Target tracking in infrared imagery using a novel particle filter

Fanglin Wang (王芳林)[1*], Erqi Liu (刘尔琦)[2], Jie Yang (杨 杰)[1], Shengyang Yu (郁生阳)[1], and Yue Zhou (周 越)[1]

[1]*Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China*

[2]*Institute of the Second Academy, China Aerospace Science and Industry Corporation, Beijing 100854, China*

[*]*E-mail: hardegg@sjtu.edu.cn*

Received September 11, 2008

To address two challenging problems in infrared target tracking, target appearance changes and unpredictable abrupt motions, a novel particle filtering based tracking algorithm is introduced. In this method, a novel saliency model is proposed to distinguish the salient target from background, and the eigenspace model is invoked to adapt target appearance changes. To account for the abrupt motions efficiently, a two-step sampling method is proposed to combine the two observation models. The proposed tracking method is demonstrated through two real infrared image sequences, which include the changes of luminance and size, and the drastic abrupt motions of the target.

OCIS codes: 100.0100, 110.3080, 150.0150.
doi: 10.3788/COL20090707.0576.

Robustly, tracking target in infrared video sequences is usually a challenging problem. Due to the characteristics of infrared imagery, there are mainly two challenging issues: how to model the target to adapt appearance changes in cluttered background, and how to solve the drastic abrupt motions incurred by ego-motions of the sensor.

To model an infrared target, the frequently used observation models include intensity histogram, standard deviation (stdev) histogram[1], edge and shapes[2]. However, it is difficult for these models to solve intensity and size variations. The eigenspace model[3−4] performs robustly to size and intensity variations in visual images, but in infrared ones, it is easily affected by background clutter. To solve the drastic abrupt motions, a naive particle filter need to increase samples, which adds significant computational overhead. To solve this problem, two separate global motion compensation modules are integrated with two separate tracking modules[2]. Venkataraman *et al.*[5] incorporated two dynamic models in the target's kinematics.

In this letter, we solve the above two issues by proposing a novel particle filter based tracking algorithm. There are two main contributions. (1) We propose a novel salience observation model which is effective to distinguish the salient target from background. It is combined with an eigenspace learning based model to adapt target appearance changes. (2) To account for the drastic abrupt motions and retain the efficiency meantime, an effective two-step sampling algorithm is proposed.

The particle filter is a sequential Monte Carlo method to recursively approximate the state $\mathbf{X}_t$ of a system[6−9]. In tracking, the objective is the posterior distribution $p(\mathbf{X}_t|\mathbf{Y}_{1:t})$, where $\mathbf{Y}_{1:t} = (\mathbf{Y}_1, \cdots, \mathbf{Y}_t)$ denote the observations up to current time step. The basic idea of the particle filter is to approximate the posterior $p(\mathbf{X}_t|\mathbf{Y}_{1:t})$ using a set of N samples (also named particles) $\left\{\mathbf{X}_t^{(i)}\right\}_{i=1}^{N}$ with importance weights $\left\{w_t^{(i)}\right\}_{i=1}^{N}$. The samples are drawn from a proposal distribution $q(\mathbf{X}_t|\mathbf{X}_{1:t-1}, \mathbf{Y}_{1:t})$ which may depend on the old state and the new measurements. The weight is recursively updated as

$$w_t^{(i)} = w_{t-1}^{(i)} \frac{p(\mathbf{Y}_t|\mathbf{X}_t^{(i)}) p(\mathbf{X}_t^{(i)}|\mathbf{X}_{t-1}^{(i)})}{q(\mathbf{X}_t^{(i)}|\mathbf{X}_{1:t-1}^{(i)}, \mathbf{Y}_{1:t})}. \tag{1}$$

