Improving Handgun Detectors with Human Pose Classification

Jesus Ruiz-Santaquiteria, Oscar Deniz, Noelia Vallez, Alberto Velasco-Mata y Gloria Bueno
VISILAB, ETSII, Avda Camilo Jose Cela, 13071, Ciudad Real, Spain
{jesus.ralegre, oscar.deniz, noelia.vallez, alberto.velasco, gloria.bueno}@uclm.es

Abstract

Unfortunately, attacks with firearms such as handguns have become too common. CCTV surveillance systems can potentially help to prevent this kind of incidents, but require continuous human supervision, which is not feasible in practice. Image-based handgun detectors allow the automatic location of these weapons to send alerts to the security staff. Deep learning has been recently used for this purpose. However, the precision and sensitivity of these systems are not generally satisfactory, causing in most cases both false alarms and undetected handguns, particularly when the firearm is far from the camera. This paper proposes the use of information related to the pose of the subject to improve the performance of current handgun detectors. More concretely, a human full-body pose classifier has been developed which is capable of separating between shooting poses and other non-dangerous poses. The classified pose is then used to reduce both the number of false positives (FP) and false negatives (FN). The proposed method has been tested with several datasets and handgun detectors, showing an improvement under various metrics.

Keywords: Handgun detection, human pose classification, deep learning, CCTV surveillance, human pose estimation.

1 Introduction

The use of CCTV surveillance systems is widespread nowadays. In these systems, a human operator can observe the images captured by cameras looking for threats or security risks. It has been demonstrated that early detection of threats is crucial to reduce the possible damage caused [3].

In this paper we focus on crimes caused by firearms such as handguns. Unfortunately, these events have become commonplace in our society. Examples of these unacceptable situations are gunfire incidents on school grounds [4], terrorist attacks [15] or mass shootings in public places like airports, train stations, museums, churches or government buildings [8].

Several works have proposed the use of machine learning and computer vision techniques to help with this situation through the creation of automated surveillance systems, which can be applied to the CCTV surveillance images to automatically detect dangerous situations and notify the security staff. Novel deep learning methods, specially Convolutional Neural Networks (CNN) have achieved significantly better results than previous machine learning approaches in many image-based classification, detection or segmentation tasks. Because of this, in recent years several deep learning image-based handgun detectors have been proposed [7, 11]. Still, when these detectors are applied in a new scenario, for example a specific CCTV camera, the fact is that the false alarm ratio usually increases [19].

In addition, detecting handguns in CCTV images is a challenging task due to the features of these particular images. Usually, the camera is located far from the object of interest and the images retrieved present poor quality in terms of detail (low image resolution), blurriness, artifacts or overexposure. We have to consider that a handgun is a relatively small sized object, which hinders its localization in this context even for security staff.

In this paper, we propose using information related to the pose of the subject to improve the performance of current handgun detectors. Our main hypothesis is that the body posture taken during shooting is characteristic enough to be considered as useful information to detect handguns. Not only person location but pose keypoints relative to a detected person can be successfully obtained with modern deep learning-based pose estimators, even in CCTV image conditions. In this work, a full-body pose classifier has been developed which is capable of separating between shooting poses and other non-dangerous poses. The identified pose information is then applied to reduce both the number of false positives and the false negatives. In this work, a human pose is considered as a shooting pose when the individual is actually fir-

https://doi.org/10.17979/spudc.9788497498418.1040

ing a handgun or in the moments leading up to the shot (carrying a handgun in a non-shooting pose is not considered as a shooting pose). On the other hand, the rest of human poses are considered as non-dangerous poses.

The paper is organized as follows. Related work is summarized in Section 2. In Section 3, the datasets used in this study are described. The details of the proposed method are presented in Section 4. The experiments and results obtained are summarized in Section 5. Finally, conclusions and future work are given in Section 6.

2 Previous Work

Recent deep learning methodologies have been applied for detecting handguns in surveillance images. Usually, deep learning detection methods are divided into two main categories: sliding window and region proposals. The first approach consists on selecting a large number of region candidates from the input image, at multiple scales and locations. Then, a CNN classifies each one of the candidate regions. Several authors have used this approach to successfully detect handguns [5, 7]. On the other hand, region proposal methods select a lower number of region candidates. The selective search method [17] or the Region Proposal Network (RPN) are examples of this family of methods. The first method applies a variety of complementary image partitionings to deal with as many image conditions as possible. The RPN, which is included in the Faster R-CNN architecture [13], is a fully-convolutional network that simultaneously predicts object bounds and objectness scores at each position to generate high quality region proposals. In the seminal work [11], both sliding window and Faster R-CNN approaches were tested for handgun detection. The best results were obtained with the Faster R-CNN method, using a VGG-16 pre-trained on the ImageNet dataset. Another dataset composed of 3000 images of handguns from several YouTube videos was used for the fine-tuning process.

