HUMAN ACTIVITY RECOGNITION USING K-NEAREST NEIGHBOR MACHINE LEARNING ALGORITHM

Vallabhaneni Rakesh Sai Chowdary[1] B.Tech, Student, Dept. of ECE, Rama Chandra College Of Engineering and Technology, *JNTUK*
Simhadri Raja Sri[2] B.Tech, Student, Dept. of ECE, Rama Chandra College Of Engineering and Technology, *JNTUK*
Veeravarapu Vamsi[3] B.Tech, Student, Dept. of ECE, Rama Chandra College Of Engineering and Technology, *JNTUK*
Vadlamudi Likhitha[4] B.Tech, Student, Dept. of ECE, Rama Chandra College Of Engineering and Technology, *JNTUK*
Ms.K.Noha Namrata[5] M.Tech, Assistant Professor, Dept. of ECE, Rama Chandra College Of Engineering and Technology, *JNTUK*

Abstract

Recognizing Human Activity is now an important concept now a days. Machine learning algorithms (ML) can help to accurately identify human activities, as long as ML algorithms are well designed and trained for high performance recognition. This paper introduces a nearest neighbouring algorithm k (KNN) for classifying human activities, namely running, walking, boxing, hand movements, cycling and surfing. For this purpose, we had used KTH dataset. This algorithm is trained and the algorithm parameters are tuned precisely to achieve maximum accuracy when compared to many existing algorithms. The results show that the KNN algorithm provides high performance in classifying human activities in all cases.

Keywords: Machine learning, KNN, Human activity recognition·

I.INTRODUCTION

Due to advances in laptop vision, computer systems are increasingly deciding on some very difficult issues (including information and photographs). Models are created in which the version can expect what the photograph is or if the version is given a photograph can find out whether the item selected within the photograph is a gift or not. These fashions are called neural networks (or synthetic neural networks) that are stimulated by the shape and function of the human mind. Gaining in-depth knowledge, the subfield of the knowledge system is to take a look at the neural networks that have distributed numerous versions of those networks over the years for various problems. For video recognition, this method uses in-depth information - in the case of certain classified images, a version is created so that it is able to generate a guess label for the brand-new video. Steps were taken for the execution, including video dataset download, extraction and pre-processing, then splitting the dataset into schooling, examining the information and then educating it on schooling information by examining the arrival of the neural community and earlier or later versions. Take a look at the information [1].

Fig 1 Different Actions

The fourth business revolution (Industry 4.0) is characterized by the use of Net of Things (IoT), cyber-body systems, giant statistics, and the use of synthetic intelligence (AI), which had a significant impact on clever manufacturing processes [1]. Moreover, humans in agile manufacturing facilities are generally chargeable for production processes. Therefore, it is necessary to classify human games in order to understand human performance at certain stages of production. In other words, with the latest virtual transformation involving the exponential growth of data and communication-push technology, it is very important to find human games seamlessly in clever factories [2, 3]. In this regard, this paper offers a strong KNN set of rules for classifying human sports, especially for laying, down stair, seating, up, standing and walking, with the aim of increasing the accuracy of the set of rules through quality tuning. Dimensions of a set of rules. The most important contributions of the paper are as follows:

(i) Executing of a stronger KNN set of rules with finetuned parameters to categorise human sports,

(ii) Estimating the overall performance of the applied set of rules for several assessment metrics the use of the HAR dataset.

(iii) A approach that accurately identifies a human from a video or an picture is a tough task.

Over the past few years, "PC imaginative and foreshadowing" scientists have been engaged in the popularity of human motion, especially due to the wide variety of commercial applications with human-PC interaction [1], counter-terrorism programs [2], and site visitor monitoring [3]. , Automobile safety [4], video surveillance [5], real-time tracking [6], rescue missions [7] and human-robotic interaction [8]. These pictures emphasize the search for man from the movies. Moreover, various demanding conditions like lighter versions and history versions are also noted. The first goal of the popularity of human motion suggests the study of enter pictures or video (non-stop series of frames) for ongoing human games. In a simple scenario, a video containing unmarried human entertainment is successfully categorized into its maximum associated entertainment category. The movement and position of the frame elements are used in the human version to understand human movement - primarily the methods we use. Positive factors of body composition, joint motion and human frame can be applied to detect human movement [10-12]. The fragmentation of human frame elements is effectively achieved in movies [13, 14], however, when the search for those frame elements is required in a real-time environment, it becomes more difficult.