During tracking, it is necessary to resample the particles to avoid the degeneracy problem. In practical applications, $q(\mathbf{X}_t|\mathbf{X}_{1:t-1}, \mathbf{Y}_{1:t})$ is often chosen to be the system dynamic model $p(\mathbf{X}_t|\mathbf{X}_{t-1})$, and the weights become the observation likelihood $p(\mathbf{Y}_t|\mathbf{X}_t)$. The final state $\mathbf{X}_t$ is often estimated as

$$\hat{\mathbf{X}}_t = \sum_{i=1}^{N} w_t^{(i)} \mathbf{X}_t^{(i)}. \tag{2}$$

Here the state vector of the particle filter is defined as $\mathbf{X} = (\mathbf{x}, s)$, where $\mathbf{x} = (x, y)$ indicates the location in the frame image and s is the scale of the target.

The observation model is used to measure the observation likelihood of samples. Combining multiple models has been proved to be effective to improve tracking performance. Here we combine saliency model and eigenspace model. The proposed saliency model can distinct the target from its surrounding background. The eigenspace model can solve the target appearance changes. Such a combination significantly improves robustness and accuracy.

Spectral residual based salience detection algorithm[7] (SRSD) is a general purpose detection algorithm. It is based on the hypothesis that the log Fourier spectrum of different images share similar trends, though each containing statistical singularities. Those singularities are

Table 1. Algorithm of the Proposed Particle Filter with Two-Step Sampling

1. Initialization: at time $t = 0$, select the target's initial state $\mathbf{X}_0$ manually; for $i = 1, ..., N$, set the initial sample set $\left\{\mathbf{X}_0^{(i)}, 1/N\right\}$ by assigning $\mathbf{X}_0^{(i)} = \mathbf{X}_0$, and initialize the eigenspace model according to Ref. [4].

2. Frame tracking: for time $t = 1, 2, ...$, repeat the following steps

(1) **Two-step sampling**:

- **Prediction**: for $i = 1, ..., N$, draw $\mathbf{X}_t^{\mathrm{s},(i)}$ according to Eq. (8), $\mathbf{X}_t^{\mathrm{s},(i)}$ should be diffused enough to cover the expected range of possible states.
- **Weighting samples** with saliency model: for $i = 1, ..., N$, compute $w_t^{\mathrm{s},(i)} = p(\mathbf{Y}_t^{\mathrm{s}}|\mathbf{X}_t^{\mathrm{s},(i)})$ according to Eq. (5), and normalize them so that $\sum_{i=1}^{N} w_t^{\mathrm{s},(i)} = 1$.
- **Resample** on sample set $\left\{\mathbf{X}_t^{\mathrm{s},(i)}, w_t^{\mathrm{s},(t)}\right\}_{i=1}^{N}$, new sample set $\left\{\tilde{\mathbf{X}}_t^{\mathrm{s},(i)}, 1/N\right\}_{i=1}^{N}$ is obtained.
- **Diffuse** the obtained samples to maintain more state hypothesizes, draw new samples $\mathbf{X}_t^{\mathrm{e},(i)}$ according to Eq. (9).
- **Weighting samples** with eigenspace model: $w_t^{\mathrm{e},(i)} = p(\mathbf{Y}_t^{\mathrm{e}}|\mathbf{X}_t^{\mathrm{e},(i)})$ according to Eq. (6), and normalize them.

(2) **Estimate** $\mathbf{X}_t$: output $\hat{\mathbf{X}}_t = \sum_{i=1}^{N} w_t^{\mathrm{e},(i)} \mathbf{X}_t^{\mathrm{e},(i)}$.

(3) **Resample** on sample set $\left\{\mathbf{X}_t^{\mathrm{e},(i)}, w_t^{\mathrm{e},(i)}\right\}_{i=1}^{N}$ and output the obtained $\left\{\mathbf{X}_t^{(i)}, 1/N\right\}_{i=1}^{N}$.

