

지능형 오디오 및 비전 패턴 기반 1인 가구 이상 징후 탐지 알고리즘

Intelligent Abnormal Event Detection Algorithm for Single Households at Home via Daily Audio and Vision Patterns

정 주 호¹ 안 준 호^{1*}
Juho Jung Junho Ahn

요 약

1인 가구의 수가 증가함에 따라 1인 가구의 구성원이 집안에서 심각한 부상을 당할 경우 혼자 도움을 청하기 쉽지 않다. 본 연구는 집안에서 1인 가구의 구성원이 심각한 부상을 당했을 때 비일상적인 상태를 탐지한다. 홈 CCTV를 기반으로 수집된 영상을 통해 패턴을 분석 및 인식하는 영상 탐지 알고리즘을 제안한다. 또한, 스마트폰을 기반으로 집안에서 발생하는 소리의 패턴을 분석 및 인식하는 음성 탐지 알고리즘도 제안한다. 각각의 알고리즘만 사용할 경우, 단점이 존재하여 넓은 영역에서 심각한 부상과 같은 상황을 탐지하기 어렵다. 그래서 두 알고리즘을 효율적으로 결합한 융합 방식을 제안한다. 각각 탐지 알고리즘의 성능과 제안된 융합 방식의 정확한 탐지 성능을 평가했다.

☞ 주제어 : 비전, 영상 처리, 오디오, 융합, 패턴, 1인 가구, 일상 행동

ABSTRACT

As the number of single-person households increases, it is not easy to ask for help alone if a single-person household is severely injured in the home. This paper detects abnormal event when members of a single household in the home are seriously injured. It proposes an vision detection algorithm that analyzes and recognizes patterns through videos that are collected based on home CCTV. And proposes audio detection algorithms that analyze and recognize patterns of sound that occur in households based on Smartphones. If only each algorithm is used, shortcomings exist and it is difficult to detect situations such as serious injuries in a wide area. So I propose a fusion method that effectively combines the two algorithms. The performance of the detection algorithm and the precise detection performance of the proposed fusion method were evaluated, respectively.

☞ keyword : Vision, Image Processing, Audio, Fusion, Pattern, Single Households, Daily behavior

1. 서 론

노인 가구의 33.5%가 1인 가구로 살고 있으며 이는 총 130만 명[1]에 이른다. 또한, 2015년에 집에서 홀로 죽음을 맞이한 1인 가구 구성원은 338명[1]에 이른 것으로 확인된다. 전문가들은 한국에서 1인 가구의 숫자가 급격하게 늘어나고 있으며 2020년에는 3,200명에 이를 것으로 전망하고 있다.

이러한 1인 가구의 증가는 한국만의 문제가 아니라 전 세계적인 문제로 급부상하고 있다. 기사에 인용된 OECD

국가들의 1인 가구 구성원 비율에 관한 자료[2]를 통해 확인한 결과, 1960년대에 대부분의 나라는 1인 가구의 비율이 20% 이하로 구성되어 있었다. 최근 2015년의 통계 자료에는 덴마크, 핀란드, 독일, 노르웨이 순서로 1인 가구의 비율이 40%가 넘는 수치를 기록하고 있다. 이처럼, 우리나라뿐만 아니라 전 세계적으로 1인 가구의 비율이 늘어나고 있는 것을 확인할 수 있다.

1인 가구 구성원이 집안에서 심각한 부상을 당해 집안에서 홀로 방치되는 문제점을 해결하기 위해, 최근 서울시에서는 60대 이상 노인들을 대상으로 정책[3]을 시작하였다. 사물인터넷을 활용하여 IoT 장치를 집 안에 설치한 후 설치된 장치의 센서 데이터를 수집해 지자체로 데이터를 전송한다. 담당생활관리사는 수집된 센서 데이터를 앱을 통해 확인하며 8시간 동안 움직임 센서의 값이 발

¹ Computer Information Technology, Korea National University of Transportation, Chungju-si, 27469, Korea

* Corresponding author (jhahn@ut.ac.kr)

[Received 6 September 2018, Reviewed 10 September 2018(R2 27 December 2018), Accepted 7 January 2019]

생하지 않을 경우 이상을 의심하여 119와 연계하며 긴급 조치를 취하게 된다. 이 기기는 실시간으로 사람의 움직임 여부도 판단할 수 있고 집안에 불이 나거나 가스가 누출이 되었을 경우를 대비하는 가스센서도 존재하여 탐지 시 알람을 보낸다.

이렇게 1인 가구 구성원이 집안에서 심각한 부상을 당했을 때의 대처 방안들이 존재함을 확인할 수 있었다. 하지만 서울시 정책[3]도 별도의 센서를 집안에 설치해야 하며, 실제로 움직임을 감지하는 센서가 사람만을 탐지하는게 아니라 반려동물이나 다양한 물체의 움직임을 탐지하는 문제점이 발생하게 된다. 본 연구에서는 이러한 문제점들을 개선하고 보다 정확하게 사람이 심각한 부상을 당했을 때의 비일상적인 상태들을 탐지하고자 한다.

알고리즘을 제안하기에 앞서서 “이상 징후”라는 단어를 정의하고자 한다. “이상 징후”란, 사람이 집안에서 심각한 부상을 당하여 비일상적인 상태인 경우, 그리고 장시간 동안 카메라 영역 내에서 사람의 움직임이 탐지되지 않거나 일정시간 이상 방 안에서 아무 소리가 나지 않거나 또는 소리가 멈추지 않고 지속적으로 나는 경우와 같은 평소 사람의 정상적인 패턴과 다른 상황을 말한다.

