Video Analysis for Weapon Detection and Alerting

First Author:Mr. SM.RAFI
Second Author:A.Lavanya
1.Associate professor,Dept of MCA,
Audisankara college of engineering and technology (AUTONOMOUS),Gudur,AP, India.
2.PG Scholor,Dept of MCA, Audisankara college of engineering and technology
(AUTONOMOUS), Gudur,AP, India.

ABSTRACT

Security and safety is a big concern for today's modern world. For a country to be economically strong, it must ensure a safe and secure environment for investors and tourists. Having said that, Closed Circuit Television (CCTV) cameras are being used for surveillance and to monitor activities i.e. robberies but these cameras still require human supervision and intervention. We need a system that can automatically detect these illegal activities. Despite state-of-the-art deep learning algorithms, fast processing hardware, and advanced CCTV cameras, weapon detection in real-time is still a serious challenge. Observing angle differences, occlusions by the carrier of the firearm and persons around it further enhances the difficulty of the challenge. This work focuses on providing a secure place using CCTV footage as a source to detect harmful weapons by applying the state of the art open-source deep learning algorithms. We have implemented binary classification assuming pistol class as the reference class and relevant confusion objects inclusion concept is introduced to reduce false positives and false negatives.

1. INTRODUCTION

The crime rate across the globe has increased mainly because of the frequent use of handheld weapons during violent activity. For a country to progress, the law-and- order situation must be in control. Whether we want to attract investors for investment or to generate revenue with the tourism industry, all these needs is a peaceful and safe environment. The crime ratio because of guns is very critical in numerous parts of the world. It includes mainly those countries in which it is legal to keep a firearm. The world is a global village now and what we speak or write has an impact on the people. Even if the news they heard is crafted having no truth but as it gets viral in a few hours because of the media and especially social media, the damage will be done. People now have more depression and have less control over their anger, and hate speeches can get those people to lose their minds. People can be brainwashed and psychological studies show that if a person has a weapon in this situation, he may lose his senses and commit a violent activity.

CCTV cameras play an important role to overcome this problem and are considered to be one of the most important requirements for the security aspect. [3]. CCTVs are installed in every public place today and are mainly used for providing safety, crime investigation, and other security measures for detection. CCTV footage is the most

important evidence in courts. After a crime is committed, law enforcement agencies arrive at the scene and take the recording of footage with them [4].

In previous years, though having surveillance cameras installed, to use them for security purposes was not an easy and dependable method. A human has to be there all the time to monitor screens. CCTV operator has to monitor 20- 25 screens for 10 hours. He has to look, observe, identify, and control the situation that can be harmful to the individuals and the property. As the number of screens increases, the concentration of the person decreases considerably to monitor each screen with time. It is impossible for the person monitoring the screens to keep the same level of attention all the time [7].

The solution to aforementioned problem is to install surveillance cameras with the ability to automatically detect weapons and raise alarm to alert the operators or security personals. However, there is not much work done on algorithms for weapon detection in surveillance cameras, and related studies are often considering concealed weapon detection (CWD), mostly using X-rays or millimeter waves images employing traditional machine learning techniques [8-12]. In the past few years, deep learning in particular convolutional neural network (CNN) has given groundbreaking results in object categorizing and detection. It has achieved finest results thus far in classical problems of image processing such as grouping, detection and localization. Instead of selecting features manually, CNN automatically learns features from given data.

2. LITERATURE SURVEY

The problem of detection and classification of objects in real- time started after major developments in the CCTV field, processing hardware, and deep learning models. Very little work has been done in this field before and most of the previous effort was related to concealed weapon detection (CWD).

Starting with concealed weapon detection (CWD), before its use in weapon detection, it was used for luggage control and other security purposes at airports and was based on imaging techniques like millimeter-wave and infrared imaging [13]. Sheen et al. suggested CWD method based on a three-dimensional millimeter (mm) wave imaging method, for detecting hidden weapons at airports and other safe locations in the body [14]. Z. Xue et al. suggested a CWD technique based on a fusion-based technique of multi-scale decomposition, which combines color visual picture with infrared (IR) picture integration [15]. R. Blum et al. suggested a CWD method based on the inclusion of visual picture and IR or mm wave picture using a multi-resolution mosaic technique to highlight the hidden weapon of the target picture [16].

E. M. Upadhyay et. al. suggested a CWD technique using image fusion. They used IR image and visual fusion to detect hidden weapons in a situation where the image of the scene was present over and under exposed area. Their methodology was to apply a homomorphic filter captured at distinct exposure conditions to visual and IR pictures [17]. Current techniques attain high precision by using various combinations of extractors and detectors, either by using easy intensity descriptors, boundary detection, and pattern matching [18] or by using more complicated techniques such as cascade classifiers with boosting.

