Mean-Shift Tracking of Moving Objects Using Multi-Dimensional Histograms

Halil I. Cüce, and A. Enis Çetin

Department of Electrical and Electronics Engineering, Bilkent University, Ankara, 06800, Turkey

ABSTRACT

In this paper, a moving object tracking algorithm for infrared image sequences is presented. The tracking algorithm is based on the mean-shift tracking method which is based on comparing the histograms of moving objects in consecutive image frames. In video obtained after visible light, the color histogram of the object is used for tracking. In forward looking infrared image sequences, the histogram is constructed not only from the pixel values but also from a highpass filtered version of the original image. The reason behind the use of highpass filter outputs in histogram construction is to capture structural nature of the moving object. Simulation examples are presented.

Keywords: Mean-shift tracking, image histogram, FLIR images, moving objects in video

1. INTRODUCTION

Moving object tracking in video is a critical task in many applications including surveillance in CCTV systems, vision based human-computer interaction, and infrared imaging systems [1-6]. Recently, tracking using the mean shift method became popular (see e.g. the references in [3]) and Yilmaz et al. [4] for combined kernel tracking with global motion compensation for forward-looking infrared (FLIR) imagery.

The mean shift method relies on the intensity distributions or smoothed histograms of the target region and estimates the location of the target center in the image space in an iterative manner. The original method [1-3] is based only on the one-dimensional histograms constructed from the image intensity values and therefore it completely misses the texture information of the object. One-dimensional histograms approximating probability density functions (pdf) do not incorporate spatial relation of the image intensity values. This produces problems especially in small targets in FLIR sequences [4-5] because the number of target intensity values are relatively small compared to large targets. In order to achieve a more realistic target modelling two-dimensional histograms or estimates of the two-dimensional probability density functions (pdf) are used to model targets in [4-5]. In addition, the authors used an estimate of the pdf of the local standard deviation of the target region to highlight the low contrast between the target and the background in FLIR images. In,⁸ edges of the object are highlighted by including an additional bit in the histogram characterizing the object. If there is a significant change in consecutive pixels then the additional bit is set to 1.

In this paper, the two-dimensional image data is processed using a highpass filter to capture object features including edges on the object. Let the random vector $[x, xh]$ represent the original and highpass filtered version of the original image x . In this article, the histogram or the pdf of this vector is estimated and the tracking is carried out using the composite histogram which characterizes the object not only from the image pixels but also from the pixels of the high-pass filtered image. The main advantage of this approach over the regular intensity value based tracking is that the first entry captures the intensity information of the target and the second entry captures (i) the pixel intensity variation information, and (ii) highlights the contrast between the target and the background around the edges of the target.

In the next section, the mean shift tracking algorithm [1] is reviewed. In Section III, the new tracking feature vector is discussed in detail and simulation results are presented in Section IV.

Further author information: This work was supported in part by Turkish Academy of Sciences (TUBA)- GEBIB program. Send correspondence to A.E.C.

A.E.C.: E-mail: cetin@bilkent.edu.tr

H.I.C.: E-mail: hic@ee.bilkent.edu.tr

2. MEAN-SHIFT TRACKING ALGORITHM

Mean-shift tracking algorithm is an iterative scheme based on comparing the histogram of the original object in the current image frame and histogram of candidate regions in the next image frame. The aim is to maximize the correlation between two histograms.

Let x represent an image $x : Z^2 \rightarrow \{0, 1, 2, \dots, N-1\}$ where N is the dynamic range of the pixel values. Let O represent the support of an object. Let y_o be the center of mass of the support O which is a subset of Z^2 . The histogram of the object O in image frame n is defined as follows

$$h_O(\ell) = \sum_{(n_1, n_2) \in O} (\delta(x(n_1, n_2) - \ell)) \quad (1)$$

where $x(n_1, n_2)$ is the value of the (n_1, n_2) -th pixel of the image x , $\delta(\cdot)$ is the discrete Dirac-delta function, and $\ell = 0, 1, 2, \dots, N-1$. The histogram is smoothed and normalized in several ways in.¹ In this paper, an ordinary low-pass FIR filter is used to smooth the histogram.

