How to Cite:
Deepshikha, K. M., Verma, S. K., & Chabaque, A. (2022). Multi action recognition using machine learning. *International Journal of Health Sciences*, *6*(S5), 7288–7301.
https://doi.org/10.53730/ijhs.v6nS5.10334

Multi action recognition using machine learning

KM Deepshikha
Department of CSE, Galgotias University
Email: deepshikhadhillon123@gmail.com

Dr. Shiv Kumar Verma
Department of CSE, Galgotias University
Email: shiv.verma@galgotiasuniversity.edu.in

Dr. Aakash Chabaque
Department of CSE, KITE Group of Institions, Meerut
Email: accheckmail@yahoo.com

***Abstract*---**One of the most exciting and useful computer vision research topics is automated human activity identification. The majority of existing research, as well as most traditional approaches and classic neural networks, disregard the in video sequences, the appearance and patterns of motion are important. Individuals are unableIn video sequences, the appearance and patterns of motion are important. This paper outlines a system for detecting, recognising, and summarising diverse human actions.The multiple action detection method takes the silhouettes of human bodies and then uses motion detection and tracking to create a unique sequence for each one.Based on the Each of the recovered sequences is then separated into shots based on the similarity between each pair of frames show the sequence's homogenous activity. The activity is identified by looking at thedata oriented gradient (HOG) histogram of the frames in each shot's Temporal Difference Map (TDMap). Comparing generated HOG to previous HOGs in the early stages of training. Making use of the TDMap image and a proposed CNN model, we can additionally recognise the action. Action summarising is carried out for each person found.

***Keywords*---**Disease prediction, Artificial Intelligence, Expert Systems, Machine Learning Algorithms, Future Recommendations.

I. Introduction

Human action detection in videos is becoming a hot topic in computer vision since it aids in video Management, summarization, and retrieval are all tasks that must be completed. The constructionrepresentation of local invariant characteristics and bag of features has made significant progress during the last two decades. Because of disparities in action performance, backdrop settings, and interpersonal variables, the task is difficult. Two things must be recognised in order to interpret the activity in the video: action recognition and action localisation. The initial is the activity in the video, and the second is the location of the action of interest.The process of categorising films into one of several predetermined action classes is known as action recognition, whereas the process of determining the spatiotemporal content of a video is known as action localization. For action recognition, we break the videos into a variety of categories. ThePositive and negative samples are included in the training. We may put the videos to the test after the training. We currently must manually annotate the appropriate area of the video's action of interest throughout the training phase. Convolutional Neural Networks (CNN) extract information from each frame and combine it from multiple frames to generate a video-level forecast. This method has the drawback of not gathering enough motion data.

CNNs (Convolutional Neural Networks (CNNs) are a type of deep neural network that uses convolutional neural networks use a layering and filtering strategy to effectively classify objects. Some of the issues in activity recognition can be addressed with thethe application A recurrent neural network is a type of neural network that uses recurrent neural networks to solve problems (RNN). When it comes to RNNs, fact, have a recursive loop that saves the data gathered in earlier times. The RNN only keeps track of one prior step, which is considered a drawback. As a result, LSTM was developed to keep track of information from multiple phases [8]. .In order to deal with the two fundamental challenges of bursting and disappearing gradients, Long-term dependencies should presumably be preserved by RNNs, whereas LSTMs should not handles these concerns more efficiently [1, 14,].

In the parts that follow, we go through our concepts in greater depth and provide additional information. The following is how The remainder of the paper is laid out as follows: This section is about that follows looks at some one of the most significant studies in the subject The proposed approach and processes are explained in the next section.We'll go over the experimental and assessment outcomes before moving on to the last section, where we'llCompare and contrast them with eight cutting-edge techniques. The report's final portion concludes with recommendations for future research.

