Data in Brief 52 (2024) 110030

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Firearm-related action recognition and object detection dataset for video surveillance systems

Jesus Ruiz-Santaquiteria*, Juan D. Muñoz, Francisco J. Maigler, Oscar Deniz, Gloria Bueno

VISILAB, E.T.S. Ingeniería Industrial, University of Castilla-La Mancha, Avda. Camilo José Cela, Ciudad Real, Spain

A R T I C L E I N F O

Article history:
Received 29 November 2023
Revised 22 December 2023
Accepted 2 January 2024
Available online 5 January 2024

Dataset link: Action recognition and object detection dataset for firearm-related actions (Original data)

Keywords:
CCTV video surveillance
Firearm detection
Human action recognition
Object detection
Video classification

A B S T R A C T

The proposed dataset is comprised of 398 videos, each featuring an individual engaged in specific video surveillance actions. The ground truth for this dataset was expertly curated and is presented in JSON format (standard COCO), offering vital information about the dataset, video frames, and annotations, including precise bounding boxes outlining detected objects. The dataset encompasses three distinct categories for object detection: "Handgun", "Machine_Gun", and "No_Gun", dependent on the video's content. This dataset serves as a resource for research in firearm-related action recognition, firearm detection, security, and surveillance applications, enabling researchers and practitioners to develop and evaluate machine learning models for the detection of handguns and rifles across various scenarios. The meticulous ground truth annotations facilitate precise model evaluation and performance analysis, making this dataset an asset in the field of computer vision and public safety.

* Corresponding author.
E-mail address: Jesus.RAlegre@uclm.es (J. Ruiz-Santaquiteria).
Social media: @GrupoVisilab (J. Ruiz-Santaquiteria)

https://doi.org/10.1016/j.dib.2024.110030

Specifications Table

Subject	Computer Vision and Pattern Recognition
Specific subject area	Human action recognition and object detection
Data format	Raw, Analyzed, Filtered .mp4 files (video sequences) .json files (object detection data in COCO format)
Type of data	Video, Text
Data collection	The data were collected from various surveillance cameras, with a focus on CCTV cameras. We used a PTZ camera (HiWatch Series, specifically the model HWP-N2204IH-DE3 [1], with a 640×480 resolution and 25FPS), as well as mobile phones (OnePlus 6T [2], with a 1920×1080 resolution and 30FPS; OnePlus 5T [3], with a 1920×1080 resolution and 30FPS). The data collection involved recording 398 videos, each featuring individuals engaging in different actions, including the presence or absence of firearms. The ground truth annotations were manually generated by an expert using MATLAB (Video Labeler) [4], and each video included JSON files with precise bounding boxes around detected objects.
Data source location	The data were collected in the University of Castilla-La Mancha, Avda. Camilo José Cela, Ciudad Real, Spain Latitude: 38.99205199635681 Longitude: −3.9190148190388827
Data accessibility	Repository name: Action recognition and object detection dataset for firearm-related actions [5] Data identification number: 10.17632/bbzpxhd22j.2 Direct URL to data: https://data.mendeley.com/datasets/bbzpxhd22j/2 The dataset is organized as follows: 1. Main Dataset Folder (Gun_Action_Recognition_Dataset): contains three subfolders corresponding to the three categories/classes: "Handgun", "Machine_Gun", and "No_Gun". 2. Category Subfolders (e.g., Handgun): each category folder contains multiple subfolders, one for each video recorded. The names of these subfolders depend on the category and the action carried out by the subject. Within each video subfolder, there are the following components: - "video.mp4": the video in .mp4 format, featuring an individual performing specific actions related to the category. - "label.json": contains the details of the "video.mp4″ video about bounding boxes (xywh format), frame IDs, and object categories (the “No_gun” category does not contain annotations), as well as general information and licenses.

1. Value of the Data

- Valuable for research in firearm detection and human action recognition: these data provide a diverse collection of real-world surveillance footage, contributing to the development and assessment of machine learning models in these areas.
- Benefits researchers and practitioners: researchers in computer vision, security, and surveillance stand to gain from this dataset, allowing them to advance the state-of-the-art in firearm detection and human action recognition.
- Reusable for model development: other researchers can utilize these data to train and validate their own firearm detection and action recognition algorithms, thus improving public safety and security.
- Supports model benchmarking: these data enable the benchmarking of firearm detection and action recognition models across various scenarios captured by surveillance cameras, ensuring a standardized evaluation platform.
- Encourages interdisciplinary collaboration: the dataset’s accessibility can foster collaboration between computer vision experts and those in law enforcement and public safety, driving innovative solutions.
- Promotes ethical AI applications: these data can contribute to the responsible and ethical deployment of AI technologies in surveillance and security contexts.

