Invariant Feature Matching based Adaptive Bandwidth Mean Shift and Its Application to Infrared Object Tracking

Fangzhou Zhao\textsuperscript{1,2}, Junshan Li\textsuperscript{1}, Yinghong Zhu\textsuperscript{1}, Wei Yang\textsuperscript{1}
1. Xi’an Research Inst. Of High-tech.
2. The Engineering College of CAPF
Xi’an, China
ark.fangzhou@yahoo.com.cn

Abstract—Mean shift algorithm has grained great success in object tracking domain due to its ease of implementation, real time response and robust tracking performance, however, the fixed kernel bandwidth may cause tracking failure for size changing objects. A novel object tracking algorithm for FLIR imagery is proposed based on mean shift with adaptive bandwidth. The scale invariant feature transform is employed to compute the affine model between the successive frames. Then, the scale and orientation of the kernel can be estimated by the gained parameters. Experiment results verify the effectiveness and robustness of this extraction algorithm which can improve the tracking performance efficiently.

Keywords—SIFT; mean shift; FLIR; affine model; object tracking

I. INTRODUCTION

Object tracking in the forward-looking infrared (FLIR) imagery has remained an important and challenging research task in many military applications including precise guidance and surveillance. Unfortunately, in contrast to visual images, FLIR images have extremely low signal-to-noise (SNR) ratios, poor object visibility, low dynamic range of gray level, and non-repeatability of the object signature\cite{1}. In addition, competing background clutter, the artifacts due to the weather conditions and high ego-motion of the sensor make the object tracking even harder.

Most tracking algorithms depend on the hot spot technique, which assume that the object is brighter than the background and acceptable levels of noise. Dawoud \cite{2} proposed a decision fusion algorithm for infrared object tracking using the weighted composite reference function. Strehl et al \cite{3} compensated the global motion using a multiresolution scheme based on affine motion model, but this scheme is unable to capture the skew, pan and tilt of the scene. Yilmaz \cite{1} utilized fuzzy clustering, edge fusion and local texture energy for detection infrared objects. Intensity and local standard deviation distributions are computed for object tracking.

Mean shift is a nonparametric statistical method for seeking the nearest mode of a point sample distribution, which originally advocated by Fukunaga. Cheng \cite{4} notes that mean shift is fundamentally a gradient ascent algorithm with an adaptive step size. Since Comaniciu \cite{5} first introduced mean shift-based object tracking, it has proven to be a promising alternative to popular particle filtering based trackers. The standard mean shift and other improved algorithm\cite{6}, however, use the fixed kernel bandwidth. In this paper, we use scale invariant feature transform to calculate the affine model between the successive frames. And then, the scale and orientation of the kernel can be estimated by the gained parameters. The kernel bandwidth can be adjusted by the scale value adaptively to improve the tracking performance.

The rest of the paper is organized as follows. Section 2 presents the mean shift based infrared object tracking algorithm. Section 3 illustrates the strategy of obtaining affine parameters based on SIFT. Experiment results and conclusions are presented in Section 4 and Section 5, respectively.

II. MEAN SHIFT BASED INFRARED OBJECT TRACKING

Tracking objects using mean shift algorithm is performed by iteratively translating a kernel in the image space such that the past and current target observations are similar\cite{5}. The target representation plays an important role in infrared target tracking. We choose the grey space as the feature space in the infrared grey image sequence to represent the target. The m-bin histogram is introduced to estimate the discrete density in the chosen feature space. Let \( \{x_i, i = 1, \ldots, n\} \) be the pixel locations of the target model. An isotropic kernel with a convex and monotonic decreasing kernel profile \( k(x) \) is employed to assign a smaller weight to the locations far from the target center \( x_0 \). The probability of the quantized grey vector \( u \) can be defined as

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

where \( b:R^2 \rightarrow \{1 \cdots m\} \) represents the index of the histogram bin at the location \( x_i \), and \( C_m \) is the normalization constant.

Similarly, the target candidate centered at \( y \) in the current frame can be defined by
\[
\hat{p}_u(y) = C_h \sum_{i=1}^{n_u} k \left\| \frac{y - x_i}{h} \right\|^2 \delta[b(x_i) - u]
\]

(2)

where \( C_h \) is also the normalization constant.

Then, we choose Bhattacharyya coefficient to measure the similarity between the distributions of the target model and the candidates. This metric is given by

\[
\hat{\rho}(y) = \rho[\hat{p}(y), \hat{q}] = \sum_{u=1}^{n} \sqrt{\hat{p}_u(y) \hat{q}_u}
\]

(3)

We can get the approximation of \( \hat{\rho}(y) \) using Taylor formula at \( \hat{p}_u(y_0) \)

\[
\hat{\rho}(y) = \frac{1}{2} \sum_{u=1}^{m} \sqrt{\hat{p}_u(y_0)} q_u + \frac{1}{2} \sum_{u=1}^{m} \hat{p}_u(y) \frac{\hat{q}_u}{\hat{p}_u(y_0)}
\]

(4)

Taking (2) into (4), we can get

\[
\hat{\rho}(y) = \frac{1}{2} \sum_{u=1}^{m} \sqrt{\hat{p}_u(y_0)} q_u + \frac{C_h}{2} \sum_{u=1}^{m} w_k \left\| \frac{y - x_i}{h} \right\|^2
\]

(5)

where \( w_k = \sum_{u=1}^{m} \left[ \frac{\hat{q}_u}{\hat{p}_u(y_0)} \delta[b(x_i) - u] \right] \). Maximizing the second part of (5), we can obtain the new location of infrared object.

\[
\hat{y}_i = \frac{1}{\sum_{i=1}^{n_u} x_i w_i} \sum_{i=1}^{n_u} x_i w_i \left( \frac{y_0 - x_i}{h} \right)
\]

(6)

where, \( g(x) = -k'(x) \). So the whole tracking procedure can be described as follows: First, initialize the candidate object location \( \hat{y}_0 \). Then, calculate the weight \( w_i \). Compute the (6) iteratively, until \( \| y_{r+1} - y_r \| < \varepsilon \) or the iterative number equal to the determined maximum. The final \( \hat{y} \) is then the tracking result of object location.