II. Related Works

Action recognition features can be manually extracted or learned automatically. using a deep learning model from start to finish. Apart from Processed photographs have extracted attributes straight from raw photos received a lot of interest as well. For motion recognition, Watanabe et al. use the MHI to obtain shift invariant characteristics. To remove unrelated motions, the authors

presented a filtered global and local motion approach. As basic representations, MHIs are utilised.Two action recognition multi-modal fusion approaches are proposed by Ni et al. The feature representation is made up of STIPs and MHIs are spatial-temporal interest points. When it comes to feature quantization, the author employs sparse coding with limitations, as well as For feature representation, temporal pyramid matching is used as well as feature representation via temporal pyramid matching Dictionary learning has two restrictions: Geometry restrictions and group sparsity Chen and his coworkers TheDMMs (depth motion maps) are projections of motion patterns onto three reference planes characteristics. To compactly express them, local binary patterns are used.There are two methods for fusion that have been proposed. Wang et al. use depth maps to try to recognise human actions using shortdatasets for training As input, the model uses weighted hierarchical depth motion maps extracts features using a three-channel CNN.

In features are extracted using a pre-trained deep CNN, followed by action recognition using aSVM and KNN classifiers are combined Using a short training data set,For action recognition, a pre-trained CNN on a large annotation data set can be supplied. As a result, deep CNN transfer learning could be a good option for training models with small datasets.HAR is represented in an unique MRST stands for Spatio-Temporal Mutual Reinforcement Convolutional Tube. By emphasizing on informative spatial appearance and temporal motion while decreasing the structure's complexity, the uses the interaction of spatial and temporal information, the The model enriches 3D inputs by decomposing them mutually into spatial and temporal representations.

Over fitting will be prevented by expanding the size of the data set; however, giving a significant It's difficult and expensive to manage a vast volume of data that has been annotated The term "transfer learning" refers to the ability to suitable in these circumstances. proposes a method for creating a new architecture based on a pre-trained model that has proven to be successful. Over fitting will be prevented by expanding the size of the data set; however, giving a significant It's difficult and expensive to manage a big number of annotated documents data. In these situations, transfer learning is appropriate. [38] presents a way for constructing a new architecture built on the foundation of a previously successful pre-trained model.

Many applications,Human action recognition is used in applications such as video surveillance, video indexing and retrieval, sports applications, and multimedia are all examples of video surveillance are all examples of video surveillance are all examples of video surveillance. Many more employment are possible benefit from action detection, recognition, and summarization.Recognizing and interpreting player stances, for example, assists judges to make appropriate conclusions the When it comes to Fouls by players, particularly in football games, necessitate accurate decisions in a variety of situations are required are required For the summarising of actions, several techniques have been presented. developed a method for summarising sporting poses in self-recorded RGB-D movies. Because games feature a series of complex activities, the authors picked games to assess the approaches' performance. Deep neural networks are used in an improved version of this approach can from a

video, extract two sets of action-related data clips and classify them as engaging or uninteresting. The writers made a presentation.strategy for recognizing behaviors that can result in a variety of interesting and useful summaries There is also a reciprocal responsibility that acknowledges the other prepared video summary actions. The creators did this by combining a SVM framework with a latent structure action inference technique. Methods of video summarization in [1–9] If the objective of the summarization is not indicated, selecting Without examining the video's content, key frames or brief fragments are used may result in the loss of information. In addition, for these methods, summarising is done on films that have different scenes changing over time, such as movies video. If the videos contain some relevant information, summary based on scene variety is still insufficient.

Video summarization is used in content-based techniques for particular scenarios that describe the aim of summarising, such as summarization of an anomalous a possible event in a monitored scene [11, 15]. Finding a general system that can manage all types of events, on the other hand, remains a challenge. The majority of action-based summarising methodologies are found in for which there are a few research studies, arerestricted to a summary of sporting events Furthermore, the summation of many activities carried out at the same time isn't addressed.

Action Recognition

Feature trajectories have been shown to be an excellent way to represent video. However,These trajectories were both insufficient in terms of quality and quantity. Wang et al.[1] advocated using dense trajectories when dense sampling became popular for picture classification to describe films. TheEach frame's dense points are sampled and traced using displacement data. To boost performance, Wang et al.[2] take camera motion into account.By matching To assess the density of optical flow, Camera motion is based on feature points between frames, SURF descriptors, and dense optical flow. To replicate the motion connection, another method was adopted. The approach does not require exact foreground-background separation because it is based on visual codewords created from local patch trajectories.