II. RELATED WORK

inside the next consultation, we in short assessment unique movement reputation strategies. Over the previous few many years, RGB photos has been notably applied in human movement recognition techniques.

the usage of every depth maps and skeleton joints, D. Thombre, J. Nirmal, and D. Lekha, [1] "Human detection and monitoring the usage of image segmentation and Kalman filter out,". category of human sports is one of the growing research regions withinside the area of computer vision.

M. Sharif, M. A. Khan, T. Akram, M. Y. Javed, T. Saba, and A. Rehman, [2] Human interest tracking withinside the video sequences is an fascinating laptop imaginative and prescient location which includes tremendous applications, e.g., surveillance systems, human-laptop interplay, and placement traffic control systems. on this studies, attention is in offering a hybrid method for green category of human sports activities from a given video series.

J.-W. Hsieh, S.-H. Yu, Y.-S. Chen, and W.-F. Hu,[3] proposed an automatic web site traffic surveillance device to estimate vital web site traffic parameters from video sequences the usage of best one virtual digicam. as soon as all abilities are extracted, an maximum efficient classifier is then designed to robustly categorize motors into exclusive training. Experimental outcomes display that the proposed technique is greater sturdy, correct, and powerful than distinct conventional techniques, which make use of best the automobile duration and a unmarried frame for car category.

G. C. Dedes and k. C. Mouskos [4] independent automobile adventure making plans is increasingly gaining hobby from scientists and professionals, who are addressing the combination of self enough motors into the general metropolis transportation tool.

G. Kapidis, R. Poppe, E. van Dam, L. P. Noldus, and R. C. Veltkamp, [5] egocentric vision is an growing area of computer vision that is characterised with the resource of using the acquisition of photographs and video from the primary man or woman perspective.

J. Cacace, A. Finzi, V. Lippiello, M. Furci, N. Mimmo, and L. Marconi, [7] An shape appropriate for the control of more than one unmanned aerial cars deployed in search & Rescue missions is furnished on this paper. the usual tool has been tested in a real worldwide undertaking with drones equipped with on-board cameras.

M. A. Goodrich and A. C. Schultz [8] Human–robotic interaction (HRI) has nowadays acquired sizeable hobby withinside the instructional network, in labs, in generation compa- nies, and thru the media. Q. Ye, Z. Han, J. Jiao, and J. Liu, [9] With the improvement of port automation, most operational fields making use of heavy system have often turn out to be unmanned. V. Parameswaran and R. Chellappa, " [11] Human movement recognition is an crucial trouble in laptop imaginative and prescient. It has a huge sort of programs in surveillance, human-pc interplay, augmented fact, video indexing, and retrieval.

Althloothi et al. [12] proposed a way primarily based mostly on units of abilties, shape abilities, kinematic structure. shape abilties have been extracted from intensity maps and the

kinematic capabilities extracted from the 3-d joint positions of a human frame and fused collectively using multiple kernel gaining knowledge of (MKL) techniques for human motion recognition.

M. V. da Silva and A. N. J. A. S. C. Marana, "Human motion reputation in movies based mostly on spatiotemporal competencies and bag-of-poses," presently, there can be a massive wide sort of strategies that use 2d poses to represent and recognize human movement in movies. Nunez, R. Cabido, J. J. Pantrigo, A. S. Montemayor, and J. F. Velez, [9] "Convolutional neural networks and lengthy brief-time period reminiscence for skeleton-based totally absolutely human interest and hand gesture reputation," Experimental testing confirmed that our education technique obtains better results than a single-degree education method.

The remainder of the paper is organized as follows: Sect. 4 presents the proposed machine learning algorithm with the dataset description. The experimental results are discussed in Sect. 7. Finally, conclusion is drawn in Sect. 8.

III. HAR MODEL

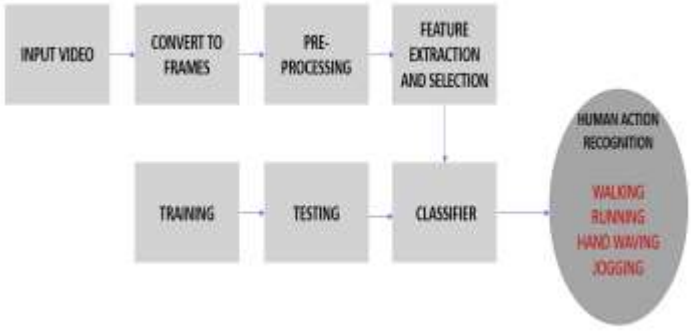

Fig 2 Existing flow chart Human action Recognition

Table I. Action Labels Coding

Action	Coding
Boxing	0
Handclapping	1
Hand waving	2
Jogging	3
Running	4
Walking	5

