Proc. of Int. Conf. on Current Trends in Eng., Science and Technology, ICCTEST

Multiple human tracking using Blob analysis and Kalman filtering

Jagadeesh B[1] and Dr. Chandrashekar M Patil[2]
[1]Vidyavardhaka College of Engineering, Department of Electronics & Communication, Mysuru, India
Email: jagadeesh.b@vvce.ac.in
[2]Vidyavardhaka College of Engineering, Department of Electronics & Communication, Mysuru, India
Email: patilcm@vvce.ac.in

***Abstract*—Computer Vision research field is gaining more importance with wide applications in video surveillance, video retrieval and analysis. Video surveillance provides continuous monitoring enhancing security and control. The proposed algorithm detects the moving object; classify the moving object as human and keep track of the human. Human tracking plays an integral role in many fields of human–machine interaction. Even though there is increase of progress in this field, visual tracking still remains a challenging task. Proposed tracking algorithm uses a combination of Blob extraction and Kalman filtering. Precise tracking of moving human beings, identification and estimation of their future location in an unknown scene are the main objectives of the proposed work. The input video is split into individual frames. Blob extraction is used for modeling followed by Kalman filtering where positions of humans are detected and tracked, followed by validation.**

***Index Terms*— Human tracking; Blob extraction; Kalman filtering**

I. INTRODUCTION

During the past few years, the performance of human tracking systems has increased steadily. The advancement in robust detectors along with other extensions of the data association methods has become two main factors that contribute this improvement. The proposed work concentrates on the importance of the appearance model, which is independent of the approaches used by previous works on tracking multiple targets. The nearest match for a given target object is detected in an image and the position of the object is predicted, meanwhile its parameters are adapted to different types of appearances of the object, which serve the purpose of the Human tracking algorithms.

There has been a slow progress in the development of algorithms for multiple human tracking beyond applications of the particle filter. The accuracy of the detector used limits the performance for this category of tracking algorithms. In computer vision, creating a detector that can give best performance has become an unsolved problem. Based on the data structures obtained, an objective function is computed which is global in nature and the best set of tracks is searched to minimize this function. In order to contract the search space and rectify the errors produced during detection stage, many constraints like track continuity and motion smoothness are being imposed. However, tracking algorithms show better efficiency and encode with much ease and filter high-order states. Tracking algorithms usually need lot of tuning process for the process of model selection. If it is not employed, the optimal solution developed for the prescribed objective will not be

Grenze ID: 02.ICCTEST.2017.1.151

the desired solution.
When it is difficult to discriminate the features of human beings in a given scene against those retrieved from various other objects, challenges are faced. Challenges will also be seen during occlusions variations in the appearance of human beings. The uniformity of human tracking depends on all of these factors. They will be usually affected by variation in the pose, changes in illumination, motion of camera, difference in viewpoints, etc. A defined template for each human being is maintained. Such approaches will face the problem in updating an existing human being such that will remain as a representative model. An effective model is not obtained when the position of the human being is fixed.
Human beings often suffer occlusion by other objects or sometimes they move out of camera’s field of view. The detection of the human being, independent of his previous position in the given image is required, to deal with such circumstances. The required execution time adds difficulty to this. The main aim is to draw the bounding box automatically around the human being, which is not required in every following frame. The video must be processed at a defined frame rate while this procedure must be executed as long as possible (long-term tracking). In the proposed work, an approach that combines the key strategies of tracking is been put forth.

Tracking of human beings in videos

Human tracking is implemented on a video sequence, which involves the analysis of the presence, size, shape and position of human beings. Human tracking is generally used in applications like video surveillance, crowd monitoring. Videos are series of images, called frames, which are displayed in faster frequency, such that human eyes will not be able to percept the continuous contents. Usually, the information of the contents present in two consecutive frames will be closely related. The first step is to identify the regions of interest. A general human detection algorithm is desired, but sometimes it will be difficult to handle the humans properly with considerable amount of variations. Hence, many practical systems prefer stationary camera, which makes human detection procedure more straightforward.
The different steps for human tracking described in many works is as shown in figure 1. It is difficult because the definition is totally specific to the application. Usually, foreground objects are nothing but moving objects and everything else will be considered as background. Sometimes even shadow will be classified as a foreground object, that can result in improper results.

