JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 27 No. 5, pp. 47-54, May 2022
https://doi.org/10.9708/jksci.2022.27.05.047

A Study on Recognition of Dangerous Behaviors using Privacy Protection Video in Single-person Household Environments

ChaeHyun Lim*, Myung Ho Kim*

*Student, Dept. of Software, Soongsil University, Seoul, Korea
*Professor, Dept. of Software, Soongsil University, Seoul, Korea

[Abstract]

Recently, with the development of deep learning technology, research on recognizing human behavior is in progress. In this paper, a study was conducted to recognize risky behaviors that may occur in a single-person household environment using deep learning technology. Due to the nature of single-person households, personal privacy protection is necessary. In this paper, we recognize human dangerous behavior in privacy protection video with Gaussian blur filters for privacy protection of individuals. The dangerous behavior recognition method uses the YOLOv5 model to detect and preprocess human object from video, and then uses it as an input value for the behavior recognition model to recognize dangerous behavior. The experiments used ResNet3D, I3D, and SlowFast models, and the experimental results show that the SlowFast model achieved the highest accuracy of 95.7% in privacy-protected video. Through this, it is possible to recognize human dangerous behavior in a single-person household environment while protecting individual privacy.

▸ **Key words**: Deep Learning, Privacy, Action Recognition, YOLOv5, Single-person household

[요 약]

최근 딥러닝 기술의 발달로 사람의 행동을 인식하는 연구가 진행 중에 있다. 본 논문에서는 딥러닝 기술을 활용하여 1인 가구 환경에서 발생할 수 있는 위험 행동을 인식하는 연구를 진행하였다. 1인 가구의 특성상 개인의 프라이버시 보호가 필요하다. 본 논문에서는 개인의 프라이버시 보호를 위해 가우시안 블러 필터가 적용된 프라이버시 보호 영상에서 사람의 위험 행동을 인식한다. 위험 행동 인식 방법은 객체 검출 모델인 YOLOv5 모델을 활용하여 영상에서 사람 객체 검출 및 전처리 방법을 적용한 후 행동 인식 모델의 입력값으로 활용하여 위험 행동을 인식한다. 실험에는 ResNet3D, I3D, SlowFast 모델을 사용하였고, 실험 결과 SlowFast 모델이 프라이버시 보호 영상에서 95.7%로 가장 높은 정확도를 달성하였다. 이를 통해 개인의 프라이버시를 보호하면서 1인 가구 환경에서 사람의 위험 행동을 인식하는 것이 가능하다.

▸ **주제어**: 딥러닝, 프라이버시, 행동 인식, YOLOv5, 1인 가구

• First Author: ChaeHyun Lim, Corresponding Author: Myung Ho Kim
*ChaeHyun Lim (immanual1995@naver.com), Dept. of Software, Soongsil University
*Myung Ho Kim (kmh@ssu.ac.kr), Dept. of Software, Soongsil University
• Received: 2022. 02. 21, Revised: 2022. 04. 21, Accepted: 2022. 04. 21.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

최근 우리나라 1인 가구의 수가 늘어나면서 1인 가구 환경에서 발생하는 문제도 증가하고 있다. 1인 가구 환경에서 발생하는 문제는 고독사와 주거 침입 후 폭행 등이 있다[1].

1인 가구의 고독사 문제는 독거노인뿐만 아니라 20-50세 사이에서도 많이 발생하고 있다. 2020년 고독사 통계에 따르면 전체 고독사 923명 중 독거노인의 고독사 비중은 42%, 65세 미만은 58%로 고독사는 1인 가구에 주거하는 모든 연령층의 문제임을 알 수 있다[2].

1인 가구 환경에서 발생하는 또 다른 문제는 범죄가 있다. 위험 상황 발생 시 주변에 도움 요청이 다른 가구 형태에 비해 어려운 1인 가구를 대상으로 하는 범죄가 매년 증가하고 있고, 1인 가구 밀집지역이 비밀집지역보다 범죄율이 약 2배 높다는 연구 결과도 있다[3].

기존의 연구들은 이러한 문제들을 해결하기 위해 센서 정보나 영상 정보를 활용하여 위험 행동을 인식한다. 하지만 센서 정보의 경우에는 한 사람에 대한 행동만을 인식할 수 있으므로 주거 침입 후 공경 행동 등 범죄에 대한 문제는 인식할 수 없다. 또 영상 정보의 경우에는 노출되지 말아야할 개인의 프라이버시가 노출될 수 있기 때문에 이를 해결해야 한다. 이에 1인 가구 환경에서 프라이버시 보호 영상을 활용한 위험 행동을 감지하는 연구가 필요하다.

본 논문의 구성은 다음과 같다. 2장에서는 사람의 행동을 인식에 관한 관련 연구에 대해 알아본다. 3장에서는 위험 행동 인식 방법에 대해 소개하고, 4장에서는 실험 및 결과를 알아보고, 5장에서는 결론 및 향후 연구로 마무리 한다.

