Preparing a Novel Dataset For Improved Suspicious Activity Detection in Academia

Asad Hameed Soomro[1*], Rafaqat Hussain Arain[2], Riaz Ahmed Shaikh[3], Vijay Kumar Premy[4], Samar Abbas Mangi[5], Waseem Ahmed Buriro[6]

***Abstract*: Human Activity Recognition (HAR) is one of the important research areas with essential security applications in a variety of environments, including academic environments. However, available datasets frequently lack the real-world diversity required for effective training of HAR models. To bridge this gap, this paper presents the CampusWatch dataset, specifically collected to capture real-world activities within academic institutions for Suspicious Activity Detection (SAD). The dataset, collected through smartphone cameras, spans both indoor and outdoor scenes from academia. It includes ten activity categories, covering nine suspicious behaviors—such as kicking, punching, running—and tenth one is "normal" activity class. This paper also outlines key steps in data preparation, including data collection, challenges, and techniques such as video annotation, noise reduction, and preprocessing, providing a robust foundation for advancing research in SAD.**

***Keywords:* Human Activity Recognition; Suspicious Activity Detection; security applications; HAR models; academic institutions; THUMOS.**

INTRODUCTION

The Human Activity Recognition (HAR) domain has significantly emerged over the past few years [1]–[10]. HAR focuses on the identification and classification of human activities from video recordings collected through various surveillance or smartphone cameras. These video recordings contain the subject's behavior while performing activities such as walking, laughing, jogging, punching, and so on. The data about the behavior of people is used by the researchers to meet certain needs from various domains like health care, fitness or home automation [11] and for accurate identification of suspicious and non-suspicious human activities.

Suspicious or abnormal activities can vary depending on the situation and surroundings. Behaviors like running, colliding, falling, jumping, fighting, or slipping [12] can be classified as suspicious if the environment is an office, an airport, or a bank. Additionally, breaking into someone's home, not paying the fee on a metro bus, kidnapping, Shoplifting, and robberies, are examples of anomalous activities that can occur in indoor or outdoor contexts [13]. Similarly, kicking, pushing, punching, etc. could be marked as suspicious activities which must be identified and detected during their occurrence in a video surveillance system. For instance, Figure 1 illustrates a scenario in which an object is running in a classroom.

[1-2-3-5]Shah Abdul Latif University, Khairpur Mirs
[1]The Benazir Bhutto Shaheed University of Technology and Skill Development Khairpur Mirs
[4]The begum Nusrat Bhutto Women University, Sukkur
[6]Sukkur IBA University, Sukkur.
Country: Pakistan
Email: asasd.soomro31@yahoo.com

As Suspicious Activity Detection (SAD) continues to emerge as a prominent research area, researchers are increasingly exploring novel techniques to improve activity detection in different domains. However, the development of effective SAD models is often limited by the availability of high-quality datasets. This article presents a research study on the data collection and preparation stage for SAD in an academic environment. The study focuses on the following tasks:

- Collection of a real-world dataset from academia, named CampusWatch.
- Preparation and pre-processing of the CampusWatch dataset for Human Activity Recognition (HAR).

The organization of the rest of the paper is as follows. Section 2 offers an overview of available datasets for HAR. Section 3 gives a thorough explanation of the proposed dataset, as well as the data collection process along with the characteristics of collected data, challenges faced during data collection, and contribution of collected data. Section 4 presents and discusses the data preparation steps and techniques; while finally, Section 5 contains the conclusions and future work lines.

Figure 1: Running Scenario

DATASETS

This section provides a brief overview of available datasets used for SAD.

a) UCF

The UCF Crime Dataset is a sizable dataset made up of 1,900 films of various crime scenes that were recorded by security cameras [14][15]. The dataset includes 13 categories of criminal behavior, including stealing, violence, and vandalism. There are a variety of resolutions for the videos in the collection, but 640x480 and 720x480 pixels are the most common standard definition (SD) formats. The lengths of the video clips in the dataset, which depict various stages of suspicious activity, often range from a few seconds to several minutes. The UCF Crime Dataset was compiled using footage from openly accessible security cameras, imitating real-world surveillance circumstances to research the detection of suspicious behavior. Researchers who are creating algorithms for identifying and categorizing illegal activity in surveillance footage can benefit greatly from the collection.