Unfortunately, although the results shown by the aforementioned detectors are promising, when they are deployed in a real surveillance scenario, a high rate of false positives and missed detections is to be expected.

In this work, the use of additional information related to the body pose is proposed to improve the performance of these handgun detectors. In the literature, several methods have been proposed for estimating 2D body poses. For multi-person pose estimation common approach, also named as top-down approach, is to first apply a person detector and then perform a single-person pose estimation for each detected person [6, 12]. In these methods, the more people there are, the greater the computational cost. On the other hand, bottom-up approaches do not need a person detector and the computational cost does not depend on the number of people appearing in the image. OpenPose [1] is a multi-person bottom-up pose estimation method based on Part Affinity Fields (PAFs). These PAFs are a set of 2D vectors which encode location and orientation of limbs over the image domain. The OpenPose architecture is able to jointly learn keypoint locations and keypoint associations.

In the proposed method, a human pose image-based classifier is trained to separate between shooting poses and other non-dangerous poses using the 2D pose keypoints as estimated with the OpenPose method.

3 Materials

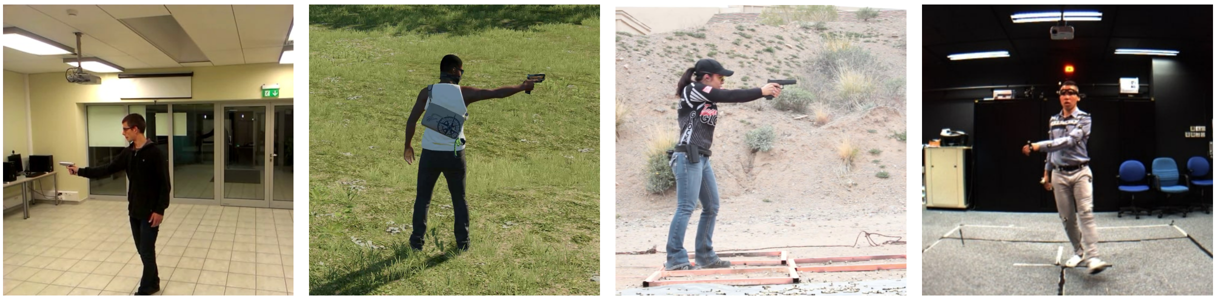
In this section, the datasets used in the experiments are presented. In order to consider different contexts and image features, four databases have been collected and labelled.

3.1 Data sources

- **Dataset A**: The first dataset is composed of 2525 handgun images of size 640x480 extracted from videos of the publicly available Guns Movies Database [7]. These clips show a man holding a handgun in various shooting poses in a single room.

- **Dataset B**: The second dataset was created with the popular shooter videogame Watch Dogs 2 [16] in a PC platform. Through the novel NVIDIA Ansel feature [10] it is possible to record in-game clips. In this way, video sequences can be captured from different positions, distances or angles. Moreover, a wide variety of shooting and non-shooting character poses can be obtained. Videos were recorded performing a full rotation of the camera around the subject at two different heights with 15 character animations. In this dataset there are 3418 images of size 3840x2160, including 1445 positives (images containing handguns) and 1973 negatives (images without handguns).

- **Dataset C**: The third dataset is a compilation of 2783 images of size 1920x1080 extracted from 11 YouTube videos, including 2135 positives (people holding handguns and



(a) Dataset A (b) Dataset B (c) Dataset C (d) Negatives dataset

Figure 1: Dataset samples

shooting in a variety of scenarios and different camera locations) and 648 negatives.

- **Negatives dataset**: A negatives dataset (images without any handgun or non-shooting pose) is needed to balance and complete the training, validation and test datasets. For this purpose, a set of 1240 randomly selected images of size 512x384 from the MADS (Martial Arts, Dancing and Sports) [20] database were collected. In these images, people appear with poses of different activities like dancing Hip Hop or Jazz, practicing TaiChi or playing football.

