"APPLICATION AND CHALLENGES TO ANALYSIS OF HUMAN ACTION AND ACTIVITY FOR BEHAVIOR UNDERSTANDING USING COMPUTER VISION TECHNIQUE"

R. Y. Gedam[1], R. J. Ramteke[2]

[1]Assistant Professor & Head, Department of Computer Science DRB ,Sindhu Mahavidhyalaya, RTM Nagpur University ,India,7972448439.

[2] Professor, Head of Department of Information Technology, School of Computer Sciences, North Maharashtra University, Jalgaon,India ,9284275691.

[1] *Email – Id :mr.rakeshgedam1@gmail.com*

[2] *Email-Id:rakeshj.ramteke@gmail.com*

Abstract-

Human Activity/Action Recognition (HRA) attract many researcher all over the globe due to its wide range of application. To understand human behavior through computer or machine recognition human action and activity properly to classify and feature extraction to derive conclusion what action has been done to carried particular activity which assist better understand HBA .The prime challenges for HRA Intra class variation and Inter class, Recognition under Real-World Settings has interpreted. The proposed framework to understand action and activity from video has been discussed.

Key words- ***Human Behavior Understand, Human Activity Reorganization, Computer Vision***

1. Introduction :

The term action and activity may be interchangeable in Human Activity Reorganization (HAR). Human action /activity recognition (HAR) is a highly dynamic and challenging research topic from last two decade for computer vision, Artificial Intelligence, machine learning, pattern reorganization, video analysis and image & signal processing and interdisciplinary community due to its huge range of application in various aspect of life. HRA aims to determining the activities of a person or a group of persons based on sensor and/or video observation data, as well as on knowledge about the context within which the observed activities take place. Human Behavior Understanding /Recognition using Human activity/action analysis have wide range of application area like Security and Surveillance, Human Computer Interaction(HCI), Intelligent Environment, Elder Health Care, Indoor Navigation, Shopping Experience and many more [1]. Human Activity Recognition (HRA) is a complex task for computer vision (CV), machine learning, pattern reorganization, Human computer interaction (HCI) and interdisciplinary area due to its various limitation. Human Activity Recognition (HAR) attract many researcher of computer science & engineering, psychology and interdisciplinary subject after analyzes numerous application of various field. Human Activity Recognition (HAR) use to find activity or action carried by one or many persons like walking, running, jumping, handshake etc. Our focus is observe and then analysis action and activity perform by human being to understand its behavior through The difference between computer vision and image analysis is not clearly defined. The two fields use largely the same algorithms and theory. In addition, there are several other fields that are closely related and often with unclear separation – for example pattern recognition, signal processing, machine vision, image processing and photogrammetric.

However, computer vision tends to focus on 3D analysis of images (e.g. structure from motion) while image analysis mainly deals with 2D pixel-wise operations (e.g. edge detection)

Basic Terminology: After review the literature we may define the following terms

Action: "actions" can be a simple motion patterns usually executed by a person or human being and typically lasting for short durations of time, may be tens of seconds or less like jumping, running, bending, walking, swimming, picking etc

Activity: "activities" can be complex sequence of actions performed by several humans who could be interacting with each other in a constrained manner for much longer temporal durations, e.g. playing, making tea, kissing, handshaking etc

Intent: human intention is based on the human actions along with the scene changes that occur due to the human actions like kicking, boxing

Motion: Motion is nothing but change in position of an object over time like running, jumping, seating, picking etc.

Behavior: The way in which an animal or person behaves in responses to particular situation are stimulus like happy, sad, eager, fear, surprised etc.[16]

Figure 1. Human activity types scaling from simple action to event

(a) (b) (c)

Figure 2. Different behaviors generated by similar spatio -emporal features corresponding to the action of running: (a) sports; (b) robbery; (c) catching a bus.

