Real-Time Infrared Object Tracking
Based on Mean Shift

Cheng Jian and Yang Jie

Institute of Image Processing & Pattern Recognition, Shanghai Jiaotong University,
Shanghai 200030, China
{ch_jian,jieyang}@sjtu.edu.cn

Abstract. The mean shift algorithm is an efficient method for tracking object in the color image sequence. However, in the infrared object-tracking scenario, there is a singular feature space, i.e. the grey space, for representing the infrared object. Due to the lack of the information for the object representation, the object tracking based on the mean shift algorithm may be lost in the infrared sequence. To overcome this disadvantage, we propose a new scheme that is to construct a cascade grey space. The experimental results performed on two different infrared image sequences show our new scheme is efficient and robust for the infrared object tracking.

1 Introduction

Object tracking is a hot topic in computer vision, image sequence analysis, video processing and compression, and perceptual user Interface. The essence of the object tracking is to determine the position of the object in the images. The mean shift algorithm is a nonparametric statistical method for seeking the nearest mode of a point sample distribution, which originally advocated by Fukunaga [1], and extended and brought to the attention of the image processing community by Yizong Cheng [2]. Recently, this algorithm has been adopted as an efficient technique for a wide variety of application, such as image segmentation, and appearance-based object tracking by G. R. Bradski [3], D. Comaniciu and Peter Meer [4, 5, 6].

Real-Time infrared object tracking is an important and challenging research task in many military applications. Infrared object tracking primarily addresses to localize and track the infrared thermal object in the infrared image sequences [7, 8]. Unfortunately, there are some disadvantages for the infrared image, such as extremely low signal to noise ratio (SNR), severe background clutter, non-repeatability of the target signature, and high ego-motion of the sensor. Commonly, many infrared object tracking algorithms are imposed different constraints, such as no drastic change of the object feature [7], and no sensor ego-motion [8]. In this paper, we follow the 'mean shift' tracking approach, which was proposed by D. Comaniciu et al. [5, 6]. Due to the nature of the infrared image, we choose the grey space as the feature space. In contrast to the color space, the lack of the information to represent the object will lead to the failure to track the infrared object. In order to overcome the disadvantage, we propose a new scheme that is to construct a cascade grey space. Then, the infrared object can be represented in the cascade grey space. The experimental results performed on two different infrared image sequences show our new scheme is efficient

and robust for the infrared small object tracking, and infrared object tracking in the severe clutter background.

2 Nonparametric Density Estimation and Mean Shift

There are several nonparametric methods available for probability density estimation [9, p.81]. Kernel estimation is one of the most popular techniques. Let $\{x_i\}_{i=1\dots n}$ be an arbitrary set of n points in the d-dimensional space R^d, the multivariate kernel density estimator with kernel K(x) and windows radius (i.e. bandwidth) h, in the point x is defined as [9, p.110]

$$\hat{p}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K(\frac{x - x_i}{h}) \tag{1}$$

There are several commonly used kernels, such as Epanechnikov kernel, and Normal kernel, etc. The Epanechnikov kernel is an optimum kernel that yields minimum mean integrated square error (MISE)

$$K_E(x) = \begin{cases} \frac{1}{2} c_d^{-1}(d+2)(1-\|x\|^2) & if \ \|x\| < 1 \\ 0 & otherwise \end{cases} \tag{2}$$

where c_d is the volume of the unit d-dimensional sphere, e.g., $c_1=2$, $c_2=\pi$, $c_3=4\pi/3$.

Let us introduce the profile [2] of a kernel K as a function k: $[0,\infty) \to R$ such that $K(x) = k(\|x\|^2)$. For example, according to (2) the Epanechnikov kernel is

$$k_E(x) = \begin{cases} \frac{1}{2} c_d^{-1}(d+2)(1-x) & if \ x < 1 \\ 0 & otherwise \end{cases} \tag{3}$$

Employing the profile notation we can write the density estimate (1) as

$$\hat{p}_K(x) = \frac{1}{nh^d} \sum_{i=1}^{n} k(\left\|\frac{x - x_i}{h}\right\|^2) \tag{4}$$

Let us denote

$$g(x) = -k'(x) \tag{5}$$

assuming that the derivative of k exists for all $x \in [0,\infty)$, except for a finite set of points. A kernel G can be defined as

$$G(x) = Cg(\|x\|^2) \tag{6}$$

where C is normalization constant.

