iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** II **Month of publication:** February 2020

DOI: http://doi.org/10.22214/ijraset.2020.2004

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 8 Issue II Feb 2020- Available at www.ijraset.com

Human Activity Recognition in Video Surveillance – A Survey

Shravani Shirish Urankar[1], Suresh K[2], Shubham Bhat[3], Ranjeet Kumar[4], Madhura J[5], Dr. Kavitha A S[6]
[1 2 3 4]*B. Tech. Student,* [5]*Assistant Professor,* [6]*Associate Professor Dept of Information Science and Engineering, Dayananda Sagar College of Engineering, Karnataka, INDIA*

Abstract: ***The act of detection of the activity of humans has gained traction nowadays and is a hot field for research and development. The ability of human activity detection (HAR) is a key prospect of development in fields of human-computer interface, computer vision and in mass video surveillance in public areas such as train stations, ATM machines, schools and colleges and at traffic signals on streets. HAR offers a cost-effective method for surveillance. Analysis based on methods used for classification for different datasets is presented in this survey paper.***
Keywords: HAR, CNN, SVM, K fold cross-validation, wearable sensor, classification, Smartphone, Accelerometer.

I. INTRODUCTION

Activity recognition for humans is done using various approaches such as deep learning, usage of neural networks or using models such as CNN and SVM. These approaches are used from the fields of computer vision to that of mass surveillance and even in the development of human-computer interfaces. This detection must be carried out as and when it occurs in the real world and is captured by the camera being used to capture this real-world input. From the input obtained by the camera feed connected to the system, we must extract the features of various humans present and obtain the relevant information required to process the activities they may be performing based on actions that we take into consideration during the building of models as per our requirements. We follow a probability-based approach to help distinguish one activity from another. The inputs must be captured over a short interval of time wherein the activity pertaining to the detection was performed. This has historically proved to be a challenging process as many factors constantly change in real-world during the detection process such as camera angles, lighting scenarios and variance with respect to the human action being performed. To construct models that detect activities with credible and relevant accuracy we must build models from the ground up keeping in mind the relevant factors and parameters that help in the activity detection. Various steps are involved in this process which usually includes the pre-processing of the data input, segmenting the relevant data while ignoring the unnecessary data. Further processes involve the extraction of key features for the building of the model that is later used to predict the human activity in real-time. Popular approaches include a learning-based approach wherein we train our model to learn automatically without the help of human intervention. This is a key phase in the construction of the model as if improper features are extracted the further detection steps cannot be carried out appropriately to obtain the desired results. Approaches such as recurrent neural networks and convolutional neural network have shown high promise and offered credible results. This paper has been divided into 4 parts. The first part consists of the introduction of the survey. The second part consists of the Literature review of multiple approached already carried out. The third part details a comparative study of the approached already carried out, while the final part helps establish a result of the approaches and details the various enhancements that can be done in future projects.

II. LITERATURE SURVEY

In this section, we will study the different approaches for human activity recognition based on neural networks.
Erhan BÜLBÜL [1] and Akram Bayat [10] proposed separate support vector machines(SVM) which make the best use of hyper-dimensional planes in order to separate examples. Although SVN can be used with and without supervision, it is usually quicker and more successful to use supervised SVN. When supervising SVM with a cubic polynomial kernel used to identify tuples in the dataset, a high success rate of 99.4 per cent was achieved.
Zhenguo Shi [2] explored an approach that included the study of recurrent neural network (RNN) in which node-to-node links shape a directed graph along a timing chain. This type of neural network is usually used in examples including timing series. A model is trained using this approach in order to obtain the temporary dynamic behaviour. It has a segment called a memory segment which mainly processes variable length of input sequence. In order to transform and extract the inherent features from the input data collected from CSI-HAS, Deep Learning networks are used. To further reduce and optimize the feature, a Sparse auto-encoder

(SAE) network is utilised. One of the demerit of using this is that the sensing performance of SAE is susceptible to input quality. To overcome this problem an approach of Recurrent Neural network based on the concept of Long short term memory (LSTM) is used by taking a CSI packet as a raw input. Adding to it , LSTM-RNN based model can also be used to extract features at the later part of the entire process. The classifier here being used is the Softmax Regression Algorithm. The coefficients of the trained network is used as an equipment for the further training process.