2. Literature Survey:

In last two decade various surveys paper in the human action and activity recognition published. Some researcher focus on 2D (with and without explicit shape models) and 3D approaches with new taxonomy was presented focusing on human motion analysis, tracking from single view and multitier cameras, and recognition of human activities [3]. Based on previous work we proposed a hierarchical action categorization hierarchy [4]. The survey for fourfold taxonomy by author i.e pose-based action recognition methods including initialization of human motion, tracking, pose estimation, and recognition methods [5]. The author Akata etl give separation meanings of "action" and "activity" was proposed where the activity recognition methods were categorized according to their degree of activity complexity [6]. According to author Human activity recognition methods characterized into two main categories i.e "top-down" and "bottom-up." [7] whereas other presented a tree-structured taxonomy, where the human activity recognition methods were categorized into two big sub-categories, the "single layer" approaches and the "hierarchical" approaches, each of which have several layers of categorization [8]. New approch for action and activity focus on modeling 3D data was extensively studied by people as the human body consists of limbs connected with joints, identified the stronger features using depth cameras, and create a 3D representation of the human body provide more efficient information rather than 2D activities carried out in

the image plane. Recently presented a categorization of human activity recognition methods from 3D stereo and motion capture systems with the main focus on methods that exploit 3D depth data. To this end, Microsoft Kinect has played a significant role in motion capture of articulated body skeletons using depth sensors[10].Although much research has been focused on human activity recognition systems from video sequences, human activity recognition from static images remains an open and very challenging task. Most of the studies of human activity recognition are associated with facial expression recognition and/or pose estimation techniques. Methods for human activity recognition from still images and categorized them into two categories according to the level of abstraction and the type of features each method uses [9]. A survey for multimodal human computer interaction focusing on affective interaction methods from poses, facial expressions, and speech. Study in human affective state recognition methods that incorporate non-verbal multimodal cues, such as facial and vocal expressions. Studied several state-of-the-art methods of human behavior recognition including affective and social cues and covered many open computational problems and how they can be efficiently incorporated into a human-computer interaction system. presented a review of state-of-the-art affective recognition methods that use visual and audio cues for recognizing spontaneous affective states and provided a list of related datasets for human affective expression recognition [12].

3. Application of HAR:

There are many applications of successful human action and activity recognition using computer or machine . In this section, we will focus only few applications but list endless due to emerging new advancement in technology. Now a days area like Visual Surveillance, Human-Robot Interaction, Entertainment, Autonomous Driving Vehicle and Video Retrieval are huge demand of HAR

3.1 **Visual Surveillance:** Due to video surveillance camera available private and public places has earned great attention within the computer vision community. For safety, protection and security services has led to more challenging intelligent surveillance work. It has a wide variety of applications, from tracking human activities in public spaces such as waiting rooms, railway stations, clinics, nursing homes, campuses. Evaluation of behavior consists of analysis and identification of movement patterns to provide a high-level interpretation of actions and interactions among objects. Surveillance camera footage is most often used to track suspicious incidents or activities and is sometimes used to avoid potentially dangerous circumstances. [13][14]

3.2 **Human-Robot Interaction Robots** : Human Robot Interaction are used to do wised range of tasks, particularly in human environments which make analysis of characteristic human-robotic interactions more significant. It play vital role in personal indoor assistants, robots must be able to deduce human activities and determine whether or not human assistance is needed. Such an interaction calls for contact between robots and humans. Visual communication [16] is one of the cleverest ways of applying such communication.

3.3 **Entertainment:** HRA became the prominent area for gaming industry for introducing different kinds of gaming consoles specifically games focused on full body motion centered play like dance and sports games increased the gaming platform's attractiveness. To achieve the accurate perception of human behavior, these gaming consoles provide low-cost RGB-Depth sensors such as kinect camera that provide information on the human skeleton in detail. This promotes the action recognition function, which optimizes intra-class activity variations and reduces cluttered background noise[17].

3.4 **Autonomous Driving Vehicle:** Recognition of human behavior within vehicles are becoming increasingly important due to the successful mapping of driver and passenger activities [18] in autonomous vehicle. With the accurate classification about what the driver is doing, the human machine interaction can be directed to the suitable modalities. If the car knows the full body movement of its passengers, the safety functions can be adapted to the in-the-moment deployment for example braking, airbag, crash avoidance etc. In-vehicle activities and driver attention in the driving task can be used by the autonomous vehicle (AV) to decide whether perform a

safe stop. A major concern about fully automated vehicles is that it should provide a design of the interaction between the user and the automatic system or vehicle.

