Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

A SURVY of video datasets for anomaly detection and human activity recognition

Vean fuaad abd al-rasheed [a], Dr. Narjis Mezaal Shati[b]

[a]*Department of Computer Science, College of Sciences, Mustansiriyah University, Baghdad, Iraq, vean.fuaad@uomustansiriyah.edu.iq*

[b]*Department of Computer Science, College of Sciences, Mustansiriyah University, Baghdad, Iraq, dr.narjis.m.sh@uomustansiriyah.edu.iq*

ARTICLE INFO

Article history:
Received: 24 /03/2022
Revised form: 06 /05/2022
Accepted : 15 /05/2022
Available online: 04 /06/2022

Keywords:

Anomaly detection,

videoDataset,

human activity recognition (HAR),

Surveillance system,

scene-type,

Benchmark Datasets.

ABSTRACT

The computer vision researchers concentrated on the automation of the surveillance system. Many datasets suited for diverse applications have been proposed by research in this subject. in a number of different domains of application Human action recognition may be used effectively, which has the potential to improve many facets of daily life.These contain, among other things, preventing violent act and detecting crimes such as murder, stealing, and property damage, as well as predicting pedestrian activity in traffic. And this study addresses the properties of public datasets used for Human Action Recognition. Researchers develop these distinct anomaly datasets as a result of the availability of security cameras installed in various areas. For researchers to comprehend and develop in this field, It is required to review anomaly detection video datasets. Therefor This study presents a survey for video surveillance activity recognition.

MSC..

https://doi.org/10.29304/jqcm.2022.14.2.931

1. Introduction

Video surveillance has emerged as a useful method for preventing security issues. By developing more accurate and powerful algorithms an artificial surveillance system has been exploring with video surveillance data in order to improve image processing tasks for object detection and tracking, and also human activity recognition. [1]. Surveillance tasks can be conducted in real-time by analyzing video, or footage can be archived and reviewed later as needed. In gaming video surveillance, video surveillance may also be utilized to detect illegal activities and evaluate player behavior [2]. There are several survey and review publications on human activity recognition; these papers should be highlighted as in [1][2][4][6][10][21].In recent years, a growing number of video datasets dedicated to

∗Corresponding author: *Vean f. abd al-rasheed*

Email addresses: *vean.fuaad@uomustansiriyah.edu.iq*

Communicated by 'sub etitor': *Dr. Narjis Mezaal Shati*

human action and activity detection are being developed. [3], and the action recognition problem may be Action analysis and action representation illustrate this. Antennae such as, radar, RGB, range, and wearable sensors are used to capture these actions. A manual HAR task, like as recognizing anomalous activity in a video clip, takes a long time. As human activities are required throughout multi-camera perspectives, such jobs are costly and difficult. [2] Various learning approaches are used in recognition operations. All of these algorithms do some of the same tasks, feature extraction, action recognition, action learning, action classification, and action segmentation are some examples. [4]

2. *Basic Definitions*

To conclude, the following are the primary components of this survey [5]:

- Scenes: A scene is a dynamic environment that contains people and cameras. Each camera can only see a portion of the scene. Even when only one camera is used, the scene is frequently greater than what the camera can record. In a lingering experiment, for example, a subject may exit the camera's field of vision but will not depart the scene. Each experiment is linked to a certain setting .Second point

- View: it is described as a method of looking at or observing something through the use of a recording device. It is classified as either single or multi-view. Only one recording device is utilized in a single view, whereas several recording devices are used in a multi-view.[6]

- A frame: is a single image in a video.

- A trace is a collection of blob characteristics (a blob is a continuous group of foreground pixels) from several frames pertaining to the same object.

- Characters to Track: The group of characters to track includes persons of various ethnicities and genders. Currently, there are 4 topics available, however, more may be introduced in the future.

- *Crowds*: The amount of stationary and moving persons in a scene typically has an influence on the capacity to follow the target topic; it would be able to evaluate whether the presence of other people has a detrimental effect on the tracking, or whether the inaccuracies are due to other reasons.

