Fast Global Kernel Density Mode Seeking with application to Localisation and Tracking

Chunhua Shen, Michael J. Brooks, Anton van den Hengel
School of Computer Science, University of Adelaide, SA 5005, Australia
{*chhshen,mjb,anton*}*@cs.adelaide.edu.au*

Abstract

We address the problem of seeking the global mode of a density function using the mean shift algorithm. Mean shift, like other gradient ascent optimisation methods, is susceptible to local maxima, and hence often fails to find the desired global maximum. In this work, we propose a multi-bandwidth mean shift procedure that alleviates this problem, which we term annealed mean shift*, as it shares similarities with the annealed importance sampling procedure. The bandwidth of the algorithm plays the same role as the* temperature *in annealing. We observe that the over-smoothed density function with a sufficiently large bandwidth is uni-modal. Using a continuation principle, the influence of the global peak in the density function is introduced gradually. In this way the global maximum is more reliably located.*

Generally, the price of this annealing-like procedure is that more iterations are required. Since it is imperative that the computation complexity is minimal in real-time applications such as visual tracking. We propose an accelerated version of the mean shift algorithm. Compared with the conventional mean shift algorithm, the accelerated mean shift can significantly decrease the number of iterations required for convergence.

The proposed algorithm is applied to the problems of visual tracking and object localisation. We empirically show on various data sets that the proposed algorithm can reliably find the true object location when the starting position of mean shift is far away from the global maximum, in contrast with the conventional mean shift algorithm that will usually get trapped in a spurious local maximum.

1. Introduction & Motivation

Kernel-based density estimation techniques for computer vision have recently attracted a great deal of attention. One example is the mean shift technique which has been applied to image segmentation and visual tracking [1–6], *etc.* Mean shift is a versatile nonparametric density analysis tool introduced in [7–9]. In essence, it is an iterative mode detection algorithm in the density distribution space. The mean shift algorithm uses kernels to compute the weighted average of the observations within a smoothing window. This computation is repeated until convergence is attained at a local density mode. This way the density modes can be elegantly located without explicitly estimating the density.

Cheng [8] notes that mean shift is fundamentally a gradient ascent algorithm with an adaptive step size. Recently Fashing *et al.* show the connection between mean shift and the Newton-Raphson optimisation algorithm [10]. They also discover that mean shift is actually a quadratic bound optimisation both for stationary and evolving sample sets [10]. Mean shift is also a fixed-point iteration procedure.

Since Comaniciu *et al.* first introduced mean shift based object tracking [2], it has proven to be a promising alternative to popular particle filtering based trackers. Incremental research has been reported in the literature. In [3] the selection of kernel scale via linear search is discussed. Elgammal *et al.* reformulate the tracking framework as a general form of joint feature-spatial distributions [4, 6]. Compared with the approach of Comaniciu *et al.* [2], the advantage is that spatial structure information of the tracked region is incorporated into the measure.

In [5] multiple spatially distributed kernels are adopted to accurately capture changes in the target's orientation and scale. Another approach is developed in [11] for the same purpose. Furthermore Fan *et al.* present a theoretical analysis of similarity measure and arrive at a criterion, leading to kernel design strategies with prevention of singularity in kernel visual tracking [12]. All of these trackers adopt mean shift or similar optimisation strategies to achieve tracking. Despite successful applications, mean shift trackers require that the displacement of the tracked target in consecutive frames is small. If this is not the case, they are likely to become trapped in spurious local maxima of the multi-modal density distribution space[1]. This happens because mean shift is a purely *local* optimisation method.

Fundamentally, mean shift has two important inherent drawbacks. First of all, it can only be used to find local modes. Being trapped in a local maximum/minimum is a common problem for traditional nonlinear optimisation algorithms. Simulated annealing is a well-known strategy which aims to achieve global optimisation. It starts by initially sampling with a reduced sensitivity to the underly-

[1] In contrast, particle filtering based trackers (*e.g.* [13]) perform better in this situation. However weak dynamical modelling also presents challenges to particle filters.

Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 **IEEE**

ing modes (on a flattened cost function surface) and then progressively increasing the sensitivity to drive samples towards peaked cost regions [14]. Recently the idea of annealing has been merged into importance sampling, yielding annealed importance sampling [15] and it has been introduced to 3D articulated tracking [16].

Motivated by the success of both simulated annealing and annealing importance sampling, we propose a novel multi-bandwidth mean shift procedure, termed *annealed mean shift* (ANNEALEDMS). It shares similarities with the annealed importance sampling procedure in the sense that it also gradually smooths the cost function surface and gently introduces the influence of the global peak. We observe that the over-smoothed density function with a sufficiently large bandwidth[2] h_M is uni-modal. Then, with a continuation principle, we slowly decrease the bandwidth $h = h_M > h_{M-1} > \cdots > h_0$ and, at each bandwidth, we maximise the density (cost) function with mean shift, starting from the convergence position of the previous run. This multi-bandwidth mean shift iteration process is similar to the multi-layered annealing procedure of annealed importance sampling. The main differences are: (1) In ANNEALEDMS, it is the degree of smoothness of the cost function that is annealed, while in annealed importance sampling, it is the degree of flatness of the cost function; (2) Most importantly, in ANNEALEDMS, the number and positions of the modes evolve slowly while in the annealed importance sampling, the temperature does not change the number of modes or their positions.

In theory, as long as the change of the bandwidth is sufficiently slow, the global maximum can usually be found successfully.[3] We provide technical details later.