The video's spatial dimension (width x height) can be found to be 160 x 120 pixels. Also, on loading a single video into a NumPy array in python, the shape of the array obtained was (1, 515, 120, 160, 3), this indicates that the video has 515 frames with the spatial dimension of the video is 160 x 120 (width x height) pixels and each frame has 3 channels Red(R), Green(G) and Blue(B).

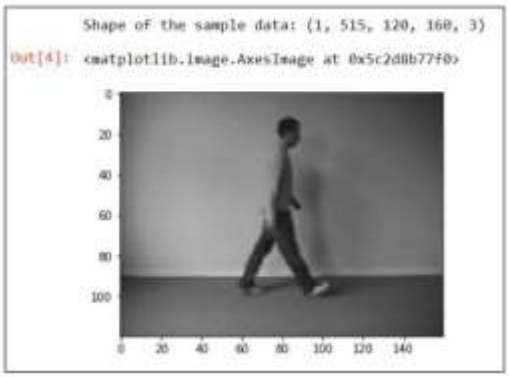

Fig. 3. Single frame of a sample video of walking of KTH dataset with shape

Dataset pre-processing operations

i. Reading in the video frame-by-frame.

ii. The videos were captured at a frame rate of 25fps.

This technique that for every 2d of the video, there can be 25 frames. We keep in mind that internal a 2nd, a human body does now not carry out very

huge movement. which means most of the frames (constant with 2nd) in our video can be redundant. consequently, best a subset of all of the frames in a video choice to be extracted. this will additionally lessen the dimensions of the enter records so as to in flip help the version educate faster and also can prevent over-becoming. distinctive techniques may be used for body extraction like: x Extracting a hard and fast extensive variety of frames from the entire frames withinside the video - say only the number one hundred frames (i.e., first eight seconds of the video). x Extracting a hard and fast huge form of frames every second from the video – say we want most effective five frames constant with second from a video whose length is of 10 seconds. this may pass back a whole of 50 frames from the video. This technique is better withinside the feel that we're extracting the frames cautiously and uniformly from the whole video. iii. every frame desires to have the same spatial dimensions (pinnacle and width). as a result every body in a video will need to be resized to the preferred length. iv. on the way to simplify the computations, the frames are converted to grayscale. v. Normalization - The pixel values stages from zero to 255. these values should ought to be normalized that permits you to assist our version converge quicker and get a better performance.

A. Implementation

The video facts set was loaded and the vital pre-processing measures have been one of the maximum crucial components of the project. consequently, we created a class (motion pictures) which became renamed (look at movies)) (to play and technique pictures. the answer over such trouble is obtained with the resource of the usage of utilizing picture AI set of policies which ends up in kind the frames having human frame observed in decided on .

Class Label	0	1	2	3	4	5
Boxing	1	0	0	0	0	0
Handclapping	0	1	0	0	0	0
Handwaving	0	0	1	0	0	0
Jogging	0	0	0	1	0	0
Running	0	0	0	0	1	0
Walking	0	0	0	0	0	1

Fig.4 Labelling of parameters

IV. PROPOSED ALGORITHM

In this article, an efficient model for HAR in uncontrolled surroundings is proposed. The proposed model consists of steps like eliminating redundant frames from movies, extracting Segments of interest (SoIs).

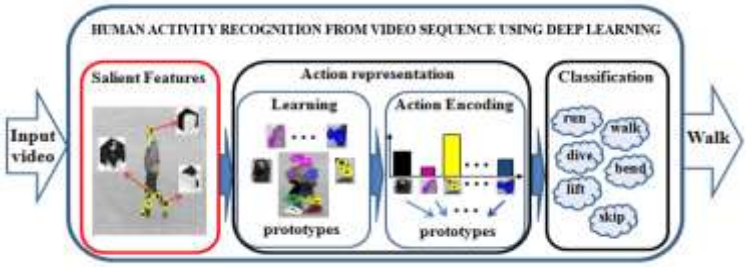

Fig. 5 Proposed architecture of Human action recognition

Fig. 5 shows the structure of the proposed KNN algorithm. It is implemented using Scikit-learn framework. It consists of five steps: the first step is a pre-processing using filters to remove the noise from the utilized dataset, the second step

is a feature extraction of the dataset. Also, another extracted features of the dataset are lx, ly, and lz, which represent the gyroscope angles. The third step is the selection of a specific feature, the fourth step is the training on the dataset, and the fifth step is the testing/classification for the activities. Fig 6 illustrates a mechanism of the KNN algorithm. For instance, it assumes two categories or classes "A and B" and a new data point x1 which colored blue. So, the KNN algorithm can detect the data point in which category will be lied.

