International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 03 | Mar-2018 www.irjet.net p-ISSN: 2395-0072

i-SURVEILLANCE
CRIME MONITORING AND PREVENTION USING NEURAL NETWORKS

Malavika Nair[1], Mathew Gillroy[2], Neethu Jose[3], Jasmy Davies[4]

[1,2,3] B. Tech student, Department of Computer Science and Engineering, SCET, Kerala, India
[4]Assistant Professor, Department of Computer Science and Engineering, SCET, Kerala, India

---

Abstract - *Closed Circuit Television Systems (CCTV) are becoming more and more popular and are being deployed in many housing estates, offices, and also in most public spaces. CCTV monitoring systems have been implemented in many American and European cities. As the number of camera views a single CCTV operator can handle is limited by human factors, such monitoring systems makes for an enormous load for the CCTV operators. Here, in this project named i–Surveillance or Intelligent Surveillance System, we propose algorithms that are able to alert the human operator when: (1) Presence of a dangerous act, after setting zones of interest and danger zones within those zones of interest,, the danger is detected when an object trespasses the danger zone, which can reduce the number of accidents in factories and suicides in certain places, (2) An abnormal behavior of a person such as handling some weapons or act of abuse or molestation is detected, which might be a potential threat. In this project, in order to allow for a real life application of the system, we focus on limiting the number of false alarms.*

Key Words: Behavioral Analysis, Security, Machine Learning, Image Processing, Neural Networks.

1. INTRODUCTION

In recent years, we've seen that there has been a marked and sustained growth in the use of Closed Circuit Television (CCTV) surveillance cameras in order to prevent crimes in public places in the USA and other Western nations. Amidst the associated public expenditure and the expansion, as well as concerns about their social costs and efficacy, there is an increasing need for an evidence-based approach to inform CCTV practices and policies. Cases of harassment in work places are also becoming very serious. With the ever growing installation of advanced CCTV infrastructure, almost entire cities can now be monitored, through the major purpose served by the same is purely evidential. It would only be natural to expect an alert or warning system for ongoing (or about to happen) mishaps and crimes, where timely action can be the difference between life and death. Such scenarios are expected to be monitored and identified by personnel viewing live footage. But as the number of CCTVs per unit are keeping rising, this approach is becoming increasingly impractical. Thus what we require is a surveillance unit capable of thriving in these situations with negligible human input.

We shall define a "situation of interest" or a "critical situation" as any sensitive situation that could possibly lead to the afore-mentioned predicaments. Consider the idea of a smart surveillance which would be triggered 'active' only when the statistical chances of the situation being of "interest" are high. In office environments each room being independently monitored can respond to such a trigger by enabling the recording. At all other times the CCTV is not recording which helps in maintaining privacy and, if required, confidentiality of work. Cutting edge research and development leads to cut-throat competition and one can easily realize why traditional surveillance techniques might prove risky. The video feed would be recorded only under a "situation of interest" in case it needs to be documented for a legal investigation. In the response to the above trigger could be an alert to be issued to the appropriate authorities along with certain alarms which could help in preventing the situation from escalating further. So, this validates the requirement for a system which could provide smart surveillance, while ensuring privacy and confidentiality.

The surveillance camera activated for recording only when there is a situation of interest. The camera is inactive or it is not recording the video when there is no human presence. The human presence is checked using motion detection algorithm. When a crime is about to be committed, then the human is notified and an alarm system connected to the main system will be activated.

2. EXISTING SYSTEM

The existing system of CCTV monitoring is done by humans and also the automated smart CCTV monitoring systems are not fully capable of making decisions and triggering actions. Surveillance in dynamic scenes attempts to recognize, detect and track certain objects from image sequences, and more generally to understand object behaviors. The objective is to develop an intelligent visual surveillance system for replacing the traditional passive video surveillance system, since the traditional passive video surveillance system proves ineffective as the number of cameras exceeds the capability of human operators to observe them. In addition to placing cameras in the place of human eyes, the goal of visual surveillance is to accomplish, as automatically as possible, the surveillance task.

