Multi-information incorporation approach to kernel-based infrared target model construction with application to target tracking

Jianguo Ling,[a,b] Erqi Liu,[b] Lei Yang,[a] and Jie Yang[a]
[a]Shanghai Jiaotong University, Institute of Image Processing and Pattern Recognition, No. 800 Dongchuan Road, Shanghai, 200240, China
E-mail: lingjianguo76@sjtu.edu.cn
[b]Institute of the Second Academy, China Aerospace Science and Industry Corporation, Beijing, 100854, China

Abstract. We present an approach that incorporates multi-information, including intensity value, spatial relation, and local standard deviation information of the pixels in target region, into kernel density estimation for constructing the kernel-based infrared (IR) target model. The incorporated information can complement each other for a target-tracking task. This constructed target model is evaluated based on the relative entropy of the two classes and is applied in a mean shift tracking system for IR target tracking to verify the effectiveness.
[DOI: 10.1117/1.2388341]

Subject terms: infrared target tracking; target model; kernel density estimation; multi-information incorporation; mean shift algorithm.

Paper 060275LR received Apr. 12, 2006; revised manuscript received Jul. 28, 2006; accepted for publication Sep. 8, 2006; published online Nov. 10, 2006.

1 Introduction

Recently, kernel-based target-tracking methods have received considerable attention in the computer vision field.[1–3] A key issue in the development of those methods is the construction of a target model. Comaniciu et al. designed the target model with an isotropic kernel.[1] Yilmaz et al. defined the target model by cascading two Epanechnikov kernels.[2] Hager et al. constructed the target model with multiple kernels of different tracking structures.[3] IR images are the thermal images that are extremely noisy due to rampant systemic noise or color noise sources incurred by the sensing instrument and the noise from the environment.[4] In most cases, the target region with a common tracker is vaguely located because of the noise. A target model based on the located target region is thus improperly computed. This may cause the tracker to fail to capture the target completely or even to lose the target in the successive tracking process. Thus, it is required to identify a more realistic target model of the IR target for the tracking task. This letter aims to extend the current kernel-based target-tracking method to achieve a robust tracking performance with a well-designed target model.

0091-3286/2006/$22.00 © 2006 SPIE

2 Multi-Information Incorporation Kernel-Based Target Model

Let $\{x_i\}_{i=1\cdots n}$ be the normalized pixel locations in the target region with center c in the current frame. The function $b:R^2\rightarrow\{1\ldots m\}$ (m-bin histograms are used) associates to the pixel at location x_i the index $b(x_i)$ of its bin in the quantized feature space. The probability of the feature (intensity values are commonly used) $u=1\ldots m$ in the target model is computed as[1]

$$q_u = C\sum_{i=1}^{n} k\left(\left\|\frac{x_i-c}{h}\right\|^2\right)\delta[b(x_i)-u], \tag{1}$$

where δ is the Kronecker delta function, C is the normalization constant, $k(\bullet)$ is the common profile used in corresponding feature domain, and h is the kernel bandwidth. Cascading two kernels is another way to estimate the kernel density in the target region.[2,5] In Ref. 5, the kernel is defined as:

$$K_{h_s,h_r}(x) = \frac{C}{h_s^2 h_r^p} k_s\left(\left\|\frac{x^s}{h_s}\right\|^2\right) k_r\left(\left\|\frac{x^r}{h_r}\right\|^2\right), \tag{2}$$

where x^s is the spatial part, x^r is the range part of a feature vector, $k_s(\bullet)$ and $k_r(\bullet)$ are the common profiles used in corresponding domain, h_s and h_r are the employed kernel bandwidths, and p is the image vector dimension. Thus, the probability of the feature $u=1\ldots m$ in the target model is given by

$$q_u = \frac{C}{h_s^2 h_r^p}\sum_{i=1}^{n} k_s\left(\left\|\frac{x_i^s-c}{h_s}\right\|^2\right) k_r\left(\left\|\frac{x_i^r-v}{h_r}\right\|^2\right)\delta[b(x_i)-u], \tag{3}$$

where c and v are the centers of the corresponding kernels. Here $k_s(\bullet)$ is used to define the spatial relation of the intensity values through the Euclidean distance of its spatial position from the target center, and $k_r(\bullet)$ is used as a weighting factor in the intensity values histogram.

Equations (1) and (3), do not pay much attention to the uneven distribution of the intensity values of the pixels in the target region. Moreover, the target center and the kernel bandwidth, which are important parameters for kernel density estimation, are not clearly shown. Here we present a new method for designing a well-performing kernel-based target model. The final target model with the kernel density estimate method incorporates intensity value, spatial relation, and local standard deviation information of the pixels in the target region. Furthermore, the computed kernel density is more approximate to the true distribution of the intensity values of the tracked target.

