International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 6 | ISSN : 2456-3307

Integral and Indexing based Feature Extraction Method for Human Detection and Tracking

Jaykumar Limbasiya[1], Hitul Marvaniya[2]
[1]PG Student, Information Technology Department, V.V.P. Engineering College, Rajkot, Gujarat, India
[2]Assistant Professor, Information Technology Department, V.V.P. Engineering College, Rajkot Gujarat, India

ABSTRACT

The detection and tracking of the human being in Video surveillance is a vibrant research topic in computer vision. It replaces old traditional method of monitoring camera feed by the human being by automatically detect and tracking along with the understanding of human behavior. Surveillance systems must be self-directed to improve the performance and eliminate such operator errors. Ideally, an automated surveillance system should only require the objectives of an application, in which there are real-time understanding of activities. Human detection includes detecting humans in each frame of video. Either each tracking method requires an object detection mechanism in every frame or when the object first appears in the video. Human tracking is the method of detecting human over time on the camera feed. The fast and efficient camera, high-quality videos, low priced camera and increasing need for the automatic system for detection and tracking will create interest in human tracking algorithms.

Keywords: Human Tracking, Video Surveillance, Detection, Object Detection and Tracking, Feature Extraction, keypoints detection, Point tracking

I. INTRODUCTION

Object detection is techniques used in computer vision applications such that navigation of robot automatically, navigation of vehicle and other surveillance application which mainly used for security. Surveillance systems must be self-directed to improve the performance and eliminate such operator errors. Ideally, an automated surveillance system should only require the objectives of an application, in which there are real-time understanding of activities. Then, the challenge is to provide robust and real-time performing surveillance systems at a reasonable price. As the cost of hardware decreases for sensing and computing, and the processor speed increase, surveillance systems have become commercially available, and they are now applied to a number of different applications, such as traffic monitoring, airport and bank security, etc.

However, occlusions, shadows, weather conditions, etc. likes shortcoming still affect machine vision algorithms (especially for a single camera). After studying the literature, it is found that detecting the object from the video sequence and track them throughout the video is a challenging task. Object tracking can be a time-consuming process due to the amount of data that is contained in the video. From the literature survey, it is found that there are many background subtraction algorithm exist, which work efficiently in both indoor and outdoor surveillance system [1].

It will be better if the shadow will be removed at the time of the foreground object detection by designing an efficient algorithm, which can properly classify the foreground object and background by removing false foreground pixel from detection [2]. Then there will no extra computation needed for shadow detection and removal.

After studying the above literature, it can be finalized that the detection of the object is a very crucial task for the real world applications. The time required for the object detection and tracking might be too high. So, it is very important to reduce or to control the time required for object detection and tracking. From the Literature survey it might be concluded that there are

many object detection techniques is available and object tracking methods too in the market.

The power consumption can also be the main point of interest when the particular method is going to be implemented in the low power device.

Background subtraction is the very basic method for object tracking through the video sequence [3]. Lighting in the scene might affect the object detection and tracking process, therefore the algorithm should be well equipped that it can handle the lighting condition. If the algorithm is very well developed then there will be no extra computation time for this process.

Video surveillance system is nowadays very important in many aspects such that security, privacy. The development in this field is very rapid and vibrant. In today's era, the aim of the developer of this system is to develop a very intelligent system, which can interpret the input and evaluate the output from it automatically. The output is only not necessary but to act accordingly is important for this system too.

The Objective of this research paper is to investigate and improve feature extraction method for human detection and tracking from surveillance video. Which originally consist of two components feature extraction and human detection and tracking.

Automatic tracking of objects is a very important task, which leads to many interesting applications. The accuracy and efficiency in tracking methods capability are essential for building applications contains higher-level vision based intelligence application. Tracking is not a trivial task given the non-deterministic nature of the subjects, their motion, and the image capture process itself.

Automated surveillance is very emerging field due to their applications in different area like security, traffic control, and human activity detection [4]. These fields are very critical for the application of automated surveillance. This system can also be important for many lives because it can direct threaten the human lives. Surveillance of vehicular traffic and human activities offers a context for the extraction of significant information such as scene motion and traffic statistics, object classification, human identification, anomaly detection, as well as the analysis of interactions between vehicles, between humans or between vehicles and humans.

The area is new of its kind needs a lot of attention and experiments for real-life implementation, so it can be considered by the computer scientists as the area of research for the betterment of humanity, especially concerning better lifestyle.

Object occlusion, complex scene unintentional motion in the scene, noise different illumination condition makes the object tracking process very complex and difficult to implement. In [2], the modification is done to overcome the problem of illumination variation and background clutter such as fake motion due to the leaves of the trees, water owing, or flag waving in the wind. There is no need for every object tracking in the available scene; there might be a single object of interest, which should be detected in the scene using normalized correlation coefficient. Another process might include the update to the template.

In [3] very basic framework to detect and track moving object is very straightforward. When the video is ready for the video analysis step, the feature extraction method extracts the candidate features from the frame extracted from the sequence of the frame from the video.

