International Journal of Engineering Research in Electronic and Communication Engineering (IJERECE) Vol 3, Issue 5, May 2016

Multiple Human Tracking Using Blob Extraction and Action Recognition Using HOG

[1] Shyma Zaidi, [2] Jagadeesh B
[1] PG Scholar, [2] Asst. Professor,
Department of ECE, Vidya Vardhaka College of Engineering [VVCE],
[1] shymazaidi@gmail.com [2] jagadeesh.b@vvce.ac.in

Abstract: **- Human tracking is an important research area in computer vision with potential applications in many other fields like augment reality, human–machine interaction, and advanced driver assistance systems. Despite a lot of progress in the field, visual tracking remains a difficult problem due to many challenges. The proposed work provides a solution for multiple human tracking methods, which uses a combination of Blob extraction and Kalman filtering. The main objectives of the proposed work are to precisely track moving or static human beings and to identify and estimate their future location in an unknown scene. A pre recorded input video is split into individual frames which are actually processed to track the human beings and recognize their actions. The background modeling is done using Blob extraction followed by Kalman filtering & Optical flow to detect the positions of humans & track them. HOG features are used to classify and recognize various human actions. The results are the frames of the video file, which consists of marking the borders of the human beings appearing in the video, tracking them and displaying their actions such as walking, sitting and throwing on the command window of Matlab.**

Keywords **– Human Tracking, Blob extraction, Kalman filter, Optical flow, Background Modeling, Motion vector.**

I. INTRODUCTION

The performance of human tracking systems has steadily increased during the past few years. Two factors mainly contributed to the improvement: the advance of robust detectors and various extensions of the data association technique. The proposed work focuses on the importance of the appearance model, which is orthogonal to the approaches of previous works on multi-target tracking, but is a key problem in single object tracking applications. Human tracking algorithms aim to find the most similar region of an image as compared with the target, by detection, predicting the position of an object and simultaneously adapting its parameters to different appearances of object.

Challenges will be seen during occlusions. Challenges are also faced when features for the human beings are difficult to discriminate against that extracted from other objects in the scene. Another challenge is the variant appearance of the human beings, e.g., dress color and texture. All these factors are important in the uniformity of human tracking. They are commonly affected by pose variation, illumination change, camera motion and different viewpoints. Approaches maintaining a template for each human being typically face how to update an existing human being such that it remains a representative model. If the human being position is fixed, it can't be an effective model for different situations.

Sometimes human beings often are occluded by other object(s) or they leave the field of view of the camera. To deal with these circumstances, the detection of the human being independently of his previous position in the image is required; in addition, the required execution time is another difficulty. Our goal is to automatically determine the bounding box of human being, which is not necessary in every frame that follows. The video stream is to be processed at a frame rate and the process should execute indefinitely long. This task is called as long-term tracking. In the proposed work, a novel approach combining the strategies of tracking, detection and recognizing the actions of the human beings.

II. LITERATURE SURVEY

Many works have been proposed on multiple objects tracking in general and multiple human tracking in particular. Similarly the human action recognition is also a domain of scholars' interest. As the proposed project deals with both multiple human tracking and human action recognition, papers on both the above specified topics have been studied and listed below.

Guicong Xu, Xiangmin Xu, Xiaofen Xing, Bolun Cai, Chunmei Qing[1] combined intensity information

(cross-bin distribute field, CDF), texture information (enhance histograms of oriented gradients, EHOG) and color information (color name, CN) in a tracking-by detection framework, in which a simple tracker called CSK is extended for multi-dimension and multi-cue fusion. The approach improved the baseline single-cue tracker by 4.4% in distance precision. Furthermore, the proposed approach achieving 75.4% is better than most recent state-of-the-art tracking algorithms.

.Zhi Liu, Fujie Wang, Yun Zhang[2] focused on a problem of adaptive visual tracking control for an uncalibrated image-based visual servoing manipulator system with actuator fuzzy dead-zone constrain and unknown dynamic. Without a prior knowledge of the system, fuzzy logic systems were employed to approximate the unmodeled nonlinear manipulator dynamics and external disturbances. It was shown that by using the recursive Newton–Euler method, the total number of fuzzy rules can be reduced significantly as compared with the traditional fuzzy logic system. Experimental results were carried out to test the visual tracking performance of the proposed controller and the boundedness of the closed-loop system.