For an IR image, the local standard deviation of the pixel x_i can be computed as[2]

$$S(x_i) = \left\{\frac{1}{|M|-1}\sum_{X\in M}[I(x_i)-I(X)]^2\right\}^{1/2}, \tag{4}$$

where $I(x_i)$ and $I(X)$ denote the gray values of pixel x_i and pixel X, respectively (pixel X is the pixel around pixel x_i in a predefined window), and $|M|$ denotes number of pixels in the neighborhood. Figure 1 shows the target region and rough contour in the local standard deviation images are

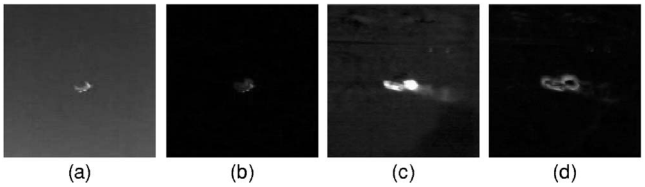

Fig. 1 Original images and the corresponding local standard deviation images: (a) and (c) original IR images and (b) and (d) the corresponding local standard deviation images.

clearly emphasized, and this is an indication that we can use the information to set the target center and the kernel bandwidth. For a discrete 2-D local standard deviation image, the zeroth moment can be defined as

$$M_{00} = \sum_{i=1}^{\text{rows}} \sum_{j=1}^{\text{cols}} S(i,j), \tag{5}$$

where rows and cols are the sizes of the analyzed target region along different orientation, and $S(i,j)$ is the local standard deviation of a pixel at position (i,j). The first moment is given by

$$M_{10} = \sum_{i=1}^{\text{rows}} \sum_{j=1}^{\text{cols}} iS(i,j),$$

$$M_{01} = \sum_{i=1}^{\text{rows}} \sum_{j=1}^{\text{cols}} jS(i,j). \tag{6}$$

Then the component of center $c=(c_x,c_y)$ of kernel $k_s(\bullet)$ is computed as

$$c_x = \frac{M_{10}}{M_{00}},$$

$$c_y = \frac{M_{01}}{M_{00}}. \tag{7}$$

In addition, center v of kernel $k_r(\bullet)$ is defined as the quantized intensity value at position (c_x,c_y). Zeroth-moment information is also used to set the search window size in Ref. 6. Illumined by this work, we set the kernel bandwidth based on a function of the zeroth moment of the local standard deviation image. If the maximum local standard deviation value is denoted as $\delta_{\max}$ in the target region of a certain IR image, the kernel bandwidth $h_s=h_r=(h_x,h_y)$ is defined as

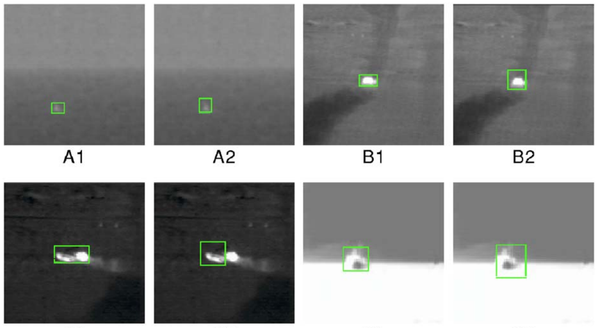

Fig. 2 Eight typical IR images and the marked target regions: A1, B1, C1, and D1, IR images with the appropriate target regions; A2, B2, C2, and D2, target regions poorly located of the corresponding IR images.

$$h_x = \min\left[\alpha\left(\frac{M_{00}}{\delta_{\max}}\right)^{1/2}, \text{rows}\right] \Big/ 2,$$

$$h_y = \min\left[\beta\left(\frac{M_{00}}{\delta_{\max}}\right), \text{cols}\right] \Big/ 2, \tag{8}$$

where α and β are the factors that are determined by our understanding of the target distribution. Thus, the target model representation is then defined by

$$q_u = \frac{C}{h_s^2 h_r^p} \sum_{i=1}^{n} k_s\left(\left\|\frac{x_i^s - c}{h_s}\right\|^2\right) k_r\left(\left\|\frac{x_i^r - v}{h_r}\right\|^2\right) \delta[b(x_i) - u], \tag{9}$$

where $c=(c_x,c_y)$ and $h_s=h_r=(h_x,h_y)$ are computed by Eqs. (7) and (8), respectively; and v is obtained with the value at position (c_x,c_y) in the quantized intensity value space.

3 Experimental Results

In our experiments, an outer margin of 10 pixels from the target region forms the background sample. For the target region, a Gaussian kernel is adopted, while for background region, we use a reverse Gaussian kernel.

Our insight is that the best-designed target model can best distinguish between target and background for a robust tracking task. The discrimination of different target models can be embodied by relative entropy values which are given by

Table 1 Relative entropy values of different target model representations.

