Adaptive Bandwidth Mean Shift Object Tracking

Xiaopeng Chen
Institute of Automation, Chinese Academy of Sciences
Beijing China
xpchen@hitic.ia.ac.cn

Youxue Zhou
Faculty of Information Science and Technology
Jiujiang University
Jiujiang China
Xiaoliu161@tom.com

Xiaosan Huang
Institute of Software, Chinese Academy of Sciences
Beijing China
xiaosan@nfschina.com

Chengrong Li
Institute of Automation, Chinese Academy of Sciences
Beijing China
lich@hitic.ia.ac.cn

***Abstract*—In this paper, a novel adaptive bandwidth mean shift algorithm toward 2D object tracking is proposed. It can simultaneously tracks the scale and orientation besides position in real time. The feature histogram weighted by a kernel with adaptive bandwidth is used for representing the target and the candidate target. The similarity of the target model and the candidate model is measured by the Bhattacharyya coefficient. A two step method is used iteratively to find the most probable target position, scale and orientation. The first step is to find the position using a mean shift iteration, the second step is to find the bandwidth which best describes the region of the object. Its convergence is proved theoretically. Experiments show that it can successfully track the position, scale and orientation in real time.**

***Keywords*—object tracking, mean shift, adaptive bandwidth, vision**

I. INTRODUCTION

THE mean shift based object tracking algorithms have achieved considerable success due to its simplicity and robustness. Many of them are based on Dorin Comaniciu's work [1][2]. The idea of kernel based tracking has been exploited and developed forward by various researchers. Paper [3] used a different measure: K-L distance and [4] used sample based similarity measure. It is easy to prove that CAMSHIFT [5] is a kind of mean shift tracking algorithm with different weight for each pixel to seek the maximum of similarity. Some of them selected different feature spaces. Dorin Comaniciu extended the mean shift tracking approach using a joint spatial-color histogram [6]. A new spatial-color histogram based mean shift object tracking method was proposed in [7] for robustness in scale change. Stanley T. Birchfield [8] used spatiogram instead of color histogram to describe the model. A novel color model was proposed in [9] for single color tracking. Joint feature-spatial spaces can describe both rigid and non-rigid objects. Tracking in joint feature-spatial spaces with mean shift is proposed in [10][11]. Adaptive feature selection for mean shift tracking is proposed in [12]. Some authors updated the target model dynamically. The target model is updated according to the similarity changes [13] or using Kalman filter to keep track of the target model[14]. Many algorithms [2][15] also took background for consideration to improve tracking robustness. Many authors combine mean shift tracking with other tools. Kalman filter [2][16] or particle filter [17] were added to smooth the tracking trajectory. Mean shift algorithm was boosted using random sampling in [18]. Annealed mean shift is proposed in [19] for global mode seeking.

However, these algorithms lack the ability of simultaneously keep tracking scale, orientation and position. The method proposed in [2] suggests repeating the mean shift algorithm at each iteration using window sizes of plus or minus 10 percent of the current size for tracking scale, which was proved to be unfeasible and in [20]. Tracking through scale space [20] is also computationally too expensive and can not reach real time requirement. Paper [9] can track the scale efficiently, however, it can only be used to track single color object. CAMSHIFT [5] can track scale and orientation, but this algorithm can only be used to track single color object. In [21][22], EM-like algorithm can track 5-DOF of non-rigid object, however, it uses Gaussian kernel which may consumes lots of computation time. What's more, its convergence did not be proved. In [23] level set based asymmetric kernel was used to track scale and orientation. In [24], anisotropic kernel mean shift is used, however, the equations of (20)~(22) are proved wrong: obviously, the shape of the ellipse is only decided by Σ; and from (9), we can not get equation (10). What's more, (19) is computationally expensive.

In our paper, we extend the traditional mean shift to an adaptive bandwidth mean shift. And then we apply the adaptive mean shift algorithm with Epanechnikov profile and Gaussian profile for object tracking adopting the object tracking framework in [2]. The tracking algorithm is called ABMSOT. Besides the mean shift step for seeking the best position, a second step is added in each iteration for finding the best bandwidth matrix, which describes the shape of the candidate object. It can simultaneously tracking position, orientation and scale in real time. The tracking effectiveness and efficiency are verified by experiments.

The paper is organized like this: in section II, we extends the mean shift algorithm presented in [25] to general case. In section III, adaptive bandwidth mean shift in 2D case is analysed. In section IV, based on the work of [21][22][24], we extends the algorithm proposed in [2] using adaptive bandwidth mean shift in 2D case. In section V, experiments are given in detail. Section VI draws the conclusion and we

978-1-4244-1676-9/08 /$25.00 ©2008 IEEE

discuss the future work in section VII.

II. ADAPTIVE BANDWIDTH MEAN SHIFT

A d-variate general kernel for multivariate kernel density estimator is

$$K_{\boldsymbol{H}}(\boldsymbol{x}) = \left|\boldsymbol{H}\right|^{-\frac{1}{2}} K(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}^{-1}\boldsymbol{x}) \tag{1}$$

Where $\boldsymbol{H}$ is a symmetric matrix. Profile $K(x):[0,\infty]\to R$ satisfies these conditions: nonnegative, nonincreasing and piecewise continuous [25].

The weighted density estimate using d-variate general kernel is defined like this: $q(\boldsymbol{x}) = \sum_{\boldsymbol{s}\in S} K_{\boldsymbol{H}}(\boldsymbol{x}-\boldsymbol{s})w(\boldsymbol{s})$ (2)

where $S\subset X$ is a finite set that represents the search space, $w: S\to(0,\infty)$ is a weight function.

