Recognition of Theft by Gestures using Kinect Sensor in Machine Learning

[1]Surbhi Kasat, [2]Kanchan Tupe, [3]Vaishnavi Kshatriya, [4]Roshani More, [5]Ms. Prachi S. Tambe

[1,2,3,4]Student, [5]Assistant Professor
Computer Department, SVIT, Nashik, Maharashtra, India

***Abstract*: Detection of crimes is the prime necessity nowadays. Cameras have been popular as measures against thefts. They are useful for deterrence of crimes such as theft and murders, it is possible to detect crimes by installing artificial intelligence to surveillance cameras. In this paper, Kinect motion sensor Device is used to recongnize the gesture of theft. As a real-time recognition of large set of dynamic gestures are needed for that algorithm is used to recognize the gesture and translate them. It includes training cameras. A large amount of gesture and posture data is used to analyze the behavior of person that means it's a Theft or not by the image processing and the data size is larger as they are image data, for that Dynamic time wrapping is used to analyse the object and person. We propose an idea to judge a behavior gesture of a person using Kinect sensor in order to detect crimes at real-time.**

***Keywords*: Abnormal behaviour, Kinect, Posture Recognition, Speech to Text, Alarm, SMS, Image Processing.**

1. INTRODUCTION

In recent years, cases such as theft and murder have been steadily occurring, and the number of occurrences is also increasing. The number of placed surveillance cameras has been increasing in order to deter crimes or identify suspects after the incident. In addition, due to the development of computer technology such as image processing and artificial intelligence, suspicious people can be detected. The camera will recognize the action being performed by the user, it gives skeleton of human body when user stand in front of kinect sensor. This actions are compared with action stored it dictionary. A dictionary is maintain with all gestures and related speech are stored. If match is found then the SMS and alarm is generated. That means we are making a desktop application that will recongnize the theft action means the gun holding action and many more gesture were catched by the kinect sensor.(kinect camera).

2. Problem Statement-

Surveillance cameras have been popular as measures against criminals. They are useful for deterrence of climes such as theft and murder, but they cannot prevent crimes since real-time watching must be required for real-time detection of the crimes this system is proposed for detection of such Harmful incidences.

3. System Architecture

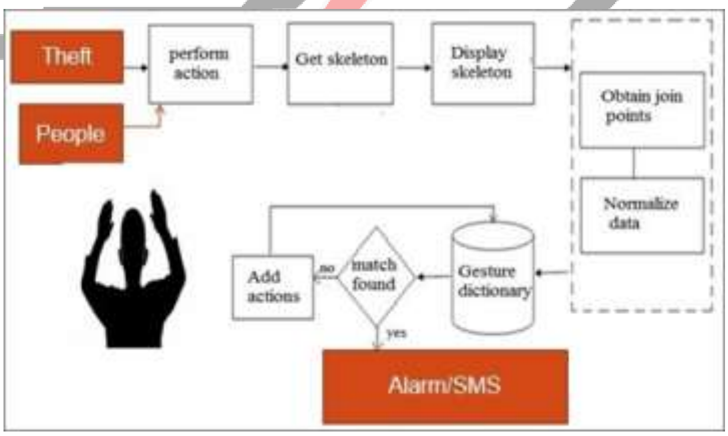

Fig. System Architecture

Consider the block diagram from Fig 2. The user is in front of the camera doing a action or getting ready to do so. With a frame rate of 20fps, a new frame is obtained and the video stream is updated with the skeleton of the user overlapped onto it. At that point, if the user wants to record a sequence (otherwise, the system asks the camera to get the next frame), three main blocks are executed: the first block consists of obtaining the data of the joints of interest (JoI) required for the frame descriptor, the second block consists of normalizing these data, and the third one consists of building the frame descriptor. Then, if the working mode is set to TRAINING (meaning that the user is adding a new gestues to the training set), the frame descriptor is added to the correspondent file of the dictionary. Otherwise, if the mode is set to TESTING the frame descriptor is added to the current test sample. Then, the system checks if the current frame is the last frame of the sign. After a sign is finished and if the working mode is TESTING, the test sign is compared using a classifier with the signs from the dictionary and the corresponding output is displayed so that the any unwanted incidence is happen is known. After that, the system keeps going with the next frame and the flow of the block diagram is repeated again.