In Figure 1 some sample images from all data sources are shown. The datasets analysed during the current study are available from the corresponding author on reasonable request.

3.2 Dataset preparation

The images from the previously described databases have been collected to create the final dataset used for training and validating the proposed full-body pose classifier, as well as a different test set to check overall method performance.

The training dataset is composed of randomly selected positive and negative images from dataset B. Horizontal flip data augmentation is also applied to increase the dataset size. In Table 1 the dataset composition is presented.

Table 1: Training dataset composition after the data agumentation process. All positive and negative images have been extracted from dataset B.

	Train	**Valid**	**Test**	**Total**
Positive	2400	245	245	2890
Negative	2400	773	773	3936
Total	4800	1018	1018	6836

Also, to check the overall performance of the method, another dataset is created collecting images from the rest of the presented databases (A, C and negatives). Dataset B, which is used to train the full-body pose classifier, is not included. In Table 2 the dataset composition is summarized. Images from each dataset are randomly selected in order to obtain balance between positive and negative images.

Table 2: Testing dataset composition. All positive and negative images have been extracted from datasets A, C and Negatives.

	A	**C**	**Negatives**	**Total**
Positive	816	305	0	1121
Negative	0	0	1240	1240

4 Methodology

The proposed method is divided in three main steps. First, 2D full-body pose information is obtained from images with the OpenPose framework. Then, after a normalization step, the human pose is classified by a trained CNN to separate between shooting poses and other non-dangerous poses. Finally, the predicted pose is combined with the image-based handgun detector results to reduce both false positives and false negatives, see Figure 2.

4.1 Pose estimation

The 2D human pose can be defined as a set of 2D locations of anatomical keypoints, such as the eyes, neck, elbows or wrists. OpenPose, which is the framework used in this step, takes as an input an RGB image and generates a list of 25 2D keypoints for each person detected in the image. The predicted confidence for each point is also available. Along with the previous information, a hand detector [14] output is also included to produce the final pose skeleton. The hand detector works in the same way, 20 keypoints and a bounding box are generated for each detected hand. Figure 3 shows an example of the pose information

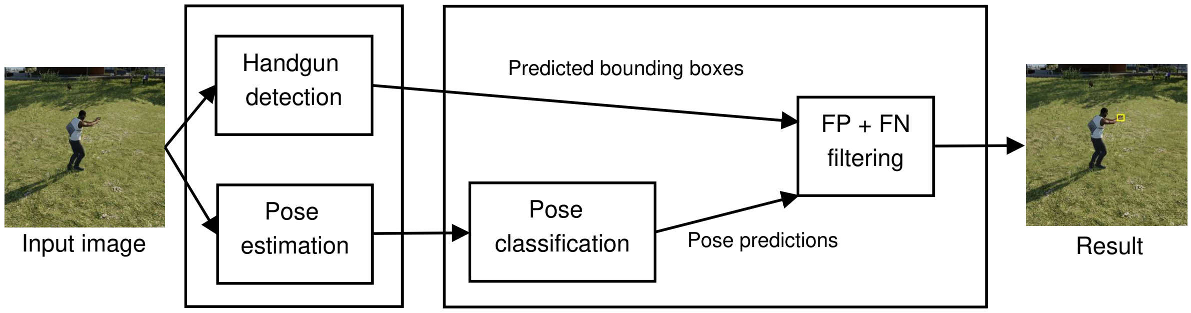
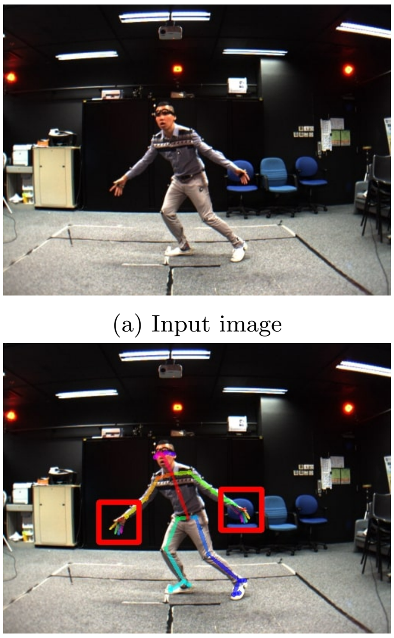


Figure 2: Overview of the proposed method. Left block, composed of the handgun detector and the pose estimator, contains the necessary external components for the proposed method (right block, which is composed of the human pose classifier and FP+FN filtering process)

obtained with OpenPose. The pose information is displayed over the original image (Figures 3a and 3b), along with the bounding box for each detected hand.