The sample mean shift kernel is defined:

$$G_{\boldsymbol{H}}(\boldsymbol{x}) = \left|\boldsymbol{H}\right|^{-\frac{1}{2}} G(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}^{-1}\boldsymbol{x}) \tag{3}$$

where $G(x) = -K'(x)$ (4)

The mean shift vector is defined:

$$\boldsymbol{m}(\boldsymbol{x}) - \boldsymbol{x} = \frac{\sum_{\boldsymbol{s}\in S} G_{\boldsymbol{H}}(\boldsymbol{x}-\boldsymbol{s})w(\boldsymbol{s})\boldsymbol{s}}{\sum_{\boldsymbol{s}\in S} G_{\boldsymbol{H}}(\boldsymbol{x}-\boldsymbol{s})w(\boldsymbol{s})} - \boldsymbol{x} \tag{5}$$

THEOREM 1 the cosine of the angle between the mean shift vector defined in (5) and the gradient of the weighted density estimate defined in (2) is strictly positive if and only if the bandwidth matrix $\boldsymbol{H}$ is positive definite.

PROOF. The gradient of (2) is

$$\nabla q(\boldsymbol{x}) = \sum_{\boldsymbol{s}\in S} \nabla K_{\boldsymbol{H}}(\boldsymbol{x}-\boldsymbol{s})w(\boldsymbol{s}) \tag{6}$$

as $\nabla K_{\boldsymbol{H}}(\boldsymbol{x}-\boldsymbol{s}) = 2\left|\boldsymbol{H}\right|^{-\frac{1}{2}} K'((\boldsymbol{x}-\boldsymbol{s})^{\mathrm{T}}\boldsymbol{H}^{-1}(\boldsymbol{x}-\boldsymbol{s}))\boldsymbol{H}^{-1}(\boldsymbol{x}-\boldsymbol{s})$ (7)

we can get $\nabla q(\boldsymbol{x}) = -2\boldsymbol{H}^{-1}\sum_{\boldsymbol{s}\in S} G_{\boldsymbol{H}}(\boldsymbol{x}-\boldsymbol{s})w(\boldsymbol{s})(\boldsymbol{x}-\boldsymbol{s})$ (8)

What's more, equation (5) can be rewritten as

$$\boldsymbol{m}(\boldsymbol{x}) - \boldsymbol{x} = -\frac{\sum_{\boldsymbol{s}\in S} G_{\boldsymbol{H}}(\boldsymbol{x}-\boldsymbol{s})w(\boldsymbol{s})(\boldsymbol{x}-\boldsymbol{s})}{p(\boldsymbol{x})} \tag{9}$$

where $p(\boldsymbol{x}) = \sum_{\boldsymbol{s}\in S} G_{\boldsymbol{H}}(\boldsymbol{x}-\boldsymbol{s})w(\boldsymbol{s})$ (10)

is the density estimate using $G_{\boldsymbol{H}}$.Compare (9) with (8), it is not difficult to find that the relationship between the mean shift vector (5) and the gradient of (2) is

$$\boldsymbol{m}(\boldsymbol{x}) - \boldsymbol{x} = \frac{\boldsymbol{H}\nabla q(\boldsymbol{x})}{2p(\boldsymbol{x})} \tag{11}$$

The cosine of the angle between the two vectors is

$$\frac{<\boldsymbol{m}(\boldsymbol{x})-\boldsymbol{x}, \nabla q(\boldsymbol{x})>}{\left\|\boldsymbol{m}(\boldsymbol{x})-\boldsymbol{x}\right\|\bullet\left\|\nabla q(\boldsymbol{x})\right\|} = \frac{\nabla q(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{H}\nabla q(\boldsymbol{x})}{\left|\det(\boldsymbol{H})\right|\bullet\left\|\nabla q(\boldsymbol{x})\right\|^2} \tag{12}$$

From (12), it is obvious that if and only if $\boldsymbol{H}$ is positive definite, the cosine of the angle between two vectors is positive.

The dist $(\boldsymbol{x}-\boldsymbol{s})^{\mathrm{T}}\boldsymbol{H}^{-1}(\boldsymbol{x}-\boldsymbol{s})$ is called Mahalanobis dist. When $\boldsymbol{H}$ is a positive definite symmetric matrix, $S = \left\{\boldsymbol{s}\middle|(\boldsymbol{x}-\boldsymbol{s})^{\mathrm{T}}\boldsymbol{H}^{-1}(\boldsymbol{x}-\boldsymbol{s}) < r^d\right\}$ defines a super elliptic ball in d-variate space where to search the samples. The bandwidth matrix $\boldsymbol{H}$ is optimized for maximizing $q(\boldsymbol{x})$, these equations must be satisfied:

$$\begin{cases} \dfrac{\partial q(\boldsymbol{x})}{\partial \boldsymbol{H}} = 0 \\ \dfrac{\partial^2 q(\boldsymbol{x})}{\partial \boldsymbol{H}^2} < 0 \end{cases} \text{ or } \begin{cases} \dfrac{\partial q(\boldsymbol{x})}{\partial (\boldsymbol{H}^{-1})} = 0 \\ \dfrac{\partial^2 q(\boldsymbol{x})}{\partial (\boldsymbol{H}^{-1})^2} < 0 \end{cases} \tag{13}$$

Then we can draw the adaptive bandwidth mean shift algorithm for seeking the mode of (2).

1) Initialize the center point $\boldsymbol{x}_0$ and bandwidth matrix $\boldsymbol{H}_0$, S_0 is calculated according to $\boldsymbol{H}_0$ and $\boldsymbol{x}_0$.

2) Use (5) to shift the position $\boldsymbol{x}_0$ to $\boldsymbol{x}_1$ ($\boldsymbol{x}_1 = \boldsymbol{m}(\boldsymbol{x}_0)$), recalculate S_0' according to $\boldsymbol{H}_0$ and $\boldsymbol{x}_1$.