Xinwu Liang, Hesheng Wang, Yun Hui Liu, Weidong Chen[3] addressed the issue of the adaptive image-based trajectory tracking control problem of rigid-link electrically driven (RLED) robotic manipulators. Adaptive laws were designed to handle uncertainties existing in the camera parameters, manipulator dynamic parameters and motor dynamic parameters. The proposed approach can deal with 3D motion control problem of uncertain robotic manipulators such that time varying depths of feature points can be allowed, which is the main contribution and also the major difference compared with previous works.

Sarang Khim, Sungjin Hong, YoongYoung Kim, Phill Kye Rhee[4] published the paper in the year 2014, in which an adaptive visual tracking method that combines the adaptive appearance model and the optimization capability of the Markov decision process was introduced. In this paper, the Markov decision process, which has been applied successfully in many dynamic systems, is employed to optimize an adaptive appearance model-based tracking algorithm. The adaptive visual tracking is formulated as a Markov decision process based dynamic parameter optimization problem with uncertain and incomplete information. The high computation requirements of the Markov decision process formulation are solved by the proposed prioritized Q-learning approach. Extensive experiments using realistic video sets were carried out, and they achieved very encouraging and competitive results.

Peng Wang, Jianhua Su, Wang Li, Hong[5] presented a robust adaptive visual tracker which is able to capture the varying appearance of target under different environments without gradual drift. They proposed a novel and flexible feature space evaluation function which was formed by the weighted sum of two components: the similarity measure and the discriminating ability measure. The proposed discriminative feature selection and target model updating mechanism was embedded in a Mean-shift tracking system which iteratively finds the nearest local optimal localization of target. Experimental results on a mobile robot system demonstrated the robust performance of the proposed algorithm under different challenging conditions.

Tahir Nawaz and Andrea Cavallaro[6] proposed measure evaluated tracking accuracy and failure, and combined them for both summative and formative performance assessment. The proposed protocol is composed of a set of trials that evaluate the robustness of trackers on a range of test scenarios representing several real-world conditions. The protocol was validated on a set of sequences with a diversity of targets (head, vehicle, person) and challenges (occlusions, background clutter, pose changes, scale changes) using six state of- the-art trackers, highlighting their strengths and weaknesses on more than 187000 frames. The software implementing the protocol and the evaluation results are made available online and new results can be included, thus facilitating the comparison of trackers.

Pushe Zhao, Hongbo Zhu, He Li, and Tadashi Shibata[7] provided a solution to the object tracking task that considers an efficient implementation was explored as the first priority. A hardware-friendly tracking framework was established and implemented on field-programmable gate array (FPGA), thus verifying its compatibility with very large scale integration (VLSI) technology. They proposed a real-time object tracking system, which was based on the multiple candidate-location generation mechanism. The system employed the directional edge-based image features and also an online learning algorithm for robust tracking performance.

Samuele Salti, Andrea Cavallaro, Luigi Di Stefano[8] proposed a unified framework for model adaptation in video tracking as well as a comprehensive evaluation of recent solutions on a heterogeneous dataset was presented. The evaluation was based on a novel objective methodology that allowed one to compare adaptive trackers without the need of thresholds. Performance of adaptive trackers was discussed, together with the various design choices using the building blocks identified in the proposed framework. This systematic

survey allowed us to identify important open challenges, related to on-line feature evaluation, on-line parameter tuning and dealing with multi-modality.

Christina J. Howard, David Masomb, Alex O. Holcombe[9] presented the paper in the on multiple object tracking (MOT). In which it was observed that, in the MOT task, observers can typically keep track of up to four moving objects. But very little is known about the extent to which object motion is used by observers during MOT. Using a range of different motion trajectory patterns, observers tracked 1–4 targets among distracters and reported the final position of one of the targets. On average, reports corresponded to previous positions rather than the final position. The significant increase in lag with speed of the object was consistent with slow or intermittent updating of object positions during tracking.

Alexandros Makris, Dimitrios Kosmopoulos, Stavros Perantonis, Sergios Theodoridis [10] proposed a framework for building particle filtering based tracking algorithms. These algorithms enhanced significantly previous approaches in terms of robustness and speed by fusing several cues hierarchically in order to achieve robust tracking in various challenging situations including occlusions, deformations, abrupt motion, illumination changes and cluttered background. The proposed work acknowledged the problem and tried to tackle it by building priors using low dimensional object models at the beginning and using them to guide the most complex ones.

