ISSN 2320 - 2602
Volume 5 No.8, August 2016
International Journal of Advances in Computer Science and Technology
Available Online at http://www.warse.org/IJACST/static/pdf/file/ijacst04582016.pdf

Mean-Shift Tracking Algorithm for Salient Object Detection in videos

[1]T.Saikumar, [2]Mamatha Nalavapani,
[1]Associate Professor, Department of Electronics &Communication Engineering, CMR Technical Campus, Hyderabad.
[2]M.Tech Scholar, Department of Electronics & Communication Engineering, CMR Technical Campus, Hyderabad.
tara.sai437@gmail.com, mamathanalavapani@gmail.com

Abstract: Salient object detection is an interesting subject in the field of video tracking and its applications. The main purpose of salient object detection is to estimate the position of the object in images in a continuous manner and reliably against dynamic scenes. When object is moving then its detection is a challenging task in much vision area. Two tasks are performed to detect salient object, in videos. First detection and second following the object path. This can be beneficially achieved by using the mean shift object tracking algorithm. In this approach, a rectangular target window is defined in an initial frame for a target in a video, and after that, the salient object is separated from the back ground by processing the data within that window. The proposed method has good accuracy to track a moving object in successive frames under some difficulties such as appearance changes due to image noise, etc. We will employ this approach for increasing the capability of moving object tracking.

Keywords: video, salient object, kernel, mean-shift, histogram.

1. INTRODUCTION

Salient Object detection is a complex task in the area of computer vision. Object tracking is a basic requirement for visual analysis. Various problems such as noises, clutters and appearance changes, etc occurs while detecting salient object. All the above mentioned problems can be solved by several algorithms. Any video tracking is completed by three steps:

- Detect the object
- Tracks the movement of object
- Observe the behavior of object

In various object tracking methods and salient object detection method, the mean shift tracking method, is a simple and popular method. The Mean Shift Tracking is an accurate and fast object tracking for small sequence. Mean shift method is used in some situations such as segmentation, target representation and localization. In mean shift method a kernel function is used. This kernel function is a rectangle region. By this rectangle region we can detect and track the objects path. Target model and target candidate are two important factor of mean shift method.

In video tracking, when the objects are moving fast relative to the frame rate then it is somewhat difficult to associate target location in consecutive video frames [2]. For this, various approaches for salient object detection (object-tracking) have been proposed. And which approach is to be implemented is completely depending upon the context in which the tracking is performed. Based on the changes to the properties of the object being tracked, object-tracking is a type of technique to track an object and to perform an immediate action on some other object which has no relation to the tracked object.

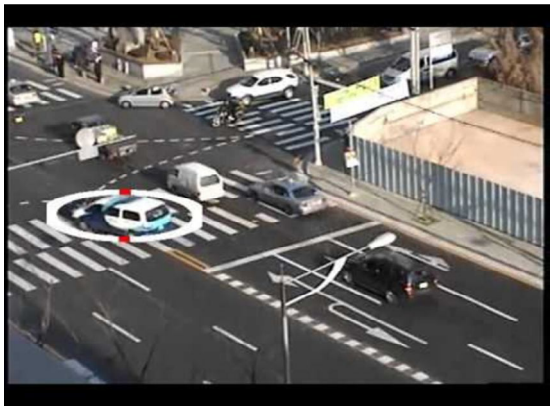

Figure 1: Tracking a moving car(salient object)

A typical visual tracker consists of two main components, which can be distinguished as:
A. Target Representation and Localization.
B. Filtering and Data Association.
Target Representation and Localization is a type of bottom up approach which has to deal with the variations in the appearance of the target. Filtering and Data Association is a type of top-down approach which has to cope with the dynamics of the tracked (salient) object, evaluation of different hypothesis and learning of scene priors [1].

2. SALIENT OBJECT DETECTION IN VIDEO SEQUENCE

Salient Object Detection in video is defined as the process of locating the position of a moving object or multiple objects over the time using a camera. But it is not similar to Camera tracking. It has a number of uses, which includes: video communication and compression, human computer

interaction, security and surveillance, traffic control, medical imaging and video editing.

Salient Object tracking may be a time taking technique because of the quantity of data contained in video.