3) Search the maximum of $\boldsymbol{H}_1$ according to (13), x_1 and calculate S_1 according to $\boldsymbol{H}_1$ and $\boldsymbol{x}_1$.

4) If $\left\|\boldsymbol{x}_0 - \boldsymbol{x}_1\right\| < \varepsilon$ and $S_0 = S_1$, stop. Else $\boldsymbol{x}_0 \leftarrow \boldsymbol{x}_1$, $S_0 \leftarrow S_1$, $\boldsymbol{H}_0 \leftarrow \boldsymbol{H}_1$, goto 2.

III. ADAPTIVE BANDWIDTH MEAN SHIFT IN 2D CASE

A. 2D bandwidth matrix and ellipse

We are only interested in the positive definite bandwidth matrix for estimating 2-variate kernel density. Given any positive definite matrix $\boldsymbol{H}$, there is a real unitary matrix $\boldsymbol{U}$, such that $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{U} = \boldsymbol{D}$,where $\boldsymbol{D}$ is a diagonal matrix. Without loss of generality, we can rewritten $\boldsymbol{H}$ as [9]

$$\boldsymbol{H} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} \tag{14}$$

where $\boldsymbol{A} = \boldsymbol{R}(\phi)\boldsymbol{Diag}(a,b)$ (15)

$$\boldsymbol{R}(\phi) = \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix} \tag{16}$$

Given a matrix equation $\boldsymbol{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{bmatrix} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ (17)

we get two solutions of the parameters of the ellipse.

$$\begin{cases} a = \sqrt{\dfrac{1}{2}(h_{11}+h_{22}+\sqrt{4h_{12}^2+(h_{11}-h_{22})^2})} \\ b = \sqrt{\dfrac{1}{2}(h_{11}+h_{22}-\sqrt{4h_{12}^2+(h_{11}-h_{22})^2})} \\ \phi = \dfrac{1}{2}\operatorname{atan}2(2h_{12}, h_{11}-h_{22}) \end{cases} \tag{18}$$

and

$$\begin{cases} a = \sqrt{\frac{1}{2}(h_{11}+h_{22}-\sqrt{4h_{12}^2+(h_{11}-h_{22})^2})} \\ b = \sqrt{\frac{1}{2}(h_{11}+h_{22}+\sqrt{4h_{12}^2+(h_{11}-h_{22})^2})} \\ \phi = \frac{\pi}{2}+\frac{1}{2}\operatorname{atan}2(2h_{12}, h_{11}-h_{22}) \end{cases} \quad (19)$$

And set $$S = \left\{ \boldsymbol{s} \middle| (\boldsymbol{x}-\boldsymbol{s})^{\mathrm{T}} \boldsymbol{H}^{-1} (\boldsymbol{x}-\boldsymbol{s}) < \sigma^2 \right\} \quad (20)$$

defines an ellipse with its first axis length σa and second axis length σb, and a rotate angle ϕ from x axis. σ is a factor deci ded by the kernel function, $\boldsymbol{x}$ is the center of the ellipse. See fi gure 1 for details.

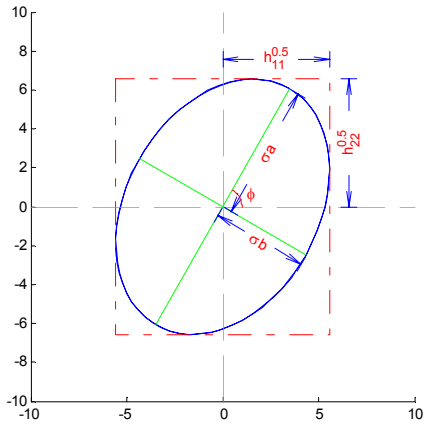

Figure 1. A rotated ellipse defined by equation (20).

If $\boldsymbol{H}$ or σ varies, the region S defined by (20) varies. σ is chosen according to the kernel profile. When Epanechnikov profile

$$K(x) = \begin{cases} c(1-x) & if \quad 0 \le x < 1 \\ 0 & if \quad x > 1 \end{cases} \quad (21)$$

is used, we choose $\sigma = 1$ for convenience. When Gaussian profile

$$K(x) = c \exp(-\frac{x}{2}) \quad (22)$$

is used, $\sigma = 2.5$ is used. Once σ is selected, $\boldsymbol{H}$ decides the orientation and scale of the ellipse region. But how to update the bandwidth matrix $\boldsymbol{H}$?

B. *Optimal bandwidth selection problem*

Our objective is to find $\boldsymbol{x}$ and $\boldsymbol{H}$ where $q(\boldsymbol{x})$ in (2) reaches its maximum. When $\boldsymbol{H}$ is held, we use (5) to update $\boldsymbol{x}$; we can also hold $\boldsymbol{x}$, and find $\boldsymbol{H}$ or $\boldsymbol{H}^{-1}$ to maximize $q(\boldsymbol{x})$. In 2-variate case, we rewrite $$\boldsymbol{H}^{-1} = \begin{bmatrix} \hat{h}_{11} & \hat{h}_{12} \\ \hat{h}_{12} & \hat{h}_{22} \end{bmatrix} \quad (23)$$

if

$$\frac{\partial q(\boldsymbol{x})}{\partial \hat{h}_{ij}} = 0 \quad (24)$$

and

$$\frac{\partial^2 q(\boldsymbol{x})}{\partial \hat{h}_{ij}^{\ 2}} < 0 \quad (25)$$

are satisfied, the solution (24) $\boldsymbol{H}$ or $\boldsymbol{H}^{-1}$ will maximize $q(\boldsymbol{x})$.

