Infrared target tracking using bandwidth adaptive mean shift

Qiao Liyong[1], Xu Lixin[1], Gao Min[2]

(1. School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China;
2. Missile Engineering Department, Ordnance Engineering College, Shijiazhuang 050003, China)

Abstract: A multiple features fusion and bandwidth adaptive mean shift tracking algorithm had been proposed. The iterative solution expressions of location and bandwidth had been established with fully parameterized bandwidth matrix based on M-estimator. The fused weight image had been produced with intensity and local standard deviation. The target template model had been generated by the combination of previous target template model and the mean value of the determined target models in the previous frames. An enlarged bandwidth matrix had been employed in the iterative solution of location vector to ensure location accuracy. To prevent the bandwidth from exploding in the presence of background clutter or imploding on self-similar target, regularization terms had been introduced. The visual results and evaluation measures show that the proposed tracking algorithm has the best performance compared with other three scale adaptive mean shift tracking algorithms.

Key words: infrared; tracking; bandwidth adaptive; M-estimator; mean shift

CLC number: TP391.4 **Document code:** A **Article ID:** 1007-2276(2015)01-0354-09

带宽自适应均值偏移红外目标跟踪

乔立永[1],徐立新[1],高 敏[2]

(1. 北京理工大学 机电学院,北京 100081; 2. 军械工程学院 导弹工程系,河北 石家庄 050003)

摘 要: 提出一种多特征融合和带宽自适应均值偏移跟踪算法。基于 M-估计器建立位置和带宽关于完整参数型带宽矩阵的迭代更新公式。分析权值图像的本质,基于灰度和局部标准差建立融合权值图像。通过先前目标模板模型和确定的目标模型的平均值生成当前目标模板模型。在位置向量的迭代公式中,采用扩大的带宽矩阵,确保定位精度。为防止由于背景杂波导致带宽膨胀或者由于目标自我相似导致带宽收缩,引入规范准则。跟踪的视觉结果和评估尺度表明,提出的跟踪算法相比于另外三种尺度自适应均值偏跟踪算法,具有最好的性能。

关键词: 红外; 跟踪; 带宽自适应; M-估计器; 均值偏移

收稿日期: 2014-05-05; **修订日期:** 2014-06-12

基金项目: 军械工程学院科研项目

作者简介: 乔立永(1982-),男,博士生,主要从事红外成像制导自动目标识别与跟踪方面的研究。Email:lile-116@163.com

导师简介: 徐立新(1969-),男,教授,博士生导师,博士,主要从事模式识别、复杂系统方面的研究。Email:lxxu@bit.edu.cn

0 Introduction

Being characterized by extremely low signal-to-noise ratio (SNR), poor target visibility, and time varying target model, infrared target tracking remains a challenging problem in pattern recognition and tracking applications[1-4].

To resolve the scale updating problem in classic mean shift tracking algorithm, Comaniciu proposed a scale plus-minus approach ($MS_{\pm}$)[5]. Zivkovic proposed an EM like kernel based tracking algorithm[6], which simultaneously estimated the position of the local mode and the covariance matrix that described the approximate shape of the local mode. Ning proposed a scale and orientation adaptive mean shift tracking algorithm (SOAMST)[7], which utilized the estimated area and the second order moment to adaptively estimate the width, height and orientation changes of the target.

In this paper, a robust infrared target tracking algorithm using bandwidth adaptive mean shift based on M-estimator in complex background (BAMS) has been proposed. To improve the robustness of mean shift algorithm while tracking scale-changing target in challenging infrared sequences, four measures have been taken.

1 Mean shift tracking with full bandwidth matrix

1.1 Location vector estimation

Let $\{x_i,\ i=1,\cdots,n\}$ be the pixels locations of the target region, the probability density function (PDF) of the target model based kernel function $k(x)$ can be defined as

$$\hat{q}^u(\hat{y}_0;H_0)=C\sum_{i=1}^{n}K_{H_0}(\hat{y}_0-x_i)\,\delta[b(x_i)-u] \tag{1}$$

Where $\hat{y}_0$ is the target center; δ is the Kronecker delta function; $b:R^2\rightarrow\{1,2,\cdots,m\}$ represents the index of the histogram bin at the location x_i; C is the normalization constant; the multivariate kernel $K_{H_0}(x)$ is employed to assign a smaller weight to the locations far from the target center, the multivariate kernel $K_{H_0}(x)$ is generated from the rotating of the univariate kernel $k(x)$ in R^d, i.e.

$$K_{H_0}(\hat{y}_0-x_i)=|H_0|^{-1/2}k\{\|H_0^{-1/2}(\hat{y}_0-x_i)\|^2\} \tag{2}$$

H_0 is the bandwidth matrix of target region, which determines the target size and orientation; In this paper, the full bandwidth matrix is chosen.

