Performance analysis for tracking of variable scale objects using mean-shift algorithm

Ning Song Peng
Henan University of Science and Technology
Luoyang 471039, China

Jie Yang
Shanghai Jiaotong University
Shanghai 200436, China

Zhi Liu
Shanghai University
Shanghai 200436, China

Abstract. Classic mean-shift trackers have no integrated scale adaptation, which limits their performance in tracking variable scale objects. By analyzing the similarity of object kernel histograms, we found that the changes of object scale and position within the fixed kernel make the Bhattacharyya coefficient monotonic decreasing. The work plays a guiding role in solving scaling problems within the mean-shift framework.
[DOI: 10.1117/1.1985487]

Subject terms: mean-shift tracking; Bhattacharyya coefficient; scaling problem.

Paper L040566RR received Aug. 20, 2004; revised manuscript received Oct. 17, 2004 and May 11, 2005; accepted for publication Jun. 6, 2005; appeared online Jun. 7, 2004; published online Jul. 22, 2005.

The mean-shift algorithm[1] is an efficient method for mode seeking without doing an exhaustive search, which leads to a real-time property. It has been introduced recently for tracking applications.[2–5] However, the fixed kernel bandwidth is always leading to poor localization in tracking objects changing in scale. A moment is used to compute the size of the tracking windows.[2] However, the computational complexity is too high to meet the real-time requirement. In general, an object scale is detected by calculating the Bhattacharyya coefficient for three different sizes (same scale, ±5% change) and choosing the size that gives the highest similarity to the target model.[5] Since it is a naive method for scale adaptation without considering the underlying relationship between the similarity and the object scale changes, the size of the tracking windows cannot always keep up with the object scale changes. In this paper, this relationship is theoretically analyzed for a possible total solution in the future.

Definition 1. A round region T containing the whole object region F and some background region B is called a tracking window. Function $c(T)$ and $c(F)$ denote the center of T and F, respectively. Their distance is measured by $d(T,F)=\|c(T)-c(F)\|$.

Definition 2. Let $\{x_i\}_{i=1\ldots n}$ be the pixel locations with $c(T)$ as the origin point. The kernel histogram[5] of T with m bins is defined by $P=\{p_\mu\}_{\mu=1\ldots m}$ where

$$p_\mu = C\sum_{i=1}^{n} k(\|x_i/r\|^2)\,\delta[q(x_i)-\mu]. \tag{1}$$

Here k is the kernel function and r is the kernel bandwidth, which determines the radius of T. Function $q:R^2 \rightarrow\{1\ldots m\}$ associates the pixel at location x_i to the index $q(x_i)$ of the kernel-histogram bin corresponding to the color of that pixel. C is derived by imposing the constraint $\Sigma_{\mu=1}^{m} p_\mu=1$. Suppose the color distribution of F is distinguished from B. It can be approximately satisfied in many applications, e.g., traffic surveillance, and described by $\Sigma_{\mu=1}^{m} p_\mu^i \cdot p_\mu^j=0$ where the color distribution of F and B are represented by $P_i=\{p_\mu^i\}_{\mu=1\ldots m}$ and $P_j=\{p_\mu^j\}_{\mu=1\ldots m}$, respectively.

Definition 3. The similarity of two kernel histograms P_i and P_j with m bins is measured by the Bhattacharyya coefficient[5]

$$\rho(i,j)=\sum_{\mu=1}^{m}\sqrt{p_\mu^i p_\mu^j}, \quad i\neq j, \tag{2}$$

where p_μ^i and p_μ^j are the value of bin μ in P_i and P_j, respectively.

Theorem 1. Given T_1 with $c(F_1)=c(T_1)$ in frame i and T_2 with the same position of T_1 in frame $i+1$ where object scale and position are changed, $\forall T_3 \in$ frame $i+1$, if $d(T_2,F_2)<d(T_3,F_3)$ then $\rho(2,1)>\rho(3,1)$.

Proof. By assuming without loss of generality that (1) the object shrinks its scale from frame i to $i+1$. (2) F_i, $i=1,2,3$ consists of u subregions with different intensity levels, i.e., $F_i=\{f_j\}_{j=1\ldots u}$, while B_i, $i=1,2,3$ consists of ν_i subregions with different intensity levels, i.e., $B_i=\{b_j\}_{j=1\ldots \nu_i}$. (3) Consider T_i; suppose its kernel histogram $P_i=\{p_\mu^i\}_{\mu=1\ldots m}$ consists of two entries, sets $\{fp_j^i\}_{j=1\ldots u}$ and $\{bp_j^i\}_{j=1\ldots \nu_i}$, corresponding to the subregion $\{f_j\}_{j=1\ldots u}$ and $\{b_j\}_{j=1\ldots \nu_i}$, respectively, where $u+\max(\nu_1,\nu_2,\nu_3)\leqslant m$.