- *Interactions*: It is defined as one-on-one communication between two or more people or objects. Human–human interactions and human–object interactions are the two kinds of interactions. The first kind includes fights, high-fives, and other interactions. The second might include person preparing tea, answering phones, and the crowd. [6]

3. Human Activity recognition

The method of detecting an activity based on input from a certain sensor that records certain motions of an individual or objects is known as human activity recognition. Human activity detection systems based on video use footage acquired by cameras, drones, or radars to generate their output. It is highly difficult and complex to get and process these inputs. Various ways may be established in this context by taking into account various strategies. The applied technique, in general, comprises of three basic processes: data capture, pre-processing and feature extraction, and activity recognition. [4][7] This review is mostly focused on video datasets, however there are certain databases designed for extremely specialized action identification, such as crowd behavior, detection of human falls, gait analysis, or posture and gesture recognition. Heterogeneous and Specific Actions are two of them that are connected to the type of actions offered by the dataset. A third group, others, is characterized by the precise approaches used to capture the actions, as illustrated in fig (1) [3].

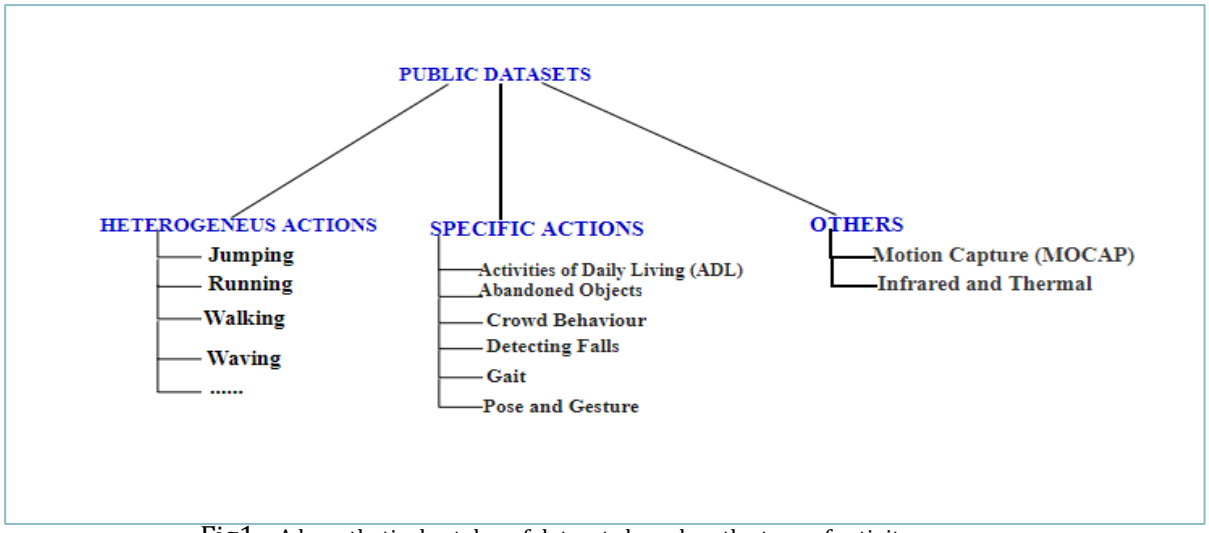

Fig1 - A hypothetical catalog of datasets based on the type of activity.

4. Anomaly video dataset

The job of detecting anomaly events is separated into global and local anomaly [2] there are presently more than 30 publicly accessible video datasets being utilized for anomaly identification. Initial datasets are made up of simple events and scenarios with a limited number of abnormalities. Because the anomalies are dramatized by a group of actors, the usual flow of events is disrupted. They are also extremely short. The existence of a few unusual objects or events in unimodal background events was seen as an abnormality in these datasets. Web (panic-escape and mob fighting are recognized as anomalies), Canoe (a boat happening once in the scene is treated as an anomaly), UMN (few individuals acting for quick evacuation is viewed as abnormality here) [8], Subway entrance/exit are some instances of such datasets (movement in the wrong direction regarded as an anomaly). [9], fig (2) shows type of anomaly datasets Categorization based on scene-type. Fig (3). Illustrate examples from several anomaly video dataset.