Apart from the individual confidence values for each keypoint, general confidence is also calculated for the whole pose, which is a weighted average of all OpenPose and hand keypoint scores. Low confidence poses, that are probably due to false positives or incomplete poses, are rejected.



(a) Input image

(b) OpenPose image

Figure 3: Pose estimation example. Best viewed in color

4.2 Pose classification

Using the pose information retrieved by OpenPose, a human pose classifier has been developed, which is capable of separating between shooting poses and other non-dangerous poses. The dataset used for training this classifier consists of positive and negative images extracted from dataset B, as explained in Subsection 3.2.

From this dataset, human pose images are generated. Variable factors such as the size of the pose detected, which depends on the distance from the subject to the camera, the position of the subject within the image or the image size may adversely affect the classification of the pose. To handle these variations and focus only on the relative position between the keypoints, a normalization procedure is needed for each detected pose. First, local body coordinates are calculated for each pose. The original neck keypoint j_0 is taken as reference and the distance between the neck and the lumbar spine keypoint j_1 is used as the scale factor for the normalization. In this way, the new keypoints k_n are calculated following Equation 1:

$$k_n = \frac{j_n - j_0}{|\overrightarrow{j_0 j_1}|} \tag{1}$$

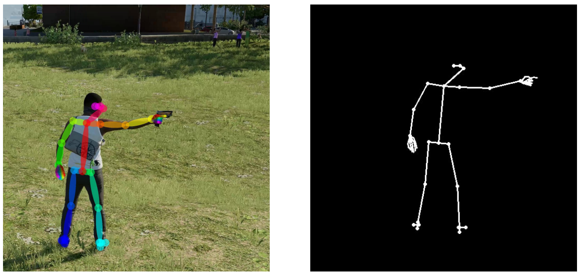
where j_n is the original 2D point and $|\overrightarrow{j_0 j_1}|$ is the distance between the neck and the lumbar spine keypoints (j_0 and j_1).

Finally, synthetic binary images of size 256x256 are created, moving and scaling the local coordinates to fit the new image size. An example of a normalized pose image (in this case a shooting pose sample) is presented in Figure 4b, along with the original OpenPose image (Figure 4a).

The binary pose image is then classified with a custom CNN-based architecture. It is formed by two Convolution-MaxPooling layers for the feature extraction step and a fully connected layer for the final classification.

The ReLU activation function is applied after each convolution layer to introduce non-linearity. Dropout is also used in the fully connected layers to reduce overfitting and categorical cross entropy is used as loss function. The training process

https://doi.org/10.17979/spudc.9788497498418.1040



(a) Original OpenPose image

(b) Normalized pose image

Figure 4: Pose classification example.

was carried out in a NVIDIA GeForce GTX 1060 GPU. The accuracy reached on the test images after 10 epochs was around 96% using the Adadelta optimizer.

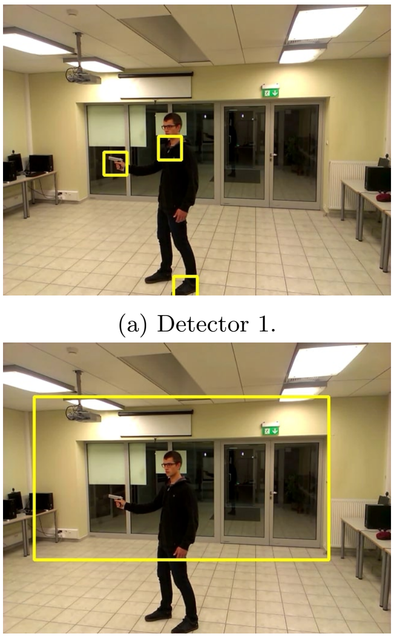
4.3 Handgun detection

In this work, two image-based handgun detectors have been used to test the proposed method. The first detector [18] is based on a Faster R-CNN architecture with a ResNet network backbone. The training dataset was composed of 871 images provided by the University of Seville [3], which were acquired from two CCTV cameras located in different college halls.

The second detector tested is also based on a Faster R-CNN but with a pre-trained VGG-16 network backbone in this case. The training dataset is composed of 3000 handgun images downloaded from several web-sites. This dataset was presented in [11].