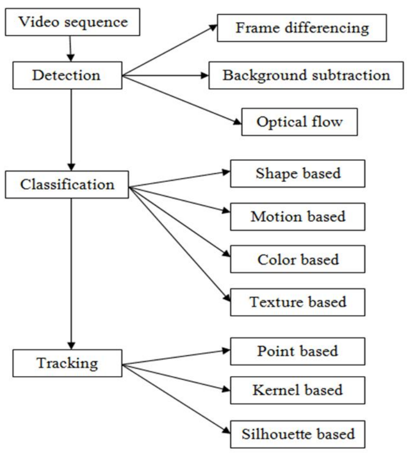

Figure 1: Basic steps in human tracking

Human Detection

This step involves identification of human beings in a given video sequence by clustering similar pixels. Since human moving is basically the primary source of information, many methods concentrates on motion detection. This can be achieved by using various methods such as:
i. Frame differencing: Here moving humans are detected by computing the difference of consecutive images. This method has its own simplicity and easily implemented. But generally it is tough to draw the complete outline of human motion because of the dynamic environments and hence result the obtained is not accurate.

ii. Optical flow: This method will compute the optical flow field of the image, and carries out clustering procedure as per the distribution characteristics of optical flow. This method gives information of movement and detects the moving human from the background better. But this method is sensitive to noise and possess more calculations, and hence not preferred for real time applications.
iii. Background subtraction: In this method, background modeling remains as the core of the algorithm. The modeling should be sensitive enough to identify moving humans. The differences present in the current video frame with a reference frame, which should be in terms of pixels represents moving humans. This method is sensitive to the changes in external environment. However, it gives the best information, provided background should be known. Background subtraction can be Recursive or Non Recursive.

Human classification

In this step, the different approaches are considered to classify as human beings. Classification methods are usually based on any of the methods described below.
i. Shape-based classification: Different information that describes the shape of the motion regions like blob, box and point representations, are used for classification of moving humans. Blob area extracted from the image, aspect ratio of blob bounding box will be the inputs to this method. Classification is carried out on each blob in each frame and the results are displayed in histogram.
ii. Motion-based classification: Optical flow an example for this type of classification. Residual flow is used for the analysis of rigidity as well as periodicity of moving entities. Usually rigid objects will present very less flow of residue while a moving which is not rigid object (human being) show larger values of average residual flow.
iii. Color-based classification: Color is easy to acquire and will be relatively constant under viewpoint changes. This algorithms possess low computational cost, and histogram based techniques are used to detect humans.
iv. Texture-based classification: This method counts the number of occurrence of gradient orientation in those parts of an image which are localized. Uniformly spaced cells on a grid compute this, which are densely spaced and make use of overlapping local contrast normalization so as to improve the accuracy.

Human Tracking

Tracking is defined as the approximate the path traversed by the human being in the given image plane as its moving around a scene. The objective is to generate the route for the human being by finding his position in each frame. Human tracking classification is done as:
i. Point tracking: Here moving humans during tracking phase are represented by their feature points. Point tracking results in false detection during occlusions. Examples are Kalman Filtering method and Particle filtering methods.
ii. Kernel based tracking: Here tracking is achieved by computation of moving human, which is actually represents an embryonic region between successive frames. Motion will be seen as parametric motion like affine, translation, etc. Sometimes parts of the human could be rejected outside the shape that has been defined. Few background points sometimes could stay inside. Examples are SVM and Simple Template matching.
iii. Silhouette based tracking approach**:** Some humans will possess complicated shapes in their hands, fingers, shoulders that could not be defined properly by simple geometrical shapes. This method provides accurate shape description. The main aim is to identify the human in every frame by a model which is created by previous frames. This method can deal with variety of shapes and Occlusions.