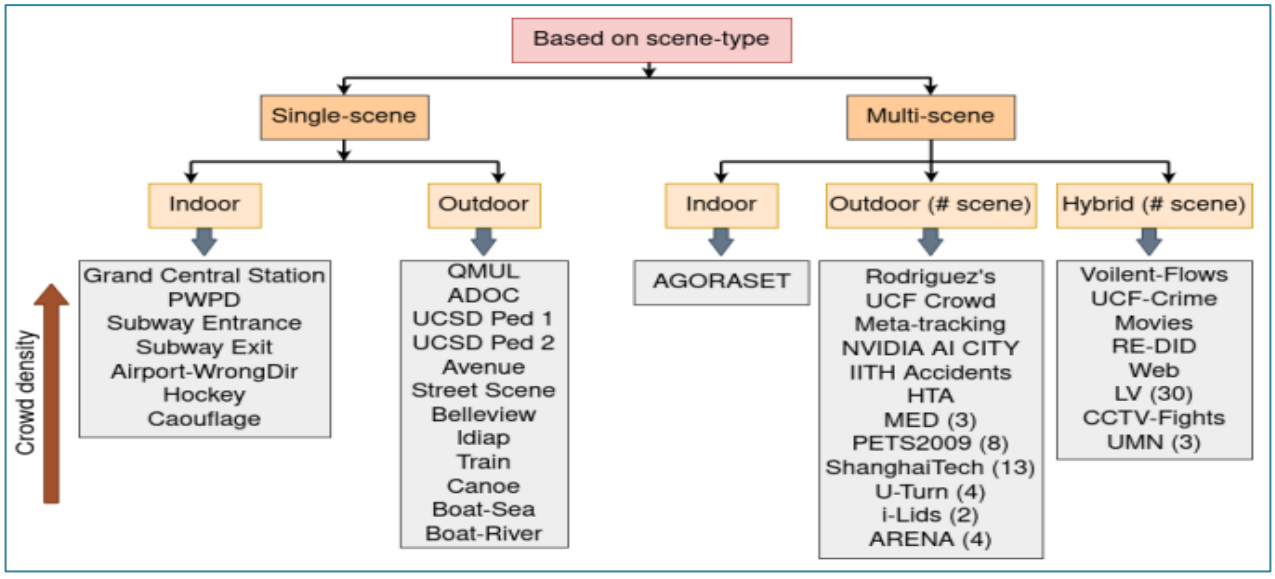

Fig2 -. Categorization of video anomaly datasets based on scene-type

Fig3 - Sample images from various anomaly video dataset [10]

5. Benchmark HAR Datasets

To put it simply, creating benchmark datasets for visual identification has always been a tough and time-consuming operation [11]. Benchmarking is one method of determining the optimal performance. When compared to modern datasets, should be labeled, verified, and demonstrated to be accurate in detecting human activities the benchmark video datasets for action recognition are properly organized. On these well-prepared datasets, several action recognition systems performed admirably and with greater accuracy. [12]. As a result, these datasets are referred to as benchmark datasets. To present, there are at least 26 human action video datasets accessible, however, the most typical benchmark datasets used by computer-vision researchers are HMDB51 and UCF101. Kinetics datasets have recently become attractive choices for researchers, since it presented a huge number of event classes by using public YouTube films. The micro-videos dataset examines an open-world language for video interpretation using social media videos. The VLOG dataset records people's daily actions in their natural spatiotemporal setting. While the "something something" dataset and Charades relied on crowd-sourced employees to capture video datasets, [13], Because of crowded backdrops, view-variation, occlusion, intra-class similarity, and application scenarios, the proposed algorithms should be able to tackle the issues provided by the videos dataset. We may separate the existing technique to activity recognition into two parts: Figure. (4) Displays handmade solutions, learnt feature representation, and comprehensive categorization taxonomy. [6].

