Real Time Object Tracking System with Automatic Pan Tilt Zoom Features for Detecting Various Objects

Sunig Kale, Ketki Patil, Poonam Satghare, Deepak Dharrao

Abstract: PTZ camera has been widely applicable to various surveillance systems. Focusing on controlling the PTZ camera automatically, an entire system capable of object detecting and tracking is proposed in this paper. The PTZ camera in the System is supposed to locate and identify an object which may considered as a threat. The basic idea is that the camera must be capable to detect and keep track of a target objects. Even though the place is crowded the system should not get affected to track the target object

Keywords: PTZ control, Image Acquisition, Object Detection, Object Classification, Background Subtraction, Noise Removal, MATLAB and OpenCV, Object Tracking.

I. INTRODUCTION

Object Tracking is one of the most significant task of smart video surveillance system. The importance of Object Tracking comes from the fact that it has a wide range of real time applications, including surveillance and monitoring of human activities in residential areas, smart rooms, traffic flow monitoring, hand gesture recognition, etc. The main problem in detection process is the influence of external environment. The algorithms completely differ for the day and night vision of the camera. Algorithms like viola-jones [1] used for face detection explicitly, cannot be used for object detection apart from the human face. The algorithms to detect and track the object varies from the shape, the color, the structure, appearance and state of the object. Object Tracking is one of the most significant task of smart video surveillance system. The importance of Object Tracking comes from the fact that it has a wide range of real time applications, including surveillance and monitoring of human activities in residential areas, smart rooms, traffic flow monitoring, hand gesture recognition, etc. The main problem in detection process is the influence of external environment. The algorithms completely differ for the day and night vision of the camera. Algorithms like viola-jones [1] used for face detection explicitly, cannot be used for object detection apart from the human face. The algorithms to detect and track the object varies from the shape, the color, the structure, appearance and state of the object.

Revised Version Manuscript Received on January 15, 2018.

Sunig Kale, Department of Computer Engineering, Indira College of Engineering and Management, Parandwadi (Maharashtra)-410506, India. E-mail: sunigkale@gmail.com

Ketki Patil, Department of Computer Engineering, Indira College of Engineering and Management, Parandwadi (Maharashtra)-410506, India. E-mail: patilketki009@gmail.com

Poonam Satghare, Department of Computer Engineering, Indira College of Engineering and Management, Parandwadi (Maharashtra)-410506, India. E-mail: satgharepoonam@gmail.com

Prof. Deepak Dharrao, Department of Computer Engineering, Indira College of Engineering and Management, Parandwadi (Maharashtra)-410506, India. E-mail: deepakdharrao@gmail.com

II. DESIGN OF PROPOSED SYSTEM

This paper aimed to design an Real-time object tracking system with automatic PTZ features for detecting various objects. The objects can vary from different data-sets along with the human face. Our primary concern is a tracking of a target object with ease rather than detection of the target object. Once the object is detected and successfully classified then the tracking of the target is done.

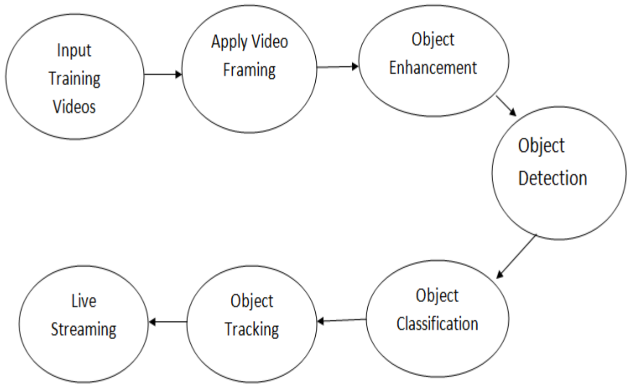

Figure 1. The Diagram of Tracking System Designed

A. Image Acquisition

The first stage of any vision system is the image acquisition stage. After the image has been obtained, various methods of processing can be applied to the image to perform the many different vision tasks required today. However, if the image has not been acquired satisfactorily then the intended tasks may not be achievable, even with the aid of some form of image enhancement. PTZ cameras are widely famous to track and acquire high resolution object and face image at a distance. The image resolutions vary according to camera used. but typically, 720×486 is considered to be enough for PTZ views. All image acquisition and processing modules are based on OpenCV and MALAB respectively.