(4) **Incremental learning**: using $\hat{\mathbf{Y}}_t^{\mathrm{e}}$ update eigenspace model. $\hat{\mathbf{Y}}_t^{\mathrm{e}}$ is the image patch vector corresponding to the estimated state $\hat{\mathbf{X}}_t$.

regarded to be caused by salient objects. The detector detects these salient objects with

$$I^s(\mathbf{x}) = g(\mathbf{x}) * F^{-1}[\exp\left(\mathrm{P}(f) + \Pi(f)\right)]^2, \qquad (3)$$

where F^{-1} denotes the inverse Fourier transform, $\Pi(f)$ denotes the phase spectrum of the Fourier transform, $P(f)$ is the spectral residual defined, and $g(\mathbf{x})$ is a Gaussian filter (see Ref. [7] for more details). The obtained I^{s} is the saliency map image.

Through vast experiments, we found SRSD was effective in most infrared images. In Fig. 1 (b), we show the saliency map image generated by Eq. (3). As a comparison, the local standard deviation image[1] is also presented. Median filtering is previously operated to ease the background noise. As can be seen, the target regions are clearly emphasized and meantime the background regions are suppressed.

Now we show how to build the saliency model. Suppose we have obtained the salience map image I^{s}, as shown in Fig. 1(b). Given the state vector $\mathbf{X}$, a rectangular target region M centered at the location $\mathbf{x}$ is obtained in I^{s}. We compute the average intensity firstly by

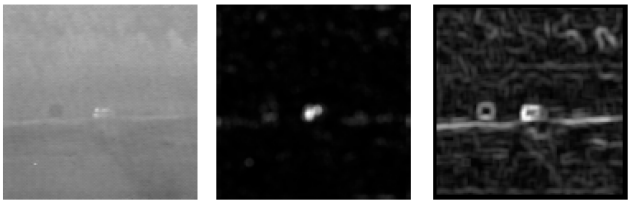
$$\mu(\mathbf{x}) = \frac{1}{|M|} \sum_{\mathbf{x}_i \in M} I^{\mathrm{s}}(\mathbf{x}_i), \qquad (4)$$

where $I^{\mathrm{s}}(\mathbf{x}_i)$ corresponds to the intensity value at location $\mathbf{x}_i$, and $|M|$ denotes the number of pixels in region M. Then we use a Gaussian distribution to describe the saliency likelihood as

$$p(\mathbf{Y}^{\mathrm{s}}|\mathbf{X}) \propto \exp\left(-\frac{\left(\mu(\mathbf{x})/I_{\max}^{\mathrm{s}} - 1\right)^2}{2\sigma_{\mathrm{s}}^2}\right), \qquad (5)$$

where $I_{\max}^{\mathrm{s}}$ is the max intensity value in I^{s}, and σ_{s} is an empirically chosen value. Equation (5) defines a rule that if a candidate region is the target, it must be firstly a salient target in the image, and the model describes how salient a target candidate is in the current image.

For eigenspace model[4], we model image observations using a probabilistic interpretation of principal component analysis. Suppose at time t, the observation matrix $\mathbf{\Psi}_t = \{\mathbf{Y}_1, \cdots, \mathbf{Y}_t\}$ is obtained, here $\mathbf{Y}_t$ is the target image patch decided by state $\mathbf{X}_t$, then the sigular value decomposition (SVD) $\mathbf{\Psi}_t = \mathbf{U}_t \mathbf{\Sigma}_t \mathbf{V}_t^T$ and the mean vector $\mu_t = \sum_{i=1}^{t} \mathbf{Y}_i$ can be computed. Accordingly, the target's eigenspace observation is assumed to be located in a subspace centered at μ_t and spanned by $\mathbf{U}_t$. To adapt the appearance changes, the incremental learning method designed[4] is invoked to learn the appearance changes online. The eigenspace observation density is defined as the distance from a sample to the eigen space. Suppose $\mathbf{U}$ and μ are the eigenvectors and mean vector, the



(a) (b) (c)

Fig. 1. (a) Input image; (b) saliency map image; (c) local standard deviation image in Ref. [1]. In (b) and (c), the high intensity corresponds to a high saliency or local standard deviation value.