2. Data Description

The timely identification of hazardous objects, such as firearms, within images is of utmost importance in mitigating potential harm [6–9]. The dataset presented herein offers a comprehensive assortment of authentic surveillance footage, thereby facilitating the advancement and evaluation of machine learning models within these domains.

The dataset is organized as follows:

1. Main Dataset Folder (Gun_Action_Recognition_Dataset)

Contains three subfolders corresponding to the three categories: "Handgun", "Machine_Gun", and "No_Gun".

2. Category Subfolders (e.g., Handgun)

Each category folder contains multiple subfolders, one for each video recorded. The names of these subfolders depend on the category and the action carried out by the subject. See the explanation below (Tables 1–3).

Every video contains a single person performing a specific action:

- **If it is a negative class ("No_gun"):**

Table 1
Description of "No_Gun" actions.

Negative class (No Gun)	
N1	Walking with empty hands
N2	Jogging
N3	Running
N4	Sneaking up with empty hands
N5	Holding a phone with a single hand (Right/Left) in a relaxed state
N6	Holding a phone with a single hand (Right/Left) looking at it
N7	Holding a phone with both hands looking at it
N8	Holding a phone with a single hand recording a video
N9	Holding a phone with both hands recording a video
N10	Holding a water bottle or soda can with one hand in a relaxed state
N11	Drinking from a water bottle or soda can
N12	Holding something heavy with both hands

Table 2
Description of "Handgun" actions.

Positive class (Handgun)						
Code	Action	Hands used	Target	Arm position	Pose	Mode
PCH1	Carry	One	Any	Any	Standing	Normal
PCH2	Carry	Both	Floor	Any	Standing	Normal
PCH3	Carry	Both	Ceiling	Any	Standing	Normal
PCH4	Carry	Both	Floor	Any	Crouched	Stealth
PCH5	Carry	Both	Ceiling	Any	Standing	Stealth
PCH6	Carry	Both	Floor	Any	Crouched	Normal
PCH7	Carry	Both	Ceiling	Any	Crouched	Normal
PAH1	Aim	One	Any	Fully extended	Standing	Normal
PAH2	Aim	One	Any	Fully extended	Standing	Stealth
PAH3	Aim	One	Any	Tucked	Standing	Normal
PAH4	Aim	One	Any	Tucked	Standing	Stealth
PAH5	Aim	Both	Floor	Any	Standing	Normal
PAH6	Aim	Both	Forward	Any	Standing	Normal
PAH7	Aim	Both	Floor	Any	Standing	Stealth
PAH8	Aim	Both	Forward	Any	Standing	Stealth
PAH9	Reload	Both	Any	Fully extended	Standing	Normal

Table 3
Description of "Machine gun" actions.

Positive class (Machine Gun)						
Code	Action	Hands used	Target	Arm position	Pose	Mode
PCW1	Carry	Both	Any	Tucked	Standing	Normal
PCW2	Carry	Both	Floor	Any	Standing	Normal
PCW4	Carry	Both	Floor	Any	Standing	Stealth
PCW5	Carry	Both	Ceiling	Any	Standing	Normal
PCW7	Carry	Both	Floor	Any	Crouched	Normal
PCW8	Carry	One	Floor	Any	Standing	Normal
PCW9	Carry	One	Ceiling	Any	Standing	Normal
PAW1	Aim	Both	Floor	Fully extended	Standing	Normal
PAW2	Aim	Both	Forward	Fully extended	Standing	Normal
PAW4	Aim	Both	Forward (tilted to one side)	Fully extended	Standing	Normal
PAW5	Aim	Both	Floor	Any	Standing	Stealth
PAW6	Aim	Both	Forward	Fully extended	Standing	Stealth
PAW7	Kneel down while aiming	Both	Any	Fully extended	Crouched	Stealth
PAW8	Reload while aiming	Both	Any	Fully extended	Standing	Normal

- **If it is a positive class, it could be “Handgun” or “Machine_Gun”:**

Every video is classified depending on the subject filmed from **V1** to **V4** (4 subjects). Also, there are different levels of brightness:

- **HB**: High brightness.
- **LB**: Low brightness.

Two cameras have been used: **C1** (IP camera) or **C2** (smartphone camera). Finally, the cameras have been placed in different locations: from **P1** to **P5** (5 different locations).