Similarly, let $\{y_i,\ i=1,\cdots,n_H\}$ be the sample points of the candidate region, the candidate target model centered at y can defined by

$$\hat{p}^u(y;H)=C_H\sum_{i=1}^{n_H}|H|^{-1/2}k\{\|H^{-1/2}(y-y_i)\|^2\}\,\delta[b(y_i)-u] \tag{3}$$

Where H is the bandwidth of candidate target region; C_H is also the normalization constant.

The similarity between the target and candidates is measured by Bhattacharyya distance.

$$d(y;H)=\sqrt{1-\rho(y;H)} \tag{4}$$

Where $\rho\ (y;H)$ is the Bhattacharyya coefficient given by

$$\rho(y;H)=\rho[\hat{p}(y;H),\hat{q}]=\sum_{u=1}^{m}\sqrt{\hat{p}^u(y;H)\hat{q}^u} \tag{5}$$

The candidate at location y possessing minimum distance with the target template model is determined as the target region in the current frame. Using Taylor expansion around the value $\hat{p}^u\ (\hat{y}_0;H_0)$ and after some manipulations, the linear approximation of the Bhattacharyya coefficient Eq.(5) is:

$$\rho[\hat{p}(y;H),\hat{q}]\approx\frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{p}^u(\hat{y}_0;H_0)\hat{q}^u}+\frac{C_H}{2}\sum_{i=1}^{n_H}\omega_i|H|^{-1/2}k\{\|H^{-1/2}(y-y_i)\|^2\} \tag{6}$$

Where

$$\omega_i=\sum_{u=1}^{m}\sqrt{\frac{\hat{q}^u}{\hat{p}^u(\hat{y}_0;H_0)}}\,\delta[b(x_i)-u] \tag{7}$$

To minimize the distance Eq. (4), the second term in Eq. (6) has to be maximized, the first term being independent of y. Observe that the second term represents the multivariate kernel density estimate[8]

$\hat{f}(y;H)$ computed with kernel profile $k(x)$ at y in the current frame, with the data being weighted by ω_i.

$$\hat{f}(y;H)=\frac{C_H}{2}\sum_{i=1}^{n_H}\omega_i|H|^{-1/2}k\{||H^{-1/2}(y-y_i)||^2\} \tag{8}$$

By taking the partial derivative of the $\hat{f}(y;H)$ with respect to location vector y, we can obtain the expression of density gradient estimators:

$$\nabla f(y;H)=\nabla\hat{f}(y;H)=\frac{\partial\hat{f}(y;H)}{\partial y}=$$

$$\frac{C_H}{n_H}\sum_{i=1}^{n}\omega_i|H|^{-1/2}k'\{(y-y_i)^{\mathrm{T}}H^{-1}(y-y_i)\}H^{-1}(y-y_i) \tag{9}$$

Where $g(x)=-k'(x)$, assuming that the derivative of $k(x)$ exists for all $x\in[0,\infty)$, except for a finite set of points.

After some manipulations, we can obtain the iterative solution expression of location vector y as following:

$$\hat{y}_1=\frac{\sum_{i=1}^{n_H}\omega_i g\{(\hat{y}_0-y_i)^{\mathrm{T}}H^{-1}(\hat{y}_0-y_i)\}y_i}{\sum_{i=1}^{n_H}\omega_i g\{(\hat{y}_0-y_i)^{\mathrm{T}}H^{-1}(\hat{y}_0-y_i)\}} \tag{10}$$

When the kernel profile function is Epanechnikov kernel function

$$k(x)=\begin{cases}1-x & 0\leqslant x\leqslant 1\\ 0 & x>1\end{cases} \tag{11}$$

$g(x)=-k'(x)=1$, substituting $k(x)$ and $g(x)$ to Eq.(10), the iterative solution expression of location vector y is:

$$\hat{y}_1=\sum_{i=1}^{n_H}\omega_i y_i/\sum_{i=1}^{n_H}\omega_i \tag{12}$$

1.2 Relation to M-estimator and full bandwidth matrix estimation

Huber[9] extended Maronna′s definition[10] to the most general form of M-estimators of multivariate location and scatter as following which are affine equivalent.