In our test dataset both methods show the common problems of this kind of detectors, essentially an unacceptable rate of false positives and false negatives. For illustrative purposes, Figure 5 shows an example output for each detector. The first detector (Figure 5a) is capable of locating the handgun, but false positives are also detected. In this case, the use of the detection confidence as the criterion for removing FPs is not adequate, as all confidence values are similar. On the other hand, the second detector (Figure 5b) produces an extremely large bounding box, being unable to locate the handgun accurately.

The poor performance reached by these handgun detectors in our test dataset is also related with the datasets used for the training step. Some images from these datasets are significantly different (e.g., profile handgun images close to the camera) to the ones tested in our work. These kind of images are useful for classification purposes, but not for detecting handguns in CCTV images (actual video surveillance environments). The Region Proposal Network (RPN) of the Faster R-CNN detector will learn extremely large bounding box proposals, preventing the localization of the small handguns in CCTV images. To solve this issue both detectors have been retrained with another custom dataset composed of 1946 images extracted from dataset C in a balanced way. Images are resized before the feature extraction step (min dimension set to 600 pixels). These images have a similar appearance in terms of camera distance and handgun bounding box size to the test dataset ones, reaching a higher performance. Also, two different implementations for these Faster-RCNN methods have been tested. Implementation 1 is a Faster-RCNN implementation for Keras-TensorFlow deep learning library [9] and implementation 2 is the OpenMMLab Detection Toolbox [2], which includes Faster-RCNN models among others.



(a) Detector 1.

(b) Detector 2.

Figure 5: Handgun detectors performance examples

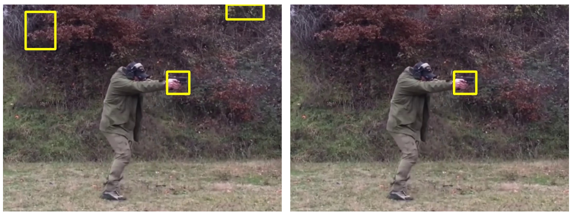
However, the number of FP and FN is still unacceptable for an automated handgun detector for video surveillance applications.

4.4 FP filtering

In the proposed method the information related to the pose of the subject is used. To reduce the FP detection rate, only detections close to hand locations are taken into account.

The selected criterion for this FP filtering stage

https://doi.org/10.17979/spudc.9788497498418.1040



(a) Handgun detector image (b) Output image after FP filtering

Figure 6: FP filtering example.

is the intersection over Union (IoU) between the hand locations retrieved by the pose estimation framework and the predicted handgun locations. This index is a measure that evaluates the overlap between two bounding boxes. It is also known as the Jaccard similarity index, and follows Equation 2.

$$IoU = \frac{area(BB_a \cap BB_b)}{area(BB_a \cup BB_b)} \tag{2}$$

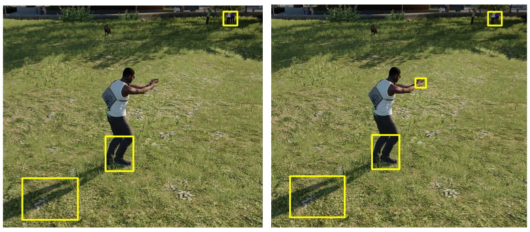
This metric is applied to establish if a predicted handgun bounding box is overlapped with the hand location. In this case, the decision threshold is 0.5. In this way, most false alarms can be removed. It is important to clarify that this approach assumes that the handguns will be detected (and the alert will be sent to the security staff) only if they are held by a person.

Although a large number of FP detections will be filtered using this approach, there are a few false positives on the hand areas. To manage this situation, the detection confidence in these cases will be modified taking into account the pose classifier decision. The new detection confidence will be the average between the handgun detection confidence and the predicted confidence for a shooting pose. Thus, if the subject has a shooting pose with a high confidence and a detected handgun on any hand area, the detection will be considered as a true detection with a high probability. On the other hand, if the pose classifier predicts a non-dangerous pose, the handgun detection will have a lower confidence value.

An example of the proposed FP filtering is presented in Figure 6. The output of the handgun detector is shown in Figure 6a and Figure 6b shows the output image after the FP filtering. After the filtering, only the correct detection is maintained.

4.5 FN filtering

Another significant drawback of the tested detectors is the high rate of undetected handguns. As in the case of the FP filtering, FNs can be also reduced applying the pose information retrieved from the pose estimation framework and the developed pose classifier.



(a) Handgun detector image (b) Output image after FN filtering

Figure 7: FN filtering example.

