iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** IX **Month of publication:** September 2020

DOI: https://doi.org/10.22214/ijraset.2020.31222

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IX Sep 2020- Available at www.ijraset.com

Activity Recognition System for Smart Campus

Manasi S. Khade[1], Hetal V. Gandhi[2]
[1, 2]Detartment of Computer Science & Engineering, Walchand College of Engineering, Sangli, India

Abstract: Long video data resulted from the use of cameras indoor and outdoor areas as a action recording. We present a new framework for recognizing student activities from the class. In this system we improve intelligent mechanism, top low level motion detection algorithm and feature extraction. We recognize the frame difference and feature selection for human activities that permits recognition. The detection of human activity from videos is very complicated appropriate to the complex reality of events, the situation in which activities took place, the require of available size of abnormal ground truth training data and other factors correlated to environmental disparity, light conditions and the working position of the captured cameras. The objective of this paper is to research and inspect machine and deep learning techniques by using videos for recognition of students indoor and outdoor activities. The importance has been on a variety of activity detection systems with machine learning techniques as their prime move toward.
Keywords: Activity, framework, motion detection, anomaly, machine learning, variation.

I. INTRODUCTION

Human Activity Recognition (HAR) system can even use the computer system to detect human activity or movements. The human activity recognition system has multitude of applications in atheletic competitions, medical treatments, smart home, transport services and elderly care [1]. The system uses machine learning algorithm for classification task. Imagine the scenario where the robot actively detects all human activities, avoiding crowd disturbances and acrimony before getting worse. In [2], the survey of different categories for activity detection and the features and techniques used for each category is provided. To study the view of many individual persons activity, the model requires not only explaining the individual act of every person in the framework but also signifying their collective behavior [3].

Many other earlier studies focus on single object actions or between pairs of objects, and for these methods it is important to successfully extract good characteristics. So we must focus on the actions of more than three people participating in this system, called group activities.

The main aim of the system is detecting the group activity through video based data and recognize the appropriate activity. The number of people is increasing in such group activites, and interactions between individuals are more complicated to find it harder to define group actvity. This system used two types of activities as shown in TABLE 1 one is indoor activity and other is outdoor activity. In that indoor activity contains *Empty class*, *Group discussion*, *Lecture is going on*, where as outdoor activity contains *Bucket overflow* and *Dumping correctly*. This study proposes a video analysis computer code system to improve the detection and recognition of hard and time activities.

Many of the papers contain the field of human activity recognition, crowded scene analysis and the behavioral comprehension, which are directly or indirectly correlated to video based activity detection. So the main objective of this paper is to study video based student activity recognition in college premises using machine learning and deep learning techniques.

Table I. Indoor and outdoor classes of activity recognition system

Location	The types of activity
Indoor	Empty class, Group Discussion, Lecture is going on
Outdoor	Bucket overflow, Dumping correctly

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IX Sep 2020- Available at www.ijraset.com

II. LITERATURE SURVEY

Human activity recognition done by using all the ways through the wearable devices which can be used for the different purposes like sports, fitness care and the event detection. The author Min et al. [4] designed a two different models in which first one is used the data from acceleration sensors and the second one is used for the location information. But in this paper before using the method of feature extraction, the sensor data is firstly splits into the different time segments which are basically the sequential time segments. In which the sliding window technique is used for the purpose of change in streaming data.

Human activity recognition system used to classify persons daily activity using the sensors that are more expensive for the human movement. Erhan et al. [5] has been proposed a different supervised machine learning algorithms like K-nearest neighbor (KNN), Support vector machine and the Decision tree and many of the methods such as Stacking, Bagging and the Boosting. In that for the classification purpose binary decision tree method is used and it gives the accuracy of 53.1\% . At the point when the branching limit is greater than the before 100, that time accuracy is increased to 94.4\. In this system were support vector machine gives an accuracy of 99.4\% and it uses the hyper dimensional planes. KNN algorithm gives a better accuracy like 97.1\% with the 3 k value. In many of the classifiers uses different agreement in which boosting technique is used. Basically AdaBoost is used for the miscalculated probability with the accuracy of 97.4\%. One more method is bagging and it gives the results of the sensitive learning algorithms with output accuracy over 98.1\%. And the last one method is stacking, in that 30 different classifiers are used and it predicts the 98.6\% of accuracy.