B. Object Detection and Extraction

It is the process of finding the area in each frame where pixels of object are placed. Method includes 2 techniques. Which are Background Subtraction and Noise removal. Background subtraction is common yet efficient method to extract the object plus it can be used to track the object too. Background image is subtracted from the current frame, If the pixel difference is greater than the predefined Threshold value, then it determines the moving object.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Noise removal focuses on improving the frame quality. Since pixels with better strength are easier to be processed upon. Target Detection based on GMM model is the technique used in this paper. Considering the moderate cost and requirement and relative good foreground detection result in both grayscale and color video sequence, so Gaussian Mixed Model (GMM) is implemented to target detection in this system.

$$p(x|\lambda) = \Sigma_{i=1}^{M} wi\ g\ (x|\mu_i\ , \Sigma_i),$$

where x is a D-dimensional continuous-valued data vector (i.e. measurement or features), w_i, where i = 1, . . ., M, are the mixture weights, and g $(x|\mu_i , \Sigma_i)$, i = 1, . . . , M, are the component Gaussian densities. Each component density is a D-variate Gaussian function of the form.The complete Gaussian mixture model is parameterized by the mean vectors, covariance matrices and mixture weights from all component densities. These parameters are collectively represented by the notation,

$$\lambda = \{w_i , \mu_i , \Sigma i\} \text{ where } i = 1, . . ., M.$$

III. PROPOSED FILTER FOR IMPULSE NOISE REMOVAL

A novel morphological mean (MM) filter is proposed for high-density impulse noise removal. The flowchart of the proposed MM filter consists of two important modules: a noise-free pixel counter (NPC) module and a morphological pixel dilation (MPD) module. In the proposed NPC module, all pixels of the input image are examined in order to collect both the position and number of the noise-free pixels. Next, the dilatation operation of these pixels is performed to fill into the neighbor noise pixels in the proposed MPD module for best recovery of the image corrupted by a high density of noise.

IV. ALGORITHMS USED

A. Noise Free Pixel Counter Module

The proposed NPC module aims to detect noise-free pixels and as a result it restores the neighbor noise pixels for images with high densities of noise. The position of the noise-free pixels in the input corrupted image are detected by using Eq. (a).

Steps:

1: Input: Noise Image I.
2: Output: Number of noise free pixels Ni, Binary noise mask Bi.
3: Initialize Bi ← 0
4: Ni ← Number of pixels P
5: i ← 1
6: **for** each pixel x **do**
7: **if** I(x) is noise pixels **then**
8: Bi ← 1
9: **else**
10: Bi ← 0
11: end if
12: Ni ← Ni – Bi(x)
13: end for

B. Morphological Pixel Dilation Module

After the noise-free pixels of the corrupted image are detected in the NPC module, the detected pixels are then used to replace the eight-connected noise pixels within the scanned window. By doing so, the image corrupted by high densities of noise can be effectively restored.

Steps:

1: Input: Noise image I, The ith binary noise mask Bi
2: Output: The ith reconstructed image Oi
3: Initialize: NΩ← Number of pixels in window of radius s
4: Weight map Wi ← 0
5: **for** each pixel x **do**
6: **if** Bi (x) = 0 & NΩ P k=1 Bi (k) ≥ 1 **then**
7: **for** each pixel e in the window of radius s **do**
8: Oi(e) ← I(x)
9: Wi(e) ← Wi(e) + 1
10: **end for**
11: **else**
12: Oi(x) ← I(x)
13: **end if**
14: **end for**
15: **for** each pixel x **do**
16: **if** Wi(x) > 0 **then**
17: Oi (x) ← Oi (x)/Wi (x)
18: **else**
19: Oi (x) ← Oi (x)
20: **end if**
21: **end for**

V. BACKGROUND SUBTRACTION ALGORITHM

A. The Universal Multimode Background Subtraction:

The system comprises of multiple innovative mechanisms in background modeling, model update, pixel classification and the use of multiple color spaces. The system first creates multiple background models of the scene followed by an initial foreground/background probability estimation for each pixel. Next, the image pixels are merged together to form mega-pixels, which are used to spatially denoise the initial probability estimates to generate binary masks for both RGB and YCbCr color spaces. The masks generated after processing these input images are then combined to separate foreground pixels from the background.

