International Journal of Sociotechnology and Knowledge Development
Volume 12 • Issue 1 • January-March 2020

Automatic Gun Detection Approach for Video Surveillance

Mai Kamal el den Mohamed, Faculty of Computer and Informatics, Cairo, Egypt
Ahmed Taha, Faculty of Computer and Informatics, Cairo, Egypt
Hala H. Zayed, Faculty of Computer and Informatics, Cairo, Egypt

ABSTRACT

The immense crime rates resulting from using pistols have led governments to seek solutions to deal with such terrorist incidents. These incidents have a negative impact on public security and cause panic among citizens. From this point, facing a pandemic of weapon violence has become an important research topic. One way to reduce this kind of violence is to prevent it via remote detection and to give an appropriate response in a short time. Video surveillance is the process of monitoring the behavior of people and objects. Surveillance systems can be employed in security applications as legal evidence. Moreover, it is used widely in suspicious activity detection applications. Intelligent video surveillance systems (IVSSs) are the use of automatic video analytics to enhance the effectiveness of traditional surveillance systems. With the rapid development in Deep Learning (DL), it is now widely used to address the problems existing in traditional detection techniques. In this article, an approach to detect pistols and guns in video surveillance systems is proposed. The presented approach does not need any invasive tools in the weapon detection process. It uses DL in the classification and the detection processes. The proposed approach enhances the obtained results by applying Transfer Learning (TL). It employs two different DL techniques: AlexNet and GoogLeNet. Experimental results verify the adaptability of detecting different types of pistols and guns. The experiments were conducted on a benchmark gun database called Internet Movie Firearms Database (IMFDB). The results obtained suggest that the proposed approach is promising and outperforms its counterparts.

KEYWORDS

Closed Circuit Television, Convolutional Neural Network, Deep Neural Network, Transfer Learning, Video Surveillance Systems

1. INTRODUCTION

The high rate of crime and violence among people is considered the third leading cause of death in 53 countries according to the report of the World Health Organization (WHO) European Region (Sethi et al., 2010). These alarming rates force governments to try to find solutions for such dangerous problems. Video surveillance systems are used for analyzing the objects behavior (Amira & Zagrouba, 2018). It involves object classification to understand the events (normal or abnormal) in videos. Abnormal activity detection plays a crucial role in surveillance applications (Huang et al., 2017; Wang et al., 2018; Cosar et al., 2017; Lloyd et al., 2017; Tripathi et al., 2019). The large-scale presence of surveillance systems is a real source of inspiration for the development of an automated system to detect problems of anti-social behavior such as vandalism, fights, gun killings, etc. In most current surveillance systems, monitoring depends on the existence of a human element. This

DOI: 10.4018/IJSKD.2020010103

makes monitoring a very challenging task. In addition, it is labor-intensive and prone to errors. These traditional systems have many problems such as weak security, low intelligence, high cost, and poor stability. Most of these systems are based on human operators. It is difficult for these operators to watch and analyze all the dangerous situations, especially with the long observation periods and a large number of cameras (Research, 2003). The reports included in (Research, 2003; Cohen et al., 2009; Dadashi, 2008) confirm that the Closed-Circuit Television (CCTV) operator suffers from video blindness after 20 to 40 minutes of active monitoring. In the last two decades, researchers and professionals of the industry have devoted their studies to develop surveillance systems that discover suspicious actions (Zhou & Tan, 2010; Liwei et al., 2010; Kishore et al., 2012; Mandrupkar et al., 2013). Automation is required in complex situations to reduce the workload of the human operator and improve the performance. Hence, surveillance systems still require intervention, improvement, and conversion from traditional surveillance systems to intelligent and smart systems (Shah et al., 2007; Tian et al., 2008). There is no human intervention at all in IVSS. The smart surveillance system automatically triggers an alert if any suspicious action or any illegal activity occurs. Accordingly, the operator focuses his attention only on the video feed and takes the convenient action.

The goal of the proposed approach is to design a system capable of automatically detect the presence of dangerous firearms especially, guns and pistols in real-time in the CCTV images. The proposed approach uses the Convolutional Neural Network (CNN) trained to determine the presence of the guns. CNN is a DL algorithm (Abdelouahab et al., 2018). DL is a subfield of machine learning. It is a technique that educates computers to perform what humans do naturally. Recently, with the emergence and successful deployment of DL techniques in image classification, researchers have emigrated from traditional techniques to DL techniques. DL has recently enriched its high ability in detection and classification. It has the ability to detect the dominant features automatically rather not manually (Tiwari & Verma, 2015; Halima & Hosam, 2016; Tiwari & Verma, 2015; Sheen et al., 2001; Xue et al., 2002; Li et al., 2008). This is the main reason prompted us to use it in our proposed approach. Nevertheless, DL suffers from two drawbacks: first, it requires very large datasets. Second, it needs high-performance computing resources. In order to overcome these two constraints. TL through fine-tuning is employed in the proposed approach. It is the improvement of learning in a new task through the transfer of knowledge from a learned task. TL means re-utilizing the knowledge learned from one problem to another one (Torrey & Shavlik, 2009). Network weights are initialized randomly if a network training is from scratch. However, the weights are initially set to the weights of the pre-trained network if fine-tuning is used. TL technique seeks to save time and get better performance. Figure 1 explains how TL improves the training performance rate.

In the proposed approach, DL has been employed to provide a greater level of performance than other traditional techniques (Mery et al., 2013; Blum et al., 2004; Upadhyay & Rana, 2014; Glowacz et al., 2015; Darker et al., 2007; Blechko et al., 2009; Darker et al., 2008; Arslan et al., 2015). The proposed approach allows the detection in noisy images with a low-quality resolution. Applying CNN in detecting firearms achieves efficient feature extraction results and accurate classification results. This increases the robustness of the presented approach. The presented approach uses two different pre-training networks (AlexNet and GoogLeNet). To avoid overfitting and to accelerate the process of training, the proposed approach uses TL. This training style demonstrates the ability of the TL in achieving tremendous results.

The overall organization of this paper is presented as follows: under section 2, the related work is explored. Section 3 describes the methodology of the proposed approach. The fourth section provides a detailed picture of the experimental results. The conclusion is approached in the last section.

2. RELATED WORK

Nowadays, automatic visual surveillance is an elementary need for security. Today, CCTV is employed as a monitoring and surveillance tool for fighting crimes. CCTV footage\films recently grow to be critical evidence in courts. All weapons, including firearms, pose very serious intimidation and risks to