In the field of activity recognition system Pavel Dohnalek et al. [6] proposed the identification of three different states first is static likes lying, sitting, standing, second is dynamic, likes walking, running and third one is transition which are standing and walking. In which it is more important to pre-process the data for improve the classification model. Where classification methods are used for the classic algorithms such as Classification and Regression Tree (CART) and the k-Nearest Neighbour (KNN) and many techniques are used like Adaptive Neuro Fuzzy Interference System (ANFIS). For this purpose different datasets are used which are available in the UCI repository. The datasets are Physical Activity Detection for Elderly People2, OPPORTUNITY Activity Recognition Dataset, Localization Individual Dataset, and PAMAP2.

The main aim of the paper is used to spread a model which can recognize the numerous sets of the everyday activities under the actual conditions, by using the composed data through the single triaxial accelerometer which is already built into cell phone. Akram et al. [7] has proposed a technique in which persons many of the activities are used for the building of classification model by using the feature selection method. In Weka toolkit, Random Forest, LSTM, Multilayer Perceptron and the Simple Logistic Regression were used to compare single and collaborative classifiers and after that K-fold cross validation is used. Activity is recognized through mobile in the hand of human and the position of pocket. The efficiency is increased when the SVM classifier provide better accuracy. For find out the phone location SVM classifier gives 91.1\% accuracy but for the pocket position it gives 90.3\% accuracy. Single triaxial accelerometers were used to achieve the exact recognition accuracy of 91.1\% for the day to day activities.

Amin Ullah et al. [8] has proposed a technique for the recognition of activities in the various fields of video based surveillance which are captured over different industrial networks. In which the firstly the video stream is divided into many critical shots, after that these shots are selected using human saliency function of convolution neural network (CNN). Next, the temporal characteristics of the operation in the frame sequence are extracted using the convolution layer of a FlowNet2 CNN model. For the activity recognition the temporal optical flow feature provides the multilayer long-term memory which is used for understanding the long-term sequence. In which on various datasets experiments are performed for test action and these datasets are INRIA person dataset, UCF101 dataset, HMDB51 dataset, Hollwood2 dataset, and YouTube dataset. For these datasets model provides different accuracy such as LSTM gives 88.6%, RLSTM gives 86.9%, Hierarchical clustering multi-task gives 76.3%, Factorized spatio-temporal CNN gives 88.1% and DB-LSTM gives 92.84%.

Recognition of human activity in the various fields of ubiquitous computing is the way to implement the deep learning model to substitute the well-known different analytical techniques which are extraction techniques and classification features. Nils Y. Hammerla et al. [9] has proposed a detailed analysis of deep, convolution and recurrent approaches using different datasets which containing the data obtained with many wearable sensors. For training the recurrent network in this context, novel regularization approach is used and it also demonstrates how they outperform on the benchmark datasets. The datasets such as Opportunity dataset, PAMAP2 dataset and the Dephnet Gait dataset.

Channel State Information (CSI) is mainly developed for the activity recognition for the human events. Zhenguo Shi et al. [10] proposed a scheme which is used for identifying the human interaction with enhanced State Information using the Deep Learning Networks (DLN-eCSI). Author developed a CSI feature enhancement scheme (CFES), which included two models such as context reduction and correlation feature which are used to pre-process data and fed into DLN as a input. After that once the signals are

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IX Sep 2020- Available at www.ijraset.com

compressed using the CFES and recurrent neural network (RNN) is applied automatically for the extraction of deeper feature and for that softmax regression algorithm is used for classification of different activities.