If the detector is not capable of locating the handgun (when it is far from the camera or the image does not have enough detail) but the pose classifier predicts a potentially dangerous pose (identified as a shooting pose with high confidence), an artificial handgun area will be added to the detector output. This artificial detection will be added to the farthest detected hand with respect to the pose reference keypoint (neck keypoint) and will have a fixed confidence value lower than the average detections (0.8), as it is not a true detection from the handgun detector.

An example of this FN filtering is shown in Figure 7. The output of the original handgun detector is shown in Figure 7a and in Figure 7b the output image after the FN filtering is presented. After the process, a new artificial detection is included. In this particular example, the combination of the proposed FP and FN filtering will lead to a correct handgun detection with no FP.

5 Experiments and results

The proposed pose-based method has been tested with the test dataset described in Subsection 3.2. The handgun detection methods as well as the implementations used are presented in Subsection 4.3. Evaluation results for both implementation 1 (Keras-TensorFlow [9]) and implementation 2 (Open-MMLab Detection Toolbox [2]) are shown in Table 3 and Table 4, respectively.

For each detector and implementation, evaluation metrics have been calculated in four different situations. The first one, named as “Original” in the tables, illustrates the detectors performance without any post-processing method applied. The second one, labelled as “Filtered (FP)”, shows how only the pose-based FP filtering method (Subsection 4.4) significantly reduces the number of false positives. It can be observed that Precision met-

https://doi.org/10.17979/spudc.9788497498418.1040

Table 3: Evaluation results for detector 1 [18] and detector 2 [11] using the implementation 1 (Keras-TensorFlow). Precision and recall metrics have been calculated with a 0.5 confidence threshold.

		Precision	**Recall**	**mAP**
Detector 1	Original	0.5398	0.5950	52.19
	Filtered (FP)	**0.8832**	0.5421	51.48
	Filtered (FN)	0.5687	**0.8683**	73.34
	Filtered (FP+FN)	0.7845	0.8154	**74.86**
Detector 2	Original	0.1580	0.7392	54.37
	Filtered (FP)	**0.8759**	0.6891	64.85
	Filtered (FN)	0.1762	**0.8701**	58.14
	Filtered (FP+FN)	0.7842	0.8208	**75.89**

Table 4: Evaluation results for detector 1 [18] and detector 2 [11] using the implementation 2 (Open-MMLab Detection). Precision and recall metrics have been calculated with a 0.5 confidence threshold.

		Precision	**Recall**	**mAP**
Detector 1	Original	0.6939	0.7312	64.01
	Filtered (FP)	**0.8930**	0.6953	63.65
	Filtered (FN)	0.6623	**0.8719**	74.74
	Filtered (FP+FN)	0.8022	0.8360	**75.76**
Detector 2	Original	0.3070	0.4247	33.10
	Filtered (FP)	**0.7648**	0.4050	35.22
	Filtered (FN)	0.4210	**0.8029**	64.01
	Filtered (FP+FN)	0.7432	0.7832	**65.53**

ric increases when applying this method, obtaining the highest mark. On the other hand, "Filtered (FN)", shows the results for the FN filtering method (Subsection 4.5), which increases the number of detected handguns through the human full-body pose classifier, obtaining the highest Recall values. Finally, "Filtered (FP+FN)" shows the effect of using both FP and FN filtering methods combined. In this case, metrics are more balanced in terms of Precision and Recall, achieving the highest mAP marks.

6 Conclusions

In this work, a new approach to improve the performance of current deep learning-based handgun detectors is proposed. This method consists of using human pose information to reduce both false positive detections and undetected handguns. A CNN image-based classifier has been trained to distinguish between shooting poses and other non-dangerous poses. Then, this predicted pose along with the pose skeleton itself is used to remove false alarms and predict the potential undetected handgun locations.

Four datasets have been collected and two well-known handgun detectors under two different implementations have been applied to demonstrate the ability of the proposed method to improve the detection performance. The evaluation metrics show that in the test images the proposed method reduces the number of false positives and false negatives, obtaining significantly higher results in terms of Precision, Recall and mAP. Although the proposed method was originally designed to be applied with CCTV surveillance cameras, the large variety of images in which the method has been tested, such as synthetic videogame images, indoor places with artificial light and outdoor locations in different contexts and camera positions, show that this approach can be applied in a wide number of scenarios.