Then, the estimate of the density gradient can be defined as the gradient of kernel density estimate (4)

$$\hat{\nabla}p_K(x) \equiv \nabla\hat{p}_K(x) = \frac{2}{nh^{d+2}}\sum_{i=1}^{n}(x-x_i)k'(\left\|\frac{x-x_i}{h}\right\|^2) = \frac{2}{nh^{d+2}}\sum_{i=1}^{n}(x_i-x)g(\left\|\frac{x-x_i}{h}\right\|^2)$$

$$= \frac{2}{nh^{d+2}}\left[\sum_{i=1}^{n}g(\left\|\frac{x-x_i}{h}\right\|^2)\right]\left[\frac{\sum_{i=1}^{n}x_i g(\left\|\frac{x-x_i}{h}\right\|^2)}{\sum_{i=1}^{n}g(\left\|\frac{x-x_i}{h}\right\|^2)} - x\right] \tag{7}$$

where $\sum_{i=1}^{n}g(\left\|\frac{x-x_i}{h}\right\|^2)$ can be assumed to be nonzero. Then, the last bracket in (7)

contains the sample mean shift vector

$$M_{h,G}(x) \equiv \frac{\sum_{i=1}^{n}x_i g(\left\|\frac{x-x_i}{h}\right\|^2)}{\sum_{i=1}^{n}g(\left\|\frac{x-x_i}{h}\right\|^2)} - x \tag{8}$$

Obviously, the value of $\hat{p}_K(x)$ will reach maximum when $M_{h,G}(x)$ equals zero. Moreover, the mean shift procedure is defined recursively by computing the mean shift vector $M_{h,G}(x)$, and a proof of the convergence can be found in [2].

3 Infrared Object Tracking

In our infrared object-tracking scenario, we follow the 'mean shift' tracking approach, which was proposed by D. Comaniciu et al. [5, 6]. In the mean shift object-tracking algorithm, there are three basic steps. Firstly, in order to characterize the object, a specific feature space must be chosen. In [3] the HSV color space was chosen, and in [5, 6] the RGB color space was chosen. In infrared object tracking, we choose the grey space as the feature space. After choosing the feature space, the object model q and the object candidate p are be represented by its probability density function (pdf) in the feature space. Secondly, for measuring the similarity between the object model and the object candidate, the similarity function must be defined. Alike [5, 6], the Bhattacharyya coefficient is chosen. Thirdly, it is to determine the location corresponding to the object in the current frame. In mean shift tracking algorithm, this step uses the gradient information, which is provided by the mean shift vector.

In our infrared object-tracking algorithm, there is a new idea: We choose the grey space as the feature space. In contrast to the color space, the lack of the information to represent the object will lead to the failure to track the object. In order to overcome the disadvantage, we propose a new scheme that is to construct a cascade grey space. The core of this scheme is that the infrared image is convoluted in the x-direction and y-direction to generate two grey sub-spaces by a specific one-dimensional filter. The cascade grey space is made of these two grey sub-spaces.

3.1 Infrared Object Representation

To represent the infrared object, we choose the grey space as the feature space in the infrared grey image sequence. Moreover, to satisfy the low-computational cost imposed by real-time processing, we employ the m-bin histogram to estimate the discrete density in the chosen feature space.

(1) Cascade Grey Space Construction

There are three sub-spaces, i.e. Red, Green, and Blue in the RGB color space, or Hue, Saturation, and Value in the HSV color space. However, there is a singular grey space in the grey infrared image. Due to this nature of the infrared grey image, we have to choose the grey space as the feature space. In contrast to the color space, there is a disadvantage of the information lack for the object representation in the grey space. With this disadvantage, the infrared object tracking will be lost in singular grey space. In order to overcome this shortcoming, we propose a new scheme in which we try to construct a cascade grey space. How to construct the cascade grey space through the singular grey space?

Our scheme to construct the cascade grey space mainly relies on the filtering method. Firstly, two specific one-dimensional filters must be defined. Here, we define two one-dimensional filters, i.e. filterX and filterY. Especially, filterX equals to the transpose of the filterY (i.e. filterX = filterY^T). One sub-space will be obtained by performing filtering with filterX in the x-direction of the infrared grey image. Similarly, performing filtering with filterY in the y-direction will make another subspace. After these two filtering operation, we obtain two infrared grey sub-spaces, which generate a cascade grey space. Then, the representation of the infrared object is credible in the cascade grey space. Moreover, the infrared object tracking is robust in our new scheme. We found that the performance of the tracker will be determined to a great extent by the choice of the filter in the infrared object-tracking scenario. Because there are low signal to noise ratio (SNR) and severe background clutter in the infrared image, the chosen filter to construct the cascade grey space should make the infrared small object prominent and weaken the background clutter, such as the Gabor filter, and the high-pass Butterworth filter.