Mahmudul Hasan and Amit K. Roy-Chowdhury [11] proposed a activity recognition system for the purpose of streaming the videos by collecting deep hybrid feature and the active learning. The segmented streaming video operations and unsupervised learning apps exploit the deep hybrid networks that have the more capability to take the advantage of local hand enginnered feature and the deep learning model. In which active learning is used to train the classifier for activity using the manually labelled instances. At the end rigorous experiments were conducted on the different datasets of human activity recognition to show the usefulness of framework. For that KTH human motion dataset, UCF50 dataset, VIRAT dataset and the TRECVID dataset are used.

Jun-Ho Choi and Manri Cheon [12] has proposed to examine the result of the video survillance on the activity recognition of multi view behaviour of humans using spatio temporal extraction feature and the deep learning method for using convolutional and recurrent neural networks. The actual work focused on to understanding the events from different perspective. For such purpose, author used the Berkeley Multimodal Human Action Dataset (Berkeley MHAD), which contains many videos recorded from the 12 cameras at a time. It contains the 11 different activities like jumping, bending, punching, waving hands, waving one hand, clapping, throwing a ball, sitting down, jumping jacks and standing up.

Jian Bo Yang et al. [13] has proposed a technique of systematic learning feature for the problem of human activity recognition. Such type of approach uses a convolutional neural network (CNN) for the use of systematically automate feature learning from the raw inputs. This features are think for the higher level of abstract reprentation of low level of raw time series signals using the deep learning. Author founds two datasets for appreciative the human performance. In which first dataset is Opportunity Activity Recognition dataset which is related to the movement of the each person and the second dataset is the Hand Gesture dataset which is used to focus on the particular action of the hand. For such type of classifications mostly SVM, kNN, Mean and variance techniques are used.

Following TABLE II is a description of researcher's literature work on the basis of the dataset and the techniques used .

Table II. Summary of literature work done by researchers

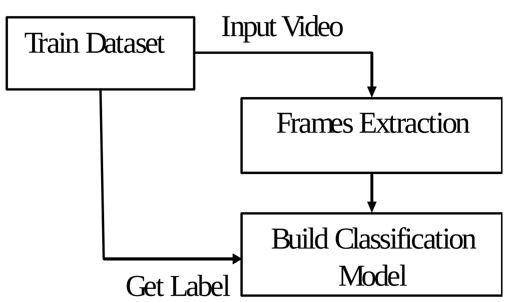
Author	Dataset	Techniques	Results
Kwon Min-Cheol, and Sunwoong Choi [4]	Sensor data	Decision tree,Random forest, Support vector machine	95% Acccuracy
Erhan Bulbul et al. [5]	Sensor data	Decision tree, Support vector, KNN	Not mentioned
Pavel Dohnálek et al. [6]	Opportunity Dataset, PAMAP2 dataset, Localization dataset	KNN, Classification and regression tree, Random forest, Linear discriminant analysis	98% Accuracy
K. P. Sanal Kumar and Dr. R. Bhavani [7]	Not mentioned	KNN, Support vector machine	88.47% Accuracy
Amin Ullah et al. [8]	UCF101, HMDB51, Hollywood2, UCF50, YouTube dataset	CNN, LSTM	69.5% 94.9% 95.8%
Nils Y. Hammerla et al. [9]	Opportunity, PAMAP2, Daphnet Gait dataset	RNN, CNN	Not mentioned
Zhenguo Shi et al. [10]	Not mentioned	RNN	Not mentioned
Mahmudul Hasan and Amit K. Roy-Chowdhury [11]	KTH, UCF50, VIRAT, TRECVID dataset	Deep networks	Not mentioned
Jun-Ho Choi and Manri Cheon [12]	MHAD dataset	Deep learning	Not mentioned
Jian Bo Yang et al. [13]	Opportunity, Hand gesture dataset	CNN, Support vector machine, KNN	Not mentioned