Action Detection

To conduct recognition, the The density of optical flow is estimated using Between-frame feature points, SURF descriptors, and dense optical flow were used. SVMs and SVMs AdaBoost are two recent prominent methods for action recognition that employ machine learning techniques to incorporate the information providedin a series of training scenarios proposes the Action MACH filter, an action recognition approach based on templates that synthesises a single Action to capture intra-class diversity. [5] proposes another approach SMILE-SVM is a method based on inaccurate action location for learning human action detectors (Multiple Instance Learning Support Vector Machines with Simulated Annealing). A figure-centric visual word representation was employed by Wang et al. [6]. In order to recognise the action, localization is treated as a latent variable in this scenario. It is necessary to learn a spatiotemporal model. Model parameters are estimated during training, and the appropriate percentage is

determined [7]. To identify human actions from background motion, an independent motion evidence element was included.

Methodology

We divided this part into numerous subsections because the process has multiple steps and sub-modules. The model's overall architecture is discussed first for a general comprehension, The learning phase comes next, followed by the transfer learning phase. multi-modal The task is accomplished using a multilayer model. Multi-modal refers to the use of more than one modality. Our modalities are text (the script matching to the action) and video of the action, as described above.

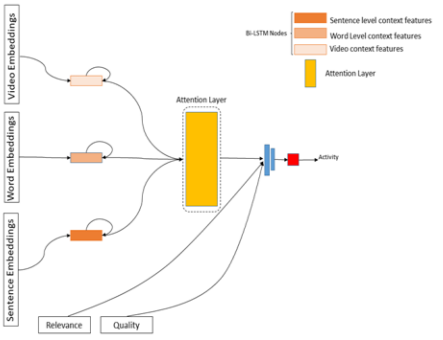


Fig 1.The architecture is multi-modal

Literature Survey/Project Design

Human action recognition has been studied extensively for many years. The majority of action recognition systems include manually annotating the relevant area of the video'sactivity of curiosity It has been researched in recent years whether the relevant component of an activity of interest may be automatically discovered and recognised. We can go over the many methods of action recognition. Its versatility allows it to be used in a variety of areas, including health, security and surveillance, entertainment, and intelligent settings, Identification of human activities has been increasingly common in recent years. Human activity recognition has gotten a lot of interest, and researchers have tried a range of approaches, including wearables, object-tagged approaches, and device-free approaches. We give a detailed assessment of work done in many domains of human activity recognition from 2010 to 2018, with a Concentrate on solutions that do not require the use of a device. Because the person is unique not obliged to carry anything, the device-free approach is gaining popularity. Instead, devices are placed throughout the environment to collect the necessary data. We propose a new taxonomy for activity recognition research that divides it into three categories: action-based, motion-based, and interaction-based. These categories are broken down into ten sub-topics, with the most recent study findings shown in each. Unlike in the past, studies that concentrated on a particular type of vehicle, this one looked at all types of vehicles activity, we include all of the activity recognition sub-areas and give a comparison of the most

recent research effort in these sub-areas to the previous research effort in these sub-areas best of our knowledge.Specifically, we highlight the work's essential characteristics and design techniques. Then, using ten key criteria, we conduct an in-depth analysis to provide the reader with a comprehensive study of current issues approaches and developments in numerous Recognized sub-areas of human activity Finally, we talk about unsolved research problems difficulties Recognition in the area of human action as well as future research goals.

Functionality Working of Project

Proposed methods

In this study, we offer a new approach for identifying and identifying and identifying and identifying and identifying and identifying and identifying and identifying and summarising several human actions in combination. We begin by employing a background-subtraction-based technique that has been proposed detect human bodies strategy in our system. After then, each one was observed separately in order to create a video sequence of each person in the scene throughout his time there.The training section was developed to depict all types The target's motion history is represented via a series of Temporal Difference Map (TDM) Histograms of Oriented Gradient (HOG). We took a series of each moving individual from the scene for the action recognition stage.Then, using a histogram of similarity between frames, shots representing homogenous sections of the series were chosen. The transactions in the sequence are represented by the peaks of the histogram, andThe subsequence between each pair of peaks defines a shot. Then, using a training set of HOGs for each action, Each shot was used to determine the planned action. The categorising of actions is done in two ways:(1)The The current action's HOG and the training set's HOG are compared using the cosine similarity measure.(2)Using TDMap photos as input, a To classify actions, a convolutional neural network (CNN) model is used.