Each video subfolder is composed of:

- **Annotations JSON File (label.json)**
  - Manually generated ground truth annotations in JSON COCO standard format.
  - Contains details about bounding boxes (xywh format), frame IDs, and object categories (the “No_gun” category does not contain annotations), as well as general information and licenses.
- **Video File (video.mp4)**
  - The corresponding video in .mp4 format, featuring an individual performing specific actions related to the category.

This structured layout allows users to navigate and access the dataset, its annotations and videos for each category, making it suitable for research and analysis in the field of firearm detection and human action recognition.

The number of videos recorded per class and additional information is shown in Table 4.

Table 4
Additional information about the recorded videos per class.

Class	Nº of videos	Average frames per video	Average video length (s)
Handgun	141	249	8,82
Machine gun	139	257	9,33
No gun	118	246	8,65

Also note that we have carried out a process of anonymization of the individuals appearing in the videos. We have done this by following the next steps:

- Read the actual frame of the video.

- Obtain the keypoints of the individual using a pose estimator (YOLOv8x-pose [10] in our case).
- Extract the face keypoints.
- Calculate a bounding box including those keypoints (with a certain margin).
- Apply a Gaussian filter to the bounding box.

As far as the authors know, this process is not reversible.

To analyze whether this process has any impact on the pose keypoints localization in the images, a series of checks have been carried out.

To this end, we have compared two videos. Both are the same video, but one includes the anonymization process, while the other does not. Specifically, we have analyzed the video stored in the PAH1_C1_P1_V1_HB_3 folder, as it is a representative example.

Using a pose detector (e.g., YOLOv8x-pose [10]) we can conclude that the anonymization of the individual influences the pose detected, mostly the five keypoints located on the head (some of them are not recognised and, if recognised, the confidence value is smaller). The confidence value for each keypoint obtained from the frame 60 of the video is included in Table 5.

Table 5
Confidence values for the keypoints of the frame 60 of the video PAH1_C1_P1_V1_HB_3.

Keypoint	Confidence value (original video) (%)	Confidence value (anonymized video) (%)
Nose	99.27	79.94
Left eye	89.26	0.00
Right eye	99.53	77.88
Left ear	0.00	0.00
Right ear	98.67	81.64
Left shoulder	99.66	98.70
Right shoulder	99.87	99.68
Left elbow	96.68	93.30
Right elbow	99.42	99.28
Left wrist	96.92	91.33
Right wrist	99.31	98.60
Left hip	99.92	99.93
Right hip	99.95	99.97
Left knee	99.91	99.89
Right knee	99.95	99.96
Left ankle	99.55	99.49
Right ankle	99.72	99.75

As can be shown in the table, the anonymization process mainly affects the face keypoints, while the rest of them are barely modified in terms of confidence values and keypoint locations. Depending on the specific application, this anonymization process may interfere or not. If the aim is to detect weapons such as handguns, as usually they are not located in the head area, the detection performance should not be affected. Futhermore, for human pose estimation, if the head keypoints are not necessary, the remaining body keypoints are detected without any problem.

The general structure of the dataset is graphically summarized in Fig. 1.

The dataset is licensed under CC BY NC 3.0. However, we do allow commercial use, but only if prior permission is requested.

A comparison has also been made with other related datasets. In Table 6, a summary of relevant data is shown.

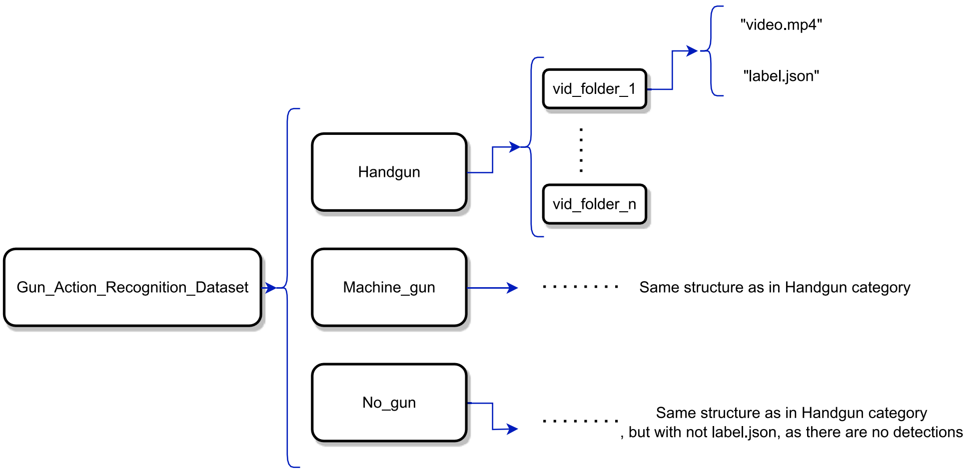

Fig. 1. General structure of the dataset. In the figure, where "vid_folder_1″ to "vid_folder_n" are the previously explained video names.