In this algorithm, a neighbourhood function is used to classify the activities. The best selection of k depends on the dataset. So, the largest k reduces the noise applied to a classification. Furthermore, the KNN is a supervised machine learning algorithm used to obtain the nearest data point from the same class and nearest data from different classes. The dataset readings are extracted with a 50 Hz sampling rate. The dataset has 6 attributes with information correlated to the activities of the humans' activities: time and x-, y-, z-accelerations. The dataset is divided into an 70% training set and a test set of 30%. The test set is utilized to evaluate the proposed algorithm.

Fig. 6 The proposed KNN algorithm structure.

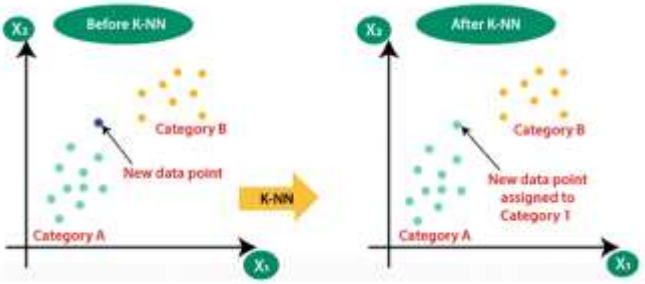

Fig. 7 The mechanism of the KNN algorithm.

$$d_E(x, y) = \sqrt{\sum_{i=1}^{n}(x_1 - y_1)^2}$$

i. Reading in the video frame-by-frame.

ii. The videos were captured at a frame rate of 25fps.

iii. Each frame needs to have the same spatial dimensions (height and width). Hence each frame in a video will have to be resized to the required size.

iv. In order to simplify the computations, the frames are converted to grayscale.

v. Normalization - The pixel values ranges from 0 to 255. These values would have to be normalized in order to help our model converge faster and get a better performance.

Detecting Redundant Frames

nowadays, virtual cameras acquire video information at 60 frames in step with second (fps) for 4K motion pictures or 30 fps for excessive-resolution films to make sure comprehensive

smoothness. Surveillance cameras uninterruptedly file films whenever; for this reason, hundreds of heaps or maybe millions of video frames are saved each hour. plenty computational time and power are required to process those frames correctly. From those movies' conduct activities, it is found that there exist many static, vain, and redundant frames, which gradual down the processing velocity of any efficient set of rules. those hobby-points are then combined with temporal and neighborhood constraints to locate the static frames. residences of SIPs appear that if the quantity and vicinity of interestpoints have not changed in a number of the video frames, it's far considered that the content material of the situation video has not changed in the course of those frames. This belongings may be hired by retaining a single frame instead of redundant frames. The proposed set of rules detects saual points at temporal and spatial scales by means of enforcing.

A. Extracting Segments of Interest

Uncontrolled surroundings, in comparison to regulated surroundings in which all parameters and environment are tuned to the most efficient level, lacks all of those. Many unsure matters may be located in a video shot in uncontrolled surroundings, inclusive of constantly moving backdrops, situations, display settings, and lots of (relevant or irrelevant) objects. those variables have a vast impact at the overall performance of an efficient set of rules. To address those boundaries, an effective approach for extracting SoI is proposed. The cautioned approach separates a video body into awesome subparts based totally on installed policies, and these subparts are called SoI. Semantic functions (along with interest, shape, and geometry) are incorporated with low-level statistics to extract SoIs (comparison, sharpness and colour). The Relevance is then calculated the usage of this aggregate.

Frames are transformed to grayscale photographs and then a low-bypass filter is carried out to them to determine the contrast rating (CS). The picture is smoothed through using a low-pass clear out to eliminate high-spatial frequency noise. The sharpness of a photograph is appeared as an vital criterion for judging the first-rate of a body due to the fact it's miles closely linked to humans' subjective feelings. Sharpness score (SS) is calculated through calculating the rectangular of difference between two adjacent pixels in a grayscale photo. colour is also critical in reaching a a success SoI extraction.

