ISSN (Online): 2230-7893
www.IJCEM.org

Object tracking in a video sequence using Mean-Shift Based Approach: An Implementation using MATLAB7

Madhurima[1], Madhulika[2]
[1] Computer Science & engineering, Apeejay Stya University,
Sohna, Gurgaon 122 103, India
10madhurima@gmail.com

[2] Computer Science & Engineering, Skyline Institute of Engineering & Technology[3],
Plot no. 3, Knowledge Park-2, Greater Noida, G.B. Nagar (U.P.)
madhulikabhatia@gmail.com

Abstract

Object-tracking is one of the most popular areas of video processing because of its applicability to daily problems and ease of production, e.g. surveillance cameras, adaptive traffic lights, plane detection, vehicle navigation, human-computer interaction, object-based video compression, smart rooms, driver assistance, perceptual user interface, augmented reality etc. The purpose of object tracking is to determine the position of the object in images continuously and reliably against dynamic scenes. The bases of the work is the mean shift object-tracking algorithm for a moving target in a video by defining a rectangular target window in an initial frame, and then process the data within that window to separate the tracked object from the background. The paper also includes experimental results of the tracking using the mean shift based algorithm with certain improvements to make it suitable for tracking fast moving object.

Keywords: *Object-tracking, Mean-shift, video, frame, color model.*

1. Introduction

Object tracking is a mechanism to track an object and to take an action on another object with no relationship to the tracked objects, based on changes to the properties of the object being tracked. Object tracking is important because it enables several important applications such as: Security and surveillance - to recognize people, to provide better sense of security using visual information; Medical therapy - to improve the quality of life for physical therapy patients and disabled people; Retail space instrumentation - to analyze shopping behavior of customers, to enhance building and environment design; Video abstraction - to obtain automatic annotation of videos, to generate object-based summaries; Traffic management - to analyze flow, to detect accidents; Video editing - to eliminate cumbersome human-operator interaction, to design futuristic video effects; Interactive games- to provide natural ways of interaction with intelligent systems such as weightless remote control. The main difficulty in video tracking is to associate target locations in consecutive video frames, especially when the objects are moving fast relative to the frame rate [1]. Numerous approaches for object tracking have been proposed. The approach to be used depends on the context in which the tracking is performed and the end use for which the tracking information is being sought.

2. Mean shift based tracking approach

Mean shift (MS), which was proposed by Fukunaga and Hostetler in 1975. The mean shift algorithm is a nonparametric clustering technique that does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters. The mean-shift algorithm is a non-parametric density gradient estimator. It is basically an iterative expectation maximization-clustering algorithm executed within local search regions [2].
Given n data points x_i, i = 1, ..., n on a d-dimensional space Rd, the multivariate kernel density estimate obtained with kernel K(x) and window radius h is given below:

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

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

$$K(x) = c_{k,d}k(\| x \|^2) \qquad (2)$$

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

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

$$\nabla f(x) = \frac{2c_{k,d}}{n\,h^{d+2}} \left[\sum_{i=1}^{n} g\left(\| \frac{x - x_i}{h} \|^2 \right) \right] \left[\frac{\sum_{i=1}^{n} x_i g(\| \frac{x - x_i}{h} \|^2)}{\sum_{i=1}^{n} g(\| \frac{x - x_i}{h} \|^2)} - x \right] \qquad (3)$$

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

$$m_h\ (x) = \frac{\sum_{i=1}^{n} x_i g(\| \frac{x - x_i}{h} \|^2)}{\sum_{i=1}^{n} g(\| \frac{x - x_i}{h} \|^2)} - x \qquad (4)$$

is the mean shift..

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).

After executing the Mean shift algorithm 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. For this problem, traditional Mean shift algorithm easily failed to track fast moving object, some solutions had proposed like combining Kalman filter or Particle filter with Mean shift algorithm [3]. It predicted the direction and speed of the object first, and then adjusting the search window center of the Meanshift convergence, the center of convergence will become more accurate. However, such methods need high CPU costing to predict the moving object location, it's unsuited to be using in real time tracking system. The improved algorithm is based on an adaptation of mean shift that finds the mean (mode) of the distribution of a given probability density image and then iterating in the direction of maximum increase in probability density. Figure 1(a) and 1(b) shows the RGB coordinate system and RGB color Model respectively. Figure 2(a) and 2(b) shows the HSV coordinate system and HSV color Model respectively. In this Algorithm, a probability distribution image of the desired color in the video sequence is created. It first creates a model of the desired hue using a color histogram and uses the Hue Saturation Value (HSV) color system [4] that corresponds to projecting standard RGB color space along its principal diagonal from white to black. Figure 3 shows the obtainable HSV colors that lie within a triangle whose vertices are defined by the three primary colors in RGB space.