In mean-shift tracking algorithm histograms of regions are compared to each other using the Bhattacharya coefficient:

$$\rho(h_O, h) = \sum_{\ell} \sqrt{h_O(\ell)h(\ell)} \quad (2)$$

In the next image frame, the histogram of the same region is estimated and the following weights $w(n_1, n_2)$ for each pixel in the region O are computed.

$$w(n_1, n_2) = \sum_{\ell=0}^{N-1} \sqrt{\frac{h_O(\ell)}{h(\ell)}} (\delta(x(n_1, n_2) - \ell)) \quad (3)$$

After this step, the center of mass $\mathbf{y}_t$ of the weights are determined. The histogram of the region centered at $\mathbf{y}_t$ is computed and compared it to the original histogram using the Bhattacharya coefficient. If

$$\rho(h_O, h_{y_o}) < \rho(h_O, h_{y_1}) \quad (4)$$

then the first candidate region for the object in the $(n+1)$ -st image frame is determined as the region centered at $(y_o + y_1)/2$. The above procedure is repeated until a satisfactory convergence level is reached. In practice, the average number of iterations is 5.

If a background subtraction based moving object detection method is used to determine the moving blobs in the image then the initial starting point of the iterations can be the blob determined by the background subtraction algorithm.⁷ Mean-shift tracking iterations improve the accuracy of the detected region obtained using background subtraction.

3. IMAGE FEATURE SELECTION

The efficiency of the mean-shift tracking algorithm depends on how representative the choice of random variables discriminating an object. The original algorithm is implemented in color image sequences and it uses color information. In infrared image sequences there is only the intensity information and the histogram constructed from the intensity information may not be satisfactory to distinguish objects. In a FLIR image there may be image regions with similar intensity histograms and this makes tracker to loose the moving object.

The ultimate goal is to find a sufficient statistic whose distribution characterizes the moving object in a unique manner. Since ordinary and FLIR images have no underlying statistical random process, it is almost impossible to define a sufficient statistic.

By using a high-pass filtered version of the image together with the actual image pixels the aim is to increase the discriminating capability of the mean-shift tracker. Pixels of the high-pass filtered image can capture structure on the moving object including the edges.

Figure 1. FLIR image of a plane. Image size is 71 by 118.

Figure 2. High-pass filtered version of the above plane image.

In Figure 1, FLIR image of a plane is shown. The image shown in Fig. 2 is a high-pass filtered version of this image. A high-pass filter with cut-off frequency at $\pi/2$ is used. This image does not clearly reflect the edges of the object. In Fig. 3, another high-pass filter with cut-off $\pi/4$ is used. This image clearly shows the edges. Therefore, it is necessary to use a high-pass filter with a lower cut-off frequency to extract the image structure.

High-pass information can be incorporated into the histogram. Since the dimension of the data is doubled compared to pixel-only histograms, the dynamic range of the new histogram is N^2 . In order to reduce the range the highpass filtered image is quantized to 4 levels, and this leads to a $4 \times N$ size histograms.

As it can be seen from the above images high-pass filtering captures (i) the pixel intensity variation information, and (ii) highlights the contrast between the target and the background around the edges of the target. This robustifies the performance of the tracking algorithm. Otherwise the histogram of the object shown in Fig. 1 essentially contains a single peak around the mean value of the object region and this is not enough to uniquely determine the object.

In Figure 4 another FLIR image is shown. This image sequence is downloaded from the web page: www.dii-llc.com/FLIR_multimedia.html. In this image there are edges within the object as well. Therefore it is important to include these features in the histogram characterizing the object in a unique manner. The highpass histograms of the plane in Figure 3 and the moving car in Figure 4 are shown in Figure 5-a and 5-b, respectively. The dynamic range of the graph in Figure 5-b is higher than Figure 5-a. As it can be seen from these graphs high-pass information provides additional information about the moving object.

A composite histogram describing both regular pixel values and pixel values of the high-pass filtered image can be constructed in several ways. The easiest approach is to concatenate the normalized histograms of two images.

Figure 3. High-pass filtered version of the above plane image. The cut-off frequency of the high-pass filter is $\pi/4$.

Figure 4. A FLIR image from the web page: www.dii-llc.com

In this case, the dynamic range of the concatenated histogram is $2N$. One can also create a two-dimensional data set by combining the pixels of the two images. In this case, the dimension of the data is doubled compared to a pixel-only histogram, therefore the dynamic range of the new histogram is theoretically N^2 . However, actual dynamic range of the high-pass filtered image is much smaller than dynamic range of intensity pixels. Quantized version of the high-pass filtered image pixels can be used. It is experimentally observed that 4 to 8 levels are found satisfactory, and this leads to a $4 \times N$ to $8 \times N$ size histograms. The second approach provides a more robust description of the object. This comes at the expense of additional computation. However, in both cases, the computational complexity does not increase drastically with the additional high-pass information in both approaches.