A second drawback of mean shift is that in many cases it converges slowly. The proposed ANNEALEDMS involves even more iterations, and especially when it is applied to localisation, in which case we have no knowledge of where to start searching. It is imperative that the computational complexity is minimal in real-time applications such as visual tracking. To our knowledge, few attempts have been made to speed up the convergence to mean shift. In [1] locality sensitive hashing is used to reduce the computational complexity of finding the nearest neighbours of a sample point involved in mean shift. Although a dramatic decrease in the execution time is achieved for high-dimensional clustering, this technique is not that attractive for relatively low-dimensional problems such as visual tracking. The speedup of [1] is not obtained by reducing the iteration steps. In this paper, we advance an accelerated version of the mean shift algorithm. Compared with the conventional mean shift algorithm, it can significantly decrease the number of iterations to convergence. The accelerated mean shift is inspired by the successful accelerated variants of bound optimisation algorithms such as Expectation Maximisation (EM). An over-relaxed strategy is then adopted to accelerate the convergence. Much effort has been expended to improve the efficiency of bound optimisation algorithms (*e.g.* EM, [17–19]). A theoretical analysis of the convergence properties for a class of bound optimisation algorithms has been given in [19], and is used as the basis for a novel adaptive over-relaxed scheme. Our proposal is inspired by this approach. Based on the findings in [10], which bridge the gap between mean shift and general bound optimisation algorithms, we promote an adaptive over-relaxed mean shift algorithm which is simple to implement yet significantly more efficient than the standard counterpart.

As applications of the proposed fast, globally mode-seeking mean shift, a fast ANNEALEDMS based object localiser and a visual tracker are developed. Substantially more promising results have been achieved over the conventional mean shift based algorithms. In summary, our key contributions comprise:

1. Development of a novel annealed mean shift algorithm which can reliably find the global mode of a density distribution. This is introduced in Section 3.
2. Reinterpretation of the mean shift algorithm, resulting in a faster version of mean shift. We discuss these issues in Section 4.
3. Application of ANNEALEDMS to the problem of visual tracking using kernel-weighted colour histogram features. Given a target model, the tracker is able to initialise automatically. It also has the capability to recover from tracking failures caused by occlusions, drastic illumination changes, *etc.*, in that the tracker itself can also be a localiser. In contrast, conventional mean shift trackers lack these desirable properties. These developments, including experimental results, are presented in Sections 5.1 and 5.2.

A brief review of the standard mean shift algorithm is presented in Section 2. We conclude the paper in Section 6 with a discussion of some important issues.

2. Mean Shift Analysis

We first review the basic concepts of the mean shift algorithm with notation similar to [9]. One of the most popular nonparametric density estimators is *kernel density estimation*. Given n data points $\mathbf{x}_i$, $i = 1, \cdots, n$, drawn from a population with density function $f(\mathbf{x})$, $\mathbf{x} \in I\!R^d$, the general multivariate kernel density estimate at $\mathbf{x}$ is defined by

$$\hat{f}_K(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{x}_i), \tag{1}$$

[2] By a sufficiently large bandwidth, we mean a bandwidth which is much larger than the optimal bandwidth with the minimum asymptotic mean integrated square error (AMISE).

[3] For continuous variables, the assertion of success is probabilistic.

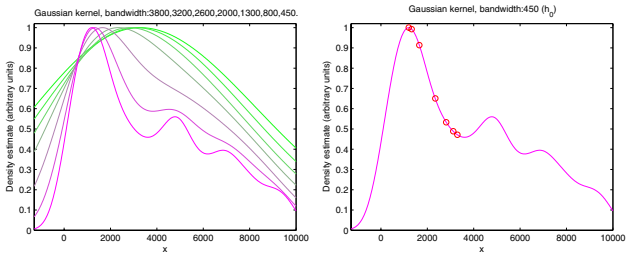

Figure 1: Multi-bandwidth density estimate on 1D galaxy velocity data. (*left*) Curves from outside to inside indicate the *annealing* process with successively decreasing bandwidths. In this case, the optimal bandwidth is $h_0 = 450$. The evolution of the modes is clearly shown: with a multi-bandwidth mean shift mode detection, it is possible to find the global maximum without being distracted by local modes. (*right*) Convergence positions at each bandwidth are marked with circles in the last curve. Note that the unit of the vertical axis is arbitrary.

where $K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-\frac{1}{2}} K(\mathbf{H}^{-\frac{1}{2}}\mathbf{x})$. Here $K(\cdot)$ is a kernel function (or window) with a symmetric positive definite bandwidth matrix $\mathbf{H} \in I\!R^{d\times d}$. A kernel function is bounded with support satisfying the regularity constraints as described in [8, 9]. For simplicity one usually assumes an isotropic bandwidth which is proportional to the identity matrix, *i.e.* $\mathbf{H} = h^2\mathbf{I}$. Employing the profile definition, the kernel density estimator becomes

$$\hat{f}_K(\mathbf{x}) = \frac{c_k}{nh^d} \sum_{i=1}^{n} k\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right), \qquad (2)$$

where $k(\cdot)$ is the profile of the kernel $K(\cdot)$ and c_k is a normalisation constant. The optimisation problem of seeking the local modes is solved by setting the gradient equal to zero. Thus we have

$$\hat{\nabla} f_K(\mathbf{x}) \stackrel{\text{def}}{=} \nabla \hat{f}_K(\mathbf{x}) = \frac{2c_k}{h^2 c_g} \hat{f}_G(\mathbf{x}) \cdot \mathbf{m}_G(\mathbf{x}) = 0, \qquad (3)$$

where

$$\hat{f}_G(\mathbf{x}) = \frac{c_g}{nh^d} \sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right), \qquad (4)$$

$$\mathbf{m}_G(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x}, \qquad (5)$$

and $g(x) = -k'(x)$. Here $k(\cdot)$ is defined to be the shadow of the profile $g(\cdot)$ [10], and $\mathbf{m}_G(\mathbf{x})$ is the mean shift vector. Clearly $\hat{\nabla} f_K(\mathbf{x}) = 0 \rightsquigarrow \mathbf{m}_G(\mathbf{x}) = 0$, and the incremental iteration scheme is obtained immediately:

$$\mathbf{x} \leftarrow \frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}. \qquad (6)$$