The continuous form of Eq. (1) is as follows:

$$\begin{cases} fp_j^i = C_i \iint_{\sigma=f_j} k(\|x/r\|^2)d\sigma, & i=1,2,3; \quad j=1\ldots u \\ bp_j^i = C_i \iint_{\sigma=b_j} k(\|x/r\|^2)d\sigma, & i=1,2,3; \quad j=1\ldots \upsilon_i \end{cases},$$

where

$$C_i = 1 \Big/ \left[\sum_{j=1}^{u}\iint_{\sigma=f_j} k(\|x/r\|^2)d\sigma + \sum_{j=1}^{\upsilon_i}\iint_{\sigma=b_j} k(\|x/r\|^2)d\sigma\right].$$

By using integral theorem of mean, we have

$$\begin{cases} fp_j^2 = C_2\cdot S_{f_j}^2 \cdot k(\|\xi_{f_j}^2/r\|^2), & \xi_{f_j}^2 \in f_j \text{ in } F_2 \\ fp_j^3 = C_3\cdot S_{f_j}^3 \cdot k(\|\xi_{f_j}^3/r\|^2), & \xi_{f_j}^3 \in f_j \text{ in } F_3 \end{cases} \tag{3}$$

where $S_{f_i}^2$ and $S_{f_i}^3$ are areas of subregion f_j in F_2 and F_3, respectively.

The fixed kernel bandwidth leads to $C_2=1/\iint_{\sigma=T_2} k(\|x/r\|^2)d\sigma=C_3$, and it is clear that $S_{f_j}^2=S_{f_j}^3$ owing to $F_2=F_3$. Since k is monotonic decreasing[1] and

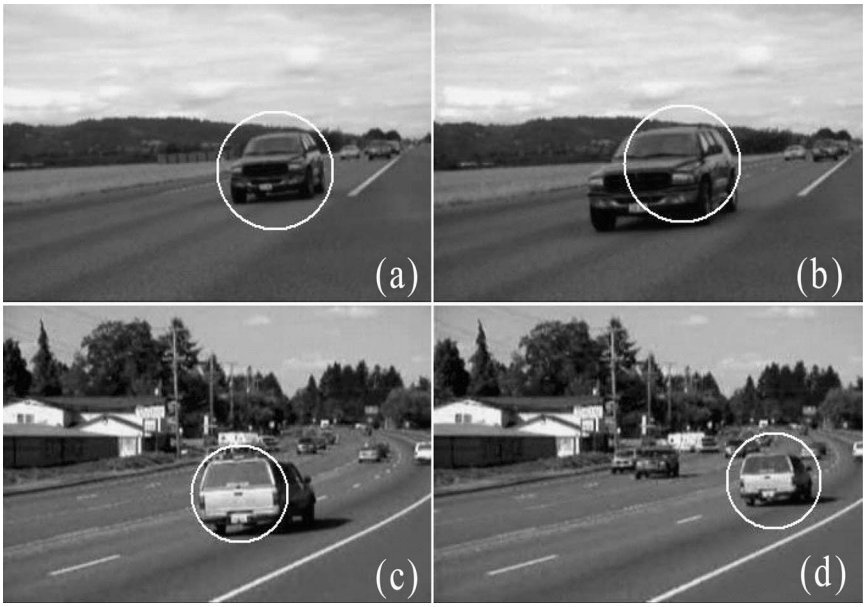

Fig. 1 Tracking results with fixed kernel bandwidth (left to right).

$d(T_2,F_2)<d(T_3,F_3)$, we have $k(\|\xi_{f_j}^2/r\|^2)>k(\|\xi_{f_j}^3/r\|^2)$. Consequently, we obtain $fp_j^2>fp_j^3$. Moreover, $\Sigma_{j=1}^{\nu_2}bp_j^2<\Sigma_{j=1}^{\nu_3}bp_j^3$ holds owing to the constraint

$$\sum_{j=1}^{u} fp_j^i + \sum_{j=1}^{v_i} bp_j^i = 1, \quad i = 1,2,3. \tag{4}$$

Since the scale of F_2 is less than F_1, the area of B_2 is greater than B_1. Thus, $\Sigma_{j=1}^{\nu_1}bp_j^1<\Sigma_{j=1}^{\nu_2}bp_j^2$ holds and then $\Sigma_{j=1}^{u}fp_j^2<\Sigma_{j=1}^{u}fp_j^1$. Therefore,

$$\begin{cases} 1 > \sum_{j=1}^{u} fp_j^1 > \sum_{j=1}^{u} fp_j^2 > \sum_{j=1}^{u} fp_j^3 > 0 \\ 0 < \sum_{j=1}^{v_1} bp_j^1 < \sum_{j=1}^{v_2} bp_j^2 < \sum_{j=1}^{v_3} bp_j^3 < 1 \end{cases}. \tag{5}$$