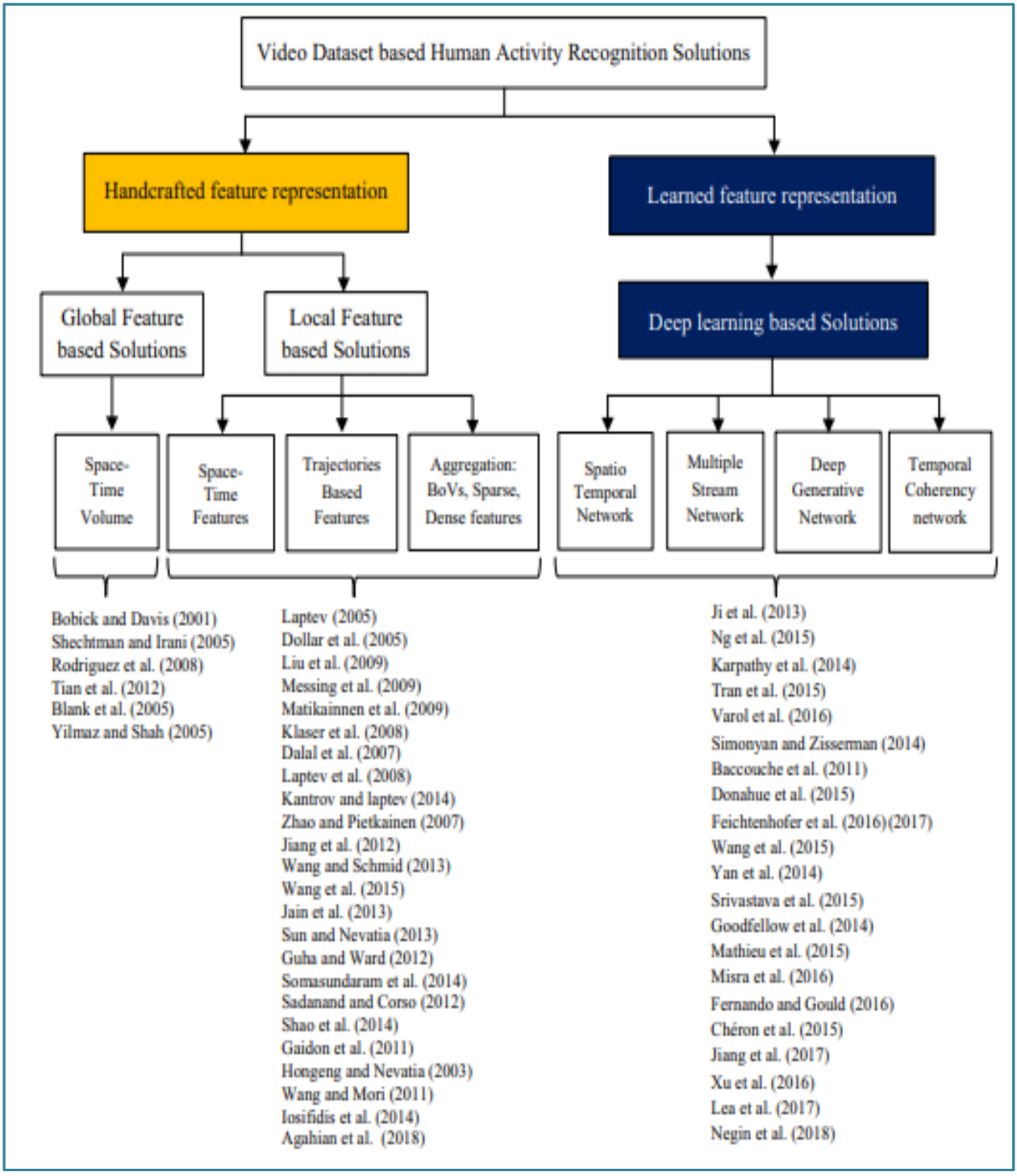

Fig3 - Human Activity Recognition Solutions Based on Video Datasets: A Taxonomy

6. Datasets

Datasets are essential for evaluating different algorithms used to achieve a certain goal. The evaluation of task-specific algorithms is dependent on factors unique to each dataset. [2] The most frequent datasets will be discussed, and they are presented in table 1. It displays a summary of the accessible datasets. [14]

Table 1- a summary of the accessible datasets.

