Journal of Southeast University (English Edition) Vol. 23, No. 2, pp. 196 – 201 June 2007 ISSN 1003—7985

Face tracking algorithm based on particle filter with mean shift importance sampling

Gao Jianpo Yang Hao An Guocheng Wu Zhenyang

(School of Information Science and Engineering, Southeast University, Nanjing 210096, China)

Abstract: The condensation tracking algorithm uses a prior transition probability as the proposal distribution, which does not make full use of the current observation. In order to overcome this shortcoming, a new face tracking algorithm based on particle filter with mean shift importance sampling is proposed. First, the coarse location of the face target is attained by the efficient mean shift tracker, and then the result is used to construct the proposal distribution for particle propagation. Because the particles obtained with this method can cluster around the true state region, particle efficiency is improved greatly. The experimental results show that the performance of the proposed algorithm is better than that of the standard condensation tracking algorithm.

Key words: face tracking; particle filter; importance sampling; condensation; mean shift

Nowadays, face tracking in video sequences attracts more and more attention in the computer vision fields. In essence, face tracking is a dynamic estimation problem, that is to say, the task of face tracking is to recover the unknown values (position and scale) of the face target using the observed video sequence. Recently, the particle filter has gained prevalence in the tracking literature, which provides a robust tracking framework as it requires neither the system to be linear nor the noise to be Gaussian.

However, the efficiency and accuracy of the particle filter depend drastically on the proposal distribution used to reallocate the particles. Currently, there are many ways to design the proposal distribution[1-6], but unfortunately, maybe they are not suitable for object tracing because the choice of a proposal distribution always depends on what special problems need to be solved. In the general condensation algorithm[7] for visual tracking, the transition prior probability is used as the proposal distribution, which is simple, but the main shortcoming is that the most recent observation is neglected, this often results in poor performance.

This paper proposes a new particle filter with mean shift importance sampling for face tracking, and it uses the mean shift tracker to construct the proposal distribution. The essential idea is to make the proposal distribution admit the most recent observation via the mean shift tracker. The experimental results demonstrate the efficiency of this algorithm for face tracking.

1 Particle Filter and Condensation

The particle filter[1-2] is a useful tool to perform dynamic state estimation via Bayesian inference. Its main task is to estimate the unknown state vector $\boldsymbol{x}$ from a collection of observations $\boldsymbol{Z} = \{z_1, z_2, \dots, z_t\}$. The particle filter is composed of two important components:

State transition model

$$\boldsymbol{x}_t = f(\boldsymbol{x}_{t-1}) + \boldsymbol{v}_t \longleftrightarrow p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}) \tag{1a}$$

Observation model

$$z_t = h(\boldsymbol{x}_t) + n_t \longleftrightarrow p(z_t \mid \boldsymbol{x}_t) \tag{1b}$$

The whole inference process of particle filter comprises two stages:

Prediction

$$p(\boldsymbol{x}_t \mid \boldsymbol{Z}_{t-1}) = \int p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}) p(\boldsymbol{x}_{t-1} \mid \boldsymbol{Z}_{t-1}) \mathrm{d}\boldsymbol{x}_{t-1} \tag{2a}$$

Update

$$p(\boldsymbol{x}_t \mid \boldsymbol{Z}_t) = K_t p(z_t \mid \boldsymbol{x}_t) p(\boldsymbol{x}_t \mid \boldsymbol{Z}_{t-1}) \tag{2b}$$

The key idea of the particle filter is to represent the posterior distribution by a weighted particles set $\boldsymbol{S} = \{(\boldsymbol{x}^{(n)}, w^{(n)}) \mid n = 1, 2, \dots, N\}$.

For the first order Markov process, suppose that we can obtain the particles set $\{\boldsymbol{x}_t^{(i)} \mid i = 1, 2, \dots, N\}$ at time t via a proposal distribution $q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, z_t)$, the particles can be weighted as[1]

$$w_t^i \propto w_{t-1}^i \frac{p(z_t \mid \boldsymbol{x}_t^i) p(\boldsymbol{x}_t^i \mid \boldsymbol{x}_{t-1}^i)}{q(\boldsymbol{x}_t^i \mid \boldsymbol{x}_{t-1}^i, z_t)} \tag{3}$$

Condensation[7] is a special version of a particle filter for object tracking in video sequences, which uses the transition prior probability $p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$ as a proposal distribution. Roughly speaking, the realization of

Received 2006-11-13.