우리는 1인 가구가 심각한 부상을 당하여 비일상적인 상태인 이상 징후를 탐지하기 위한 알고리즘을 제안하려고 한다. 홈 CCTV를 기반으로 하여 사람의 패턴을 분석 및 인식하는 영상 탐지 알고리즘과 스마트폰을 기반으로 하여 집 안에서 발생할 수 있는 소리를 기반으로 하여 소리의 패턴을 분석 및 인식하는 음성 탐지 알고리즘이다. 영상 탐지 알고리즘과 음성 탐지 알고리즘의 각각의 단점을 보완하고 높은 정확도로 이상 징후를 탐지하고 두 알고리즘을 효율적으로 결합한 융합 방식도 제안하려고 한다. 각각의 알고리즘은 실험을 통하여 성능을 평가하였으며, 융합 방식은 다양한 시나리오를 이용하여 정확도를 평가했다.

2. 관련 연구

영상을 기반으로 하는 기존의 탐지연구[4, 5, 6, 7, 8, 9, 10]들은 사람과 물체를 효율적으로 인식하려는 연구들이 많이 진행되고 있다. 해당 연구[4]를 통하여, 각자 개인의 패턴을 SVM(Support Vector Machine)을 이용하여 학습시킨다. 그리고 사람의 자세에 대한 실루엣을 이용하여 분류하고 CNN(Convolution Neural Network)을 이용하여 학습하는 연구[5]도 있다. 인식되는 사람의 각도,

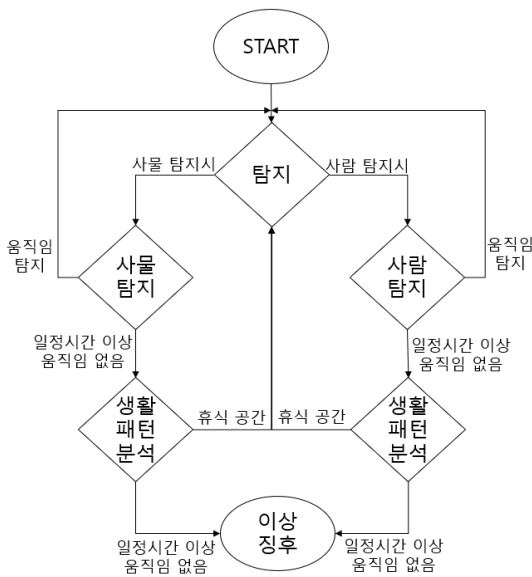
속도, 비율을 이용하며 KNN(K-Nearest Neighbors) 알고리즘을 활용하는 연구[6]가 있으며, 또한 사람의 실루엣의 중심점을 5등분으로 나누어 각도를 구하고 이를 분류하는 알고리즘으로는 MEWMA(Multivariate Exponentially Weighted Moving Average) 알고리즘을 이용하는 연구[7]가 있다. 인식하는 물체를 ssd-mobilenet 과 Faster R-cnn 모델을 사용하여 이에 따른 높은 정확성을 이용하는 연구[9]도 있다. 화재가 날 때 빠르게 연기가 나게 되는데 이 연기를 탐지하여 초기에 화재를 신속하게 인식해서 도움을 주는 연구[10]도 있다.

오디오를 기반으로 하는 이벤트 탐지연구[11, 12]는 발생하는 오디오를 인식하여 어떤 오디오인지 탐지하고 탐지되는 오디오를 분류하는 연구들이 많이 진행되고 있다. 이러한 연구들은 오디오가 발생되면 이에 대한 이벤트를 감지하여 발생된 오디오의 특징을 추출하고 이에 맞는 오디오가 어떤 오디오인지 찾는 것에 집중되어 있다. 오디오 이벤트를 기반으로 하는 연구들 중 대중적으로 사용되고 있는 알고리즘은 MFCC-GMM 알고리즘[11]이다. MFCC를 이용하여 주파수의 특징 및 값을 추출한 후 GMM을 이용하여 추출된 주파수의 특징들을 분석해 소리를 분류시킨다. 그리고 CNN(Convolution Neural Network) 기법을 이용하여 발생되고 있는 오디오 이벤트를 탐지하고 분류하는 연구[12]도 있다.

본 연구를 통해 높은 정확성으로 이상 징후를 탐지하기 위해 이전 연구[13, 14]를 확장하였다. 이전의 연구를 통해 우리는 다양한 센서를 이용하여 이상 징후를 탐지하기 위해 노력해왔다. 그중에서도 영상은 카메라 영역이라는 제한된 범위 안에서 사람의 행동에 대해 인식하는 것으로 가장 적합하였다. 왜냐하면, 실제로 사람이 어떤 행동을 하고 있는지 직관적으로 탐지할 수 있기 때문이다. 또한, 사람이 집 안에 있으면 아무런 소리도 내지 않기로 쉽지 않다. 그래서 집 안에서 발생하는 소리를 기반으로 하여 사람의 움직임을 판단하려고 한다. 우리는 다양한 센서를 이용하여 이상 징후를 탐지하기 위해 노력해왔다. 하지만 한가지 센서만 이용하여 이상 징후를 탐지하는 것은 한계점이 존재하였다. 그래서 본 연구를 통해 두 가지 센서를 결합하여 한가지의 센서를 이용하는 것보다 높은 정확성으로 이상 징후를 탐지할 수 있었다. 따라서 본 연구는 각각의 영상 탐지 알고리즘과 음성 탐지 알고리즘을 개별적으로 이용하는 것보다 두 알고리즘을 결합한 융합 방식을 제안한다.

3. 알고리즘

영상 탐지 알고리즘은 홈 CCTV를 통해 수집된 영상을 기반으로 했다. Tensorflow Object Detection을 기반으로 하여 사람 또는 사물을 탐지하였고 Faster R-cnn inception 모델을 사용하였다. 영상 탐지 알고리즘에 대한 Flow Chart는 그림 1과 같다. 홈 CCTV를 통해 인식된 영상이 사람을 탐지하게 되면 사람을 탐지한 상태가 된다. 사람이라고 탐지된 상태에서 사람의 움직임을 탐지하게 되면 정상 상태라고 분류가 된다. 그 후에는 다시 정상 탐지상태로 돌아간다. 사람 탐지상태에서 사람이 장시간 움직임이 없으면 이 사람에 대한 생활 패턴을 분석한다. 생활 패턴 분석 단계에서 이 사람이 항상 휴식을 취하는 공간이면 사람은 정상인 상태로 분류가 된다. 그 후에는 다시 정상 탐지로 돌아간다. 일정 시간 이상 탐지된 사람과 사물의 움직임이 없으면 이상 징후라고 탐지하게 된다.