(2) Object Model and Candidate Representation

Let $\{x_i\}_{i=1\ldots n}$ be the pixel locations of the object model, centered at x_0. Before calculating the histogram, we employ an isotropic kernel with a convex and monotonic decreasing kernel profile k(x) to assign a smaller weight to the locations farther from the center. Here, we choose the Epanechnikov kernel to represent the infrared object. Furthermore, the function b: $R^2 \to \{1\cdots m\}$ is the index of the histogram bin at the location x_i in the quantized feature space, h is the radius of the kernel profile and δ is the Kronecker delta function. The probability of the quantized grey vector u in the object model is given by

$$\hat{q}_u(x_0) = C_m \sum_{i=1}^{n} k(\left\|\frac{x_0 - x_i}{h}\right\|^2)\delta[b(x_i) - u] \tag{9}$$

where C_m is the normalization constant, and $C_m = \dfrac{1}{\sum_{i=1}^{n} k(\left\|\frac{x_0 - x_i}{h}\right\|^2)}$.

Similarly, the object candidate, centered at y in the current frame, is defined by

$$\hat{p}_u(y) = C_c \sum_{i=1}^{n} k(\left\|\frac{y - x_i}{h}\right\|^2) \delta[b(x_i) - u] \qquad (10)$$

where C_c is the normalization constant, and $C_c = \dfrac{1}{\sum_{i=1}^{n} k(\left\|\frac{y - x_i}{h}\right\|^2)}$. Note that C_c

does not depend on y, since x_i and y are organized in a regular image lattice. There-fore, C_c is a constant with a given kernel and window radius h.

3.2 Similarity Measurement Based on Bhattacharyya Coefficient

The mean shift object-tracking algorithm is an appearance based tracking method and it employs the mean shift iterations to find the object candidate that is the most similar to a given model in terms of intensity distribution. In [5, 6] the similarity between the distributions of the object model and candidates is expressed by a metric based on the Bhattacharyya coefficient. At location y, the sample estimate of the Bhattacharyya coefficient for object model density q_u and object candidate density $p_u(y)$ is given by

$$\hat{\rho}(y) \equiv \rho[\hat{p}(y), \hat{q}] = \sum_{i=1}^{m} \sqrt{\hat{p}_u(y)\hat{q}_u} \qquad (11)$$

Then, the distance between two distributions can be defined as

$$d(y) = \sqrt{1 - \rho[\hat{p}(y), \hat{q}]} \qquad (12)$$

In the infrared object tracking, to measure the similarity between the object model and candidate, we likewise define the metric based on the Bhattacharyya coefficient.

3.3 Object Localization

In the mean shift object-tracking algorithm, the object localization procedure starts from the position y_0 of the object in the previous frame and searches in the neighbor-hood. According to section 3.2, the most probable location y of the object in the cur-rent frame is obtained by minimizing the distance (12). Minimizing the distance is equivalent to maximizing the Bhattacharyya coefficient (11). In this procedure, the probabilities $\{\hat{p}_u(\hat{y}_0)\}_{u=1...m}$ of the object candidate at location y_0 in current frame must be computed at first. Then, using Taylor expansion around the value $\hat{p}_u(\hat{y}_0)$, the first-order approximation of the Bhattacharyya Coefficient is obtained

$$\rho[\hat{p}(y),\hat{q}] \approx \frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{p}_u(\hat{y}_0)\hat{q}_u} + \frac{1}{2}\sum_{u=1}^{m}\hat{p}_u(y)\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)}} \tag{13}$$

Then, introducing (10) in (13), we obtain

$$\rho[\hat{p}(y),\hat{q}] \approx \frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{p}_u(\hat{y}_0)\hat{q}_u} + \frac{C_c}{2}\sum_{i=1}^{n}w_i k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \tag{14}$$

where $\quad w_i = \sum_{u=1}^{m}\delta[b(x_i)-u]\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)}}$ (15)

Therefore, to minimize the distance (12), the second term in (14) has to be maximized, the first term being independent of y. The second term in (14) represents the density estimate computed with kernel profile k(x) at y in the current frame, with the data being weighted by (15). According to (8), the maximization can be efficiently achieved based on the iterative procedure of the mean shift vector

$$\hat{y}_1 = \frac{\displaystyle\sum_{i=1}^{n}x_i w_i g\left(\left\|\frac{\hat{y}_0-x_i}{h}\right\|^2\right)}{\displaystyle\sum_{i=1}^{n}w_i g\left(\left\|\frac{\hat{y}_0-x_i}{h}\right\|^2\right)} \tag{16}$$

where y_1 is the new location of the object in the current frame.