Anjana Wijekoon [3] and Wenchao Xu [11] explored various approaches out of which convolution neural network (CNN, or ConvNet) is commonly used to process visual image. They are commonly referred by move invariant or Space Invariant Artificial Neural Networks, that mainly concentrate on invariance characteristics of translation and common design. To equalize the performance of every secret sheet an approach of Batch Normalisation is applied over the dataset. They define three different segments of the classifier and sensors and each produce an output vector of 100 size. DCT-1D input: ACT and ACW. Changes with respect to 1 Dimension and further into the max-pooling hidden layer where the channel count is 1 and input feature is 180 bits long. Raw-1D: 1-dimensional convolutions and the levels of pooling. ACT and ACW uses a window frame of size 500 (window frames per second) raw data with 3 channels (3 accelerometer axes). The vector is formed by flattening the frame and a timing window is obtained by processing the frames. Single-dimensional display feature-length frame width vector for frame width height/frame height/frame per second for 1 path.–2D: This approaches mainly uses two-dimensional convolution and total levels of pooling. 1 channel in a timing window along with a 2 dimensional vector are obtained after extracting the information via PM as well as DC.

Akbar Dehghani [4] proposed a model that uses k-fold Cross Validation as the primary methodology for testing model output and adjusting hyperparameters. This means measurements are Random and Identically Distributed, i.e. data points are taken from the same source separately. Nonetheless, samples belonging to the same topic are likely to be interrelated in HAR datasets because of the underlying physical, biological and demographic factors. Additionally, there is often a temporal correlation between a subject's samples due, for example, to fatigue, training, experience. Consequently, k-fold CV could overestimate the performance of the classification by relying on correlations within subjects.

Stefanie Anna Baby [5] proposed a model which includes a dynamic vision sensor, it processes only the foreground objects that being the humans captured the camera while completely ignoring the unnecessary processing of backgrounds thus decreasing processing requirements and greatly enhancing the detection efficiency of the given model. The difference in intensity of every pixel solely depends upon the texture edge and hence the output extracted from DVS is very sparse. To obtain higher confidence in Activity Detection, various segments of the DVS data is prepared also represented as motion maps which are further used to describe features of interest although while using the sparse data and segments of movement captured by the DVS. Moreover motion maps along with the features of interest are combined together to process them thereby increasing the activity recognising accuracy. The event stream from DVS is processed using a timing window to obtain a video segment. Hence by obtaining the mean over the time axis, x-y, x-t, y-t projection is obtained.

Gaming Ding [6] used a model that uses a decision tree-based approach to classify different human activities The classification accuracy of single tree classifier is enhanced by the concept of bagging which helps randomize the selection of the partitioning data nodes in the tree construction. We take a majority vote based on the decision tree provided and assign the vector to the given class. This process is done for each of the trees in the forest. This particular process requires a large collection of labelled data to achieve the required high accuracy. For classification, we make use of accelerative data. The model with its hights accuracy case outperforms the classification capability of SVM and Naive Bayes classifier.

Artur Jordao [7] used a model in which ample amount of data is collected from certain sensors named tri-axial accelerometer, magnetometer and gyroscope by which human activities are detected and classified into different classifiers based on recognition. To achieve this one methodology is the way of representing the raw signal to be able to clearly distinguish between human activities exploring at the classification stage. The signals used are signalLTCV and SNLS . They mainly emphasizes on rendering the leave which was proposed as leave-one-trial-out. This approach is widely used along with cross validation. This proposed methods also holds higher confidance and accuracy level for statistical models.

Fabien Baradel [8] proposed the model that uses the method of human activity detection with reference to the feature points captured along with the input data. We do not need to directly perform pose detection but rather perform two important processes, one to extract the relevant data with respect to the feature points and the other to derive inference based on the values of these points over time intervals. An unstructured set of data is collected as glimpses with the values extracted from trackers and recognizers. The pose that is detected is used in the training of the model while keeping the focus of the model on the structure of the human body based on the feature points. The attentional mechanism is carried out in the feature space which is used in calculation with the global model for the complete processing of the image obtained fro the camera feed.

Sumaira Ghazal [9] used an SVM classifier to extract human skeleton information such as the location of joints of the human body shape in skeletal from the camera input. This data can be used in the estimation of the human pose and can also be used to detect the activity associated with the motion of this data points based on the input obtained by the camera feed.