3.5 **Video Retrieval:** Video is the most useful components for information retrieval because sometime picture unable to produce all necessary and sufficient information for event occur but due to the rapid growth of technology, the number of videos available on the Internet is increasing significantly. Videos have become popular with widely available equipment that can make, upload, and share images, high-speed Internet access, and free storage servers. Due to the huge amount of video data being uploaded to the Web every second, the viewing of videos and movies on the Web seems to be quite prevalent today. Such a great deal of video data and its popularity have led to video understanding and action recognition for the retrieval of relevant videos. Content-based video retrieval is used for finding the user desired items among these big video data. Majority of the videos are related to the human actions. So retrieval of human action based videos from the big data of videos is becoming more popular[19].

4. Challenges in HAR :

The Human activity and action are the compliment to each other while considering case of HBU .After analysis the action and activity we summarized following challenges in domain

A) Intra class variation and Inter class Similarities:

Due to no availability of grammar and strict definition for human activities we suffer from two kind of confusions.

i)Same activity may vary from subject to subject, which leads to the intra class variations.

ii) Deferent activities may express similar shapes (e.g., using a laptop and reading) lead to interclass similarities phenomenon in activity recognition. To deal with these problems accurate and distinctive features need to be designed and extracted from videos where the activity takes place.

B) ***Recognition under Real-World Settings***

i. Complex and Various Backgrounds: Generally static cameras are used in video surveillance and fall detection system. In case of Sports event broadcast dynamic recording is use. Gaining popularity of smart devices such as smart glasses and smartphones, people tend to record videos with embedded cameras from wearable devices anytime

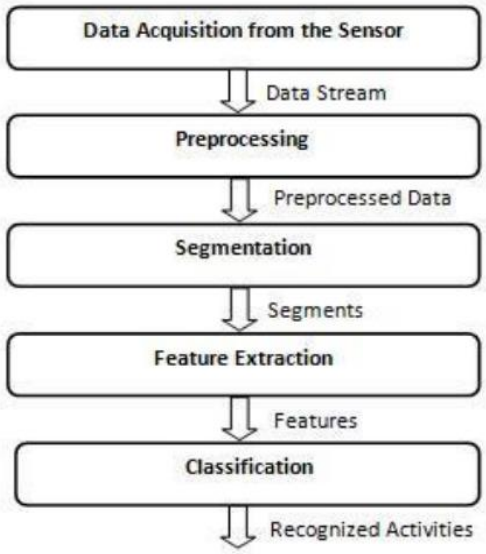
ii.Multi subject Interactions and Group Activities: Early day's researcher focus on low-level human activities such as jumping, running, and waving hands etc. activities consisting a single subject without any human-human or human-object interactions.

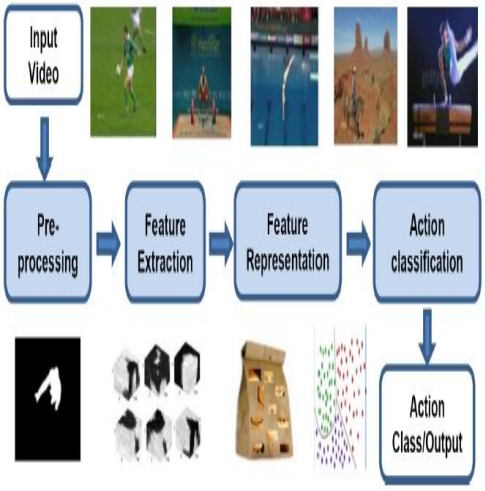
iii. Low-Quality Long-Distance and Videos: Large and crowded places like the railway station, Buss stand , Market, or public place we observed occlusions happen frequently due to surveillance cameras installed in high places cannot provide high-quality videos like present datasets in which the target person is clear and obvious Though we do not expect to track everyone in these cases, some abnormal or crime- related behaviors should be recognized by the HAR system

5. Proposed System :

To reorganization human activity /action from video or image first we capture the data either by wearable or no wearable sensor. Once data has been captured by sensor then based on data stream send for preprocessing. Further video has been break into frame based on scenes and slots shown in fig . for image segmentation. After successfully segmented the image necessary feature need to be extracted while applying different machine learning algorithms under different testing methodology and based on that feature

classification has to done for recognized or predict the particular Action or actity for given video or image or other sample for testing.