II. Related Works

1. Object Detection

객체 검출은 영상 내에 객체 분류(Classification)와 객체의 지역화(Localization)를 수행하는 작업을 말한다. 딥러닝 기반의 객체 검출 방법은 Two-Stage 방법과 One-Stage 방법으로 나눌 수 있다.

Two-Stage 방법은 지역 제안(Region Proposal) 수행 후 객체 분류를 진행한다. Selective Search와 같은 지역 제안 알고리즘을 사용하여 영상 내에 객체가 있을 법한 영역을 찾아내고, 찾아낸 영역의 객체를 분류하는 두 단계로 나눠서 수행된다. 그래서 Two-Stage 방법은 One-Stage 방법보다 속도는 느리지만 정확도가 높다는 장점이 있다. Two-Stage의 대표적인 모델로 R-CNN이 있다[4].

One-Stage 방법은 지역 제안과 객체 분류를 동시에 수행하는 방법이다. One-Stage 방법은 지역 제안과 객체 분류를 동시에 수행하기 때문에 Two-Stage 방법보다 정확도는 떨어지지만 속도가 빠르다는 장점이 있다. One-Stage 방법의 대표적인 모델로 YOLO가 있다[5].

2. Action Recognition

2.1 Keypoint-Based Action Recognition

관절 키포인트 기반의 행동 인식은 자세 추정(Pose Estimation) 방법을 활용하여 영상에서 사람의 관절 키포인트를 추출하고, 추출한 키포인트들의 정보를 활용하여 행동을 인식한다. 자세 추정 방법은 하향식 추정 방법과 상향식 추정 방법이 있다. 하향식 추정 방법은 영상에서 사람의 바운딩 박스 영역을 검출한 후 바운딩 박스 내부에서 사람의 관절 키포인트를 추출하는 방법으로 대표적인 모델로 AlphaPose가 있다[6]. 상향식 추정 방법은 영상 내에 사람 객체의 관절 키포인트를 모두 추출하고, 추출된 키포인트 간의 상관관계를 추정하는 방법으로 대표적인 모델로 OpenPose가 있다[7]. 자세 추정 방법을 통해 추출한 사람의 관절 키포인트 정보는 순환신경망(Recurrent Neural Network) 계열의 LSTM(Long Short-Term memory)이나 GRU(Gated Recurrent Units) 모델의 입력값으로 사용되어 사람의 행동을 인식한다[8, 9]. 최근에는 사람의 관절을 노드로, 연결을 엣지로 간주하여 그래프를 구성하고, 구성한 그래프를 그래프 합성곱 신경망(Graph Convolution Network)의 입력값으로 사용하여 행동을 인식하는 연구가 있다[10, 11].

2.2 Video-Based Action Recognition

비디오 기반 행동 인식 방법은 RGB 영상이나 Depth 영상과 같은 카메라 영상을 활용하여 행동을 인식한다. 기존 방법에서는 CNN과 RNN를 동시에 사용하거나 연속된 프레임 간 물체의 이동 정보를 추정하는 광학 흐름(Optical Flow)을 활용한 Two-Stream 네트워크를 구성하여 사람의 행동을 인식하였지만 최근에는 3D CNN을 사용하여 영상 내 사람의 행동을 인식한다. 3D CNN은 공간적 특징에 대한 학습이 가능한 2D CNN 구조에서 시간축을 확장한 구조로 시공간적 특징을 학습 할 수 있다. 대표적인 3D CNN 모델로는 I3D와 SlowFast가 있다[12, 13].

3. Research Trends

최근 딥러닝 기술의 발달로 CCTV 및 카메라 영상을 활용하여 사람의 행동을 인식하는 연구가 진행 중에 있다. 특히 영상에서 사람의 위험 행동을 감지하기 위한 연구가 주를 이룬다.

사람의 넘어짐 행동을 감지하기 위해 입력된 영상에서 PoseNet 모델을 활용하여 사람의 관절 키포인트를 추출하고, 추출된 정보에서 키포인트들의 위치 정보와 위치 변화 가속도 변화 정보로 가공하여 GRU 입력값으로 사용함으로써 넘어짐을 인식하는 연구가 있다[14].

현장 안전 관리를 위해 AlphaPose와 LSTM를 활용하여 물건 옮기기, 넘어짐, 걷기, 물건 들기, 서기, 눕기에 대한 행동을 인식하는 연구가 있다[15]. 영상에서 사람의 관절 키포인트를 추출하고, 추출된 키포인트를 LSTM의 입력 데이터로 사용하여 행동을 인식한다. 이 논문에서 행동 인식은 9 프레임 단위로 이루어진다.