2. Implementations:

A. Joints of interest (JoI) :

OpenNI/NITE can track up to 15 joint positions. After carefully studying the signs of the proposed default dictionary for the system, only 4 joints out of the 15 resulted to be significant for the description of a sign: both hands and both elbows. There is no point in tracking others joints such as the shoulders, the knees, the feet, etc. because they remain almost static during the execution of the sign. Adding these joints to the sign descriptor will be the same as adding redundant information. Even though the description step can be done using the four previously mentioned joints, some other joints are also required for the normalization and the sign modeling steps. There are head and tarso joint.

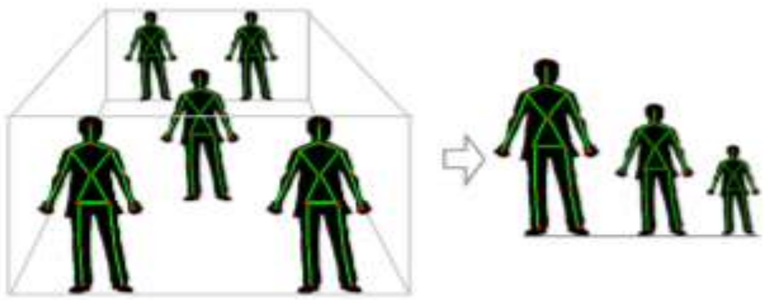

Fig. Normalize user data

3. Normalization **of the data:**

1) Invariant to the user's position:

The normalization must take into account the position of the user. The slight variation in depth can cause a considerable variation of the X and Y values. The distances between one joint and another one can drastically vary depending on the position of the user. Instead of directly storing the Cartesian coordinates X,Y, and Z (which can be obtained using OpenNI/NITE), the proposal consists in normalizing all the joint coordinates with respect to the position of the TORSO. This position remains always constant along the sign frames and is the right one to be used to make the system position-invariant. In mathematics, a spherical coordinate system is a coordinate system for three-dimensional space where the position of a point is specified by three numbers: the radial distance of that point from a fixed origin (r), its polar angle measured from a fixed zenith direction (θ), and the azimuth angle of its orthogonal projection on a reference plane that passes through the origin and is orthogonal to the zenith, measured from a fixed reference direction on that plane (ϕ).
The radial distance r will be expressed by d and defines a vector between the TORSO and the correspondent joint. (θ and ϕ) are the angles that describe the direction of this 3D vector.

Given the set of joints J = {LE, RE, LH, RH} and considering T as the TORSO, the set of distances
D = {dLE, dRE, dLH, dRH}, and the sets of orientations
Θ = {θLE, θRE, θLH, θRH} and
Φ = {ϕLE, ϕRE, ϕLH, ϕRH} are defined as follows:

$$\sum_{i=1}^{n} D(i) = \sqrt{(J(i)x - Tx)^2 + (J(i)y - Ty)^2 + (J(i)z - Tz)^2}$$

$$\sum_{i=1}^{n} \theta(i) = \text{atan2}\left(\sqrt{(J(i)x - Tx)^2 + (J(i)y - Ty)^2}, (Tz - J(i)z)\right)$$

$$\sum_{i=1}^{n} \emptyset(i) = atan2\ (J(i)y - Ty), (J(i)x - Tx)$$

where n is the number of joints from J.

2) Invariant to user's size:

Given a sign, its description must be the same no matter if the user is tall or short. Although the way that the dictionary is built allows it to have several samples (meaning that we can have the same gesture described for different user's sizes), there is no way to add the samples for all the possible user's sizes to the dictionary. Otherwise the classification process will become slower and less accurate. After the normalization of the user's position, every joint is expressed by its relative distance d to the TORSO joint and the two angles θ and ϕ that describe the orientation of this distance. The proposal shown in Fig 6(b) consists of normalizing all the relative distances d by the factor that is defined by the distance between the HEAD and the TORSO joints (dHT). This factor tells about the size of the user and all the distances D can be normalized accordingly to this value.