III. COMPARATIVE STUDY

Table 1

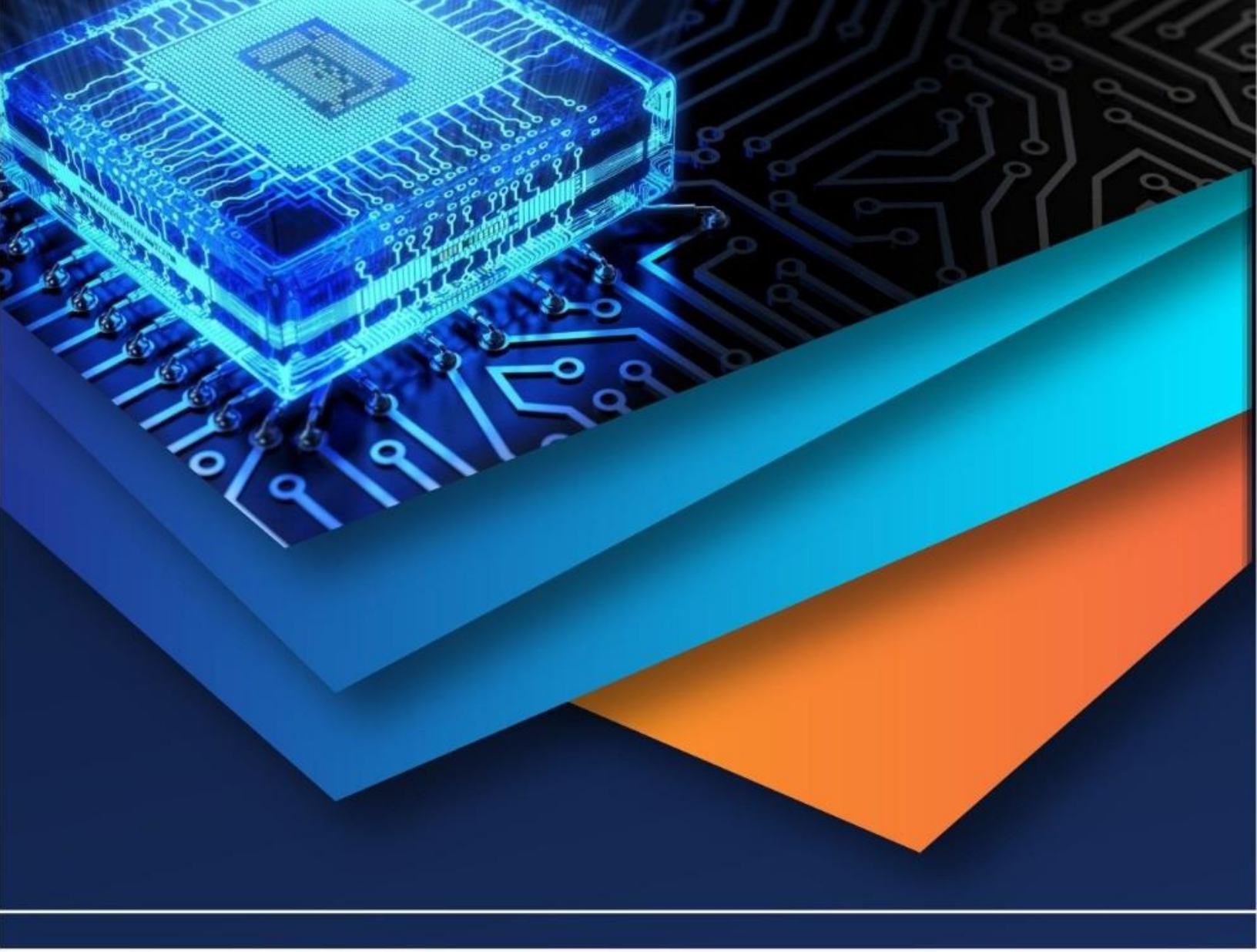
Sl No	Research Paper	Classifier	Dataset	Accuracy
1	Akram Bayat.2014 [10]	SVM	Triaxial accelerometer data	88.76%
2	Akbar Dehghani.2017 [4]	K Fold Cross-Validation	9 Xsens measurement data	96%
3	Erhan BÜLBÜL.IEEE 2018 [1]	SVM	Accelerometer and gyroscope data	96%
4	Zhenguo Shi.2018 [2]	LSTM - RNN	CSI packet	94.7%
5	Wenchao Xu.2018 [11]	CNN	Raw data	91.97%
6	Stefanie Anna Baby.2018 [5]	Motion maps and Motion Boundary Histogram	UCF YouTube Action Data Set,DVS gesture dataset.	79.33%
7	Gaming Ding.2018 [6]	RF Classifier	Raw data	93.01%
8	Sumaira Ghazal.2018 [9]	Rule based classifier	INRIA and Freiburg Dataset MPII Human Pose Dataset	95%
9	Fabien Baradel.2018 [8]	Global model and Glimpse Clouds.	NTU RDB+D Dataset, Northwestern-UCLA Multiview Action 3D Dataset	87.6% 89.9%
10	Anjana Wijekoon.2019 [3]	CNN	Accelerometer and Pressure sensor data.	81.82%
11	Artur Jordao.2019 [7]	LTCV and SNLS combinations	MHEALTH, PAMAP2 datasets	83%