Foundation item: The National Natural Science Foundation of China (No. 60672094).

Biographies: Gao Jianpo (1975—), male, graduate; Wu Zhenyang (corresponding author), male, professor, zywu@seu.edu.cn.

condensation comprises mainly three steps: resampling, dynamic prediction, and observation update. Fig. 1 shows the diagram of condensation.

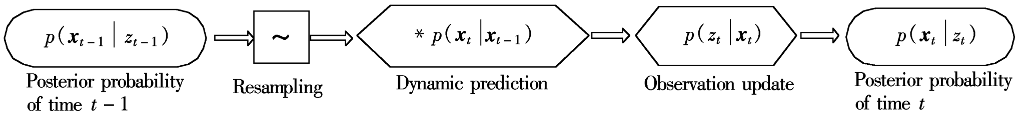

Fig. 1 The diagram of condensation

2 Particle Filter with Mean Shift Importance Sampling

2.1 Construct proposal distribution using mean shift tracker

In the aforementioned condensation tracking algorithm, a transition prior probability $p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$ is used as the proposal distribution for visual tracking. The main shortcoming of this kind of proposal distribution is that the most recent observation is neglected. It frequently results in poor tracking performance. In order to improve the tracking performance, a better proposal distribution should be chosen. Obviously, the most recent observation can provide useful information for the proposal distribution, if the proposal distribution can embody the most recent observation, it is likely to direct particles to the ideal place.

Now the problem is how to make the proposal distribution embody the most recent observation. Assume that we can get the coarse target state estimation using the most recent observation. Then we can use this coarse state estimation to construct the proposal distribution. Mean shift tracker[8] is an ideal choice to fulfill the above task due to its high efficiency for object tracking.

Mean shift[8-9] is an effective method for mode seeking in probability space, which is based on the theory of nonparametric kernel probability density estimation. Comaniciu et al. proposed a simple and efficient object tracking algorithm[9] based on the mean shift theory, in which the kernel histogram is used as the tracking cue.

Let $\{\boldsymbol{x}_i^*\}_{i=1,2,\ldots,n}$ represent the pixel locations relative to the center of the target model region, and $u = 1, 2, \ldots, m$ be the bin index of the color histogram; then the kernel color histogram of the target model can be denoted as

$$\hat{q}_u = C \sum_{i=1}^{n} k(\|\boldsymbol{x}_i^*\|^2)\delta[b(\boldsymbol{x}_i^*) - u] \tag{4a}$$

$$C = \frac{1}{\sum_{i=1}^{n} k(\|\boldsymbol{x}_i^*\|^2)} \tag{4b}$$

where δ is Kronecker delta function, and $b(\boldsymbol{x}_i^*)$ is the function that gives the bin index of point $\boldsymbol{x}_i^*$.

In the same way, let $\{\boldsymbol{x}_i\}_{i=1,2,\ldots,n_h}$ be the pixel locations relative to the center $\boldsymbol{x}$ of the candidate target region; then the kernel color histogram of the candidate target can be denoted as

$$\hat{p}_u(\boldsymbol{x}) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{\boldsymbol{x} - \boldsymbol{x}_i}{h}\right\|^2\right)\delta[b(\boldsymbol{x}_i) - u] \tag{5a}$$

$$C_h = \frac{1}{\sum_{i=1}^{n_h} k\left(\left\|\frac{\boldsymbol{x} - \boldsymbol{x}_i}{h}\right\|^2\right)} \tag{5b}$$

The similarity of the target model and the candidate target can be computed using the following Bhattacharyya coefficient.

$$\hat{\rho}(\boldsymbol{x}) = \rho[\hat{p}(\boldsymbol{x}), \hat{q}] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\boldsymbol{x})\hat{q}_u} \tag{6}$$

Searching the target in frame t is essentially to seek $\boldsymbol{x}$ that makes Eq. (6) maximal.