V. RESULTS ANALYSIS

Fig. 8 Consecutive frames of a sample video of 'Walking' of KTH dataset

Here in Fig. 8 we had given an video as input in which we had taken 50 frames of the input video and the analysis of the input parameters is done using KNN.

Fig. 9 The frames of a sample video of 'Walking' of KTH dataset with Image AI

Fig. 10 The frames of a sample video of 'Hand Waving' of KTH dataset with Image AI

Fig. 11 The frames of a sample video of 'Hand Waving'

Processing frame no.1

Processing frame no. 2……..

……..Processing frame no.48

Processing frame no.49

Processing frame no.50

Hand Waving

Classification Accuracy: 97.7778 %

Processing Time (s): 8.674555

VI. CONCLUSION

The essential motive of human movement reputation is to look for persevering with human sports in pictures or video (a continuous series of frames). on this paper, we gift an more suitable HAR version based on strategies including deleting superfluous frames from movies, extracting SoIs, and using characteristic descriptors. The implementation of a KNN gadget studying algorithm for the recognition of daily human movements become described in this have a look at. This algorithm has a 97.46 percent testing accuracy and a 2.54 percentage checking out loss charge. because the datasets utilized in deep gaining knowledge of video processing are generally massive, superior processing requires GPU computing with higher overall performance.

REFERENCES

[1] D. Thombre, J. Nirmal, and D. Lekha, "Human detection and tracking using image segmentation and Kalman filter," in 2009 International Conference on Intelligent Agent & Multi-Agent Systems, 2009, pp. 1-5: IEEE.

[2] M. Sharif, M. A. Khan, T. Akram, M. Y. Javed, T. Saba, and A. Rehman, "A framework of human detection and action recognition based on uniform segmentation and combination of Euclidean distance and joint entropybased features selection," EURASIP Journal on Image and Video Processing, vol. 2017, no. 1, p. 89, 2017.

[3] J.-W. Hsieh, S.-H. Yu, Y.-S. Chen, and W.-F. Hu, "Automatic traffic surveillance system for vehicle tracking and classification," IEEE Transactions on Intelligent Transportation Systems, vol. 7, no. 2, pp. 175- 187, 2006.

[4] G. C. Dedes and K. C. Mouskos, "GPS/IMU/video/radar absolute/relative positioning communication/computation sensor platform for automotive safety applications," ed: Google Patents, 2014.

[5] G. Kapidis, R. Poppe, E. van Dam, L. P. Noldus, and R. C. Veltkamp, "Egocentric Hand Track and Object-based Human Action Recognition," arXiv preprint arXiv:1905.00742, 2019.

[6] A. Veenendaal, E. Daly, E. Jones, Z. Gang, S. Vartak, and R. S. Patwardhan, "Sensor tracked points and HMM based classifier for human action recognition," Computer Science and Emerging Research Journal.

[7] J. Cacace, A. Finzi, V. Lippiello, M. Furci, N. Mimmo, and L. Marconi, "A control architecture for multiple drones operated via multimodal interaction in search & rescue mission," in 2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), 2016, pp. 233-239: IEEE.

[8] M. A. Goodrich and A. C. Schultz, "Human–robot interaction: a survey," Foundations and Trends® in Human–Computer Interaction, vol. 1, no. 3, pp. 203-275, 2008.

[9] Q. Ye, Z. Han, J. Jiao, and J. Liu, "Human detection in images via piecewise linear support vector machines," IEEE transactions on image processing, vol. 22, no. 2, pp. 778-789, 2012.

[10] S. Ali, A. Basharat, and M. Shah, "Chaotic invariants for human action recognition," in 2007 IEEE 11th International Conference on Computer Vision, 2007, pp. 1-8: IEEE.

[11] V. Parameswaran and R. Chellappa, "View invariance for human action recognition," International Journal of Computer Vision, vol. 66, no. 1, pp. 83-101, 2006.

[12] A. Yilmaz and M. Shah, "Recognizing human actions in videos acquired by uncalibrated moving cameras," in Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, 2005, vol. 1, pp. 150-157: IEEE.

[13] D. Ramanan, D. A. Forsyth, and A. Zisserman, "Tracking people by learning their appearance," IEEE transactions on pattern analysis and machine intelligence.

[14] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction

for human pose estimation," in 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1-8: IEEE.