The main purpose of salient object detection is to track target objects in successive video frames. The tracking can be specifically tough when the objects are in motion fast relative to the frame rate. Other condition that steps up the difficulty of the issue is when the tracked object frequently changes orientation and location over time. For these conditions salient object tracking systems usually need a motion model which explains how the image of the target might change for various possible motions of the object.

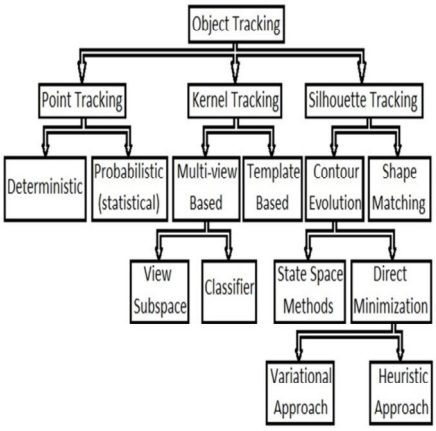

Figure 2: Classification of Salient Object Tracking

A. Video processing

Video processing is a specific type of signal processing that frequently uses video filters having the input and output signals as video streams or video files. Video processing methods are employed in DVDs, video scalars, video codec, television sets, video players, VCRs, and other devices [11]. This technique also needs a stream processing construction, in which video frames coming from an ongoing stream are taken into account for processing one or more at an instant. This type of method is found to be a bit difficult in systems that have current video or where the video data is so huge that cramming the overall set into the counter is incompetent[12].

B. Algorithms

In order to operate salient object detection, an algorithm examines consecutive video frames and produces the motion of targets between the frames as output. There are a wide variety of algorithms, each with advantages and disadvantages. There are two main components of a visual object detection system: target representation and localization, as well as filtering and data association. Target representation and localization provide many tools for recognizing the moving object. Locating and tracking the target object is completely based on the algorithm. For example, the use of blob tracking is helpful for analyzing human motion as a person's contour changes effectively. Generally, the estimated complexity for these algorithms is low. Some common target representation and localization algorithms are as follows:

1) Mean-Shift tracking: It is also known as Kernel-Based tracking. It is an iterative positioning process built on the augmentation of a parallel measure (Bhattacharyya coefficient) [6].

2) Contour tracking: It is also known as Condensation Algorithm and is mainly used to estimate the object boundary. Contour tracking process iteratively develops an initial contour occurred from the foregoing frame to its new location in the present frame. This algorithm of contour tracking directly provides the output of the contour by reducing the contour energy using gradient descent.

3. MEAN-SHIFT BASED TRACKING ALGORITHM

In 1975, Fakunaga and Hostetler proposed an algorithm known as Mean-shift (MS) object-Tracking Approach. It is basically iterative expectation maximization – clustering algorithm executed within local search regions [3]. In other words, it is a type of non-parametric clustering algorithm that does not need prior information of the number of clusters and also does not limit the shape of the clusters. That is, the Mean-shift algorithm is a nonparametric density gradient estimator.

The following steps are iterated in order to track the object by using the Mean-Shift algorithm:

A. Select a search window size and the initial position of the search window.

B. Estimate the mean position in the search window.

C. Center the search window at the mean position estimated in Step B.

D. Repeat Steps B and C until the mean position moves less than a preset threshold. That is, until convergence is achieved.

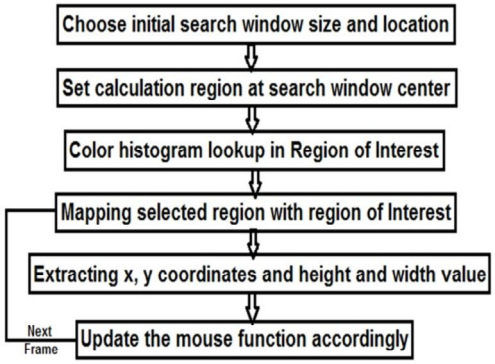

Figure 3: Procedure of Object Tracking

After the Mean-Shift Algorithm is executed on various videos it is concluded that when the target

moves so fast that the target area in the two neighboring frame will not overlap, tracking object often converges to a wrong object. Because of this issue, traditional Mean-Shift Algorithm gets failed to track fast moving object. But there are some solutions like combining Kalman filter or particle filter with Mean-shift Algorithm[4]. The center of convergence will become more accurate as at first it predicts the direction and speed of the object and then adjusting the search window center of the mean shift convergence. However, such methods are not suitable to use in real time tracking systems as these require high CPU costing to estimate the moving object location. The kernel-based tracking algorithm, when combined with prior task-specific information, can gain adequate results. This enhanced model could profitably detect and track a human subject in arbitrary motion and in a situation where there is a certain alter in radiance. It was constituted that a well built alteration in radiance induced the procedure to experience quite major deformation in the probability distribution image. Hence, the non-adaptive behavior of the mean shift algorithm may lead to a wrong tracking conclusion.

