되어 있고 대걸레질과 빗자루질 혹은 한 손으로 컵 안의 물을 마시거나 양손을 사용하여 물을 따르는 등 유사한 행

동을 담은 영상으로 구성되어 있다. 데이터 세트는 RGB 영상뿐만 아니라 깊이 정보와 스켈레톤 정보를 포함하나 본 논문에서는 RGB 영상만을 사용하여 실험을 진행하였다.

2. 실험 결과

본 논문에서는 위 3가지 데이터세트를 구성하는 각 영상의 RGB 데이터에 ArtTrack을 사용하여 스켈레톤을 추출하고 변위 벡터와 관계 벡터로 변형하였다. 그리고 샘플링을 통해 시작 지점을 정하고 시작지점에서 연속된 프레임에 이를 적용하여 결과 값을 저장하였다. Northwestern-UCLA Multiview Action 3D Dataset의 경우 32개를 샘플링하고 가장 짧은 영상의 프레임이 4 이므로 3개의 연속된 프레임을 각 샘플링 마다 저장하였다. MSR-Daily Activity 3D Dataset은 32개를 샘플링하고 가장 짧은 영상의 프레임에 따라 40개의 연속된 프레임을 각 샘플링 마다 저장하였다. SYSU 3D Human-Object Interaction Set는 40개를 샘플링하고 한 샘플링 당 50개의 연속된 프레임을 각각 저장하였다. 이후 저장된 데이터에서 RNN 모델과 활성화 함수로 SELU를 적용하여 학습을 진행하였다. 학습된 모델에서 샘플링에 관하여 각 클래스 분류 결과를 얻고 각 샘플링 결과를 클래스의 분포로 변환하였다. 얻은 클래스의 분포를 Linear SVM(Support Vector Machine)^[13]을 사용하여 학습하고 테스트 데이터의 클래스 분포를 입력으로 넣어 테스트 데이터의 SVM 출력을 최종 클래스로 지정하였다. 이후 테스트를 5번 반복시행 하여 평균 정확도를 구하였다. 성능 비교에 있는 다른 알고리즘의 경우 데이터 세트의 RGB 영상뿐만 아니라 깊이 정보와 스켈레톤 정보를 모두 포함하여 행동 인식의 정확도를 구하였으나 본 논문에서 제안한 방법은 RGB 영상만을 사용하여 행동 인식 정확도를 구하였다. 여러 정보를 사용한 다른 알고리즘과 비교하였을 시 RGB 영상만을 사용한 본 논문의 알고리즘은 비교적 우수한 행동인식 정확도를 확인할 수 있었다.

그림 4는 Northwestern-UCLA Multiview Action 3D Dataset의 실험결과로 각 행동에 대한 클래스가 1~10까지 할당되어 실험 결과값의 정확도를 나타낸 것이다. 예를 들

어 Pick up with one hand는 클래스 1로 분류되어 나온 백분율이 0.75이고 클래스 2로 나온 경우는 0.24이므로 정확도는 75%이며 10가지 행동의 평균값이 표 3의 제안한 방법의 정확도이다.

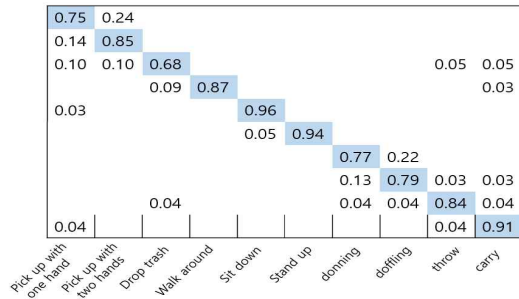


그림 4. Northwestern-UCLA Multiview Action 3D Dataset의 confusion matrix

Fig. 4. The confusion matrix of the Northwestern-UCLA Multiview Action 3D Dataset

표 3. Northwestern-UCLA Multiview Action 3D Dataset에 관한 성능 비교
Table 3. Performance Comparison of Northwestern-UCLA Multiview Action 3D Dataset

Method	Average accuracy(%)	Method	Average accuracy(%)
Virtual View ^[14]	50.7	Poselet ^[17]	54.9
Hankelet ^[15]	54.2	MST-AOG ^[18]	81.6
Action Bank ^[16]	24.6	제안한 방법	83.5

그림 5와 표 4는 MSR-Daily Activity 3D Dataset이며 성능이 MST-AOG방법보다는 수치가 낮지만 다른 세 방법보

다는 훨씬 높으므로 의미가 있다고 본다.