To assume $\hat{\boldsymbol{x}}_0$ is the target location in frame $t-1$, and $\{\hat{p}_u(\hat{\boldsymbol{x}}_0)\}_{u=1,2,\ldots,m}$ is its kernel color histogram. When the kernel color histogram $\{\hat{p}_u(\hat{\boldsymbol{x}})\}_{u=1,2,\ldots,m}$ in frame t does not change drastically from $\{\hat{p}_u(\hat{\boldsymbol{x}}_0)\}_{u=1,2,\ldots,m}$, we can search the target in frame t by the following iterative process $\hat{\boldsymbol{x}}_0 \to \hat{\boldsymbol{x}}_1 \to \hat{\boldsymbol{x}}_0$[9]:

$$\hat{\boldsymbol{x}}_1 = \frac{\sum_{i=1}^{n_h} \boldsymbol{x}_i w_i g\left(\left\|\frac{\hat{\boldsymbol{x}}_0 - \boldsymbol{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} w_i g\left(\left\|\frac{\hat{\boldsymbol{x}}_0 - \boldsymbol{x}_i}{h}\right\|^2\right)} \tag{7}$$

Suppose that the coarse tracking result given by the mean shift tracker at time t is $\boldsymbol{x}_{t0} = \{x_{t0}, y_{t0}, s_{t0}\}$, where x_{t0}, y_{t0}, and s_{t0} are the location and scale of the face target; then we can use this result to construct the proposal distribution of the particle filter with mean shift importance sampling as follows:

$$q(\boldsymbol{x}_t) = \frac{1}{(2\pi)^{\frac{3}{2}}|\Sigma|^{\frac{1}{2}}} \cdot \exp\left[-\frac{1}{2}(\boldsymbol{x}_t - \boldsymbol{x}_{t0})\Sigma^{-1}(\boldsymbol{x}_t - \boldsymbol{x}_{t0})^{\mathrm{T}}\right] \tag{8}$$

where $\boldsymbol{x}_t$ is the state vector of the tracked face, and Σ is the proposal distribution covariance which can be determined in accordance with the reliability of the mean shift tracker or the empirical knowledge.

2.2 Weight the particles

When we use the proposal distribution of Eq. (8) to produce the particles of time t, it is obvious that the point to point propagation rule in the standard particle filter is broken, so the particle weighted formula (3) should be modified. According to the theory of the Bayesian filter and importance sampling[1,10], if we use the proposal distribution $q(\boldsymbol{x}_t)$ given by Eq. (8) to generate a particles set $\{\boldsymbol{x}_t^{(i)} \mid i=1,2,\ldots,N\}$ at time t, then the weight of the i-th particle $\boldsymbol{x}_t^i$ can be computed as

$$w_t^i = \frac{p(z_t \mid \boldsymbol{x}_t^i) p(\boldsymbol{x}_t^i \mid z_{t-1})}{q(\boldsymbol{x}_t^i)} \tag{9a}$$

$$p(\boldsymbol{x}_t^i \mid z_{t-1}) = \sum_{j=1}^{N} p(\boldsymbol{x}_{t-1}^j \mid z_{t-1}) p(\boldsymbol{x}_t^i \mid \boldsymbol{x}_{t-1}^j) = \sum_{j=1}^{N} w_{t-1}^j \, p(\boldsymbol{x}_t^i \mid \boldsymbol{x}_{t-1}^j) \tag{9b}$$

From Eqs. (9a) and (9b) we can see that, for each particle's weight computation, the sum in Eq. (9b) must be evaluated. It is more complex than the standard particle filter. In order to reduce the computational burden, we give the following theorem.

Theorem 1 Suppose that $\sum_{i=1}^{N} w_{t-1}^i = 1, \bar{\boldsymbol{x}}_{t-1} = \sum_{i=1}^{N} w_{t-1}^i \boldsymbol{x}_{t-1}^i$, $\boldsymbol{x}_t$ is fixed, $\hat{p}(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$ is the first order Taylor expansion of $p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$ around $\bar{\boldsymbol{x}}_{t-1}$, the residual error of the Taylor expansion at point $\boldsymbol{x}_{t-1}^i$ is R_t^i, then $\sum_{i=1}^{N} w_{t-1}^i \, p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}^i) = p(\boldsymbol{x}_t \mid \bar{\boldsymbol{x}}_{t-1}) + \sum_{i=1}^{N} w_{t-1}^i \, R_t^i$.

