RESEARCH REVIEW JOURNALS

Volume-03
Issue-10
October -2018

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com [UGC Listed Journal]

Agricultural Activity Pattern Recognition System Based on Deep Learning

Manpreet Singh Gill

Asst. Prof. & Head, Dept. of Computer Science, Akal Degree College, Mastuana (PB.) INDIA

ARTICLE DETAILS

Article History
Published Online: 10 October 2018

Keywords
Agricultural, Pattern Recognition System, Deep Learning

***Corresponding Author**
Email: gillkotra@gmail.com

ABSTRACT

There are many activities are the farm level which cannot be simply detected and measured with sensors only and for that we need to use video based systems which can capture a video stream and analyze what is happening in context to agriculture pretext[10]. The many automations and AI based future tools will work in the company of humans and will have to recognize the basic farm and agriculture-based activities to make better and effective decisions.

Overview

Video analytics is a process of analyzing video feeds and understanding temporal and spatial events. Through video analytics, we understand the events happening in a video scene and try to predict and evaluate these scenes based on some predefined models and processes. Video analytic is an essential part of the surveillance that is used to surveillance the area like a security guard with a very minimal involvement of human[1] [2]. The main aim of video analytic is to find any non-appropriate incident from the video automatically.

But in the current stage, video analytic is growing day by day and spread its applications in various domains. By using this technology, security-oriented companies are designing and developing various applications and hardware products like active cameras, smart digital video recorders and many others to maintain security in various public or private domains [3] [4] [5]. Now video analytics not only used to find out the abnormality or irregularity from a video clip even it has the capability to drawing out the meaning of that particular video clip.

By concentrating on these fruitful improvements in the field of passive video analytics, now, this field is known as Video Content analysis and spreading their roots to various domains like machine intelligence, computer-vision, healthcare, smart assessment etc[6]. Video analytics is using various artificial intelligence-based algorithms to make able to see and interpret the scene to make decision on the basis of that event. The meaning of event understandability is completely different from the motion detection and recognition. This domain not only classifies the motion of the subject but also understand various contexts to track the object through the entire scene. Following these major advantages, now video analytics is moving to the area of healthcare or assessment. Various cameras are used to capture videos to detect the presence of the subject to generate the meaning for their activities.

The age of Artificial Intelligence and machine learning is here and is being used in the various fields of agriculture[7]. Companies are utilizing sensors and different IoT based advancements to screen yield and soil wellbeing [8] [9]. Various AI and machine learning instruments are being utilized to anticipate the ideal time to sow seeds, get cautions on dangers from bothering assaults, and that's only the tip of the iceberg. Companies are utilizing information streams originating from numerous sources to manufacture an effective inventory network for production management and control.

Literature Review

Activity recognition is a combination of various preprocessed steps. We divide our literature according to all those steps which we need to cover to achieve a proper output. The base step for action recognition system is object detection and classification. Isolated the examination in 3D and 2D (with and without express shape models) approaches. In [46], another scientific categorization was introduced concentrating on human movement examination, following from a solitary view and recognition of human exercises and multi-view cameras. In soul to the past scientific categorization, Wang et al. (2003) [47] presented a comparative of various leveled activity classification chain of command. The review of Moeslund et al. (2006) [48] essentially centered around present based activity acknowledgment techniques and proposed a fourfold scientific categorization, including instatement of human movement, following, present estimation, and acknowledgment strategies.

A fine partition between the implications of activity and action was proposed by [49], where the action acknowledgment techniques were ordered by their level of movement many-sided quality. Poppe (2010) [50] portrayed human action acknowledgment techniques into two principle classifications, depicting them as best down and base up. Then again, Aggarwal and Ryoo (2011) [51] displayed a tree-organized scientific categorization, where the human action acknowledgment techniques were arranged into two major sub-classifications, the various leveled approaches and the one layer approaches, each of which has a few layers of classification.

Concept of Video Analytics

Video analytics are being utilized for a scope of utilizations, such as, reducing shrinkage, improving public safety, providing operational intelligence and increasing security, across several vertical markets [1]. Video Analytics is a

robotized way to deal with break down and oversee video, without the cost or worker hours beforehand required. Video Analytics can be coordinated with other data and security frameworks to make new conceivable results for overseeing and using video information. Video Analytics software can most likely be layered onto your current surveillance system, so including the capability is an incremental improvement [2].

Video Analytics is empowering a rapidly creating number of implanted video items, for example, insightful advanced video recorders (DVRs) and savvy cameras with robotized abilities that just a few years back would have required human checking. Extensively, video investigation is the extraction of significant and important data from the advanced video [3].

For most video analytics applications, it presents the essential processing path, on the other hand the objective is to automatically comprehend the activities happening in a monitored scene. Subsequently, the performances of these applications are significantly reliant on the precision and robustness [4].

Video preparing can be portrayed as the examination of the substance of the video to procure an appreciation of the scene that it delineates. It is a fundamental segment of various innovations, including sight and sound, video observation and mechanical autonomy. From an essential science point of view, techniques in video investigation are persuaded by the need to make machine calculations that can imitate the capacities of human (and another creature) visual frameworks [5].

Developing health informatics systems guarantee to reform health and health service programs around the world. However, changing this expectation filled vision into a reality will require a tremendous exertion from countless authorities, medical professionals, software engineers, designers and analysts [6].

Detection, Tracking And Segmentation of Objects

A clip of a video reveals many facts which contain interesting and useful information. With the help of that information, we can serve mankind in many ways. Researchers and professionals are motivated to study and explore the information, which captured in the conditioned and unconditioned environment. There are various types of methods which describe the different techniques of human detection, tracking and segmentation. In this part of literature review, we will discuss various kinds of methods which define the process of detection, tracking and segmentation.

Detection and Classification Of Objects

Over the ongoing years, identifying individuals in a video scene for observation framework is drawing in more consideration. The explanation behind considerably more consideration is because of its extensive variety of uses in fall detection for elderly people, human step portrayal, and individual counting in an unusual occasion discovery, individual ID, sexual orientation order, and so on. Generally, human or object detection is classified into 2 categories: object identification and object categorization. Object recognition is performed by foundation deduction, optical stream and spatiotemporal separating. What's more, question arrangement could likewise be performed by a Motion-based Method, Texture-based Method and Shape-oriented technique [14].