II. Blob Analysis And Kalman Filtering

The algorithm applied for multiple human tracking is blob extraction. Blob extraction is carried out using Background modeling. Figure 2 shows the representation of the proposed work for human tracking.

BLOB Anaylsis

Blob Analysis is a fundamental method which analyses the image regions that show consistency. They are used for the accurate analysis that can determine humans clearly from the background. Blob Analysis provides solution for a variety of visual inspection problems. A key advantage includes greater flexibility and it can perform efficiently. But it requires clear background-foreground relationship and pixel-precision, which can be considered as limitations.

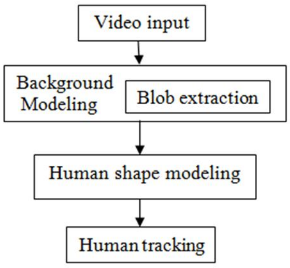

Figure 2 : Representation of the proposed work for human tracking.

The common idea of background segmentation is to involuntarily create a binary mask which splits the group of pixels into two sets, foreground and background pixels. A stationary background frame can be matched to current frame in uncomplicated cases. Pixels with high deviation are identified as foreground. This simple method might work in few specialized scenarios. Every pixel is characterized by its intensity in the RGB space. The procedure for Blob Analysis consists of the following steps:

i. Extraction: In this step, Image Thresholding is carried out to obtain a region which corresponds to the objects (or single object) under inspection.
ii. Refinement: The region which is extracted is degraded by different types of noises. Region transformation techniques are used for Region enhancement.
iii. Analysis: Here, the refined region is subjected to measurement and the final results are determined. If the region shows multiple objects, then it will be split into respective blobs where separate inspection on every blob is carried out.

Blob extraction for object tracking

There are two methods that permit the extraction of regions in an image:

i. Image Thresholding: It is a commonly used method that computes a region as a set of pixels that will arrive at few conditions depending on the particular operator (e.g. region of pixels showing higher brightness in comparison with either a given value or the neighborhood pixels which show average brightness). The output obtained shall be a single region, which represents multiple objects.
ii. Image segmentation: It computes a set of blobs relevant to those areas in the image that will meet few conditions. The output data obtained shall be an array of connected regions (blobs).

Blob extraction for human tracking

The purpose of BLOB extraction is to separate the connected pixels called BLOBs (objects) in the obtained binary image. Whether the two pixels are connected or not, their connectivity is defined by those pixels that are neighbors to those pixels and those are not. The 8-connectivity provides better accuracy compared to the 4-connectivity, but the 4-connectivity requires fewer computations and hence it is often applied, and hence processing of the image will be faster.

Extracting BLOBs will be the first step in detecting the target object. The next step is to carry out the classification into different BLOBs. In the next case, the classification is all about figuring out if the BLOBs are human or non-human. The classifier works in two steps. In the first step, number of characteristics and features are used to represent every BLOB. In the second step, certain matching methods are implemented so as to compare the characteristics of every BLOB to the characteristics of the target object. Feature extraction is nothing but converting every BLOB into some representative numbers. It is done so that relevant information is retained while ignoring the rest. Before calculating any of the features, every BLOB has to be excluded which will be connected to the image border. The main reason to do so is that there will be no information available regarding the part of object which is external to the image.

Kalman Filtering

Kalman filter is a region based method to find the regions of the object in forthcoming frame. Center of object is determined first and then Kalman filtering is made use for predicting position of the same in the next frame. Kalman filtering is preferred to compute the state of a linear system. Kalman filtering consists of two steps, prediction and correction. It provides optimal estimate to generate the position for the motion model for a moving object even if the video contains some amount of dynamic noise and noisy observations regarding the position at each time step. For Gaussian noises, Kalman filtering will provide an optimal solution. The filter reduces the mean square errors of the parameters such as position, velocity, etc. The Kalman filter is basically an online process, which means the new observations will be processed as and when they arrive. A discrete time dynamic linear system having an additive white noise that can model unexpected disturbances will be required to formulate a Kalman filter problem.