The Detection of Human is the very crucial task and the severity of application is very high compared to the other applications of the object or human detection. The time factor required to detect and track that human being is very important in this application. There are many types of object detection and tracking method are available and the main focus of the researcher is on the time required for the computation of detection process and tracking process as well. Therefore, the main problem of any human detection and tracking technique is the Time factor.

II. METHODS AND MATERIAL

In this research paper, the proposed method comes with few concepts like Integral Image, Indexing etc. The concept of Integral Image is introduced in 1984, and properly introduced to the world of Computer Vision in 2001 by Viola and Jones [5].

When creating an Integral Image, we need to create a Summed-Area Table. In this table, if we go to any point (x,y) then at this table entry we will come across a value. This value itself is quite interesting, as it is the sum of all the pixel values above, to the left and of course including the original pixel value of (x,y) itself. The peak value in the scale space is detected using indexing. An additional orientation vector is also assigned to each keypoint, which increase the accuracy of oriented objects. The actual inputs required and the steps, which have followed, are described in the following algorithm:

Algorithm 1: I^2 SIFT

1. procedure ***I^2 SIFT****(Image I, amount of blurring, k factor to obtain DoG)*
2. Generate Scale Space of an Image using concept of Integral Image.
3. Detect Peak in scale space by comparing pixel by its index.
4. Localize Interest Point.
5. Remove Outliers and truncate the floating-point value of descriptor vector.
6. Assign Orientation to each key interest point with the new additional orientation component
7. Compute relative Orientation and Magnitude at Key points (Local Image Descriptors)

The video is the collection of frames. Initially for the testing purpose, a frame is extracted from the surveillance video. The basic SIFT algorithm is implemented and tested on that frame only to test the different parameter such as sigma value and others.

The following step is being carried out as parts of feature extraction using SIFT:

A. Generate Scale Space of an Image
B. Keypoint Localization
C. Accurate Keypoint Localization
D. Orientation Assignment
E. Keypoint descriptor

A. Generate Scale Space of an Image

This is the first stage to get the feature from an initially loaded image. The DoG (Difference of Gaussian) is generated from the image. As described before the keypoints is detected with the help of cascading filtering approach, which ultimately generates the Scale-space of the image from the input image using DoG function [6]. The first stage of keypoint detection is to identify locations and scales that can be repeatable assigned under differing views of the same object.

B. Keypoint Localization

Each image of DoG is being searched for the extreme points, which we can call interest point, sometimes it also called as Candidate keypoints. These points are further filtered to reduce the number of points and gating better accuracy.

C. Accurate Keypoint Localization

Initially the keypoint is calculated by comparing the candidate points with their neighbours following to that process detail investigation is carried out to the data location, scale and the ratio of principal curvature [7]. The information gathered from the investigation is helping the algorithm to discard the points with lower contrast value and the points on the edges, which sometimes poorly localized. Thus, the keypoints from the set of candidate point is reduced and leads to better accuracy for the localization of the keypoints.

D. Orientation Assignment

The orientation factor assigns to each keypoint is assigned a consistent orientation, which uses the local image properties [8]. Therefore keypoint descriptor is projected relatively the orientation factor which ultimately helps the keypoint descriptor to be invariance to the rotation of the image.

E. Keypoint descriptor

The keypoint descriptor is weighted by a Gaussian window, which initially calculates the magnitude and the other orientation for each point, which is sampled around the location of keypoints. The histogram is built based on the orientation factor assign to each sample, which comes from the accumulation of samples gathered [9]. The final histogram is of 4x4 sub regions. The height of the histogram bar is calculated from the summation of the magnitude of the arrow assign to the corresponding direction of the particular region.

III. RESULTS AND DISCUSSION

This section includes the details about the datasets used in the experiment for this research paper, how the task is implemented and the results for the table which contain the time-based comparison between the existing system and proposed system.

A. Dataset Information

We are using the datasets provided by Active Vision Laboratory Department of Engineering Science, University of Oxford. The project Coarse Gaze Estimation in Visual Surveillance focuses on the problem of obtaining passive course gaze estimates in surveillance video [10]. The 'Town Centre' dataset was used to test tracking performance in both the CVPR 2011 and the BMVC 2009 papers.

B. Tasks for Human Detection and Tracking

To detect and track the human, the dataset of surveillance video is taken. Initially the first frame is extracted from the surveillance video. Before that feature from the available template is stored in, the dataset, which is used to match and detect the object from the frame, extracted from the surveillance video.
Initial frame extracted from the video is shown in Figure 1. A better result can be obtained by extracting the intermediate frame from a video and apply the feature-extraction algorithm on it to change the interesting point detected runtime. This method can be used to detect the multiple humans, which are new to the detection and tracking system.

Another approach of machine learning might be used to get better assumption about the naive user comes into the front of video surveillance because machine learning is the better option for building training set and applying that training set on actual dataset available for testing.