CWD though had worked for some sort of cases but it had many limitations. These systems were based on metal detection; non-metallic guns cannot be detected. They were costly to use in many locations

because they need to be coupled with X-ray scanners and conveyor belts and responds to all metallic objects, so were not accurate. Economic cost and health risks limited the practical implementation of such methods. Furthermore, video-based firearm detection was a preventive measure for acoustic detection of gunshot and can be combined with it for implementation [19-20].

The idea of automated image processing for public security purposes in many fields has been well recognized and studied. CCTV was the ultimate need for this kind of work to progress. CCTV was first used back in 1946 in Germany and at that time, these cameras were installed to observe the launch of a rocket named V2 [21]. Although it had been used earlier, major improvements happened in the last two decades. With the advancement in CCTV technology, visual object recognition and detection for surveillance, control, and security were performed. In 1973, Charge-Coupled Device (CCD) was developed, which made the deployment of surveillance cameras possible by 1980 [22]. If we go a bit forward in time, a company named Axis Communication developed the first-ever network camera, which enabled the transformation of surveillance cameras from analog to digital [22]. This transformation of analog to digital video made it possible for everyone to apply image processing, machine learning, and computer vision techniques on videos recorded from surveillance cameras. In 2003, Royal Palm Middle School in Phoenix used facial recognition for the first time for tracking missing children.

3. PROPOSED WORK

Gun detection in real-time is a very challenging task. As our desired object has a small size so, detecting it in an image is also very challenging in presence of other objects, especially those objects that can be confused with it. Deep learning models faced several below mentioned challenges for detection and classification task:

- The first and main problem is the data through which CNN learn its features to be used later for classification and detection.
- No standard dataset was available for weapons.
- For real-time scenarios, making a novel dataset manually was a very long and time-consuming process.
- Labeling the desired database is not an easy task, as

all data needs to be labeled manually.

- Different detection algorithms were used, so a labeled dataset for one algorithm cannot be utilized for the other one.
- Every algorithm requires different labeling and pre-

processing operations for the same-labeled database.

- As for real-time implementation, detection systems require the exact location of the weapon so gun blocking or occlusion is also a problem that arises frequently and it could occur because of self, inter- object, or background blocking.

The main contributions of this work are: presentation of a first detailed and comprehensive work on weapon detection that can achieve detection in videos from real-time CCTV and works well even in low resolution and brightness because most of the work done earlier is on high definition training images but realtime scenario needs realtime training data as well for better results, finding of the most suitable and appropriate CNN based object detector for the application of weapon detection in real-time CCTV video streams, making of a new dataset because real-time detection also needs real-time training data so we

made a new database of 8327 images and preprocessed it using different OpenCV filters i.e. Equalized, Grayscale and clahe that helped in detecting images in low brightness and resolution, introducing the concept of related confusion classes to reduce false positives and negatives, training and testing of our novel database on the latest state of the deep learning based classification and detection models among them Yolov4 performed best in terms of both speed and accuracy and our selected trained model predict images at almost every orientation, angle, and view, achieving the highest mean average precision of 91.73% along with a F1-score of 91% on Yolov4.

A. *OBJECT RECOGNITION*

As the name suggests, it is the process of predicting the real class or category of an image to which it belongs by making probability high only for that particular class. CNN's are used to efficiently perform this process. Many state of the art Classification and Detection algorithms uses CNN as a backend to perform their tasks.

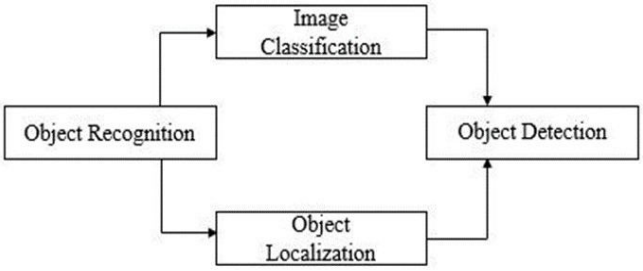

FIGURE 1. Object Recognition to detection Hierarchy

Fig. 1 depicts that classification and localization come under the category of recognition and combined classification and localization is performed to do object detection. Let us have a brief overview of the object classification, localization, and detection.

1) IMAGE CLASSIFICATION

The classification model takes an image and slide the kernel/filter over the whole image to get the feature maps. From the feature extracted, it then predicts the label based on the probability.

2) OBJECT LOCALIZATION

This method outputs the actual location of an object in an image by giving the associated height and width along with its coordinates.