Let $\{y_i,i=1,2,\cdots,n\}\in R^d$ be a data set sampled from density $f(y)$, $f(y)=f(|y|)$ is a spherically symmetric probability density in R^d, the prototype distribution $f(y)$ is only approximately known. We apply general non-degenerate affine transformations $y_i\rightarrow H^{-1/2}(y-y_i)$ to obtain a d-dimensional location and scale family of "elliptic"densities[10]

$$f(y;H)=|H|^{-1/2}f(||H^{-1/2}(y-y_i)||) \tag{13}$$

Where $||\cdot||$ stands for Euclidean norm, $H\in P_d$, P_d denote the set of all $d\times d$ positive definite symmetric matrics. The problems are to estimate the location vector y and the scatter matrix H from the sampled data set $\{y_i,i=1,2,\cdots,n\}$.

Assume $\rho(y;H)=-\log f(y;H)$, the M-estimate of (y,H) can be obtained by minimizing $\sum_{i=1}^{n}\rho(y;H)$.

The most general form of affine equivariant M-estimator of multivariate location and scatter are usually defined implicitly as the solutions $y\in R^d$ and $H\in P_d$, respectively, to the simultaneous equations:

$$y=\frac{\sum_{i=1}^{n}u_1(s_i)y_i}{\sum_{i=1}^{n}u_1(s_i)} \tag{14}$$

$$H=\frac{\sum_{i=1}^{n}u_2(s_i)(y-y_i)(y-y_i)^{\mathrm{T}}}{\sum_{i=1}^{n}u_3(s_i)} \tag{15}$$

Where $s_i=(y-y_i)^{\mathrm{T}}H^{-1}(y-y_i)$, and with u_1, u_2 and u_3 being real-valued functions.

Kernel density estimation $\hat{f}(y;H)$ in Eq.(8) represents the probability density at data point y,. The relation between mean shift algorithm and M-estimator for location can be obtained by

$$\hat{y}=\arg\max_y\hat{f}(y;H)=$$

$$\arg\max_y\left\{\frac{C_H}{2}\sum_{i=1}^{n_H}\omega_i|H|^{-1/2}k\{||H^{-1/2}(y-y_i)||^2\}\right\}=$$

$$\arg\min_y\left\langle\frac{C_H}{2}\sum_{i=1}^{n_H}\omega_i\cdot|H|^{-1/2}\{1-k[||H^{-1/2}(y-y_i)||^2]\}\right\rangle=$$

$$\arg\min_y\left\{\frac{C_H}{2}\sum_{i=1}^{n_H}\omega_i\cdot|H|^{-1/2}\rho[||H^{-1/2}(y-y_i)||^2]\right\} \tag{16}$$

We can see that mean shift algorithm is equivalent to M-estimator for location estimation, which also had been proved in the seminal paper for mean shift tracking.

Equally, we can obtain the bandwidth matrix estimate based on scatter M-estimator for mean shift algorithm.

$$\frac{\partial \hat{f}(y;H)}{\partial H}=\frac{C_H}{4}\sum_{i=1}^{n_H}\omega_i\cdot|H|^{-1/2}H^{-1}k\{(y-x_i)^{\mathrm{T}}H^{-1}(y-x_i)\}-$$
$$\frac{C_H}{2}\sum_{i=1}^{n_H}\omega_i\cdot|H|^{-1/2}k'\{(y-x_i)^{\mathrm{T}}H^{-1}(y-x_i)\}\cdot$$
$$\{H^{-1}(y-x_i)(y-x_i)^{\mathrm{T}}H^{-1}\} \quad (17)$$

After some manipulations, we can obtain the iterative solution expression of bandwidth matrix as follows:

$$H_1=\frac{2\sum_{i=1}^{n_H}\omega_i g\{(y-y_i)^{\mathrm{T}}H_0^{-1}(y-y_i)\}\{(y-y_i)(y-y_i)^{\mathrm{T}}\}}{\sum_{i=1}^{n_H}\omega_i k\{(y-y_i)^{\mathrm{T}}H_0^{-1}(y-y_i)\}} \quad (18)$$