표 4. MSR-Daily Activity 3D Dataset에 관한 성능 비교
Table 4. Performance Comparison of MSR-Daily Activity 3D Dataset

Method	Average accuracy(%)	Method	Average accuracy(%)
STIP ^[19]	54.5	Poselet ^[17]	23.6
Action Bank ^[16]	23.0	MST-AOG ^[18]	73.1
제안한 방법	70.6		

그림 6과 표 5는 SYSU 3D Human-Object Interaction Set의 결과 값으로 정확도가 다소 낮은 듯 하나 다른 알고리즘의 경우 데이터 세트의 RGB 영상뿐만 아니라 깊이 정보와 스켈레톤 정보를 모두 포함하여 행동 인식의 정확도를 구하 것이고 본 논문에서 제안한 방법은 RGB 영상만을 사용하여 행동 인식 정확도를 구한 것이지만 수치상 비슷한 결과가 나왔으므로 의미가 있다고 본다.

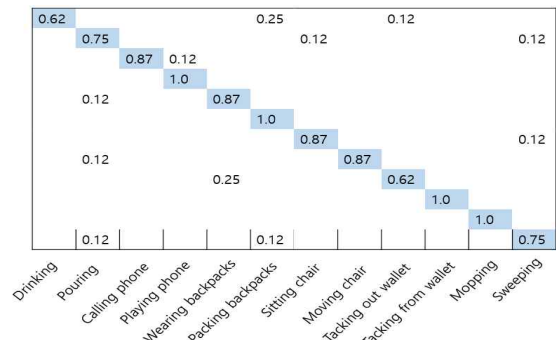


그림 6. SYSU 3D Human-Object Interaction Set의 confusion matrix
Fig 6. The confusion matrix of SYSU 3D Human-Object Interaction Set

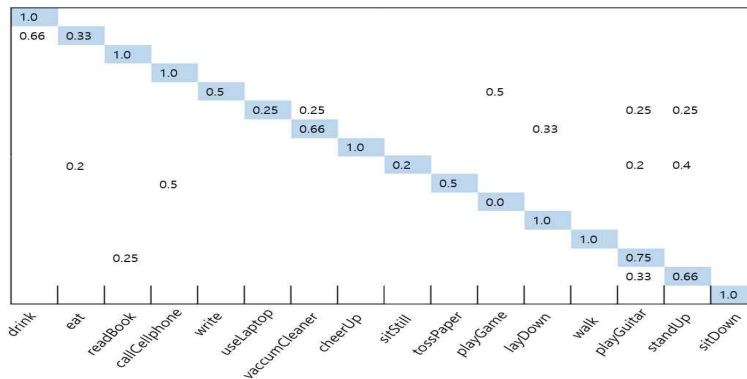


그림 5 MSR-Daily Activity 3D Dataset의 confusion matrix
Fig 5. The confusion matrix of MSR-Daily Activity 3D Dataset

표 5. SYSU 3D Human-Object Interaction Se에 관한 성능 비교
Table 5. Performance Comparison of SYSU 3D Human-Object Interaction Set

Method	Average accuracy(%)	Method	Average accuracy(%)
HON4D ^[19]	79.22	SVM ^[4]	82.78
HFM ^[20]	76.74	MTDA ^[22]	84.21
MPCCA ^[21]	80.72	JOULE ^[4]	84.89
제안한 방법	83.12		

IV. 결 론

본 논문에서는 RGB 영상 입력만으로 행동 인식이 가능한 방법을 제안하였다. 그리고 학습의 성능을 높이기 위해 활성화 함수로 SELU를 적용하고 데이터 증대를 적용하여 학습을 하고 클래스의 분포를 SVM을 적용하여 얻은 결과를 최중클래스로 지정하여 행동 인식 분류 성능을 높일 수 있었다. 데이터 세트에 스켈레톤 정보를 같이 활용한 기존의 행동 인식 알고리즘과 비교하면 더 우수하거나 비슷한 성능을 지님을 확인할 수 있었다. 본 논문의 알고리즘의 경우 벡터의 변화를 사용함으로써 관점에 따른 영상 내 행동 인식에 장점이 있으나 한 영상 내 여러 가지 행동을 하거나 여러 사람의 행동을 인식하는데 한계점을 지닌다. 앞으로 이러한 점을 보완한다면 다양한 영상 내 행동 인식에 적용할 수 있을 것이다.

