Kinect Based Suspicious Posture Recognition for Real-Time Home Security Applications

Mandikal Vikram*, Aditya Anantharaman*, Suhas BS*, Ashwin TS, Ram Mohana Reddy Guddeti
Department of Information Technology
National Institute of Technology Karnataka, Surathkal, India 575025
Email: {15it217.vikram, 15it201.aditya.a, 15it110.suhas}@nitk.edu.in, {ashwindixit9, profgrmreddy}@gmail.com

***Abstract*—Suspicious posture recognition is a paramount task with numerous applications in everyday life. We explore one such application in real-time home security using the Microsoft Kinect depth camera. We propose a novel method where the remote device itself detects the suspicious activity. The server is alerted by the remote device in case of a suspicious activity which further alerts the home owners immediately. We show that our method, works in real-time, is robust towards changing lighting conditions and the computations happen on the remote device itself which makes it suitable for real-time home security.**

***Keywords*—Posture recognition, Home security, Kinect, Android app**

I. Introduction

Home security is a very important domain of security. The growing crime rates across cities showcase the bitter reality. Many people underestimate the need to put in place appropriate home security measures. A burglary or theft can often lead to devastating consequences for the homeowners. The current security systems rely largely on visual features such as facial features which can easily be masked. The criminals could mask themselves with sunglasses, caps or use more sophisticated techniques to deceive such visual feature dependent security systems. Also, these systems fail in adverse lighting conditions.

Hence, we need reliable intelligent systems which can alert homeowners at real-time, irrespective of the lighting conditions of the place or the clothes worn by the criminals. In this work, we propose a real-time system for home security using the Microsoft Kinect depth camera which checks for suspicious postures outside the door and alerts the homeowners using a push notification on their phones. Since our method uses the skeletal information captured by the Kinect depth camera (which is done using infrared rays) and not the visual features, it is thus robust towards lighting conditions and clothes of criminals (intruders), which makes it suitable for home security applications.

Our main contribution in this paper is a novel real-time home security solution for suspicious posture detection which can be deployed remotely with minimalistic interaction with the web server.

*** equal contribution**

978-1-5386-8325-7/18/$31.00 ©2018 IEEE

Our Approach for real-time home security consists of:

- Using Kinect depth camera to detect human postures from the skeleton.
- Classify postures as suspicious or normal using logistic regression.
- In case the posture is suspicious, then alert the homeowners by a push notification on their phones.

All the computations involving the inference part of the machine learning algorithm, occur on the remote device itself. This avoids unnecessary communication overhead between the server and the remote device. The remote device communicates with the server only when it detects a suspicious activity.

II. Related Works

Kinect is extensively used by researchers for gesture and posture recognition, fall detection, in-home security and surveillance and 3D modeling. The IR Sensor or the depth camera of Kinect was used by [1] for fall detection. In this work the authors have demonstrated a real-time application to detect walking falls accurately and robustly. The depth camera was used by [2] for gesture recognition. Here, the authors used the body-joint positions obtained from Kinect and then applied machine learning techniques on these, to classify the gestures. In [3], authors developed a home security surveillance system which can detect abnormal activities such as intrusion detection, fire, smoke etc. using sensors and CMOS camera. A Short Message Service (SMS) or Multimedia Messaging Service (MMS) notification is sent on detection of suspicious activity. The authors in [4] developed a multi-sensor fusion algorithm that includes Kinect devices also, to detect abnormal activities among the elderly. They used numerous sensors such as bands, Zenith cameras and other wearables which contain accelerometer, gyroscope, heart rate sensor etc. to monitor the patient. The work [5] discussed about monitoring elderly home occupants. The authors used visual features from Kinect devices placed at multiple locations inside the house to learn a few fuzzy parameters. These parameters are further used to detect abnormal behavior patterns. A disadvantage of this work is the usage of multiple cameras for different locations, which might not be practically viable.

To the best of our knowledge, there have been no works on home security using Kinect's depth camera. Table I

summarizes some of the works that use Kinect, with their methodology, merits and demerits.

III. Proposed Methodology

We propose a novel technique for real time home security using suspicious posture recognition with the help of Kinect depth camera and logistic regression. The proposed methodology consists of the following steps:

- Data Acquisition: Capture skeleton information using Kinect depth sensor.
- Data Processing and Feature Extraction: Features extracted by calculating the joint angles.
- Posture Recognition using Logistic regression.
- Alerting the homeowners using a push notification.