기존의 위험 행동 인식 방법은 자세 추정 모델을 통해 영상에서 사람의 관절 키포인트를 추출하고, 추출된 정보를 통해 사람의 행동을 인식한다. 지금까지의 연구에서는 일반 카메라 영상을 사용하기 때문에 개인의 프라이버시를 보호하지 못한다. 또한 한 사람에 대한 위험 행동만을 인식이 가능하며, 관절 키포인트 기반의 특성상 사람의 행동에 대한 정보만을 포함하기 때문에 물건과의 상호 작용 행동에서 인식률이 다소 떨어지는 것을 확인할 수 있다.

III. The Proposed Scheme

1. Dangerous Behavior Definition

1인 가구 환경에서 발생할 수 있는 위험 행동을 인식하기 위해 일반 행동 및 위험 행동에 대한 정의가 필요하다. 1인 가구 환경에서 발생할 수 있는 행동들을 크게 1인 상황일 때와 두 사람 이상일 때도 나눠서 정의하였다.

1인 상황일 때에 대한 행동은 일반 행동인 앉기, 서기, 걷기, 눕기와 위험 행동인 넘어짐으로 구분하였다. 앉기는 서 있다가 바닥 및 의자에 앉는 행동, 서기는 바닥 및 의자에 앉아 있다가 서는 행동, 걷기는 걷는 행동, 눕기는 바닥에 눕거나 누워있는 행동, 넘어짐은 바닥에 넘어지는 행동으로 정의하였다.

2인 이상 상황일 때에 대한 행동은 일반 행동인 모임과 위험 행동인 주먹질, 발길질, 집단 폭행, 칼로 찌르기로 구분하였다. 모임은 2명 이상이 어떠한 공격 행동 없이 모여 있는 행동, 주먹질은 한 사람이 다른 사람을 주먹으로 치는 행동, 발길질은 한 사람이 다른 사람을 발로 차는 행동, 집단 폭행은 2명 이상의 사람이 한 사람에게 일방적으로 폭행을 가하는 행동, 칼로 찌르기는 한 사람이 칼을 들고 다른 사람을 찌르는 행동으로 정의하였다.

2. Gaussian Blur Filter

블러 필터는 영상을 흐릿하게 하거나 노이즈를 제거하기 위해 사용된다. 본 논문에서는 영상을 흐릿하게하기 위해 가우시안 블러 필터를 사용한다. 가우시안 블러 필터는 중앙에 가까울수록 값이 크고, 멀어질수록 값이 작아지는 가우시안 분포 형태의 필터이다. 많은 필터 중 가우시안 블러 필터를 사용하는 이유는 영상에서의 경계선(Edge) 영역을 가장 부드럽게 조정하는 필터이기 때문에 프라이버시를 보호하는데 가장 적합하다고 판단하였다. 본 연구에서는 OpenCV를 사용하여 가우시안 블러 필터를 적용하였으며, Sigma 값으로 7을 적용하였다. Fig. 1은 영상에 가우시안 블러 필터를 적용한 예시이다.

Fig. 1. Example of applying Gaussian Blur Filter

3. Preprocessing Method using YOLOv5 model

본 논문에서는 영상에서 사람 객체를 검출하기 위해 YOLOv5 모델을 사용한다. YOLOv5 모델은 One-Stage 방법이지만 기존의 Two-Stage 방법보다 높은 정확도를 달성하였고, 실시간 객체 검출이 가능하다. Fig. 2는 YOLOv5를 활용하여 4 프레임을 기준으로 검출한 사람의 바운딩 박스 영역을 나타낸 그림이다[16]. Fig. 2에서 파란색 네모 박스는 각 프레임에서 YOLOv5 모델이 검출한 사람의 바운딩 박스를 의미하고, 빨간색 화살표가 가르키는 이미지는 각 프레임에서 검출된 사람의 바운딩 박스를 기준으로 추출한 사람 이미지를 의미한다.

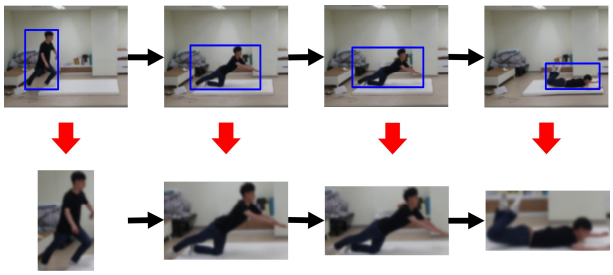


Fig. 2. Example of person area detection using YOLOv5

Fig. 2에서 추출한 사람의 바운딩 박스 영역은 사람의 행동 변화에 따라 그 크기가 모두 다르다. 이렇게 추출된 바운딩 박스를 일정 크기로 조정하여 비디오 기반 행동 인식 모델에 입력값으로 사용할 경우에는 시간공적 특징이 손실 될 수 있다. 그래서 본 논문에서는 시공간적 특징을 보존하면서 영상 내 1인 상황일 때와 2인 이상 상황일 때로 나눠 전처리 방법을 적용한다.