4. SIMULATIONS EXAMPLES

The object in Figure 1 can be tracked very easily by trackers using image pixels, and image and high-pass filtered image pixels because the plane is flying on an open sky.

The car in the parking lot (Figure 2) can be also tracked very easily by both trackers using image pixels, and image and high-pass filtered image pixels because the background is not cluttered in this scene, either. A couple of overlaid images showing the tracking is shown in Figure 6.

Figure 5. (a) Histogram of the highpass image shown in Figure 3, (b) highpass histogram of the moving car in Figure 4. Mean values of high-pass histograms are very close to zero. The variance of the histogram in part (b) is higher than the variance of the histogram in part (a). In the above plots, histograms are shifted by 100.

Figure 6. Tracking the car in the parking lot.

In Figure 3, the IR image of an aeroplane is shown. In this case, the ordinary mean-shift tracker fails to track the object as shown in Figure 7. It is clear that gray-scale pixel histogram is not enough to uniquely characterize the plane in this case. On the other hand, the tracker using both the FLIR pixel and high-pass information successfully tracks the object as shown in Figure 8.

In Figure 9 the frequency response of the FIR high-pass filter used in filtering the image is shown. This image is filtered both horizontally and vertically using the filter shown in Figure 9 to obtain the high-pass filtered image. The computational cost of filtering operation is low because the FIR filter has only 7 nonzero coefficients. To obtain the high-pass histogram there is no need to filter the entire image frame. Only a window around the moving object is sufficient because means-shift tracker performs a local search to determine the next location of the object.

Figure 7. Mean-shift tracking of a plane using only FLIR pixel information. The tracker fails to track the plane.

Figure 8. Tracking the plane using both highpass information and regular pixels. The plane is successfully tracked.

Figure 9. Frequency response of the high-pass filter.

5. CONCLUSION

Figure 8 clearly shows the superior quality of the proposed tracker using both the pixel information and the high-frequency information over the ordinary mean-shift tracker using only pixel information. The same behavior is observed in all the examples tried. Whenever the background is cluttered the ordinary mean-shift tracker fails to track the moving object due to the lack of information. The histogram constructed from the gray-level information of FLIR images is not sufficient to uniquely characterize the object.

By including high-pass information to tracking process the robustness of the tracker is improved against noise and cluttered background. Additional high-pass information leads to unique characterization of the moving object by the histogram in all the examples tried.

The computational cost of the proposed scheme is low. It consists of only an additional high-pass filtering operation and an increased size in histogram construction. High-pass filtering operation may not be carried out over the entire image. It is enough to perform high-pass filtering only around the moving object.

ACKNOWLEDGMENTS

We gratefully acknowledge *Digital Imaging Infrared LLC* of Apopka, Florida for letting us to use their FLIR videos in our research.

REFERENCES

1. D. Comaniciu, V. Ramesh, and P. Meer, "Real-Time Tracking of Non-Rigid Objects Using Mean Shift," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. II, pp. 142-149, June 2000.
2. D. Comaniciu, et.al, "Real-Time Tracking of Non-Rigid Objects Using Mean Shift", US patent pending, 2000.
3. Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer, "Kernel-Based Object Tracking," *IEEE Trans. on Pattern Anal. and Machine Intelligence*, Vol. 25, No. 5, pp. 564-579, May 2003.
4. A. Yilmaz, K. Shafique, N. Lobo, X. Li, T. Olson, and M. Shah, "Target Tracking in FLIR Imagery Using Mean Shift and Global Motion Compensation," *IEEE Workshop on Computer Vision Beyond Visible Spectrum*, 2001.
5. A.Yilmaz, K .Shafique, and Mubarak Shah, "Target Tracking in Airborne Forward Looking Infrared Imagery," *Journal of Image and Vision Computing*, Vol. 21, No. 7, 2003.

6. Murat Bagci, Yasemin Yardimci, A. Enis Cetin, "Moving object detection using adaptive subband decomposition and fractional lower order statistics in video sequences," *Signal Processing*, October 2002.
7. F.M. Porikli, O. Tuzel, "Fast Object Tracking by Adaptive Background Models and Mean-Shift Analysis", *International Conference On Computer Vision Systems (ICVS)*, April 2003.
8. M.D. Flickner, I. Haritaoglu, "Method of detecting and tracking groups of people," US Patent application, No. 0107649 June 12, 2003.