Despite the promising results achieved, there is room for improvement. The first major limitation of the proposed work is related to lack of publicly available datasets of mass shootings and handgun assaults, which prevents evaluating the proposed method under more realistic and challenging conditions (e.g. crowds and hidden body parts). Regarding the pose estimator used (OpenPose), although the overall performance is satisfactory, there are situations in which the pose, completely or partially, is not detected correctly. Thus, the next steps of the method, such as the pose classifier, will not work properly if the estimation is not sufficiently accurate. On the other hand, the extra computational cost added by the pose estimator and the pose classifier must be studied in order to analyze its impact when deployed in real time systems, as the number of images per second

https://doi.org/10.17979/spudc.9788497498418.1040

that the system is able to process may be affected. Finally, different ways to encode pose information or alternative CNN architectures can be studied to improve the performance of the pose classifier.

Acknowledgement

This work was partially funded by projects PDC2021-121197-C22 (funded by MCIN/AEI/ 10.13039/501100011033 and by the European Union NextGenerationEU/PRTR) and SBPLY/21/180501/000025 (funded by the Autonomous Government of Castilla-La Mancha and the European Regional Development Fund -ERDF-). The first author is supported by Postgraduate Grant PRE2018-083772 from the Spanish Ministry of Science, Innovation, and Universities.

References

[1] Cao, Z., Hidalgo, G., Simon, T. et al. (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE transactions on pattern analysis and machine intelligence*, *43*(1), 172–186.

[2] Chen, K., Wang, J., Pang, J., Cao, Y. et al. (2019). MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*.

[3] Enriquez, F., Soria, L. M., Alvarez-Garcia, J. A., Caparrini, F. S. et al. (2019). Vision and crowdsensing technology for an optimal response in physical-security. *International conference on computational science*, 15–26.

[4] Everytown for Gun Safety Support Fund. (2020). Gunfire on School Grounds in the United States [Accessed: 28/06/2022].

[5] Gelana, F., & Yadav, A. (2019). Firearm detection from surveillance cameras using image processing and machine learning techniques. *Smart Innovations in Communication and Computational Sciences*, 25–34.

[6] Gkioxari, G., Hariharan, B., Girshick, R., & Malik, J. (2014). Using k-poselets for detecting people and localizing their keypoints. *Procs. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 3582–3589.

[7] Grega, M., Matiolanski, A., Guzik, P., & Leszczuk, M. (2016). Automated detection of firearms and knives in a CCTV image. *Sensors*, *16*(1), 47.

[8] Gun Violence Archive. (2021). Mass Shootings [Accessed: 28/06/2022].

[9] Kim, Y. (2020). Keras-FasterRCNN implementation [Accessed: 28/06/2022].

[10] NVIDIA Corporation. (2020). NVIDIA Ansel website [Accessed: 28/06/2022].

[11] Olmos, R., Tabik, S., & Herrera, F. (2018). Automatic handgun detection alarm in videos using deep learning. *Neurocomputing*, *275*, 66–72.

[12] Pishchulin, L., Jain, A., Andriluka, M., Thormählen, T. et al. (2012). Articulated people detection and pose estimation: Reshaping the future. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3178–3185.

[13] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 91–99.

[14] Simon, T., Joo, H., Matthews, I., & Sheikh, Y. (2017). Hand keypoint detection in single images using multiview bootstrapping. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 1145–1153.

[15] Tessler, R. A., Mooney, S. J., Witt, C. E., O'Connell et al. (2017). Use of firearms in terrorist attacks: Differences between the United States, Canada, Europe, Australia, and New Zealand. *JAMA internal medicine*, *177*(12), 1865–1868.

[16] Ubisoft Entertainment. (2020). Watch Dogs 2 official website [Accessed: 28/06/2022].

[17] Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. *International journal of computer vision*, *104*(2), 154–171.

[18] Vallez, N., Velasco-Mata, A., Corroto, J. J., & Deniz, O. (2019). Weapon detection for particular scenarios using deep learning. *Iberian Conference on Pattern Recognition and Image Analysis*, 371–382.

[19] Vállez, N., Bueno, G., & Déniz, O. (2013). False positive reduction in detector implantation. *Conference on Artificial Intelligence in Medicine in Europe*, 181–185.

[20] Zhang, W., Liu, Z., Zhou, L., Leung, H. et al. (2017). Martial arts, dancing and sports dataset: A challenging stereo and multi-view dataset for 3D human pose estimation. *Image and Vision Computing*, *61*, 22–39.