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IX Sep 2020- Available at www.ijraset.com

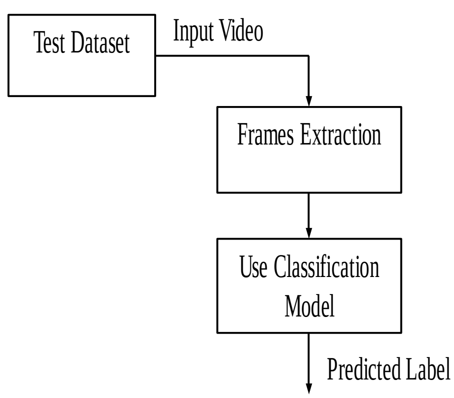
III. METHODOLOGY

Recognizing human activities depends directly on the characteristics derived from movement analysis. In the section of methodology, we explain the activity recognition framework for the student activities in detail as shown in Fig 1. In this framework input is a video segment containing one activity and output is recognizing appropriate activity.

The model provides a framework for the activity recognition system by using the video surveillance which are recorded in college campus, where the behaviours of human are identified in the stream of surveillance by using the convolution layers of AlexNet, LeNet5 and the ResNet CNN models. Activity recognition is a wide field of study focused on recognizing a person's specific movement or action based on knowledge from the video recording. Such type of movements or actions are often typically a activities performed indoors as well as outdoor, such indoor activities are empty class, group discussion, lecture is going on and outdoor activities are dumping correctly and bucket overflow. The data may be remotely recorded, such as videos. The continuous stream of video is firstly splits into frames, where cetain frames are selected using the convolution neural network (CNN). CNNs are used for image classification and recognition because of its high accuracy. It is challenging problem because there are no clear or straightforward ways to link the recorded video data to actual human activities and each subject may perform an operation with considerable variability, resulting in variations in the recorded video data.



(a)



(b)

Fig. 1. Block Diagram of Activity Recognition System. (a) Building model during training phase. (b) Predicting labels during testing phase.

A. Dataset

Collect different CCTV videos of indoor and outdoor activities from college. Indoor activity contains three classes like empty class, group discussion and lecture is going on. Where as outdoor activity contains two classes which are dumping correctly and bucket overflow. For using these CCTV videos generate a frames. For the indoor classes total 4073 frames are used and for outdoor classes total 1063 frames are used. TABLE III describes the size of sample frames for each class under indoor and outdoor category.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IX Sep 2020- Available at www.ijraset.com

Table III. Dataset of indoor and outdoor classes

Category	Classes	Total Images
Indoor	Empty class	1379
	Group discussion	1194
	Lecture is going on	1500
Outdoor	Bucket overflow	771
	Dumping correctly	292

B. Video To Frame Conversion

Any video or animation you see is made up of a sequence of still images on your screen, camera, tablet or even at the movie theatre. The faster the images are played, the more natural and smooth the motion appears. Most of the videos are captured at approximately 24-30 frames per second, each image is called a frame where you see frames per second (FPS). In our system framerate is 1.0 means it will capture image in each 1.0 second.

C. Pre-processing

In pre-processing resize all images with same width and height. Pre-processing is apply only when one of the frame having different height and width as compare to other generated frames. After pre-processing divide frames into two folders such that 80% of frames present in training folder and 20% of frames present in testing folder and each folder contains different indoor and outdoor classes.

D. Classification Model

Classification is the process of predicting the type of data points given. A classifier uses some training data to understand how class relates to given input variables. When the classifier is accurately trained, the output is then given. In this framework 20% frames are used for the testing and 80% frames for training data. Many classification algorithms are available now, but it can not be proven which one is superior. In this system CNN classifiers like AlexNet, LeNet5 and ResNet are used to classify the indoor and outdoor activities with different accuracy for both indoor and outdoor classes.