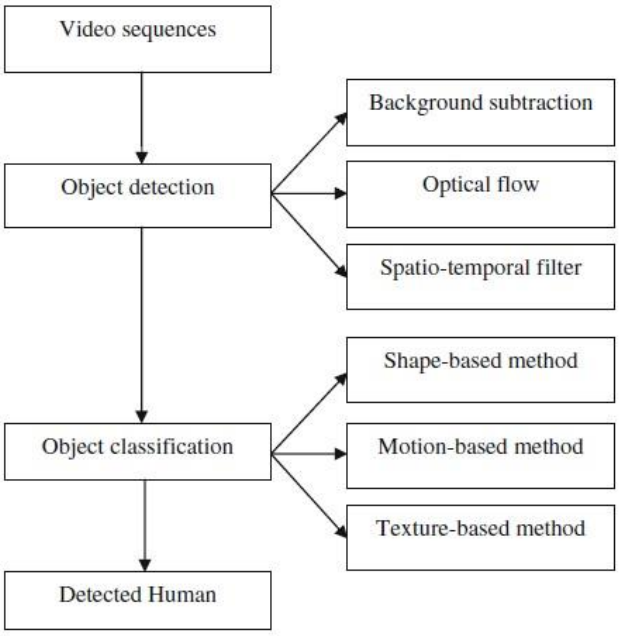

Fig 1. Techniques of object detection and classification

Background Subtraction Techniques:

There are two or three open approaches to manage to perform subtraction. The most well-known ones are a warping background [19], temporal differencing [18], hierarchical background [20], non-parametric background [17], and adaptive Gaussian mixture [16] models.

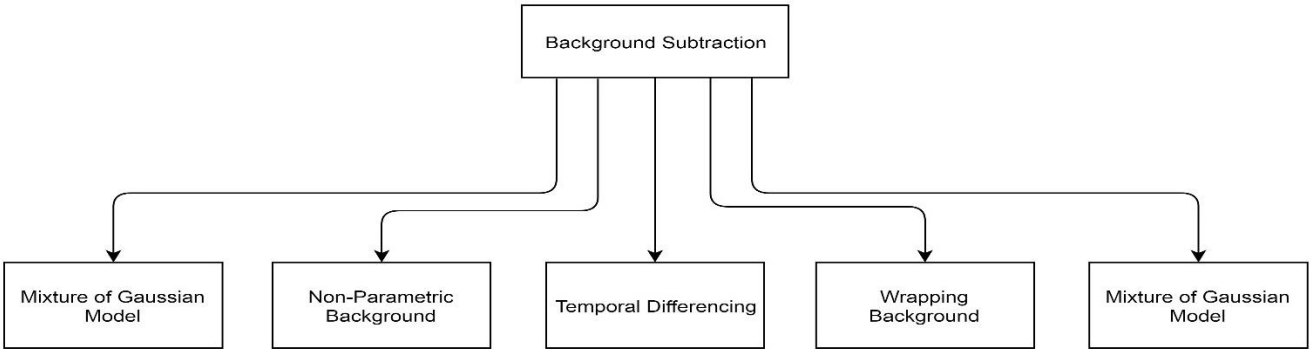

Fig 2: Background Subtraction Techniques

Optical Flow:

That appraisals movement in the video by coordinating focuses over objects in a picture frame(s). Optical flow is utilized for the spatial smoothness and brilliance consistency to portray the lucid movement of focuses or highlights between picture outlines [14]. Optical flow- based movement division utilizes qualities of stream vectors of moving items after some time to identify moving locales in a picture arrangement. One major advantage of utilizing optical flow is: it is powerful to different and synchronous cameras and protest movements, making it perfect for swarm examination and conditions that contain thick movement.

Spatio-temporal filter:

To recognize motion in view of spatiotemporal investigation, the activity or movement is portrayed by means of the whole 3D spatiotemporal information volume spread over by the moving individual in the picture arrangement. These strategies for the most part consider the movement in general to describe its spatiotemporal disseminations [22].

Object Classification Techniques.

A protest in movement should be arranged precisely for its recognition as an individual [14]. The accessible order techniques could be separated into three principle classifications: surface based technique, movement oriented technique and shape-based strategy.

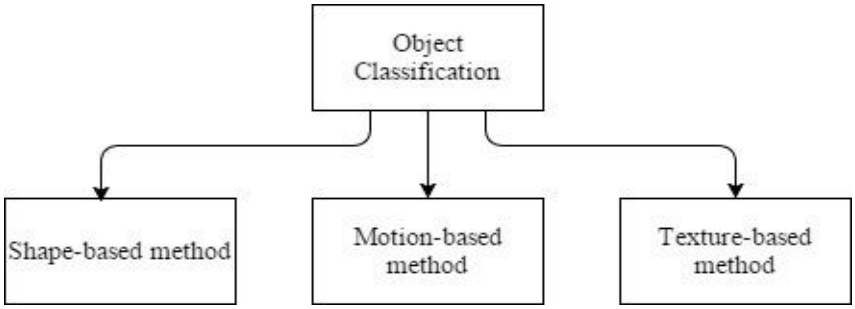

Fig 3: Object Classification

Object Tracking:

Tracking can be characterized as a issue of approximating the way of a question in the picture plane as it moves around a scene. The ways to deal with following the items are piece following, point tracking and outline [23].

Tracking are a portion of the difficulties that ought to be taken care in question following as depicted in [24]:

1. Difficult object motion,
2. Complex objects structures,
3. Imperfect and entire object occlusions,
4. Noise in an image,
5. Evidence loss caused by estimate of the 3D realm over a 2D image.

Choosing the exact highlights has a basic impact in tracking. Feature choice is entirely connected with the question portrayal. For example, shading is utilized as an element for histogram based portrayals, while for shape based portrayal, question edges are typically utilized as features [24].

1. **Color:** All video outline groups depend on various color spaces show. The information of various edge can be put away in different shading spaces going from dim scale, RGB, YCbCr and HSV color spaces. The information that is put away in each edge is the brilliance in each ghostly band. Color pictures are meant as red (R), green (G) and blue (B) layers or RGB. Other unmistakable shading spaces utilized in the territory of Object following are YCbCr and HSV [25].
2. **Edges:** Object limits are much of the time created solid changes in picture forces. These progressions are recognized by Edge identification. A key property of edges is that they are less touchy to brightening changes contrasted with shading highlights. The

majority of the calculation will utilize the edge as the fundamental specialist in object tracking.

3. **Texture**: Texture is a level of force uniqueness of a surface which lists properties, for example, smoothness and consistency. Contrasted with shading space show, surface requires a handling step. On premise of shading, the surface highlights are less delicate to enlightenment changes as same as to edge highlights.