The proposed methodology is illustrated in Fig. 1. The Microsoft Kinect depth camera captures the skeletal information of the subject. The angles between the joints are used as the features to classify the postures. This is done on a frame to frame basis - if the classifier classifies the posture as suspicious for a certain threshold number of continuous frames, a photo is taken using a camera - not necessarily the Kinect camera, it could be any camera which will have a good view of the subject. This photo is then forwarded to the server which stores it and also notifies the concerned homeowners using a push notification to their phone.

The premise assumed for our approach is shown in Fig. 2. The Kinect is placed in front of the door such that any person approaching the door would be in between the door and the Kinect. Thus, the Kinect will capture the position of the subject from behind. Another camera which would have a good view of the subject would be required to take a photo in case of an alarm. A secondary camera such as a CCTV camera can be used for this purpose.

A. Feature Extraction from Skeletal Information

Skeletal information is obtained from Kinect depth camera directly and is plotted on the depth image. It provides the coordinates of these joints. The joints which we consider are - left shoulder, right shoulder, head, neck, left hand, right hand, left elbow and right elbow. The following vectors are calculated between the joints - neckHead (between neck and head), neckLeftShoulder (between neck and left shoulder), neckRightShoulder (between neck and right shoulder), rightShoulderElbow (between right shoulder and right elbow), leftShoulderElbow (between left shoulder and left elbow), rightElbowHand (between right elbow and right hand), leftElbowHand (between left elbow and left hand) and torsoNeck between the torso and the neck. These vectors are further used to calculate the following angles - between rightShoulderElbow and rightElbowHand, between leftShoulderElbow and leftElbowHand, between neckrightShoulder and rightShoulderElbow, between neckleftShoulder and leftShoulderElbow, between torsoNeck and the y-axis and between torsoNeck and neckHead, these 6 angles are the features which are used for classifying the postures.

B. Classification using Logistic Regression

Logistic regression is a simple classifier that is widely used for binary classification and the details are shown in Algorithm 1. It can describe and find relationships between a dependent and an independent variable. The sigmoid or logistic function is defined as follows:

$$\sigma(z) = \frac{1}{1 + \exp(-z)} \tag{1}$$

The function(1) also serves as the hypothesis. It can also be observed that the range of this function is [0, 1]. So, this function provides a probability measure for classification. For labeled training examples $(x^i, y^i) : i = 1..m$, the cost function is defined as follows:

$$J(\theta) = -\frac{1}{m}(\sum_{i=1}^{m} y^{(i)} log(h_\theta(x^{(i)})) + (1-y^{(i)}) log(1-h_\theta(x^{(i)})) \tag{2}$$

The cost function (2) measures how well the hypothesis fits the data. The objective is to minimize the cost function. We use gradient descent for this purpose. The weights are modified as follows:

$$\theta_j := \theta_j + \alpha(y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)} \tag{3}$$

where α is the learning rate.

Algorithm 1: Logistic Regression

```
initialize vector θ to random values ;
m = number of training examples;
while iter < maxiter do
    calculate cost J(θ);
    i = 0;
    while i < m do
        calculate the hypothesis h_θ = σ(θ^(T) X^(i));
        update θ using (3);
        i = i+1;
    end
    iter = iter+1;
end
```

Here, the problem is a two way classification into suspicious and normal actions with six features which are the angles as mentioned before. Logistic regression suits well for this data.

C. Scenarios Considered

In this work, we considered the below mentioned scenarios while training the model. We created the data based on these scenarios.

1) Normal Activity: These include the normal postures without any suspicion. It is important for the system to not unnecessarily send alarms to the homeowner. The image from the secondary camera and the depth camera can be seen in Fig. 3(a) and Fig. 3(e). Many casual poses of ordinary people were taken to train the model to identify these normal postures. The number of normal samples in the generated dataset is equal to the sum of the the number of samples in all other scenarios.