3) OBJECT DETECTION

This task uses the properties of the aforementioned algorithms. The detection algorithm tells us the bounding box having x and y coordinates with associated width and height along with the class label. Non-max suppression is used to output the box with our desired threshold [60]. This process gives the following results altogether:

- Bounding Box
- Probability

In past object detection was very limited because of less data and low processing power of computers but with the passage of time the computing power of computers increased and world moved from CPU's to Graphic Processing Units (GPU). GPU's were firstly made for increasing the graphic quality of the systems and for gaming but later GPUs were used extensively for deep learning. In ImageNet, competitions started and contained about 1000 classes [61]. This was the evolution of machine learning and deep learning. In the beginning, the models were not very deep, means there were not many layers as they are now in an algorithm. Because of the aforementioned developments, in

4. CONCLUSION

For both monitoring and control purposes, this work has presented a novel automatic weapon detection system in real- time. This work will indeed help in improving the security, law and order situation for the betterment and safety of humanity, especially for the countries who had suffered a lot with these kind of violent activities. This will bring a positive impact on the economy by attracting investors and tourists, as security and safety are their primary needs. We have focused on detecting the weapon in live CCTV streams and at the same time reduced the false negatives and positives. To achieve high precision and recall we constructed a new training database for the real-time scenario, then trained, and evaluated it on the latest state-of-the-art deep learning models using two approaches, i.e. sliding window/classification and region proposal/object detection. Different algorithms were investigated to get good precision and recall.

Through a series of experiments, we concluded that object detection algorithms with ROI (Region of Interest) perform better than algorithms without ROI. We have tested many models but among all of them, the state-of-the-art Yolov4, trained on our new database, gave very few false positive and negative values, hence achieved the most successful results

5. REFERENCES

[1] “Christchurch mosque shootzings”, En.wikipedia.org, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Christchurch_mosque_shootings. [Accessed: - 10 Jul- 2019].

[2] Unodc.org. (2019). Global study on homicide. [online] Available at: https://www.unodc.org/unodc/en/data-and-analysis/global-study-on-homicide.html [Accessed 10 Jul. 2019].

[3] Deisman, Wade. "cctv: Literature review and bibliography." In Research and Evaluation Branch, Community, Contract and Aboriginal Policing Services Directorate. Ottawa: Royal Canadian Mounted. 2003.

[4] Ratcliffe, Jerry. Video surveillance of public places. Washington, DC: US Department of Justice, Office of Community Oriented Policing Services, 2006.

[5] Grega, Michał, Andrzej Matiolański, Piotr Guzik, and Mikołaj Leszczuk. "Automated detection of firearms and knives in a CCTV image." Sensors 16, no. 1 (2016): 47.

[6] "China’s CCTV surveillance network took just 7 minutes to capture BBC reporter – TechCrunch", TechCrunch, 2019. [Online]. Available: https://techcrunch.com/2017/12/13/china-cctv- bbc-reporter/. [Accessed: 15- Jul- 2019].

[7] Cohen, Neil, Jay Gattuso, and Ken MacLennan- Brown. CCTV operational requirements manual 2009. Home Office Scientific Development Branch, 2009.

[8] Flitton, Greg, Toby P. Breckon, and Najla Megherbi. "A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery." Pattern Recognition 46, no. 9 (2013): 2420-2436.

[9] Gesick, Richard, Caner Saritac, and Chih-Cheng Hung. "Automatic image analysis process for the detection of concealed weapons." In Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security

and Information Intelligence Challenges and Strategies, p. 20. ACM, 2009

[10] Tiwari, Rohit Kumar, and Gyanendra K. Verma. "A computer vision based framework for visual gun detection using harris interest point detector." Procedia Computer Science 54 (2015): 703-712.

[11] Tiwari, Rohit Kumar, and Gyanendra K. Verma. "A computer vision based framework for visual gun detection using SURF." In 2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO), pp. 1-5. 2015.

[12] Xiao, Zelong, Xuan Lu, Jiangjiang Yan, Li Wu, and Luyao Ren. "Automatic detection of concealed pistols using passive millimeter wave imaging." In 2015 IEEE International Conference on Imaging Systems and Techniques (IST), pp. 1-4. IEEE, 2015.

[13] Flitton, Greg, Toby P. Breckon, and Najla Megherbi. "A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery." Pattern Recognition 46, no. 9 (2013): 2420-2436.

[14] Sheen, David M., Douglas L. McMakin, and Thomas

E. Hall. "Three-dimensional millimeter-wave imaging for concealed weapon detection." IEEE Transactions on microwave theory and techniques 49, no. 9 (2001): 1581-1592.

[15] Xue, Zhiyun, Rick S. Blum, and Y. Li. "Fusion of visual and IR images for concealed weapon detection." In Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002. (IEEE Cat. No. 02EX5997), vol. 2, pp. 1198-1205. IEEE, 2002.

[16] Blum, Rick, Zhiyun Xue, Zheng Liu, and David S. Forsyth. "Multisensor concealed weapon detection by using a multiresolution mosaic approach." In IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004, vol. 7, pp. 4597-4601. IEEE, 2004.

Author's profile:

SM.RAFI M.Tech

Currently he is working Associate professor in Audisankara college of engineering and technology Gudur, Tirupati (Dt)

He is done M.Tech from Gokula Krishna college of engineering sullure pet.He is dedicated to teaching field from the last 14 years .

A.LAVANYA is pursuing MCA from Audisankara college of engineering and technology, Gudur, Affiliated to JNTUA in 2024, Andhra Pradesh, India .