likelihood of eigenspace mode is defined as[4]

$$p(\mathbf{Y}^{\mathrm{e}}|\mathbf{X}) \propto \exp\left(-\frac{\| (\mathbf{Y}^{\mathrm{e}}-\boldsymbol{\mu}) - \mathbf{U}\mathbf{U}^{\mathrm{T}}(\mathbf{Y}^{\mathrm{e}}-\boldsymbol{\mu})\|^2}{2\sigma_{\mathrm{e}}^2}\right), \tag{6}$$

where σ_{e} is a fixed value chosen empirically.

Sometimes, target motions become drastic and unpredictable due to ego-motion of the sensor. The standard importance sampling (IS) used in particle filter (PF) will lead to gradual departure of the sample set from the real target state, which eventually results in tracking loss. So if we simply combine the two features as

$$p(\mathbf{Y}|\mathbf{X}) = p(\mathbf{Y}^{\mathrm{s}}|\mathbf{X})p(\mathbf{Y}^{\mathrm{e}}|\mathbf{X}), \tag{7}$$

the above mentioned problem can be solved by increasing samples. However, that will decrease the efficiency. To solve the problem efficiently, we propose a two-step sampling method. The algorithm is illustrated detailedly in Table 1.

In the prediction step, we use a random walk model as the dynamic model

$$\mathbf{X}_t^{\mathrm{s}} = \mathbf{X}_{t-1} + \mathbf{v}_t^{\mathrm{s}}, \tag{8}$$

where the 3×1 vector $\mathbf{v}_t^{\mathrm{s}}$ denotes a random Gaussian noise with zero mean and fixed variance $\left(\sigma_{\mathrm{s},x}^2, \sigma_{\mathrm{s},y}^2, \sigma_{\mathrm{s},s}^2\right)$. The variance of scale is set to 0, i.e., $\sigma_{\mathrm{s},s}^2 = 0$ since the salience model can not accurately describe the target size variation. Another dynamic model is used to diffuse the sample set obtained after re-sampling the samples at prediction step

$$\mathbf{X}_t^{\mathrm{e}} = \mathbf{X}_t^{\mathrm{s}} + \mathbf{v}_t^{\mathrm{e}}, \tag{9}$$

where $\mathbf{v}_t^{\mathrm{e}}$ also stands for a random Gaussian noise with fixed variance $\left(\sigma_{\mathrm{e},x}^2, \sigma_{\mathrm{e},y}^2, \sigma_{\mathrm{e},s}^2\right)$. Here $\sigma_{\mathrm{e},x}^2, \sigma_{\mathrm{e},y}^2$ are small values to diffuse the sample set for maintaining more state hypothesizes, and $\sigma_{\mathrm{e},s}^2$ is not 0 to capture target size changes.

Our method is evaluated on two real infrared sequences. The experiments are carried out using MATLAB 7.0 software, on a personal computer with Pentium IV 2.4-GHz central processing unit (CPU) and 512-M random-access memory (RAM). The tracking targets are selected manually in the first frame. The first sequence, containing 120 frames with size of 160×120, is about a flying plane, and the second one is a car in 320 frame images sized of 128×128. For the two sequences, the initial target sizes are chosen as 22×12 and 12×10, the numbers of samples N are 100 and 200, and the average tracking speed of our algorithm are 6.9 and 7.1 fps, respectively. The results of intensity histogram model based particle filter (HMPF)[8] and eigenspace model based particle filter (EMPF)[4] are also presented as comparisons. The variances for dynamic models in HMPF and EMPF are both set as $\left(5^2, 5^2, 0.01^2\right)$. The variances for our algorithm depicted by Eqs. (8) and (9) are set as $\left(5^2, 5^2, 0^2\right)$ and $\left(1^2, 1^2, 0.01^2\right)$, respectively. For each setting, we repeat each algorithm ten times to get a statistical reflection of the behavior of the algorithms.