b) THUMOS

The THUMOS dataset contains approximately 20,000 movies spanning various activities such as sports, daily routines, and human actions [16]. The intervals between these films range from a few seconds to several minutes. Videos are available in both regular and high-definition quality. The dataset has a variety of challenges such as occlusion, blurred background, moving cameras, occlusion with barriers, variation in size and appearance of subjects, etc., thus the dataset is extremely useful for testing activity detection.

c) HMDB5

HMD20151 is a large-scale video dataset and has become popular in the application of HAR. It encompasses over 2000 video clips in 51 action categories under sport, traffic, and other numerous actions [17] [18]. The collection contains a diverse set of events, perspectives, and actors, as well as movies of varied lengths and quality, ranging from standard definition (SD) to high definition (HD). It can be used to assess the performance of activity detection systems in the real world since it contains such complexities including complex location, change in camera direction and other forms of obstruction.

d) Kinetics

The Kinetics dataset is a large-scale video collection, that consists of individuals performing activities collected from different contexts. It contains 600 action classes with a minimum of 600 video clips per class [19]. These clips range from a few seconds to about several minutes and in quality from the standard definition (SD) to the high definition (HD). This dataset is a valuable resource for researchers working on action recognition systems. Its size, variety of movements, and video quality make it an important benchmark for training and testing machine-learning models in human activity recognition.

e) ShanghaiTech

The ShanghaiTech dataset is a collection of vehicle-annotated videos and images of complex urban scenes. There are 1,198 images and 336 videos of which 330,165 individuals are annotated. The media in the dataset are of high definition/standard definition and are obtained from surveillance videos taken at different parts of Shanghai, China. It provides a befitting solution for using the advanced algorithms of crowd analysis and a host of other applications in computer vision. Further cross-tabulation analysis of the data collected is presented in Table 1 which highlights the dataset features.

Table 1. Characteristics Of Datasets

Data set	Number of Classes	Number of videos/Photos	Resolution	Video Clip Duration	Data collection source
UCF Crime Data set	13 classes	approximately 1,900 videos	640x480 or 720x480 pixels	few seconds to several minutes	camera feeds, simulating real-world surveillance scenarios
THUMOS	101 action classes	approximately 20,000 videos	720p (1280x720 pixels) and 1080p (1920x1080 pixels)	few seconds to several minutes	online video sharing platforms, sports events, and movies
HMDB51	51 action classes	approximately 7000 videos	720p (1280x720 pixels) and 1080p	few seconds to several minutes	online video sharing platforms, movies, and web videos
Kinetics	400 action classes	approximately 306,245 videos	720p (1280x720 pixels) and 1080p	a few seconds to a couple of minutes	online video platforms and other public repositories
Shanghai Tech	focuses on crowd counting and does not have distinct classes	1198 images	720p (1280x720 pixels) and 1080p (1920x1080 pixels)	a few seconds and others spanning several minutes	surveillance cameras and recording crowd scenes

While the available datasets discussed in this article provide valuable insights into human suspicious activity detection, there remains a need for a more realistic and comprehensive dataset that accurately represents the complexities of real-world scenarios. Figure 2 showcases a selection of frames extracted from the datasets discussed in this section, offering a visual representation of the diverse range of activities captured within these datasets.

PROPOSED DATASET

The proposed Dataset-I, collected for this research, aims to identify suspicious behavior in individuals by using footage from diverse indoor and outdoor settings, captured by smartphone cameras in academic environments. This dataset focuses on activities linked to suspicious behaviors, encompassing 23,041 frames categorized into ten classes: "Kicking," "Punching," "Running," "Normal," "Pushing," "Smoking," "Throwing," "Jumping," "Falling," and "Talking." This resource provides researchers with extensive real-time video footage. Figure 3 displays various activity frames from the collected dataset. Further, Table 2 provides detailed breakdown of the different camera makes and models used for data collection.

Table 2. Data Collection Equipment Overview

Serial No.	Camera Type	Camera Model	Camera Resolution	Frame Rate	Lens
1	Smartphone Camera	Samsung A31s	1080	30 fps	26 mm
2	Smartphone Camera	OPPO Reno 3	4080/1080	30/60 fps	26/52 mm
3	Smartphone Camera	iPhone 12 Pro Max	4080/1080	24/30/60 fps	6/65/13 mm

Figure 2: Sample Images of Various Activity

Characteristics of the Dataset CampusWatch

The dataset CampusWatch provides several distinct characteristics that make it novel for enhancing the identification and classification of human suspicious actions. The key characteristics and specifics of the dataset are as follows:

i. ***Data Collection Method:*** Smartphone cameras were utilized to record video footage in indoor and outdoor settings. This method of data gathering is flexible and convenient, allowing for the capturing of a wide range of ambiguous actions.