4. MEAN SHIFT BASED TRACKING APPROACH

Given n data points xi,i = 1, ..., n on a d-dimensional space Rd, the multivariate kernel density estimate obtained with kernel K(x) and window radius h is :

$$f(x)=\frac{1}{nh^{d}}\sum_{i=1}^{n}k\left(\frac{x-x^{i}}{h}\right) \quad (1)$$

For radially symmetric kernels, it suffices to define the profile of the kernel k(x) satisfying

$$k(x)=c_{k,d}k\left(\|x\|^{2}\right) \quad (2)$$

Where, $c_{k,d}$ is a normalization constant which assures K(x) integrates to 1. The modes of the density function are located at the zeros of the gradient function.

∇f(x) = 0. The gradient of the density estimator Eq.(1) is

$$\nabla f(x)=\frac{2c_{k,d}}{nh^{d+2}}\sum_{i=1}^{n}(x_i-x)\,g\left(\left\|\frac{x-x_i}{h}\right\|^{2}\right)$$

$$\nabla f(x)=\frac{2c_{k,d}}{nh^{d+2}}\left[\sum_{i=1}^{n}g\left(\left\|\frac{x-x_i}{h}\right\|\right)\right]\left[\frac{\sum_{i=1}^{n}(x_i)g\left(\left\|\frac{x-x_i}{h}\right\|^{2}\right)}{\sum_{i=1}^{n}g\left(\left\|\frac{x-x_i}{h}\right\|^{2}\right)}-x\right]$$

where g(s) = −k'(s). The first term is proportional to the density estimate at x computed with kernel $G(x)=c_{k,d}g\left(\|x\|^{2}\right)$ and the second term is the mean shift.

$$m_h(x)=\left[\frac{\sum_{i=1}^{n}\left(x_i g\left(\left\|\frac{x-x_i}{h}\right\|^{2}\right)\right)}{\sum_{i=1}^{n}g\left(\left\|\frac{x-x_i}{h}\right\|^{2}\right)}-x\right] \quad (4)$$

To track the target using the Mean Shift algorithm, it iterates the following steps:

1. Choose a search window size and the initial location of the search window.
2. Compute the mean location in the search window.
3. Center the search window at the mean location computed in Step 2.
4. Repeat Steps 2 and 3 until convergence (or until the mean location moves less than a preset threshold).

Color distributions derived from video image sequences change over time, so the mean shift algorithm has to be modified to adapt dynamically to the probability distribution it is tracking.

Procedure used for Implementation

The procedure first executes a file to convert the selected avi file video in to frames. This file generates a series of frames of the video file and creates a folder in which these frames are stored for further processing. These frames are then converted from RGB scale to Gray Scale png files. Then the absolute difference between the consecutive frames is calculated and stored in the same folder created above. Then the Threshold of convergence (in pixels for each deg. of freedom) is set. Here we have used the value as:

The number of pixels to expand search window is provided. The value used here is 5. Initial search window size and Initial location of search window is set. The x and y coordinates are set for plotting motion. The image is converted from RGB space to HSV space. Hue information is extracted from the converted image. The search window box is created on the image. The centroid of search window is computed. The threshold is checked. The known information about the centroid and Mean convergence is used to alter the search window size. Window size is adjusted according to the new altered window size. AVI movie parameters are displayed on the screen as shown in experimental results.