Dataset/ YEAR	Scenes	Application Domain	Source	No Of Actors	Types Of Anomaly	No Of Videos
Avenue/(2013) [14]	outdoor	anomaly detection	capture CUHK Campus Avenue	1-3	8-12	37
CAVIAR/(2004)[3][9]	indoor	Target detectors clustering of trajectories multi agent activity recognition	Recorded in a shopping center in Lisbon.	24	<5	28
(UCF)crime/ (2018) [16,17]	In/Outdoors	anomaly detection action classification	video surveillance cameras	Unspecified	13	950(N)+950(A)
(UCSD)/ (2010) [1][9]	Outdoors	anomaly detection crowd density estimation action classification	stationary camera	Unspecified	**Peds1:** 5 **Peds2 :5**	**Peds1:** 34(N)+36(AN) **Peds2 :** 16(N)+14(AN)
KTH/(2004) [21]	In/Outdoors	Action recognition feature extraction	Recorded videos	25	6-10	600
movie-fight and	In/Outdoors	violence detection	Recorded videos	Unspecified	1	200

Hockey-fight/(2011)[9]	indoor				1	1000
Subway Entrance and Subway Exit/(2008)[9]	Indoor indoor	Anomaly detection abnormal behavior modeling	Recorded videos subway platform of the Daegu Metro	2-4 2-4	<3 <3	1 1
UCF 50 UCF101/(2010)[21]	In/Outdoors	Event, action recognition	Videos from web as YouTube	Unspecified	>20	13,320 ucf101
VIRAT/(2011) [3][10]	Outdoor	Annotation and tracking human-vehicle interaction recognition action recognition	Drone, surveillance	Unspecified	11-15	550 videos.
Weizmann standard/(2001_2005)	Outdoor	Action recognition person identity Temporal segmentation	Recorded videos	9	1-5	90

This section offers several common human action detection video datasets that are widely used in computer vision research. As an illustration: **The Moments in Time Dataset** contains about a million 3-second films that correspond to 339 different verbs. Over 1,000 videos are related to each word. Importantly, the dataset is designed to have, and will grow toward, a very huge collection from both inter-class and intra-class variety that captures a dynamic action at multiple levels of analysis, such as "opening" eyes, lips, doors, curtains, or even flower petals.. [13]

The Hollywood dataset 2008, Human Actions datasets: consists of video snippets derived from Hollywood films Professional actors execute twelve different action categories, resulting in more realistic scenes than previous basic action datasets. [11], Each sample is labeled with one or more of Answer Phone, Get Out of Car, Hand Shake, Sit Up ,

Kiss, Sit Down, Hug Person, and Make A stand are the 8 different action classes. The dataset is divided into two sections Includes a test set of 20 videos and 2 training sets of 12 videos that varied from the test set. The so-called automated training set, which comprises of 233 video clips with around 60% accurate classifications, is generated via automated script-based activity labeling. [3]

EduNet, A dataset including 20 distinct action types from a classroom activity monitoring dataset. Each class features at least 200 clips culled from YouTube videos and taken in a actual classroom atmosphere of 1 to 12 regular schools with a Nikon DSLR D5600 camera from Nikon Corporation in Tokyo, Japan. There were 7851 video clips totaling nearly 12 hours in length. EduNet, based on what we know, is the biggest and maybe the most widely available dataset in the classroom environment, comprising both student and teacher behaviors. [12]

Charades v1.0 Analysis in 15 scenarios, 40 objects and 30 actions are combined to form Charades. This narrow vocabulary, paired with open-ended writing, results in a dataset that covers a wide range of topics. Furthermore, these combinations constitute action classes that may be standardly benchmarked. [18].

UCF Sports Action Dataset: comprises of multiple activities gathered in 2008 from various athletic events broadcast on broadcast television networks such as the BBC and ESPN. The video samples were sourced from several stock footage websites, such BBC Motion Gallery and Getty Images. [3]. the dataset is made up of a natural pool of behaviors that appear in a variety of scenarios and views. This dataset contains 16 diving videos, 25 golf swinging videos, 25 kick videos, 15 pulling videos, 35 hitting videos, 14 horse-riding videos, 15 jogging videos, 15 skateboarding videos, and 35 walk videos (22 videos).There is a basic ground truth provided in the activity notes. Action recognition is utilized in this dataset. The dataset has been utilized in a variety of applications, including action identification, reallocation, and saliency detection. [19]