1. Determine the set of values for $h_m, (m = M \cdots 0)$ (*a.k.a.* the annealing schedule).
2. Randomly select an initial starting location for the first annealing run and get the convergence location of $\hat{f}_{h_M,K}(\cdot)$, which is $\widehat{\mathbf{x}}^{(M)}$, using *mean shift*.
3. *for* each $m = M-1, M-2, \cdots, 0$, run *mean shift* to get the convergence position $\widehat{\mathbf{x}}^{(m)}$ with the initial position $\widehat{\mathbf{x}}^{(m+1)}$, *i.e.*, the convergence position from the previous bandwidth. $\widehat{\mathbf{x}}^{(0)}$ is then the final global mode.

Figure 2: The ANNEALEDMS algorithm.

3. Annealed Mean Shift

Let $h_m(m = M, M-1, \cdots, 0)$ be a monotonically decreasing sequence of bandwidths such that h_0 is the optimal bandwidth for the considered data set and usually $h_M \gg h_0$.[4] A series of kernel density functions $\hat{f}_{h_M,K}(\cdot)$, $\hat{f}_{h_{M-1},K}(\cdot)$, $\cdots$, $\hat{f}_{h_0,K}(\cdot)$ are applied to the sample data, where the subscripts of $\hat{f}_{h,K}(\cdot)$ denote the bandwidth and kernel type respectively.

Figure 1 illustrates a 1D example[5], where $M = 6$. With a large bandwidth, the function $\hat{f}_{h_M,K}(\cdot)$ is uni-modal, merely representing the overall trend of the density function. Thus the starting point of the first annealing run does not affect the mode detection.

The ANNEALEDMS algorithm is given in Figure 2.

3.1. Remarks

1. Apparently the annealing schedule is a trade off between efficiency and efficacy: slow annealing is more likely to find a global maximum, but could also be prohibitively expensive.
2. A justification of why ANNEALEDMS works is that the number of modes of a kernel density estimator with a Gaussian kernel is monotonically non-increasing [22]. In order to convey convergence information, the monotonicity of number of modes with respect to bandwidths is compulsory. Note that the monotonicity result only applies to the Gaussian kernel: compactly supported kernels such as the Epanechnikov kernel may not have this property. However, as pointed out in [22], the lack of monotonicity happens only for relatively very small bandwidths. The notion of a critical bandwidth[6] for the popular kernels such as the

[4] There is a tremendous amount of literature on how to select the optimal bandwidth in order to produce a minimum AMISE estimate (see, *e.g.*, [20, 21].) In this work, we assume h_0 can be obtained by existing techniques.

[5] The 1D galaxy velocity data set is also used in [10].

[6] The smallest bandwidth above which the number of modes is monotone.

Epanechnikov kernel is still well defined. Moreover, "just as in the Gaussian case, the critical bandwidth is of the same size as the bandwidth (h_0) that minimises mean square error of the density estimator" [22]. This conclusion serves as one of the theoretical bases of our ANNEALEDMS: we are not interested in the bandwidth under h_0. Rather, we only take advantage of the property of *over-smoothness* at bandwidths *above* h_0. Therefore, for the applications we are interested in, *e.g.*, visual localisation and tracking, the problem of nonmonotonicity does not arise. We have also empirically shown this important proposition.

3. Unless otherwise specified, in our examples the (truncated) Gaussian kernel is used because it is well suited to fast computation [6, 23]. When Gaussian kernel is adopted, the mechanism is related with the well-developed multi-scale theory [24, 25]. The over-smoothed kernel density is essentially a Gaussian smoothed version of the true density, via convolution with extra Gaussians.

We note that, in the statistics literature, [24] has proposed an algorithm *SiZer* to explore the significant modes in an estimated curve across multiple scales. *SiZer* employs a similar principle. In computer vision, a similar strategy, variable-bandwidth density-based fusion (VBDF), has also been adopted to find the most significant mode of a density function in the context of information fusion for multiple motion estimation [26]. However, there are no theoretical details given in [26]. We independently develop ANNEALEDMS mainly inspired by simulated annealing and annealed importance sampling. We have shown a connection between ANNEALEDMS and these annealing techniques. Furthermore, we use it in a novel way to solve some problems in robust visual localisation and tracking.

3.2. Numerical Examples

1D example. Figure 1 shows a simple 1D example on the galaxy data [10]. Because of the density estimator's uni-modal property at a large bandwidth (h_M), the start position at h_M has no effect on the final convergence. Figure 1 shows that the global maximum is successfully located with a rough seven-step annealing schedule. For this particular case, actually only two steps are needed to locate the global mode.

2D example. For this example, the data are drawn from a Gaussian mixture $0.1 \cdot \mathcal{N}\left([-1,0]^\top, 0.13\mathbf{I}\right) + 0.2 \cdot \mathcal{N}\left([1,2]^\top, \mathbf{I}\right) + 0.7 \cdot \mathcal{N}\left([1,-2]^\top, \mathbf{I}\right)$. A four-step ANNEALEDMS with bandwidths $\{2, 1.02, 0.66, 0.45\}$ is used to locate the global mode. Figure 3 depicts the annealing process. Again due to the uni-modal property, no matter from which initial position ANNEALEDMS starts, the global mode is always obtained eventually. A video sequence (*GMM2D.avi*)[7] is also generated to show the mode evolution process more elaborately.