In the sequence shown in Fig. 2, a plane poses illumination changes and so does the background. Figure 2(a) shows tracking results of EMPF with one hundred samples. The method is effective when target appearance just changes, but it may fail when the background changes drastically (such as frame 24). As shown in Fig. 2(b), HMPF with 100 samples also fails quickly when the background intensity changes. Although with 100 samples, our algorithm is robust to the background changes thanks to the saliency model, and is robust to illumination and size changes due to the eigenspace model. Figure 3 plots the RMSE error which is computed by $\sqrt{(x_t - \hat{x}_t)^2 + (y_t - \hat{y}_t)^2}$, where (x_t, y_t) and $(\hat{x}_t, \hat{y}_t)$ are the ground truth and the estimated location, respectively. For EMPF, we increase samples to 500 until it can give satisfying results. The average RMSEs for HMPF, EMPF, EMPF with 500 samples, and our algorithm are 28.07, 31.53, 1.40, and 1.45, respectively. As can be seen, our algorithm can present robust and accurate tracking results.

The second sequence is sequence "rng17_01" in the AMCOM data set. To better evaluate the ability of solving drastic abrupt motions, we do tracking once every three frames. The results of EMPF with 500 samples are also shown for comparison. In Fig. 4(a), the solid and dashed black rectangles stand for tracking results in current and previous frame respectively, where we can find target displacements between two consequent frames are frequently very large. Figure 5 shows the trajectories obtained through the ground truth, as well as the estimated positions obtained by EMPF and our algorithm. Even though with 500 samples, EMPF loses the target quickly and can not resume right tracking. Only with 200 samples, our algorithm can deal with the drastic abrupt motions well (such as frame 16, 55, 70, and 82), and adapt the illumination and size changes (such as frame 244 and 316). Figure 4(c) illustrates the two-step sampling process when drastic abrupt motion happens, where the black rectangle in latter two images stands for a sample. We can see that the samples drawn in prediction step are widely diffused to cover possible locations, but only little percentage of them are effective.

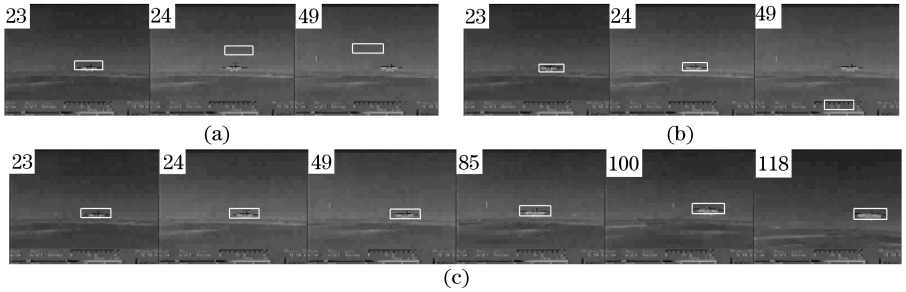


Fig. 2. Tracking results on sequence "plane". (a), (b), and (c) are the tracking results of EMPF, HMPF, and our algorithm.

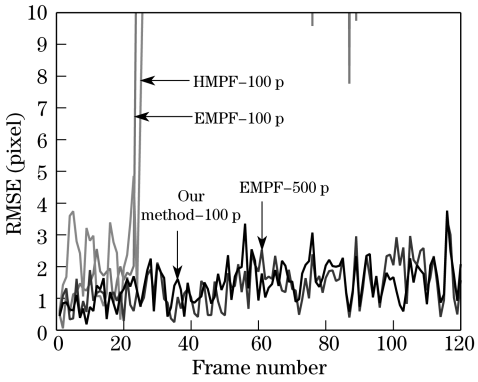


Fig. 3. Plot of RMSE error.