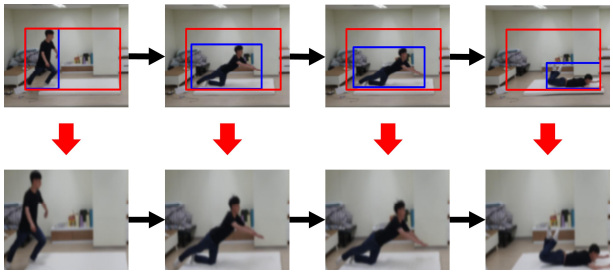


Fig. 3. Example of preprocessing method when one person

영상 내 한 사람일 경우에는 그림 Fig. 3과 같은 전처리 방법을 적용한다. Fig. 3은 4 프레임에 대해 전처리 방법을 적용한 예시이고, 파란색 박스는 현재 프레임에서 검출된 사람 영역을 의미하고, 빨간색 박스는 전처리 방법으로 크기를 조정한 최종적으로 추출할 영역을 의미한다. 한 사람에 대한 전처리 방법은 YOLOv5 모델을 통해 추출된 사람 영역 좌표를 활용하여 프레임 별로 왼쪽 상단 좌표 (x_l, y_l)와 오른쪽 하단 좌표 (x_r, y_r)를 구한다. 그 후 각 프레임에서 구한 사람 영역 좌표들을 비교하여 가장 작은 왼쪽 상단 좌표 (x_l', y_l')와 가장 큰 오른쪽 하단 좌표 (x_r', y_r')를 구하여 각 프레임에서 (x_l', y_l', x_r', y_r') 영역을 추출한다.

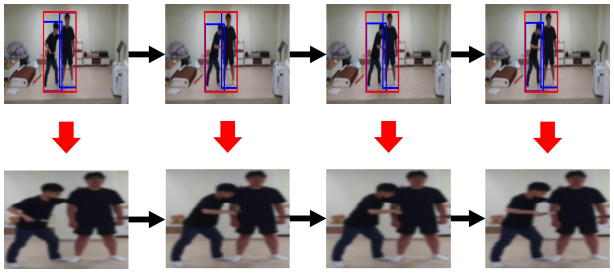


Fig. 4. Example of preprocessing method when there are more than two people

영상 내 두 사람 이상일 경우 전처리 방법은 Fig. 4와 같은 전처리 방법을 적용한다. 사람 간의 상호작용이 일어날 수 있는 상황은 서로 가까이 있을 때이다. YOLOv5 모델을 활용하여 사람 영역을 검출하고, 검출된 영역끼리 IOU(Intersection Over Union)를 계산하여 0보다 클 경우 사람 간의 상호작용이 발생할 수 있는 상황이다. 프레임 별로 검출된 영역들을 비교하여 가장 작은 왼쪽 상단 좌표 (x_l, y_l)와 가장 큰 오른쪽 하단 좌표 (x_r, y_r)를 구한다. 그 후 전체 프레임 기준으로 가장 작은 왼쪽 상단 좌표 (x_l', y_l')와 가장 큰 오른쪽 하단 좌표 (x_r', y_r')를 구하여 모든 프레임에서 (x_l', y_l', x_r', y_r') 영역을 추출한다. 만약 사람 수가 세 명 이상일 경우 x_l 좌표 기준으로 정렬 후 전처리 방법을 적용한다. 이와 같은 방법으로 영상에서 사람 영역을 추출할 경우 시공간 특징을 보존하면서 행동 인식이 가능하다.

4. Dangerous Behavior Recognition Method

Fig. 5는 위험 행동을 인식하는 흐름도이다. 위험 행동 인식 단계는 입력된 카메라 영상에서 YOLOv5 모델을 활

Fig. 5. Dangerous Behavior Recognition Flow

용하여 영상 내 사람 수를 확인한다. 영상 내 사람 수가 1명일 때에는 Fig. 3과 같은 전처리 방법을 적용하고, 2명 이상일 때에는 Fig. 4와 같은 전처리 방법을 적용한다. 전처리 방법은 16 프레임 기준으로 적용한다. 전처리 방법을 통해 새롭게 추출된 16 프레임은 행동 인식 모델의 입력값으로 사용되어 위험 행동을 인식한다. 위험 행동을 인식하기 위한 실험에 사용한 행동 인식 모델은 ResNet3D, I3D, SlowFast이다.

IV. Experiments and Results

1. Dataset

본 논문에서 인식하는 행동 클래스는 앉기, 서기, 걷기, 눕기, 넘어짐, 모임, 주먹질, 발길질, 집단 폭행, 칼로 찌르기로 총 10개의 클래스가 있다.