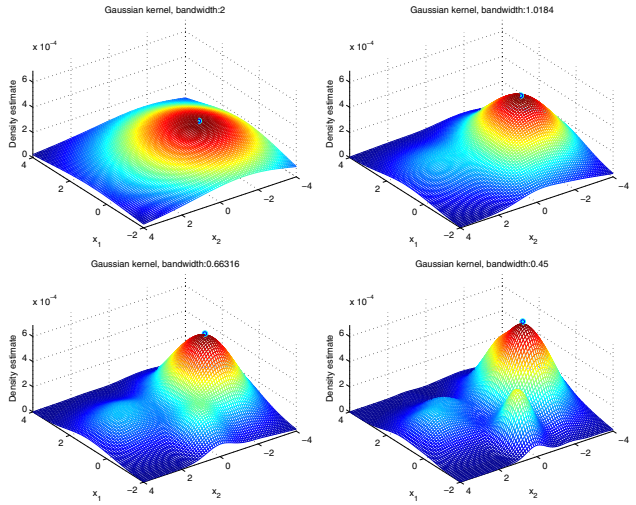

Figure 3: Multi-bandwidth density estimate on 2D artificial Gaussian mixture data. A four-step annealing schedule is employed to find the global mode. The modes found by mean shift across bandwidths are marked with circles. See the sequence *GMM2D.avi* which demonstrates a slower evolution across bandwidths.

For these two examples, we do not assume any prior information about the distribution structure of the data. The only information needed is the approximate range of the data, which is usually available.

4. Fast Mean Shift

Generally, the price of global convergence of ANNEALEDMS is that more iterations are required. This is the case particularly when the start point is far away from the convergence position. It is imperative that the computational complexity is minimal in real-time applications such as visual tracking. We introduce an adaptive over-relaxed accelerated mean shift in this section.

4.1. Adaptive Over-Relaxed Mean Shift

The following two theorems serve the bases of the adaptive over-relaxed mean shift algorithm.

Theorem 4.1 (Cheng [8]): *Mean shift with kernel $G(\cdot)$ finds the modes of the density estimate with kernel $K(\cdot)$,* i.e. *$\hat{f}_K(\cdot)$, where $K(\cdot)$ is the shadow of the kernel $G(\cdot)$.*

With the analysis in Section 2, Theorem 4.1 is evident.

Theorem 4.2 (Fashing *et al.* [10]): *Mean shift with kernel $K(\cdot)$ is a quadratic bound optimisation over a density estimate with a continuous shadow of $K(\cdot)$.*

[7] The videos mentioned in this paper can be accessed at http://www.cs.adelaide.edu.au/~vision/demo/index.html.

These two theorems show that mean shift is actually a bound maximisation. One step of the mean shift procedure of Equation (6) finds the exact maximum of the lower bound of the objective function $\hat{f}_K(\mathbf{x}^{(\kappa)})$.[8] From Equation (3) we have $\mathbf{m}_G(\mathbf{x}) \propto \frac{\hat{\nabla} f_K(\mathbf{x})}{\hat{f}_G(\mathbf{x})}$, which means mean shift is a gradient ascent algorithm with adaptive step size. Hence its convergence rate is better than conventional fixed-step gradient algorithms. As we will see, however, from the viewpoint of bound optimisation, the learning rate can be over-relaxed to make its convergence faster.

It is well known that, in order to guarantee increasing the cost function at each iteration, bound optimisation methods must usually build conservative bounds, leading to slow convergence [17, 19]. A lot of work has been carried out to speed up bound optimisation methods, especially for the EM algorithm due to its popularity [18]. In [19] it is shown that by over-relaxing the step size, acceleration can be achieved. Denote the bound function as $\rho(\mathbf{x}, \mathbf{x}^{(\kappa)})$, then the over-relaxed bound optimisation iteration becomes:

$$\mathbf{x}^{(\kappa+1)} = \mathbf{x}^{(\kappa)} + \beta \left[\arg\max_{\mathbf{x}} \rho(\mathbf{x}, \mathbf{x}^{(\kappa)}) - \mathbf{x}^{(\kappa)} \right]. \quad (7)$$

Apparently when the learning rate $\beta = 1$, over-relaxed optimisation reduces to the standard bound optimisation algorithm. It is easily seen that when $\beta > 1$ acceleration is realised. Nevertheless by simply assigning a fixed value to β, no convergence is secured and it seems quite difficult, if not impossible, to obtain the optimal value for β. Xu proves that in the case of the Gaussian Mixture Model (GMM) parameter estimation with EM, convergence can be guaranteed using this method when we are *close* to a local maximum and $0 < \beta < 2$ [17]. This conclusion is generalised to the case of general bound optimisation methods in [19]. Based on this important proposition, a simple adaptive over-relaxed bound optimisation is readily available: the learning rate β can be adjusted by evaluating the cost function. When one observes for some $\beta > 1$ that the cost function becomes worse, then β has been set too large and needs to be reduced. By simply setting $\beta = 1$ immediately, convergence can be achieved. By regarding mean shift as a special case of bound optimisation, these theoretical conclusions also apply to mean shift.