Proof When $\boldsymbol{x}_t$ is fixed and $\boldsymbol{x}_{t-1}$ is variable, $p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$ is the function of variable $\boldsymbol{x}_{t-1}$. For clarity, $p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$ and $\hat{p}(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$ can be denoted as $p_{x_t}(\boldsymbol{x}_{t-1})$ and $\hat{p}_{x_t}(\boldsymbol{x}_{t-1})$, respectively. Because $\hat{p}(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$ is the first order Taylor expansion approximation of $p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$, $p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}^i)$ can be expressed as

$$p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}^i) = p_{x_t}(\boldsymbol{x}_{t-1}^i) = \hat{p}_{x_t}(\boldsymbol{x}_{t-1}^i) + R_t^i = p_{x_t}(\bar{\boldsymbol{x}}_{t-1}) + \frac{\partial p_{x_t}(\boldsymbol{x}_{t-1})}{\partial \boldsymbol{x}_{t-1}}\bigg|_{x_{t-1}=\bar{x}_{t-1}} (\boldsymbol{x}_{t-1}^i - \bar{\boldsymbol{x}}_{t-1}) + R_t^i \tag{10}$$

Then we can obtain

$$\begin{aligned}\sum_{i=1}^{N} w_{t-1}^i p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}^i) &= \sum_{i=1}^{N} w_{t-1}^i [\hat{p}_{x_t}(\boldsymbol{x}_{t-1}^i) + R_t^i] = \\ \sum_{i=1}^{N} w_{t-1}^i \, p_{x_t}(\bar{\boldsymbol{x}}_{t-1}) &+ \sum_{i=1}^{N} w_{t-1}^i \frac{\partial p_{x_t}(\boldsymbol{x}_{t-1})}{\partial \boldsymbol{x}_{t-1}}\bigg|_{x_{t-1}=\bar{x}_{t-1}} (\boldsymbol{x}_{t-1}^i - \bar{\boldsymbol{x}}_{t-1}) + \\ \sum_{i=1}^{N} w_{t-1}^i \, R_t^i &= p_{x_t}(\bar{\boldsymbol{x}}_{t-1}) + \frac{\partial p_{x_t}(\boldsymbol{x}_{t-1})}{\partial \boldsymbol{x}_{t-1}}\bigg|_{x_{t-1}=\bar{x}_{t-1}} \left(\sum_{i=1}^{N} w_{t-1}^i \, \boldsymbol{x}_{t-1}^i - \bar{\boldsymbol{x}}_{t-1}\right) + \\ \sum_{i=1}^{N} w_{t-1}^i R_t^i &= p_{x_t}(\bar{\boldsymbol{x}}_{t-1}) + \sum_{i=1}^{N} w_{t-1}^i \, R_t^i = p(\boldsymbol{x}_t \mid \bar{\boldsymbol{x}}_{t-1}) + \sum_{i=1}^{N} w_{t-1}^i \, R_t^i\end{aligned} \tag{11}$$

Based on theorem 1, we can approximate Eq. (9b) in the sense of the first order Taylor expansion as

$$p(\boldsymbol{x}_t^i \mid z_{t-1}) = \sum_{j=1}^{N} w_{t-1}^j \, p(\boldsymbol{x}_t^i \mid \boldsymbol{x}_{t-1}^j) \approx p\left(\boldsymbol{x}_t^i \,\middle|\, \sum_{j=1}^{N} w_{t-1}^j \, \boldsymbol{x}_{t-1}^j\right) = p(\boldsymbol{x}_t^i \mid \bar{\boldsymbol{x}}_{t-1}) \tag{12}$$

Then Eq. (9a) becomes

$$w_t^i \approx \frac{p(z_t \mid \boldsymbol{x}_t^i) p(\boldsymbol{x}_t^i \mid \bar{\boldsymbol{x}}_{t-1})}{q(\boldsymbol{x}_t^i)} \tag{13}$$

Obviously, this reduces the computational complexity greatly.