사용한 데이터 셋은 개인 촬영 데이터와 NTU-RGB+D 데이터 셋, Ai Hub의 이상행동 CCTV 영상 데이터 셋으로 구성하였다[17, 18]. 개인 촬영 데이터는 앉기, 서기, 걷기, 눕기, 넘어짐, 칼로 찌르기를 촬영하였고, NTU-RGB+D 데이터 셋에서는 앉기, 서기, 넘어짐, 주먹질, 발길질 행동 데이터를 추출하였다. 이상행동 CCTV 영상 데이터 셋에서는 집단 폭행과 모임 행동 데이터를 추출하였다.

비디오 기반 행동 인식은 연속된 프레임을 입력값으로 활용하기 때문에 16 프레임 기준으로 영상 내 사람 수에 따라 전처리 방법을 적용한 후 클립 영상을 생성하였고, 클래스 당 800개 클립의 학습 데이터와 240개 클립의 테스트 데이터로 구성하였다.

2. Hyperparameter Configuration

Table 1. Hyper parameters of Model Configuration

Hyper parameter	Value
Input Size	(224, 224)
Batch Size	8
Learning Rate	0.01
Epoch	100
Optimizer	SGD
Loss Function	Cross-Entropy

Table 1은 비디오 기반 행동 인식 모델을 학습에 사용된 하이퍼 파라미터(Hyper parameters) 세부 설정에 대한 표이다. 모델의 입력 크기(Input Size)는 (224, 224)으로 설정하였고, 배치 사이즈(Batch Size)는 8, 학습률(Learning Rate)은 0.01, 에폭(Epoch)은 100으로 설정하였다. Optimizer는 SGD(Stochastic Gradient Descent), Loss Function은 Cross-Entropy로 설정하였다.

3. Experimental Results of the Model trained with the Original Video

이 절에서는 원본 영상으로 학습한 비디오 기반 행동 인식 모델들에 대한 실험 결과를 분석한다. 원본 영상으로 학습한 모델을 활용하여 원본 영상에 대한 테스트 결과와 가우시안 블러 필터 영상에 대한 테스트 결과를 비교하여 원본 영상으로 학습한 모델이 가우시안 블러 필터 영상에 대한 실험에서 성능 저하의 유무를 확인하였다.

Table 2. Experimental Results on the Original Video

Model	Accuracy(%)
ResNet3D	89.5
I3D	94.3
SlowFast	96.8

Table 2는 원본 영상으로 학습한 모델들의 원본 영상에 대한 실험 결과이다. I3D와 SlowFast 모델은 각각 94.3%, 96.8%로 94.0%가 넘는 정확도를 달성하였다. 하지만 ResNet3D 모델은 90.0%를 넘지 못하는 89.5%로 다른 모델들에 비해 낮은 정확도를 달성하였다. 원본 영상에 대한 실험 결과 SlowFast 모델이 가장 높은 정확도를 달성하였다.

Table 3. Experimental Results on Privacy Video using Model trained on the Original Video

Model	Accuracy(%)
ResNet3D	80.9
I3D	81.8
SlowFast	85.6

Table 3은 원본 영상으로 학습한 모델들의 가우시안 블러 필터가 적용된 영상에 대한 실험 결과이다. 원본 영상에 대해 높은 정확도를 달성한 I3D와 SlowFast 모델은 가우시안 블러 필터가 적용된 영상에서는 각각 81.8%, 85.6%로 90.0%를 넘지 못하는 정확도를 달성하였다. 특히 I3D 모델의 경우에는 약 12% 정도의 정확도가 감소하였다. ResNet3D 모델도 80.9%로 원본 영상에 대한 실험 결과보다 약 8%정도의 정확도가 감소하였다. 본 실험을 통해 원본 영상으로 학습한 비디오 기반 행동 인식 모델들은 원본 영상보다 가우시안 블러 필터가 적용된 프라이버시 보호 영상에서의 정확도가 감소하는 것을 확인하였다. 이를 통해 프라이버시 보호 영상에 대한 비디오 기반 행동 인식 모델의 학습 및 성능 개선이 필요함을 확인하였다.

4. Confusion Matrix

이 절에서는 비디오 기반 행동 인식 모델이 1인 행동뿐만 아니라 2인 이상 행동, 사람과 물건 간의 상호작용 행동에 대해서도 바르게 인식하는지 확인하기 위하여 혼동행렬을 분석한다. 혼동행렬은 원본영상에 대한 실험에서 가장 높은 정확도를 달성한 SlowFast 모델을 기준으로 결과를 도출하였다. Fig. 6은 SlowFast 모델을 활용한 각 행동들에 대한 혼동행렬 그림이다.