5. EXPERIMENTAL RESULTS

For experiment purpose "MATLAB "version 2012, is used to execute all the required source code. The procedure discussed is applied on a video file. The search window size used here is 5.The first code allows the user to select a video file, when

executed and it allows the user to browse the required video file. When the user selects a particular video file then the corresponding frames are generated and stored in a folder in the current directory. Then the next code is executed that takes the generated frames as input and calculates the absolute difference between each frame and stores it in the same folder that was created in the previous step. In this approach, a rectangular target window is defined in an initial frame for a moving target in a video, and after that the tracked object is separated from the background by processing the data within the window. At the end the actual mean shift code is executed to track the object in the selected video in each subsequent frame stored. The output video is stored in the current directory. The screenshots of the output video are:

Figure 4: Target Selection

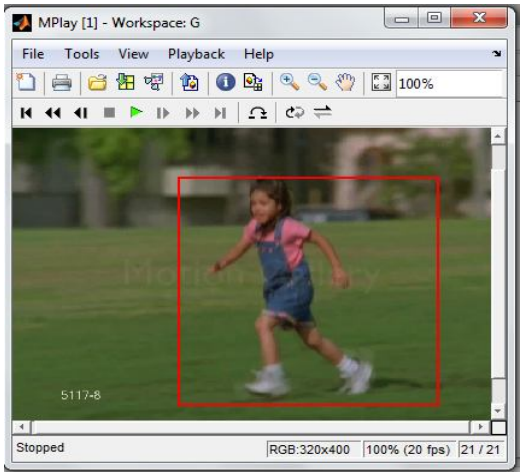

Figure 5: Salient Object Detection

6. CONCLUSION

Object tracking in an untidy surrounding remains a demanding investigation subject. In this paper we reviewed the mean shift algorithm with some definite improvements. At first, the frames of the video file are created. And after that, the window size is estimated to track a target precisely when the target's shape and location are altering. Numerous consequences can be obtained by implementing mean-shift algorithm over the target object.

According to the motion model, one should begin the tracker in several different positions in the neighborhood of basin of attraction if the movement of the target from frame to frame is well known to be greater than the operational basin of attraction. If total closure is present, one should take on a more revealing motion filter. In the same way, one should check that the selected target description is adequately different for the application domain or not. Hence our review of tracking an object will surely make us to analyze new areas of investigation and moreover, also helps to improve its applications in the already existing areas.

REFERENCES

[1] Comaniciu, V. Ramesh and Peter Meer, "*Kernel-Based Object Tracking*," IEEE TRANSACTIONS ON PATTERN ANALYSISAND MACHINE INTELLIGENCE, VOL. 25, NO. 5, MAY 2003.

[2] W. Hu, T. Tan, L. Wang, S. Maybank, "*A survey on visual surveillance of object motion and behaviors*", IEEE Trans. Syst. Man Cyber.-C 34 (3) (2004) 334–352.

[3] R. T. Collins, "*Mean-shift blob tracking through scale space*", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2003.

[4] R. Bradski. "*Computer vision face tracking for use in a perceptual user interface*", Intel Technology Journal, 2nd Quarter, 1998.

[5] R. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade, "*Algorithms for cooperative multisensor surveillance*, "*Proceedings of the IEEE*, vol. 89, no. 10, pp. 1456–1477, 2001.

[6] Comaniciu, D.; Ramesh, V.; Meer, P., "*Real-time tracking of nonrigid objects using mean shift*," Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on, vol.2, no., pp.142, 149 vol.2, 2000.

[7] V. Ferrari, T. Tuytelaars, and L. V. Gool, "Real-*time affine region tracking and coplanar grouping*," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Kauai, Hawaii, volume II, 2001,pp.226–233

[8] S. Intille, J. Davis, and A. Bobick, "*Real-time closed-world tracking*," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, 1997, pp. 697–703.Aparna Shivhare et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (4) , 2015, 3774-3777

www.ijcsit.com 3776

[9] D. Bue, D. Comaniciu, V. Ramesh, and C. Regazzoni, "*Smart cameras with real-time video object generation*," in *Proc. IEEE Intl. Conf. on Image Processing*, Rochester, NY, volume III, 2002,pp.429–432.

[10] U. Handmann, T. Kalinke, C. Tzomakas, M. Werner, and W. von Seelen, "*Computer vision for driver assistance systems*," in *Proceedings SPIE*, volume 3364, 1998, pp. 136–147.

[11] Wang, Yao, Jörn Ostermann, and Ya-Qin Zhang. *Video Processing and Communications*. Signal Processing Series. Upper Saddle River, N.J.: Prentice Hall, 2002. ISBN 0-13-017547-1.

[12]http://in.mathworks.com/solutions/image-video-processing/videoprocessing.Html

[13] G.R.Bradski. "Computer vision face tracking for use in a perceptual user interface", Intel Technology Journal,2nd Quarter,1998.