When Epanechnikov profile is used, after some complicated calculations we get the solution of $\boldsymbol{H}$ using Epanechnikov profile :

$$\boldsymbol{H}_{\mathrm{e}} = \frac{4\sum_{\boldsymbol{s}\in S} w(\boldsymbol{s})(\boldsymbol{x}-\boldsymbol{s})(\boldsymbol{x}-\boldsymbol{s})^T}{\sum_{\boldsymbol{s}\in S} w(\boldsymbol{s})} \quad (26)$$

The second differential is verified to be minus. So $q(\boldsymbol{x})$ will reach its maximum at $\boldsymbol{H}_e$.

When Gaussian profile is used, $\ln q(\boldsymbol{x})$ is calculated instead.

$$\ln q(\boldsymbol{x}) = \ln \sum_{\boldsymbol{s}\in S} c \left|\boldsymbol{H}\right|^{-\frac{1}{2}} \exp(-\frac{(\boldsymbol{x}-\boldsymbol{s})^{\mathrm{T}} \boldsymbol{H}^{-1} (\boldsymbol{x}-\boldsymbol{s})}{2}) \quad (27)$$

By Jensen's inequality [9], we get the lower bound of $\ln q(\boldsymbol{x})$:

$$\ln q(\boldsymbol{x}) \ge L = \sum_{\boldsymbol{s}\in S} w(\boldsymbol{s})(\ln c + \frac{1}{2}\ln\left|\boldsymbol{H}\right|^{-1} - \frac{(\boldsymbol{x}-\boldsymbol{s})^{\mathrm{T}} \boldsymbol{H}^{-1} (\boldsymbol{x}-\boldsymbol{s})}{2}) \quad (28)$$

if substitute $q(\boldsymbol{x})$ with L in (24) and (25) we can get the solutions of $\boldsymbol{H}_g$:

$$\boldsymbol{H}_{\mathrm{g}} = \frac{\sum_{\boldsymbol{s}\in S} w(\boldsymbol{s})(\boldsymbol{x}-\boldsymbol{s})(\boldsymbol{x}-\boldsymbol{s})^{\mathrm{T}}}{\sum_{\boldsymbol{s}\in S} w(\boldsymbol{s})} \quad (29)$$

For the solution of Gaussian profile or Epanechnikov profile, $\forall \boldsymbol{y} \ne 0$,

$$\boldsymbol{y}^{\mathrm{T}} \boldsymbol{H} \boldsymbol{y} = \frac{\sum_{\boldsymbol{s}\in S} w(\boldsymbol{s}) \boldsymbol{y}^{\mathrm{T}} (\boldsymbol{x}-\boldsymbol{s})(\boldsymbol{x}-\boldsymbol{s})^{\mathrm{T}} \boldsymbol{y}}{\sum_{\boldsymbol{s}\in S} w(\boldsymbol{s})} = \frac{\sum_{\boldsymbol{s}\in S} w(\boldsymbol{s}) \gamma(\boldsymbol{s})^2}{\sum_{\boldsymbol{s}\in S} w(\boldsymbol{s})} \quad (30)$$

Because $\exists \boldsymbol{s}, \gamma(\boldsymbol{s}) = \boldsymbol{y}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{s}) \ne 0$, so $\boldsymbol{H}$ in expression (26)(29) is positive definite. From THEOROM1 the mean shift vector will point to the direction where $q(\boldsymbol{x})$ climb to its local maximum.

IV. Tracking using adaptive bandwidth meanshift

A. *Target model*

The target model is represented by its probability distribution function q in the feature space, the same as the definition given in [2]. Given a target with initial region $\hat{S}_0$ (ellipse) centered at position $\hat{\boldsymbol{x}}_0$, and we set $\sigma_g = 2.5$ when Gaussian profile is used or $\sigma_e = 1$ if Epanechnikov profile is used, and the initial bandwidth matrix $\hat{\boldsymbol{H}}_0$ can be initialized from expression (14)(15), where $\hat{a}_0 = \frac{x\_axis\_len}{\sigma}$, $\hat{b}_0 = \frac{y\_axis\_len}{\sigma}$. $\hat{\phi}_0$ is the rotating angle of the ellipse

$\hat{S}_0$ .Usually $\hat{\phi}_0 = 0$. The feature histogram is weighted according to the Mahalanobis distance from point $\boldsymbol{s}$ to the region center $\hat{\boldsymbol{x}}_0$ by matrix $\hat{\boldsymbol{H}}_0$, The function $b: R^2 \to \{1...M\}$ associates to the pixel at location $\boldsymbol{s}$ of its bin in the feature space [2]. The probability of the feature index $u = 1...M$ of the target model is computed as

$$\hat{q}_u = C\sum_{\boldsymbol{s}\in\hat{S}_0} \left|\hat{\boldsymbol{H}}_0\right|^{-\frac{1}{2}} K((\hat{\boldsymbol{x}}_0 - \boldsymbol{s})^{\mathrm{T}} \hat{\boldsymbol{H}}_0^{-1}(\hat{\boldsymbol{x}}_0 - \boldsymbol{s}))\delta[b(\boldsymbol{s}) - u] \quad (31)$$

Where $K(\bullet)$ is an Epanechnikov profile (see (21)) or a Gaussian profile (see (22)).

as $\left|\hat{\boldsymbol{H}}_0\right|^{-\frac{1}{2}}$ is invariant to $\boldsymbol{s}$, (31) can be rewritten as

$$\hat{q}_u = C\sum_{\boldsymbol{s}\in\hat{S}_0} K((\hat{\boldsymbol{x}}_0 - \boldsymbol{s})^{\mathrm{T}} \hat{\boldsymbol{H}}_0^{-1}(\hat{\boldsymbol{x}}_0 - \boldsymbol{s}))\delta[b(\boldsymbol{s}) - u] \quad (32)$$

where $C = \dfrac{1}{\sum_{\boldsymbol{s}\in\hat{S}_0} K((\hat{\boldsymbol{x}}_0 - \boldsymbol{s})^{\mathrm{T}} \hat{\boldsymbol{H}}_0^{-1}(\hat{\boldsymbol{x}}_0 - \boldsymbol{s}))}$ (33)

is the normalization constant.