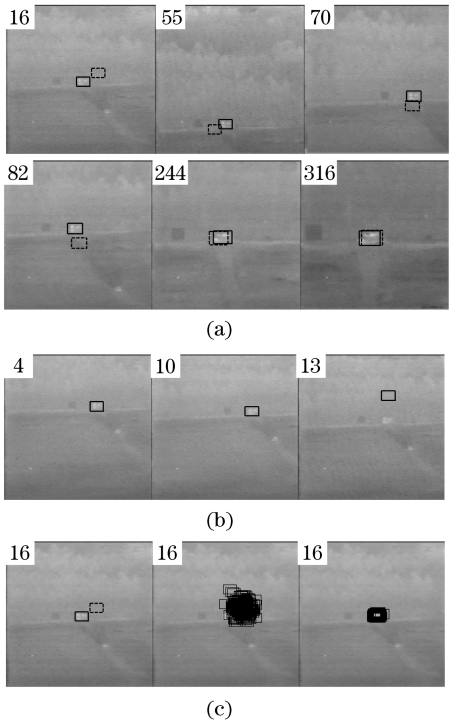


Fig. 4. Tracking results on sequence "rng17_01". (a) Our algorithm with 200 samples; (b) EMPF with 500 samples; (c) the proposed two-step sampling process: tracking result, the first step sampling, and second step sampling.

After the second sampling, most samples are effective, which helps to lead to satisfying tracking results.

In conclusion, a novel infrared target tracking method is proposed. Two different observation models are used to describe the target. The proposed saliency model can distinguish the target from the background surroundings, and the invoked eigenspace model can solve the target appearance changes. To account for the abrupt motions, the two complementary observation models are combined efficiently within a two-step sampling method in a particle filter framework. The experimental results demonstrate that the proposed tracker is robust to appearance changes and abrupt motions.

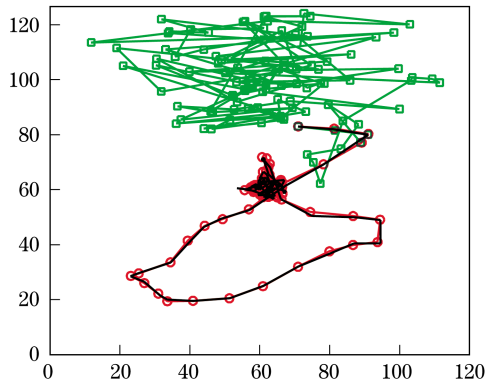


Fig. 5. Trajectories on sequence "rng17_01" every 3 frames. The thick line denotes the true trajectory, the line with marked square is obtained with EMPF, and the one with marked circle is obtained by our algorithm.

The authors are thankful to Xiaodi Hou for his valuable discussion and the centre of imaging science, ARO DAAH049510494 for providing the AMCOM infrared data sets. This work was supported by the National "863" Project of China (No. 2007AA01Z164) and the National Natural Science Foundation of China (Nos. 60675023 and 60602012).

References

1. A. Yilmaz, K. Shafique, and M. Shah, Image and Vision Computing **21,** 623 (2003).
2. J. S. Shaik and K. M. Iftekharuddin, in *Proceedings of IEEE International Joint Conference on Neural Networks* 1201 (2003).
3. M. J. Black and A. D. Jepson, International Journal Computer Vision **26,** 63 (1998).
4. D. A. Ross, J. Lim, R.-S. Lin and M.-H. Yang, Int. J. Vis. **77,** 125 (2008).
5. V. Venkataraman, G. Fan, and X. Fan, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* 3466 (2007).
6. M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, IEEE Transaction on Signal Processing **50,** 174 (2002).
7. X. Hou and L. Zhang, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* 2280 (2007).
8. J. Cheng, Y. Zhou, and N. Cai, J. Infrared Millim. Waves (in Chinese) **25,** 113 (2006).
9. B. Zhang, W. Tian, and Z. Jin, Chin. Opt. Lett. **4,** 569 (2006).