The common drawbacks from all the previous works are listed as follows:

- They were efficient for detecting single human being, not good results were recorded for multiple targets.
- As the crowd number increased, the algorithms were bound to become less accurate and record more false detections.
- Occlusions were not addressed in most of the works.
- Practical implementation of Robust & Real time algorithm was very complicated.
- Computational complexity was very high.

The objectives of the proposed work are:

- To design a simple and stable algorithm that can detect multiple human beings.
- To be capable of predicting & tracking the motion of the human beings in the frame under processing.
- To yield better results for crowds.
- To deliver steady performance even in case of sudden movements.
- To recognize the actions performed by the human beings in the frames under consideration.

III. DESIGN METHODOLOGY

The work carried out has two parts: Multiple human tracking and their action recognition. Tracking is performed using blob analysis with Kalman filter and optical flow technique. Recognition of actions is done by features extracted by HOG (Histogram Of Gradients) which uses SVM (Support Vector Machine) for classification of actions.

Fig 1 shows the block diagram of proposed tracking method. Blob extraction is carried out using Background modeling which is used for Human tracking.

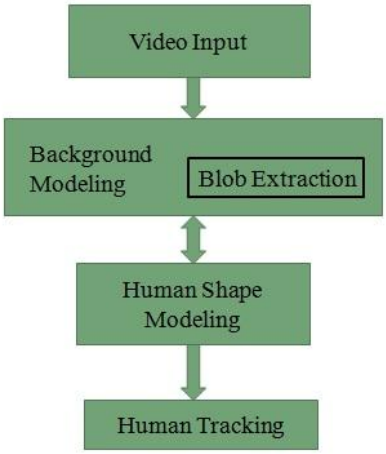

Fig 1: Block diagram of tracking method

The common idea of background segmentation is to involuntarily create a binary mask which splits the set of pixels into the set of foreground and the set of background pixels. BLOB stands for Binary Large OBject. It refers to a group of connected pixels in a binary image. Term “Large” indicates that only objects of a certain size are of interest. “Small” binary objects are usually noise. There are three fundamental steps: Extraction, Refinement and Analysis. In extraction stage each blob is represented by characteristic denoted features and a comparison method is applied to compare features. Refinement is done for removal of noisy areas and gaps. It uses morphological operations like opening, closing and fill techniques. In analysis stage tracking of wanted blobs and calculation the centroid and compactness of the blobs are done.

After the parameters initialization is done, a first foreground detection can be done and then the parameters are updated effectively. A background pixel keeps up a correspondence to a high weight with a weak variance due to the fact that the backdrop is more present than moving objects and that its value is practically steady. Once the

parameters maintenance is made, foreground detection can be made and so on.

Kalman filter & Optical flow methods are used to track the humans as they have many advantages than other tracking methods. It is observed that most of the algorithm dependent on application environment and very less resistant to noise. A good filtering algorithm can remove noise from the data and retain useful information mainly Kalman Filter. An optimal recursive data processing algorithm has been taken for this tracking problem. The Kalman filter not only works well in practice, but it is theoretically attractive because it can be shown that of all possible filters, it is the one that minimizes the variance of the estimation error.

A filter which is based on state space technique and recursive algorithm is a recursive predictive filter. It is named so that it does not require storing all the previous measurement and process the data optimally. Since the time of its introduction, the Kalman filter has been the subject of extensive research and application, particularly in the area of autonomous or assisted navigation. The main advantages of Kalman filter innovations are:

- The variance of the Kalman filter innovations is smaller than the variance of the deterministic innovations.
- The computation time of Kalman filter is less.

Generally it is two step process (i) Prediction (ii) Correction. First the state is predicted with the dynamic model then it is corrected with the observation model so that the error co-variance is minimized. The process is repeated with the each time with the step as the previous step as initial value.

Optical flow is the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer (an eye or a camera) and the scene. Here this method is used to track human beings. It is the displacement field for each of the pixels in an image sequence. It is the distribution of the apparent velocities of humans in an image. By estimating optical flow between video frames, one can measure the velocities of objects or humans in the video.