3.1 Pre-processing

Action detection technologies now on the market are made to recognise human actions in a single-person situation. The This paper proposes the detection and recognition of a large number of human bodies. The purpose is to obtain each person's information silhouette individual while they are present in the scenario. Then, using a background-subtraction-based strategy, we employ motion detection to Make detection masks for the stuff we find and identify all of the persons in the scenario. We begin the background subtraction approach by inputting the N first frames from the movie into the background model, which is based on each frame being divided into 1616 pixel blocks The humans are created once the backdrop model is created discovered in each frame of the video using object segmentation and background subtraction.

A bounding box is used to keep track of each individual based on observed masks. The Kalman-based tracker is employed in this study, which employs the To estimate the each track's centroid in the current frame and update its circumference, use the Kalman filter as a result. The enhancedThe Each recognised individual's silhouette is eliminated to create a new sub-sequence of

the silhouette in question. while their presence in the scene is tracked by a modified version of to the extraction of Oriented Gradients, displays the proposed method for motion detection and tracking pictures

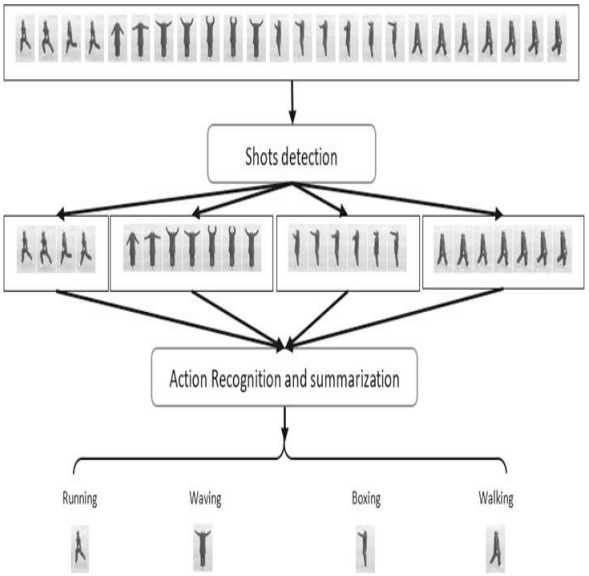


Fig 2. Steps for summarization based on recognised activities

Action Recognition

Feature trajectories have proven to be effective in depicting video. These trajectories' quality and quantity, however, were insufficient. Wang et al.[1] advocated using dense trajectories to represent films when dense sampling became popular for image classification. Each frame's dense points are sampled and traced using displacement information. In order to boost productivity, The Wang et al. take camera motion into consideration. [2]. By matching To assess the density of optical flow, feature points between frames, SURF descriptors, and dense optical flow are used camera motion. Another method [3] was used to simulate the motion connection. Because the method is based on It is not necessary to have a precise foreground-background separation when using visual codewords created from local patch trajectories.Another strategy for locating informative regions was proposed by Dorr et al.[4]. They employed algorithms for saliency mapping. This research suggests the use of a collaborative learning paradigm for determining the a specific action's spatial and temporal scopes a novel strategy.

Action Detection

SMILE-SVM stands for Multiple Instance Support Vector Machines with Simulated Annealing, and it's a multi-instance learning method based on inaccurate action location for learning human action detectors. A figure-centric visual word representation was employed by Wang et al. [6]. Localization is considered as a latent variable in this case in order to recognise the action. It is

learned a spatio-temporal model.The relevant fraction is selected and model parameters are estimated during the training[7]. For detecting human actions from background motion, [8] presented an independent motion evidence element. The majority of the solutions necessitate the use of bounding boxes to annotate the relevant area of the video. It took a long time for humans to intervene. To get around The video was split into subgroups by Brendel et al.[9]. using the bounding box and then constructed a model to identify the appropriate subgroup. This work provides an overview of a strategy for improving detection It understands both geographical and temporal scales To show human action, a dense trajectory is used as a local feature.