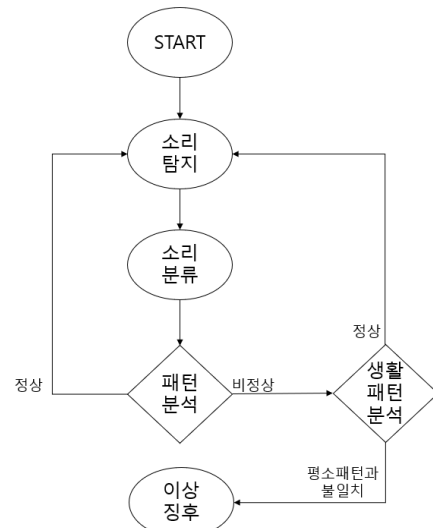


(Figure 1) Flow chart for the proposed vision detection algorithm

홈 CCTV를 통해 인식된 영상이 사물을 탐지하게 될 경우 사물 탐지상태가 된다. 사물 탐지상태에서 사물이 움직여서 사물의 위치가 변경되면 사람이 움직임을 가한 것이기 때문에 이는 정상 상태라고 분류가 되어 정상 탐지상태로 돌아가게 된다. 사물이 장시간 동안 움직임이 없으면 생활 패턴 분석상태로 넘어가게 된다. 생활 패턴

분석상태에서 휴식 공간에서 사용하는 소파나 침대일 경우에는 정상인 상태로 분류하고 다시 정상 탐지상태로 분류하게 된다. 일정 시간 이상 사람과 사물의 움직임이 탐지되지 않으면 이상 징후라고 탐지한다. 본 연구는 Tensorflow Object detection 알고리즘의 Faster R-cnn inception 모델의 이용한 결과물을 통하여 탐지된 사람의 x, y 좌표와 길이, 탐지된 사람의 각도 및 정확도와 사물의 x, y 좌표를 기반으로 하는 별도의 그래프를 생성하여 확인하고 이를 분석하였다.

본 연구의 음성 탐지 알고리즘은 이전의 연구[13, 14]를 확장하고 스마트폰을 기반으로 하여 집에서 발생하는 소리를 이용해 이상 징후를 탐지한다. 음성 탐지 알고리즘에 대한 Flow Chart는 그림 2와 같다.

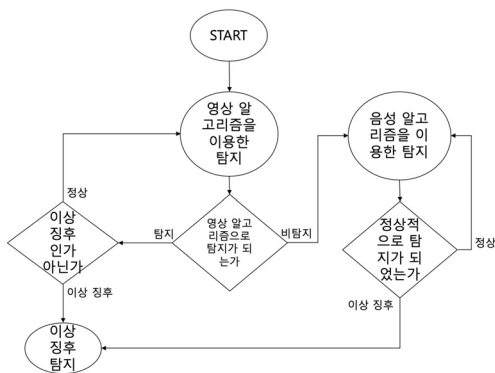


(Figure 2) Flow chart for the proposed voice detection algorithm

사람이 별도의 장치를 집 안에서 휴대하지 않고 집에서 발생하는 소리를 기반으로 하여 분석하고 패턴을 탐지하고 사람의 움직임 여부 또는 이상 징후를 탐지하려고 한다. 음성 탐지 알고리즘은 집에서 소리가 발생하게 되면 발생된 소리는 MFCC-GMM 알고리즘을 이용하여 특징 추출 및 모델을 기반으로 하여 소리를 분류시킨다. GMM 알고리즘을 이용하여 0.5초마다 Sampling rate는 22050을 이용하며 12개의 데이터 지점을 수집하여 소리를 3가지 유형 중 하나로 분류하게 된다. 각 데이터 지점은 MFCC 알고리즘으로 추출한 13개 주파수로 구성되어

있다. 본 연구에서는 학습시킨 모델을 기준으로 소리를 분류한다. 첫 번째는 아무 소리도 나지 않는 소리(조용한 소리)이며, 두 번째는 집 안에서 날 수 있는 노이즈가 낀 소리(예: 에어컨, 냉장고 등 팬이 돌아가는 소리)와 세 번째는 일상적으로 말하는 소리를 탐지하도록 하였다. 아무 소리도 나지 않는 소리를 분류한 이유는 이상 징후가 발생하였을 때, 사람이 움직이지 않기 때문에 이를 탐지하기 위하여 아무 소리도 나지 않는 소리를 분류하였다. 노이즈가 낀 소리는 아무것도 나지 않는 소리와 말하는 소리 간의 정확성을 높이기 위하여 일상생활 속에서 날 수 있는 노이즈 소리를 추가하였다. 말하는 소리는 일반적으로 사람이 움직이거나 살아있음을 탐지하는 것을 목표로 하였다. 음악이나 영상을 재생시켜놓은 상태에서 이상 징후가 발생할 경우, 지속적으로 말하는 소리만 탐지 될 것이다. 그래서 패턴을 탐지하기 위해 말하는 소리를 분류하였다. 우리는 분류된 소리를 기반으로 하여 각각의 소리별 비율 및 평소 사람이 내는 소리의 패턴을 분석하여 사람의 움직임 및 이상 징후를 탐지한다.