III. ADAPTIVE BANDWIDTH WITH SIFT

The SIFT features\(^{[7]} \) are local and based on the appearance of the object at particular interest points, and are invariant to image scale and rotation. They are also robust to changes in illumination, noise, and minor changes in viewpoint. In addition to these properties, they are highly distinctive, relatively easy to extract, allow for correct object identification with low probability of mismatch and are easy to match against a (large) database of local features. Object description by set of SIFT features is also robust to partial occlusion; as few as 3 SIFT features from an object are enough to compute its location and pose.

This is the stage where the interest points, which are called keypoints in the SIFT framework, are detected. For this, the image is convolved with Gaussian filters at different scales, and then the difference of successive Gaussian-blurred images are taken. Keypoints are then taken as maxima/minima of the Difference of Gaussians (DoG) that occur at multiple scales. Specifically, a DoG image \( D(x, y, \sigma) \) is given by

\[
D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, k\sigma)
\]

(7)

where \( L(x, y, k\sigma) \) is the original image \( I(x, y) \) convolved with the Gaussian blur \( G(x, y, k\sigma) \) at scale \( k\sigma \)

\[
L(x, y, \sigma) = G(x, y, k\sigma) * I(x, y)
\]

(8)

Hence a DoG image between scales \( k_i \sigma \) and \( k_j \sigma \) is just the difference of the Gaussian-blurred images at scales \( k_i \sigma \) and \( k_j \sigma \). For scale-space extrema detection in the SIFT algorithm, the image is first convolved with Gaussian-blurs at different scales. The convolved images are grouped by octave (an octave corresponds to doubling the value of \( \sigma \)), and the value of \( k \) is selected so that we obtain a fixed number of convolved images per octave. Then the Difference-of-Gaussian images are taken from adjacent Gaussian-blurred images per octave.

Once DoG images have been obtained, keypoints are identified as local minima/maxima of the DoG images across scales. This is done by comparing each pixel in the DoG images to its eight neighbors at the same scale and nine corresponding neighboring pixels in each of the neighboring scales. If the pixel value is the maximum or minimum among all compared pixels, it is selected as a candidate keypoint.

This keypoint detection step is a variation of one of the blob detection methods by detecting scale-space extrema of the scale normalized Laplacian, that is detecting points that are local extrema with respect to both space and scale, in the discrete case by comparisons with the nearest 26 neighbours in a discretized scale-space volume. The difference of Gaussians operator can be seen as an approximation to the Laplacian, here expressed in a pyramid setting.

Figure 1. Keypoints location and scale selection
And then, each keypoint is assigned one or more orientations based on local image gradient directions. This is the key step in achieving invariance to rotation, as the keypoint descriptor can be represented relative to this orientation and therefore achieves invariance to image rotation. Previous steps found keypoint locations at particular scales and assigned orientations to them. This ensured invariance to image location, scale and rotation. Now we want to compute descriptor vectors for these keypoints such that the descriptors are highly distinctive and partially invariant to the remaining variations, like illumination, 3D viewpoint, etc. This step is image closest in scale to the keypoint's scale. Just like before, the contribution of each pixel is weighted by the gradient magnitude, and by a Gaussian with $\sigma$ 1.5 times the scale of the keypoint. Histograms contain 8 bins each, and each descriptor contains a 4x4 array of 16 histograms around the keypoint. This leads to a SIFT feature vector with $(16 \times 8 = 128)$ elements. This vector is normalized to enhance invariance to changes in illumination.

Finally, we can use the SIFT to recovery the parameters of affine model between successive frames.

IV. EXPERIMENT RESULTS

Here, we exhibit the experimental results of the different infrared image sequences in our proposed tracking algorithm. These experiments are implemented on the MATLAB 2008 platform with the Pentium IV 3.0GHZ. The object model is chosen by hand in the first frame in our experiments. In the first experiment, we track the infrared tank with the proposed algorithm and mean shift tracking algorithm, respectively. The DoG images of frames and matching results are shown as follows.

Then, we can use the matching results to estimate the kernel bandwidth effectively. Part of the frames in the tank tracking experiment is shown in Figure 6. As shown in Figure 1, mean shift algorithm has a larger tracking error because of the blurred boundary and low dynamic range of gray level of infrared images. Comparatively, the proposed algorithm outperforms the mean shift algorithm. The object location is well amended by the feature matching method with the adaptive kernel bandwidth.

V. CONCLUSIONS

We propose a robust approach for tracking objects in FLIR imagery based on mean shift and SIFT. The infrared object is represented by kernel histogram in grey space. Mean shift algorithm is employed to track the object firstly. Because of the fixed kernel bandwidth, the object location obtained by mean shift algorithm always has some errors. So, in the second tracking stage, feature matching is adopted to amend the location error. We use SIFT to extract the feature points of template object and candidate area. And then, the bandwidth can be adjusted by the estimation of the parameters of affine model. Finally, we can capture the infrared object accurately. Experiment results show our
novel scheme is efficient and robust for tracking infrared objects.

ACKNOWLEDGMENT
This work was supported by National Natural Science Foundation of China under Grant No.60772151.

REFERENCES