When the kernel profile function is Epanechnikov kernel function as in Eq.(11), substituting $k(x)$ and $g(x)$ to Eq. (18), the iterative solution expression of bandwidth matrix is:

$$H_1=\frac{2\sum_{i=1}^{n_H}\omega_i\{(y-y_i)(y-y_i)^{\mathrm{T}}\}}{\sum_{i=1}^{n_H}\omega_i\{1-(y-y_i)^{\mathrm{T}}H_0^{-1}(y-y_i)\}} \quad (19)$$

2 Target and local background model

2.1 Estimating target parameters from full bandwidth matrix

For the symmetric positive definite bandwidth matrix H, the following Eq. (20) defines a unique ellipse region[11] which is used to represent the target.

$$(x-x_i)^{\mathrm{T}}H^{-1}(x-x_i)\leqslant 1 \quad (20)$$

The axes of the ellipse point along the eigenvectors of H, the half-length of these axes equal to $\sqrt{\lambda_j}$, and the center is x. The elliptical parameter, namely, the width, height, and the orientation of the target can be well estimated by the singular value decomposition (SVD) of bandwidth matrix H. Because H is symmetric positive definite, it is decomposed as follows:

$$H=R\Lambda R^{\mathrm{T}} \quad (21)$$

Where

$$R=\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix} \quad (22)$$

R is the eigenvector matrix; θ is defined as the orientation of the ellipse which is the angle between the major axis and the horizontal axis, as in Fig.1.

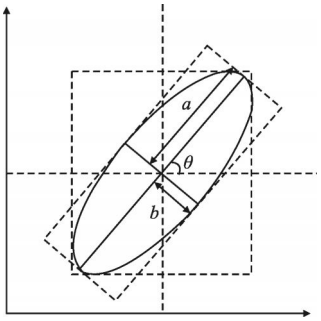


Fig.1 Parameters of an ellipse

$$\Lambda=\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix} \quad (23)$$

Λ is the eigenvalue matrix the major axis of th e ellipse corresponds to the maximum eigenvalue, the minor axis corresponds to the minimum eigenvalue, as in Fig.1.

$$a=\sqrt{\lambda_1} \quad (24)$$

$$b=\sqrt{\lambda_2} \quad (25)$$

2.2 Target window and local background region

The target window is initialized manually with an ellipse to determine the target center and bandwidth matrix simultaneously. As is shown is Fig.2, an ellipse can provide more precise shape in for mation than a rectangle bounding box, which is tighter for the target contour and defines the orientation of the target region.

Fig.2 Ellipse and bounding box

The "center-surround" approach has been used to sample pixels from the target and the background[12]. That is, a compact set of pixels composed by an ellipse covering the target is chosen to represent the target pixels, while a larger annulus of neighboring pixels surrounding the ellipse is chosen to represent

the background, whose center is the same as the target region. If the bandwidth of the target region is H_T, then the local background region will be defined by bandwidth

$$H_{LB}=k*H_T \tag{26}$$

Fig.3 Target and local background

2.3 Weight image and multiple features fusion

The weight ω_i calculated in Eq.(7) can actually be considered as the likelihood of a pixel in the current frame being part of the tracked target. A likelihood map or weight image where the value at a pixel is proportional to the likelihood that the pixel comes from the target we are tracking can be computed through weight ω_i.

The intensity and local standard deviation are employed to label pixels with the likelihood that they belong to the target, and the sum rule has been employed to produce the fused weight image because of its excellent properties.

The fusion weight ω_i is obtained by

$$\omega_i=k_1\omega_{1i}+k_2\omega_{2i} \tag{27}$$

Where k_1 and k_2 are the weight coefficients for intensity weight ω_{1i} and local standard deviation weight ω_{2i} respectively. An adaptively weight coefficients updating is achieved by the Bhattacharyya distance between target and local background model. The target and local background model are represented by their original one dimensional histograms without kernel weight, I and J, respectively.