Fig, Different size of user

Given the set of distances D = {dLE, dRE, dLH, dRH}, the normalized set of distances Dnorm is obtained as follows:

$$\sum_{i=1}^{n} Dnorm(i) = \frac{D(i)}{dHT}$$

where n is the number of distances from D and dHT is the HEAD-TORSO distance . There is no need to normalize the angles θ and ϕ since they are expressing the direction and the direction remains the same after the normalization.

C. Gesture descriptor

Once the JoI data are obtained and normalized, the next step is building a descriptor for each posture. The descriptor must be able to describe a posture in a way that this descriptor will be unique and sufficiently different from the other descriptors of the dictionary.

D. Data processing and Feature extraction:

With the skeleton tracked by Kinect, joint positions are obtained. Since, joint vector has 3 coordinates and a skeleton consist of 20 joints, the feature vector has 60 dimensions. Apart from the feature described above, another feature can be extracted by calculating the joint angles. While working with postures, we observed that ten joints, namely Torso, Neck, Head, Left shoulder, Left elbow, Left wrist, Right shoulder, Right elbow, Right wrist, Left hip and Right hip, are the most important joints for representing postures. From these joints, we can calculate different sets of angles. Subsequently, we defined the following angle based features to recognize desired postures: angle between vector joining left hand to left elbow and vector joining left elbow to left shoulder, angle between vector joining left elbow to left shoulder and left shoulder to torso, angle between vector joining right hand to right elbow and right elbow to right shoulder, and angle between vector joining right elbow to right shoulder and right Shoulder to torso. The calculation of the subject's posture is based on the fundamental consideration that the orientation of the subject's torso is the most characteristic quantity of the subject during the execution of any action.

E. Classifier

The classifier is the function that will output the corresponding alarm once the user inputs is done
Given an input sequence of frames, the classifier will match it with the closest sequence of frames (from the default dictionary.

Class Diagram:

Class diagrams are the most popular UML diagrams used for construction of software applications. So it is very important to learn the drawing procedure of class diagram.
Class diagram have a lot of properties to consider while drawing but here the drawing will be considered from a top level view. Class diagram is basically a graphical representation of static view of the system and represents different aspect of the application.so collection of class diagrams represent the whole system.