Generatingdense Trajectories

We offer a library of instructional videos. Our goal is to find out what action was taken as well as the location of the corresponding action in the video. The relevant portion is automatically identified by our approach. All we have to do now is annotate whether it's a positive or bad training video.Local patch trajectories will be used to produce the video depiction. Wang et al. created dense trajectories and demonstrated that it outperformed KLT tracking of sparse local features. As a result, we use a dense trajectory technique in this paper.We use the Big Displacement Optical Flow Tracker for extracting trajectories in order to represent the video since it can track objects with quick motion and large displacement. The grid step size is set at 5. We calculate four types of descriptors Each trajectory has four sorts of descriptors: trajectory, MBH, HOG, and HOF. It is possible to do so while searching, calculate the camera motion between two frames for the three sorts of descriptors.The second frame is distorted due to camera movement. As a result, calculating The optical transfer between frames and descriptions is straightforward. The camera motion is estimated using the changing object borders in the foreground. The typical BOW technique is utilised to encode the local features. After that, the descriptors of trajectories in a movie are quantized using the k-means technique. The nearest vocabulary term is assigned based on Eucledian distance when quantifying. Finally, the RootSIFT technique normalises codeword occurrences,and the video description is the result of their concatenation.

Generating Object Level Segmentation

To create dense trajectories, On a grid with regular spacing, feature points are densely sampled and tracked at various spatial scales. In a dense optical flowfield, median filtering is used to track each point over future frames. We assume there are F frames in the input video and extract the P feature point trajectories. T(p) is separated into p=1P, is subdivided into a number of categories. each with a high degree of resemblance. Make a matrix out of Take T(p) and break it down into M and N are two characters. Using DCT, we can reduce Error and N(p). Calculate the affinity matrix N(p)[9]. The scene should then be divided into foreground and backdrop. The backstory has a reduced dimensional space. The foreground element N(p),p can be retrieved. after multiple iterations. Clustering foreground trajectories into an undetermined number of groups is the final step in our split stage. We receive a foreground moving item after numerous iterations. As a result, we've created motion-consistent clusters,

each of which is linked to a specificIn the scene, there is a foreground moving object.

Training and Testing the Videos

Learning Videos

We suppose we have M1 is a positive training film, while M2 is a negative training film. for each action. To handle the trajectories collected from each positive training video, we apply our split-and-merge trajectory approach This technique has the ability to be used to obtain positive training content segmentation at the object level film. Following that, we'llTo understand the geographical and temporal extents of the positive instances' behaviours, utilise the training set to train an SVM model. The M1 positive training videos Positive training films (V1+, V +,...VM1+) and M1– –

Testing the action

When you're given an action video, you must figure out What is the action in the film and where is it? We use the divide and merge method for the trajectories test video. The movie will be segmented at the object level after this operation. We'll be able to distinguish between several front moving items. The conventional BOW technique is utilised to create descriptors for each moving object in the foreground The video for the test is descriptor is d(Vt, I sti, eti), where i=1N is the number of moving things in the front and N is the number of moving items in the background. the number of moving things in the background. Multiple descriptors are created if a foreground object is detected.

The M2 negative training videos include V1, V2,...VM2. Suppose The foreground moving items in V + are N.s denotes the frame to begin with The matching score of the description with the highest matching score is used iii
Let ei represent the last frame and li represent the action's performer. li can be a part of a moving item in the foreground. The spatial and temporal extent is represented by these factors.The video descriptions are obtained using the usual BOW method. d(Vi+,li,si,ei) is the symbol for it. After that, it's fitted to an SVM is a statistical learning model.
pick how long the score

(V +,l,s,e)=K(w, d(V +,l,s,e))+b

action will last.

Data use

We ran benchmarking tests on the two most cutting-edge human posture estimation models, OpenPose and AlphaPose, in this section. We experimented with several modes in single-person and multi-person scenarios.tf-pose-estimation-based real-time multi-person human action recognition The following is the pipeline: tf- pose-estimation allows for real-time multi-person pose

estimation. Extraction of Features TensorFlow / Keras for multi-person action recognition.