TABLE I
RELATED WORKS

Author	Methodology	Merit	Demerit
Mastorakis et al. [1]	Features extracted from the 3D bounding boxes obtained from the IR camera are used for fall detection	It is a robust walking fall detection system that requires no preknowledge of the scene	This deals with safety inside the home while our work is towards security at the home door.
Biswas et al. [2]	Kinect used for obtaining depth images which are used for gesture classification.	Gesture recognition - the computations are cheaper since depth images are used instead of RGB images. Classification into 8 classes using SVM which can be extended to more classes	This work addresses the problem of gesture recognition into a few very simple classes, which is already has many established techniques. No direct practical application of this work.
Zhang et al. [6]	Recognition of activities of daily life using RGB-D cameras including fall detection for the elderly.	Uses depth cameras which can handle illumination changes and protects privacy	Handles just 5 types of fall detection which might not take into account all types of falls.
Nar et al. [7]	Obtains the skeletal features from the Kinect and classifies them into suspicious or normal activities.	Effectively uses just six angles as features to classify the postures.	No notification to administrators or anyone else. ATMs already have well established security mechanisms - this raises questions about the practical applicability of this work.

2) Aggressive Activity: It is important to identify aggressive activities as they are an indication of a possible threat. An aggressive posture can be seen in Fig. 3(b) and Fig. 3(f). It can be seen in the depth image that the classifier detected it as an abnormal activity. The generated dataset included more such aggressive postures.

3) Pushing the door Activity: An activity of a person pushing the door is a direct threat. It is almost certain that the person has malicious intent. One such activity captured can be seen in Fig. 3(c) and Fig. 3(g). As seen the classifier has detected it as an abnormal activity and an alarm was raised. It is important the homeowner must be alerted about such activities immediately.

4) Break open the door Activity: This another activity which is an indication of a definite threat to the home. The homeowner needs to be alerted about this activity to take immediate action. One such activity is shown in Fig. 3(d) and Fig. 3(h). It can be clearly be seen in the Fig. 3(d) that the intruder is trying to break the door. Hence, the depth image shows that the activity has been classified as abnormal.

It is important that the classification is binary into normal and suspicion activities. The aggressive activities, the pushing the door activities and breaking the door activity comprise of the abnormal behavior. It is not important to further classify these suspicious activities into particular scenarios as it would increase the latency involved in alerting the homeowner which is not desirable. The main objective is to take a picture of the suspected activity and notify the same to the owner with as less delay as possible - a two way classification into suspicious and not suspicious activities serves this purpose.

D. Raising Alarm

The successive frames captured by the Kinect are processed to obtain the above mentioned features. These features are then passed to the logistic regression classifier. If the classifier predicts that the posture is suspicious for a certain threshold number of frames, then an alarm needs to be raised. The secondary camera (CCTV camera) immediately captures a picture which would give a good view of the suspect - this is immediately sent as a post request to a web-server. We hosted the website on Azure-Web Services in this proposed work. The web-server on receiving the image, stores it and sends a notification to the homeowners on their android phones. The azure notification hub was used for this. Thus, the homeowners are alerted at real-time and can take the necessary action if required.

E. System Implementation

The classifier was trained on 360 data points which were generated by us. The calculated weights are then used for real-time classification using Processing and SimpleOpenNI toolbox.

Algorithm 2: Inference Time Computations

```
score = 0;
while joint coordinates are incoming from the Kinect do
    compute X using the joint coordinates;
    h_θ = σ(θ^(T) X^(i));
    if h_θ > confidence_threshold then
        score = score + 1;
        if score > num_frame_threshold then
            capture picture and send alarm to server;
            score = 0
        end
    else
        if score > 0 then
            score = score − 1;
        end
    end
    iter = iter+1;
end
```