IV. CONCLUSIONS

The goal of this survey paper was to help establish the different kinds of techniques and approaches that can be used for activity detection. As described above various promising approaches including deep learning and neural networks have been used. Nowadays human masses are monitored 24*7 through CCTV cameras. These are placed at sensitive places such as public gathering areas, train station and in schools and colleges. They provide aid to the security agencies in helping them monitor multiple locations at the same time while assisting in the detection of activities that require immediate law enforcement. As per the analysis conducted here, we have found some useful solutions which have scope for future enhancements. More experimentation must be performed on a variety of datasets to help improve the detection capabilities of various proposed models.

REFERENCES

[1] Erhan BÜLBÜL, Aydın ÇETİN and İbrahim Alper DOĞRU "Human Activity Recognition Using Smartphones"Gazi University Ankara, TURKEY IEEE 2018.
[2] Zhenguo Shi, J. Andrew Zhang, Richard Xu, and Gengfa Fang "Human Activity Recognition Using Deep Learning Networks with Enhanced Channel State information" IEEE 2018.
[3] Anjana Wijekoon, Nirmalie Wiratunga, Kay Cooper "MEx: Multi-modal Exercises Dataset for Human Activity Recognition" arXiv:1908.08992v1 [cs.CV] 13 Aug 2019.
[4] Akbar Dehghani, Tristan Glatard and Emad Shihab "Subject Cross-Validation in Human Activity Recognition" Conference 17, July 2017, Washington, DC, USA.
[5] Stefanie Anna Baby, Bimal Vinod, Chaitanya Chinni, Kaushik Mitra "Dynamic Vision Sensors for Human Activity Recognition" arXiv:1803.04667v1 [cs.CV] 13 Mar 2018.
[6] Gaming Ding*, Jun Tian*, Jinsong Wu**, Qian Zhao*, Lili Xie "Energy-Efficient Human Activity Recognition Using Wearable Sensors" 2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW): IoT-Health 2018: IRACON Workshop on IoT Enabling Technologies in Healthcare.
[7] Artur Jordao, Antonio Carlos Nazare, Jessica Sena, William Robson Schwartz " Human Activity Recognition Based on Wearable Sensor Data: A Standardization of the State-of-the-Art" arXiv:1806.05226v3 [cs.CV] 1 Feb 2019.
[8] Fabien Baradel, Christian Wolf, Julien Mille, Graham W. Taylor "Glimpse Clouds: Human Activity Recognition from Unstructured Feature Point " arXiv:1802.07898v4 [cs.CV] 21 Aug 2018.
[9] Sumaira Ghazal, Umar S. Khan "Human Posture Classification Using Skeleton Information".2018 International Conference on Computing, Mathematics and Engineering Technologies – iCoMET 2018.
[10] Akram Bayat∗ , Marc Pomplun, Duc A. Tran "A Study on Human Activity Recognition Using Accelerometer Data from Smartphones" The 11th International Conference on Mobile Systems and Pervasive Computing (MobiSPC-2014).
[11] Wenchao Xu, Yuxin Pang, Yanqin Yang and Yanbo Liu "Human Activity Recognition Based On Convolutional Neural Network" sensors." IEEE International Conference on Consumer Electronics-Asia IEEE, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY