Anomalous Motion Identification for Bank Surveillance

Shefali Shahane[1] Shreya Dajgude[2] Shreya Sisodia[3] Aditi Darade[4] Mr. Rahul Chakre[5]
[1,2,3,4]Student [5]Assistant Professor
[1,23,4,5]Department of Computer Engineering
[1,2,3,4,5]GES's R. H. Sapat College of Engineering, Management Studies and Research, Nashik, India.

Abstract— Posture recognition is one of the most exciting areas of computer vision, due to the wide range of applications in various fields. The problems of the previous system, a simple camera, can be resolved with the help of a 3-d camera. In this study, methods for the use of the structure of the information to be provided by the Kinect 3-d camera for the purpose of recognition for the successful real-time intelligent monitoring and control. To achieve the orientation of the recognition, the system can make use of the Kinect to track the bones, joints, and the position. Through the analysis of the information contained on the site, the system will detect improper conduct.

Keywords: Kinect, Posture Recognition, Abnormal Behaviour, Skeleton Discovery

I. INTRODUCTION

Bank-level security is a very important aspect of safety and security. With the growing crime rate in the cities is a bitter reality. A lot of people underestimate the need for the rest of the bank's security procedures. A robbery or burglary, it can often lead to disastrous consequences for the banking system. Modern security systems are highly dependent upon visual cues, such as facial features, which can be easily disguised. Criminals are disguising themselves as sunglasses, hats, and whether the use of more sophisticated techniques, in order for people to be fooled by this appearance of the bound systems. In addition, these systems fail in low-light conditions.

That is why there is a need for a reliable, intelligent systems, which warn you of the banks in real-time, regardless of the lighting conditions of the place or the dress of the play. So, the course offers real-time bank security system using the Microsoft Kinect, underground to the camera, check that goods are in the city, and warnings to banks via push notifications or the bank's records of the alarms on their mobile phones. Because of this, the method allows the use of wireframe of stolen information provided by the Kinect depth camera, which occurs by using infrared rays), in the place of the appearance, so it can withstand the conditions and the clothes of the criminals (robbers), to make it suitable for home security applications.

The approach to the provision of real-time banking, is as follows:

1) Use of the Kinect camera to detect human poses in the structure.
2) Classification of postures as abnormal or normal using logistic regression.
3) If the situation is suspicious, the bank has the authority to an alarm or a push notification on their phones.

All the calculations, which include the logical portion of the machine-learning algorithm that can be performed on the device itself. This allows you to avoid the unnecessary cost for the communication between the server and the remote device. It is an external device that communicates with the server when it detects suspicious activity.

II. LITERATURE SURVEY

Antonio Chella, Haris Dindo, Ignazio Infantino [1] proposed a method for the simultaneous tracking of people in the real world, as well as the detection of posture deficiencies in the environment, in the context of human-computer interaction. As soon as you have a tracking algorithm that tracks a person's system to assess their attitudes as well.

Bernard Boulay [2] is a proposed for a general operating practice for the recognition of a person's position in real-time, with the help of a static camera. 2D methods are to represent the silhouette of the observed person is to be provided by real-time data processing.

Chhavi Dhiman, Dinesh Kumar Vishwakarma [3] proposed a system that provides for the recognition of human activity and disturbance, in the film, as it relates to the context of applications, such as the fall of the investigation, internal security, surveillance, crowd analysis using RGB, depth and skeleton have to proof.

Shalini Nehra, Jagdish Lal Raheja [4] proposed a system that works in real-time, and has the ability to detect motion, drop, sessile drop, accurately, and safely, without taking into account any false actions.

G. Hernandez-Penaloza, A. Belmonte-Hern ´andez, M. Quintana, and F. Al- ´ varez. [5] proposed for a system consisting of a set of algorithms used for data collection and processing for the detection of abnormal events in the scenarios.

Shuoyan LIO, Hao XUE, Chunjie XU, Kai FANG [6] presented an approach that measures the degree of similarity between the rules need to change in order for the motion of the incoming video data in order to determine whether the unusual-of action.

D. Martinez [7] presented the project which was the development of a machine translation of sign language so that the computer generates a word, a sign, acting on behalf of deaf and the user is in front of the camera.

Paul D. Rosero *et al.* [8] presented an intelligent electronic glove system able to detect numbers of sign language in order to automate the process of communication between a deaf-mute person and others.