	Relative Entropy Values							
Information Used in Kernel Density Estimation	A1	A2	B1	B2	C1	C2	D1	D2
Spatial relation [Eq. (1)]	−8.03	−4.59	−9.11	−5.10	−3.54	−1.76	−3.95	−2.64
Intensity+spatial [Eq. (3)]	−8.15	−4.61	−11.28	−8.09	−2.70	−1.45	−4.60	−2.84
Our method [Eq. (9)]	−11.67	−10.18	−12.67	−11.00	−4.83	−6.62	−4.97	−3.99

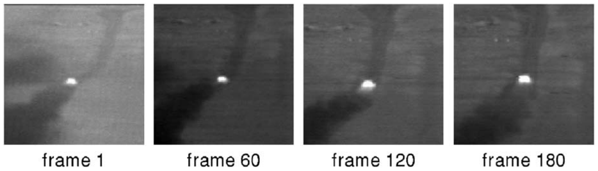

Fig. 3 Test sequence.

$$W(p,b) = -\sum_{u=1}^{m} p(u)\log[p(u)/b(u)] - \sum_{u=1}^{m} b(u)\log[b(u)/p(u)] \tag{10}$$

where p and b are the target kernel density distribution and the background kernel density distribution, respectively. Here $W(p,b)$ is a negative value, and a small $W(p,b)$ means a high separation power for target and background by the corresponding kernel density estimation method. In Fig. 2, eight typical 128×128-pixels IR images are selected to confirm the validity of our approach. The rectangles in the IR images show the target regions. Table 1 shows $W(p,b)$ values of different representations of the target model with several kernel density estimation methods. Here, we find that the method, which incorporates multi-information of target region, is more effective where used in a tracking framework because the discrimination of the target and background indicated by $W(p,b)$ values. Moreover, when the target region is poorly located, the superiority of our method over two other methods is evident.

We also embedded the proposed kernel density estimation in a mean shift tracking system. Figure 3 shows some selected frames of a 180-frame test video sequence where each frame is 128×128 pixels. Here, the intensity space is taken as the feature space and it is quantized into 32 bins. The tracking algorithm with different target model constructions was developed in Matlab7.0 on a Pentium 4 platform. In Figs. 4(a) and 4(b), the rectangles on the left IR images show the initial target bounding box and the plots on the right show the tracking performances of mean shift tracking algorithm with different target model representations. It is shown that the proposed method is more effective to help to track the target with minor prediction errors, and the superior performance is obvious when the initial selected target region is poorly located. Undoubtedly, the additional computational complexity incurred by the proposed target model representation per frame is dominated by the computation of local standard deviation and moment. Based on the target information in the previous frames, we perform this computation in a region that is 2 to 3 pixels larger than the actual target region size. The current implementations of the mean shift tracking algorithm with the initial target bounding box illustrated in Fig. 4(a) are capable of tracking at 15, 14, and 13 frames/s for the target models obtained with Eqs. (1), (3), and (9), respectively. As such, if the tracking algorithm adopts the initial target bounding box shown in Fig. 4(b), the target models represented by Eqs. (1), (3), and (9) enable tracking at frame rates of 15, 14, and 12 frames/s, respectively. From this, we find that the tracking algorithm with the proposed target model construction is competent and a little more complex with respect to computational complexity and cost of implementation.

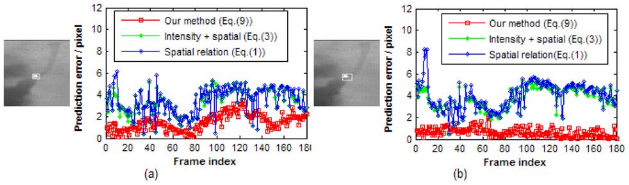

Fig. 4 Prediction errors (between prediction and ground truth) of different target model representations: (a) appropriate target region initially selected (target region is 13×9 pixels) and (b) the initial target region is poorly located (target region is 18×12 pixels).

4 Conclusions

A new method that incorporates multi-information into the kernel density estimation of an IR target model was proposed. The local standard deviation information was designed to select the appropriate target center and kernel bandwidth. This constructed target model was evaluated based on the relative entropy of two classes and applied in a mean shift tracking system for IR target tracking to verify the effectiveness.

Acknowledgments

We would like to thank the anonymous reviewers for their valuable comments. This work is partially supported by Aeronautics Science Fund (China) under Grant No. 04F57004.

References

1. D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(5), 564–577 (2003).
2. A. Yilmaz, K. Shafique, and M. Shah, "Target tracking in airborne forward looking infrared imagery," *Image Vis. Comput.* **21**(7), 623–635 (2003).
3. G. D. Hager, M. Dewan, and C. V. Stewart, "Multiple kernel tracking with SSD," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 790–797 (2004).
4. J. Wei and I. Gertner, "Discrimination, tracking, and recognition of small and fast moving objects," *Proc. SPIE* **4726**, 253–266 (2002).
5. D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(5), 603–619 (2002).
6. G. R. Bradski and C. Santa, "Computer vision face tracking for use in a perceptual user interface," *Intel Technol. J.* **2**(2), 12–21 (1998).