Object tracking using Kalman filtering

Multiple objects can be detected and tracked simultaneously by using Kalman filtering method. There are three basic steps in video analysis for tracking objects: **-** detection of motion of interested objects, tracking them in successive frames and performing the analysis of object tracks so as to identify their behaviors. The complexity of object tracking algorithms increases due to the noises in the images, illumination changes, and complex motion of the object and partial/full occlusions. Most of the tracking algorithms make an assumption that the motion of the moving object is smooth and there will not be sudden changes observed. Kalman filter estimates recursively the state of the target object. This method is vastly used in various domains like object tracking, economics and navigation systems.

Human tracking using Kalman filtering

Many methods are available for tracking. Some of them are Mean Shift algorithm, surf method, optical flow method etc. Here Kalman filter is used to track the humans as it is more advantageous than other tracking methods. It is seen, majority of the algorithms will depend on area of application are not much resistant to noise. An effective filtering algorithm can remove noise retaining the required information mainly Kalman Filter. The data processing algorithm developed is recursive as well as optimum, which has been implemented in this tracking challenge. The Kalman filter gives excellent results for practical applications because it is the only filter which reduces the variance of the estimation error to the best. It is a recursive predictive filter and it need not store its past measurements so as to deliver optimal processing of the data. Ever since this filter has been introduced, it has given opportunities for research, specifically in the area of autonomous navigation.

III. PROPOSED METHODOLOGY

Multiple human tracking is done using blob extraction algorithm. Blob analysis technique involves thresholding and morphological operations. The flowchart of the algorithm is shown in the figure 3.

A number of videos containing multiple human beings with/without the movement of objects are considered for validating the performance of the Matlab code. Measures are taken such that the algorithm detects and tracts only human beings and not any other moving objects. The given video file is initialized by reading the video file using **videoreader** command. Two video player windows are created, one of the video players is used to display the actual video while the other one plays the foreground masks. An empty array is created for the purpose of tracking human beings. Array is updated as and when the human beings are detected. System objects are also created which are used for blob detection. Morphological operations are carried out on these objects which involve noise removal and filling holes. In order to carry out filling the holes, first the connected components must be identified. These connected components are the blobs which are obtained by using adjacency of neighboring pixels which show similar features. The current location of tracking stage is predicted by using the features that describe the nature of human beings. These characteristics are obtained by area, centroid of blobs.

For every blob which should be tracked, a boundary box is defined. The position of the boundary box is updated and shifted based on the predicted location. Centroid calculation plays a key role in updating the boundary box location. After done with predicting the position of boundary box, a new detection is carried out to determine and update the estimate of human location in tracking. If the predicted boundary box is different with respect to area and position compared to the boundary box obtained by new detection, then the predicted one is replaced by the newly detected boundary box.

The previous track is deleted since it is replaced by the new values of human detection. After deleting the lost track, a new track is created using Kalman filtering method. The new track is loaded with all the information of the ongoing tracking and is updated to the arrays of track which were created in the beginning stages of the algorithm. To conclude and validate the tracks obtained, the tracks which are visible for more than minimum number of frames are only considered. Here the minimum number of frames so that the tracks should be identified positive is considered as 8. If tracking does not occur for at least 8 frames then tracking is ignored since those short lived tracks may be due to noise.