E. AlexNet Architecture

AlexNet architecture consist of total 8 layers in which 5 are convolution layers and remaining 3 are fully connected layers. It examine the problem of image classification. The input of the AlexNet is a picture which having 1000 number of classes, and output contains a vector of 1000 numbers. The input size of AlexNet is 256×256 RGB image. It means that all images contains in the training set need to be size of 256×256. For the image classification, features are extracted using multiple convolutional kernels (filters). There are many kernels available in single convolution layer with same size. In which AlexNet's first convolution layers contains 96 kernels with 11×11×3. In that overlapping max pooling layer followed by the first two convolution layer. There is connection between third, fourth and fifth convolution layer. After that second fully connected layer feeds into a softmax classifier with the 1000 number of class labels.

F. LeNet5 Architecture

The LeNet5 architecture consist of two sets of convolution layers and average pooling layers, which followed by a flattening layer, then fully connected layer and finally a softmax classifier is used. In which convolution layer having 5×5 convolution with stride is 1. The sub sampling layer having 2×2 average pooling layer. In this network Tanh sigmoid activation function is used. There are many architectures are available that were made in LeNet5 which are not very common in the era of deep learning.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IX Sep 2020- Available at www.ijraset.com

IV. EXPERIMENTAL RESULTS

A subset of deep learning networks is the convolutional neural network (CNN). CNNs constitute a huge breakthrough in the classification of images. Classification of images is the process of taking an input (like a picture) and outputting a class. For the image classification different models are used which are AlexNet, LeNet5 and ResNet. Using these models indoor and outdoor classes are classified as shown in following image.

A. Models

Classification of Indoor and Outdoor Activities

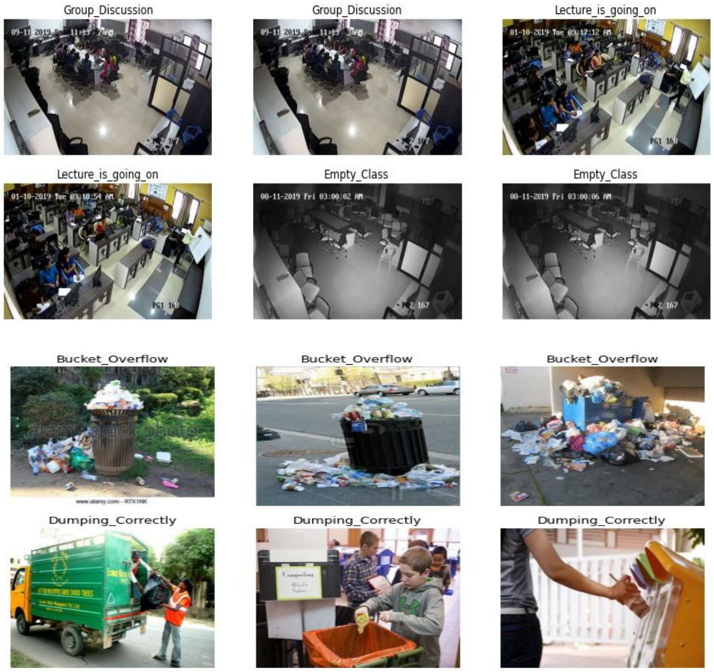


Fig. 2. Classification Results of Indoor and Oudoor Classes of Activity Recognition System

The performance of the proposed CNN approach on Hand Movement dataset shown for comparision purpose in TABLE IV [13].

Table IV. The AF, NF and AC results of the proposed cnn method for the hand gesture dataset.

	Average F-measure	Normalized F-measure	Accuracy
SVM	76.0	85.0	85.6
KNN	64.8	73.2	71.8
MV	87.5	91.3	91.2
DBN	71.8	82.2	82.8
CNN	89.2	92.0	92.2

B. Confusion Matrix

Following graph shows that number of samples in test data for each class and out of that how many are correctly classified for both indoor and outdoor activity.