우리는 홈 CCTV를 기반으로 하여 영상 데이터를 수집하고 사람의 패턴을 분석 및 인식하여 이상 징후를 탐지하는 영상 탐지 알고리즘을 제안하였다. 그리고 스마트폰을 기반으로 하여 집에서 발생하는 소리를 분석하고 소리 패턴을 분석 및 인식하여 이상 징후를 탐지하는 음성 탐지 알고리즘을 제안하였다. 본 연구에서는 영상 탐지 알고리즘과 음성 탐지 알고리즘을 결합한 융합 방식을 제안한다. 융합 방식의 Flow Chart는 그림 3과 같다.



(Figure 3) Flow chart for the proposed Fusion method

영상 탐지 알고리즘은 홈 CCTV 영역 밖에서 발생하는 일상생활 및 이상 징후를 탐지하지 못하는 단점을 가지고 있다. 음성 탐지 알고리즘은 소리가 발생되다가 사람이 이

상 징후가 발생되어 소리가 멈출 경우 정상으로 판단하여 이상 징후를 탐지 못하는 단점 및 거리가 먼 곳에서 소리가 날 경우에 이를 정상적으로 탐지하지 못하는 단점을 가지고 있다. 융합 방식은 각 알고리즘의 단점을 서로 보완하여 높은 정확성으로 이상 징후를 탐지할 수 있다.

본 연구는 각각의 탐지 알고리즘은 실험을 통하여 성능을 평가했고 융합 방식은 시나리오를 이용하여 알고리즘의 정확성을 평가했다.

4. 실험 및 평가

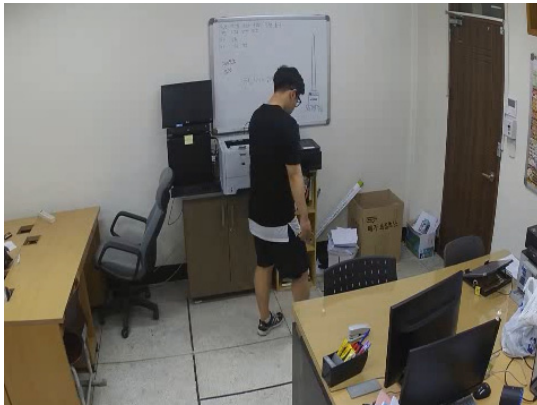
영상 탐지 알고리즘은 Youtube를 통해 수집한 영상 및 실험을 기반으로 하여 테스트를 진행하였고 이에 대한 Recall, Precision, Accuracy는 표 1과 같다.

(Table 1) Performance evaluation results for the proposed vision detection algorithm

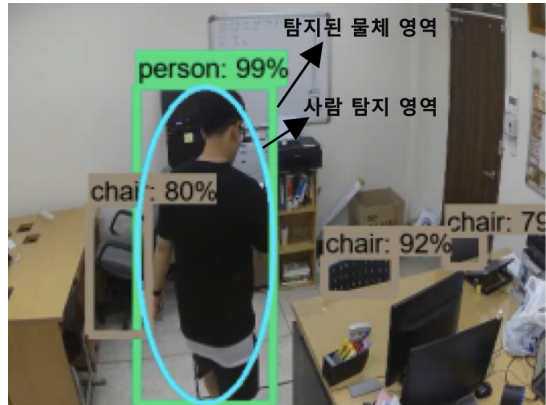
알고리즘	Recall	Precision	Accuracy
영상 탐지 알고리즘	0.95	0.95	0.93

Youtube와 실험을 통해 수집한 영상 데이터는 총 96개이며 사람이 정상적으로 움직이고 있음을 True로 하여 Recall, Precision, Accuracy를 구하였다. Recall은 $tp / tp + fn$ 이기 때문에 70 / 73을 하여 0.95가 나왔고, Precision은 $tp / tp + fp$ 이기 때문에 70 / 73을 하여 0.95가 나왔다. Accuracy는 $tp + tn / tp + tn + fp + fn$ 을 하여 0.93 이 나오게 되었다.

실험은 총 5번의 실험을 진행하였고, 카메라 영역 내에서 정상적인 활동을 하고 있음을 알 수 있는 영상 2개와 이상 징후 상태를 나타내는 영상 2개, 카메라 영역 밖에서의 상황 1개를 하여 총 5번의 실험을 진행하였고 그림 4, 5, 6, 7은 실제 영상 탐지 알고리즘을 적용한 모습이다. 그림 4은 아무것도 적용하지 않은 원본 영상이다. 그림 5는 Tensorflow Object Detection 알고리즘의 Faster R-cnn inception 모델이 적용된 영상이다. 그림 6은 사람이라고 탐지가 된 부분에는 파란색 마커를 칠하고, 물체라고 탐지가 된 영역은 빨간색 마커가 칠해지게 하여 이를 기반으로 적용된 모습이다. 그림 7은 영상의 첫 부분을 이미지 한 장으로 캡처한 후 이를 배경으로 저장한 뒤 흑백 처리하고 Gaussian Blur를 적용한 후 Opencv의 Motion Analysis를 통하여 배경을 움직임이 있는 물체에