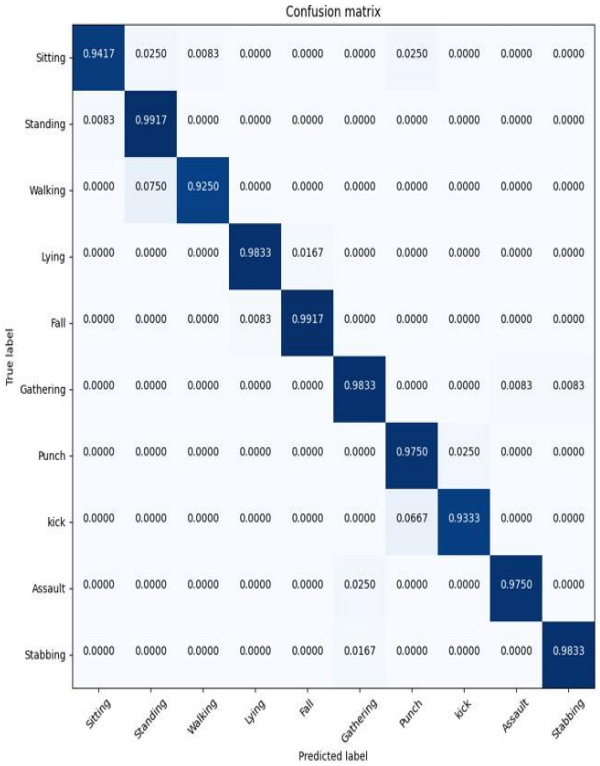


Fig. 6. Confusion Matrix of slowfast model

1인 행동의 경우에는 대체적으로 높은 인식률을 보인다. 앉기와 걷기 행동은 다른 행동들에 비해 낮은 인식률을 보이지만, 앉기의 경우 오인식한 행동은 서기, 걷기, 모임이다. 서기의 경우 앉기와 유사한 행동이며, 걷기와 모임 행동의 경우에는 매우 낮은 오인식률을 보인다. 하지만 이는 개선해야할 점으로 보인다. 걷기의 경우 오인식한 행동은 서기 행동으로, 서기에서 걷기로 이어질 수 있는 행동이기 때문에 오인식한 것으로 보인다.

2인 이상 행동에서도 높은 인식률을 보인다. 발길질 행동의 경우에는 주먹질로 오인식한 경우가 많다. 이는 주먹질과 발길질 행동이 서로 유사한 행동이기 때문으로 보인다. 사람과 물건 간의 상호작용 행동인 칼로 찌르기의 경우에도 98.3%로 높은 인식률을 보인다. 칼로 찌르기 행동도 높은 인식률을 보이는 이유는 키포인트 기반 행동 인식과 다르게 칼과 같은 물건에 대한 정보가 포함되기 때문으로 보인다. 혼동행렬을 통해 비디오 기반 행동 인식은 1인 행동뿐만 아니라 2인 이상 행동, 사람과 물건 간의 상호작용 행동에서도 높은 인식률은 보임을 확인하였다.

5. Experimental Results of the Model trained with Privacy Video

이 절에서는 가우시안 블러 필터가 적용된 프라이버시 보호 영상으로 학습한 비디오 기반 행동 인식 모델들에 대한 실험 결과를 분석한다. 실험은 크게 두 가지로 나눠서 진행하였다. 가우시안 블러 필터 영상 학습에 대한 실험 결과와 전처리 방법을 통해 사람 수를 나눌 수 있으므로 영상 내 사람 수에 따라 1인 행동과 2인 이상 행동으로 나누어 두 개의 모델을 학습하고, 최종적으로 두 개의 모델에 대한 평균 정확도를 계산하여 사람 수에 따라 두 개의 모델로 나누는 것에 대한 성능 개선의 여부를 확인하는 실험을 진행하였다.

Table 4. Experimental Results on Privacy Video using Model trained on Privacy Video

Model	Accuracy(%)
ResNet3D	87.3
I3D	91.7
SlowFast	94.7

Table 4는 프라이버시 보호 영상으로 학습한 모델의 프라이버시 보호 영상에 대한 실험 결과이다. Table 3에서의 확인한 정확도보다 ResNet3D 모델은 약 7%, I3D 모델은 약 9% 정도의 성능이 향상된 것을 확인하였다. 특히 SlowFast 모델의 정확도는 9%정도 향상된 것을 확인하였으며, Table 2에서의 원본 영상 학습 및 실험 결과인 96.8%와 비슷한 94.7%를 달성하였다. 이를 통해 가우시안 블러 필터가 적용된 프라이버시 영상에서도 위험 행동을 잘 인식하는 것을 확인하였다.