Similar to object classification, object tracking is also divided into 3 subparts or techniques:

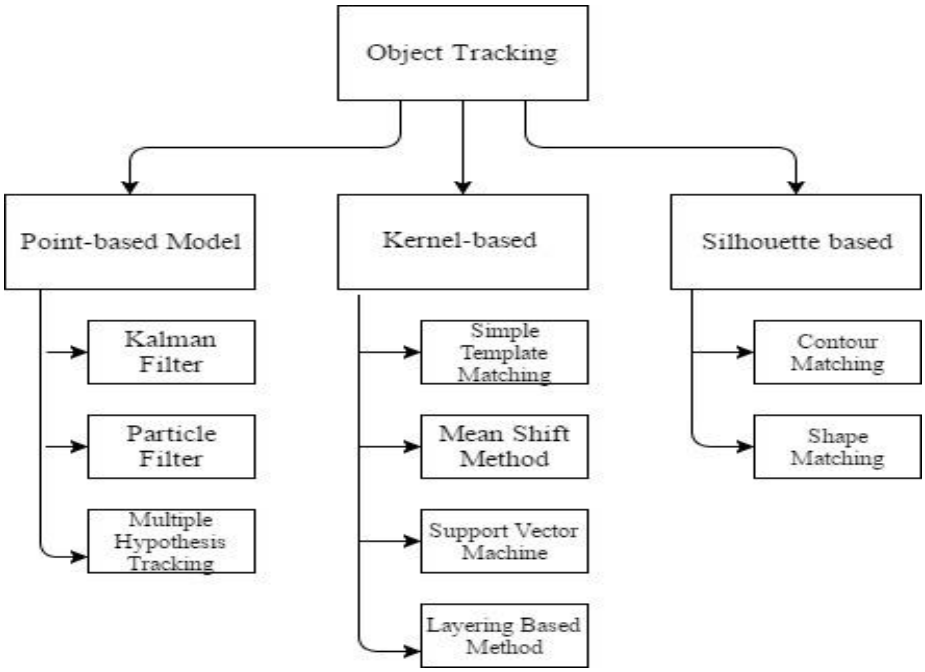

Fig 4: Object Tracking Methodologies [24]

Object Segmentation:

Segmentation is the way toward recognizing changes and extricating significant changes for further qualification and analysis. Pixels that have changed are named as "Foreground Pixels"; those that do not change are known as "Background Pixels". Along with that, segmentation is also named as "Background Subtraction". Pixels left after the background has been deducted are the foreground pixels. The degree of "change" which is used to recognize foreground pixels is a key factor in segmentation that can vary relying upon the application [3].

Image segmentation alludes to the standout amongst the most essential procedures of picture preparing. It is the way toward parceling or separating a picture into parts, known as portions. It is generally utilized for applications like object recognition or picture pressure on the grounds that for such applications, it is wasteful to process the complete picture thus the parts from the picture is utilized for additionally handling [26].

Object Segmentation Techniques:

There are distinctive existing strategies which are used for picture division. These all procedures have their own particular vitality. These all procedures can be drawn nearer to two noteworthy systems of division i.e. region based or edge-based methodologies. Each system can be related with various pictures to perform required division. These all systems besides can be planned into three courses of action [27], [28].

A. **Stochastic Segmentation Techniques:** The stochastic methods are those strategies of the picture division that works away at the discrete pixel estimations of the photo as opposed to the helper information of the zone [26].

B. **Structural Segmentation Techniques:** The fundamental methods are those frameworks of picture division that rely on the data of the structure of a required piece of the picture i.e. the required district which is to be divided [26].

C. **Hybrid Techniques:** The hybrid strategies are the blend of the two picture division procedures that utilization the ideas of both above methods i.e. these utilizations basic data and discrete pixel together [29].

The well-known procedures utilized for image segmentation are clustering based techniques, watershed-based techniques, thresholding method, region-based techniques, edge detection based techniques, artificial neural network based and partial differential equation based techniques etc. These all procedures are unique in relation to each other as for the strategy utilized by these for segmentation.

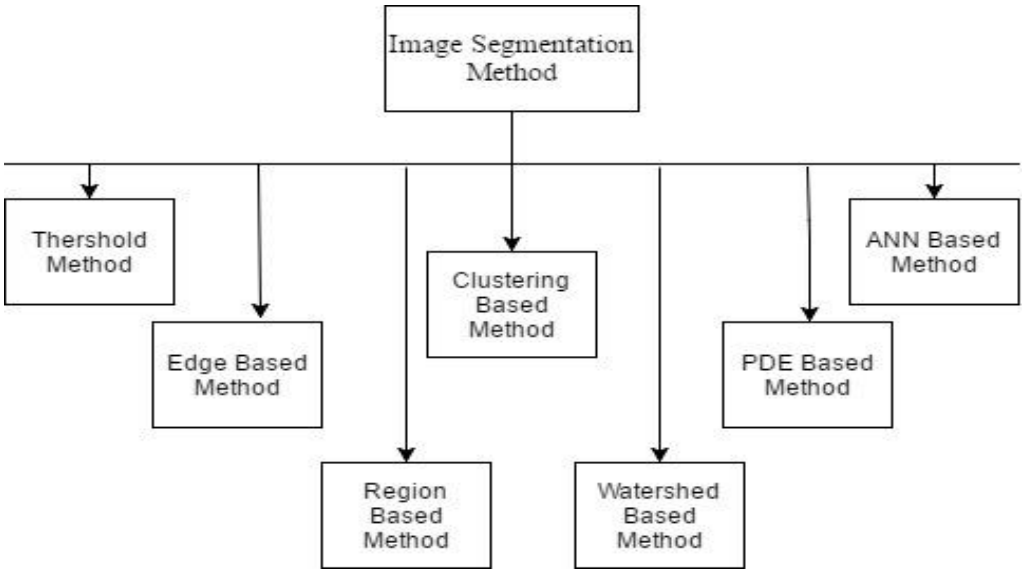

Fig 5: Object Segmentation Techniques

A. **Edge-Based Segmentation Method:** The edge detection systems are all around made strategies for image preparing alone. The edge-construct segmentation techniques are based with respect to the quick difference in power an incentive in an image on the grounds that a solitary force esteem does not give great data about edges. Edge recognition methods find the edges where either the principal subsidiary of power is more noteworthy than a specific limit or the second subordinate has zero intersections. In edge-based segmentation strategies, as a matter of first importance, the edges are distinguished and they are associated together to shape the protest limits to portion the required areas. The fundamental two edge-based segmentation techniques are Gray histograms and Gradient-based strategies. To recognize the edges one of the essential edge location systems like Sobel administrator, shrewd administrator and Robert's administrator and so on can be utilized. The consequence of these techniques is fundamentally a paired image. These are the auxiliary procedures in light of discontinuity detection [31].