Visual surveillance in dynamic scenes has a wide range of potential applications, such as a security guard for communities and important buildings, traffic surveillance in cities expressways, detection of military target objects, etc. We focus in this paper on applications involving the

surveillance of people, vehicles, as they are typical of surveillance applications and include the full range of surveillance methods. Surveillance applications which involve people or vehicles include the following.

a) Access control in special areas : Only those people with a special identity are allowed to enter in some security-sensitive locations such as important governmental offices and military bases. A biometric feature database including legal visiting members is built beforehand using biometric techniques. When someone is about to enter to this security-sensitive location, the system automatically obtains the visitor's features, such as their facial appearance, height and walking gait from those images taken in real time situation. Based on this, the system is able to decide whether the visitor can be cleared for entry.

b) Person-specific identification in certain scenes : The police can catch suspects with the help of personal identification at a distance by a smart surveillance system. The cops may build a database containing the biometric feature of suspects, and they may place visual surveillance systems at those places where the suspects are usually seen, such as subway stations, casinos, etc. The systems are able to automatically recognize whether or not the people in view are suspects. If yes, then the alarms will be given immediately. Even though such systems with face recognition have already been used at public sites, their reliability is too low for police requirements.

c) Crowd flux statistics and congestion analysis : The flux of people at important public areas such as stores, can be automatically computed by the surveillance systems, using techniques for human detection. It can then provide congestion analysis to assist in the management of the people. Similarly, expressways and junctions of the roads can be monitored through visual surveillance systems, and further analyze the status of road congestion and traffic.

d) Anomaly detection and alarming : At certain times, it is necessary to analyze the behaviors or characteristics of vehicles and people and to determine whether these characteristics are normal or abnormal. For example, abnormal behaviors indicative of theft, can be analyzed using the visual surveillance systems, which can be placed in supermarkets and parking lots. Normally, there are many ways of giving an alarm. One of the ways is to make a recorded announcement automatically, whenever any abnormal behavior is detected. Another method is to contact the police automatically.

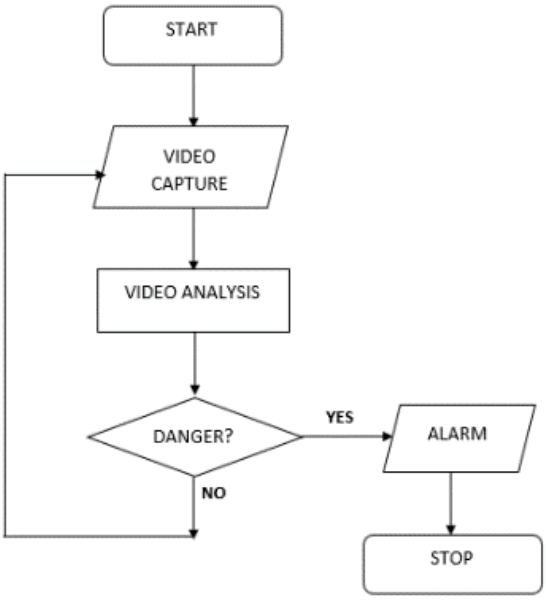

Fig-1 : Proposed System

The image of a person being captured, tying him to a particular place, or the vehicle number plate being registered on footage are probably the most obvious and perhaps widespread cause of loss of privacy. As the numbers of CCTV cameras are increasing, the likelihood of the above happening is sure to increase.

Drawbacks

- The system is used only on the army borders, but not into public places such as metro stations, big malls, etc.
- It is a complex mechanism. Individual safety level is not considered.
- Video is captured continuously, which can lead to privacy issues.
- The system used also doesn't have an efficient alarm system that triggers automatically on an event of abuse, harassment or bullying.

Lack of privacy as the image of victim is not blurred.

3. SYSTEM DESIGN

In this paper, we use classification method for analyzing the video and to detect anomalies. A Graphical User Interface (GUI) is implemented for the automatic video surveillance. Here, when the login screen appears, the user can login to the application with the registered username and password. If the user credentials are found to be correct, the application will show the options BROWSE and VIEW. Here, the user can either provide a live webcam feed or choose a video by browsing the folders, and providing the correct path.

The fed video is then converted into frames, and it is analyzed for checking abnormalities. We first check whether there are some human actions present in the frames. This checking is done with the help of the trained models. These trained models are compared with the actions present in the frame, in order to detect human presence in that frame. For this purpose of training, we require a dataset containing a set of images, and the dataset used here is HMDB51 (Human Motion Database 51). Next, we annotate these frames using a tool labelImg. Since there are quite a large number of videos, all these videos can be annotated only if it is converted into frames. This is done using a program, and the code is written in OpenCV using Python. Then, after saving all the frames as xml files, the train and test directories are separated. The usual procedure for separating train and test directories is as follows :

- 90% of the total xmls comprises the train directory
- 10% of the total xmls comprises the test directory

The training data is used for creating the model, and while the training data is being trained, each model from the test directory is taken and tested with each model of the test directory. From this we can get the percentage loss in accuracy. At a particular point when the loss percentage graph becomes steady, the training may be stopped.