The accelerated mean shift algorithm obtained in this way is shown in Figure 4. One can easily check that the following relation holds (up to a translation and a scale factor):

$$\begin{aligned} \hat{f}_K(\mathbf{x}^{(\kappa+1)}) &= \rho(\mathbf{x}^{(\kappa+1)}, \mathbf{x}^{(\kappa+1)}) \geq \rho(\mathbf{x}^{(\kappa+1)}, \mathbf{x}^{(\kappa)}) \\ &\geq \rho(\mathbf{x}^{(\kappa)}, \mathbf{x}^{(\kappa)}) = \hat{f}_K(\mathbf{x}^{(\kappa)}). \end{aligned}$$

Note that in the above analysis we do not take the mean shift with a weight function into consideration, but the accelerated algorithm also applies for the weighted case, because the two theorems concerned are derived from the weighted mean shift [8,10]. The only overhead is the evaluation of the cost function. However, we will see that for the (truncated) Gaussian kernel, its special structure means that computing the mean shift iteration with Equation (6) also results in evaluation of the cost function $\hat{f}_K(\mathbf{x})$. Because the shadow of the Gaussian kernel is itself, we have $\hat{f}_K(\mathbf{x}) = \hat{f}_G(\mathbf{x})$.

[8] $\kappa = 1, 2, \cdots$, denotes the iteration index.

1. *Initialisation:*
   Set the iteration index κ = 1, the learning rate $\beta = 1$, and the step parameter $\alpha > 1$.
2. *Iterate until convergence condition is met:*
   (a) Calculate $\widetilde{\mathbf{x}}^{(\kappa+1)}$ with Equation (6). Calculate the mean shift $\mathbf{m}_G(\mathbf{x}^{(\kappa+1)}) = \widetilde{\mathbf{x}}^{(\kappa+1)} - \mathbf{x}^{(\kappa)}$.
   (b) $\mathbf{x}^{(\kappa+1)} = \mathbf{x}^{(\kappa)} + \beta \cdot \mathbf{m}_G(\mathbf{x}^{(\kappa+1)})$.
   (c) *if* $\hat{f}_K(\mathbf{x}^{(\kappa+1)}) > \hat{f}_K(\mathbf{x}^{(\kappa)})$,
   Accept $\mathbf{x}^{(\kappa+1)}$ and $\beta = \alpha \cdot \beta$;
   else
   Reject $\mathbf{x}^{(\kappa+1)}$, $\mathbf{x}^{(\kappa+1)} = \widetilde{\mathbf{x}}^{(\kappa+1)}$, and $\beta = 1$.
   (d) Set $\kappa = \kappa + 1$. Start a new iteration.

Figure 4: The over-relaxed adaptive mean shift algorithm.

A question naturally arises, what if a kernel other than a Gaussian, *e.g.*, Epanechnikov, is adopted? The observation that we can reliably judge the behaviour of $\hat{f}_K(\mathbf{x})$ through the estimate $\hat{f}_G(\mathbf{x})$ is only satisfied when these two kernel functions generate density estimates of the same degree of smoothness. For different kernels, as long as the bandwidths are adjusted accordingly, all the kernels are asymptotically equivalent under the AMISE error criterion. Therefore the kernel type is not of importance in mean shift analysis but the bandwidth plays a critical role. For a non-Gaussian kernel, the shadow is different from itself ($\hat{f}_K(\mathbf{x}) \neq \hat{f}_G(\mathbf{x})$). The smoothness of two kernel density estimates with the same bandwidth but different kernels might be quite different. As a consequence, usually we cannot reuse the density $\hat{f}_G(\mathbf{x})$ calculated in Equation (6) and an extra evaluation of the cost function $\hat{f}_K(\mathbf{x})$ needs to be made.

In fact, if the bandwidths of two different kernels h_A, h_B satisfy $\frac{h_A}{h_B} = \frac{\delta_{0,A}}{\delta_{0,B}}$, where δ_0 is a kernel's *canonical bandwidth*, then the density estimates based on these two kernels have the same degree of smoothness [27]. Utilising this knowledge, in practice, if the canonical bandwidths associated with a kernel and its shadow kernel are comparable, we still can reuse $\hat{f}_G(\mathbf{x})$. Although no details on this topic are presented in this paper, we have validated this conclusion with numerical experiments. However, one should be aware that the measurement of *comparable* is application dependent.

4.2. Numerical Experiments

We compare the performance of the proposed accelerated mean shift algorithm with the standard mean shift algorithm on both synthetic data and real application data sets. Note that rejected iterations are also counted for the accelerated mean shift algorithm.

Due to limited space, the detailed description of the data sets is omitted, which can be found in [23]. In all the tests, we use $\alpha = 1.25$ and the convergence tolerance $\varepsilon = \frac{\hat{f}_K(\mathbf{x}^{\kappa+1}) - \hat{f}_K(\mathbf{x}^{\kappa})}{\hat{f}_K(\mathbf{x}^{\kappa})} = 0.001$. The resulting mode locations found by the two algorithms are so close that the difference is negligible. We run the comparison with three arbitrarily selected start points on each data set. The experiment results are reported in Table 4.2. The proposed algorithm is significantly more efficient than the standard mean shift. The evaluation results are promising: a speedup by a factor of about $2 \sim 5$ can be achieved in these evaluations. We have also developed an accelerated mean shift tracker, which outperforms the conventional mean shift tracker [23].

The accelerated mean shift's performance with fewer convergence iterations has proven commensurate with its standard counterpart. In theory when the start point is extremely close to the local maximum, the rejection in the proposed accelerated mean shift procedure might happen frequently, resulting in a resource waste. In practice these cases are very rare. Moreover one can devise smarter step-size adjustment strategies to survive in this extreme case.

data set	*initial*	*number of iterations*	
		fast mean shift	mean shift
data set #1	-0.8	13	51
	1.5	16	77
	3.6	11	33
data set #2	9800	12	49
	-1005	8	15
	3200	10	31
data set #3	$(-5, 20)$	12	34
	$(-10, 16)$	11	29
	$(20, 10)$	13	35
data set #4	$(1, -1.4)$	29	119
	$(1.5, 0.4)$	17	65
	$(0.3, 0.3)$	12	36

Table 1: Comparison of number of iterations for convergence. The initial location for each run is shown in the second column.