B. Target candidates

With the position $\boldsymbol{y}$ and bandwidth matrix $\boldsymbol{H}$ given in previous frame, the point set S is defined by (20), and the probability distribution

$$\hat{p}_u(\boldsymbol{y}) = C_{\boldsymbol{H}} \sum_{\boldsymbol{s}\in S_{\boldsymbol{H}}} \left|\boldsymbol{H}\right|^{-\frac{1}{2}} K((\boldsymbol{y} - \boldsymbol{s})^{\mathrm{T}} \boldsymbol{H}^{-1}(\boldsymbol{y} - \boldsymbol{s}))\delta[b(\boldsymbol{s}) - u] \quad (34)$$

where $C_{\boldsymbol{H}} = \dfrac{1}{\sum_{\boldsymbol{s}\in S_{\boldsymbol{H}}} \left|\boldsymbol{H}\right|^{-\frac{1}{2}} K((\boldsymbol{y} - \boldsymbol{s})^{\mathrm{T}} \boldsymbol{H}^{-1}(\boldsymbol{y} - \boldsymbol{s}))}$ (35)

is the normalizeation constant which is not depend on y .

C. Bhattacharyya Coefficient

The similarity of the target and the candidate is measured by Bhattacharyya coefficient

$$\rho(\hat{p}(\boldsymbol{y}), \hat{q}) = \sum_{u=1}^{M} \sqrt{\hat{p}_u(\boldsymbol{y})\hat{q}_u} \quad (36)$$

Using Taylor expansion (36) can be the form

$$\rho(\hat{p}(\boldsymbol{y}), \hat{q}) \approx \frac{1}{2}\sum_{u=1}^{M} \sqrt{\hat{p}_u(\boldsymbol{y}_0)\hat{q}_u} + \frac{1}{2} C_{\boldsymbol{H}} \sum_{\boldsymbol{s}\in S_{\boldsymbol{H}}} w(\boldsymbol{s}) \left|\boldsymbol{H}\right|^{-\frac{1}{2}} K((\boldsymbol{y} - \boldsymbol{s})^{\mathrm{T}} \boldsymbol{H}^{-1}(\boldsymbol{y} - \boldsymbol{s})) \quad (37)$$

where $w(\boldsymbol{s}) = \sum_{u=1}^{M} \sqrt{\dfrac{\hat{q}_u}{\hat{p}_u(\boldsymbol{y}_0)}}\delta[b(\boldsymbol{s}) - u]$. (38)

So our objective is to find the maximum of the similarity between the target and candidate model. As the first part of (37) is not depend on y , our objective is to find the max of

$$q(\boldsymbol{y}) = \sum_{\boldsymbol{s}\in S_{\boldsymbol{H}}} w(\boldsymbol{s}) \left|\boldsymbol{H}\right|^{-\frac{1}{2}} K((\boldsymbol{y} - \boldsymbol{s})^{\mathrm{T}} \boldsymbol{H}^{-1}(\boldsymbol{y} - \boldsymbol{s})) \quad (39)$$

We do not care about $C_{\boldsymbol{H}}$ because it is just a normalization factor. When (39) reaches its maximum, so it is of (36). According to the statements in section II and III, a two step method can be iterated for seeking the mode of (36). In the first step, a mean shift iteration (5) is calculated once; second recenter the window and calculate the optimal bandwidth matrix according to (26) or (29), the bandwidth matrix then decides the search window for the next loop. The detail of the algorithm is written below.

D. Algorithm for adaptive mean shift tracking

As mentioned in last part, we use a two step method for seeking the mode of (39) for mean shift tracking. We write the algorithm step by step.

Init: Select an initial region $\hat{S}_0$ and its center $\hat{\boldsymbol{x}}_0$, calculate the initial bandwidth matrix $\hat{\boldsymbol{H}}_0$ according to (14)(15), and then calculate the target model $\{\hat{q}_u\}_{u=1...M}$ using (32).

For every frame in the image sequence, do:

1) Given the bandwidth matrix $\boldsymbol{H}_0$ and position $\boldsymbol{y}_0$ in the previous frame, the point set S_0 is defined by (20)

2) Derive the weights $\{w(\boldsymbol{s})\}_{\boldsymbol{s}\in S_0}$ according to (38).

3) Find the next location $\boldsymbol{y}_1$ of the target candidate according to $\boldsymbol{m}(\boldsymbol{x})$ in (5) :

$$\boldsymbol{y}_1 = \frac{\sum_{\boldsymbol{s}\in S_0} \boldsymbol{s} w(\boldsymbol{s}) G((\boldsymbol{y}_0 - \boldsymbol{s})^{\mathrm{T}} \boldsymbol{H}_0^{-1}(\boldsymbol{y}_0 - \boldsymbol{s}))}{\sum_{\boldsymbol{s}\in S_0} w(\boldsymbol{s}) G((\boldsymbol{y}_0 - \boldsymbol{s})^{\mathrm{T}} \boldsymbol{H}_0^{-1}(\boldsymbol{y}_0 - \boldsymbol{s}))} \quad (40)$$

the factor $\left|\boldsymbol{H}_0\right|^{-\frac{1}{2}}$ is eliminated in both denominator and numerator of (40) as it does not depend on s .