참 고 문 헌 (References)

- [1] C. Jung, and D. Kang, "A Recognition Algorithm of Suspicious Human Behaviors using Hidden Markov Models in an Intelligent Surveillance System," *Journal of multimedia information system*, Vol.11, No.11, pp.1491-1500, Nov. 2008.
- [2] TRAN, Du, et al. "Learning spatiotemporal features with 3d convolutional networks." *arXiv preprint arXiv:1412.0767*, 2014.
- [3] LI, Wanqing, ZHANG, Zhengyou, LIU, Zicheng. "Action recognition based on a bag of 3d points". *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010 IEEE Computer Society Conference on. IEEE, p. 9-14.2010.
- [4] HU, Jian-Fang, et al. "Jointly learning heterogeneous features for RGB-D activity recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 5344-5352. 2015.
- [5] INSAFUTDINOV, Eldar, et al. ArtTrack: "Articulated multi-person tracking in the wild," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [6] Diogo Carbonera Luvizon, Hedi Tabia, David Picard, "Learning features combination for human action recognition from skeleton sequences," *Pattern Recognition Letters*, Volume 99, pp 13-20, ISSN 0167-8655. 2017.
- [7] CHO, Kyunghyun, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.
- [8] IIOFFE, Sergey, SZEGEDY, Christian, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *International Conference on Machine Learning*. pp. 448-456. 2015.
- [9] SRIVASTAVA, Nitish, et al. Dropout, "A simple way to prevent neural networks from overfitting," *Journal of machine learning research*, 15.1: 1929-1958. 2014.
- [10] KINGMA, Diederik; BA, Jimmy. Adam: "A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [11] NAIR, Vinod, HINTON, Geoffrey E, "Rectified linear units improve restricted boltzmann machines," *Proceedings of the 27th international conference on machine learning (ICML-10)*. pp. 807-814. 2010.
- [12] KLAMBAUER, Günter, et al. "Self-Normalizing Neural Networks," *arXiv preprint arXiv:1706.02515*, 2017.
- [13] CHANG, Chih-Chung, LIN, Chih-Jen. LIBSVM, "A library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, 2.3: 27, 2011.
- [14] LI, Ruonan, ZICKLER, Todd. "Discriminative virtual views for cross-view action recognition," *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on. IEEE, pp. 2855-2862. 2012.
- [15] LI, Binlong; CAMPS, Octavia I.; SZNAIER, Mario. "Cross-view activity recognition using hanklets," *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on. IEEE, pp. 1362-1369. 2012.
- [16] SADANAND, Sreemananth, CORSO, Jason J. Action bank "A high-level representation of activity in video,". *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on. IEEE, pp. 1234-1241. 2012.
- [17] MAJI, Subhransu, BOURDEV, Lubomir, MALIK, Jitendra. "Action recognition from a distributed representation of pose and appearance," *Computer Vision and Pattern Recognition (CVPR)*, 2011 IEEE Conference on. IEEE, pp. 3177-3184. 2011.
- [18] WANG, Jiang, et al. "Cross-view action modeling, learning and recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2649-2656. 2014.
- [19] OREIFEJ, Omar, LIU, Zicheng. Hon4d "Histogram of oriented 4d normals for activity recognition from depth sequences," *Computer vision and pattern recognition (CVPR)*, 2013 IEEE conference on. IEEE, pp. 716-723. 2013.
- [20] CAO, Liangliang, et al. "Heterogeneous feature machines for visual recognition," *Computer Vision*, 2009 IEEE 12th International Conference on. IEEE, pp. 1095-1102. 2009.

- [21] CAI, Zhuowei, et al. "Multi-view super vector for action recognition," *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. pp. 596-603. 2014.
- [22] ZHANG, Yu; YEUNG, Dit Yan. "Multi-task learning in heterogeneous feature spaces," *25th AAAI Conference on Artificial Intelligence and the 23rd Innovative Applications of Artificial Intelligence Conference*, AAAI-11/IAAI-11, San Francisco, CA, United States, 7-11 August 2011, Code 87049, Proceedings of the National Conference on Artificial Intelligence. pp. 574. 2011.
- [23] S.Shin, and J.Cha, "Human Activity Recognition System Using Multimodal Sensor and Deep Learning Based on LSTM," *Transactions of the Korean Society of Mechanical Engineers - A* 42(2), pp 111-121. 2018.2,
- [24] S.Jo, H.Kang, "Real-time object tracking in Multi-Camera environments," *Journal of Computing Science and Engineering*, 2004.10, Vol. 31, No. 2 (II),pp691-693

저 자 소 개



김 미 경

- 부산대학교 전산학과 (공학석사)
- 부산대학교 멀티미디어협동과정(공학박사)
- 부산대학교 전자공학과 강의전담교수
- 현재 : 부산대학교 소프트웨어교육센터 강의전담교수
- ORCID : <https://orcid.org/0000-0002-3884-1812>
- 관심분야 : 컴퓨터비전, 물체 추적, 기계학습, 딥러닝



차 의 영

- 서울대학교 자연대학 전자계산학 이학석사
- 서울대학교 공과대학 컴퓨터공학 박사
- 한국전자기술연구소 연구원
- 현재 : 부산대학교 공과대학 정보·컴퓨터공학과 교수
- 관심분야 : 지능형 로봇, 자율 주행, Deep Learning, Image/Signal Processing