Figure 3: Sample Images of Each Activity of CampusWatch.

ii. ***Video Quality and Frame Rate:*** The dataset consists both in Standard Definition (SD) and High Definition (HD) quality video recordings.

iii. ***Video Duration:*** Each video is recorded at minimum 30 frames per second to ensure accurate analysis and recognition of subtle and ambiguous behaviors.

iv. ***Data Sources:*** Video recordings were collected in real-time from academia in both indoor and outdoor environments, while objects were performed various activities. The diverse environment contributes to the richness of the dataset, capturing a wide range of scenarios and contexts where suspicious activities may occur.

v. ***Activity Classes:*** The dataset contains a set of nine suspicious activity classes, each of them corresponds to the certain human behavior that should be recognized. These categories consist of kicking, punching, running, pushing, smoking, throwing, jumping, colliding and talking. Furthermore, all the activities except these nine will be classified as normal activity which is marked as the tenth activity class.

vi. ***Sample Size:*** Each suspicious activity class is represented by enough video frames, ensuring a balanced distribution across categories. This allows for robust analysis, training, and testing of suspicious activity recognition models, with adequate representation of each behavior. Moreover, Table 2 provides a detail breakdown of samples for each activity.

vii. ***Challenges for Activity Recognition:*** CampusWatch have several issues to test the capabilities of activity detection algorithms. Figure 4 shows the challenges in dataset like differences in camera angle, lighting conditions, occlusion objects, and background clutter. Further, in section 3.2 challenges have been discussed in detail.

Figure 4: Occlusion objects in running scenario

Finally, the CampusWatch dataset can be considered a valuable resource for HAR as it contains real-time activity recordings from academia. Its distinguishing properties, including high-quality video recordings, different data sources, well-defined suspicious activity classifications, and a large volume of video clips provide a great opportunity for the optimization, evaluation, and improvement of security surveillance algorithms and models. This dataset also serves as a solid foundation for future studies on suspicious activity detection and surveillance system effectiveness.

Challenges

i. ***Privacy Concerns:*** Ensuring compliance with privacy regulations and obtaining consent from individuals who appeared in videos was a significant challenge. In this regard, permission was asked by every individual before recording the video.

ii. ***Ethical Considerations:*** The research team followed ethical standards to collect data that upholds individual rights and academic integrity.

iii. ***Environmental Variations:*** During video recordings, numerous environmental conditions like high or low luminosity, occlusion objects, and high background clutter mimic surveillance occur. Henceforth, these environmental variations were managed through meticulous planning to ensure the quality and consistency of the data.

DATA PREPARATION

I. *Video Frames Annotation*

An automatic video annotation approach is described in [20], has been employed in this study. The approach divides the video into individual frames and assigns an action label, such as kicking, punching, running, pushing, smoking, throwing, jumping, colliding, talking, and normal. The output frames are then stored in the database to feed the model precise and accurate data for accurate activity classification and identification.

II. *Data Cleaning*

A thorough cleaning procedure was applied to the dataset to eliminate any corrupted or missing video segments. To guarantee the accuracy and consistency of the data, videos containing technical problems—such as excessive blurriness or slow frame rates—were also excluded.

a. ***Frame Extraction:*** The first fundamental step in the data preparation process is frame extraction, where each video is divided into individual frames. Given a video $V = \{v_i\}^N$ i=1 with N clips (individual frames within the video) the annotated video-level label Y ∈ {1, 0} indicates whether a suspicious activity exists in this video. We take a video V as a bag and clips $v_i$in the video as instances. Specifically, a negative bag (i.e. Y = 0) marked as $B^n = \{v_i^n\}^N$ $i = 1$ has no suspicious instance, while a positive bag (i.e. Y = 1) denoted as $B^a = \{v_i^a\}^N$ $i = 1$ has at least one. Further, **Error! Reference source not found.** depicts Bag-of-Instance classification as negative or positive.

Figure 5: Bag-of-Instance

b. ***Noise Reduction:*** To enhance the clarity and quality of the video frames, denoising filters—such as Gaussian and median filters—were applied to reduce noise, artifacts, and visual disruptions. This process is illustrated in Figure 6 and was crucial for improving the accuracy of activity detection.