5. Fast Annealed Mean Shift Based Visual Localisation and Tracking

In this section we apply the two improvements on mean shift to visual localisation and tracking. In all the localisation and tracking experiments we use RGB colour histograms, consisting of $16 \times 16 \times 16$ bins. The tracking framework presented in [2] is adopted, but we use an annealing procedure for global mode seeking.

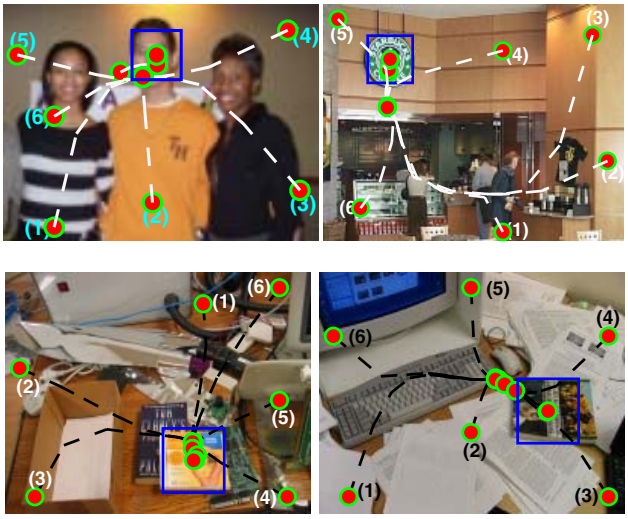

Figure 5: We locate a specified human face (*left top*) beside two spurious faces, the STARBUCKS logo (*right top*), a CD (*left bottom*) and a book cover (*right bottom*) under cluttered backgrounds. The ANNEALEDMS is started at arbitrarily selected positions. Dashed lines indicate the mean shift searching trajectories for each run. Dots indicate the start and convergence positions of mean shift for each bandwidth. See the accompanying videos *localiser*$\{1, 2, 3, 4\}$*.avi* for an intuitive demonstration on the annealing convergence processes.

5.1. Visual Localisation

Up to date, mean shift has typically been used for tracking motions with small displacements due to its lack of global mode seeking capability. Armed with ANNEALEDMS, it is possible to locate a target *no matter from which initial position the mean shift localiser starts*, given the target template.

In our experiments, the ANNEALEDMS localiser starts at arbitrarily selected positions. All result on successful location of the target. Six runs for each example are marked in Figure 5. Four objects are located successfully in different environments. For the first example, the bandwidths are $\{60, 40, 20, 10\}$. We plot the cost functions of this example in Figure 6 to explore how ANNEALEDMS works in this case. The influence of the most significant peak is introduced gradually, which guides search towards the global mode. One can see that even at $h_1 = 20$, there have been plenty of local modes which can *easily* make the search stop prematurely. At $h_0 = 10$, there are three major modes corresponding to the three faces in the figure. Note that mean shift does not converge to the exact modes in Figure 6 due to the Taylor approximation [2]. However it converges to a position close to the true mode. For localisation and tracking, this accuracy loss is negligible.

The other three examples begin at $h_4 = 80$ and a five-step annealing guarantees a global mode in these cases. For

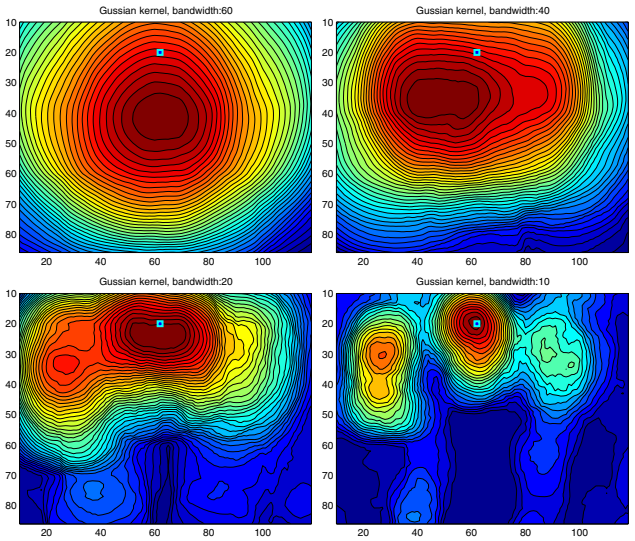

Figure 6: The cost functions (corresponding to the first example in Figure 5) at different bandwidths: 60, 40, 20, and 10, are plotted as contours of 2D translations. The true mode is marked with a square.

the CD and book cover localisation, we take the template models from other images with large geometry and slight illumination differences. The success proves that the colour histogram is a robust feature. It is straightforward to include other features, *e.g.*, intensity gradient, to make the localiser more robust. Without the annealing procedure, most runs stop at a local maximum—only when the initial positions are located in the small area close to the global mode, can standard mean shift find the target. When no prior knowledge is available about the global maximum we are seeking, it is always beneficial to employ a relatively broad bandwidth mean shift procedure, which can provide a coarse location of the global mode. In the experiments, although we do not carefully design the annealing schedule, global modes are achieved.

5.2. Visual Tracking

Many tracking algorithms fail to perform well in practice. They have several fundamental drawbacks: (1) They work well only when the displacements between consecutive frames are relatively small; (2) Usually they cannot self-start; (3) They are not robust to occlusions and are unable to recover from momentary tracking failures. Standard mean shift trackers are no exception. Our ANNEALEDMS tracker alleviates these weaknesses by incorporating an efficient bottom-up localisation functionality.