Table 6
Related datasets. Basic characteristics.

Source	#images	#videos (frames)	Handgun/Rifle	Downloadable
MGD [11]	602	250 (2857 total)	Handgun	No
URST [12]	–	- (5149 total)	Handgun	Yes
UCF [13]	–	1900 (≈ 7247 each)	Both	No
Granada [14]	3000	–	Handgun	Yes
Edgecase.ai [15]	4683	–	Handgun	Yes
Shenzen [16]	51,889	–	Both	No
IMFDB [17]	>12,758	>270,037	Both	Yes
SU [18]	2745	–	Both	No
MU [19]	–	250 (5500 total)	Handgun	No
AGH [20]	–	- (12,000 total)	Handgun	No
Youtube-GDD [21]	–	343 (5000 total)	Both	Yes
UCLM (ours)	–	392 (≈ 250 each)	Both	Yes

Note that we have not divided our dataset into train/validation/test. All data is available, and users can split the dataset depending on the purpose for which they want to use it. However, the recommended setup is to consider different locations for training/validation and test. For example, locations P1-P2-P3 could be used for training/validation and P4-P5 to test the performance of the trained models in completely new camera locations. This can be useful to check if the method suffers from overfitting to specific locations. Moreover, a cross validation setup is also suggested, performing different training/test workflows applying different locations at each step to cover all possible scenarios.

3. Experimental Design, Materials and Methods

Data Collection Process:

1. **Surveillance Footage:** the dataset is primarily composed of surveillance footage obtained from various sources:
   - PTZ camera: HiWatch Series, specifically the model HWP-N2204IH-DE3 [1]
   - OnePlus 6T camera [2]
   - OnePlus 5T camera [3]
2. **Video Recording:** a total of 398 videos were recorded. Each video features an individual performing specific actions, which relate to the presence or absence of firearms.
3. **Categorization:** the videos were categorized into three groups: "Handgun", "Machine_Gun", and "No_Gun". Each category represents a different scenario in video surveillance.
4. **Manual Annotation:** ground truth annotations were manually generated by an expert. For each video, a JSON file was created following the COCO format. This JSON file contains precise

information about the bounding boxes of detected objects (xywh format), frame details, and object categories.

Tools and Software:

- Video Recording: We used a PTZ camera (HiWatch Series, specifically the model HWP-N2204IH-DE3 [1], with a 640×480 resolution and 25FPS), as well as mobile phones (OnePlus 6T [2], with a 1920×1080 resolution and 30FPS; OnePlus 5T [3], with a 1920×1080 resolution and 30FPS).
- Annotation: the ground truth annotations were manually created using Video Labeler [4], a tool from MATLAB.
- Video Format: the recorded videos are in .mp4 format.

Experimental Conditions:

- **Light conditions:** as mentioned in the "Data description" section, there are different levels of brightness in the videos recorded: HB (high brightness) and LB (low brightness).
- No other experimental conditions have been contemplated.

Limitations

- Bias and Realism: the dataset may not fully capture the diversity of real-world scenarios in video surveillance environments, as it focuses on specific actions related to firearms in schools, which may lead to bias in the data.

Ethics statement

Data collection was carried out in line with ethical procedures and standards. Explicit consent of the individuals filmed was sought. After collection, images were all anonymized so that identification of the individuals was no longer possible. No other identification data was collected. Data collection, processing and storage has been carried out in accordance with ethical rules and principles addressed in following European legal documents: (i) Declaration of Helsinki-Ethical Principles for Research Involving Human Subjects; APA guidelines for research ethics, (ii) Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995, (iii) Convention No. 108 of the Council of Europe for the protection of individuals with regard to automatic processing of personal data, (iv) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, (v) Ethical Guidelines on Assessment List for Trustworthy Artificial Intelligence (ALTAI).

Data Availability

Action recognition and object detection dataset for firearm-related actions (Original data) (Mendeley Data)

CRediT Author Statement

Jesus Ruiz-Santaquiteria: Conceptualization, Data curation, Software, Writing – review & editing; **Juan D. Muñoz:** Conceptualization, Data curation, Software, Writing – original draft; **Francisco J. Maigler:** Conceptualization, Data curation, Software; **Oscar Deniz:** Funding acquisition, Project administration, Supervision, Conceptualization; **Gloria Bueno:** Funding acquisition, Project administration, Supervision, Conceptualization.