According to Eqs. (2) and (4), the geometric interpretation of the Bhattacharyya coefficient is the cosine of the angle between the m-dimensional unit vectors $(\sqrt{p_1^i}\ldots\sqrt{p_m^i})$ and $(\sqrt{p_1^j}\ldots\sqrt{p_m^j})$. The smaller angle they have, the more similar the two kernel histograms are. For the target tracking application, this angle is equal to the angle between two 2-D unit vectors: $\mathbf{Z}_i=[(\Sigma_{j=1}^{u}fp_j^i)^{1/2},(\Sigma_{j=1}^{\nu_i}bp_j^i)^{1/2}]$ and $\mathbf{Z}_j=[(\Sigma_{k=1}^{u}fp_k^j)^{1/2},(\Sigma_{k=1}^{\nu_j}bp_k^j)^{1/2}]$. Then, $\rho(i,j)$ can be measured by $\angle(\mathbf{Z}_i,\mathbf{Z}_j)$. Using Eqs. (4) and (5) in conjunction with the geometric relationship, it is clear that $\angle(\mathbf{Z}_3,\mathbf{Z}_1)>\angle(\mathbf{Z}_2,\mathbf{Z}_1)$. Finally, $\rho(2,1)>\rho(3,1)$.

Using theorem 1, we can easily determine that the Bhattacharyya coefficient $\rho(t,1)$ is monotonic decreasing and achieves its maximum in the case where $d(T_t,F_t)=0$. It means the image in T_t $[d(T_t,F_t)=0]$ is most similar to the image in T_1. As long as some parts of the object in the next frame reside inside the kernel, theorem 1 ensures mean-shift iterations converge to the object center.[2,5]

In our experiments, the object kernel histogram computed by the Gaussian kernel has been derived in the RGB space with $32\times32\times32$ bins. Figure 1 shows two video clips where the size of tracking window (white circle) is unchanged. The top row shows the tracking results where

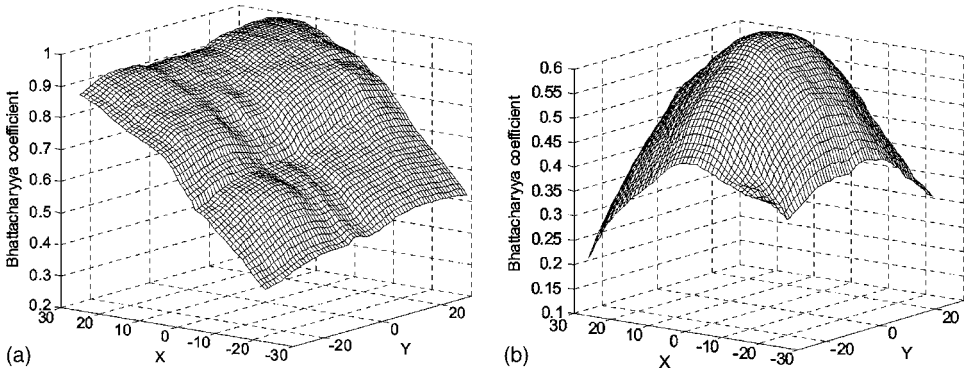

Fig. 2 Surface plot of Bhattacharyya coefficient around the object center.

the object expands its scale, while the bottom row demonstrates the results for the object shrinking its scale. In the first frame of each clip, the initial kernel histogram is obtained from the initial tracking window whose center overlaps the object center. Figure 2 shows the Bhattacharyya coefficients corresponding to the tracking windows centered in a 60×60 neighborhood around the object center. Figures 2(a) and 2(b) correspond to Figs. 1(b) and 1(d), respectively. The Bhattacharyya coefficient in Fig. 2(b) is monotonic decreasing and the maximum corresponds to the object center, which validates our theorem. In the case where the object expands its scale and can not be enwrapped by the tracking window, the monotonic decreasing profile in Fig. 2(b) no longer holds and poor localization potentially occurs; see also top row in Fig. 1. The reason lies in the fact that there are more local maxima in Fig. 2(a) and any location of a tracking window that is too small will yield a similar value of the Bhattacharyya coefficient.

In conclusion, the changes of object scale and position within the fixed kernel will not impact the localization accuracy of the mean-shift tracking algorithm. When the object scale exceeds the size of the tracking window, the tracker outputs poor localization. On the contrary, when the object shrinks its scale, the center of the tracking window locates the object center all the time. Indeed, our previous work[4] for tracking rigid objects with scale changes is based on this conclusion. We hope this paper will valuable for fully solving scaling problems within the mean-shift framework in the future.

References

1. K. Fukanaga and L. D. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans. Inf. Theory* **21**, 32–40 (1975).
2. G. R. Bradski, "Computer vision face tracking for use in a perceptual user interface," in IEEE Workshop on Applications of Computer Vision, p. 214–219, Princeton, NJ (1998).
3. A. Yilmaz, K. Shafique, and M. Shah, "Target tracking in airborne forward looking infrared imagery," *Image Vis. Comput.* **21**, 623–635 (2003).
4. N. S. Peng, J. Yang, and J. X. Chen, "Kernel-bandwidth adaptation for tracking object changing in size," *Lecture Notes in Computer Science* **3212**, 581–588 (2004).
5. D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.* **5**, 564–575 (2003).