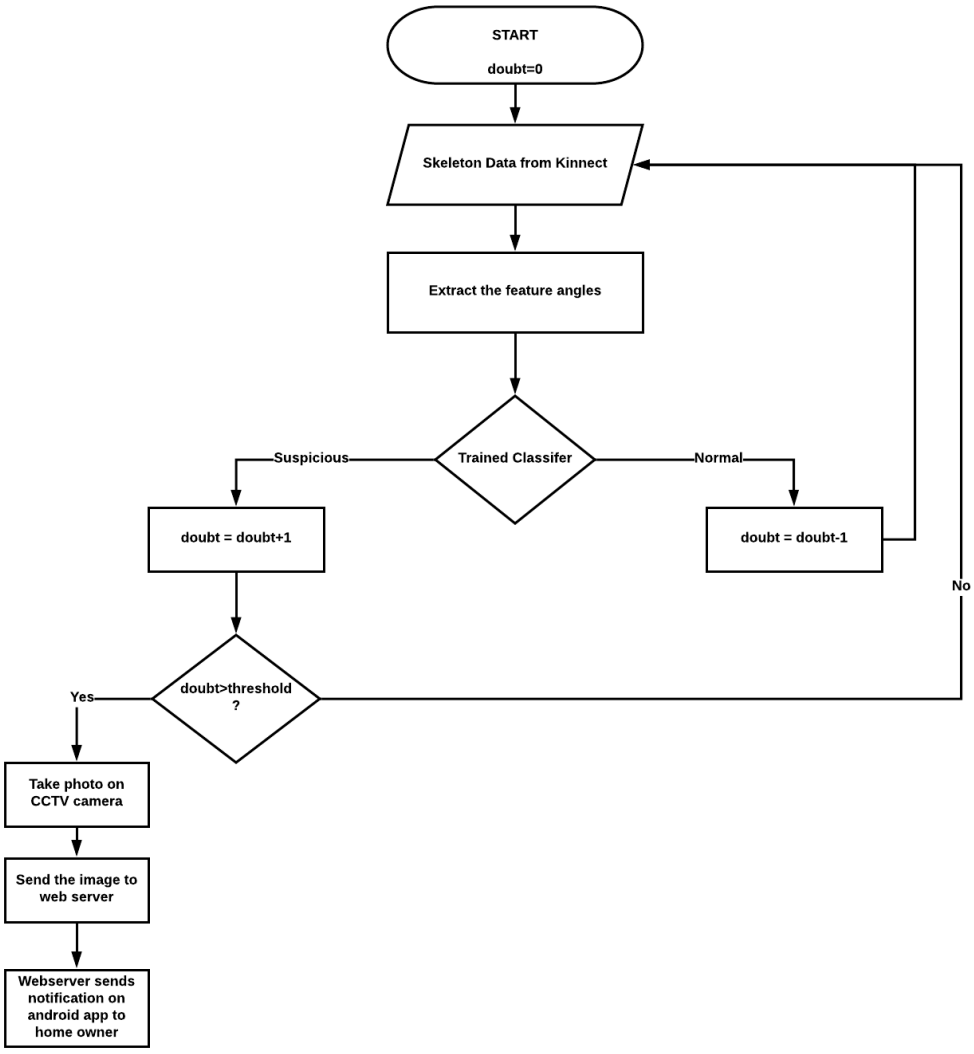


Fig. 1. Flowchart of the Methodology involved

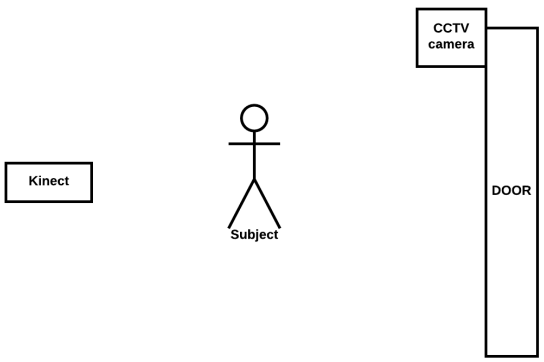
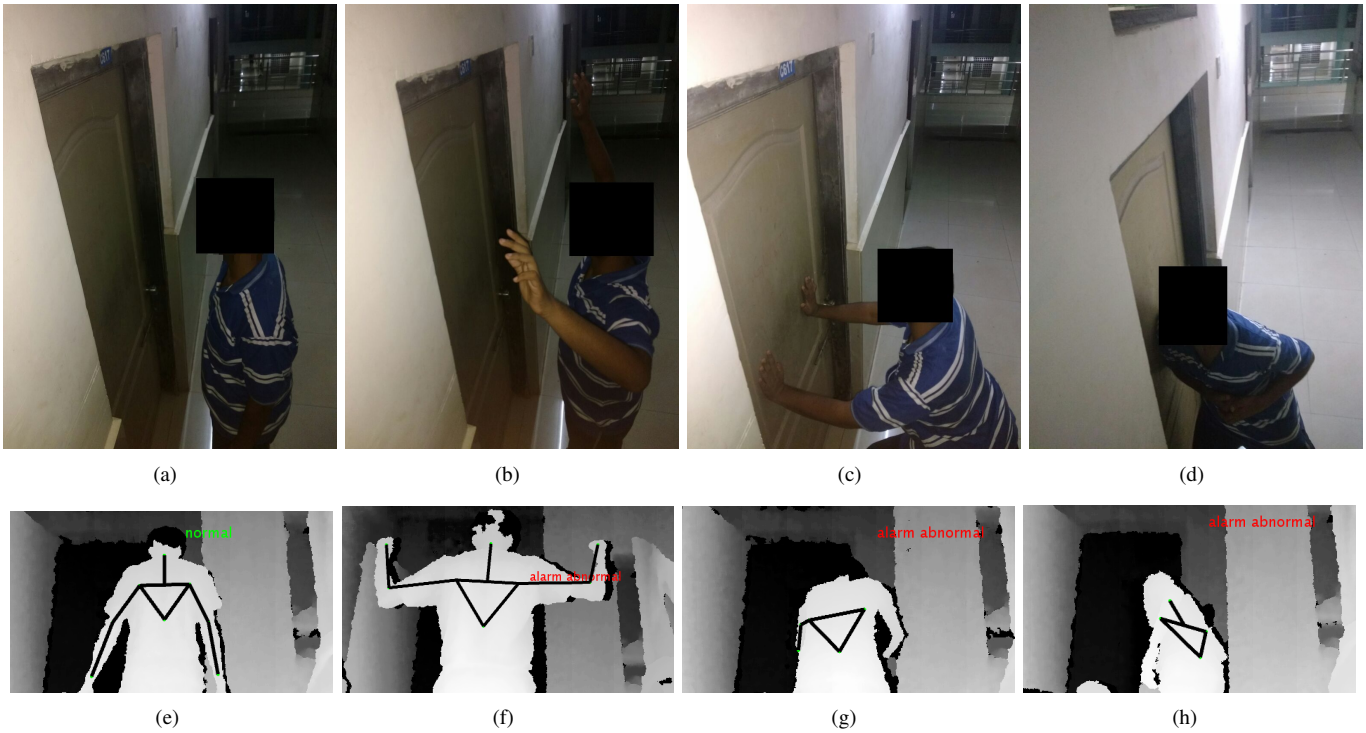