Acknowledgements

This work was partially funded by project PDC2021-121197-C22 (funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR), Horizon Europe dAIEdge Grant no. 101120726 by the European Commission and project SBPLY/21/180501/000025 by the Autonomous Government of Castilla-La Mancha.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] HWP-N2204IH-DE3 camera. Product | HiWatch. (n.d.). https://www.hi-watch.eu/en-us/product/2014/network-ptz/2-0-mp-4-ir-network-ptz-camera. Accessed in December 2023.
[2] OnePlus 6T tech specs. Tech Specs - OnePlus. (n.d.). https://www.oneplus.com/global/6t/specs. Accessed in December 2023.
[3] OnePlus 5T - technical specification. OnePlus 5T - Technical Specification - OnePlus. (n.d.). https://www.oneplus.com/us/support/spec/oneplus-5t Accessed in December 2023.
[4] Matlab Video Labeler. Label video for computer vision applications - MATLAB. (n.d.). https://www.mathworks.com/help/vision/ref/videolabeler-app.html.
[5] J. Ruiz-Santaquiteria, J.D.l Muñoz, F.J. Maigler, Ó. Déniz, G. Bueno, Action Recognition and Object Detection Dataset for Firearm-Related Actions, Mendeley Data, V2, 2023, doi:10.17632/bbzpxhd22j.2.
[6] J. Ruiz-Santaquiteria, A. Velasco-Mata, N. Vallez, G. Bueno, J.A. Alvarez-Garcia, O. Deniz, Handgun detection using combined human pose and weapon appearance, IEEE Access 9 (2021) 123815–123826.
[7] A. Velasco-Mata, J. Ruiz-Santaquiteria, N. Vallez, O. Deniz, Using human pose information for handgun detection, Neural. Comput. Appl. 33 (24) (2021) 17273–17286.
[8] J. Salido, V. Lomas, J. Ruiz-Santaquiteria, O. Deniz, Automatic handgun detection with deep learning in video surveillance images, Appl. Sci. 11 (13) (2021) 6085.
[9] J. Ruiz-Santaquiteria, A. Velasco-Mata, N. Vallez, O. Deniz, G. Bueno, Improving handgun detection through a combination of visual features and body pose-based data, Pattern Recognit. 136 (2023) 109252.
[10] G. Jocher, A. Chaurasia, J. Qiu, YOLO by Ultralytics (Version 8.0.0) [Computer software], 2023 Accessed in December 2023. https://github.com/ultralytics/ultralytics.
[11] Jun Yi Lim, Md Istiaque Al Jobayer, Vishnu Monn Baskaran, Joanne Munyee Lim, John See, Kok Sheik Wong, Deep multi-level feature pyramids: application for non-canonical firearm detection in video surveillance, Eng. Appl. Artif. Intell. 97 (2021).
[12] Jose Luis Gonz ález, Carlos Zaccaro, Juan Alvarez-Garcia, Luis Soria Morillo, Fernando Caparrini, Real-time gun detection in CCTV : an open problem, Neural Netw. 132 (2020) 297–308 09.
[13] Waqas Sultani, Chen Chen, Mubarak Shah, Real-world anomaly detection in surveillance videos, CoRR (2018) abs/1801.04264.
[14] Roberto Olmos, Siham Tabik, Francisco Herrera, Automatic handgun detection alarm in videos using deep learning, CoRR (2017) abs/1702.05147.
[15] EdgeCase.ai. World's First Synthetic Gun Detection Dataset from EdgeCase.ai. https://www.edgecase.ai/articles/worlds-first-synthetic-gun-detection-dataset-from-edgecase-ai, Accessed in December 2023.
[16] Delong Qi, Weijun Tan, Zhifu Liu, Qi Yao, Jingfeng Liu, A gun detection dataset and searching for embedded device solutions, CoRR (2021) abs/2105.01058.
[17] IMFDB. Internet Movie Firearms Database. Accessed in December 2023.
[18] Justin Lai, Sydney Maples, Developing a real-time gun detection classifier, Course: CS231n, Stanford University, 2017.
[19] JunYi Lim, Md Istiaque Al Jobayer, Vishnu Monn Baskaran, Jomanne Munyee Lim, KokSheik Wong, John See, Gun detection in surveillance videos using deep neural networks, in: 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2019, pp. 1998–2002.
[20] Michal Grega, Andrzej Matiola ński, Piotr Guzik, Mikolaj Leszczuk, Automated detection of firearms and knives in a CCTV image, Sensors 16 (1) (2016).
[21] Yongxiang Gu, Xingbin Liao, and Xiaolin Qin. YouTube-GDD: a challenging gun detection dataset with rich contextual information, 2022.