Finally, the abnormality and its accuracy will be determined and shown as the output.

4. PROPOSED SYSTEM

The proposed system deals with the current trend of security system i.e. inclusion of privacy concern, cost effectiveness, efficient alarm system. Conditional video recording is the major inclusion. At all times the camera is on but not recording. It would keep on processing the frames it receives and would not capture it unless 'triggered'.

Activation of subroutines is done only when multiple personnel are detected. These subroutines involve image processing in conjunction to detect critical instances. Once enough baseline deviation is flagged, the situation is marked as critical and recording is activated. The ideology is as follows:

1) In case the person is not alone, we activate the camera but any documentations, notes, scribbling on a board or screens (which are to be considered confidential at all times) shall be blanked out/blurred to avoid any leaks (any simple irreversible technique could be used).

2) Once it has been established that the person is not alone, we refine our analysis to identifying the number of people, their gender and other characteristics. We make use of human detection, gesture recognition and motion detection, and emotion detection.

3) Voice, just as video, is not to be recorded unless we discover a situation of interest.

In public places privacy is no longer a pressing matter. So the subroutines for tracking and human detection are always in operation in a conditional focus mode. By conditional focus we mean that we run only a preliminary test for critical situations. Once that is flagged, do we bring in our heavier classifiers involving both Video and Audio processing. This ensures a smoother real time operation.

The approach, therefore, is to wait for a trigger to a critical situation and then to take the required action. It is worth mentioning that the classification boundary in this case should be a bit less stringent than the one above, the reason being that this system is not entirely self-sufficient and hence relies on human input. Thus, it would only make sense to slack the constraints so that even if false alarms are increased, the probability of a crime going unnoticed is considerably suppressed. The preliminary indicators for instance can be: rapid and jerky movements (in addition to "excited" audio signature) between individuals closer than a particular safe proximity. Once triggered recording initiates plus an alert signal is issued either directly to the authorities or a decision making moderator.

There is not a lot of research on quantification of a strong theoretical framework for basing emotion detection through Image processing algorithms. Associative neural networks like CALM incorporate interpretation of human emotions through gestural cues. It is possible with higher resolution cameras to include facial expression analysis systems to categorize the physical expression of emotions, such as Facial Action Coding System (FACS).

5. METHODOLOGY

The objective of our project is to design and implement a system to prevent crime and suicides by tracking human behavior through the analysis of video captured using CCTV. The proposed system has multiple functionalities including preventing suicides, abnormal behavior and even attempts of abuse.

The various modules of our project are as follows :

5.1 Video Capture CCTV

Closed Circuit Television (CCTV) is used to monitor the scene. Video is not recorded until it is necessary i.e. video is recorded only when the person is not alone and when there are sudden movements or actions. Closed Circuit Television, also known as video surveillance refers to the use of a CCTV or a video camera for transmitting a signal to a certain specified place on a limited set of monitors. CCTV schemes in public settings had small and non-statistically significant effects on crime: 7% reduction in city and town centers and 23% reduction in public transport settings.

5.2 Frame Capturing

Videos are a continuous flow of frames or still images. Each frame is captured and processed to check for motion in the

scene. Motion in the scene means there's some human activity in the scene. Human activity is confirmed by checking for faces or body patterns. A continuous video has thousands of frames. In order to save storage space, only the last 20 frames are kept for reference. Old frames get deleted when new ones arrive.

5.3 Motion Detection

Motion in the scene is detected by analyzing the frames captured. This can be achieved by a technique called background subtraction wherein still objects in the scenes are eliminated by analyzing the position of each object in the stored frames. If the position of object is changed, the object is flagged to be a moving object. To ensure that it is a human, face patterns or body patterns are matched with saved shapes.

5.4 Frame Analysis

For video analysis we base our processing on a standard multiple moving objects tracking algorithm. The approach is to use a background subtraction based on Gaussian mixture models to segment out the moving objects. After the foreground is de-noised, blob detection is run and the connected pixels are bounded in a box which is to be tracked. The motion tracking is achieved by a Kalman filter. After observing the object for a set number of frames, the Kalman Filter then estimates the position of the object in subsequent frames (all the while observing the actual motion that happens and recording the errors for use in future pre dictions). Then behavior of moving character is analyzed using behavioral analysis algorithm to check for any act of abuse, bullying or harassment.