B. Threshold Method: Threshold methods are the least complex strategies for image segmentation. These techniques separate the picture pixels regarding their intensity level. These techniques are utilized for pictures having lighter objects than the background. The choice of these techniques can be manual or programmed i.e. can be founded on earlier learning or data of image features [30].

C. Region-Based Segmentation Method: These are the strategies that fragment the frame into different locales having comparable attributes. There are two fundamental procedures took after [28], [32], [33].

1. **Region splitting and merging methods**: This strategy utilizes two fundamental methods i.e. combining and part to fragment a picture into different areas. Part implies iteratively partitioning a picture into areas having comparative qualities and developing adds to consolidating the nearby comparative locales.
2. **Region growing methods:** The area developing based division strategies are the techniques that section the picture into different districts in light of the developing of seeds (introductory pixels). These seeds can be picked physically (in view of earlier learning) or consequently (in view of specific application). At that point the developing of seeds is controlled by arrange among pixels and with the help of the prior learning of the issue, this can be ended.

D. **Watershed-Based Methods:** These techniques are use the possibility of topological elucidation. The intensity represents the basins having a hole in its minima from where the water spills. Right when water accomplishes the outskirt of the bowl the neighboring bowls are combined. To keep up detachment between bowls dams are required and are the fringes of the locale of the division [34].

E. **Artificial Neural Network Based Segmentation Method:** The artificial neural system based division techniques empower the learning procedures of the human cerebrum with the end goal of basic leadership. It is utilized to confine the required picture from the foundation. Nowadays this methodology is by and large used for the division of therapeutic pictures. A neural system comprises associated nodes and each node has a particular weight. This technique has fundamentally two phases: separating highlights and division by the neural system [33].

F. Partial Differential Equation Based Segmentation Method: The partial differential condition based strategies are the quick technique for segmentation. These are legitimate for time-based applications. On a basic level there are two PDE methods: curved non-quadratic variety reclamation and nonlinear isotropic dissemination channel which are used to upgrade the edges. The outcomes of the PDE strategy is limits and obscured edges that can be moved by utilizing close administrators. The fourth request PDE strategy is utilized to diminish the commotion from the picture and the second request PDE technique is utilized to use to

all the more probable recognize the limits and edges [35].

G. Clustering Based Segmentation Method: The clustering-based methods are the systems, which fragment the picture into groups having pixels with practically identical characteristics. Information grouping is the technique that segments the data segments into bunches with the end goal that components in a similar bunch are more similar to each other than others. There are two central classes of bunching techniques: Partition based strategy and Hierarchical strategy. The parcel based strategies utilize streamlining techniques iteratively to restrict a goal work. Though, the various leveled strategies depend on the idea of trees. In this, the root of the tree represents the whole database and the internal nodes represent the clusters.

Activity Recognition

Identifying human activities in the video has numerous vital applications like visual observation, video perusing, medicinal services checking, human-PC connection, investigation of games occasions and so forth. The procedure includes marking the pictures into fitting class for perceiving the action. In this paper, we will see the procedure of movement acknowledgment and the diverse characterization methods that can be utilized to perceive the action. To start with, division is performed on the casings that are extricated from the video [36].

Prior the vast majority of the strategies that were utilized to perceive the human exercises were straightforward exercises like standing, strolling, running, sitting and so forth however did not center around perceiving exercises like games occasions, medicinal services and so forth. Different procedures are currently utilized that can be utilized to perceive distinctive sorts of complex exercises. Existing methodologies for human movement acknowledgment incorporate best down approach and the base up approach.

With a specific end goal to perceive progressively complex human activities, much research has been done in the course of recent decades. The primary spotlight was on the extraction of worldwide highlights from recordings i.e. the highlights depicting the shape or the presence of a human amid the activity.

There are different sorts of order systems that are utilized to perceive the exercises. Every classifier contrasts from others in view of the sort of utilizations chose. Edges are removed from a video and division and procedures like shading power based, movement based, grouping based, limit based can be utilized after which include extraction is performed. These are the underlying strides previously grouping [36].

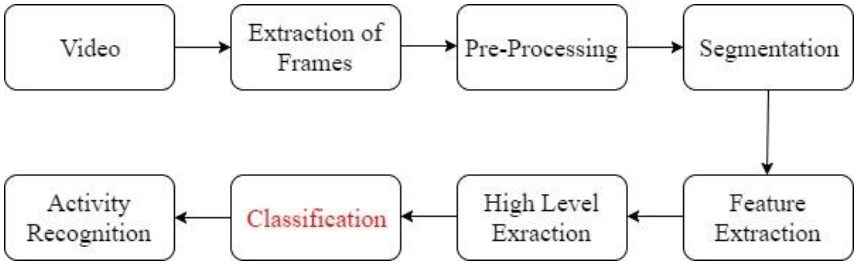

Fig. 6 Process of Activity Recognition

Activity Recognition and Classification Techniques:

Classification of images to name them into the suitable classification should be possible utilizing the methods said in figure 7.

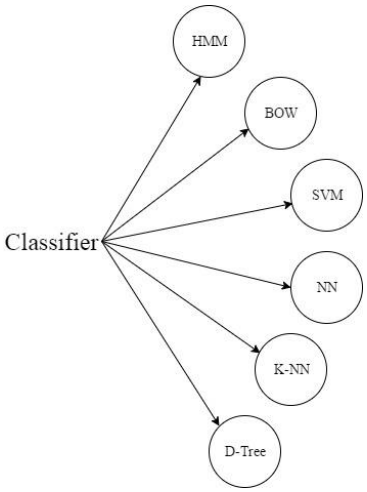

Fig 7: Classifier Types

At the point when a picture portrayal is accessible for an observed frame or sequence, human activity acknowledgment turns into a characterization issue. An activity name or appropriation over marks is given for each edge or grouping [37].

BOW (Bag of words)

1. **Integral bag of words:** It is an approach in which probabilistic movement expectation is utilized for developing vital histograms so as to speak to the

human activities. In a given video, It is required to Keep in mind the end goal to anticipate the continuous movement by processing histograms of visual words for action [38].