In general, moving objects or humans that are closer to the camera will display more apparent motion than distant objects that are moving at the same speed. Optical flow estimation is used in computer vision to characterize and quantify the motion of objects in a video stream, often for motion-based object detection and tracking systems. The experimental brightness of any object point is constant over time. Close to points in the image plane move in a similar manner (the velocity smoothness constraint). Suppose we have a continuous image; f{x, y, t) refers to the gray-level of (x, y) at time t. Representing a dynamic image as a function of position and time permits it to be expressed.

- Assume each pixel moves but does not change intensity
- Pixel at location (x, y) in frame1 is pixel at (x+Δx, y+Δy) in frame2.
- Optic flow associates displacement vector with each pixel.

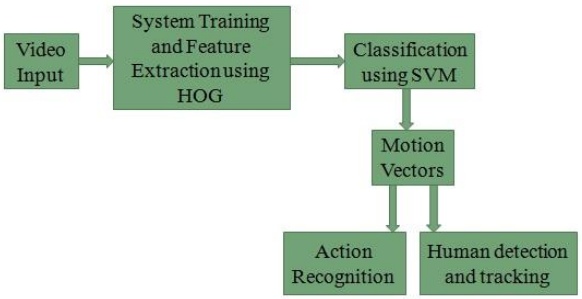

Fig 2: Block diagram of human action recognition

Fig 2 shows the block diagram representation of human action recognition. It uses HOG to extract features and by the knowledge of motion vectors SVM classifies and recognizes the actions. HOG is a feature descriptor used in image and video processing to detect an object.

HOG has five main steps. First step of calculation is the computation of the gradient values usually using sobel operators. The second step is creating the cell histograms. Each pixel within the cell casts a weighted vote for an orientation-based histogram channel based on the values of gradient computation. As for the vote weight, pixel contribution can either be the gradient magnitude itself, or some function of the magnitude.

Third step is descriptor blocks which are done to account for changes in illumination and contrast, the gradient strengths is locally normalized, which requires grouping the cells together into larger, spatially connected blocks. The HOG descriptor is then the concatenated vector of the components of the normalized cell histograms from all of the block regions. These blocks typically overlap, meaning that each cell contributes more than once to the final descriptor.

Next is Normalization factor which can be one of the following, L2-norm, L2-hys, L1-norm, and L1-sqrt depending on the characteristics of video frames. The final step in object recognition using HOG descriptors is to feed the descriptors into some recognition system based on

supervised learning. SVM classifier is a binary classifier which looks for an optimal hyperplane as a decision function. Once trained on images containing some particular object, the SVM classifier can make decisions regarding the presence of an object, such as a human, in additional test images.

IV. IMPLEMENTATION AND RESULTS

The implementation is done using Matlab 2014. Initially the given video is spilt into number of frames according to standard frame rate. The selection points are marked on the positions where humans are recognized using the generated Motion vectors. The positions & locations of the humans are marked using background modeling which is done using blob extraction, which are identified as Threshold outputs. A bounding box is drawn for the selection points obtained after Thresholding. HOG features are classified using SVM and different actions are recognized.

In Video Processing, different parameters of videos play a huge role. To account that, work is carried out on different types of videos. Here, different videos are processed and tracking algorithm is implemented.
The properties and other characteristics of the videos are shown in table 1:

Table 1: Different characteristics of the videos

Sl. No.	Size & Duration	Number of Person	Movement Speed	Presence of crowd	Presence of other objects
1.	21.7MB, 14secs	5	Moderate	No	No
2.	2.49MB, 16secs	4	Fast	No	Yes
3.	3.25MB, 20secs	13	Fast	Yes	Yes
4.	1.67MB 22secs	15	Fast	Yes	Yes
5.	2.46MB 19secs	5	Moderate	Partial	No

Fig 3 shows the output of video 1 in which 3 persons are being tracked successfully and walking action is recognized. It can be seen that the person seen behind the tree in the frame is also being tracked.

Fig 3: Tracking and walking results of video 1.

Fig 4 shows the results of video 2. In this case, a tennis video is considered which consists of 4 tennis players who exhibit sudden movements unlike the first case. The video quality is moderate which is of the duration 16 seconds and occupy 2.49Mb of memory. Even though the players are showing sudden movements, the algorithm was still able to track their positions successfully and mark the bounding box in the proper positions. This video involved movement of ball, which is of small size along with the movements of players. The algorithm was able to track only players ignoring the ball movements. The action recognized here is the person hitting the ball.