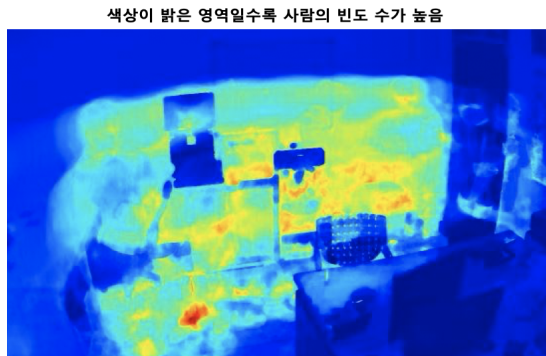
(Figure 4) An Original video



(Figure 5) Application to Faster R-cnn inception



(Figure 6) Movement classifications of humans or objects

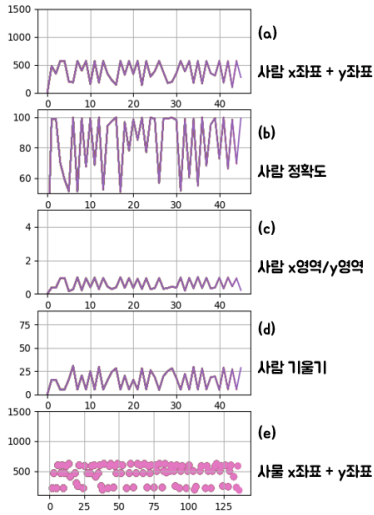


(Figure 7) Repeated motion pattern detections

대해서 감산한다. 그림 8을 통해서 Tensorflow Object detection 알고리즘의 Faster R-cnn inception 모델이 적용된 것을 바탕으로 탐지되는 것들을 기반으로 하여 그래프를 그렸다. 그림 8 (a) 그래프는 사람이라고 탐지된 물체의 x, y좌표의 합을 나타내며, 그래프의 y축에 나타나고 x축은 영상을 탐지한 프레임 수이다. 그림 8 (b) 그래프는 사람이라고 탐지된 물체의 정확도를 나타낸다. 탐지의 기준치를 정확도 50% 이상만 탐지하게 코드를 수정하였다. 왜냐하면, 정확도 50% 이하는 사람이라고 판단하기 어렵기 때문에 반영하지 않았다. 그래프의 y축은 정확도에 해당하는 값이며 x축은 영상을 탐지한 프레임 수이다. 그림 8 (c) 그래프는 사람으로 탐지된 영역의 $(x \text{ 최대값} - x \text{ 최소값}) / (y \text{ 최대값} - y \text{ 최소값})$ 의 비율이다. 보통의 경우에는 사람이 탐지 될 경우 x의 길이가 짧고 y의 길이가 길어서 1보다 작은 값으로 수렴된다. 하지만 사람이 누워있을 경우 x의 길이가 길고 y의 길이가 짧아지기

때문에 상대적으로 1보다 큰 값을 가질 수 있다. 이를 통해 사람이 서 있는 상태인지 누워있는 상태인지를 구분할 수 있다. 그래프의 y축은 비율의 값이며, x축은 영상을 탐지한 프레임 수이다. 그림 8 (d) 그래프는 사람이라고 판단되는 물체의 원형으로 탐지된 부분에 대한 점선의 기울기를 나타내는 그래프이다. 사람이 정상적으로 활동을 할 경우 일정한 기울기 값이 지속적으로 일정한 값이 나오게 되지만 서 있는 상태에서 급작스럽게 이상 징후가 발생 될 경우에는 기울기의 값이 급격하게 변하는 것을 확인할 수 있다. 그래프의 y축은 기울기의 값이며, x축은 영상을 탐지한 프레임 수이다. 그림 8 (e) 그래프는 사람 이외의 사물을 탐지했을 경우 사물의 좌표를 기반으로 하여 마커를 칠한다. 사물 탐지 상태에서 사물의 움직임을 파악하기 위하여 마커를 칠하였다. y축은 탐지된 사물의 x, y좌표의 합을 나타내며, x축은 영상을 탐지한 프레임 수이다. 이렇게 카메라 영역 내에서는 제안

된 알고리즘을 통하여 사람의 이상 징후를 탐지할 수있다. 하지만 탐지 가능한 부분도 카메라 영역 내에 사람이 있어야지만 가능한 부분이다. 만약, 사람이 카메라 영역 밖에서 이상 징후나 일상 생활을 지속적으로 이어 나갈 경우에는 영상 탐지 알고리즘만 이용해서는 정상적으로 탐지하기는 어렵다.



(Figure 8) An analyzed graph with Faster R-cnn inception algorithm

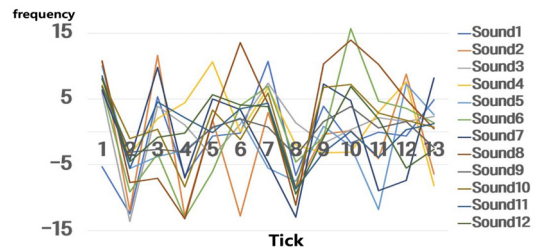
그 다음 음성 탐지 알고리즘이다. 음성 탐지 알고리즘도 Youtube를 통해 수집된 음성 데이터 및 자체적인 실험을 진행하였고, 이에 대한 Recall, Precision, Accuracy는 표 2와 같다.

(Table 2) Performance evaluation results for the proposed voice detection algorithm

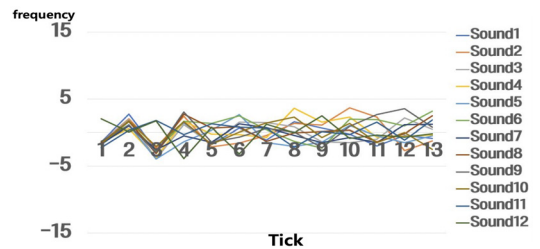
알고리즘	Recall	Precision	Accuracy
음성 탐지 알고리즘	0.95	0.97	0.95

음성 탐지 알고리즘도 Youtube와 실험을 통하여 수집한 음성 데이터는 총 66개이며 사람이 정상적으로 소리를 내고 있음을 True로 하여 Recall, Precision, Accuracy를 구하였다. Recall은 $tp / tp + fn$ 이기 때문에 46 / 48을 하여 0.95가 나왔고, Precision은 $tp / tp + fp$ 이기 때문에 46 / 47을 하여 0.97가 나왔다. Accuracy는 $tp + tn / tp + tn + fp + fn$ 을 하여 0.95가 나오게 되었다.