The first row of Fig.4 is intensity histograms of target and local background, and the second row is the local standard deviation histogram. The weight coefficients k_1 and k_2 are calculated as following:

$$k_1=\frac{d[I_{int},J_{int}]}{d[I_{int},J_{int}]+d[I_{LStd},J_{LStd}]} \tag{28}$$

$$k_2=\frac{d[I_{LStd},J_{LStd}]}{d[I_{int},J_{int}]+d[I_{LStd},J_{LStd}]} \tag{29}$$

Where $[I_{int}, I_{LStd}, J_{int}, J_{LStd}]$ are the intensity and local standard deviation one dimensional histograms of target and local background respectively, $d[I_{int},J_{int}]$ and $d[I_{LStd},J_{LStd}]$ are the intensity and the local standard deviation Bhattacharyya distances between target and local background calculated according to Eq. (4) respectively.

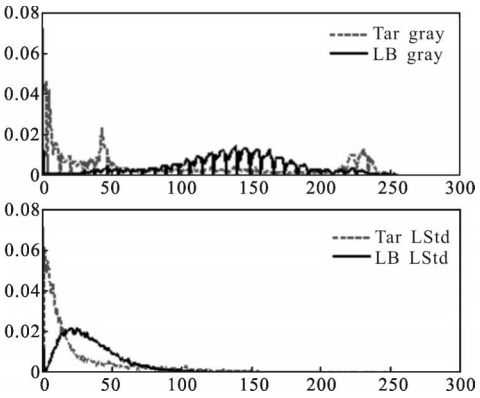


Fig.4 Gray and LStd hist

3 Target window updating and complete tracking algorithm

Target window updating includes target template model updating, target location vector updating and bandwidth matrix updating.

3.1 Target template model updating

To update target template model adaptively while preventing it from being polluted by background feature, two measures have been taken in this paper. First, the target template is updated in a fixed length frame interval n, which can be 3–5. Second, the newly updated target template model is generated by the combination of the previous target template model and the mean value of the determined target models in the previous n frames. Let ξ_i be the similarity based on Bhattacharyya coefficient between the determined target model and the target template model in the current frame,so

$$\xi_t=\sum_{u=1}^{m}\sqrt{\hat{p}_t^u(y;H)\hat{q}_t^u} \tag{30}$$

The target template model is updated for next frame by the following equation:

$$\hat{q}_{t+1}=\begin{cases}\hat{q}_t, & \xi<=0.7\\ \xi*\hat{q}_t+(1-\xi)*q, & 0.7<\xi<=0.9\\ \xi*q+(1-\xi)*\hat{q}_t, & \xi>0.9\end{cases} \tag{31}$$

Where

$$\xi=(\xi_k+\xi_{k-1}+\cdots+\xi_{k-n+1})/n \tag{32}$$

$$q=(\hat{q}_t+\hat{q}_{t-1}+\cdots+\hat{q}_{t-n+1})/n \tag{33}$$

3.2 Location vector updating

Peng have found that the changes of target scale and position within the region of previous tracking window would not impact the location accuracy of mean shift tracker[13]. Therefore, to insure the location accuracy of target center, the tracking window, i.e. the bandwidth matrix should be enlarged appropriately enough to cover the target inside the elliptical tracking window in despite that the actual target scale increases or decreases, and even the target rotates arbitrarily. In this paper, the enlargement ratio is 1.44. Assuming the previous bandwidth matrix of target template model is H_0, and then the bandwidth matrix H_{En} used in the iterative expression of location vector Eq.(10) for current frame is given by

$$H_{\mathrm{En}}=1.44*H_0 \tag{34}$$

3.3 Full bandwidth matrix updating

The bandwidth matrix will be inclined to explode in the presence of background clutter; and it will be inclined to implode on self-similar target. To make the bandwidth matrix suitable for tracking, regularization terms should be introduced[14].

Being different from location vector, the initial value in iteration expression of bandwidth matrix in Eq.(18) is the original previous bandwidth. To control the iteration of bandwidth more precisely, we do not use common the matrix norm as the stop criterion of bandwidth iteration, but analyze the variations of eigenvalues and eigenvector, which correspond to the elliptical major axis, the minor axis and the orientation.

4 Experiments results and tracker evaluation

4.1 Experiments setup

The proposed algorithm BAMS is compared with other scale adaptation target tracking algorithms originating from standard Mean-shift algorithm with available source code, including SOAMST algorithm, scale plus- minus Mean shift algorithm ($\mathrm{MS}_{\pm}$), and EM algorithm, and the parameters of the three compared tracking algorithm are left default as set by the authors.