Figure 6. Applying Gaussian filter to enhance image quality, showcasing noise reduction and edge refinement.

c. ***Data Validation:*** The cleaned dataset was subjected to extensive validation to ensure its integrity and alignment with research objectives. This process involved visual inspections, statistical analyses, and spot checks to confirm that the dataset met the required quality standards.

Frame Preprocessing

To ensure that frames were consistent and to ease the process of training the chosen models, all frames were resized. Several noise reduction techniques discussed in Section 4.2 were used as a means of enhancing the quality of the frames in the video clip a factor that improved the likelihood of detecting suspicious activities.

a. ***Frame Resizing and Standardization:*** To ensure uniformity across the dataset, all video frames were resized to a standard dimension., all the frames in the videos were resized to a fixed resolution. This approach reduces the computational complexity of the model's training and evaluation. To calculate the pixel values of the resized image as depicted in **Error! Reference source not found.** bilinear interpolation was applied. The interpolation can be stated in sigma notation as follows:

output image (x, y) = $\sum_{i=1}^{n}(w_i * Image\ (x_i, y_i))$

Where N represents the number of neighboring pixels considered, (x_i, y_i) denotes the coordinates of the i^{th} neighboring pixel, and w_i is the weight assigned to each neighboring pixel.

b. ***Temporal Segmentation:*** To capture the temporal dynamics of suspicious actions, each video clip was segmented into smaller time frames. This segmentation enables the model to better analyze the sequential nature of actions and improve detection accuracy.

Figure 7. (a) A Complete Frame with Dimensions 1280x720, (b) Resized Frame with Dimensions 224x224

Data Augmentation

To increase the diversity of the dataset and improve the model's generalization capability, data augmentation techniques such as random cropping, rotation, and flipping were employed. This approach ensures the model is robust when encountering novel situations.

Finally, the dataset comprises a comprehensive collection of 23,041 video frames, covering nine distinct suspicious activity types, with all other activities classified under a tenth category, "Normal." Detailed information about the number of frames for each activity class is provided in Table 3, demonstrating the dataset's capacity to support rigorous analysis and model training for both suspicious activity detection and the identification of normal behaviors.

Table 3. Breakdown of Activity Classes and Video Frame Counts

S. No	Category	Video Frame Count
1	Kicking	2,744
2	Punching	2,150
3	Running	2,110
4	Normal	2,700
5	Pushing	1587
6	Smoking	2,090
7	Throwing	1,950

8	Jumping	2,700
9	Colliding	2,300
10	Talking	2,710
Total =		**23041**

CONCLUSION AND FUTURE WORK

In conclusion, this research introduced the CampusWatch dataset, specifically designed for suspicious activity detection (SAD) in academic environments. The article detailed the steps involved in preparing this dataset, including data collection, video annotation, noise reduction, and preprocessing. The dataset consists of 23,041 frames, covering nine distinct suspicious activities such as kicking, punching, and running, along with a "Normal" activity class. This comprehensive dataset provides a strong foundation for developing, evaluating, and benchmarking models aimed at detecting suspicious behaviors, offering a real-world scenario-based benchmark for SAD models.

Expanding the CampusWatch dataset to include more activity classes and diverse environments will be part of future research. Moreover, incorporating advanced deep learning techniques for real-time activity detection and anomaly recognition could significantly boost the performance and scalability of surveillance systems in academic settings and beyond.

Declaration of Competing Interest:

The authors declare no competing interests associated with this work.

REFERENCES

[1] Z. Qin et al., “Fusing Higher-Order Features in Graph Neural Networks for Skeleton-Based Action Recognition,” IEEE Trans. Neural Networks Learn. Syst., vol. 35, no. 4, pp. 4783–4797, Apr. 2024, doi: 10.1109/TNNLS.2022.3201518.

[2] P. Kumar, S. Chauhan, and L. K. Awasthi, “Human Activity Recognition (HAR) Using Deep Learning: Review, Methodologies, Progress and Future Research Directions,” Arch. Comput. Methods Eng., vol. 31, no. 1, pp. 179–219, Jan. 2024, doi: 10.1007/S11831-023-09986-X/METRICS.

[3] S. C.A. Thomopoulos, “Risk Assessment and Automated Anomaly Detection Using a Deep Learning Architecture,” in Deep Learning Applications, 2021. doi: 10.5772/intechopen.96209.