4) while $\rho[\hat{p}(\boldsymbol{y}_1), \hat{q}] < \rho[\hat{p}(\boldsymbol{y}_0), \hat{q}]$

do $\boldsymbol{y}_1 = \dfrac{\boldsymbol{y}_1 + \boldsymbol{y}_0}{2}$ recalculate $\rho[p(\boldsymbol{y}_1), \hat{q}]$

5) Move the window center to $\boldsymbol{y}_1$, update S_0 recalculate $\{w(\boldsymbol{s})\}_{\boldsymbol{s}\in S_0}$.

6) Update the bandwidth matrix according to (26) or (29):

$$\boldsymbol{H}_1 = \frac{\lambda \sum_{\boldsymbol{s}\in S_0} w(\boldsymbol{s})(\boldsymbol{y}_1 - \boldsymbol{s})(\boldsymbol{y}_1 - \boldsymbol{s})^{\mathrm{T}}}{\sum_{\boldsymbol{s}\in S_0} w(\boldsymbol{s})} \quad (41)$$

Where $\lambda = \lambda_e = 4$ or $\lambda = \lambda_g = 1$ according to the profile used.

7) recalculate point region S_1 according to (20).

8) if the points in S_1 and S_0 are the same, stop; Else $\boldsymbol{y}_0 \leftarrow \boldsymbol{y}_1, S_0 \leftarrow S_1, \boldsymbol{H}_0 \leftarrow \boldsymbol{H}_1$, goto 2).

V. EXPERIMENTS

In order to verify the effectiveness and efficiency of the ABMSOT algorithm, ABMSOT with Gaussian profile (ABMSOT-G) and ABMSOT with Epanechnikov (ABMSOT-E) profile are compared with KBOT presented in [2]. They are tested by an image sequence "Walk1" which comes from the EC Funded CAVIAR project IST200137540, URL:http://homepages.inf.ed.ac.uk/rbf/caviar/. The errors and

mean errors of position, scale and orientation are defined like this:

$$\begin{cases} e_x = x(t) - x_{gt}(t) \\ e_y = y(t) - y_{gt}(t) \\ e_o = o(t) - o_{gt}(t) \\ e_a = (a(t) - a_{gt}(t)) / a_{gt}(t) \\ e_b = (b(t) - b_{gt}(t)) / b_{gt}(t) \end{cases} \tag{42}$$

$$\begin{cases} \overline{e}_x = (\sum_t abs(x(t) - x_{gt}(t))) / frame\_count \\ \overline{e}_y = (\sum_t abs(y(t) - y_{gt}(t))) / frame\_count \\ \overline{e}_o = (\sum_t abs(o(t) - o_{gt}(t))) / frame\_count \\ \overline{e}_a = (\sum_t abs((a(t) - a_{gt}(t)) / a_{gt}(t))) / frame\_count \\ \overline{e}_b = (\sum_t abs((b(t) - b_{gt}(t)) / b_{gt}(t))) / frame\_count \end{cases} \tag{43}$$

Where $x(t)$, $y(t)$, $o(t)$, $a(t)$, $b(t)$ are the tracking data of x position, y position, orientation, a-axis of the elliptic shape, b-axis of the elliptic shape. Though a_{gt} and b_{gt} are not included in the ground truth data, they can be calculated from the ground truth data of bounding rectangle and the direction. It is not difficult to prove that h_{11} and h_{22} in (17) are half width and height of the bounding rectangle of the ellipse with axis a and b. As the ground truth orientation is given, a and b can be calculated from (17). However, the solution is singular when the orientation is 45, 135, 225, 315 deg. In this case, the older a and b in previous frame are used. Now that the angle range of orientation is 0~180, we can also extend the range to 0~360 .A sudden angle change of ± 180 is smoothed by add ∓ 180. The orientation and shape estimation of KBOT are not considered as the algorithm can not track the scale and orientation.

They are slightly different from the definitions presented in [22].

Three tracking algorithms search of color histogram in Luv space of size 16 * 16 * 16. They run on a PC which has a Intel R4 3.0GHz CPU and a memory of 512MB.We tracked the object "1" in "Walk1" sequence from frame 270 to frame 450. It is very difficult to track object "1" after frame 450 as the color of the object is almost the same as the background and the illumination is much weak. The initial value is the same as the ground truth data, except for the orientation. Some tracking pictures are shown in figure 2,3 and 4. The statistical results are shown in figure 5, 6, 7, and table 1. Ground truth data is in black; tracking result of ABMSOT-E is in red; tracking result of ABMSOT-G is in green; tracking result of KBOT is in blue. Legends of figure 6 and figure 7 are the same as figure 5 so they are not drawn on the canvas.

Figure 2. Frames 300, 330, 360, 390 of tracked image sequence using ABMSOT-E.

Figure 3. Frames 300, 330, 360, 390 of tracked image sequence using ABMSOT-G.