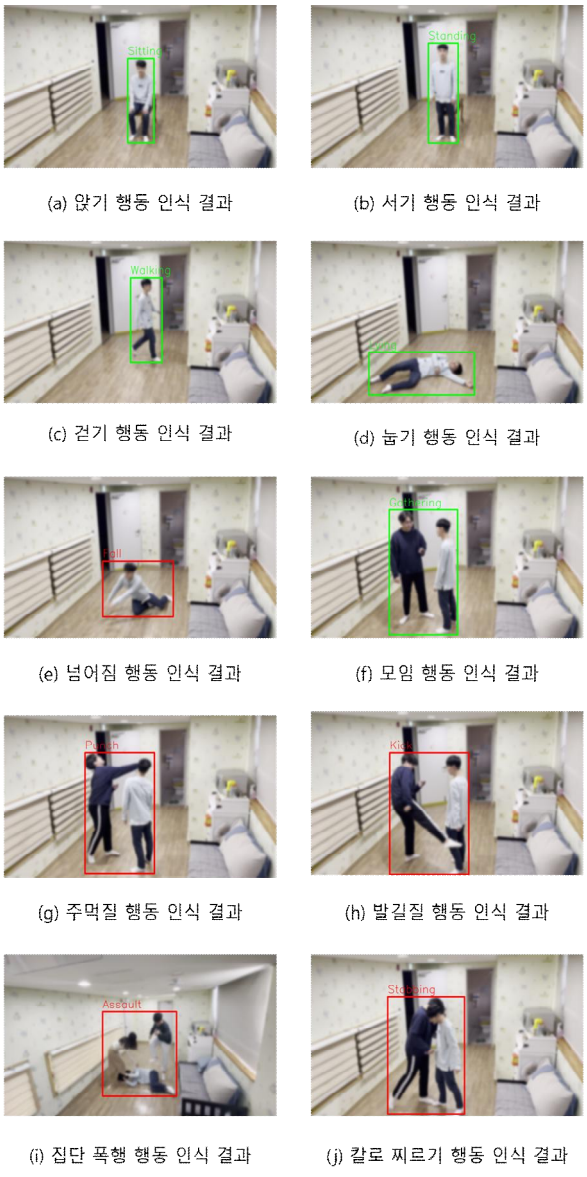
Table 5. Experimental Results of the Model according to the Number of People

Model	(a)	(b)	(c)
ResNet3D	90.8	90.5	90.6
I3D	93.6	94.1	93.8
SlowFast	96.0	95.5	95.7

YOLOv5를 활용한 전처리 방법을 통해 영상 내 사람 수를 나눌 수 있다. 이 점을 활용하여 영상 내 사람 수에 따라 두 개의 모델로 나누어서 학습을 진행하였다. Table 5

는 사람 수에 따라 나누어서 모델을 학습한 실험 결과이다. Table 5에서 (a)는 1인 행동에 대한 정확도를, (b)는 2인 이상 행동에 대한 정확도를, (c)는 (a)와 (b)에 대한 평균 정확도를 의미한다. Table 4와 Table 5의 비교를 통해 사람 수에 따라 모델을 나누어서 학습하는 것이 사람 수를 나누지 않고 학습하는 것보다 1~3%의 정확도가 향상된 것을 확인할 수 있다.

6. Dangerous Behavior Recognition Result in Single-person Household environment



(a) 앉기 행동 인식 결과 (b) 서기 행동 인식 결과

(c) 걷기 행동 인식 결과 (d) 눕기 행동 인식 결과

(e) 넘어짐 행동 인식 결과 (f) 모임 행동 인식 결과

(g) 주먹질 행동 인식 결과 (h) 발길질 행동 인식 결과

(i) 집단 폭행 행동 인식 결과 (j) 칼로 찌르기 행동 인식 결과

Fig. 7. General behavior and Dangerous behavior recognition results

1인 가구 환경과 유사한 환경에서 카메라를 통해 입력받은 영상에 가우시안 블러 필터를 적용한 프라이버시 보호 영상을 활용하여 위험 행동을 인식하고자 하였다. 위험 행동 인식 단계는 Fig. 5와 같은 방식으로 진행하였고, 한 행동을 인식하기 위한 프레임 수는 16 프레임을 사용하였고, 사용한 모델은 Table 5에서 가장 높은 정확도를 달성한 SlowFast 모델을 활용하여 실험을 진행하였다. 실험 결과 Fig. 7과 같이 일반 행동 및 위험 행동을 모두 바르게 인식하는 것을 확인하였다. 이를 통해 1인 가구 환경에서 프라이버시를 보호하면서 위험 행동을 인식이 가능하다.

V. Conclusions

본 논문에서는 1인 가구 환경에서 위험 행동을 인식하기 위해 앉기, 서기, 걷기, 눕기, 넘어짐, 모임, 주먹질, 발길질, 집단 폭행, 칼로 찌르기에 대한 행동을 인식하는 연구를 진행하였다. 또한 개인의 프라이버시를 보장하기 위해 카메라 영상에 가우시안 블러 필터를 적용하여 실험을 수행하였다. 객체 검출 모델인 YOLOv5를 활용하여 영상 내 사람 영역을 추출하고, 추출된 사람 영역에 전처리 방법을 적용하여 행동 인식 모델의 입력값으로 사용하였다. 실험을 통해 개인의 프라이버시를 보호하면서 1인 가구 환경에서 발생 할 수 있는 위험 행동을 인식할 수 있음을 확인하였다. 이를 통해 1인 가구환경에서 발생하는 고독사 및 침입자의 공격 행위를 인식할 수 있다.