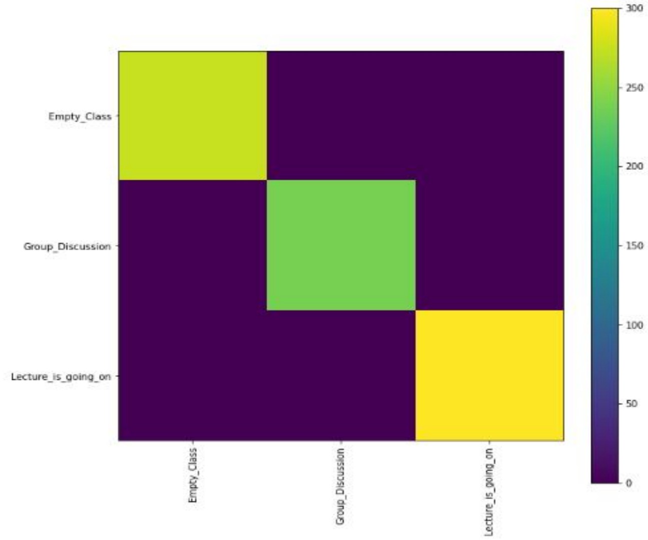


Fig. 3. Confusion matrix for indoor classes

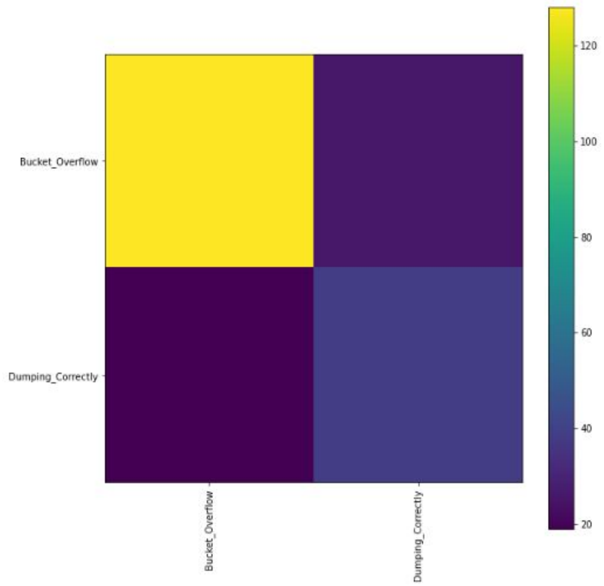


Fig. 4. Confusion matrix for outdoor classes

C. Accuracy Table

Accuracy is one metric used to assess models of classification. Informally, it's the percentage of results that shows how our model got correct. TABLE V and VI contains accuracy of indoor and outdoor activities using LeNet5 and AlexNet model.

Table V. Accuracy table of indoor activity recognition system

	Indoor
LeNet5	99%
AlexNet	99%

Table VI. Accuracy table of outdoor activity recognition system

	Outdoor
LeNet5	78%
AlexNet	78%

D. Error Analysis

Formally, error Analysis refers to the process of examining the misclassification of your algorithm, so we can understand the underlying causes of the errors. It gives us a direction for handling the errors. Typically, we start with a small model which is bound to be of low precision (high error). We can then begin evaluating this model and analyze the errors. We can grow with model as we evaluate and repair these errors.

Fig. 5. Misclassification of images

V. CONCLUSION

In this system we build the models for HAR system using a image dataset and CNN. We considered 3 indoor activities including *Empty class*, *Group discussion*, and *Lecture is going on*, and outdoor activities including *Bucket overflow* and *Dumping correctly*. The experimental results for this study showed that various indoor activities can be classified using LeNet5 and AlexNet models and these both models gives 99% of accuracy. Where as outdoor activities classified using LeNet5 and AlexNet models and these two models gives 78% of accuracy.