Figure 4. Frames 300, 330, 360, 390 of tracked sequence using KBOT

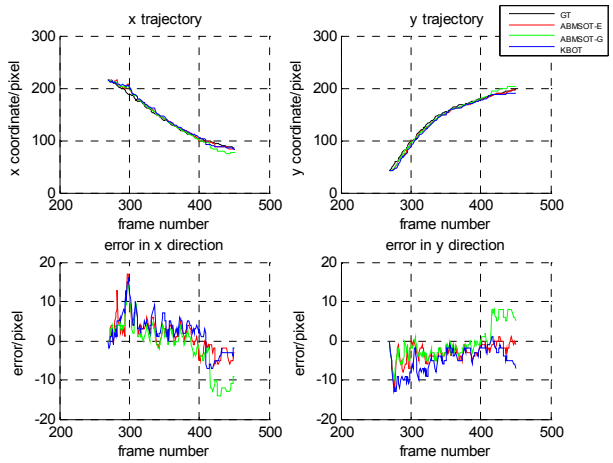

Figure 5. Position tracking trajectory and corresponding error

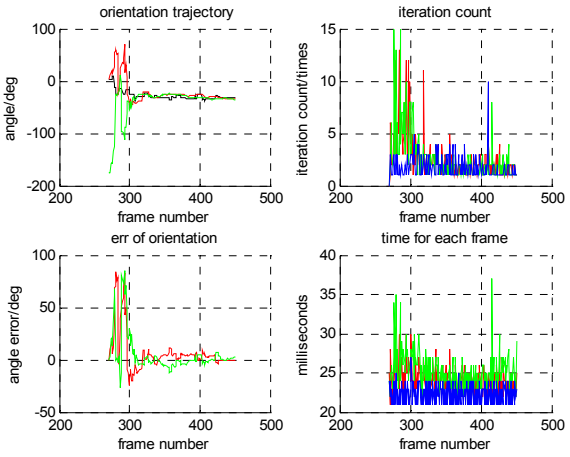

Figure 6. Orientation tracking trajectory and error, iteration count and consumed time for each frame

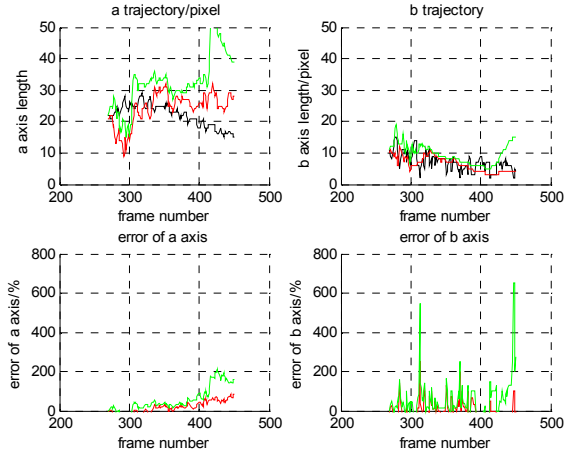

Figure 7. A-axis and error, b-axis and error

TABLE I. STATISTICAL MEAN DATA OF THREE ALGORITHMS

Mean Val/Unit	ABMSOT-E	ABMSOT-G	KBOT
Position x/pixel	3.4972	4.6630	4.3260
Position y/pixel	2.9503	3.4420	4.9613
Orientation/deg	9.9337	8.5193	None
Scale a/%	30.73	61.74	None
Scale b/%	34.05	63.93	None
Iteration Count	2.6851	2.6961	1.7348
Time/ms	24.1602	25.2597	22.3923

Figure 2, 3, 4 shows the exact tracking sequence. The boxes in red are actually calculated data. The boxes in green are ground truth data. Figure 5 shows the position tracking ability of three tracking algorithms. The tracked trajectories of three algorithms are much close to the ground truth data. The errors are shown in the second row. Obviously ABMSOT-E is better than the two other algorithms.

The first column of Figure 6 shows the orientation tracking results. At the beginning, the orientation is not stable, but after some image frames, the right orientation was found and kept. The cause of the phenomenon is that the initial position and size of the target object have great impact on the tracking results. The mean tracking errors of ABMSOT-G and ABMSOT-E are almost the same. After frame 300, orientations of objects were successfully tracked. The errors are not greater than 2 degrees.

The second column of Figure 6 shows the efficiency of the algorithms. ABMSOT-E and ABMSOT-G takes only slightly more time than KBOT. Mean iteration counts of three algorithms can be found in Table 1. Mean time consumed for search in each frame of ABMSOT-E, ABMSOT-G, KBOT is 24.1602ms, 25.2597ms, 22.3923ms. They all can track objects in real time.

Figure 7 shows the ability of tracking the shape of the object. As the calculated a and b values from bounding rectangle and orientation of the objects are not so accurate, the errors are somehow greater than other tracking properties. However, from the tracked sequence we can see that the tracking effect is not so bad. And the mean errors of them are not greater than 100%. ABMSOT-E is much better than ABMSOT-G.

In all, ABMSOT-E and ABMSOT-G can successfully track position, orientation and scale of objects in real time. ABMSOT-E is much better than ABMSOT-G.

VI. CONCLUSION

We first presented an adaptive bandwidth mean shift framework. We proved that if the bandwidth matrix is definite positive, mean shift vector points to the direction of climbing up to the local maximum of kernel density estimate function. The density kernel uses Mahalanobis dist instead of Euclidian dist so that the bandwidth matrix can be used to describe the shape of searching samples. If bandwidth matrix is definite positive, the shape is a super elliptic ball in d-variate space. In the adaptive bandwidth mean shift framework, the first step is

using mean shift to find the next position, the second step is finding the best bandwidth matrix. They are iterated until neither position nor bandwidth matrix changes. In 2D case, we found the best bandwidth matrixes for both Epanechnikov profile and Gaussian profile.

The efficiency of ABMSOT algorithm is validated through experiments. The result shows that it can successfully track the position, orientation and scale in real time. Through mathematical compare, we can conclude that the performance of ABMSOT-G is the same as the the EM-like algorithm presented in [21]. But ABMSOT-E is much better than ABMSOT-G. We suggest using ABMSOT-E for object tracking.

VII. FUTURE WORK

We use adaptive bandwidth mean shift for object tracking and have achieved considerable success. Adaptive bandwidth mean shift is a general search algorithm so it may also be used for clustering, segmentation, etc. Because it adjusts bandwidth matrix according to visited samples, we think it is more accurate than conditional mean shift.