Steps:

- BG Model Selection:

To select an appropriate BG (Background) Model for the incoming frame is the first step in this process.

- Binary Mask (BM) Generation:

In this step, the input image and the selected BG model are first used to estimate an initial probability estimate for each pixel. The input image is simultaneously passed to the MP module, which segments the image in arbitrary number of MPs. Average probability estimates are calculated for each MP using pixel-level probability estimates and then threshold to generate Binary Mask(BM) for each color channel.

- Binary Masks Aggregation/Fusion:

In the binary masks aggregation or fusion, the binary masks which are generated are aggregated in one frame and the background is substracted.

- Binary Masks Purging:

The FGD mask is then applied to each of the BMs obtained in step 3[2]. This removes all of the falsely detected foreground regions and increases our confidence in classifying FG and BG pixels in the final step.

- Foreground Mask:

In the final step of process, the FM (Foreground Mask) is obtained by the process of purging the binary masks.

VI. METHODS/TECHNIQUES USED

A. Algorithms used:

1. Noise Removal:
   - Noise free pixel module
   - Morphological pixel dilation module
2. Background Subtraction:
   - Universal Multimode Background Subtraction.
3. Preprocessing:
   - NAFSM Filter
   - MDBUTM Filter
   - AWM Filter
4. Object Recognition:

Object Recognition is a technology in the field of computer vision for finding and identifying objects in an image or video sequence. Humans recognize a multitude of objects in images with little effort, despite the fact that the image of the objects may vary somewhat in different viewpoints, in many different sizes and scales or even when they are translated or rotated. Objects can even be recognized when they are partially obstructed from view. This task is still a challenge for computer vision systems. Many approaches to the task have been implemented over multiple decades.

- Statistical approaches, which consider patterns as points in d-dimensional space.
- Structural approaches, which use arcs and segments to represent shapes.
- Neural network approaches.
- Hybrid approaches, which combine statistical and structural

In this paper object recognition is mainly based upon matching the attributes stored by using the SIFT algorithm. SIFT stands for Scale Invariant Feature Transform. SIFT can be incorporated with OpenCV. SIFT can efficiently identify the object in between the image, using feature matching, rotation, translation, Zoom. Moving objects are characterized by their color-histograms.

5. Object Tracking:

The aim of an object tracker is to generate the trajectory of an object over time by locating its position in every frame of the video. But tracking has two definition one is in literally it is locating a moving object or multiple object over a period of time using a camera [1]. Tracking is the problem of estimating the trajectory or path of an object in the image plane as it moves around a scene. A comprehensive survey on intra-camera tracking algorithms is found in [3] and it can be classified into two categories in terms of tracking strategy: deterministic and probabilistic tracking.

The tasks of detecting the object and establishing a correspondence between the object instances across frames can either be performed separately or jointly. Detection and recognition of continuous activities from video is a core problem to address for enabling intelligent systems that can extract and manage content fully automatically. Recent years have seen a concentration of works revolving around the problem of recognizing single-person actions, as well as group activities [4]. In the first case, object detection is done and after object detection tracking algorithms are applied [1]. In the latter case, the object region and correspondence are jointly estimated by iteratively updating object location and region information obtained from previous frames. There are different methods of Tracking.

A. Point is tracking:

Tracking can be formulated as the correspondence of detecting objects represented by points across frames. There are two types of Point tracking, first is the Deterministic approach and the second one is Statistical approach. Objects which are in consecutive frames are representation is done by points, and then the association of the points is basically based on the previous object state which can include object position and motion.