4.2 Evaluation measures

In this paper, we adopt four evaluation measures proposed by Maggio [15], such as Euclidian center location error, normalized center location error, dice error, and lost track ratio.

Euclidian center location error. The Euclidian distance between the center location vector y of the tracked targets and the manually labeled ground truths $\tilde{y}$ is used to evaluate the center location error.

$$d_t(y,\tilde{y})=\sqrt{(y-\tilde{y})'(y-\tilde{y})}=\sqrt{\sum_{i=1}^{2}(y_i-\tilde{y}_i)^2} \tag{35}$$

Normalized center location error. Assuming the state s of an elliptical tracker

$$s=(u,v,a,b,\theta) \tag{36}$$

is represented by the center (u,v), the length of the two semi-axes (a,b) and the rotation of the estimated elliptical target area θ. Also, assuming the ground-truth information

$$\tilde{s}=(\tilde{u},\tilde{v},\tilde{a},\tilde{b},\tilde{\theta}) \tag{37}$$

on the expected position of the ellipse is available. The center location error can be normalized by rescaling and rotating the coordinate system so that the errors in u and v become

$$e_u(s,\tilde{s})=\frac{\cos\tilde{\theta}(u-\tilde{u})-\sin\tilde{\theta}(v-\tilde{v})}{\tilde{a}} \tag{38}$$

$$e_v(s,\tilde{s})=\frac{\sin\tilde{\theta}(u-\tilde{u})+\cos\tilde{\theta}(v-\tilde{v})}{\tilde{b}} \tag{39}$$

The normalized center location error at frame t is

$$e_t(s_t,\tilde{s}_t)=\sqrt{e_u(s_t,\tilde{s}_t)^2+e_v(s_t,\tilde{s}_t)^2} \tag{40}$$

Note that, when the estimated center is outside the ground-truth area, then $e_t(s_t,\tilde{s}_t)>1$.

Dice error. A measure of the match between the set of pixels A_t and $\tilde{A}_t$, defined by the estimated and ground-truth ellipse in each frame t, is called dice error.

$$D_t=1-\frac{2|A_t\cap\tilde{A}_t|}{|A_t|+|\tilde{A}_t|} \tag{41}$$

Where $|\cdot|$ denotes the cardinality of a set.

Lost track ratio. The lost track ratio η is the ratio between the number of frames where the tracker is not successful N_L and the total number of frames in a test sequence N_T:

$$\eta=N_L/N_T \tag{42}$$

A track at time index t can be defined lost when the dice error D_t in that frame exceeds a certain value T_η:

$$D_t>T_\eta \tag{43}$$

T_η =0.8 is adopted in this paper.

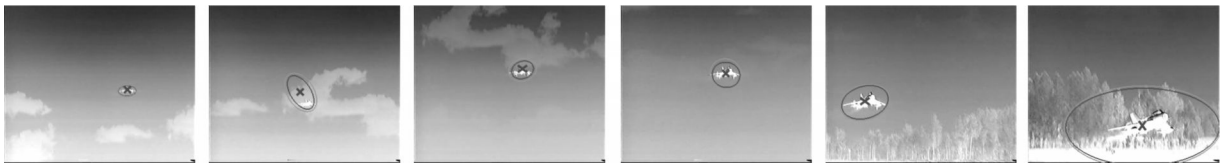
4.3 Tracking results and performance evaluation

Due to space limitations, the visual tracking results of only the proposed BAMS algorithm and SOAMST algorithm for the flying jet plane infrared sequences are shown in Fig.5. The tracking window of the proposed BAMS algorithm is shown as red ellipse, and the center is represented as red octagon. The tracking window of SOAMST algorithm is shown as cyan ellipse, the external green ellipse is the search window, and the center is represented as red crosses.

The results of evaluation measures, such as Euclidian center location error d, normalized center location error e_t, dice error D_t, are shown in Fig.6(b)−(d). In these figures, only the evaluation measures of trackers which can contain the target inside the tracking window from the first frame to the end frame through out the current infrared sequence are plotted. The intensity and local standard deviation Bhattacharyya distances between target and local background annotated manually are shown in Fig.6 (a), which show the complexity variations of the two infrared sequences.