향후 연구로는 학습된 모델들을 활용하여 1인 가구 환경에서 위험 상황 발생 시 알람을 전송하는 프라이버시 보호 영상을 활용한 위험 행동 감지 시스템을 구현하고자 한다.

REFERENCES

[1] WonJong Kim, “Single-Person Households and Crime”, Korean Journal of Law and Economics, Vol. 17, No. 1, pp.137-160, 2020. DOI: 10.46758/kjle.2020.04.17.1.137

[2] JangHo Kim, “Lonely death, safety net urgently needed”, https://www.kongje.or.kr/news/articleView.html?idxno=880

[3] JunHwi Park, “Research on for Strengthening the Efficacy ofCriminal Policy for Public Safety (II) : Improving Safety in Single-Person Household Concentration Areas”, korean institute of criminology, pp.1-938, 2017.

[4] R. Girshick, J. Donahue, T. Darrell and J. Malik. “Rich feature hierarchies for accurate object detection and semantic

segmentation," Proceedings of the IEEE conference on computer vision and pattern recognition, pp.580-587, 2014. DOI: 10.1109/CVPR.2014.81

[5] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," Proceedings of the IEEE conference on computer vision and pattern recognition, pp779-788, 2016. DOI: 10.1109/CVPR.2016.91

[6] H. S. Fang, S. Xie, Y. W. Tai and C. Lu, "Rmpe: Regional multi-person pose estimation," Proceedings of the IEEE International Conference on Computer Vision, pp.2334-2343, 2017. DOI: 10.1109/ICCV.2017.256

[7] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.7291-7299, 2017. DOI: 10.1109/CVPR.2017.143

[8] F. Gers, J. Schmidhuber, F. Cummins, "Learning to forget: Continual prediction with LSTM," Neural computation, vol.12, no.10, pp.2451-2471, 2000. DOI: 10.1162/089976600300015015

[9] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014. DOI: 10.3115/v1/D14-1179

[10] S. Yan, Y. Xiong, D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," In Thirty-second AAAI conference on artificial intelligence, 2018. DOI: 10.1186/s13640-019-0476-x

[11] M. Li, S. Chen, X. Chen, Y. Zhang, Y. Wang, Q. Tian, "Actional-structural graph convolutional networks for skeleton-based action recognition," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp.3595-3603, 2019. DOI: 10.48550/arXiv.1904.12659

[12] J. Carreira, and A. Zisserman, "Quo vadis, action recognition? a new model and the kinetics dataset," In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.6299-6308, 2017. DOI: 10.1109/CVPR.2017.502

[13] C. Feichtenhofer, H. Fan, J. Malik, K. He, "Slowfast networks for video recognition," In Proceedings of the IEEE/CVF international conference on computer vision, pp. 6202-6211, 2019. DOI: 10.1109/ICCV.2019.00630

[14] YoonKyu Kang, HeeYong Kang, DalSoo Weon, "Human Skeleton Keypoints based Fall Detection using GRU", Journal of the Korea Academia-Industrial, Vol. 22, No. 2, pp.127-133, 2021. DOI: 10.5762/KAIS.2021.22.2.127

[15] HyunJae Bae, GyuJin Jang, YoungHun Kim, JinPyung Kim, "LSTM(Long Short-Term Memory)-Based Abnormal Behavior Recognition Using AlphaPose", KIPS Trans. Softw. and Data Eng, Vol. 10, No. 5, pp.187-194, 2021. DOI: 10.3745/KTSDE.2021.10.5.187

[16] G. Jocher, K. Nishimura, K. Mineeva, T. Vilarino, YOLOv5, https://github.com/ultralytics/yolov5.

[17] Aihub, "Introduction of abnormal behavior CCTV video", https://www.aihub.or.kr/aidata/139.

[18] A. Shahroudy, J. Liu, T. Ng and G. Wang, G, "Ntu rgb+ d: A large scale dataset for 3d human activity analysis," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp.1010-1019, 2016. DOI: 10.1109/CVPR.2016.115

Authors

ChaeHyun Lim received the B.S. degree in Computer Science from MyongJi University, Korea in 2019 and M.S. degree in Software from Soongsil University, Korea in 2022. ChaeHyun Lim is currently a Student in the Department of software at Soongsil University. He is interested in Privacy, Artificial Intelligence, Deep Learning and Computer Vision.

Myung Ho Kim received the B.S. degree in Computer Science from Soongsil University, Korea, in 1989 and M.S. and Ph.D. degrees in Electronic Calculation form POSTEC, Korea, in 1991 and 1995, respectively. Dr. Kim joined the faculty of the Department of Software at Soongsil University, Seoul, Korea, in 1995. He is currently a Professor in the Department of Software, Soongsil University. He is interested in Privacy, Artificial Intelligence, System Software and Open Software.