

A Survey on Human Action Recognition

Dr. Vimuktha Evangeleen Sails¹, Ranjana S. Chakrasali², Deepika R Chimkode³

¹Associate Professor, Dept. of Computer Science Engineering, BNMIT, Karnataka, India

²Assistant Professor, Dept. of Computer Science Engineering, BNMIT, Karnataka, India

³Student, Dept. of Computer Science Engineering, BNMIT, Karnataka, India

Abstract - In today's world Surveillance system is playing an important role in the field of security. Moving object detection has been widely used in video surveillance system. As well as motion estimation is an important part of surveillance video processing such as video filtering and compression from video frames. Video Surveillance System is a powerful tool used for monitoring people and their activities for public security. The motive of having surveillance system is not only to put cameras in place of human eyes, but also making it capable for recognizing activities automatically.

Key Words: Image Processing, Video Surveillance, Motion Estimation, Compression, Filtering.

1. INTRODUCTION

Machine learning and understanding of human actions is a complex, diverse, and challenging area that has received much attention within the past ten years (2001–2011). Human activity location, movement following, scene displaying, and conduct understanding (human action acknowledgment and revelation of action designs) have gotten a great deal of consideration in the PC vision and AI people group. Applications have been in-yet not constrained to-video reconnaissance, human PC interfaces, and mixed media semantic comment and ordering. Applications have been in-but not limited to-video surveillance, human-computer interfaces, and multimedia semantic annotation and indexing. Intelligent visual surveillance has got more research attention and funding due to increased global security concerns and an ever increasing need for effective monitoring of public places such as airports, railway stations, shopping malls, crowded sports arenas, military installations, etc., or for use in smart healthcare facilities such as daily activity monitoring and fall detection in old people's homes.

In recent year's surveillance system have become more popular in the field of computer vision. Traditional surveillance systems only provide analog services in hardware. Security guards must stay at security room and look at arrays of CCTV (Closed Circuit Television) or play back the videotapes sequentially to find out the surveillance events. This demanding task is very inefficient. Real time moving object detection is core of surveillance applications. One of the main challenges in these applications is to detect moving objects competently. Moving object detection judges the change in images, captured by a camera and detects

whether any moving object present or not, if there any, extracts the object as soon as possible. Although a number of well-known methods are exist, however problem becomes more difficult to solve in presence of noise, illumination changes, and complex body motion and in real time environment. This Human activities recognition is becoming a field of great interest and relevance to a number of areas of research and applications. An automation system capable of classifying the Human physical is extremely in demand for various applications such as such as smart surveillance and monitoring systems, entertainment, virtual reality, interactive interfaces for mobile services, healthcare systems, automatic scene understanding, human-computer interaction, content-based image, and video retrieval, automatic scene understanding and other vision-based interfaces. Currently, there has been an incredible growth in the volume of computer vision research geared at understanding human activities and behaviors. There are growing interests on recognizing context generated from human such as recognizing someone walking or sitting, to the higher level task of recognizing and interpreting the global behavior of several interacting people. Human action detection and tracking in a video surveillance system is an active research area in image processing and computer vision system. Due to increase in terrorist activities and many general social problems, security has become the top most priorities of all nations. So, there is a need for effective monitoring of public places for security at airports, railway stations, shopping malls, banks, etc. To monitor the activities of a human, surveillance cameras are used. Surveillance cameras are used for monitoring banks, department stores, museums, patrolling of highways and railways for accident detection, for fire detection, patrolling national borders, monitoring peace treaties and so many other applications. They are also used for observing the activities of elderly and infirm people for early alarms. In traditional surveillance system, the video is captured by the camera and is displayed on the monitor in a control room. To monitor the videos, human resources are presents in a control room and continuously monitor the video for recognizing the activities. In many situations, it is common to find poor monitoring due to human factor like fatigue because it is very tedious or boring task to continuously monitor the scene because sometimes nothing strange or uncommon thing happens in a scene that catches the attention. So, there is a need to design an intelligent surveillance system that can automatically detect and track the object (i.e. Human in our case) in a video. The aim is to describe different methods used for the

human detection, tracking and recognition and advantages & disadvantages with specified accuracy of each method.