2. **Features and visual words:** The 3D space-time neighborhood highlights are used to anticipate the human exercises. To begin with, the element extractor changes over a video into 3D shape by linking picture outlines along the time node and after that finds movement changes. It extricates neighborhood highlights and after that structures groups in view of their appearances i.e. highlight vector esteems. These are called 'visual words', which are groups of highlights [38].

HMM (Hidden Markov Model)

For general exercises, utilizing Markov chain gives exact outcomes anyway the unpredictable responses are hard to show. For that purpose, a model is produced to work upon difficult exercises like HMM i.e. Shrouded Markov Model. Gee develops some movement shown by watching action with the impacts. The states are obvious to the spectator so the main characteristics are the state change likelihood in the Markov chains. Whereas in HMM model, the state isn't straightforwardly noticeable to the client, however, the yields are obvious which are reliant on the states [36].

SVM (Support Vector Machine)

It is a directed model in which information and perceives designs breaks down that are utilized for order. An SVM classifier manufactures a model that allots the information into one of the classifications such that the given preparation set and the information should be ordered into two classes. Human movement acknowledgment is displayed as a multi-dimensional characterization issue as a class for each activity and its point is to relegate a class mark to a given activity or action. Consequently SVM is utilized for activity order. There are many regulated learning calculations which can prepare the activity recognizer to perceive the examples of development after some time yet SVM has high speculation ability and gives high exactness [39].

NN (Neural Networks)

Neural Network comprises of a gathering of information sources and handling units known as neurons. Neurons are masterminded in three layers i.e. the information layer, the concealed layer and the yield layer. The quantity of information neurons is chosen in light of the free factors in the preprocessed include information. Various yield neurons rely upon the class names wanted. On the off chance that more exercises are to be added to the class mark, more factors will be required and thus various neurons so exchange capacities are utilized for the yield neuron. The exchange layer of any capacity decides its yield. At first, the system is prepared and tried with one concealed layer and after that the quantity of neurons is expanded. The expanding number of shrouded layer neurons results in an expansion of grouping rate. The levanberg-Marquardt calculation is utilized for preparing the neural system [40]. Neural Network utilizes movement acknowledgment to improve human memory by recording the human exercises. For every action, one neural system is prepared. Each system estimates the level of one movement and the system with the most elevated amount of certainty wins. Every neuron is appointed one grouping assignment i.e. every neuron has an order of one movement. Neural systems are utilized for the most part in human movement acknowledgment for distinguishing the falls. Other than falling, some different exercises and developments are likewise distinguished like standing up from sitting, taking a seat from standing, standing up from lying, resting from standing, sitting, running, strolling. The preparation and testing information are accumulated from a few subjects and the classifier is tried. Two separate strategies are utilized to lessen the highlights: coordinate grouping and order after element subset determination [41].

K-NN (Nearest Neighbor Classification)

K-NN classifier estimates the separation between the picture portrayal got from a watched grouping and the preparation set. The most well-known grouping from the preparation set is decided for characterization. NN arrangement is performed either at outline level or for the entire succession. 1-NN demonstrate with Euclidean separation is utilized for worldwide highlights and for histogram portrayals. However, Euclidean separation isn't much reasonable contingent upon the kind of picture portrayal. The vast majority of the strategies utilize NN grouping alongside dimensionality decrease. The distinctive separation metric that can be utilized for NN characterization is least mean casing shrewd separation, outline arrange saving variation, parametric and non-parametric thickness capacities, discriminative separation metric, and so forth. In an arrangement, numerous activities can be spoken to by key stances. There are distinctive portrayals for key stances like edge portrayal can be utilized to mark key stances, edge portrayal in an unsupervised form by physically giving activity class names in the wake of grouping, speaking to an arrangement of activity key postures by 3D portrayals. The arrangement of key postures can be discriminatively chosen. The disadvantage of K-NN is that for the extensive preparing set, figuring such separations and correlations can be costly [42].

Decision Trees

A decision tree is a various levelled show that recursively does the partition of the information space into class districts. It comprises of decision node and leaves in which every node has a test work from(x). Utilizing every one of the highlights to assemble a decision tree, first, all element data can be picked up. At that point, the mean is discovered which the best component for the primary grouping is. After a few levels of arrangement, we get classes. At that point perform testing of the example set. Every subset will generally have a similar size and same class extent as in the preparation set. Out of that, one subset is evacuated and prepared in view of other residual subsets. The primary point of the decision tree is to discover the littlest tree conceivable that would make the information in the wake of the part as unadulterated as could be expected under the circumstances. Virtue is estimated by a capacity called entropy [41].

Data Collection and Preprocessing:

As we are working on deep learning, we need a huge amount of data to train our model. For this purpose we will be collecting the

videos of farm activities directly or using the videos available online and will eventually use this collection to train our model. We will also have to preprocess these videos to reduce the resolution as it will be very hard to process videos which are above 36o frames per second. Also the unnecessary segments from the video will also be removed in order to train the model on very crisp data.

Base Model Strategy: Action acknowledgment plans to draw a portrayal of human activities and connections through the examination and comprehension of human movement designs. It contains two level methods. As delineated in below given figure, the center level expects to identify the areas of intrigue (ROIs) relating to static or moving people, while the lower level perceives fleeting movement examples of human exercises.

From a specialized perspective, human action acknowledgment can be considered as an arrangement issue utilizing time-changing element information. Visual data is removed from video groupings and spoke to in important highlights, which are utilized to coordinate with the highlights extricated from a gathering of named reference successions speaking to normal exercises.

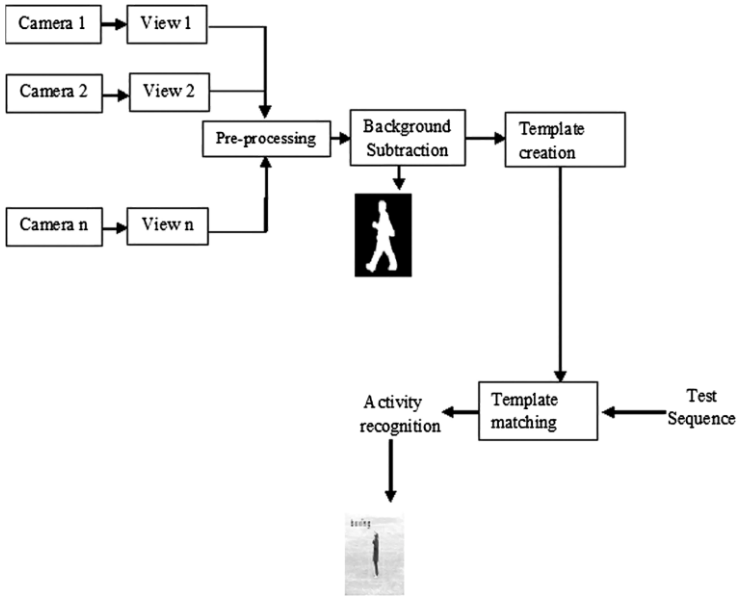

Fig 8: Base Model Strategy

Amid the extraction methodology, three sorts of highlights might be included: single question's highlights (i.e., position, speed, veins, shape, shading and so on.), worldwide highlights of numerous articles (i.e., normal speed, district inhabitance, relative positional varieties and so on.), and the connections amongst items and foundation.