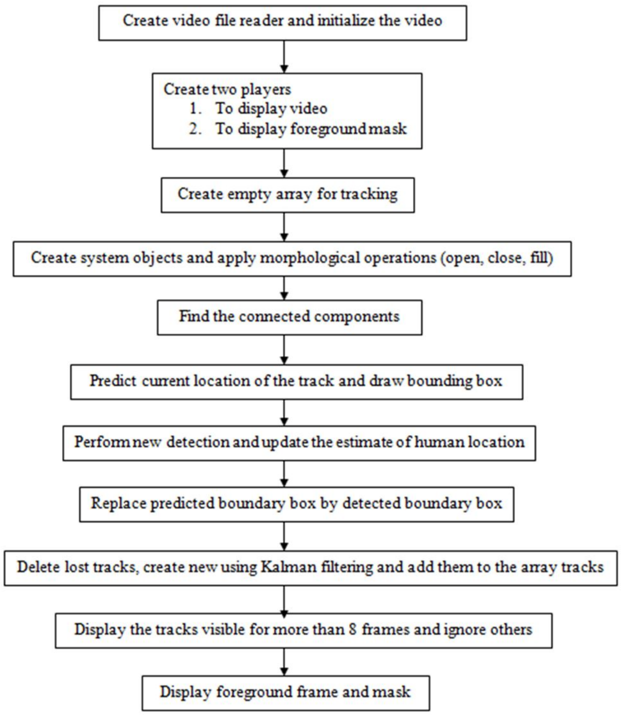

Figure 3: Flowchart of the proposed tracking algorithm

IV. RESULTS

In Video Processing, different parameters of videos play a huge role. To account that, work is carried out on different types of videos. Good quality videos are processed easily compared to compressed version of them. It's a known fact that as the number of persons in a video increases the tracking algorithm becomes less accurate. Another challenge is to successfully detect and track even when sudden movements are observed in the videos. Videos with the presence of other moving objects are considered to prove that the algorithm is capable of detecting and tracking only humans among other moving objects. The different characteristics and features of the videos considered for the work is shown in Table 1.

TABLE I. CHARACTERISTICS OF THE VIDEOS CONSIDERED FOR ALGORITHM IMPLEMETATION

Sl. No.	Size and 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
6.	9.27MB, 36secs	6	Moderate	No	Yes
7.	3.21MB, 25secs	5	Moderate	No	No
8.	5.45MB, 40secs	4	Moderate	No	No

The first video under consideration is a very good quality video of duration 14 seconds and memory size 21.7MB. The camera remains stationary and hence the area covered by the camera remains constant. The video consists of 5 persons with moderate walking speed. The tracking result and histogram representation of a frame of video is as shown in figure 4.

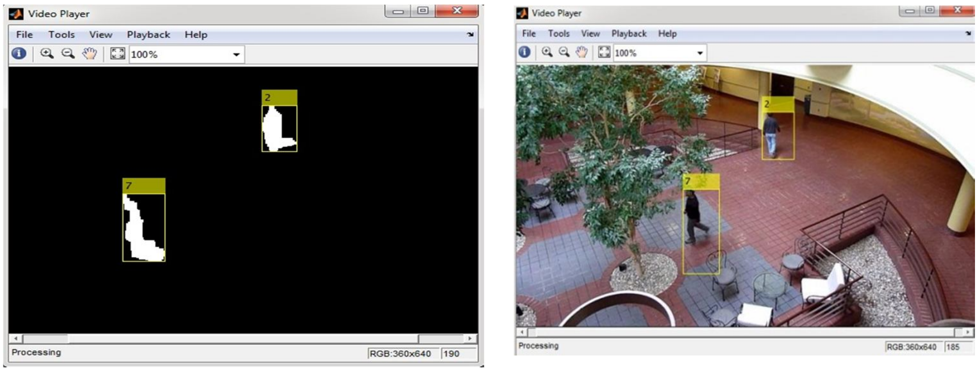

Figure 4: Tracking result and histogram representation of a frame of video 1

In test video 2, a tennis video is considered which consists of 4 tennis players who exhibit sudden movements unlike the test video 1. The video quality is moderate which is of the duration 16 seconds and is of 2.49Mb size. The tracking result and histogram representation of a frame of video 2 is as shown in figure 5.