2. Human activity detection, tracking and recognition methods:

In this paper [1] a basic and proficient reconnaissance framework dependent on movement identification with movement vector estimation from observation video outlines. Movement is recognized with another methodology edge based approach which makes identification quicker. The observation video is then handled for movement estimation utilizing optical stream with Horn-Schunck calculation for assessing movement vector for its sensible execution and straightforwardness. This technique is computationally quicker without requiring any extraordinary equipment for picture preparing. So it very well may be progressively appropriate to installed frameworks. A reconnaissance framework is proposed where movement discovery conspire is an edge based methodology. Edges are strong against enlightenment changes and clamor.

In this paper [2] a human instance acknowledgment framework for video reconnaissance utilizing one static camera. The preparation and testing stages were executed utilizing four distinct classifiers which are K Means, Fuzzy C Means, and Multilayer Perceptron Self Organizing Maps and Feedforward Neural systems. The precision acknowledgment of utilized classifiers is determined. In addition, results demonstrate that directed learning classifiers will in general perform superior to unsupervised classifiers for the instance of human stance acknowledgment. Moreover, for every individual classifier, the acknowledgment rate has been observed to be relative to the quantity of preparing stances. The camcorder is the information procurement gadget which for this situation is an advanced camera running in video recording mode. The video arrangements recorded from the camcorder are changed over into datasets of static shading pictures (one picture compares to one edge of a video succession). The pre-handled pictures are prepared and various classifiers were assessed. The exhibitions of these classifiers were thought about. In this work, the following classifiers were used:

- K means
- Fuzzy C Means (FCM)
- Multilayer Perceptron Feed-forward Neural Networks (MLP – NNs)
- Self-organizing maps (SOMs)
- K nearest neighbour (KNN)

The video is ceaselessly caught by the observation camera conveyed at the appropriate areas at better places. The stream of work can be clarified well by the well-ordered methodology as given underneath:

Background Modelling: Foundation displaying is a strategy for removing the moving item in video outlines. There are different techniques to do Background demonstrating. Amid changing over the video into casings, some of the time we can get a casing which we can use as a foundation. On the off chance that there is no single casing which we can use as a foundation, at that point we have to demonstrate the foundation by utilizing a few strategies, for example, taking mean or middle of n number of edges, Adaptive Gaussian blend display.

Human Detection: In Human Detection, the human is recognized in the zone under reconnaissance. There are different calculations that are utilized for human recognition like Optical Flow calculation, Background Subtraction calculation, Temporal Differencing, Gaussian blend display, Min-Max technique, Kernel thickness estimation (KDE), Eigen foundations, Codebook (CBRGB), thus any others. Optical stream calculation can be utilized to identify autonomously moving focuses within the sight of camera movement, anyway Optical stream technique is unpredictable and touchy to clamor and isn't material to constant calculations. Transient differencing do the pixel-wise contrast between a few back to back edges in a picture succession to separate moving districts. The upside of this system is that it is versatile to dynamic conditions, however completes a poor employment of separating all applicable component pixels. Foundation subtraction is a prominent strategy for human location where the foundation is static in nature and it endeavours to identify moving districts in a picture by differencing between current picture and a reference foundation picture in a pixel-by-pixel way. In any case, it is very delicate to changes of dynamic scenes because of helping and incidental occasions. In this way, it is very reliant on a decent foundation model to lessen the impact landscape changes.

Human Tracking: Human Tracking implies determining a correspondence of the human identified in one edge with the human recognized in the following casing. On the off chance that the highlights are coordinated, at that point the human distinguished in the current and the past edge is said to be the equivalent. Valuable scientific devices for following incorporate the Kalman channel, Condensation calculation, dynamic Bayesian system, Camshift, Meanshift, Particle channel, and so forth. Following techniques are partitioned into four noteworthy classifications: region based following, dynamic shape based following, highlight based following and model-based following. In district based following, the highlights of the mass, recognized in one picture outline are coordinated to the mass identified in the other casing. On the off chance that there is a match, at that point the distinguished picture is connected with the picture in the past casing. District based following functions admirably in scenes containing just couple of items, however can't dependably deal with impediment between articles. Dynamic form based following calculations track questions

by speaking to their diagrams as bouncing shapes. The upside of dynamic form based calculations is that they depict protests all the more basically and all the more successfully and lessen computational multifaceted nature.