Fig. 2. Premise for our approach

F. Inference time Computations

The trained weight matrix θ is used during the inference time. X denotes the 6 angles which are computed using the coordinates of the joints provided by the Kinect. The inference time computations are described in Algorithm 2. Algorithm 2, keeps track of a score which indicates how likely a suspicious activity has occurred. For each frame, the 6 angles are computed and these along with trained weights are used to predict if the particular frames consists of a suspicious activity. The computations involved in the inference are just 6 multiplications, an addition and a calculation of a sigmoid, which shows that it is suitable for deployment on remote devices and can run in real-time. A frame is classified as suspicious if the confidence is above a certain threshold referred as the $confidence\_threshold$ which was found to be 0.89 experimentally. Each frame which is classified as a suspicious activity increases the score and each frame which is classified as not suspicious reduces the score (score cannot go below 0). When the score crosses a threshold referred as the $num\_frame\_threshold$ (which was set to 6) an alarm is



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 3. (a)-(d) are camera images (with face censored). (e)-(h) are Kinect depth images. (a) and (e) denote Scenario 1 - Normal Activity. (b) and (f) denote Scenario 2 - Aggressive Activity. (c) and (g) denote Scenario 3 - Pushing Activity. (d) and (h) denote Scenario 4 - Break Open the Door Activity.

sent to the server along with captured image.

IV. Results and Analysis

Fig. 3 shows the various scenarios described in the previous section.