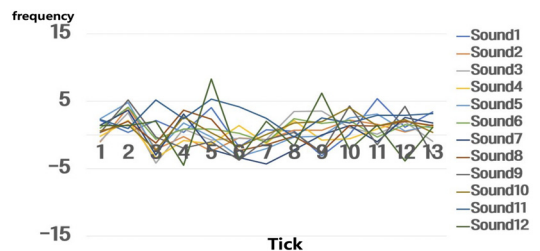
음성 탐지 알고리즘은 MFCC-GMM 알고리즘을 이용하여 소리를 분류시킨다. 분류 방법은 0.5초마다 Sampling rate는 22050을 이용하여 12개의 데이터 지점을 수집하고 각 데이터 지점을 분석하여 3가지의 소리 중 하나로 분류시킨다. 본 연구에서는 집안에서 나는 소리를 기반으로 하여 사람의 움직임을 효율적으로 탐지하기 위하여 소리를 분류시키는 별도의 모델을 만들어서 탐지했다. 분류된 소리들도 각자의 소리 모델에 따라 패턴이 존재한다. 말하는 소리 모델에 대한 패턴은 그림 8과 같고, 조용한 소리 모델에 대한 패턴은 그림 9, 노이즈가 낀 소리에 대한 모델은 그림 10과 같다. 각 그림에 대한 x축은 MFCC 알고리즘을 이용하여 추출한 데이터 포인트에 대하여 13개의 주파수(Tick)로 구성되어 있어서 x축의 값은 1부터 13까지 구성되어 있으며, y축은 탐지된 데이터 포인트에 대한 각각의 주파수 탐지 값(frequency)이다.



(Figure 9) Talking voice model patterns

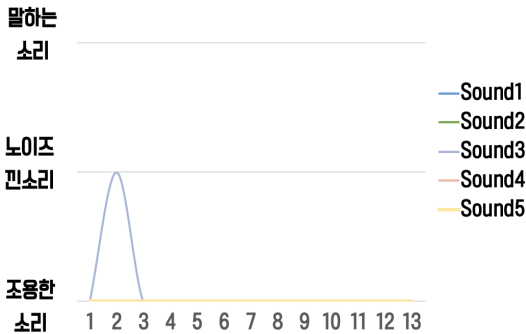


(Figure 10) Silent voice model patterns

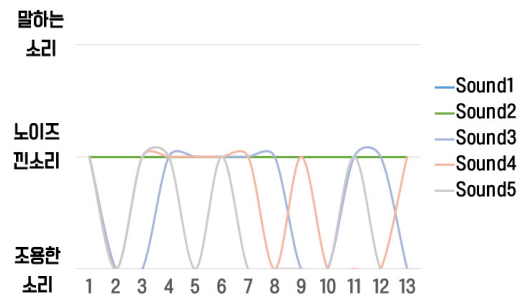


(Figure 11) Background noise model patterns

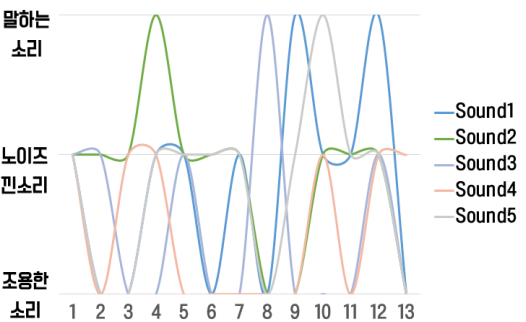
그림 12와 그림 13과 그림 14는 분류된 소리에 대한 패턴이다. 조용한 소리는 그림 12와 같은 패턴이 나오게 된다. 그림 13은 노이즈가 낀 소리를 탐지했을 때의 패턴이며, 그림 14는 말하는 소리에 대한 패턴이다. 각 그림의 그래프에 대한 x축은 탐지된 소리의 개수며 y축은 탐지된 소리이다.



(Figure 12) Silent voice patterns with the proposed voice detection algorithm

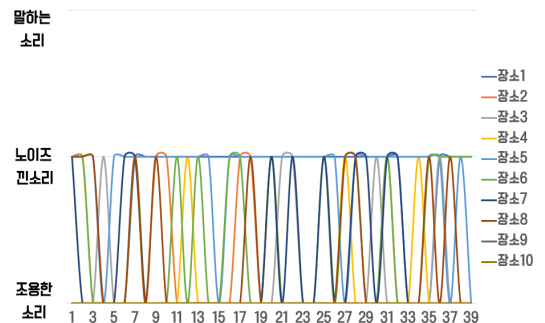


(Figure 13) Background noise patterns with the proposed voice detection algorithm



(Figure 14) Talking voice patterns with the proposed voice detection algorithm

사람이 정상적인 활동시에는 사람들이 대개 아무 말도 없이 혼자 있거나 음악이나 TV를 한 번도 안 듣지는 않기 때문에 말하는 소리가 탐지된다. 하지만, 이상 징후가 발생 될 때는 음악이나 TV가 켜져 있지 않는 상태라면 아무 소리도 나지 않는 소리와 노이즈가 낀 소리만 탐지가 된다. 본 연구에서는 이를 증명하기 위해서 서로 다른 장소 10곳에서 이상 징후와 같은 상황을 연출하여 실험하였고 이에 대한 패턴을 분석했다. 본 실험에서는 사람이 항상 집 안에서 생활을 하며 이상 징후가 발생했다고 가정하여 실험을 진행하였다. 실험에 대한 결과는 그림 15와 같다. 그림 15에서와 같이 조용한 소리와 노이즈 낀 소리만 탐지 되었고 말하는 소리는 한 번도 탐지되지 않았다. 항상 집 안에서 발생하는 소리는 사람마다 생활 패턴이 다르기 때문에 나타나는 소리 패턴은 다를 것이다. 이러한 패턴을 분석하여 사람의 이상 징후 여부를 파악할 수 있다. 하지만, 음성 탐지 알고리즘도 단점이 존재한다. 모든 상황에 대해서 정상적으로 탐지할 수는 없다. 소리가 멀리서 나는 경우에는 음성 탐지 알고리즘만 이용하여 사람이 정상 활동을 하는지 이상 징후가 발생했는지 여부를 알기란 힘들다. 소리가 발생하다가 이상 징후가 발생하여 소리가 사라진 경우에 대해서도 음성 탐지 알고리즘만 이용해서는 사람이 정상인 상태에서 소리가 나지 않는 것인지, 이상 징후가 발생하여 소리가 나지 않는 것인지에 대해서는 탐지하기 어렵다.