Akanksha Toshniwal, Kavi Mahesh, R. Jayashree [9] presented an idea that it provides a clear overview of the various tracking (ADT) technique. The real data is not available, as one would expect, therefore, choosing the right advertising and the algorithm depends on a number of factors, including the input of data, the type of violation, the output of the data and the domain knowledge. Hybrid method [10] which is the combination of supervised and unsupervised is also efficient for doing the classification and feature extraction respectively. Here, SVM is used as the supervised

component and for the feature extraction task ICA is for the multiclass data set.

Aditya Parab, Abhishek Nikam, Prajwal Mogaveera, Ashwini Save [11] presented an approach that lies in the fact that the cameras are to be used at the ATM, cages, however, the control of the law enforcement agencies cannot keep up with them. Thus, in this system, the incorrect behavior is detected with the help of the CNN and the lstm center is under video surveillance.

III. CONCLUSION FROM LITERATURE SURVEY

The above-mentioned papers, and for the recognition of a person's attributes can be carried out with the aid of a color camera. In this work, to identify the attributes of a person, it is necessary to determine which of the areas of interest in the image. The main drawback of the above work is that it is sensitive to a change in the dress and the lighting conditions. In contrast to these works, the work presented in this article is aimed at revealing the human skeleton (with the help of an in-depth camera). There are many works that have been proposed for the orientation of the recognition. However, the vast majority of this work, is making use of a scheme of any information obtained from the RGB camera. Therefore, it will react in the same way, day and night, regardless of the performance or the outdoors, no matter of the intensity of its illumination, which makes it particularly suitable for the monitoring of banking or ATM cash withdrawals.

IV. PROPOSED METHODOLOGY

The main modules of the proposed system is shown in the below figure 1. The proposed model consists of three modules: data collection, data processing and feature extraction, and the posture recognition.

The method that will be used for the determination of the posture (normal or abnormal), is a Logistic regression model. Because the system is trying to predict the probability of a false attitude to the subject of information on the frame, the regression analysis is a good approach to this problem. In addition, given that the dependent variable is dichotomous (which is normal behavior and conduct disorders), a logistic regression model, in this case, it will be an effective regression model. The whole process starts with the collection of data, to posture recognition, will be described below.

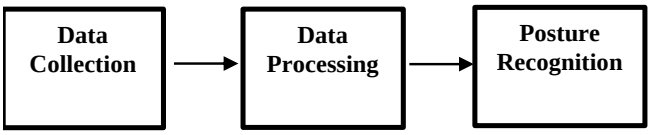

Fig. 1: Main Modules of the Proposed System

In the collection of the data, the Kinect will register a variety of information, such as the color, the thickness and the structure. The processing of personal data is aimed at finding the three-dimensional locations of the body joints with the help of the skeleton of the detection algorithms. An additional feature is a user-tracking, in which the algorithm makes use of, and processes within, the depth of the image to determine the position of each member of the body of the camera's coverage area. The matching attributes are defined on the basis of the angles made by the lines at some joints. In addition, the system normalizes the data, if necessary.

The Kinect can also be used to change the color of the stream, like a traditional camera, in addition to all the information about the subsurface and infrastructure. In addition to the depth and skeleton information, the Kinect also provides details on the color similar to that of traditional cameras. However, in this project, the focus is on exploring the possibility of the use of the skeletal structure of the information to the recognition of human posture. Human posture recognition is focused on learning the elements and are presented as a set of training data, and the classification of the test, of the pre-defined classes abnormal and normal.

A. Data Collection

The Kinect sensor is a device that reacts to movement. The name is a combination of the words "kinetic" and "type". It was originally created as a Natural User Interface (NUI) for the Microsoft Xbox 360 console to create a new, uncontrolled experience. It allows the user to communicate with and control software on the Xbox 360, a sign of recognition.