The Kinetics-400 dataset [20] it differs from current datasets in that it focuses on human actions rather than activities or occurrences. Kinetics is double the size of prior benchmark datasets. Each class has around 400 clips, including HMDB-51 and UCF-101. Kinetics is distinguished by its categorization, which includes I individual behavior (singular)—punching, laughing, sipping, pulling, and so on; (ii) person–person actions—kissing, holding hands, hugging, and so on; and (iii) person–object actions—mowing the lawn, dishwashing, opening gifts, mopping, and so on. Kinetics-600 is a 600-classification augmentation of the Kinetics human activity dataset. [12]

7. Related Work

TABLE 2- Summary of Some techniques for Abnormal Activity Recognition & detection

8. Discussion

Table 1 summarizes the essentially key features of the commonly used datasets, that provide an easy search the name of dataset, the type of scenes, the total number of actors, Application Domain, Source of dataset, the number of anomaly action,
. When it is known, the total number of actors or ordinary people engaged in all the videos stored in a dataset is provided the datasets recorded in controlled conditions normally use indoor scenarios. However, like can be seen in the 2d column, there are also datasets obtained in outdoor scenarios or
In both. Obviously outdoor scenario datasets are more challenging than indoor ones, variations Scenes test the ability of each algorithm to identify actions
Independent of the background, appearance of the actors, and the scale of the actors. The sixth column shows the number of anomaly type recording detected in each dataset. Other important information
Is the total number of videos (7th column). We also display the year of creation or first publication that describes the dataset, the amount of publications that reference the dataset has been used for benchmarking, with older datasets being more referred, whereas newer datasets being less mentioned because they are younger yet often give higher ground truth information. Older datasets have a higher likelihood of being utilized, as seen by the most prolific datasets, KTH, WEIZMANN, and CAVIAR, while others are still seldom cited in articles [6]. another type of information that can be useful, such as mentioning if this data set is a subset, or a regression of an older type of data set, as is the case in of BEHAVE is some sub-datasets with multi-view of CAVIAR, While Ucf50 is an extension of YouTube Action Dataset, ucf101 is an advanced version of UCF 50 dataset other type of datasets that are authentic repositories of large quantities of video, containing thousands of hours of footage. We are speaking of VIRAT[3] There are several important aspects to investigate when considering a comparison of existing video data sets such as: scene Density, Example anomalies/event, Resolution, Camera motion, Total No. of Frames, number name of Reference[9,21].
We Summaries Some Related Works briefly In Table 2. Which Mentioned Datasets Were Utilized. Nazir et al.

References	model	Data set	Result
Wang et al. (2016)	TSN	HMDB51 UCF101	Accuracy: 69.4
Nazir et al. (2018)	SVM	KTH UCFSports Hollywood2	Average Accuracy: 91.8 Average Accuracy: 94
Waqas Sultani et al. (2018)	C3D& MIL ranking loss	UCF Crime dataset	AUC (74.44) without constraints. AUC with constraints (75.41)
Prakhar & Vinod. (2018)	DNN	UCSD PAD1 UCSD PAD2	AUC (74.8% in UCSDPed1) and (80.2% in UCSDPed2)
Li et al. (2019)	Actional graph-based CNN	NTU-RGBD Kinetics	Accuracy: 94.2 Top-5 accuracy: 56.5
Divya Thakur et al.(2019)	CNN MSER-CNN	UCF Crime dataset	ACC 98.36% ACC 95.06%
Jian et al. (2019)	FCN	Sports video	Accuracy: 97.4
jwan Jamal Ali. (2020)	GMM, KNN	Weizmann, KTH	ACC dataset (97%) detection rate (97%) and false alarm rate (4%).
Muthana S. Mahdi and et al. (2021)	VGG16 & LSTM	CAVIAR, KTH, and YouTube scenes	ACC 95.3
Shabana and et al(2021)	CNN&LSTM	Hockey Fight, Surveillance Fight datasets	ACC 96%(Hockey Fight). ACC 81.05% (Surveillance Fight datasets)

(2018), Jwan Jamal Ali. (2020) utilize machine learning technique, and the other works is deep learning.