[4] G. Vallathan, A. John, C. Thirumalai, S. K. Mohan, G. Srivastava, and J. C. W. Lin, “Suspicious activity detection using deep learning in secure assisted living IoT environments,” J. Supercomput., vol. 77, no. 4, pp. 3242–3260, Apr. 2021, doi: 10.1007/s11227-020-03387-8.

[5] C. V. Amrutha, C. Jyotsna, and J. Amudha, “Deep Learning Approach for Suspicious Activity Detection from Surveillance Video,” in 2nd International Conference on Innovative Mechanisms for Industry Applications, ICIMIA 2020 - Conference Proceedings, 2020, pp. 335–339. doi: 10.1109/ICIMIA48430.2020.9074920.

[6] F. G. Ibrahim Salem, R. Hassanpour, A. A. Ahmed, and A. Douma, “Detection of Suspicious Activities of Human from Surveillance Videos,” in 2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering, MI-STA 2021 - Proceedings, 2021, pp. 794–801. doi: 10.1109/MI-STA52233.2021.9464477.

[7] A. Gorave, S. Misra, O. Padir, A. Patil, and K. Ladole, “Suspicious Activity Detection Using Live Video Analysis,” 2020, pp. 203–214. doi: 10.1007/978-981-15-0790-8_21.

[8] S. Sharma and V. Dhama, “Abnormal Human Behavior Detection in Video Using Suspicious Object Detection,” in Lecture Notes in Electrical Engineering, Springer, 2020, pp. 379–388. doi: 10.1007/978-981-15-1420-3_39.

[9] P. K. Roy and H. Om, “Suspicious and violent activity detection of humans using HOG features and SVM classifier in surveillance videos,” in Studies in Computational Intelligence, Springer Verlag, 2018, pp. 277–294. doi: 10.1007/978-3-319-63754-9_13.

[10] P. B. Divya, S. Shalini, R. Deepa, and B. S. Reddy, “Inspection of suspicious human activity in the crowdsourced areas captured in surveillance cameras,” Int. Res. J. Eng. Technol., 2017, Accessed: Jun. 06, 2022. [Online]. Available: https://www.academia.edu/download/55458847/IRJET-V4I12153.pdf

[11] N. Zhu et al., “Bridging e-Health and the Internet of Things: The SPHERE Project,” IEEE Intell. Syst., vol. 30, no. 4, pp. 39–46, 2015, doi: 10.1109/MIS.2015.57.

[12] K. J. Cheoi, “Temporal Saliency-Based Suspicious Behavior Pattern Detection,” Appl. Sci. 2020, Vol. 10, Page 1020, vol. 10, no. 3, p. 1020, Feb. 2020, doi: 10.3390/APP10031020.

[13] O. M. Rajpurkar, S. S. Kamble, J. P. Nandagiri, and A. V Nimkar, “Alert Generation on Detection of Suspicious Activity Using Transfer Learning,” in 2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020, 2020. doi: 10.1109/ICCCNT49239.2020.9225263.

[14] J. Abellan-abenza, “Classifying Behaviours in Videos with Recurrent Neural Networks,” vol. 7, no. 4, pp. 1–14, 2017.

[15] B. Solmaz, B. E. Moore, and M. Shah, “Identifying behaviors in crowd scenes using stability analysis for dynamical systems,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 10, pp. 2064–2070, 2012, doi: 10.1109/TPAMI.2012.123.

[16] H. Idrees et al., “The THUMOS challenge on action recognition for videos ‘in the wild,’” Comput. Vis. Image Underst., vol. 155, pp. 1–23, 2017, doi: 10.1016/j.cviu.2016.10.018.

[17] H. Kuehne, “MIT Open Access Articles HMDB : A Large Video Database for Human Motion Recognition HMDB : A Large Video Database for Human Motion Recognition,” 2012.

[18] F. De la Torre, J. Hodgins, J. Montano, S. Valcarcel, and J. Macey, “Guide to the Carnegie Mellon University multimodal activity (CMU-MMAC) database,” no. April, 2009.

[19] W. Kay et al., “The Kinetics Human Action Video Dataset,” 2017, [Online]. Available: http://arxiv.org/abs/1705.06950

[20] S. Bianco, G. Ciocca, P. Napoletano, and R. Schettini, “An interactive tool for manual, semi-automatic and automatic video annotation,” Comput. Vis. Image Underst., vol. 131, pp. 88–99, 2015, doi: 10.1016/j.cviu.2014.06.015.