Adaptive bandwidth mean shift object tracking algorithm is the extension of KBOT algorithm. So the way for improvements of KBOT can also be used to improve ABMSOT algorithm without much difficulty. For example, we can take background for consideration. We can use kalman filter to smooth the results and to improve the speed. We can use other features such as texture, edge and joint spatial-color feature. The target model can also be automatically adjusted to improve the robustness of the algorithm.

We should mention that the choice of σ has great impact on the ABMSOT algorithm. Small change of σ may result in much different results. We are going to find the reason and to find the way of automatically choosing σ.

The initial value is selected manually. It impacts the tracking results and so needs optimization.

[1] G Comaniciu, D., Ramesh, V. and Meer, P., "Real- Time Tracking of Non- Rigid Objects using Mean Shift," IEEE Computer Vision and Pattern Recognition, Vol II, 2000, pp. 142- 149.

[2] Comaniciu, D., Ramesh, V. and Meer, P., "Kernel- Based Object Tracking," IEEE Transactions on Pattern And Machine Intelligence, 25(5): 564- 577, May 2003.

[3] H. Chen and T. Liu, "Trust- region methods for real- time tracking," in Proc. 8th Intl. Conf. on Computer Vision, Vancouver, Canada, volume II, 2001, pp. 717- 722.

[4] CJ Yang, R Duraiswami and L Davis, "Efficient Mean-Shift Tracking via a New Similarity Measure", Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume I, pp. 176- 183.

[5] Bradski, G.R., "Computer Vision Face Tracking for Use in a Perceptual User Interface, " IEEE Workshop on Applications of Computer Vision, Princeton, NJ, 1998, pp. 214- 219.

[6] Comaniciu, D. "Bayesian kernel tracking," Annual Conference of the German Society for Pattern Recognition, 2002, pp. 438-445.

[7] Dong Xu, Yi Ming Wang and Jiwen An, "Applying a new spatial-color histogram in mean shift based tracking algorithm," Conference of Image and Vision Computing, 2005,

[8] Stanley T. Birchfield and Sriram Rangarajan, "Spatiograms versus Histograms for Region Based Tracking," IEEE conference on Computer Vision and Pattern Recognition, 2005, vol.2 pp. 1158- 1163.

[9] Mun-Ho Jeong, Bum-Jae You, Yonghwan Oh, etc, "Adaptive Mean Shift Tracking with Novel Color Model," IEEE International Conference on Mechatronics & Automation, 2005, vol.3, pp. 1329- 1333.

[10] C. Yang and R. Duraiswami and A. Elgammal, "Real-time kernel-based tracking in joint feature-spatial spaces," Tech Report CS-TR-4567, Univ. of Maryland, 2004.

[11] Ahmed Elgammal, Ramani Duraiswami and Larry S. Davis, "Probabilistic tracking in joint feature-spatial spaces", IEEE International Conference on Computer Vision and Pattern Recognition, 2003, vol. 1, pp. 781-788.

[12] Vladimir Nedovic, Martijn Liem, Maarten Corzilius and Mark Smids, "Kernel-based object tracking using adaptive feature selection," project report, 2005.

[13] Weng Muyun, He Mingyi, Zhang Yifan, "An Adaptive Implementation of the Kernel-Based Object Tracking Method," International Conference on Innovative Computing, Information and Control, 2006, vol. 2, pp. 354- 357.

[14] Peng Ning-song,Yang Jie,Liu Zhi."Mean shift blob tracking with kernel histogram filtering and hypothesis testing,"Pattern Recognition Letters,2005,26(5):605 ~ 614.

[15] Fatih Porikli, Oncel Tuzel, "Object Tracking in Low-Frame-Rate Video," Image and Video Communications and Processing, proceedings of the SPIE 2005, Vol. 5685, pp. 72-79.

[16] Zhiwei Zhu, Qiang Ji, Fujimura, K. and Kuangchih Lee, " Combining Kalman filtering and mean shift for real time eye tracking under active IR illumination," Proceedings. 16th International Conference on Pattern Recognition, 2002, vol. 4, pp. 318- 321.

[17] Caifeng Shan, Yucheng Wei, Tieniu Tan, and Ojardias, F., "Real time hand tracking by combining particle filtering and mean shift," Proceedings. Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004, pp. 669- 674.

[18] Alex Po Leung, Shaogang Gong, "Mean-Shift Tracking with Random Sampling ," the 17th by the British Machine Vision Conference, 2006.

[19] Shen, C. Brooks, M. J. van den Hengel, A. , " Fast Global Kernel Density Mode Seeking: Applications to Localization and Tracing," IEEE Transactions on Image Processing, 2007, Vol. 16, pp. 1457-1469.

[20] R. Collins, "Mean-Shift Blob Tracking through Scale Space," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2003.

[21] Zivkovic Z. and Krose, B, "An EM-like algorithm for color-histogram-based object tracking," Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2004, vol. 1, pp. 798- 803.

[22] Zivkovic Z. and Krose, B., "A Probabilistic model for an EM-Like Object Tracking Algorithm Using Color Histograms," In Proc. 6th IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance, 2004.

[23] Alper Yilmaz, "Object Tracking by Asymmetric Kernel Mean Shift with Automatic Scale and Orientation Selection," Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2007.

[24] SuMin Qi, Xianwu Huang and Huaifeng Yi, "Object Tracking by Anisotropic Kernel Mean Shift," JOURNAL OF ELECTRONICS & INFORMATION TECHNOLOGY, 2007, vol. 29(3), pp. 686~ 689.

[25] Cheng, Y., "Mean Shift, Mode Seeking, and Clustering," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 17(8), 1995, pp. 790-799.