(Figure 15) Abnormal events classified by locations

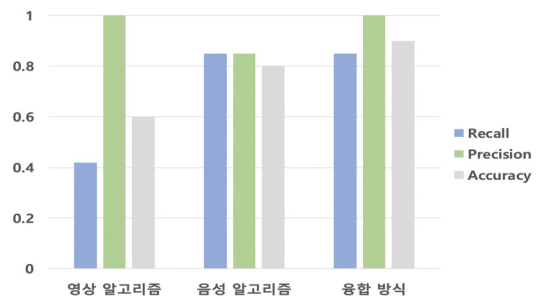
융합 방식은 영상 탐지 알고리즘과 음성 탐지 알고리즘을 결합하여 만들었다. 융합 방식은 영상 탐지 알고리즘의 단점과 음성 탐지 알고리즘의 단점을 각각 보완하여 높은 정확성으로 이상 징후를 탐지 할 수 있다. 우리는 다양한 시나리오를 통하여 융합 방식의 성능을 평가했다. 시나리오는 표 3과 같다.

본 연구에서는 다양한 시나리오를 통해서 한가지의 알고리즘을 이용했을 경우의 한계점과 두 가지 알고리즘을 결합한 융합 방식의 성능을 평가하였다. 영상 탐지 알고리즘의 경우 카메라 영역 내에서도 다양한 자세와 사람이 반만 나온 경우에도 정상적으로 탐지하는지를 실험해보았다. 음성 탐지 알고리즘의 경우 일반적인 상황에서는 정상적으로 탐지했지만 녹음되는 스마트폰이 소리 발생지점에서 멀리 있는 경우 이를 정상적으로 탐지할 수 있는지를 확인해보려고 한다. 융합 방식은 두 알고리즘을 결합하여 각 알고리즘의 단점을 서로 보완해주고 다양한 상황에서도 탐지할 수 있다는 것을 시나리오를 통해서 증명했다.

카메라 영역 내에서 발생하는 모든 상황들은 영상 탐지 알고리즘을 통해 정상적으로 탐지하였다. 하지만, 카메라 영역 밖에서 발생된 모든 상황에 대해서는 영상 탐지 알고리즘은 탐지할 수 없다. 음성 탐지 알고리즘은 소리가 발생된 상황에서 이상 징후가 발생할 때, 소리가 없어질 경우에 사람은 이상 징후가 발생하였지만, 패턴을 비교했을 때 정상이라고 판단할 수도 있기 때문에 정상적으로 탐지할 수 없다. 소리가 발생하는 가까운 거리에서는 정상적으로 탐지할 수 있지만, 거리가 멀고 소리가 작아질 경우에는 실제로 사람은 TV를 정상적으로 시청하고 있지만 음성 탐지 알고리즘은 이를 정상적으로 탐지하지 못한다. 융합 방식은 각각의 알고리즘들의 단점을 서로 보완하고, 영상과 음성 탐지 알고리즘이 사람의 행동을 정상적으로 탐지하지 못하는 시나리오 상황에서는 융합 방식도 탐지하지 못하지만 영상, 음성 탐지 알고리즘 한가지의 알고리즘만 정상적으로 탐지를 할 경우 융합 방식을 이용하여 정상적인 활동 및 이상 징후를 탐지 할 수 있었다.

각 알고리즘 및 융합 방식을 명확하게 비교하기 위하여 시나리오를 기반으로 Recall, Precision, Accuracy를 구하였다. 시나리오 중 대표적인 10개의 시나리오를 선정하여 각 알고리즘 및 융합 방식에 대한 Recall, Precision, Accuracy를 구하였고 이에 대한 결과는 그림 16과 같다.

그림 16을 통해 알 수 있듯이, 영상 알고리즘은 카메라 영역 밖의 행동을 전혀 알 수 없기 때문에 음성 알고리즘과 융합 방식에 비해 Accuracy가 현저히 떨어지는 것을 확인할 수 있다. 음성 알고리즘도 오디오를 녹음하는 기기와 가까이에서 발생하는 행동들에 대해서는 정상적으로 탐지할 수 있지만, 거리가 멀어지고 발생하는 소리가 작아진다면 정상적으로 탐지할 수 없게 된다. 융합 방식처럼 영상 알고리즘과 음성 알고리즘을 적절히 결합하고 융합하여 사용한다면 다양한 상황에서 높은 정확성으로 이상 징후를 탐지할 수 있다.



(Figure 16) Recall, Precision, Accuracy by Scenario-Based Algorithm

(Table 3) Experiment results of Individual algorithms and fusion method

상황 / 알고리즘	영상 탐지 알고리즘	음성 탐지 알고리즘	융합 방식
카메라 영역 내, 누워있기	O	O	O
누웠다 일어나기	O	O	O
뛰기	O	O	O
걸기	O	O	O
카메라 등지기	O	O	O
사람 오른쪽 반만	O	O	O
사람 왼쪽 반만	O	O	O
사람 상체만	O	O	O
소리가 있는 이상 징후	O	O	O
소리가 없는 이상 징후	O	O	O
소리가 있다가 없어지는 이상 징후	O	X	O
카메라 영역 밖, 누워있기	X	O	O
뛰기	X	O	O
TV 보기 거리 : 3M	X	O	O
TV 보기 거리 : 30cm	X	O	O
TV 보기 거리 : 4M	X	O	O
TV 보기 거리 : 4M 볼륨:20	X	X	X
음악 듣기	X	O	O
책상에 앉아 있기	X	O	O
소리 있는 이상 징후	X	O	O
소리 없는 이상 징후	X	O	O
소리가 있다가 없어지는 이상 징후	X	X	X