The system allows for the use of the Kinect, because it is highly effective in providing detailed information about the objects in comparison with other similar devices. The Kinect is composed of a variety of sensors. The Kinect features an RGB camera, which captures the twelve images with a resolution of up to 1280 x 960 pixels per second. It also features an INFRARED sensor that can capture depth images. This is reflected in several points, so that the final cell to the right of the IMAGE a depth camera, and for the calculation of the 3D environment. The unit is installed at an angle drive for the vertical angle. The Kinect is able to detect, for 2 users at the same time, and the computation of their skeleton in 3-D, 2-D and joints, which is the connection point between the body, including the legs, knees, hips, shoulders, elbows, hands, head, *etc.*

Fig. 2: Kinect Camera

The Software on the Xbox, the processes, the depth of the image to find the people and to find out the position of their bodies to be a part of it. This process is known as skeletonization, because it will determine the location of the user, the skeleton, the joints and the bones that connect them, on the basis of the data contained in the depth of the image.

With the proper handling of your library, you can access the user's location data, which gives information about the various joints of the skeletal system. The general information that is collected in the frame. For each frame, the position of the 2-D points is assessed and collected. Each connection has two of the most important parts of the information. First, the system will have the location of

contacts in the x, y, and z-axes. Finally, you can check the status of your joints. If Kinect will be able to keep track of the joint, the status is set to "tracked". If the connection cannot be completed, the system will try to get it on, the arthritis, the position of the joints. If it is successful, then the status of the connection is to be set to "inferred". Otherwise, the connection status is set to "did not track".

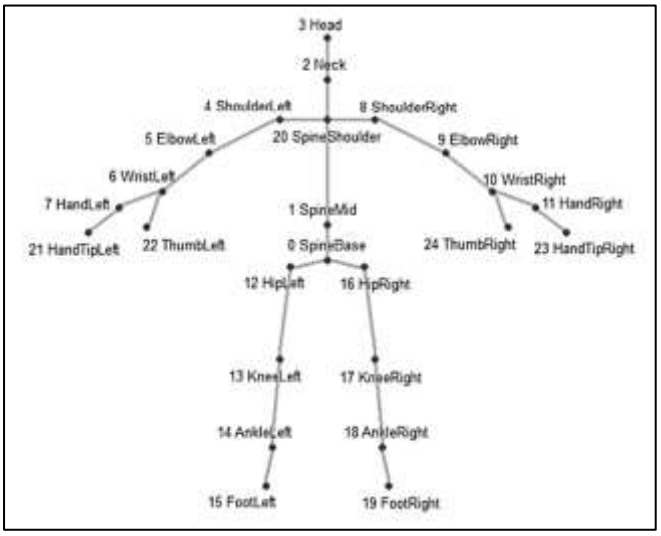

Fig. 3: Human Skeleton Tracking

B. Data Processing and Feature Extraction

System can also make use of the Kinect tracked skeleton is in general position. Since it is the common vector 3 coordinates of the skeleton and is made up of 20 joints, a feature vector is 60 size.

In addition to the features described above, there is another function that may be obtained from the calculation of the connection angles. When working with an attitude, and there are ten of joints, namely, on the Trunk, Neck, Head, Shoulder, Left Shoulder, Left Wrist, Arm, Knee, Wrist, Left Hip, and Right Hip, which is the most important joints in these poses. Of these compounds, different types of angles can be calculated. The system is then given to the next subject of a variety of item that you want, the recording functions of the angle between the vector connecting the left arm to the left, the elbow and the vector connecting the left elbow towards the arm, the angle between the vector connecting your left elbow to the left arm, the shoulder, the chest, the angle between the vector connecting the hand to the right elbow and right elbow of the right arm, and the angle between the vector connecting the right elbow right shoulder, and right shoulder, and on to the plane. The calculation of the subject's pose is based on the basic philosophy is that the orientation of the subject's body is the most typical value of the subject, the action, and because of this, it can be used as a reference.

The proposed system has used the KNN (K-Nearest Neighbour) algorithm because it takes the similarities between new cases / data and existing cases and puts the new case in exactly the same category as the existing categories. This algorithm is used to store all available data and to classify new data points appropriately. This means that where new data comes from it can be easily categorized into a functional phase using the KNN algorithm. The KNN algorithm in the training phase simply returns the data when it receives new data, and then separates that data into the exact same category as the new data.

C. Posture Recognition

The posture recognition can be achieved with the use of Logistic Regression, a machine learning technique, that is, for a given input signal, for predicts the class, and gives the probability associated with the prediction. These probabilities are extremely useful, as they will have a certain degree of confidence in the forecast.