9. Conclusions

We introduced a fundamental concept and dataset description for complicated activity recognition and anomaly detection in multimedia streams. There are several datasets available for recognizing human actions and activities. There is a growing demand for wide-ranging datasets that represent the varying nature of real-world recognition settings. For cross-dataset validation, robust evaluation approaches are necessary, which will be effective for actual scenario applications. Benchmarking in the area of human activity comprehension because one of the key aims of this survey is to allow for additional development, research, and advancement.

Acknowledgements

The authors would like to thank Mustansiriyah University in Bagdad, Iraq, for their cooperation with this study (http://uomustansiriyah.edu.iq).

References

[1] S. Vishwakarma, A. Agrawal, "A Survey on Activity Recognition and Behavior Understanding in Video Surveillance",springer,2012.
[2] P. Pareek, A. Thakkar, "A survey on video based Human Action Recognition: recent updates, datasets, challenges, and applications", Springer Nature B.V. 2020.
[3] J. Chaquet, E. Carmona, A. Fernandez, "A survey of video datasets for human action and activity recognition", research get, 2013.
[4] T. Özyer, D. Selin, R. Alhajj, "Human action recognition approaches with video datasets—A survey", ELSEVIER, 2021.
[5] D. Gorodnichy, R. Laganiere, D. Macrini, "Video analytics evaluation: survey of datasets, performance metrics and approaches", CANADA, 2014.
[6] T. Singh, D. Vishwakarma, "Video benchmarks of human action datasets: a review", Springer Nature 2018.
[7] I. Kaloskampis, Y. Hicks, and D. Marshall, "Complex activity recognition and anomaly detection in multimedia streams", Published 2014.
[8] R. Mehran, A. Oyama, M. Shah, "Abnormal Crowd Behavior Detection using Social Force Model", IEEE, 2009.
[9] P. Kumari, A. Bedi, and M. Saini, "Multimedia Datasets for Anomaly Detection: A Review", arXiv: 2112.05410v3 [cs.CV] 4 Apr 2022.
[10] N. Patil, P. Biswas, "A Survey of Video Datasets for Anomaly Detection in Automated Surveillance", IEEE©2016.
[11] F. Heilbron, V. Escorcia, B. Ghanem, J. Niebles, "ActivityNet: A Large-Scale Video Benchmark for Human Activity Understanding", IEEE, June 2015.
[12] V. Sharma, M. Gupta, A. Kumar, and D. Mishra, "duNet: A New Video Dataset for Understanding Human Activity in the Classroom Environment", MDPI, 2021.
[13] M. Monfort, A. Andonian, B. Zhou, K. Ramakrishnan, S. Bargal, T. Yan, L. Brown, "Moments in Time, Dataset: one million videos for event understanding", IEEE, 2019.
[14] M. Ali, M. Al-Berry, Z. Taha, "Comparative Study for Anomaly Detection in Crowded Scenes", IJICIS, Vol.21, No.3, 84-94 DOI: 10.21608/ijicis.2021.84588.1112.
[15] M. Roshtkhari, M. Levine, "Online Dominant and Anomalous Behavior Detection in Videos", IEEE, June 2013.
[16] W. Sultani, Chen Chen, "Real-world Anomaly Detection in Surveillance Videos,CVPR" , IEEE, 2018.
[17] S. Vosta, and K. Choong Yow, "CNN-RNN Combined Structure for Real-World Violence Detection in Surveillance Cameras", MDPI, 2022.
[18] G. Sigurdsson, G. Varol, X. Wang, I. Laptev, A. Gupta, "Hollywood in Homes: Crowdsourcing Data Collection for Activity Understanding", Springer International Publishing AG 2016.
[19] Kh. Soomro, and A. Zamir, "Chapter 9 Action Recognition in Realistic Sports Videos", Springer International Publishing Switzerland 2014.
[20] W. Kay and et al, "The Kinetics Human Action Video Dataset", arXiv: 1705.06950v1 [cs.CV] 19 May 2017.
[21] S. Kang,R. Wildes, "Review of Action Recognition and Detection Methods", arXiv:1610.06906v2 [cs.CV] 1 Nov 2016