Face tracking example. The tracked target moves fast hence leading to large displacements between consecutive frames. An annealing schedule $\{60, 30, 18\}$ is used by ANNEALEDMS. The ANNEALEDMS tracker is automatically started by a localisation process, while the mean shift tracker is manually started. As in mean shift tracking, ANNEALEDMS also starts at the position of the previous frame. Unless otherwise noted, in all the tracking experiments, the convergence tolerance is the l_2-norm distance between two iterations $\varepsilon = 0.2$ pixels. Figure 7 summarises the tracking results. The ANNEALEDMS tracker is more robust and accurate than the standard mean shift tracker: When the displacement is large, the standard mean shift tracker becomes easily stuck in spurious modes.

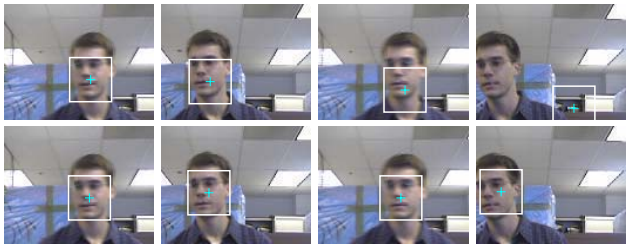

Figure 7: The face tracking sequence with standard mean shift (*top*) and ANNEALEDMS (*bottom*). Frames #5, #14, #22 and #25 are shown. The object is *accurately* detected and tracked by ANNEALEDMS despite large displacements. In contrast, mean shift is more likely to trap into local modes and gives inaccurate results (#5, #14 and #22) or even fails completely (#25). See the video *facetracker.avi* for details.

Implementation issues. Mean shift might get stuck at false modes caused by discrete colour values of pixels. Wang *et al.* observe this phenomenon in grey image histogram clustering [28]. Their analysis also applies to colour image histograms. We avoid this problem by simply ceiling the mean shift step $\lceil \mathbf{m}_G(\mathbf{x}) \rceil$ (Equation (5)). This modification increases the size of shift steps, hence leading to a quicker convergence. The drawback is that it might lose accuracy. We use the original step by Equation (5) at the last bandwidth h_0. Because we are only interested in the last convergence position, accuracy is retained. Both in localisation and tracking, it has been observed that this simple treatment results in satisfactory convergence without accuracy loss. We compare the number of convergence iterations per frame for the face tracking video in Figure 8.[9] One can see that in this example, their convergence speeds are similar. In many frames, ANNEALEDMS is even faster. The reason is that mean shift at the first few bandwidths ($h_{M\ldots1}$) can move close to the mode quickly. However, larger bandwidths of ANNEALEDMS mean that slightly more computation might be needed to build histograms. We also have implemented the proposed accelerated mean shift algorithm (Section 4) into tracking, and a considerable speedup has been achieved [23].

Basketball tracking example. This example shows the ANNEALEDMS tracker's ability to recover from temporal failures. The original sequence is down sampled by a factor of 2 to make the target's displacements larger. The mean

[9] Only frames #1…22 are compared because from #23 on, the mean shift tracker fails. For ANNEALEDMS, we count the sum of iterations at each bandwidth.

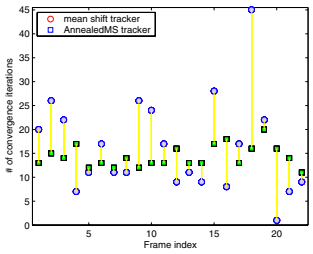

Figure 8: Comparison of the number of iterations per frame: mean shift (marked with circles) *vs.* ANNEALEDMS (marked with squares), for the face tracking sequence.

shift tracker fails as early as at Frame #6. Therefore we only show the tracking results of ANNEALEDMS in Figure 9. ANNEALEDMS tracks across bandwidths $\{30, 15, 8\}$ and works successfully. It is not always necessary to perform a global search in tracking. In ANNEALEDMS the hierarchical bandwidths control the size of the searching area in the cost function. For this video, the first bandwidth is not set very large because a global search might not be desired. At #18, ANNEALEDMS loses the target due to illumination changes. However, it recovers immediately at #19. It drifts slightly because of the basket's occlusions at #20 and recovers at the next frame. Again we observe ANNEALEDMS tracking is efficient: only an average 8.1 iterations per frame is needed.

Weetbix box tracking example. We track a part of a weetbix box, which is recorded by an extremely unstable camera. ANNEALEDMS shows its robustness over the mean shift tracker again. See *weetbixbox.avi* for tracking results.

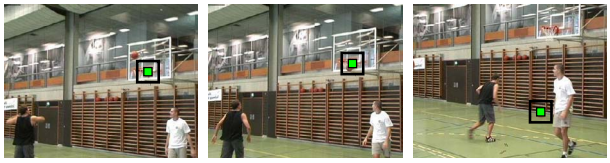

Figure 9: The basketball tracking results with ANNEALEDMS. Frames #18, #20 and #29 are shown. See *basketball.avi* for details.

6. Conclusion

We have presented a new global mode seeking mean shift, termed ANNEALEDMS. Improvements are achieved over the standard mean shift when the density has multiple peaked modes. We have also introduced the new ANNEALEDMS strategy into localisation and tracking. Promising results have been obtained in both applications, even with simple annealing schedules, which are not carefully designed.

An adaptive over-relaxed mean shift is also advanced to accelerate the convergence speed. Compared with the standard mean shift algorithm, the number of convergence iterations is almost always significantly decreased. It provides an additional speedup to existing techniques such as locality sensitive hashing [1] and fast Gaussian transformation [6].