Logistic regression is the appropriate regression analysis to conduct when the dependent variable is a binary one. Logistic regression is a type of analysis and the forecast. A logistic regression model was used to describe data and to explain the relationship between a binary dependent variable and one or more independent variables.

The Hypothesis function for Logistic regression is:

$$h\theta(x) = \frac{1}{1+e^{-z}} \quad (1)$$

Where z is θ^T X and $h\theta(x)$ is called sigmoid function.

We see that hθ (x) → 1 as z → ∞ and hθ (x) → 0 as z → −∞, and that hθ (x) is tied between (0,1). In Logistic Regression, the function below is used as a cost function:

$$cost(h\theta(x), y) = \begin{Bmatrix} -\log(h\theta(x)), if\ (y=1) \\ -\log(1-h\theta(x)), if\ (y=0) \end{Bmatrix} \quad (2)$$

Which basically means that if hθ (x) is closer to 1, the cost will be less, and if hθ (x) is closer to 0, the cost will be greater. that is, if hθ (x) is closer to 0, the cost will be less, and if hθ (x) is closer to 1, the cost will be greater. Combining these two components, it can get the following cost job:

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log(h_\theta(x^{(i)}) + (1-y^{(i)}) \log(1-h_\theta(x^{(i)}))] \quad (3)$$

The target is to minimize this cost function, so we can use the gradient descent method. For gradient descent, it just iteratively trims the θ vector:

$$\theta_j = \theta_j + \alpha(y^{(i)} - (h_\theta(x^{(i)}))x_j^{(j)} \quad (4)$$

Where, α is the learning rate. Here we chose α = 0.1. θj refers to the j th element in θ vector, i is the current training sample in use.

D. Classifier/ Gesture Recognition

Given a sequence of frames, the classifier will provide a map to the nearest order of the square of the standard dictionaries. Dynamic-Time-Warping (DTW) has been introduced in the ' 60s, and an algorithm for measuring the similarity between two sequences which may vary in time or speed. With the DTW, your computer will be able to find an optimal match between two sequences, with a number of limitations. The sequences are "warped" straight-line basis over the period of time dimension to determine a measure for the similarity. It is independent of the number of non-linear changes in the price. In this project, DTW algorithm have been used for gesture recognition purposes.

E. System Implementation

The features that were used in this project is the same made by the several bones in this position. These features make

itself felt by a pathological beliefs or intentions. The system consisted of the training data (a function value) of the algorithm in a text file.

The classifier was trained on the 360 points of data that are generated by the user. The calculated weights, and is used for real-time classification algorithms.

F. Normal and Abnormal Postures

For demonstration purposes, the system is chosen for the following positions, which shall be deemed invalid, and will use the same data as the training data set;

- Hands up gesture
- Fiddling with the camera position
- A person holding a gun

An alarm will occur if the current position is that of a person coincides with one of the above functions. The system will look for a position that does not correspond to the categories of the wrong posture, as if in a normal position. In this case, the alarming condition is not the case.

V. EXPERIMENT AND RESULT ANALYSIS

After the training of the machine and with the help of the training data, the algorithm was able to detect the abnormal and normal behavior as expected. For example, if a person is trying to destroy the bank, or attempting to cheat, or is trying to capture a camera, the system is designed to react to an alert. If there is any kind of activity, as shown in the figure 6, the alarm didn't go off.