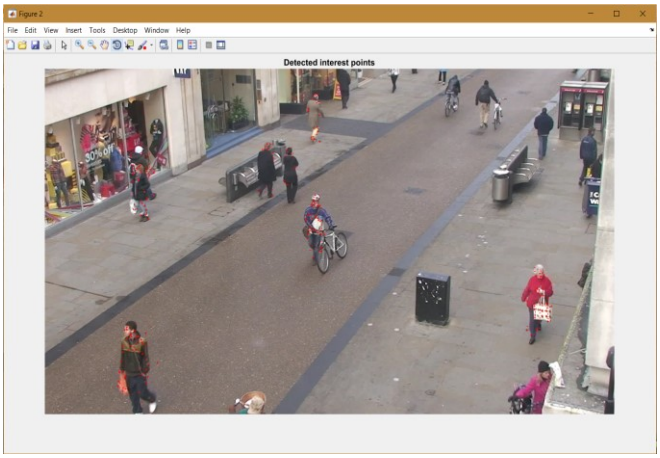

Figure 1. Detected in interest point in the frame

The detected points are shown in figure 1. Now, for video surveillance any of the tracking methods might be used. In this dissertation, the point-based tracking method is used to track the detected interest point throughout the video as shown in figure 2.

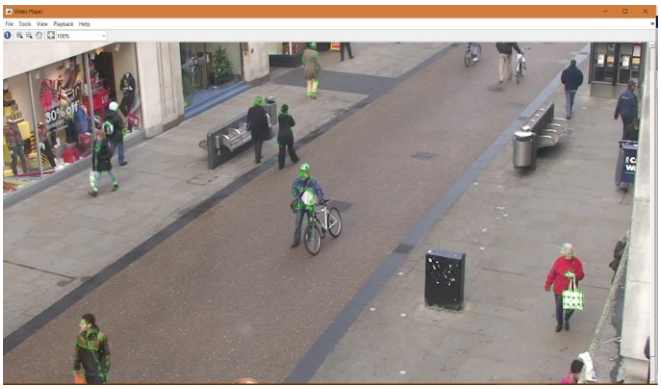

Figure 2. Interest points tracked throughout video

C. Comparing with Existing System

The Table I shows the comparisons of time required for different phase between the existing system, which is SIFT developed by David G. Lowe and I^2 SIFT, which is proposed in this research paper.

TABLE I
COMPARING WITH EXISTING SYSTEM

Different Phase	Traditional SIFT *(in seconds)*	I^2 SIFT *(in seconds)*
Time Required for Single Template Feature Extraction	0.125	0.146
Time Required for Feature Extraction from Frame	6.438	6.197
Time Required for Matching Feature	0.353	0.447
Total Time Required	**6.916**	**6.790**

IV.CONCLUSION

Ultimately, it is concluded that the modification reduces the time required for human detection from the frame, there might be other possibility is that some other techniques should be used alongside SIFT for human detection and Tracking. It is concluded that the basic feature extraction method SIFT takes more time to extract the feature from the image provided because the SIFT algorithm has higher computation; therefore it is a bit slower for video processing. The video is

consisting of different frames and these frames are moving ideally approx. 30 frames per second.

V. FUTURE WORK

There might be the possibility that the improvement in the traditional SIFT method along with parallel processing may reduce the computation time required. Other techniques like those that Background Subtraction might be used alongside improved SIFT for Human Detection and Tracking in the surveillance video.

VI.REFERENCES

[1] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking", in Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., vol. 2, pp. 246-252, IEEE, 1999.

[2] J. C. S. Jacques, C. R. Jung, and S. R. Musse, "Background subtraction and shadow detection in grayscale video sequences", in Computer Graphics and Image Processing, 2005. SIBGRAPI 2005. 18th Brazilian Symposium on, pp. 189-196, IEEE, 2005.

[3] D. T. Nguyen, W. Li, and P. O. Ogunbona, "Human detection from images and videos: a survey", Pattern Recognition, vol. 51, pp. 148-175, 2016.

[4] M. Brown and D. G. Lowe, "Invariant features from interest point groups.", in BMVC, vol. 4, 2002.

[5] Wikipedia, "Scale-invariant feature transform - Wikipedia, the free encyclopedia." http://en.wikipedia.org/wiki/Scale-invariant_feature_transform. Accessed Jan, 2017.

[6] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features", Computer vision-ECCV 2006, pp. 404-417, 2006.

[7] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors", IEEE transactions on pattern analysis and machine intelligence, vol. 27, no. 10, pp. 1615{1630, 2005.

[8] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features", Computer Vision{ECCV 2010, pp. 778{792, 2010.

[9] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An e_cient alternative to sift or surf", in Computer Vision (ICCV), 2011 IEEE International Conference on, pp. 2564{2571, IEEE, 2011.

[10] I. R. Ben Benfold, "Active vision laboratory." http://www.robots.ox.ac.uk/ActiveVision/Research/Projects/2009bbenfold_headpose/project.html#datasets. Accessed Nov, 2016.