After feature extraction, we will use feature selection strategies to select the required features which we need to generate a specific output. We will perform this section with various kinds of feature selection techniques. Then we gather the result generated by feature selection and then we classify the selected result by comparing our predetermined templates.

Systematic Architecture Model: This base model component has been divided into the following subsections.

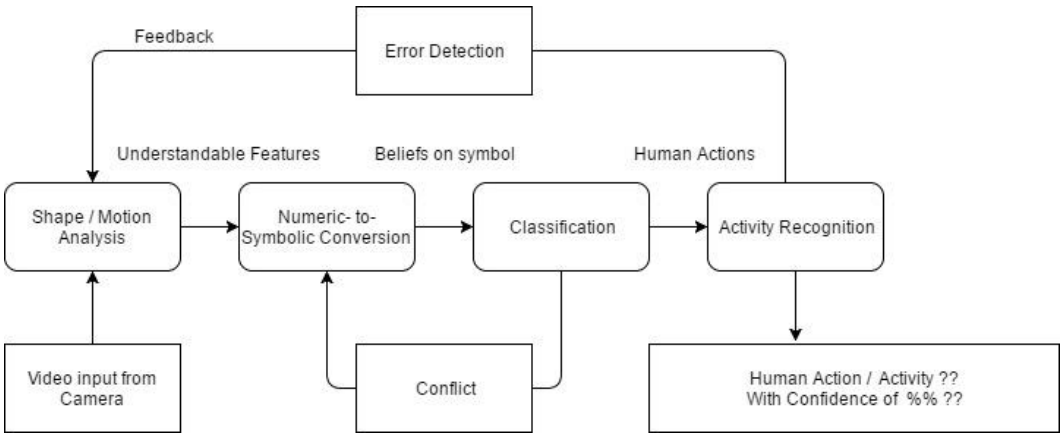

Fig 9: Systematic Architecture Model

Process Model: This diagram defines the procedural structure of our system.

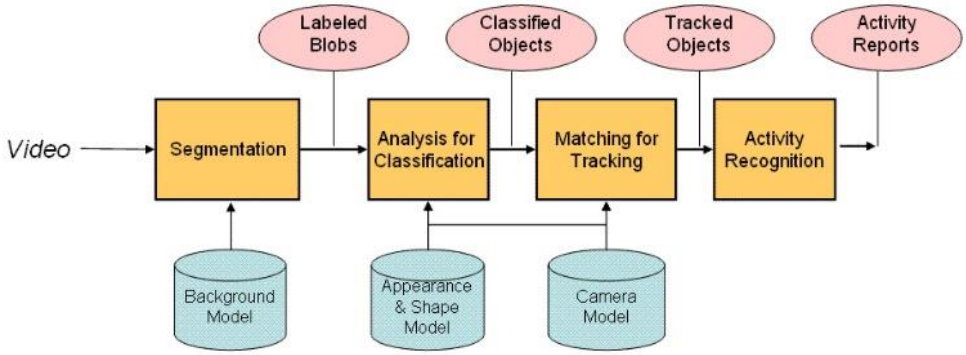

Fig 10: Process Model

Object detection: In Computer vision applications, for instance, video observation, human development examination, human-machine correspondence, and protest based video encoding (e.g., MPEG4), objects of intrigue are frequently the moving closer view questions in a picture grouping. One fruitful strategy for closer view question extraction is to smother the establishment centers around the photo plots. To accomplish this, an exact and versatile foundation display is frequently attractive. For the most part, human or protest location is grouped into 2 classifications: Object Detection and Object Classification. Object detection could be performed by foundation subtraction, optical stream and spatiotemporal filtering.

Object classification: Before applying any method to process the object, we need to know whether the object is according to our requirement or not. So, we need to classify the object first. An object can be classified as vehicles, flying creatures, gliding mists, swaying tree and other moving articles. An object in motion needs to be classified accurately for its recognition as a human being.
The available classification methods could be divided into three main categories:

- motion-based method
- Texture-based method.
- shape-based method

Object tracking: Object tracking in video structures is one of the fundamental ceaseless examination regions in the field of PC vision. The point of Object tracking is to discover a protest of a pre-characterized class in a video outline. Video structures involve various edges and colossal certainties, henceforth video following is a period superseding techniques. Following is the only ID of intrigue, especially on following of moving vehicles or walkers. Choosing the exact highlights has a basic impact in following.

Object segmentation: Segmentation is the way toward recognizing changes and separating pertinent changes for encouraging examination and capability. Pixels that have changed are alluded to as "Frontal area Pixels"; those that don't change are classified "Background Pixels". In this way, segmentation is likewise called "Background Subtraction". Pixels staying after the foundation has been subtracted are the closer view pixels. The level of "progress" which is utilized to recognize closer view pixels is a key factor in segmentation and can shift contingent upon the application.

The Image segmentation is alluded to as a standout amongst the most vital procedures of picture handling. Picture segmentation is the procedure of isolating or partitioning a picture into parts, called fragments. It is for the most part utilized for applications like picture pressure or protest acknowledgement in light of the fact that for these sorts of uses, it is wasteful to process the entire picture. Along these lines, picture segmentation is utilized to section the parts from the picture for additionally preparing.

Basically, we have three types of segmentation techniques, which we can use for object segmentation:

- Region-based segmentation
- Special theory based segmentation
- Edge-based segmentation

For feature extraction and classification we will use the concepts of Deep Learning having four major architectures:

- Convolutional Neural Networks (CNN)
- Unsupervised Pre-trained Networks (UPNs)
- Recurrent Neural Networks (RNN)
- Recursive Neural Networks

From one of this architecture, we will use to generate the features of trimmed video and predict the actions from those trimmed clips of a person. For feature extraction and activity classification, we also have many types of trained architecture like LeNet, AlexNet, GoogLeNet, ResNet etc.

A general example of feature extraction and classification is described by CNN architecture.

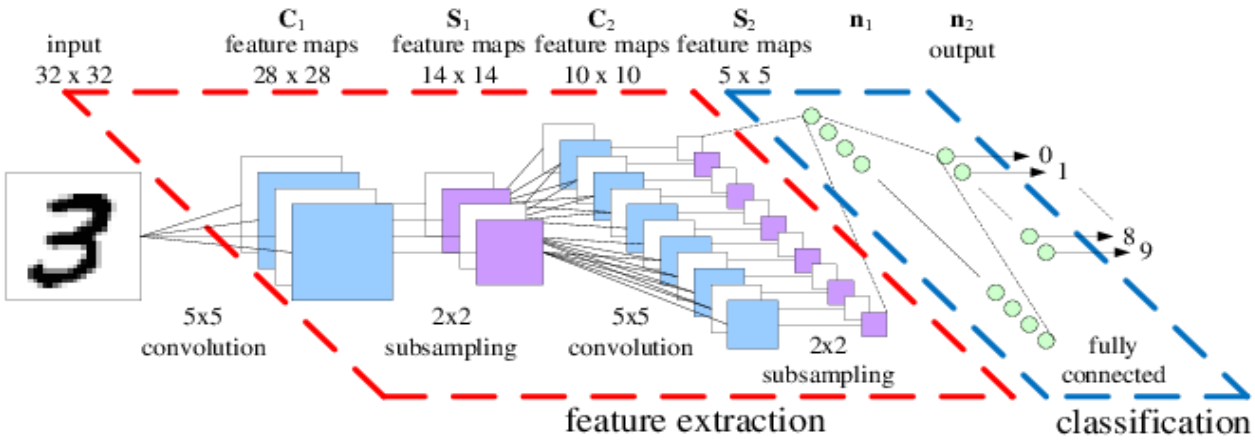

Fig11: General architecture of Convolutional Neural Network

A general example of video streaming of the previous results is described by RNN architecture.

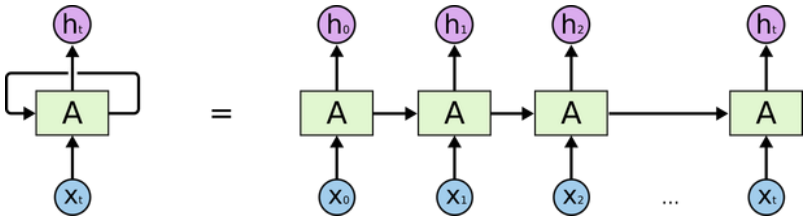

Fig12: General architecture of Recurrent Neural Network

The above given methodology will be followed to achieve the desired results for the proposed work.

References

[1] C. S. Regazzoni, A. Cavallaro, Y. Wu, and J. Konrad, "Video Analyticss for Surveillance: Theory and Practice ," no. September, pp. 16–17, 2010.

[2] R. O. I. March and W. Paper, "Video Analyticss and Security," no. March, 2008.

[3] N. Gagvani, "Introduction to video analyticss," 2008.

[4] A. Dore, M. Soto, and C. S. Regazzoni, "Bayesian Tracking for Video Analyticss " no. September, pp. 46–55, 2010.

[5] A. K. Roy-Chowdhury, *Video Processing—An Overview*, vol. 4. Elsevier Ltd, 2014.

[6] B. Shneiderman, "Improving Healthcare with Interactive Visualization," *Computer (Long. Beach. Calif).*, vol. 46, no. 5, pp. 1–13, 2013.

[7] Tripicchio, P., Satler, M., Dabisias, G., Ruffaldi, E. and Avizzano, C.A., 2015, July. Towards smart farming and sustainable agriculture with drones. In *Intelligent Environments (IE), 2015 International Conference on* (pp. 140-143). IEEE.

[8] Ryu, Minwoo, Jaeseok Yun, Ting Miao, Il-Yeup Ahn, Sung-Chan Choi, and Jaeho Kim. "Design and implementation of a connected farm for smart farming system." In *SENSORS, 2015 IEEE*, pp. 1-4. IEEE, 2015.

[9] Scherr, Sara J., Seth Shames, and Rachel Friedman. "From climate-smart agriculture to climate-smart landscapes." *Agriculture & Food Security* 1, no. 1 (2012): 12.

[10] Harvey, Celia A., Mario Chacón, Camila I. Donatti, Eva Garen, Lee Hannah, Angela Andrade, Lucio Bede et al. "Climate-smart landscapes: opportunities and challenges for integrating adaptation and mitigation in tropical agriculture." *Conservation Letters* 7, no. 2 (2014): 77-90..

[11] Thomas, Jim, and Joe Kielman. "Challenges for visual analytics." *Information Visualization* 8, no. 4 (2009): 309-314.

[12] Gubbi, Jayavardhana, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. "Internet of Things (IoT): A vision, architectural elements, and future directions." *Future generation computer systems* 29, no. 7 (2013): 1645-1660..

[13] Van Mele, Paul. "Video-mediated farmer-to-farmer learning for sustainable agriculture." *A scoping study for SDC, SAI Platform and GFRAS* (2011): 1-47.

[14] M. Paul, S. M. E. Haque, and S. Chakraborty, "Human detection in surveillance videos and its applications - a review," *EURASIP J. Adv. Signal Process.*, vol. 2013, no. 1, p. 176, 2013.

[15] D. Forsyth, "Object detection with discriminatively trained part-based models," *Computer (Long. Beach. Calif).*, vol. 47, no. 2, pp. 6–7, 2014.

[16] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," *Proc. 1999 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Cat No PR00149*, vol. 2, no. c, pp. 246–252, 1999.

[17] W. Kim and C. Kim, "Background subtraction for dynamic texture scenes using fuzzy color histograms," *IEEE Signal Process. Lett.*, vol. 19, no. 3, pp. 127–130, 2012.

[18] I. J. Of, "Research in Computer Applications and Robotics," vol. 3, no. 6, pp. 93–101, 2015.

[19] D. Viradiya, "A survey on human detection for video surveillance system," vol. 2, no. 1, pp. 154–157, 2014.

[20] M. E. Rashid and V. Thomas, "A Background Foreground Competitive Model for Background Subtraction in Dynamic Background," *Procedia Technol.*, vol. 25, no. Raerest, pp. 536–543, 2016.

[21] J. Candamo, M. Shreve, D. B. Goldgof, D. B. Sapper, and R. Kasturi, "Understanding transit scenes: A survey on human behavior-

recognition algorithms," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 206–224, 2010.
[22] Hua Zhong, Jianbo Shi, and M. Visontai, "Detecting unusual activity in video," *Proc. 2004 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, 2004. CVPR 2004.*, vol. 2, no. June, pp. 819–826, 2004.
[23] U. K. J. 3 Himani S. Parekh1, Darshak G. Thakore 2, "A Survey on Object Detection and Tracking Methods," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 2, no. 2, pp. 2970–2978, 2014.
[24] J. J. Athanesious and P. Suresh, "Systematic Survey on Object Tracking Methods in Video," vol. 1, no. 8, 2012.
[25] P. Sebastian, Y. V. Voon, and R. Comley, "Colour space effect on tracking in video surveillance," *Int. J. Electr. Eng. Informatics*, vol. 2, no. 4, pp. 298–312, 2010.
[26] D. Kaur and Y. Kaur, "Various Image Segmentation Techniques: A Review," *Int. J. Comput. Sci. Mob. Comput.*, vol. 3, no. 5, p. 809–814, date accessed: 18/05/2016, 2014.
[27] O. Badr *et al.*, "SCOPUS Indexed Journal," vol. 89, p. 65394240, 2014.
[28] A. M. Siddiqui, A. Ghafoor, and M. R. Khokher, "Image segmentation using multilevel graph cuts and graph development using fuzzy rule-based system," *IET Image Process.*, vol. 7, no. 3, pp. 201–211, 2013.
[29] I. Singh and D. Kumar, "A Review on Different Image segmentation Techniques," *Indian J. Appl. Res.*, vol. 4, no. 4, pp. 1–3, 2014.
[30] T. Lindeberg, M. Li, C. Vision, and C. Science, "Segmentation and classification of edges using minimum description length approximation and complementary junction cues," *Comput. Vis. Image Underst.*, vol. 67, no. 1, pp. 88–98, 1997.
[31] M. Agarwal and V. Singh, "A Methodological Survey and Proposed Algorithm on Image Segmentation using Genetic Algorithm," *Int. J. Comput. Appl.*, vol. 67, no. 16, pp. 7–17, 2013.
[32] S. Angelina, L. P. Suresh, and S. H. K. Veni, "Image segmentation based on genetic algorithm for region growth and region merging," *2012 Int. Conf. Comput. Electron. Electr. Technol. ICCEET 2012*, pp. 970–974, 2012.
[33] A. Khare, M. Saxena, and S. Tiwari, "Edge Detection Method for Image Segmentation – A Survey of Soft Computing Approaches," no. 4, pp. 174–178, 2011.
[34] W.-X. Kang, Q.-Q. Yang, and R.-P. Liang, "The Comparative Research on Image Segmentation Algorithms," *2009 First Int. Work. Educ. Technol. Comput. Sci.*, pp. 703–707, 2009.
[35] M. Yambal and H. Gupta, "Image Segmentation using Fuzzy C Means Clustering : A survey," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 7, pp. 2927–2929, 2013.
[36] M. Sharma and B. Joglekar, "Survey of Classification Techniques For Human Activity," vol. 1, no. 12, pp. 19–26, 2014.
[37] R. Poppe, "A survey on vision-based human action recognition," *Image Vis. Comput.*, vol. 28, no. 6, pp. 976–990, 2010.
[38] M. S. Ryoo, "Human activity prediction: Early recognition of ongoing activities from streaming video," *Iccv*, pp. 1036–1043, 2011.
[39] C. Schuldt, L. Barbara, and S.- Stockholm, "Recognizing Human Actions : A Local SVM Approach ∗ Dept . of Numerical Analysis and Computer Science," *Pattern Recognition, 2004. ICPR 2004. Proc. 17th Int. Conf.*, vol. 3, pp. 32–36, 2004.
[40] a. Sharma, Y.-D. L. Y.-D. Lee, and W.-Y. C. W.-Y. Chung, "High Accuracy Human Activity Monitoring Using Neural Network," *2008 Third Int. Conf. Converg. Hybrid Inf. Technol.*, vol. 1, pp. 430–435, 2008.
[41] M. Baccouche, F. Mamalet, and C. Wolf, "Sequential deep learning for human action recognition," *Proc. Int. Conf. Hum. Behav. Underst.*, pp. 29–39, 2011.
[42] S. KAGHYAN and H. SARUKHANYAN, "Activity recognition using K-nearest neighbor algorithm on smartphone with Tri-axial accelerometer," *Int. J. Informatics Model. Anal.*, vol. 1, pp. 146–156, 2012.
[43] J. Y. Lee and W. Yu, "Visual tracking by partition-based histogram backprojection and maximum support criteria," *2011 IEEE Int. Conf. Robot. Biomimetics, ROBIO 2011*, pp. 2860–2865, 2011.
[44] Q. Bai, "Analysis of Particle Swarm Optimization Algorithm," *Comput. Inf. Sci.*, vol. 3, no. 1, pp. 180–184, 2010.
[45] Gavrila, Dariu M. "The visual analysis of human movement: A survey." Computer vision and image understanding 73, no. 1 (1999): 82-98.
[46] Vrigkas, Michalis, Christophoros Nikou, and Ioannis A. Kakadiaris. "A review of human activity recognition methods." Frontiers in Robotics and AI 2 (2015): 28.
[47] Wang, Liang, Weiming Hu, and Tieniu Tan. "Recent developments in human motion analysis." Pattern recognition 36, no. 3 (2003): 585-601.
[48] Moeslund, Thomas B., Adrian Hilton, and Volker Krüger. "A survey of advances in vision-based human motion capture and analysis." Computer vision and image understanding 104, no. 2 (2006): 90-126.
[49] Turaga, Pavan, Rama Chellappa, Venkatramana S. Subrahmanian, and Octavian Udrea. "Machine recognition of human activities: A survey." IEEE Transactions on Circuits and Systems for Video Technology 18, no. 11 (2008): 1473-1488.
[50] Poppe, Ronald. "A survey on vision-based human action recognition." Image and vision computing 28, no. 6 (2010): 976-990.
[51] Aggarwal, Jake K., and Michael S. Ryoo. "Human activity analysis: A review." ACM Computing Surveys (CSUR) 43, no. 3 (2011): 16.