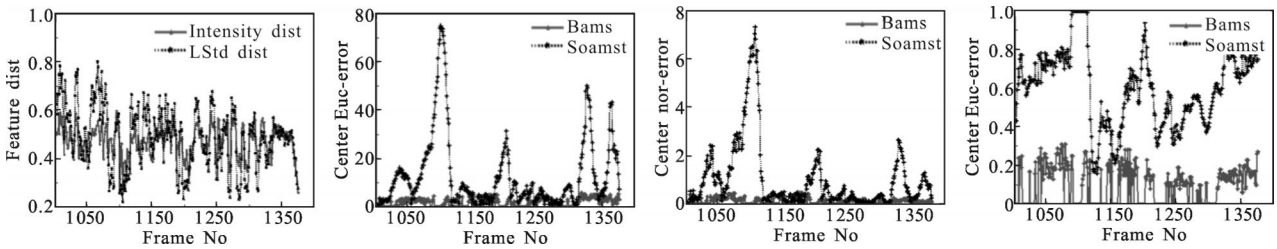
The results of the final performance vector $(\bar{\eta},\bar{d},\bar{e},\bar{D})$

(a) Proposed BAMS algorithm



(b) SOAMST algorithm

Fig.5 Visual tracking results for flying jet plane infrared sequence



(a) Feature distance (b) Center Euc-error (c) Center Nor-error (d) Dice error

Fig.6 Feature distance variations and tracking errors for flying jet plane infrared sequence

are listed in Tab.1. The values of lost track η are compared first. If the difference of the lost track value η between other tracking algorithm and the proposed algorithm is higher than 0.1, then the other evaluation measures $(\bar{d},\bar{e},\bar{D})$ of this algorithm are not listed in this table.

As shown in Fig.6 (b)–(d) and Tab.1, the proposed BAMS algorithm produces the least errors in terms of both location and target size for the flying jet plane infrared sequence.

Tab.1 Performance evaluation results of the four tracking algorithms

Algorithm	BAMS				SOAMST				$MS_{\pm}$				EM			
Sequences	η	$\bar{d}$	$\bar{e}$	$\bar{D}$	η	$\bar{d}$	$\bar{e}$	$\bar{D}$	η	$\bar{d}$	$\bar{e}$	$\bar{D}$	η	$\bar{d}$	$\bar{e}$	$\bar{D}$
Flying jet plane	0.00	1.74	0.11	0.10	0.11	–	–	–	0.91	–	–	–	0.77	–	–	–

4.4 Computation complexity

According to Comaniciu[5], the mean computation cost C_o of the location iteration is approximately given by

$$C_o=N(c_H+n_h c_S)\approx Nn_h c_S \tag{44}$$

Where N is the average iterations number per frame; c_H is the cost of the histogram; n_h is the number of target pixels, and c_S is the cost of an addition, a square root, and a division.

Because the compared tracking algorithms are all developed from the standard mean shift tracking algorithm, the mean computation costs of the location iteration of these algorithms for the same infrared sequence are dependent on the average number N of iterations per frame. In addition, because the Epanechnikov kernel function is employed in the iteration solution expression of bandwidth matrix of the proposed BAMS algorithm, as in Eq. (18), the mean computation cost of the bandwidth matrix iteration is approximate with that of the location iteration. Therefore, the total computation cost of the proposed BAMS algorithm is dependent on the total iterations number of location and bandwidth matrix.

Based on the aforementioned analysis, we compare the computation complexity of the proposed BAMS algorithm with that of other tracking algorithms by the total iterations number per frame, as shown in Fig.7. Tab. 2 shows the average iterations numbers of the four compared tracking algorithms, and the average is done over all frames in each infrared sequence. In the implementations of scale plus-minus Mean shift algorithm and EM algorithm by Zivkovic[6], the iterations iterations numbers are fixed as 6.

As shown in Fig.7, the total iterations number per frame of the proposed BAMS algorithm is smaller in general than that of other three tracking algorithms for the flying jet plane infrared sequences. As listed in Tab.2, the average interations numbers of the proposed BAMS algorithm are also the least among the four tracking algorithms.