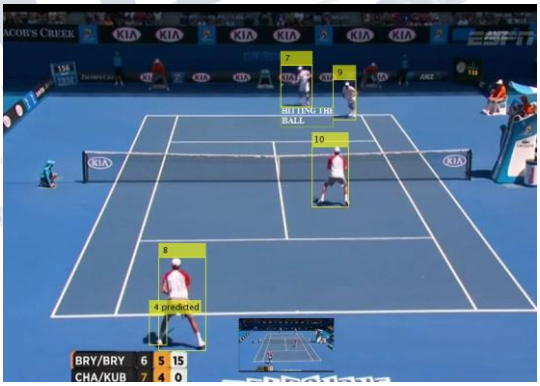

Fig 4: Results of video 2.

A real time video, extracted from the CCTV footage of the Institution is considered in video 5. Even though it does not involve any sudden movements of human or any other object movements in the video, the results obtained were quiet convincing, with humans being tracked successfully with proper area of bounding box. It has been reflected in the fig 5. Two persons sitting has been successfully recognized in fig 6.

Fig 5: Tracking Results of real-time Video 5.

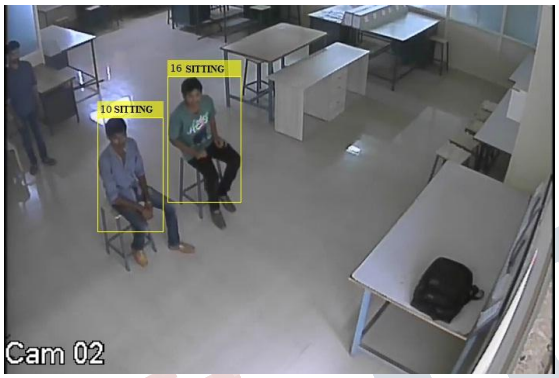

Fig 6: Sitting results of real-time video 5.

VI. CONCLUSION AND FUTURE SCOPE

The proposed work gives a simple and stable solution for Multiple Human Tracking. A combination of blob extraction and Optical flow along with Kalman filtering turned out to give effective results with very lesser rates of false negatives and action recognition using HOG-SVM combination. Future work can be done on developing a better algorithm which can effectively detect every individual in crowds.

REFERENCES

[1] Guicong Xu, Xiangmin Xu, Xiaofen Xing, Bolun Cai, Chunmei Qing, "*Multi-invariance Appearance Model for Object Tracking"* 2015.

[2] Zhi Liu, Fujie Wang, Yun Zhang**,** *"Adaptive Visual Tracking Control for Manipulator With Actuator Fuzzy Dead-Zone Constraint and Unmodeled Dynamic"* 2015.

[3] Xinwu Liang, Hesheng Wang, Yun Hui Liu, Weidong Chen, *"Adaptive Visual Tracking Control of Uncertain Rigid-Link Electrically Driven Robotic Manipulators with an Uncalibrated Fixed Camera"* 2014.

[4] Sarang Khim, Sungjin Hong, YoongYoung Kim, Phill Kye Rhee, *"Adaptive visual tracking using the prioritized Q-learning algorithm: MDP-based parameter learning approach"* 2014.

[5] Peng Wang, Jianhua Su, Wang Li, Hong Qiao, *"Adaptive visual tracking based on Discriminative Feature Selection for Mobile Robot"* 2014.

[6] Tahir Nawaz and Andrea Cavallaro, *"A Protocol for Evaluating Video TrackersUnder Real-World Conditions"* 2013.

[7] Pushe Zhao, Hongbo Zhu, He Li, and Tadashi Shibata**,** *"A Directional-Edge-Based Real-Time Object Tracking System Employing Multiple Candidate-Location Generation"* 2013.

[8] Samuele Salti, Andrea Cavallaro, Luigi Di Stefano, *"Adaptive Appearance Modeling for Video Tracking: Survey and Evaluation"* 2012.

[9] Christina J. Howard, David Masomb, Alex O. Holcombe *"Position representations lag behind targets in multiple object tracking"* 2011.

[10] Alexandros Makris, Dimitrios Kosmopoulos, Stavros Perantonis, Sergios Theodoridis, **"***A hierarchical feature fusion framework for adaptive visual tracking"* 2011