3 Face Tracking Schemes

3.1 State transition model

Define $\boldsymbol{x} = \{x, y, s\}$ as the state vector, where x, y is the face target location, and s is the face scale (The facial shape is regarded as the ellipse of 1∶1.2). In order to alleviate the poor effects of an imperfect proposal distribution constructed via the mean shift tracker due to a poor environment, the particle allocation mechanism of our face tracking system comprises three forms: importance sampling via the proposal distribution in Eq. (8), prior uniform distribution, and prior transition, as expressed by Eq. (14). This kind of sampling mechanism is similar to ICONDENSATION[10]; however, our importance sampling method is very different (ours is based on the mean shift tracker) and the probability (t_1) of importance sampling is adaptive.

$$p(\boldsymbol{x}_t \mid z_{t-1}) = \begin{cases} q(\boldsymbol{x}_t) & \alpha < t_1 \\ U(\boldsymbol{x}_t) & t_1 \leqslant \alpha < t_1 + t_2 \\ p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}) & t_1 + t_2 \leqslant \alpha \end{cases} \tag{14}$$

where t_1 and t_2 are the parameters that decide the probability of three sampling forms; α is a random number of uniform distribution in [0, 1]; $q(\boldsymbol{x}_t)$ is the proposal distribution function given by Eq. (8); $U(\boldsymbol{x}_t)$ denotes a uniform distribution in a specified region. In our face tracking system, the probability of importance sampling t_1 is not fixed, that is to say, it can be adjusted according to the reliability of the proposal distribution. The higher similarity measure between the candidate region decided by mean shift tracker and the face model, the larger t_1 is, and vice versa. It is obvious that this adaptive mechanism can alleviate the negative influence due to the imperfect proposal distribution.

3.2 Observation model

In the observation update step we weight each particle according to the current observations. The first problem in this step we face is the choice of tracking cues. In this paper, color and shape are chosen as the tracking cues.

3.2.1 Color model

We choose a color histogram in HSV space to describe the face target. In order to alleviate the effect of illumination, we use fewer bins for the V component when the color histogram is established, specifically, HSV color space is quantized as $8 \times 8 \times 4$ bins.

Suppose that $\boldsymbol{q}_{\mathrm{c}}$ and $\boldsymbol{p}_{\mathrm{c}}$ are the color histogram of the face model and candidate region decided by the state vector $\boldsymbol{x}_t$, respectively. The similarity measure between $\boldsymbol{q}_{\mathrm{c}}$ and $\boldsymbol{p}_{\mathrm{c}}$ can be computed using Bhattacharyya distance.

$$d_{\mathrm{c}} = \sqrt{1 - \rho[\boldsymbol{p}_{\mathrm{c}}, \boldsymbol{q}_{\mathrm{c}}]} \tag{15a}$$

$$\rho[\boldsymbol{p}_{\mathrm{c}}, \boldsymbol{q}_{\mathrm{c}}] = \sum_{u=1}^{m} \sqrt{\boldsymbol{p}_{\mathrm{c}}(u)\boldsymbol{q}_{\mathrm{c}}(u)} \tag{15b}$$

where u is the index of color histogram bins.

The color observation likelihood can be defined as

$$w_{\mathrm{c}}(z_{\mathrm{c}} \mid \boldsymbol{x}_t) \propto \exp(-\lambda_{\mathrm{c}} d_{\mathrm{c}}) \tag{16}$$

where λ_{c} is the constant factor, and z_{c} denotes the color observation.