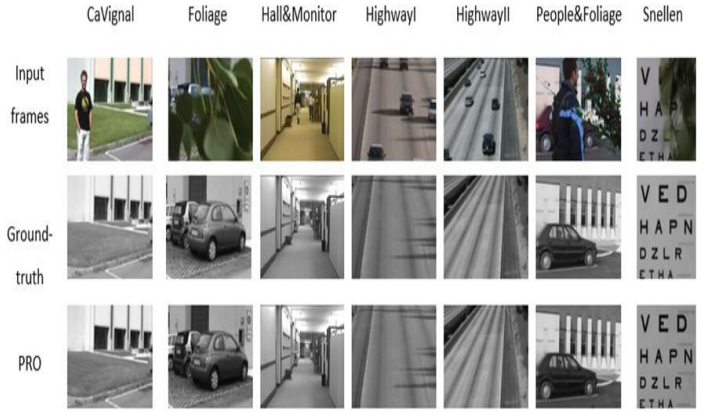
We used our laptop's webcam to acquire 3916 training photos for training the model and categorising five actions: squat, stand, punch, kick, and wave. There is only one individual performing one of these five acts in each training image. The films are captured at 10 frames per second with a frame size of 640 x 480 pixels and stored as pictures. To recognise the human pose in each training image, we used tf-pose-estimation. OpenPose's output skeleton format is available at OpenPose Demo - Output. The training data files are saved in the data folder:

1. original data (skeleton raw.csv)
2. skeleton filtered.csv: filtered data with missing postures removed.

We developed an Online-Realtime-Action-Recognition-based-on-OpenPose Deep Learning model. Keras and Tensorflow are used to implement the model in training.py. To conduct a 5-class classification, the model has three hidden layers and a Softmax output layer.

Results

The proposed method's performance is assessed in this section. The suggested background modelling strategy is evaluated on the Two background-subtraction-based techniques and the SBI dataset are compared. A description is also provided for the created multiple human action dataset.The When the suggested dataset, as well as the PET09 dataset, are utilised to test The Weizmann and UCF-ARG human action recognition datasets are utilised to train the proposed approach, as well as the performance of the multiple action recognition approach technique. The suggested action recognition algorithms' performance is also discussed. Two methods are used to recognise actions that have been detected : (1) Using the A measure of cosine similarity between TDMap HOGs A categorization is used in the training set and the current activity to recognise of actions termed (CS-HOGTDMap) is created. (2) Action classification utilising the suggested CNN model trained on TDMap pictures, dubbed (CNN-TDMap).

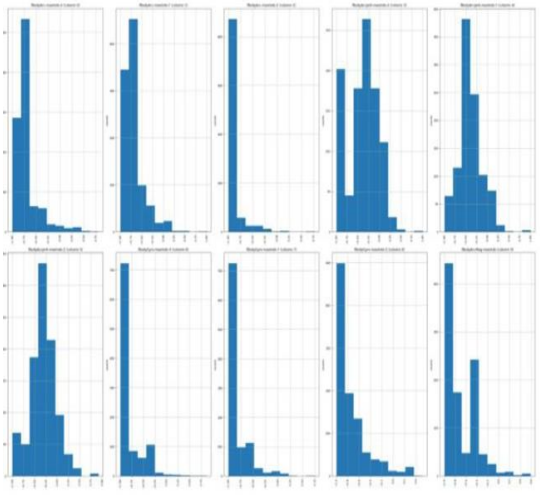


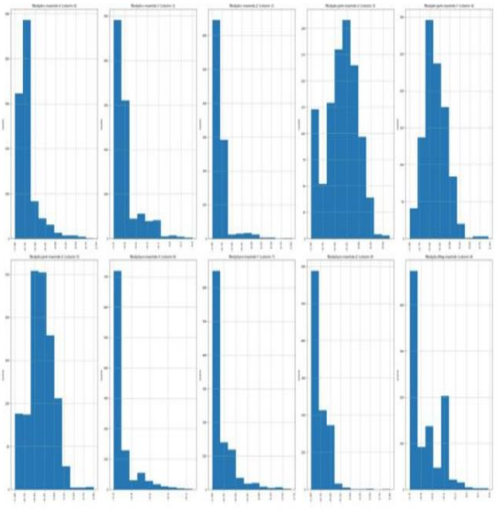
Results from the SBI dataset using the offered method as a backdrop