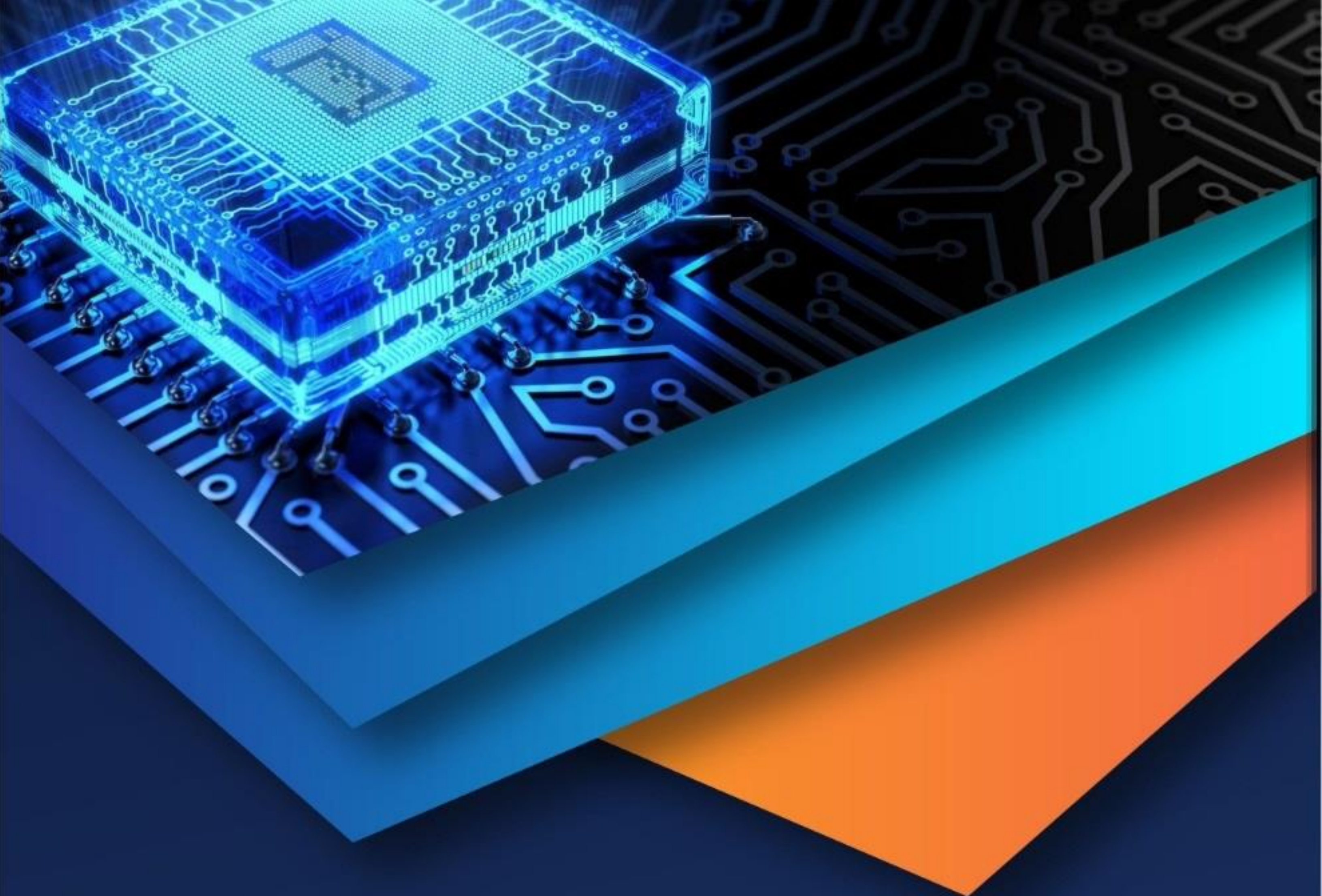
VI. FUTURE SCOPE

Future strategies will concentrate on improving our approach to identification and extending into areas of community interaction of greater interest. In existing system they used image dataset which contains single entity (person) but in our approach we are going to work on video dataset contains various indoor and outdoor group activities based on video surveillance captured over college campus.