3.2.2 Shape model

The facial shape can always be approximated by an ellipse, so in this paper a parametric ellipse is used as the facial shape model. The observation likelihood measure of the elliptical facial shape is based on the image gradient. Specifically, we first construct N measure lines along the hypothesized ellipse decided by state vector $\{x, y, s\}$, as shown in Fig. 2(a). With regard to every measure line, for instance, the k-th measure line shown in Fig. 2 (b), the edge detection is then processed on the measure line, and we use the distance between the point p_k (intersection) and the nearest edge point q_k from it to describe the shape similarity of the k-th measure line. Suppose that the coordinates of point p_k and q_k are (x_k, y_k) and $(x_{\mathrm{e}k}, y_{\mathrm{e}k})$, respectively, then the similarity measure of the k-th measure line can be computed as

$$d_k = \sqrt{(x_k - x_{\mathrm{e}k})^2 + (y_k - y_{\mathrm{e}k})^2} \tag{17}$$

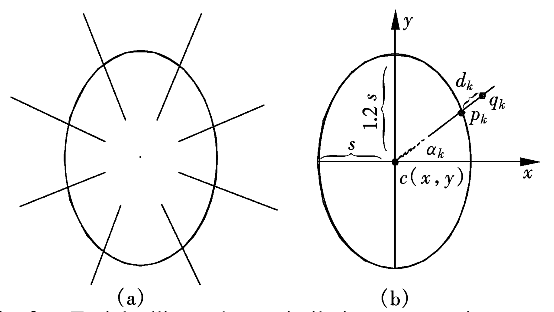

Fig. 2 Facial ellipse shape similarity computation. (a) Sketch map of measure lines construction; (b) Shape similarity computation of the k-th measure line

The whole shape similarity of the ellipse decided by state vector (x, y, s) can be obtained as

$$d_{\mathrm{s}} = \sum_{k=1}^{K} d_k \tag{18}$$

where K is the number of measure lines.

The shape observation likelihood can be defined as

$$w_{\mathrm{s}}(z_{\mathrm{s}} \mid \boldsymbol{x}_t) \propto \exp(-\lambda_{\mathrm{s}} d_{\mathrm{s}}) \tag{19}$$

where λ_{s} is a constant factor, and z_{s} denotes the shape observation.

Under the assumption that the observations from the color and shape cues are statistically independent, the entire observation likelihood of state $\boldsymbol{x}_t$ is given.

$$p(z \mid \boldsymbol{x}_t) = w_{\mathrm{c}}(z_{\mathrm{c}} \mid \boldsymbol{x}_t) w_{\mathrm{s}}(z_{\mathrm{s}} \mid \boldsymbol{x}_t) \tag{20}$$

3.3 Realization of our face tracking algorithm

Suppose that the particles set at time $t-1$ is $\{\boldsymbol{x}_{t-1}^{(n)}, w_{t-1}^{(n)} \mid n = 1, 2, \dots, N\}$, for the frame of time t, the realization of our face tracking system can be summarized as follows:

① Obtain the coarse target state estimation $\boldsymbol{x}_{t0} = \{\boldsymbol{x}_{t0}, y_{t0}, s_{t0}\}$ via the mean shift tracker, then use $\{\boldsymbol{x}_{t0}, y_{t0}, s_{t0}\}$ to construct the proposal distribution $q(\boldsymbol{x}_t)$ as Eq. (8).

② Propagate the particles as Eq. (14) and get the state particles set $\{\boldsymbol{x}_t^{(n)} \mid n = 1, 2, \dots, N\}$ of time t.

③ For every particle $\boldsymbol{x}_t^{(n)}$, give its weight according to the observation as follows:

a) If $\boldsymbol{x}_t^{(n)} \sim q(\boldsymbol{x}_t)$, then

$$w_t^{(n)} = \frac{p(z_t \mid \boldsymbol{x}_t^{(n)}) p(\boldsymbol{x}_t^{(n)} \mid \bar{\boldsymbol{x}}_{t-1})}{q(\boldsymbol{x}_t^{(n)})}$$

b) If $\boldsymbol{x}_t^{(n)} \sim U(\boldsymbol{x}_t)$ or $\boldsymbol{x}_t^{(n)} \sim p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$, then

$$w_t^{(n)} = p(z_t \mid \boldsymbol{x}_t^{(n)})$$

④ Normalize $\{w_t^{(n)} \mid n = 1, 2, \dots, N\}$ to get $\{w_t^{*(n)} \mid n = 1, 2, \dots, N\}$, and estimate the target state of time t