The The suggested background is tested using the SBI dataset approach modelling. The created background Figure depicts the proposed approach in action. The results are convincing, and all videos have a background that is generated without false ghosting utilising our technology. For the background of FThe suggested method is oliage and People & Foliage sequences was successful in estimating Even when the scenes were full with moving items, the background worked well throughout the film.We used a variety of metrics to combine the visualised results, Total number of Error Pixels (EPs), Error Pixel Percentage (pEPs), and Clustered Error Pixels (CEPs), Color image Quality Measure (CIQM), Peak-Signal-to-Noise-Ratio (PSNR), MultiScale Structural Similarity Index (MS-SSIM), and others (CQM). compares the results with the results of two backdrop modelling methodologies, [54] and [43], respectively, from IMBS-MT. In most Some of the proposed names were HighwayI, Hall & Monitor, Sallen, and Foliage. dataset movies succeeds in modelling the When compared to other procedures, it has an excellent history with good results.

The collected findings are Against The proposed method was compared against state-of-the-art methods in order to evaluate it approaches. The majority of action recognition systems work with data that just comprises There is only one individual doing something. When more than one individual is present in a scenario, the recognition process may perform badly. Furthermore, most systems are aware of the action by looking at the silhouette of the moving object. As a result, a poor binarization or segmentation can affect the recognition accuracy. Our technique identifies various actions of multiple actors to solve all of these issues. The background is used to compute the Each moving human body is represented by a mask movement. The After that, A histogram of oriented gradients for each variable (HOG) is calculated retrieved silhouettes and compared to all of the histograms created during the training phase.

The Based on the training phase, the distance between HOGs of each extracted target is calculated outcomes to recognise human actions. The histogram of the temporal difference's directed gradient The video's map frame is calculated during the training phase. As a result,We have a set of HOGs for each action, each of which represents one possibility of that activity. We do the arithmetic the right way HOG of the identified sequence during the testing phase. The distance indicated above is then used to compare HOG to all other HOGs from the training phase. The action is represented by the smallest All distance values are multiplied by this value. The datasets from KTH, Weizmann, and UCF-ARG are available compared to determine the usefulness of the suggested technique. Furthermore, we do a test between the videos in each dataset.





Shows the accuracy rate for each strategy, and As may be observed, the proposed strategy yields positive outcomes improved and more efficient The The adoption of the new data representation is related to the outcomes. An improved version of the proposed technology can be used to detect and recognise various human actions in real time.The suggested algorithm extracts the sequence of each individual in the picture and applies it to it. The proposed method is assessed based on our dataset. On the MHAD and PET datasets, the detected person and their behaviours are depicted in Fig. 10, which shows the obtained results. One sample from the The visualisations include the PET09 dataset as well as three films from our dataset. Each person's actions are detected, subsequenced extracted, recognised, and summarised by humans are all part of the proposed approach's validation process.The UIUCI dataset is also used to test the actions that have been taught some of the outcomes. Because the used datasets contain the same type of action, these movies The training phase does not include them. The recognition percentage for the UIUC dataset video was 98 percent.We employ

two datasets for human-human interaction:IXMAS and UY-interaction are two examples of IXMAS and UY-interaction. The results of human interaction detection and recognition. In the IXMAS dataset, for example, During the training phase, we show a video of two people fighting. The acknowledgement in four movies collected from different fields of view is illustrated in Fig. 11b, which demonstrates the results (FOV). In addition, Some results from the UT-interaction dataset are shown that illustrate various activities that have been found.

The suggested algorithm's accuracy is determined by the The A human body has been discovered on the site, tracked, and segmented. Furthermore, the detected individual may be in a static position, needing a huge number of movies during to represent all of the behaviours in the training phase the environment. various situations.This paper defines a video summary as a collection of action labels and keyframes from a video in which the actors perform the action.The summary offered in this study uses human activities to divide the the behaviours of each and every person who is present in the scene The whole image summary is built based on the findings of each object's recognition and extraction activity.