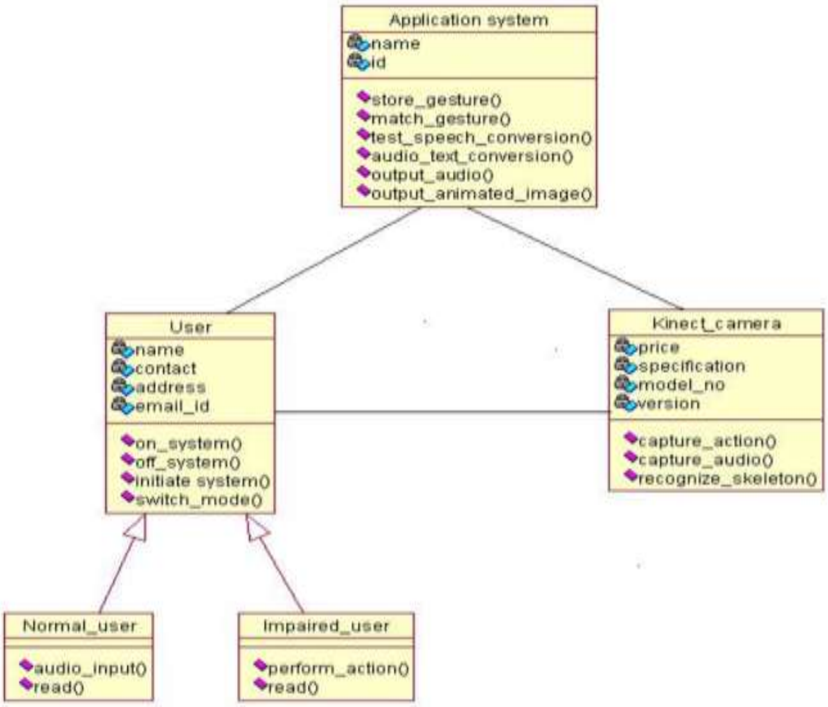

Fig 6. Class Diagram

RESULS

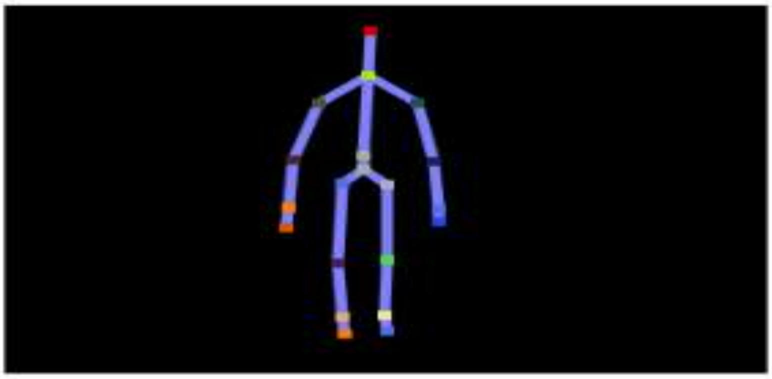

Fig 7. Skeleton Recognition with joints

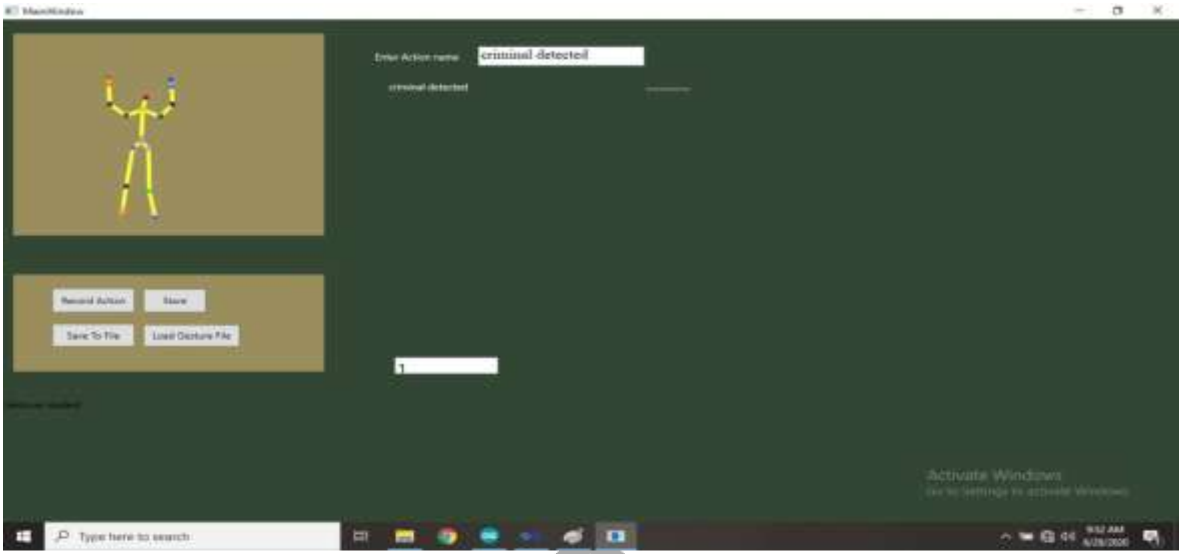

Fig 8. Criminal detected

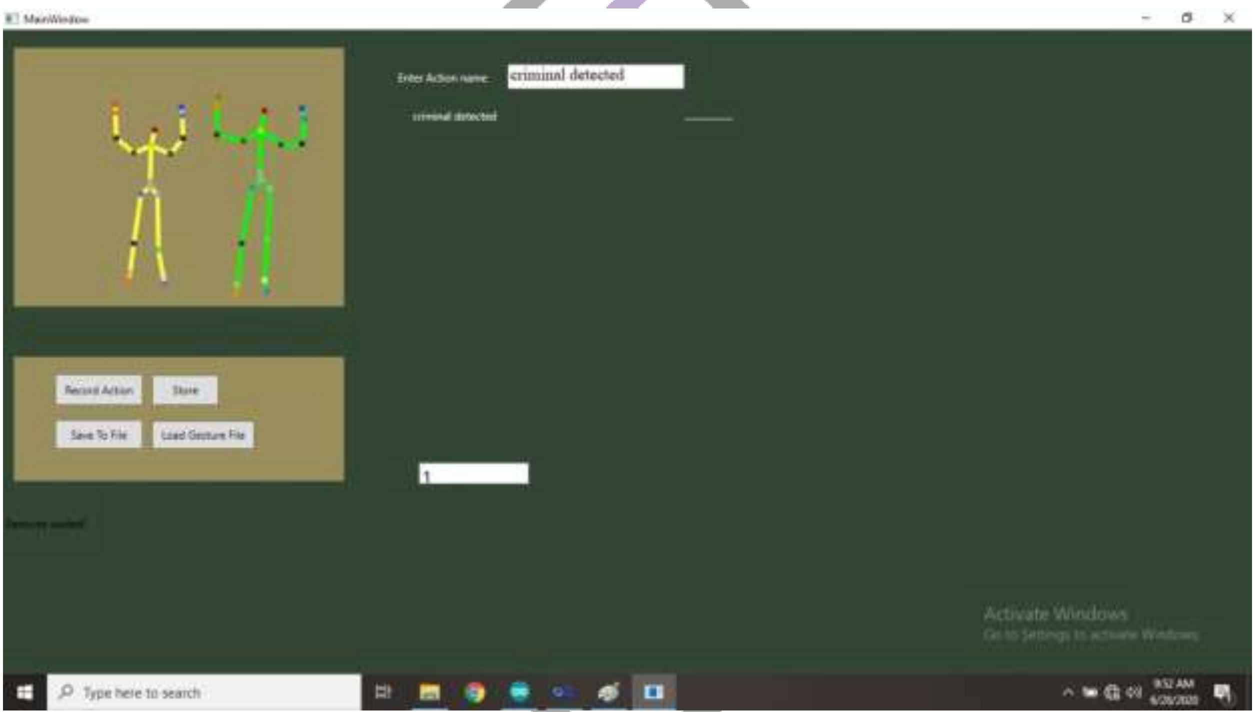

Fig 9. Multiple Criminal detected

CONCLUSION

We develope a system in which posture recognition can be used to detect abnormal behavior of a person in the Bank, malls, shops etc. We achieved this by using skeleton data that can be extracted from the depth image provided by a 3D camera like Kinect. Surveillance cameras are effective to deter theft and murder. However, they cannot prevent incidents in advance unless video is manually watching in real-time. Tech- nologies of image recognition and articial intelligence have made the surveillance cameras detect the incidents in real-time with their gestures. In Future you can develope an Android application for the same working.

REFERENCES

[1] Rajvi Nar Alisha Singal Praveen Kumar," Abnormal Activity Detection for Bank ATM Surveillance", August 2016
[2] Min Yi, "Abnormal Event Detection Method for ATM Video and its Application. Communications in Computer and Information Science", 2011.
[3] Antonio Chella, Haris Dindo, Ignazio Infantino, "People Tracking and Posture Recognition for Human-Robot Interaction. Vision Based Human-Robot Interaction", March 2006.
[4] Bernard Boulay, "Human Posture Recognition for Behaviour ,Universite Nice Sophia Antipolis", 2007. English.
[5] Masaki Shiraishi and Ryuya Uda,"Detection of Suspicious Person with Kinect by action coordinate"2019
[6] Horiuchi, Y., Shinoda, H., Makino, Y., "Computational Foresight: Forecasting Human Body Motion in Real-time for Reducing Delays in Interactive System", ISS (2017).

[7] Dehbandi, B., Barachant, A., Harary, D., Long, J.D., Tsagaris,"Using data from the microsoft kinect 2 to quantify upper limb behavior: a feasibility study", IEEE J. Biomed. Health Inform. 21(5), 1386–1392 (2017)
[8] Yun, J., Lee, S.S.,"Human movement detection and idengication using pyroelectric infrared sensors" ,(2014)