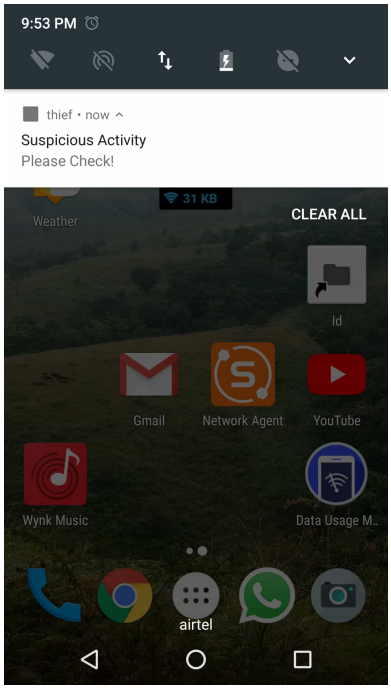


Fig. 4. Screenshot of the notification on the android phone

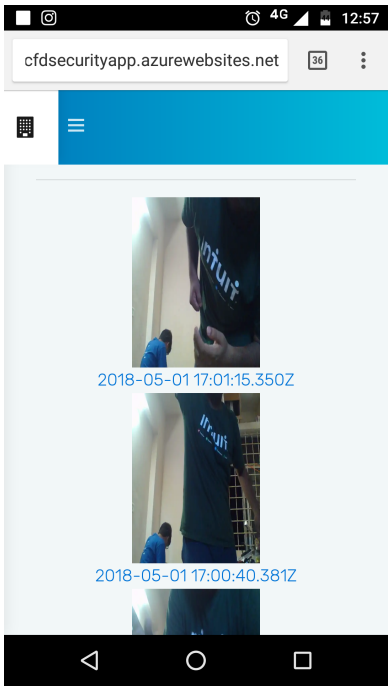


Fig. 5. Suspicious images with date and time displayed on the website

The metrics used for evaluation are defined as follows:
True Positive (TP): When a suspicious posture is detected by the device and it is a suspicious posture.
False Positive (FP): When a suspicious posture is detected by the device and it is a normal posture.

TABLE II
CONFUSION MATRIX FOR SUSPICIOUS AND NORMAL POSTURES

	Suspicious (Actual)	**Normal (Actual)**
Suspicious (Predicted)	15	2
Normal (Predicted)	3	30

TABLE III
EVALUATION METRICS

Precision(%)	**Recall(%)**	**Accuracy(%)**
88.23	83.33	90

False Negative (FN): When no suspicious posture is detected by the device and it is a suspicious posture.
True Negative (TN): When no suspicious posture is detected by the device and it is a normal posture.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

We define precision, recall with respect to detecting the suspicious posture class and accuracy with respect to both the classes as in (4), (5) and (6). We evaluated our methodology on 50 different postures which were not considered to train the model. The evaluation metrics calculated are precision, recall and accuracy. The results are recorded in Table II and Table III.

The screenshots of the notification to the homeowner can be seen in Fig. 4 and the images captured by the secondary camera as viewed on the server can be seen in Fig. 5. The average latency for the notification on the android phone after the suspicious activity was performed was 2 seconds only. This depends on the speed of the network to which the android phone is connected, hence could vary by a small margin from user to user, however this can definitely be considered as a real-time suspicious activity notification system.

V. CONCLUSION AND FUTURE WORK

In this work, we propose a technique for real-time home security applications which can be deployed directly on the remote device with minimal interaction with the web server. We show that lag between the suspicious activity and the homeowners getting notified on their android devices is just about 2 seconds, while obtaining an accuracy of 90% on real world users. The proposed method is also inherently robust towards changing lighting conditions and other factors which affect techniques relying on visual features. Hence, we emphasize that the proposed framework is highly suitable for real-time deployment in home security applications. Future directions of work could include extending the system to more security applications and online learning.

REFERENCES

[1] G. Mastorakis and D. Makris, "Fall detection system using kinects infrared sensor," *Journal of Real-Time Image Processing*, vol. 9, no. 4, pp. 635–646, 2014.

[2] K. K. Biswas and S. K. Basu, "Gesture recognition using microsoft kinect®," in *Automation, Robotics and Applications (ICARA), 2011 5th International Conference on*. IEEE, 2011, pp. 100–103.

[3] J. Hou, C. Wu, Z. Yuan, J. Tan, Q. Wang, and Y. Zhou, "Research of intelligent home security surveillance system based on zigbee," in *Intelligent Information Technology Application Workshops, 2008. IITAW'08. International Symposium on*. IEEE, 2008, pp. 554–557.

[4] G. Hernández-Penaloza, A. Belmonte-Hernández, M. Quintana, and F. Alvarez, "A multi-sensor fusion scheme to increase life autonomy of elderly people with cognitive problems," *IEEE Access*, vol. 6, pp. 12 775–12 789, 2018.

[5] H. Pazhoumand-Dar, C. P. Lam, and M. Masek, "A novel fuzzy based home occupant monitoring system using kinect cameras," in *Tools with Artificial Intelligence (ICTAI), 2015 IEEE 27th International Conference on*. IEEE, 2015, pp. 1129–1136.

[6] C. Zhang and Y. Tian, "Rgb-d camera-based daily living activity recognition," *Journal of Computer Vision and Image Processing*, vol. 2, no. 4, p. 12, 2012.

[7] R. Nar, A. Singal, and P. Kumar, "Abnormal activity detection for bank atm surveillance," in *Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on*. IEEE, 2016, pp. 2042–2046.