Conclusion

A new method for detecting, recognising, and summarising several human beings has been developed in this work, in which the activities of a person appearing in a scenario were summarised. Each person's motion/actionseach human body silhouette is identified, monitored, as well as a collection of human body silhouettes created formed for a certain scene.The process for detecting shots was created with the goal of determining the a series of acts in a sequence that has been generated after detecting and summarisingThe purpose of the shot detection technique was to figure out which activities are included in the created sequence. were two processes to recognising each activity.First, aIn each sub-sequence, a training dTDMaps had a set of HOGs employed. The HOGs that were created were then calculated and put to use during the testing phase to represent distinct scenarios for each action. The recognition is done by determining the shortest distance between the present action's HOG and the training's HOGs Furthermore, the calculated TDMaps pictures are employedIn a CNN model, the action is also classified. By overlaying each recognised individual's image on all photos during his presence in the scene,a summary was created.Multiple The proposed method could be used to identify and recognise human action technique. Because of its simplicity and the quantity of features it employs, This algorithm is capable of being used in real time.

Reference

1. Almeida J, Leite NJ, Torres R. d. S. (2012) Vison: Video summarization for online applications. Pattern Recognit Lett
2. Almeida J, Leite NJ, Torres R. d. S. (2013) Online video summarization on compressed domain. J Vis Commun Imag Represent 24(6):729–738
3. Almeida J, Torres R. d. S., Leite NJ (2010) Rapid video summarization on compressed video. In: 2010 IEEE International Symposium on Multimedia (ISM). IEEE

4. Angel B, Miguel L, Antonio J, Gonzalez-Abril L. Mobile activity recognition and fall detection system for elderly people using ameva algorithm. Pervasive Mob Comput. 2017;34:3–13.
5. Anuradha K, Anand V, Raajan NR. Identifcation of human actor in various scenarios by applying background modeling. Multimed Tools Appl. 2019;79:3879–91.
6. Bajaj P, Pandey M, Tripathi V, Sanserwal V. Efcient motion encoding technique for activity analysis at ATM premises. In progress in advanced computing and intelligent engineering. Berlin: Springer; 2019. p. 393–402.
7. Chaquet JM, Carmona EJ, Fernández-Caballero A. A survey of video datasets for human action and activity recognition. Comput Vis Image Underst. 2013;117:633–59.
8. Chavarriaga R, Sagha H, Calatroni A, Digumarti S, Tröster G, Millán J, Roggen D. The opportunity challenge: a benchmark database for on-body sensor-based activity recognition. Pattern 3(4):397–409
9. Lopez, M. M. L., Herrera, J. C. E., Figueroa, Y. G. M., & Sanchez, P. K. M. (2019). Neuroscience role in education. International Journal of Health & Medical Sciences, 3(1), 21-28. https://doi.org/10.31295/ijhms.v3n1.109
10. Martins GB, Papa JP, Almeida J (2016) Temporal-and spatial driven video summarization using optimum-path forest. In: 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images SIBGRAPI). IEEE
11. Mehmood I et al (2016) Divide-and-conquer based summarization framework for extracting affective video content. Neurocomputing 174:393–403
12. Mei S et al (2015) Video summarization via minimum sparse reconstruction. Pattern Recognit Lett 48(2):522–533
13. os Santos Belo L et al (2016) Summarizing video sequence using a graph-based hierarchical approach.Neurocomputing 173:1001– 1016
14. Recognit Lett. 2013;34:2033–42. Chen BH, Shi LF, Ke X. A robust moving object detection in multi-scenario big data for video surveillance. IEEE Trans Circuits Syst Video Technol. 2018;29(4):982–95
15. Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2022). Post-pandemic health and its sustainability: Educational situation. International Journal of Health Sciences, 6(1), i-v. https://doi.org/10.53730/ijhs.v6n1.5949
16. ujatha C et al (2014) Multilevel Framework for Summarization of surveillance videos. In: 2014 Fifth International Conference on Signal and Image Processing (ICSIP). IEEE
17. Xu Q et al (2014) Browsing and exploration of video sequences: A new scheme for key frame extraction and 3D visualization using entropy based Jensen divergence. Inform Sci 278:736–756