6. Discussion & Conclusion

In HBU the action and activity term are interchange as per requirement .We produced human activity recognition model. There is number of literature paper available for action and activity but for HBU is very limited. Finally We summarized the some challenges with proposed model Human Behavior Understanding.

REFERENCES:

1. 1.Zawar Hussain, Michael Sheng ,Different Approaches for Human Activity Recognition– A Survey, 2019
2. 2.https://www.frontiersin.org/articles/10.3389/frobt.2015.00028/full ,2015.
3. 3.Aggarwal, J. K., and Cai, Q. (1999). Human motion analysis: a review. Comput. Vis. Image Understand. 73, 428–440. doi:10.1006/cviu.1998.0744
4. 4.Aggarwal, J. K., and Ryoo, M. S. (2011). Human activity analysis: a review. ACM Comput. Surv. 43, 1–43. doi:10.1145/1922649.1922653
5. Aggarwal, J. K., and Xia, L. (2014). Human activity recognition from 3D data: a review. Pattern Recognit. Lett. 48, 70–80. doi:10.1016/j.patrec.2014.04.011
6. Akata, Z., Perronnin, F., Harchaoui, Z., and Schmid, C. (2013). "Label-embedding for attribute-based classification," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Portland, OR), 819–826.
7. Alahi, A., Ramanathan, V., and Fei-Fei, L. (2014). "Socially-aware large-scale crowd forecasting," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Columbus, OH), 2211–2218.
8. AlZoubi, O., Fossati, D., D'Mello, S. K., and Calvo, R. A. (2013). "Affect detection and classification from the non-stationary physiological data," in Proc. International Conference on Machine Learning and Applications (Portland, OR), 240–245. Amer, M. R., and Todorovic, S. (2012). "Sum-product networks for modeling activities with stochastic structure," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Providence, RI), 1314–1321.
9. Amin, S., Andriluka, M., Rohrbach, M., and Schiele, B. (2013). "Multi-view pictorial structures for 3D human pose estimation," in Proc. British Machine Vision Conference (Bristol), 1–12.
10. Andriluka, M., Pishchulin, L., Gehler, P. V., and Schiele, B. (2014). "2D human pose estimation: new benchmark and state of the art analysis," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Columbus, OH), 3686–3693.

11. Andriluka, M., and Sigal, L. (2012). “Human context: modeling human-human interactions for monocular 3D pose estimation,” in Proc. International Conference on Articulated Motion and Deformable Objects (Mallorca: Springer-Verlag), 260–272.
12. 12.Anirudh, R., Turaga, P., Su, J., and Srivastava, A. (2015). “Elastic functional coding of human actions: from vector-fields to latent variables,” in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Boston, MA), 3147–3155.
13. Michal Koperski. “Human action recognition in videos with local representation”. PhD thesis. 2017.
14. Addwiteey Chrungoo, SS Manimaran, and Balaraman Ravindran. “Activity recognition for natural human robot interaction”. In: International Conference on Social Robotics. Springer. 2014, pp. 84–94.
15. David Vogt et al. “Learning human-robot interactions from human-human demonstrations (with applications in lego rocket assembly)”. In: 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids). IEEE. 2016, pp. 142–143.
16. Chi-Hung Chuan, Ying-Nong Chen, and Kuo-Chin Fan. “Human action recognition based on action forests model using kinect camera”. In: 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA). IEEE. 2016, pp. 914–917. Cristofer Englund and Martin Torstensson. “In-vehicle Driver and Passenger Activity Recognition”. In: SSBA. 2019.
17. Cristofer Englund and Martin Torstensson. “In-vehicle Driver and Passenger Activity Recognition”. In: SSBA. 2019. Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2020 doi:10.20944/preprints202005.0146.v111
18. Mohsen Ramezani and Farzin Yaghmaee. “A review on human action analysis in videos for retrieval applications”. In: Artificial Intelligence Review 46.4 (2016), pp. 485–514.
19. Piotr Tadeusz Bili´nski. “Human action recognition in videos”. PhD thesis. Universit´e Nice Sophia Antipolis,2014