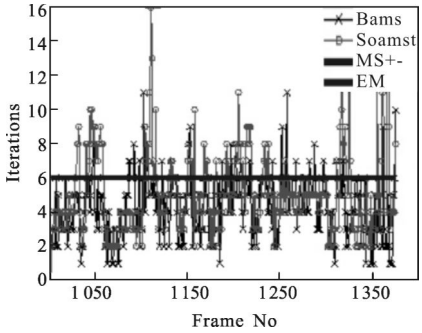


Fig.7 Iterations numbers of the four tracking algorithms

Tab.2 Aaverage iterations numbers of the four tracking algorithms

Algorithm sequences	Bams	Soamst	$MS_{\pm}$	EM
Flying jet plane	4.28	5.07	6.00	6.00

The above computation complexity analysis and comparisons between the proposed BAMS algorithm and other three tracking algorithms have proved the effectiveness of the proposed BAMS algorithm.

5 Conclusion

An adaptive multiple features fusion and

bandwidth updating mean shift tracking algorithm has been proposed. The iterative solution expressions of location and bandwidth have been deduced according to M-estimator. Fused weight image with intensity and local standard deviation has been produced. Historical information of the mean similarity between the determined target region and the target template model has been used to update the target template model. Regularization terms, based on the feature Bhattacharyya distances between target and local background, the variations of eigen-values and eigenvectors, have been introduced to prevent the bandwidth from exploding in the presence of background clutter or imploding on self-similar target.

Experiments on typical target scale changing infrared sequences have been carried out. Tracking performance evaluation measures appropriate for infrared sequences in which the target´s scale changes continuously have been introduced to compare the tracking results of the proposed algorithm with other scale adaptation mean shift tracking algorithms. The visual results and evaluation measures show that the proposed algorithm produces the least errors for the typically challenging infrared sequences.

References:

[1] Qiao Liyong, Xu Lixin, Gao Min. Influences of infrared image complexity on the target detection performance [J]. ***Infrared and Laser Engineering***, 2013, 42: 253-261.(in Chinese)

[2] Qiao Liyong, Xu Lixin, Gao Min.Fast maximum entropy thresholding based on two-dimensional histogram oblique segmentation in infrared imaging guidance [J]. ***Infrared and Laser Engineering***, 2013, 42(7): 1691-1699.(in Chinese)

[3] Qiao Liyong, Xu Lixin, Gao Min. Survey of Image Complexity Metrics for Infrared Target Recognition [J]. ***Infrared Technology***, 2013, 35(2): 88-96.(in Chinese)

[4] Yilmaz A, Shafique K, Shah M. Target tracking in airborne forward looking infrared imagery [J]. ***Image and Vision Computing***, 2003, 21(7): 623-635.

[5] Comaniciu D, Ramesh V, Meer P. Kernel-Based Object Tracking [J]. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, 2003, 25(5): 564-575.

[6] Zivkovic Z, Krose B. An EM-like algorithm for color-histogram-based object tracking [C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004: 798-803.

[7] Ning J, Zhang L, Zhang D, et al. Scale and orientation adaptive mean shift tracking [J]. ***IEEE Computer Vision***, 2012, 1(6): 52-61.

[8] Webb A R, Copsey K D. Statistical Pattern Recognition[M]. 3rd ed, America: John Wiley & Sons Ltd,2011: 198.

[9] Huber P J. Robust estimation of a location parameter[J]. ***The Annals of Mathematical Statistics***, 1964, 35(1): 73-101.

[10] Maronna R A. Robust M-Estimators of multivariate location and scatter[J]. ***The Annals of Statistics***, 1976, 4(1): 51-67.

[11] Strang G. Linear Algebra and Its Applications [M]. 4th ed. Britain:Wellesley-Cambridge Press, 2005.

[12] Collins O T, Liu Y. Online selection of discriminative tracking features[J]. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, 2005, 27(10): 1631-1643.

[13] Peng N S, Yang J, Liu Z. Performance analysis for tracking of variable scale objects using mean-shift algorithm [J]. ***Optical Engineering***, 2005, 44(7): 70505.

[14] Vojir T, Noskova J, Matas J. Robust Scale-Adaptive Mean-Shift for Tracking [M].Berlin: Springer, 2013: 7944, 652-663.

[15] Maggio E, Cavallaro A. Accurate appearance-based Bayesian tracking for maneuvering targets [J]. ***Journal Computer Vision and Image Understanding***, 2009, 113(4): 544-555.