B. Kernel tracking:

Performed by computing the motion of the object, represented by a primitive object region, from one frame to the next. Object motion is in the form of parametric motion or the dense flow field computed in subsequent frames. Based on the appearance representation used there are two types of Kernel tracking, first is the Template and second one is Density-based Appearance Model and Multi-view appearance model.

C. Silhouette Tracking:

It gives us an accurate shape description of the target objects. The goal of silhouette tracker is to find the object region in each frame by means of an object model generated using the previous frames. There are two types of Silhouette Tracking, first is the Shape matching and second one is the Contour tracking.

D. Feature Selection for Tracking:

It plays a vital role to select a proper feature in tracking. Therefore, the feature selection is closely related to the object representation. For example, color is used as a feature for histogram appearance representations, the contour-based representation, the object edges are usually used as features [4]. Basically, many tracking algorithms use a combination of these features. There are various details of common visual features and they are as follows:

• Color-Color of an object is influenced by two factors. These are mainly based on Spectral power distribution of the illuminant and Surface reflectance properties of the object. Different color models are RGB, L*u*v and L*a*b used to represent color.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

- Edges-Edge detection is used to identify strong changes in image intensities generated by object boundary. Edges which are less sensitive to illumination changes are compared with color features. Most popular edge detection approach is Canny Edge detector.
- Optical Flow-It is defined as a dense field of displacement vector which defines the translation of each pixel in a region. It is calculated using the brightness constraint, through them they assume brightness constancy of corresponding pixels in consecutive frames. Optical Flow is basically used as a feature in which motion is based on segmentation and tracking application.
- Texture-Texture is used in the measure of the intensity variation of a surface which quantifies properties such as smoothness and regularity. It requires a processing step to generate the descriptors. There are different kind of texture descriptors: They are as follows
  - The Gray-Level Co-occurrence Matrices,
  - The loss texture measures,
  - The wavelets and
  - The steerable pyramids.

E. Background:

Object tracking is basically a very challenging task where in the presence of variability Illumination condition, the background motion, the complex object shape, the partial and the full object occlusions we have to track the object. In this thesis, we are doing some modification so as to overcome the problem of illumination variation and background clutter such as fake motion due to the leaves of the trees, the water which is flowing, or the flag waving in the wind [4]. Sometimes object tracking involves tracking of a single interested object and that is done using normalized correlation coefficient and updating the template. Object tracking is an important job within the field of computer vision. Object detection is a technique which involves locating of the objects in frames of a video sequence. Tracking is also a process of locating the moving objects or the multiple objects over a period of time with the help of a camera. Computationally, tracking is the problem in which estimation of the trajectory or path of an object in the image plane as it is moving around a scene. Huge knowledge about the number and the size of objects, or the object appearance and shape, can also be used to simplify the problem [4]. . The Kalman filter is a recursive predictive filter that is based on the use of state space techniques and recursive algorithms. Kalman filtering is composed of two steps [4]. There are two steps involved in the Kalman filter

- The prediction step and
- The correction steps.

The first step state that it is predicted with the dynamic model. The prediction step is the step which uses the state model to predict the new state of the variables. In this sense it is an optimal estimator. Kalman filter is used in the vision community for tracking.

VII. ASIFT ALGORITHM

A fully affine invariant image matching algorithm needs to cover the 6 affine parameters. The SIFT method covers 4 parameters by normalizing rotations and translations, and simulating all zooms out of the query and of the search images [5]. ASIFT complements SIFT by simulating the two parameters that model the camera optical axis direction, and then applies the SIFT method to compare the simulated images, so that all the 6 parameters are covered. In other words, ASIFT consist of three parameters: the scale, the camera longitude angle and the latitude angle (which is equivalent to the tilt) and normalizes the other three (translation and rotation) [5]. ASIFT can thus be mathematically shown to be fully affine invariant. Against any prognosis, simulating the whole affine space is not prohibitive at all with the proposed affine space sampling.