5. 결 론

우리는 홈 CCTV를 기반으로 하며 영상 패턴을 분석하여 이상 징후를 탐지하는 영상 탐지 알고리즘과 집에서 나는 소리를 분석 및 패턴을 탐지하여 분석하는 음성 탐지 알고리즘을 개발했다. 이를 결합한 융합 방식을 본 논문에서 제안했다. 융합 방식은 각 알고리즘의 단점을 보완하고 집안의 넓은 영역에서 이상 징후를 탐지했다. 우리는 본 연구를 통해 별도의 비용 없이 사람들이 집에서 사용하고 있는 장비들을 이용하여 집에 혼자 사는 1인 가구의 구성원이 집 안에서 심각한 부상을 당했을 때의 이상 징후를 탐지하는 것을 목표로 했다. 추후에는 다양한 센서들을 추가하거나 인공지능 알고리즘들을 결합하여 사회에 이바지 될 수 있는 다양한 연구들을 진행해 볼 계획이다.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF), grant funded by the korea government (MSIP; Ministry of Science, ICT & Future Planning) (No. NRF-2017R1C1B5017847). This was also supported by Korea National University of Transportation in 2018

참고문헌(Reference)

- [1] The Korea Times, "Sensor to be installed to prevent lonely death", http://www.koreatimes.co.kr/www/biz/2018/01/602_243308.html, 2018.
- [2] Inter Press Service News Agency, "The Rise of One-Person Households", <http://www.Ipsnews.net/2017/02/the-rise-of-one-person-households>, 2018.
- [3] MBC News, "If you don't move, I'll warn you. Internet of Things to Prevent Solitude", http://imnews.imbc.com/replay/2018/nwdesk/article/4586838_22663.html, 2018
- [4] Glen Debard, Marc Mertens, Toon Goedemè, Tinne Tuytelaars and Bart Vanrumst, "Three Ways to Improve the Performance of Real-Life Camera-Based Fall Detection Systems", *Journal of Sensors* (2017) <https://doi.org/10.1155/2017/8241910>
- [5] Miao Yu, Liyun Gong, Stefanos Kollias, "Computer vision based fall detection by a convolutional neural network", *ACM* (2017) <https://doi.org/10.1145/3136755.3136802>
- [6] Koldo de Miguel, Alberto Brunete, Miguel Hernando and Ernesto Gambao, "Home Camera-Based Fall Detection System for the Elderly", *Multidisciplinary Digital Publishing Institute (MDPI), Sensors*, 21. (2017) <https://doi.org/10.3390/s17122864>
- [7] Fouzi Harroua , Nabil Zerroukib , Ying Suna , Amrane Houacineb, "Vision-based fall detection system for improving safety of elderly people", *IEEE Instrumentation and Measurement Society*, 21, (2017) <https://doi.org/10.1109/MIM.2017.8121952>
- [8] Philip Geismann. Georg Schneider, "A Two-staged Approach to Vision-based Pedestrian Recognition Using Haar and HOG Features", *IEEE*,6 (2008) <https://doi.org/10.1109/IVS.2008.4621148>
- [9] Huang J, Rathod V, Sun C, Zhu M, Korattikara A, Fathi A, Fischer I, Wojna Z, Song Y, Guadarrama S, Murphy K, "Speed/accuracy trade-offs for modern convolutional object detectors", *CVPR* 2017, https://github.com/tensorflow/models/tree/master/research/object_detection
- [10] Ibrahim Furkan Ince, Mustafa Eren Yildirim, Yucel Batu Salman, Omer Faruk Ince, Geun-Hoo Lee and Jang-Sik Park, "Fast Video Fire Detection Using Luminous Smoke and Textured Flame Features," *KSII Transactions on Internet and Information Systems*, vol. 10, no. 12, pp. 6048-6069, 2016. DOI: <https://doi.org/10.3837/tiis.2016.12.019>
- [11] L. Vuegen B. Van Den Broeck P. Karsmakers J. F. Gemmeke B. Vanrumste H. Van hamme, "an mfcc-gmm approach for event detection and classification", *IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events*, pp.1-3 2013, <https://www.semanticscholar.org/paper/An-Mfcc-gmm-Approach-for-Event-Detection-and-Vuegen-Broeck/bf4d54fc69e19ace82d87231c45bf5786e19bffd>

[12] Minkyu Lim, Donghyun Lee, Hosung Park, Yoseb Kang, Junseok Oh, Jeong-Sik Park, Gil-Jin Jang and Ji-Hwan Kim, "Convolutional Neural Network based Audio Event Classification," KSII Transactions on Internet and Information Systems, vol. 12, no. 6, pp. 2748-2760, 2018. DOI: <https://doi.org/10.3837/tiis.2018.06.01>

[13] Junho Ahn, Richard Han, "myBlackBox: Blackbox

Mobile Cloud Systems for Personalized Unusual Event Detection", Sensors, 2016, Volume 16, Issue 5, 753; doi: <https://doi.org/10.3390/s16050753>, 20 pages.

[14] JuHo Jung, HwiJune Park, JunHo Ahn, "Unusual Event Detection Algorithm via Personalized Video and Voice Patterns for Preventing Solitary Death", Proceedings of the 37th KSII Spring Conference, v.19 n.1, pp.7~8, 2018

● 저 자 소 개 ●



정 주 호(Ju-Ho Jung)

2013년~현재 : 한국교통대학 컴퓨터정보기술공학부 소프트웨어학전공(학·석사)

관심분야 : 인공지능, 지능형IoT, 영상 및 음성처리

E-mail : jjis1005k@ut.ac.kr



안 준 호(Jun-Ho Ahn)

2009년~2013년 University of Colorado, Boulder, Computer science, Ph.D.

2013년~2017년 ETRI 국가보안기술연구소

2017년~현재 한국교통대학 컴퓨터정보기술공학부 소프트웨어학전공 교수

관심분야 : 인공지능, 지능형IoT, 패턴처리

E-mail : jhahn@ut.ac.kr