Future work will explore the effects of annealing schedule design on the localisation and tracking performances. Other discriminative features will be adopted for better localisation and tracking performances, rather than relying solely on simple colour histograms.

References

[1] B. Georgescu, I. Shimshoni, P. Meer, Mean shift based clustering in high dimensions: A texture classification example, in: IEEE Int'l Conf. on Comp. Vision, Vol. 2, Nice, France, 2003, pp. 456–463.
[2] D. Comaniciu, V. Ramesh, P. Meer, Kernel-based object tracking, IEEE Trans. on Patt. Anal. and Mach. Intell. 25 (5) (2003) 564–577.
[3] R. Collins, Mean-shift blob tracking through scale space, in: IEEE Conf. on Comp. Vision and Patt. Recog., Vol. 2, Madison, Wisconsin, 2003, pp. 234–240.
[4] A. Elgammal, R. Duraiswami, L. S. Davis, Probabilistic tracking in joint feature-spatial spaces, in: IEEE Conf. on Comp. Vision and Patt. Recog., Vol. 1, Madison, Wisconsin, 2003, pp. 781–788.
[5] G. D. Hager, M. Dewan, C. V. Stewart, Multiple kernel tracking with SSD, in: IEEE Conf. on Comp. Vision and Patt. Recog., Vol. 1, Washington, D.C., 2004, pp. 790–797.
[6] C. Yang, R. Duraiswami, L. Davis, Efficient spatial-feature tracking via the mean-shift and a new similarity measure, in: IEEE Conf. on Comp. Vision and Patt. Recog., Vol. 1, San Diego, CA, 2005, pp. 176–183.
[7] K. Fukunaga, L. D. Hostetler, The estimation of the gradient of a density function, with applications in pattern recognition, IEEE Trans. on Information Theory 21 (1975) 32–40.
[8] Y. Cheng, Mean shift, mode seeking, and clustering, IEEE Trans. on Patt. Anal. and Mach. Intell. 17 (8) (1995) 790–799.
[9] D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE Trans. on Patt. Anal. and Mach. Intell. 24 (5) (2002) 603–619.
[10] M. Fashing, C. Tomasi, Mean shift is a bound optimization, IEEE Trans. on Patt. Anal. and Mach. Intell. 27 (3) (2005) 471–474.
[11] Z. Zivkovic, B. Krose, An EM-like algorithm for color-histogram-based object tracking, in: IEEE Conf. on Comp. Vision and Patt. Recog., Vol. 1, 2004, pp. 798–803.
[12] Z. Fan, Y. Wu, Multiple collaborative kernel tracking, in: IEEE Conf. on Comp. Vision and Patt. Recog., Vol. 2, San Diego, CA, 2005, pp. 502–509.
[13] P. Pérez, C. Hue, J. Vermaak, M. Gangnet, Color-based probabilistic tracking, in: 7th European Conf. on Comp. Vision, Vol. 2350 of LNCS, Springer, Copenhagen, Denmark, 2002, pp. 661–675.
[14] P. van Laarhoven, E. Aarts, Simulated Annealing: Theory and Applications, Springer Verlag, 1987.
[15] R. M. Neal, Annealed importance sampling, Statistics and Computing 11 (2) (2001) 125–139.
[16] J. Deutscher, I. Reid, Articulated body motion capture by stochastic search, Int'l J. of Comp. Vision 61 (2) (2005) 185–205.
[17] L. Xu, On convergence properties of the EM algorithm for Gaussian mixtures, Neural Computation 8 (1) (1996) 129–151.
[18] L. E. Ortiz, L. P. Kaelbling, Accelerating EM: An empirical study, in: Uncertainty in Artificial Intell., Stockholm, Sweden, 1999, pp. 512–521.
[19] R. Salakhutdinov, S. Roweis, Adaptive overrelaxed bound optimization methods, in: Int'l Conf. on Mach. Learning, AAAI Press, Washington DC, 2003, pp. 664–671.
[20] D. Comaniciu, An algorithm for data-driven bandwidth selection, IEEE Trans. on Patt. Anal. and Mach. Intell. 25 (2) (2003) 281–288.
[21] M. C. Jones, J. S. Marron, S. J. Sheather, A brief survey of bandwidth selection for density estimation, J. of the American Statistical Association 91 (433) (1996) 401–407.
[22] P. Hall, M. C. Minnotte, C. Zhang, Bump hunting with non-Gaussian kernels, The Annals of Statistics 32 (5) (2004) 2124–2141.
[23] C. Shen, M. J. Brooks, Adaptive over-relaxed mean shift, in: 8th Int'l Symp. on Signal Processing and Its Applications, Sydney, Australia, 2005.
[24] P. Chaudhuri, J. S. Marron, Scale space view of curve estimation, The Annals of Statistics 28 (2) (2000) 408–428.
[25] T. Lindeberg, Scale-Space Theory in Comp. Vision, Kluwer Academic Publishers, 1994.
[26] D. Comaniciu, Nonparametric information fusion for motion estimation, in: IEEE Conf. on Comp. Vision and Patt. Recog., Vol. 1, Madison, Wisconsin, 2003, pp. 59–66.
[27] W. Härdle, M. Müller, S. Sperlich, A. Werwatz, Nonparametric and Semiparametric Models, Springer Series in Statistics, Springer, 2004.
[28] H. Wang, D. Suter, False-peaks-avoiding mean shift method for unsupervised peak-valley sliding image segmentation, in: 8th Digital Image Computing: Techniques and Applications, Sydney, 2003, pp. 581–590.