REFERENCES

[1] Sun, Jian, Yongling Fu, Shengguang Li, Jie He, Cheng Xu, and Lin Tan. "Sequential Human Activity Recognition Based on Deep Convolutional Network and Extreme Learning Machine Using Wearable Sensors." Journal of Sensors 2018.
[2] Pawar, Karishma, and Vahida Attar. "Deep learning approaches for video-based anomalous activity detection." World Wide Web 22, no. 2 (2019): 571-601.
[3] Wu, Jianchao, Limin Wang, Li Wang, Jie Guo, and Gangshan Wu. "Learning Actor Relation Graphs for Group Activity Recognition." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9964-9974. 2019.
[4] Kwon, Min-Cheol, and Sunwoong Choi. "Recognition of Daily Human Activity Using an Artificial Neural Network and Smartwatch." Wireless Communications and Mobile Computing 2018 (2018).
[5] Bulbul, Erhan, Aydin Cetin, and Ibrahim Alper Dogru. "Human Activity Recognition Using Smartphones." In 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), pp. 1-6. IEEE, 2018.
[6] Dohnalek, Pavel, Petr Gajdoš, and Tomáš Peterek. "Human activity recognition: classifier performance evaluation on multiple datasets." Journal of Vibroengineering 16, no. 3 (2014): 1523-1534.
[7] Bayat, Akram, Marc Pomplun, and Duc A. Tran. "A study on human activity recognition using accelerometer data from smartphones." Procedia Computer Science 34 (2014): 450-457.
[8] Ullah, Amin, Khan Muhammad, Javier Del Ser, Sung Wook Baik, and Victor Hugo C. de Albuquerque. "Activity recognition using temporal optical flow convolutional features and multilayer LSTM." IEEE Transactions on Industrial Electronics 66, no. 12 (2018): 9692-9702.
[9] Hammerla, Nils Y., Shane Halloran, and Thomas Plötz. "Deep, convolutional, and recurrent models for human activity recognition using wearables." arXiv preprint arXiv:1604.08880 (2016).
[10] Shi, Zhenguo, J. Andrew Zhang, Rithard Xu, and Gengfa Fang. "Human activity recognition using deep learning networks with enhanced channel state information." In 2018 IEEE Globecom Workshops (GC Wkshps), pp. 1-6. IEEE, 2018.

[11] Hasan, Mahmudul, and Amit K. Roy-Chowdhury. "A continuous learning framework for activity recognition using deep hybrid feature models." IEEE Transactions on Multimedia 17, no. 11 (2015): 1909-1922.
[12] Choi, Jun-Ho, Manri Choi, Min-Su Choi, and Jong-Seok Lee. "Impact of three-dimensional video scalability on multi-view activity recognition using deep learning." In Proceedings of the on Thematic Workshops of ACM Multimedia 2017, pp. 135-143. 2017.
[13] Yang, Jianbo, Minh Nhut Nguyen, Phyo Phyo San, Xiao Li Li, and Shonali Krishnaswamy. "Deep convolutional neural networks on multichannel time series for human activity recognition." In Twenty-Fourth International Joint Conference on Artificial Intelligence. 2015.
[14] Chen, Heping, Biao Zhang, and Thomas Fuhlbrigge. "Robot Throwing Trajectory Planning for Solid Waste Handling." In 2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), pp. 1372-1375. IEEE, 2019.
[15] Rutqvist, David, Denis Kleyko, and Fredrik Blomstedt. "An Automated Machine Learning Approach for Smart Waste Management Systems." IEEE Transactions on Industrial Informatics 16, no. 1 (2019): 384-392.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY