Human Activity Recognition Methods

Dhivya Karunya S, Krishna Kumar

Abstract: *Human action in a video based application plays a significant role that alerts the researchers towards recognizing the motion of human. Other video applications also have video content extraction, summarization, and human computer interactions. The existing methods needs manual footnote of pertinent portion of actions of our interest. Recognition of human action can be done authentic without physical commentary of applicable parts of action of any one's interest. In this paper we try to update the previous reviews on many ways of recognizing Human activities in videos that had different techniques like Hidden Markov model, feature extraction, segmentation etc. the recognition of human activity in applications like visual observation in mobile, human fall detection, video conference, robotics.*

Keywords : Human action Recognition, Spatial and Temporal extent, Dense Trajectory, Support Vector Machines, Spatio-Temporal Interest Points, Hidden Markov models.

I. INTRODUCTION

Detection of human-actions in a video sequence is the recent subject discussing in machine learning. Understanding human action helps in recovery of videos, management, and depicting the main points of actions. In the earlier papers there are tremendous onward movement had been made by creating local invariant characteristics and container of attributes depiction. Thus situation is demanding because of differences in action execution, background settings and interpersonal dispute. There are two things to look forward in order to notice the action in video they are action identification and action localization. First is what type of action is carried out and secondly where these actions will be present in a video. The difficulty of allocating videos to a specific activity classes is commonly familiar as action identification and action localization will be detecting space and time of video. Training of these videos into different category for action recognition will be a challenging task. Positive and negative samples will be included in training. Testing of video is done succeeding the training phase. At training stage we physically explain the small portion of action. Explaining manually is unexciting, consumes time and will be more erroneous. Many action perception techniques that exist will recognize which Video proportions are accumulated by a person of actions initiates a technique to point out temporal range of video. However spatial conditions are neglected due to which gives rise to unwanted information from that domain.
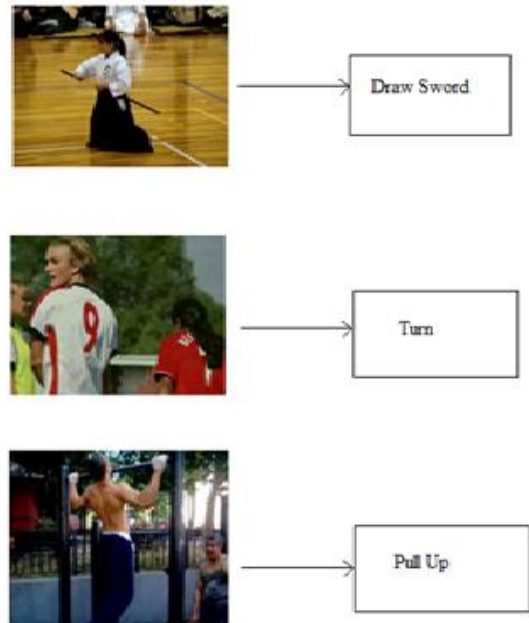

Fig.1. Detection and displaying of actions performed on video frames.

II. CLASSIFICATION OF HUMAN ACTIVITY

We classify human activity recognitions into two they are Unimodal and Multimodal methods.

A. Unimodal Methods

Human-actions can be represented from statistics of one plan of action, like images. Unimodal methods are suitable for understanding human liveliness based on motion characteristics. However the capacity to identify fundamental group only from movement will be on its own will be a difficult task. Few techniques [18] use snippets of motion trajectories and other uses full extent of movement curves by following optical flow characteristics.

Unimodal methods again divided into: (i) Space time, (ii) Stochastic, (iii) Rule based, and (iv) Shape based methods.

Dhivya Karunya S*, ECE, S.E.A College of Engineering and Technology, Visvesvaraya Technological University, Bangalore, Karnataka, India. E-mail: dhivyakarunyas@gmail.com.

Krishna Kumar, ECE, Gopalan College of Engineering and Management, Visvesvaraya Technological University, Bangalore, Karnataka, India. E-mail: njkrishnakumar@gmail.com

Retrieval Number: E9771069520/2020©BEIESP
DOI: 10.35940/ijeat.E9771.069520
Journal Website: www.ijeat.org

1024

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication
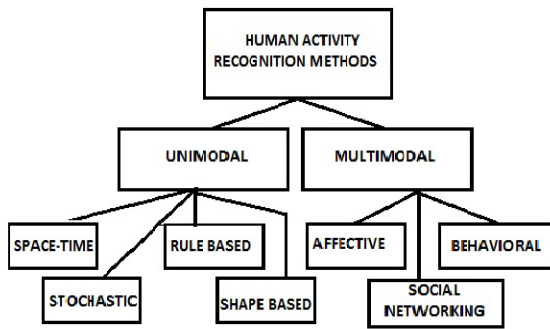*© Copyright: All rights reserved.*

Fig.2. Classification of Human Activities.

Space Time Method: This method includes activity identification technique that identifies human actions as a bunch of spatio-temporal attributes. This method concentrates on analyzing liveliness based on space-time characteristics otherwise on uniting the trajectory.

Stochastic Method: It identifies activities by applying statistical models like Hidden Markov models (HMM) that represent human behavior actions. The human behavior is like a sophisticated chain of activities. Every activity is related by a quality vector that unites details regarding velocity, local descriptors and position. HMM is used to cipher human movements, whereas identification is carried out by finding image characteristics representing an activity.

Rule based Method: They make use of a specified set of procedures to narrate human movements. This method determines in progress incidents by modeling an action using procedures or bunch of features which relates an occurrence. Every liveliness is represented as a set of primitive procedures or features that permits the building of expressive prototype for human activity identification.

Shape based Method: It represents events with great reasoning by shaping the movement of human body parts efficiently. We understand that event identification techniques are established on the human outline does a major role in identifying human activities. A individual outline comprises of limbs mutually attached to each other, obtaining body parts of human exactly from the videos is very important. This will be a challenging task and treated as a portion of action identification task. Other techniques tell us the details about solving this problem.

B. Multimodal Method

These methods merge attributes gathered from various resources. They are usually based on feature bonding that can be extracted by two plans of actions: early-fusion and late-fusion. The uncomplicated means of getting assistance of several characteristics is to instantly chain attributes in a larger attribute vector and acquire the knowledge of the fundamental action. Recognition performance can be improved by this feature fusion approach; however modern feature vector will be having massive dimension.

Multimodal signs are highly correlated with respect to time; hence a secular alliance of the fundamental actions and various processes will be a key issue for mastering the information. In those circumstances [19], audio-visual investigation will be used in numerous implementation platforms for tracing and action identification. Multimodal methods are divided into (i) Affective, (ii) Behavioral, and (iii) Social networking methods.

Affective Method: It indicates human- actions pertaining to psychological expressions and perceptual condition of an individual. The principle behind emotional observation will be mastering the similarity between an individual's affective situation and relating actions that are extremely associated to emotional condition and conveying of an individual with another. Affective calculating research represents capability of an individual to reveal, identify, and restrict their affective conditions in the form of hand signaling, facial appearances, physical changes, verbal expression, and event identification. A major issue in affective calculation [20] is precisely explained data. Moreover, the challenge is to achieve form environments, because affective incidents are conveyed in a dissimilar way by various individuals or that takes place at the same moment with other actions.

Behavioral Method: This type of method aims to identify behavioral features, non-verbal multimodal signals, like gestures, facial changes, and auditory signals. For computer vision group identifying human actions is a challenging task. A behavior identification module gives details of the identity and intellectual condition of an individual, and its significance may differ from video monitoring to computer-human association. Main aspects which can influence individual conduct are divided into few components, like feelings, humor, activities, and interactions, with other persons. Therefore identification of complicated activities can be critical for understanding individual's actions. One principle detail of individual's actions identification is the possibility of proper attributes [21] that is used to know actions in applications, like physiology and gaming. Main task in understanding actions of human is to define certain psychological features for multimodal bilateral interactivity. These features will be explanations of emotional events or perceiving events, like as valence, engagement and activation.

Social Networking Method: This method represents the feature s and actions of human in some surface of human-human interplays in public affairs from signals, speech and body movements. Communal interactivity is a major part in day today life. Basic part of the human actions is the capacity to link with different individuals with their behavior. Communal interactivity [22] is viewed as a particular kind of action where anyone can accepts their actions as per a group of people nearby them. All the communal networking systems affecting individual's behavior, like as Twitter, YouTube, and Facebook estimate social interactions and conclude how much these sites are involved in problem of social capital, education, identity, youth culture, and privacy. A current survey on human behavior identification gives entire summary of all approaches for mechanized human behavior inspection for single individual, other group of individuals, and machine-person interactions.

III. LITERATURE REVIEW

Human actions are unusual and recognizing such activities have been carried out from the past two decades. Many of the human action identification technique explain manually about interest in a small portion of videos.

But the recent era has witnessed the changes can be found easily with automation and can identify the action. Different authors use different approaches to recognize and detection of action few of them are reviewed.

A. Action Recognition: A multi-view of event identification system [10] is presented for combining in Human-Robot Interaction (HRI) framework with uncommon users. For example, children's have restricted information for guiding and other methods are struggling to find solution. Various characteristics extraction methods, encoding techniques and fusion methods are integrated and tested to make a structured arrangement that identifies children's actions.

Nweke Henry Friday et al [11] surveyed to identify actions using mobile sensor, information is gathered using suitable sensors, subdivided, required characteristics are obtained and actions classification is done using favoring models. Finding attributes is a main phase that helps to minimize calculation time and establish increase in identification.

Defining feature trajectories show more efficiency while representing videos. The trajectories of quality and quantity were not satisfactory. Then for image classification the usage of dense sampling came into existence and has become popular. Liu et al. [1] propose dense trajectories for representing the videos. Based on displacement information from each frame dense points are traced and sampled. Camera motion is taken into account for further improvement in interpretation schmiid et al. [2].

The movement of camera was approximated by uniting the feature points linking the frames by using dense optical flow and SURF descriptors. Motion relationship is focused at modeling the motion connection by the other approach Jiang et al. [3]. This approach functions on top of optical code-words extracted by local path trajectories hence any precise background and foreground partition is not required.

Ahmed et al. [4] propose a unique technique for locating instructive area. Author uses saliency mapping techniques. This paper adopts a novel technique using a space for understanding spatial and temporal spatial and temporal stretch of action.

Bishoy et al. [12] explains the evaluations that are carried out on fitness exercises and daily activities from 16 participants. The result shows that they have used naive bayes in the experiment which had performed better in accuracy and efficiency of classification.

Muhammad et al. [13] describes work carried out on recognition of physical activity during an online session in mobile phones. They have considered the sensors of mobile phone that does the local classification in real time.

B. Action Detection: For each of the actions and moment description Mahlanobis distance was carried out to identify human behavior. Machine learning techniques like AdaBoost and Support Vector Machines gives one prospect for including data accommodation in a group of training instances. [4] Initiates MACH filters, a pre-format technique for activity identification that is competent of apprehend intra category unevenness by consolidating a single activity.

An alternate technique proposed [5] is based on indistinct action locations human action detection is found by making use of imitate Multiple Instance Learning Support Vector Machines, also supporting nature of appearance features and

motion in action detection are verified.

Lan et al. [6] utilized a statistic centric ocular word depiction. Hence in this paper localization can be casted as unexpressed variable in order to identify the behavior. Spatio-temporal prototype is understood. Throughout the preparation [7] prototype parameters is approximated and applicable part is recognized. [8] Differentiating human actions from background motion a self-sustaining motion evidence feature is made use of.

Most of the techniques need the applicable part of video that will be explained with Vault boxes. People's involvement was unexciting. In order to get better of Vault box Genmert et al. [9] the author splits the video into blocks of smaller groups and a prototype is obtained which locate the applicable smaller group. Dense Trajectory has been used with local features that indicate the human action; also it initiates a practice that makes us to learn spatial as well as temporal extents in improvement of detection of actions.

IV. CHALLENGES AND APPLICATIONS

Here we deal with review of some familiar issues for event identification making use of mobile sensors.

A. Content Susceptivity: The precision of action identification, which is established on accelerometer information, that are highly simulated by the issues engaged in guiding and investigating phases, that is because of different individuals has various moving fashion. Those with similar content, even they may have unusual fashion at separate periods. Relative result shows that guiding and investigating same issue attains big precision. Guiding and investigating on similar set of groups with more contents will be next big authenticity. The precision reduces when trial information is gathered from similar content, but will be on unfamiliar days. An identification prototype trained on a diversified dataset had worked so well, when it is tested on information from new people. A cross person activity identification method suggested by Deng et al.[14] to remove sequel of user susceptivity. The guiding stage comprises of two different sections: Initial prototype which will be trained offline and adaptive prototype will be revised online.

B. Location Sensitivity: The accelerometer has a property of reading raw data in smart phones and wearable sensors which depends on positions and orientation of the individual body. The smart phone and wearable sensors has an accelerometer where its reading relays on position and orientation of sensors on individual's body. One way was found in Ref. [15] for addressing the sensitivity orientation by the use of another magnetometer sensor. This sensor gives the magnetic vector on the three axes of device co-ordinate system which will be in orthogonal directions. Therefore it is used to get device horizontal angle. The reading of accelerometer is permitted for converting into co-ordinate axes of the earth.

C. Action problem: The complication in the individual's action leads to risk further for identification in the model. People carrying out various works [16] but at the same period may puzzle the classifiers that are guided by an action per segment. The lifestyle and person's contrast will result inequality in a way that individual's executes tasks, that in turn obtains struggling in appealing the action identification models.

D. Energy and Resource Constraints: Activity identification significance need constant sensing and updating in online of the classification model, these will consume more energy. It also needs remarkable computing resources updating in online. Due to monitoring of needed sampling frequency varies for various actions, [17] create options on sampling frequency as well as categorizing attributes. Hence it decreases computing resource cost and energy also eliminates consuming of time in frequency province attribute computation

E. Inadequate guiding Set: The challenge part in subject susceptibility was advisable that guiding details should have different number of subjects as realizable. Moreover, it is difficult to synchronize individuals with distinct ages, body shapes and different behavior to gather information under a closed surveillance mechanism. Semi monitoring study is applied to adopt this issue. For activity identification, collection of unlabeled data will be easier and requires less number of user's effort. Combining virtual evidence boosting method and semi supervised learning reduces the mankind labeling cost and increases the efficiency for feature selection. Scale- invariant classifier with ―R‖ metric (SIC-R) is used to know multi scale incidents of individual's movements.

F. Day to day Life Monitoring Application: It aims to give a suitable authority for the movement registering, or facilitating with physical jerks and healthy workouts. Such appliances are enabled with sensors like gyroscope, GPS accelerometer which in line with individual's measures taken, calorie burned, how many minutes slept, travelling distance calculation, stairs climbed, and nature of sleep. For users assistance is provided through online in order to analyze information tracing and understanding data in files. In contrast to smart phone sensors the appliances are further complicated, because these sensors are planned particularly for movement identification and supervising.

G. Individual Biometric Signature: A person's movements are always distinct. For example, if two persons walk on a small track by raising their hands, it is more or less impossible for them to share exactly same movements. Though in a productive impression, the dissimilarities will be present. Accelerometers act like a sensor that can capture such differences. The action identification methods give us a feasible clarification for biometric signature of human with models in attributes. In these implementations, pattern identification techniques will be used to find distinct movement patterns that will be saved. This is possible due to the usage of mobile devices.

H. Elderly and Youth Care: Elders require care in-terms of mentally and physically due to their age. The outmost aim in the present research of human action surveillance is to introduce newest techniques for old people. These applications and technologies will help them in dangerous situations. The invention of smart phone serves the elderly with fall detection as well. Action identification and motion sensors will help elders in a dynamic way, like alarm for waking up, time for taking medicine, monitoring of living area through remote access. Another field that is benefited from the activity identification is the youth care. Many of the applications like monitoring the movements of a child, forecasting their demands for food, sleeping. In children's safety detection, activity recognition techniques are used.

I. Localization: Action identification on cell phones will assist in condition perception thus it will be used in localization. The signals of GPS is not strong inside high raised structures and below ground because signal cannot penetrate faster hence this is one of the reasons to use sensors in cell phones instead of GPS for localization. In order to assist in locating the position the action identification techniques with mobile sensors are used. However GPS localization is 2 dimension orientations that do not have any idea about the information on a person's elevation. Another intention for using sensors in mobile for localization is that accuracy of GPS reduces within the metropolis because of high raised properties. For this condition GPS based localization may puzzle among a restaurant and multiplexes, that will be too close.

J. Industry Assisting: The activity identification methods could help labors in their job. This work initiates wearable sensors to a job wearable-evaluation which is a sort of appendix of the body allowing a labor to execute exceptional work. Smart cameras are used to detect people's gestures in movies, production of cars where robots are used as assistant etc., are the other types of applications on action recognition.

V. RESEARCH GAP

Obtaining any data related to an individual in any situation is the process called as human sensing that relates only the speculation of spatiotemporal properties. It comprises of ground level elements about the location and older days of individual in the space. Appearance: It is probably the personality that is frequently looked after in existing environment, the widespread poise-sensor being proximity sensors and motion sensors. In collaborative situations although where humans will be using moveable or wearable gadgets, RFID has become common to give solutions.

Count: The number of humans present in a space can be concluded by hiring a individual or using a sensor that would describe the whole locality of interest or counting individually at entry and exit levels. However for large people-counting results vary from break-beams and thermal imagers into direct automatic hurdles like turnstiles.

Position: Where will be each individual present? Position-identification, or localization, comprises of getting the spatial correlate of humans core of collection. Localization is obtained by using equipment or completely non equipped devices, also a mesh of existence sensors may be utilized to confine humans, localization is treated to be having a more resolution of position identification.

Tracing: Tracing is the procedure of interpret the compatibility problem which will be of obtaining the spatio-temporal background of one and all in a scenario. Significantly, tracing can be related as retrieving an individual's respective identification. For an instance, if a person who is given a temporary ID on detection then tracing is a problem of identifying at each successive sampling of the space which identifying the same individual. When a individual come back on next day the temporary ID will loses its existence of sensing the gaps. In that case the previous day's person must be given an ID which is new when detected again. Circumstances that guides to depletion of a individual's comparative ID are known as obscurity.

Recognition: As we see at first look it may be an odd to a category or specification into a class of spatio-temporal properties. Moreover, recognition is no longer than a real expansion of tracing where each individual is allocated identical distinct ID instead of comparative IDs. Hence, originality observation increase tracing so it turns out to be feasible to retrieve individual's spatial secular background even above noticeable discontinuity.

In the current research technologies single human activities concept has been developed using Support Vector Machine SVM classifiers. More interestingly Spacio-temporal interest points (STIP) technique is used to develop Multitask human activities recognition. Data fusion concept has been done only using Image processing and Robotics. There is a need for combining more data in image processing technology.

VI. CONCLUSION

In this literature review we have seen many Human Activity Recognition techniques, various algorithms used for human and device interdependence in image and videos, various people-people interrelation methods and classification types are used for identification. Main aim is to contribute an immense survey and comparing other methods and procedures of Human Activity identification. In this paper challenges and applications are also discussed. The review in Human action identification represents crucial advancement in diverse point. Moreover all those works have not inscribed the different issues of action identification such as including the background of location, human-device association in videos, and online action forecasting.

REFERENCES

1. H.Wang,A Klaser,C.Schmid and C-L.Liu, "Action recognition by dense trajectories ," in Proc. IEEE Conf. Comput. Vis.Pattern Recog.,Jun. 2011,pp 3169-3176.
2. H.Wang and C Schmiid , "Action recognition with improved trajectories," in Proc.IEEE Int. Conf. Comput. Vis., Dec 2013,,pp 3551-3558.
3. Y-G Jiang,Q.Dai,X.Xue,W.Liu and C-W Ngo. "Trajectory-based modeling of human actions with motion reference points," in Proc. Eur.Conf .Comput.Vis.,Oct 2012,Vol 7576,pp.425-438.
4. M. D. Rodriguez, J. Ahmed, and M. Shah, "Action MACH: A spatiotemporal maximum average correlation height filter for action recognition,"in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2008,pp. 1–8.
5. Y. Hu, L. Cao, F. Lv, S. Yan, Y. Gong, and T. S. Huang, "Action detectionin complex scenes with spatial and temporal ambiguities," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Sep.–Oct. 2009, pp.128–135.
6. T. Lan, Y. Wang, and G. Mori, "Discriminative figure-centric models for joint action localization and recognition," in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011, pp. 2003–210.
7. M.Raptis, I.Kokkinos and S.Soatto," Discovering discriminative action parts from mid-level video representations" ,in Proc ,IEEE Conf.Comput.Vis,.Pattern Recog.,Jun 2012,pp.1242-1249
8. W. Brendel and S. Todorovic, "Learning spatiotemporal graphs of human activities," in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011,
9. M.Jain,J.van Genmert,H.Jegou , P.Bouthemy and C.Snoek ,"Action localization with tubelets from motion" ,in Proc IEEE
10. Niki Efthymiou, Petros Koutras , Panagiotis Paraskevas Filntisis , Gerasimos Potamianos , Petros Maragos "Multi-View Fusion for Action Recognition in Child-Robot Interaction" IEEE Xplore: 06 September 2018 ISSN: 2381-8549.
11. Nweke Henry Friday , Mohammed Ali Al-garadi , Ghulam Mujtaba , Uzoma Rita Alo , Ahmad Waqas, "Deep Learning Fusion Conceptual Frameworks for Complex Human Activity Recognition Using Mobile and Wearable Sensors IEEE Xplore 26 April 2018.
12. BishoySefen, Sebastian BaumbachHuman Activity Recognition Using Sensor Data of Smart phones and Smart watches/ ICAART 2016.
13. Muhammad Shoaib, Stephan Bosch , Ozlem Durmaz Incel , Hans Scholten and Paul J.M. Havinga "A Survey of Online Activity Recognition Using Mobile Phones/ Sensors" 2015.
14. Wu, Q., Wang, Z., Deng, F., Chi, Z., and Feng, D. D "Realistic human action recognition with multimodal feature selection and fusion" 2013 IEEE Trans. Syst. Man Cybern. Syst. 43, 875–885. doi:10.1109/TSMCA.2012.2226575.
15. J. R. Kwapisz, G. M. Weiss, and S. A. Moore, Cell phone based biometric identification, in Proc. 4th Int. Biometric Theory, Applications and Systems Conf., Washington, DC, USA, 2010, pp. 1-7.
16. L. Bao and S. S. Intille, Activity recognition from userannotated acceleration data, in Pervasive Computing, Springer, 2004, pp. 1-17.
17. Y. E. Ustev, O. Durmaz Incel, and C. Ersoy, User, device and orientation independent human activity recognition on mobile phones: Challenges and a proposal, in Proc. 13th Int. Pervasive and Ubiquitous Computing Adjunct Publication Conf., Zurich, Switzerland, 2013, pp. 1427- 1436
18. Matikainen, P., Hebert, M., and Sukthankar, R.. "Trajectons: action recognition through the motion analysis of tracked features," 2009 in Workshop on Video-Oriented Object and Event Classification, in Conjunction with ICCV (Kyoto: IEEE), 514–521.
19. Wu, Q., Wang, Z., Deng, F., Chi, Z., and Feng, D. D. (2013). Realistic human action recognition with multimodal feature selection and fusion. IEEE Trans. Syst. Man Cybern. Syst. 43, 875–885. doi:10.1109/TSMCA.2012.2226575.
20. Soleymani, M., Pantic, M., and Pun, T. (2012). Multimodal emotion recognition in response to videos. IEEE Trans. Affective Comput. 3, 211–223. doi:10.1109/T-AFFC.2011.37.
21. Metallinou, A., Katsamanis, A., and Narayanan, S. (2013). Tracking continuous emotional trends of participants during affective dyadic interactions using body language and speech information. Image Vis. Comput. 31, 137–152. doi:10.1016/j.imavis.2012.08.018.
22. Candamo, J., Shreve, M., Goldgof, D. B., Sapper, D. B., and Kasturi, R. (2010). Understanding transit scenes: a survey on human behavior-recognition algorithms. IEEE Trans. Intell. Transp. Syst. 11, 206–224. doi:10.1109/TITS.2009.2030963.

AUTHORS PROFILE

Dhivya Karunya S., is currently working as Assistant Professor in the department of Electronics and Communication Engineering and a research scholar at S.E.A CET, Bangalore under Visvesvaraya Technological University, Belagavi. She has done her bachelor's in Electronics and Communication Engineering at Sengunthar Engineering College Tamilnadu, affiliated to Anna University and her Masters in Digital Communication and Networking at S.E.ACET, Bangalore affiliated to Visvesvaraya Technological University, Belagavi. She has a rich experience of 10 years in Teaching. Her area of interest includes Signal Processing, Communication and Networking, Digital Electronics and Communication, Artificial Intelligence, Machine Learning. She is a member of ISTE.

Dr. Krishna Kumar is working as a professor at Gopalan College of Engineering and Management in the department of Electronics and Communication Engineering has done his Bachelor's in Electronics and Communication Engineering from S.J.C E Mysore, affiliated to University of Mysore and his Masters in Network communication and security from Dr M.G.R Educational and Research Institute Chennai and obtained doctorate from Techno global University, shillong. He has a rich experience of 20 years in teaching and research. He has worked at various capacities in affiliated universities. He has published more than 15 research papers. His area of interest includes Communication and networking, video processing, Machine learning, Artificial intelligence, Internet of Things, Data Science. He is also a member of ISTE.