Fig. 4: Bank Administration Login Page

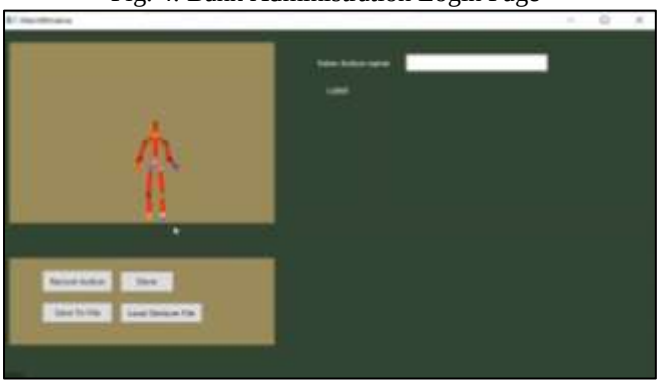

Fig. 5: Skeleton Detection of Person

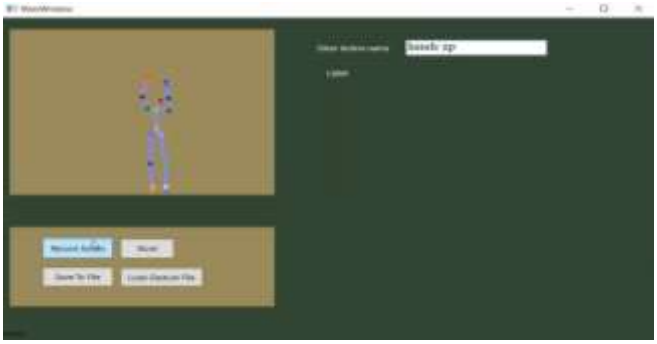

Fig. 6: Hands-Up Gesture Detected

Fig. 7: A Person Holding A Gun Gesture

VI. CONCLUSION

In this project, the system has proved to be, that the posture recognition can be used for the detection of improper conduct by a person in a bank. This will be achieved through the use of the skeletal muscles of the data that can be extracted from the depth image, which was created with the help of 3D cameras such as the Kinect. The system used the processing language to write the code.

A machine-learning algorithm, Logistic Regression, and the DTW algorithm are used to compute the probability that the position of the person being monitored is abnormal. This has been achieved by the calculation of the weights of the training data set. And, with the help of the angles between the various bones, including the features that we collect for the purpose of calculating the weights for the algorithm. It also makes use of the gradient descent method for the computation of the optimal weights. Therefore, the system of the bank is able to teach the wrong thing, with the help of an algorithm and the data.

VII. FUTURE WORK

Future activities will include the deployment of the system to more security applications and locations, such as universities, shopping malls, theaters, etc. as well as online learning.

How to enable a new gesture, and the impact of detection of the method will do its best to ensure that the system will stop automatically.

The most recent advance in the future will be related to the computation of the cost. While the current system works in real-time mode, the computational cost can be improved by reducing the number of dimensions is the most important descriptors. A Principal component analysis, can be a good option.

REFERENCES

[1] Antonio Chella, Haris Dindo, Ignazio Infantino, "People Tracking and Posture Recognition for Human-Robot Interaction", Vision Based Human-Robot Interaction, Palermo, Italy, March 2006.
[2] Bernard Boulay, "Human Posture Recognition for Behaviour", [cs. OH]. University of Nice Sophia Antipolis, 2007.
[3] Chhavi Dhiman, Dinesh Kumar Vishwakarma, "A review of state-of-the-art techniques for abnormal human activity recognition", 2019.
[4] Shalini Nehra, Jagdish Lal Raheja, "Unobtrusive and Non-Invasive Human Activity Recognition using Kinect Sensor", IEEE, 2020.
[5] G. Hernandez-Penaloza, A. Belmonte-Hern ´ andez, M. Quintana, and F. Al- ´ varez, "A multi-sensor fusion scheme to increase life autonomy of elderly people with cognitive problems", IEEE Access, vol. 6, 2018.
[6] Shuoyan LIO, Hao XUE, Chunjie XU, Kai FANG, "Abnormal behavior detection based on the motion-changed rules", IEEE, 2020.
[7] D. Martinez, MSc Thesis, "Sign Language Translator using Microsoft Kinect XBOX 360TM", Department of Electrical Engineering and Computer Science, Computer Vision Lab, University of Tennessee.
[8] Paul D. Rosero-Montalvo, Pamela Godoy Trujillo, Edison Flores-Bosmediano, Jorge Carrascal Garc, Santiago Otero-Potosi, Henry Benitez-Pereira and Diego H. PeluffoOrd, "Sign Language Recognition Based on Intelligent Glove Using Machine Learning Techniques", IEEE, 2018.
[9] Akanksha Toshniwal, Kavi Mahesh, R. Jayashree, "Overview of Anomaly Detection techniques in Machine Learning", 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020.
[10] Rahul R. Chakre, Dr. Radhakrishna Naik, "Performance Analysis of Hybrid (supervised and unsupervised) method for multiclass data set" IOSR Journal of Computer Engineering , Volume 16, Issue 4, Ver. III (Jul – Aug. 2014), PP 93-99.
[11] Aditya Parab, Abhishek Nikam, Prajwal Mogaveera, Ashwini Save, "A New Approach to Detect Anomalous Behaviour in ATMs", 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020.