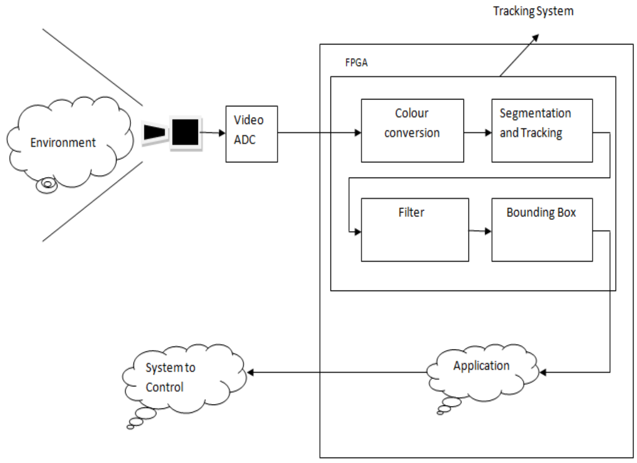

Figure: Block Diagram

VIII. DATASET

The datasets employed in this system contain synchronized and static views, captured at a frame rate of 30 fps from uncalibrated camera installed at a corner/wall in an auditorium. Each image frame has a size of 320×230 pixels. A total of 5000 frames (≈ 2-hour in length) from camera view can be used for activity based scene decomposition. Detailed description on dataset is given as follows:

A camera placed at the most effective position can cover most of the floor area of the room. The PTZ features extends the vision of the camera and allows the higher success rate of capturing the unidentified events occurred within the range. The current proposed system mainly focuses on an Auditorium where large number of crowd may gather within a **very short span of time.** The camera must keep a watch at any instance of time. The goal of an objective is to locate out the threatful objects **at a run time** (The type of object is covered under object-set). As soon as objective 1 gets complete. 2nd Objective (Object Tracking) should be done within the relay.

IX. CONCLUSION

All the review which has been done in this paper signifies the case of using the surveillance system that focuses on detecting and identifying the threat full objects out of the given subsets. The object or the person carrying the object should be traced down by the system. Another secondary objectives such as identifying the threat level or prioritizing the object or a person should be carried away.

Detection and recognition of continuous activities from video is a core problem to address for enabling intelligent systems that can extract and manage content fully automatically. This system assures agile and easy to use surveillance model with a lot of accuracy and more scope to integrate new and innovative techniques in future.

REFERENCES

1. PTZ Design and Implementation of a Detection and Tracking of Moving Object Based on PTZ Camera.
2. Multiple Object Detection using OpenCV on an Embedded Platform.
3. A Survey on Object Detection and Tracking Algorithms.
4. ASIFT: An Algorithm for Fully Affine Invariant Comparison.
5. New Object Detection, Tracking and Recognition Approaches for Video Surveillance over Camera Network,2015
6. Online Human Interaction Detection and Recognition with Multiple Cameras,2016
7. Visual Attention Based Background Modelling for Detecting Infrequently Moving Objects,2016
8. Salient Point Tracking for Key Frames Selection in Outdoor Image Sequences,2016
9. Occlusion aware Real time Object Tracking by Integrated Circulant Structure Kernels Classifier,2016
10. Multiple Sensor Fusion and Classification for Moving Object Detection and Tracking,2015
11. Face Tracking and Recognition Data Distance: A Coaxial and Concentric PTZ Camera System,2013
12. Neural Background Subtraction for Pan-Tilt Zoom Cameras,2014
13. Hierarchical Ensemble of Background Models for PTZ-Based Video Surveillance,2014

Sunig Kale: B.E Department of Computer Engineering, Indira College of Engineering and Management, sunigkale@gmail.com

Ketki Patil: B.E Department of Computer Engineering, Indira College of Engineering and Management, patilketki009@gmail.com

Poonam Satghare: B.E Department of Computer Engineering, Indira College of Engineering and Management, satgharepoonam@gmail.com

Prof. Deepak Dharrao: Department of Computer Engineering, Indira College of Engineering and Management, deepakdharrao@gmail.com

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.