5.5 TensorFlow

TensorFlow is basically an open source software library that is used for numerical computation using data flow graphs. Each node in the data flow graph represents certain mathematical operations, and the edges of the graph represent the multidimensional data arrays, commonly known as tensors, communicated between them.

TensorFlow Object Detection API : One of the core challenges in computer vision is to create accurate machine learning models, which are capable of identifying and localizing multiple objects in a single image. It is an open source framework built on top of TensorFlow that makes it easy to deploy, train and construct object detection models.

5.5 HMDB51

HMDB refers to Human Motion Database. There are a total of 6847 video clips in this dataset, and these clips are divided into certain action categories, 51 in number. Each of these 51 action categories contain a minimum of 101 video clips, which is extracted from a range of sources. These actions can be categorized as follows :

1. General facial actions : laugh, talk, smile, chew.
2. Facial actions with object manipulation : drink, smoke, eat.
3. General body movements : clapping hands, pull up, climb stairs, fall on the floor, jump, push up, run, sit down, dive, sit up, somersault, stand up, backhand flip, turn, walk, wave.
4. Body movements with object interaction : brush hair, catch, hit something, kick ball, pick, push something, ride bike, golf, ride horse, shoot ball, shoot gun, draw sword, sword exercise, throw.
5. Body movements for human interaction : hug, kiss, fencing, punch, kick someone, shake hands, sword fight.

There is no pre-processing done on the video clips.
A three - level grading of video quality is applied to evaluate the huge set of clips. Only those video samples are rated "good" which have the quality that you can identify the single fingers during the motion. Those samples which do not meet this requirement are rated either "bad" or "medium" if some of the limbs or body parts vanish while the action is being executed.

6. RESULT AND DISCUSSION

In this paper, the result is the final design of the project on topic "i - Surveillance : Crime Monitoring and Prevention Using Neural Networks".

Since there is a GUI implemented, the user is able to login and then choose from live webcam feed or browse the video from some locations. Here, algorithms have been designed that can alert the human operator when the presence of a dangerous act, or an abnormal behavior of a person is detected. This can assure safety in public places as well as in other locations.

The major advantage of the project includes efficiency, fast to access and uniqueness. The behavioral analysis algorithm also makes it easier for the CCTV operators to monitor the CCTV systems and prevent crimes.