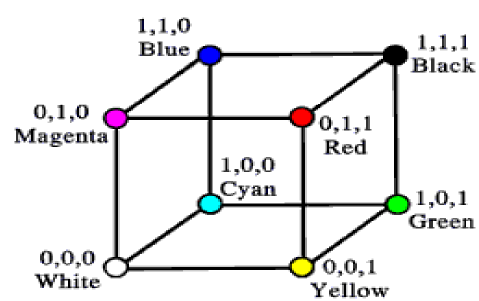

Fig.1 (a) RGB Coordinates System

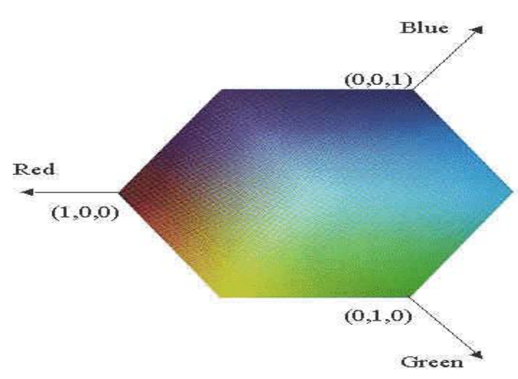

Fig.1 (b) RGB Color Model

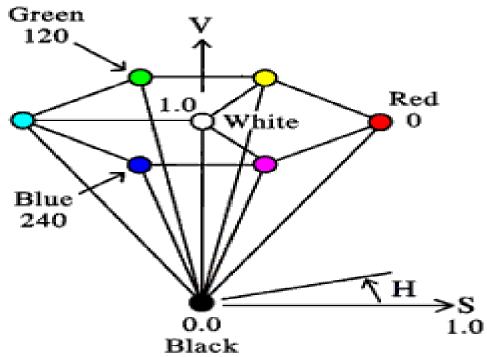

Fig.2 (a) HSV Coordinates System

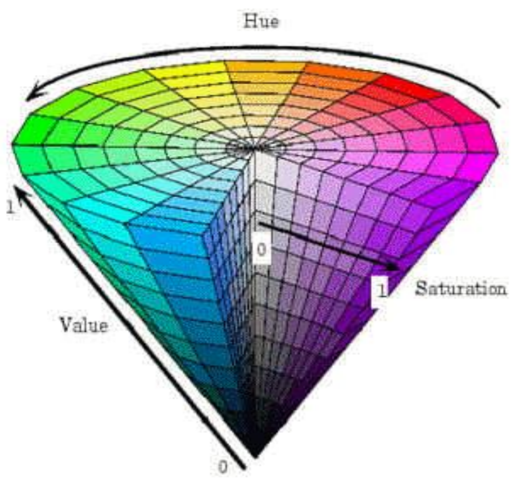

Fig.2 (b) HSV Color Model

The main disadvantage of the RGB color space in applications with natural images is a high correlation between its components: about 0:78 for r_{BR} (cross correlation between the B and R channel), 0:98 for r_{RG} and 0:94 for r_{GB} [8]. 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.

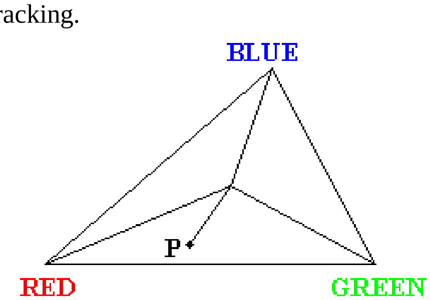

Fig 3. Obtainable HSV color from RGB color space

The algorithm discussed below is an adaptation of mean shift algorithm. And it is calculated as:

1. Choose the initial location of the search window.
2. Mean Shift as above (one or many iterations); store the zeroth moment.
3. Set the search window size equal to a function of the zeroth moment found in Step 2.
4. Repeat Steps 2 and 3 until convergence (mean location moves less than a preset threshold).

3. 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 1. 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.