A new posture recognition method is proposed in paper [3]. The method uses only two devices to achieve its function: a laptop computer and a Kinect sensor. The Kinect sensor comprises of a profundity sensor, a RGB camera, a multi exhibit amplifier and a mechanized tilt. The profundity sensor is made out of an infrared beam producer and a monochrome CMOS sensor to catch profundity pictures with a goals of 320*240 pixels; the RGB camera is utilized to catch shading pictures with a goals of 640*480 pixels. The USB port is utilized for correspondence between the Workstation the Kinect sensor. The picture preparing procedures utilized incorporate the flat and vertical projection, star skeleton, LVQ neural system and picture handling strategies. This investigation adds to look into about programmed home care frameworks. Older individuals who live alone can regularly advantage from a robot to give home consideration administrations. These robots must have a capacity to perceive the individual's poses in typical and risky circumstances, so as to send precise reports to the consideration focus.

In this paper [4] Background subtraction is a key pre-processing step for several video automatic operations. Various techniques have been proposed to perform background subtraction automatically in complex environments. Visual background extractor (ViBe) is a popular background subtraction technique that can initialize its model in a single frame, adapt to the environment changes and achieve satisfactory subtraction results. ViBe performs background subtraction on pixel level, mainly containing three parts, the background model initialization, the foreground detection and the model updating. As an advantage of the ViBe technique, the background model is initialized from a single frame, which is actually a spatial-temporal process of collecting a series of pixel samples.

3. Merits and Demerits of methods discussed

An observation framework dependent on new movement location approach and movement estimation from reconnaissance video successions. The movement recognition conspire in the observation framework is computationally quicker and movement location precision is about 98.6%. Movement estimation approach is finished by Horn-Schunck optical stream calculation. This gives high thickness of movement vector and effortlessness in usage. In this way, it needs less computational expense. Furthermore, no morphological activities are performed which makes our framework quick. It is free from light changes like indoor, outside, bright, foggy, night, day and so forth and moving camera condition. The disadvantage here is it does not work for 3D motion estimation with tracking, recognition and classification of moving object.

A strategy for human instance usage which comprises of two phases: preparing and assessment organize whereby models with ideal parameters is assembled and arrangement organize where the scholarly model is mimicked for any future contribution to ongoing. The preparation and assessment arrange, thusly, is comprised of various sub-stages what's more, in a similar token, the sending stage additionally comprises a arrangement of littler sub-stages. The fundamental assessment measure utilized in this paper is 'acknowledgment exactness' which measures the precision of a stance acknowledgment show.

Video observation framework is a functioning examination zone in PC vision framework due to its different applications in open security, in military security, bank security, sports, etc. In this way, there is a need to plan a keen reconnaissance framework that ought to be fit to distinguish, track end perceive the exercises of people naturally. In this work, as a matter

Of first importance foundation estimation/demonstrating has been done on the video caught by the camera by taking mean of n outlines. After this, human identification has been finished utilizing foundation subtraction calculation and afterward morphology activity is connected.

3. Conclusion

In this survey, we discussed about the different methods used for human activity detection. The various advantages and disadvantages of each methods and also its accuracy.

REFERENCES

- [1] Hossen, Muhammad Kamal, and Sabrina Hoque Tuli. "A surveillance system based on motion detection and motion estimation using optical flow." In *Informatics, Electronics and Vision (ICIEV)*, 2016 5th International Conference on, pp. 646-651. IEEE, 2016.
- [2] Htike, Kyaw Kyaw, Othman O. Khalifa, Huda Adibah Mohd Ramli, and Mohammad AM Abushariah. "Human activity recognition for video surveillance using sequences of postures." In *e-Technologies and Networks for Development (ICeND)*, 2014 Third International Conference on, pp. 79-82. IEEE, 2014.
- [3] Kaur, Rajvir, and Sonit Singh. "Background modelling, detection and tracking of human in video surveillance system." In *Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH)*, 2014 Innovative Applications of, pp. 54-58. IEEE, 2014.
- [4] Wang, Wen-June, Jun-Wei Chang, Shih-Fu Haung, and Rong-Jyue Wang. "Human posture recognition based on images captured by the Kinect sensor." *International Journal of Advanced Robotic Systems* 13, no. 2 (2016): 54.
- [5] Wang, Haixia, and Li Shi. "Foreground model for background subtraction with blind updating." In *Signal*

and Image Processing (ICSIP), IEEE International Conference on, pp. 74-78. IEEE, 2016.

[6] Mohanad Babiker, Othman O. khalifa, Kyaw Kyaw Htike, Aisha Hassan, Muhamed Zaharadeen. "Automated Daily Human Activity Recognition for Video Surveillance Using Neural Network." In *Smart instrumentation, measurement and applications (ICSIMS)*, IEEE International Conference on, pp. 28-30. IEEE, 2017