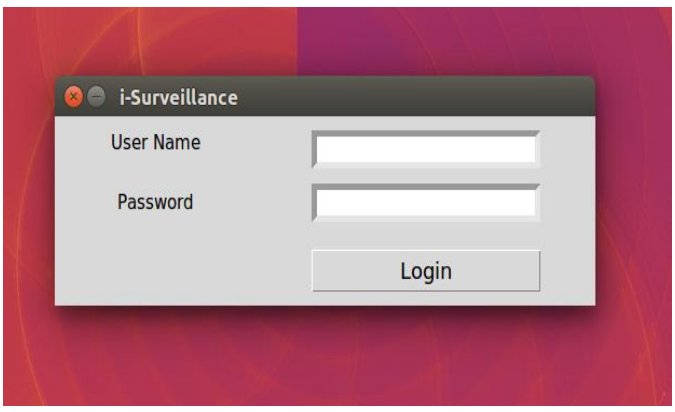

Fig-2 : Login Screen

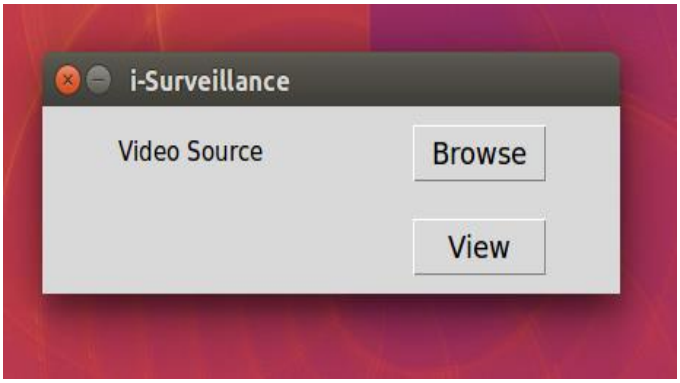

Fig-3 : Option for Live webcam or Browse from folder

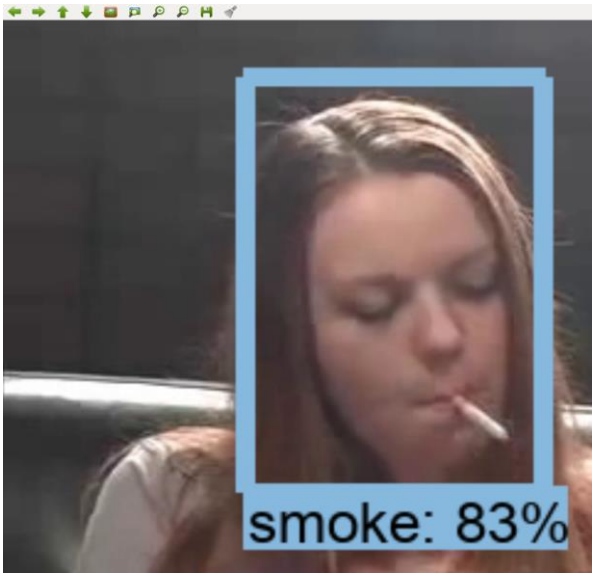

Fig-4 : Detection of smoking

Fig-5 : Detection of hitting with something

7. CONCLUSION

Surveillance by using CCTV systems has reached to its best level. Also, sending information or data through data networks to servers is common these days, but coupling these two surveillance and data transmission processes is a very challenging work. CCTV surveillance systems are mostly implemented and managed by governments. So, using CCTV systems information has many security problems and it is very difficult to handle.

In this project named i-SURVEILLANCE, we proposed algorithms that are able to alert the human operator when an abandoned luggage, presence of a dangerous act, or an abnormal behaviour of a person is detected. We focused on limiting the number of false alarms in order to allow for a real life application of the system.

This proposed system can be used at low scale in first phase where security issues are less or easy to handle. In the future, we will enhance the proposed system tracking algorithm by using the Enhanced Filter model that will consider the multiple sensing data of mobile user with network connected CCTV environment.

ACKNOWLEDGEMENT

We would like to express our immense gratitude and profound thanks to all those who helped us to make this project a great success. At the very outset we express our thanks to the Almighty God for all the blessings endowed on us. We can hardly find words to express our deep appreciation for the help and warm encouragement that we have received from our project guide Mrs. Jasmy Davies, Assistant Professor, Sahrdaya College of Engineering & Technology and also our project coordinators Mr. Willson Joseph C. and Mrs. Anusree K, Assistant Professors, Sahrdaya College of Engineering & Technology. We would also extend our deep sense of gratitude to all other faculties for their help and advice throughout the advancement of this project. We are extremely thankful to our parents and friends for being with us as a constant source of inspiration.

REFERENCES

[1] Choi Woo Chul and Na Joon Yeop, “Relative Importance for Crime Prevention Technologies as Part of Smart City based on Spatial Information”, IEEE Journal, Smart Cities Symposium Prague, 13 July 2017

[2] Robin Singh Sidhu and Mrigank Sharad, “Smart Surveillance System for Detecting Interpersonal Crime”, presented in International Conference on Communication and Signal Processing, April 6-8, 2016, published in IEEE Journal, 24 November 2016

[3] Aniket Bhondave, Bhagyawant Biradar, Vishal Suryavanshi and Harshal Nakil, “Suspicious Object Detection Using Back-Tracing Technique”, International Journal of

Advanced Research in Computer and Communication Engineering Volume 5, Issue 1, January 2016

[4] Yuchae Jung and Yongik Yoon, "Behavior Tracking Model in Dynamic Situation using the risk ratio EM", IEEE Journal, 12 March 2015

[5] Swapnali Rayte, Rohini Bhamare, Kaustubh Barhate and Mahendra Sonawane, "Crime Monitoring and Controlling System by Mobile Device", International Journal on Recent and Innovation Trends in Computing and Communication, Volume 3, Issue 1, January 2015

[6] 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

[7] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "HMDB: A Large Video Database for Human Motion Recognition", ICCV, 2011.

BIOGRAPHIES

Malavika Nair is currently pursuing Bachelor of Technology in Computer Science & Engineering in Sahrdaya College of Engineering & Technology, Kodakara affiliated to Calicut University.

Mathew Gillroy is currently pursuing Bachelor of Technology in Computer Science & Engineering in Sahrdaya College of Engineering & Technology, Kodakara affiliated to Calicut University.

Neethu Jose is currently pursuing Bachelor of Technology in Computer Science & Engineering in Sahrdaya College of Engineering & Technology, Kodakara affiliated to Calicut University.

Jasmy Davies is currently working as the Assistant Professor in Computer Science & Engineering department in Sahrdaya College of Engineering & Technology, Kodakara affiliated to Calicut University.