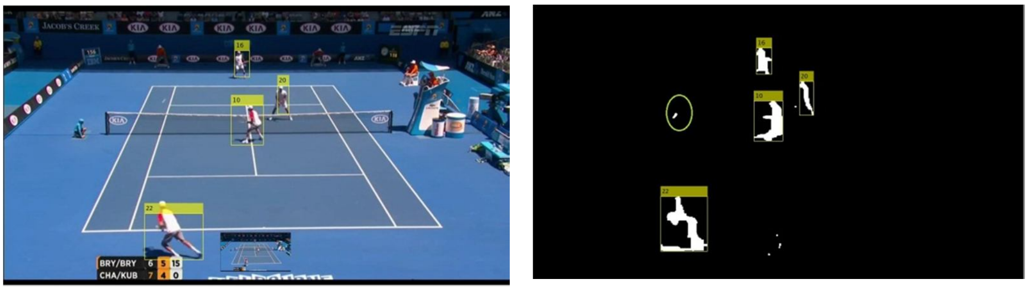

Figure 5: Tracking result and histogram representation of a frame of video 2

In test video 3, a moderate quality video is considered, which is of the duration 20 seconds. Tracking results and histogram representation of a frame of video 3 of this video is shown in the figure 6.

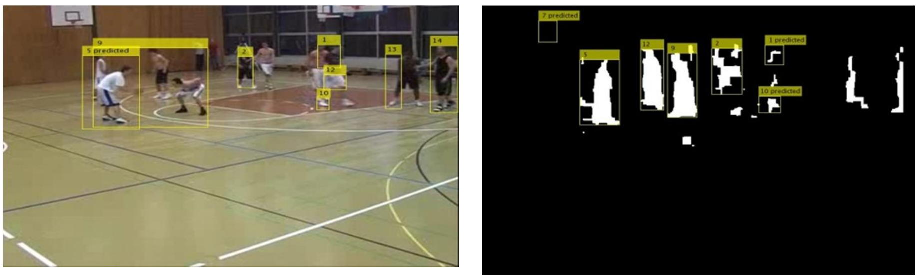

Figure 6: Tracking result and histogram representation of a frame of video 3

V. Conclusion And Future Work

The proposed work gives a simple and stable solution for Multiple Human Tracking. Although there are numerous Rworks done on the subject, the proposed work stands unique with its ability of providing excellent results for various kinds of videos considered. The combination of blob extraction with Kalman filtering gives effective results with less number of false detections. The advantages of the proposed work are simple algorithm, easy to implement, less complicated, can detect human beings in crowd and track them, sudden movements of humans do not affect system performance, only human movements are tracked ignoring the movements of any other objects in the video.

Even though the proposed work gives good results with many advantages, many improvements can be done. Future works can aim to identify false tracks of shadows and reflections and ignore them and thus reduce the rate of false detections. There is scope to work on developing an algorithm which can provide successful tracking of humans in moving camera videos.

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 Macomb, 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.

[11] Wu, Zhang, Betke “online motion agreement tracking”, 2013.

[12] Sherin Cherian, C.Senthil Singh, “Real time implementation of object tracking through webcam”.

[13] Chih-Lyang Hwang and Kuo-Ching Chang, “Adaptive Visual Tracking via Learning Detector of Specific Landmarks”, 2013.

[14] Jialue Fan, Wei Xu, Ying Wu, and Yihong Gong, “Human Tracking Using Convolutional Neural Networks” 2010.

[15] Vasilis Papadourakis, Antonis Argyros, “Multiple objects tracking in the presence of long-term occlusions” 2010.

[16] Jerome Berclaz, Francois Fleuret, Engin Turetken, and Pascal Fua “Multiple Object Tracking using K-Shortest Paths Optimization” 2009.