3. Experimental Results

For experiment purpose "MATLAB 7.0" is used to execute all the required source code. The procedure discussed above is applied on various video files like Samplevideo.avi, redcup.avi etc. The search window size used here is 5. The first code allows the user to select a video file is executed that 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. As an example, the 1st Frame, 2nd Frame, 61st Frame and 62nd Frame are shown in figure 4(a), 4(b), 4(c) and 4(d) respectively. After calculating the absolute difference between the 1st Frame and 2nd Frame and 61st Frame and 62nd Frame the resultant frames are shown in figure 5(a) and 5(b) respectively. At the end the actual mean shift code is executed that track the object in the selected video in each subsequent frame stored. And the output video is stored in the current directory. The screen shot of the output video is shown in figure 6.

Fig 4. (a) 1st Frame of "Samplevideo.avi"

Fig 4. (b) 2nd frame of Samplevideo.avi"

Fig 4. (c) 61st frame of "Samplevideo.avi"

Fig 4. (c) 62nd frame of "Samplevideo.avi"

Fig 5. (a) Absolute difference between 1st frame

Fig 5. (b) Absolute difference between 61st and 62nd frame

Fig 6. Screen shot of "Samplevideo.avi" output video

5 Conclusion

Target tracking in a cluttered environment remains a challenging research topic. In this paper we implemented the mean shift algorithm with certain improvements. Firstly, the frames of the avi video file are generated. Secondly, the window size is calculated to track a target accurately when the target's shape and orientation are changing. Various results of the experiment conducted on the "Samplevideo.avi" file are shown. For example, if the motion of the target from frame to frame is known to be larger than the operational basin of attraction, one should initialize the tracker in multiple locations in the neighborhood of basin of attraction, according to the motion model. If total occlusions are present, one should employ a more sophisticated motion filter. Similarly, one should verify that the chosen target representation is sufficiently discriminant for the application domain. The kernel-based tracking technique, when combined with prior task-specific information, can achieve reliable performance [5]. Experimental results revealed that this enhanced model could successfully detect and track a human subject in random motion and in a condition where there is a certain change in illumination. The bounding rectangular tracking window's size depends quite significantly on the probability distribution of the image within the search window. Since improved algorithm changes its search window size by depending on the Zeroth Moment, and this relies on the interested pixel values. Thus, it enhances the selection concentration and in the robustness of object tracking by eliminating noisy pixels prior to computing the Zeroth Moment within the probability distribution image. It was found that a strong change in illumination caused the procedure to suffer quite severe distortion in the probability distribution image. The non-adaptive nature of the mean shift algorithm may lead to an incorrect tracking decision.

References

[1] 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.
[2] R. T. Collins, "Mean-shift blob tracking through scale space", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2003.
[3] G. R. Bradski. "Computer vision face tracking for use in a perceptual user interface", Intel Technology Journal, 2nd Quarter, 1998.
[4] A.R. Smith, "Color Gamut Transform Pairs," SIGGRAPH 78, pp. 12-19, 1978.
[5] F. Aherne, N. Thacker, and P. Rockett, "The Bhattacharyya metric as an absolute similarity measure for frequency coded data," Kybernetika, vol. 34, no. 4, pp. 363–368, 1998.

Madhurima received the Master's Degree in Information Technology from Guru Gobind Singh Indraprastha University, New Delhi, B.Engg in Information Technology from B.S Anangpuria Institute of Technology and Management, Faridabad and Diploma in Information Technology from Government Polytechnic for Women, Faridabad. From Aug 2006 to Oct 2007, she was working as a lecturer in B.S.A.I.T.M, Faridabad. From Oct 2007 to Aug 2010 she was working as a lecturer in Apeejay College of Engg., Gurgaon. She is currently working as an Assistant Professor in Apeejay Stya University, Gurgaon. She has published approximately fourteen research papers in various international and national journals and conferences. Her research interests are in image segmentation, object tracking and object oriented testing. She has published one book with the title " Computer Networks" with Laxmi Publications.

Madhulika received the Master's Degree in Computer Science from LIMAT, Faridabad, B.Engg in CSE from B.S Anangpuria Institute of Technology and Management, Faridabad and Diploma in CSE from Government Polytechnic for Women, Faridabad. From Jan 2006 to Oct 2007, she was working as a lecturer in B.S.A.I.T.M, Faridabad. From Oct 2007 to Feb 2010 she was working as a lecturer in Apeejay College of Engg., Gurgaon. She is currently working as an Assistant Professor in Skyline Institute of Engineering & Technology, Greater Noida. She has published approximately fourteen research papers in various international and national journals and conferences. Her research interests are in image segmentation, object tracking, Skin Detection and Face Recognition Systems. She has published one book with the title "Computer Networks" with Laxmi Publications.