4   Experiments

Here, we exhibit the experimental results of the different infrared image sequences in our new scheme. Because one limitation of the mean shift tracking method is that at least some part of the object in the next frame should reside inside the kernel, we suppose that the object motion is continuous. There are two infrared image sequences: the first experiment deals with the infrared small object sequence, and the second experiment is to apply our new scheme to tracking the infrared object in the severe clutter background. These experiment are implemented on the MATLAB 6.1 platform with the Pentium IV 2.4GHz. The object model is chosen by hand in the first frame in our experiments. In the grey cascade space, the object model and the object candidate are represented by the histogram with 32×32 bins.

The infrared small object sequence has 120 frames of the 130×100 pixels in this experiment, and the radius h of the object model is $(h_x, h_y) = (14, 10)$. In order to make the infrared small object prominent, we choose one-dimensional Gabor filter to filter the infrared image in direction-x and direction-y for constructing the cascade grey space. In this experiment, the new scheme is capable of tracking at 10 fps (frames per second). The tracking result is shown every 30 frames in the Fig. 1.

The infrared object sequence with the severe clutter background has 300 frames of the 400×300 pixels, and the radius h of the object model is $(h_x, h_y) = (42, 25)$. Due to

Fig. 1. The infrared small object tracking results, from left to right, is the 1st, 30th, 60th, 90th, 120th frame.
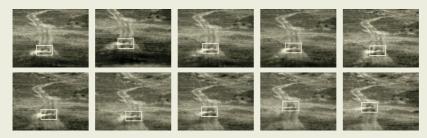

Fig. 2. The tracking result in the severe clutter background, the first row from left to right is the 1st, 30th, 60th, 90th, 120th frame, and the second row from left to right is the 150th, 180th, 210th, 240th, 270th frame.

the severe clutter background, we must try to weaken the clutter background to make the object tracking robust. In order to weaken the clutter background, we choose the one-dimensional high-pass Butterworth filter to filter the infrared image in direction-x and direction-y for constructing the cascade grey space. In this experiment, the new scheme is capable of tracking at 6 fps. The tracking result is shown every 30 frames in the Fig. 2.

The two experimental Results show that our opposed scheme is successful and indicate that our scheme is insensitive to the low SNR and the clutter background. Due to the mean shift gradient-descent seeking the object region in the feature space, the mean shift iterative processes rapidly converge in our two experiments. Therefore, our new scheme for the infrared object tracking can satisfy real-time processing.

5   Conclusion

We propose a new scheme in the infrared object tracking based on the mean shift algorithm. In our scheme, we construct a cascade grey space as the feature space for representing the infrared object. For the different infrared image sequence, we choose different filters to generate different cascade grey spaces. In order to demonstrate the performance of our scheme, we choose two kinds of infrared image sequence. The experimental results show our new scheme is efficient and robust for the infrared small object tracking and the severe clutter background.

References

1. K. Fukanaga, L. D. Hostetler: The Estimation of the Gradient of a Density Function, with Application in Pattern Recognition. IEEE Trans. Information Theory, 1975, Vol. 21: 32-40
2. Yizong Cheng: Mean Shift, Mode Seeking, and Clustering. IEEE Trans. Pattern Analysis and Machine Intelligence, 1995, Vol.17, No.8: 790-799

3. G. R. Bradski: Real Time Face and Object Tracking as a Component of a Perceptual User Interface. Proc. IEEE Workshop on Applications of Computer Vision, Princeton, (1998) 214-219
4. D. Comaniciu, P. Meer: Mean Shift: A Robust Approach towards Feature Space Analysis. IEEE Trans. Pattern Analysis and Machine Intelligence, 2002, 24(5): 603-619
5. D. Comaniciu, V. Ramesh, P. Meer: Real-Time Tracking of Non-Rigid Objects Using Mean Shift. Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2000, 142-149
6. D. Comaniciu, V. Ramesh, P. Meer: Kernel-Based Object Tracking. IEEE Trans. Pattern Analysis and Machine Intelligence, 2003, Vol. 25, No.5: 564-577
7. H.Shekarforoush, R. Chellappa: A Multi-Fractal Formalism for Stabilization, Object Detection and Tracking in FLIR Sequence. IEEE International Conference on Image Processing, 2000, Vol. 3
8. D. Daviesy, P. Palmery, M. Mirmehdiz: Detection and Tracking of Very Small Low Contrast Objects. Ninth British Machine Vision Conference, 1998
9. Andrew R. Webb: Statistical Pattern Recognition, Second Edition. Weily, 2002