$$\bar{\boldsymbol{x}}_t = \sum_{n=1}^{N} w_t^{*(n)} \boldsymbol{x}_t^{(n)}$$

4 Experimental Results

The performance of our face tracking algorithm based on the particle filter with mean shift importance sampling is tested using the video sequences downloaded from the Stanford Vision Laboratory, which are designed specially to test face tracking algorithms. 50 particles are used in the particle filter proposed in this paper. For comparison purposes, we also implement face tracking based on the general condensation tracking algorithm using 100 particles. Figs. 3(a) and (b) show the tracking results of the condensation tracking algorithm and our proposed method, respectively. From the given results we can see that the condensation is easily distracted by background clutter because its proposal distribution does not take into account the current observation. On the other hand, because the superior proposal distribution of our particle filter with mean shift importance sampling can place the limited particles more effectively, it provides obvious improvement over the condensation tracking algorithm. In addition, it should be noted that our improved particle filter only uses half the number of particles that are used by the condensation tracking algorithm .

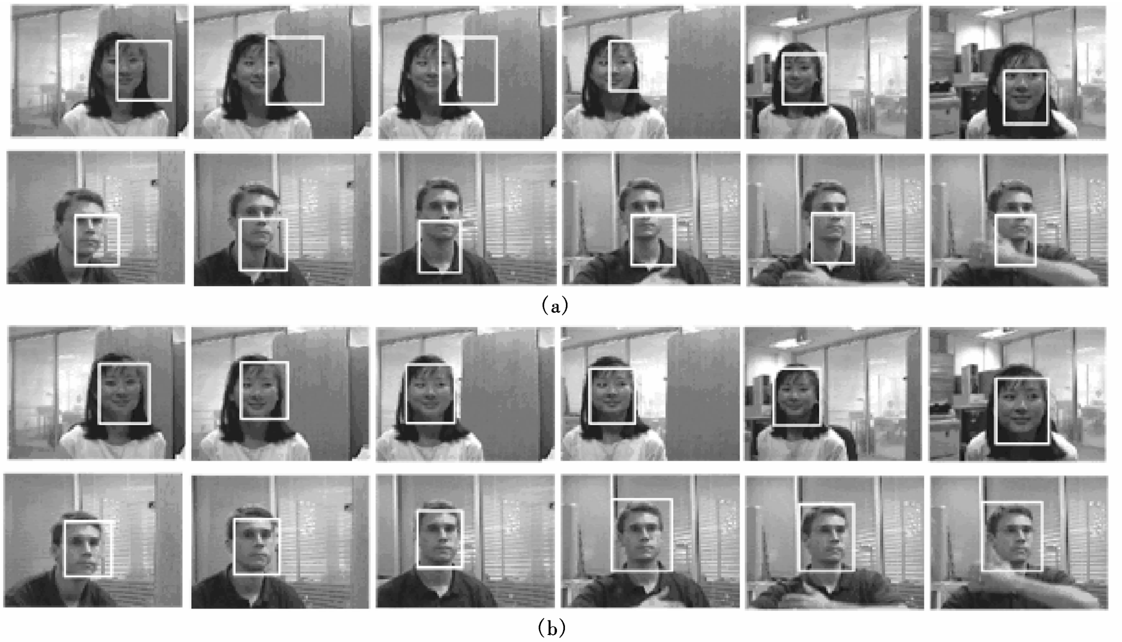

Fig. 3 Comparison of face tracking results. (a) Condensation algorithm; (b) Algorithm proposed in this paper

As for the computational complexity, although the proposed algorithm introduces the mean shift tracker to construct the proposal distribution function, the mean shift tracker is simple and efficient. Furthermore, it is only used to attain the coarse location of the face, so we can make the mean shift tracker iterate very few times in each frame (3 times in this paper). From the above discussion we can see that the mean shift tracker does not bring much computational burden. On the other hand, although the proposed algorithm is more complex when computing particle weight, the theorem given in this paper can solve this problem perfectly. On the whole, the proposed algorithm is a little more complex than the condensation tracking algorithm, but it is worthy due to good performance it brings.

5 Conclusion

The efficiency and accuracy of the particle filter depend drastically on the proposal distribution used to re-allocate the particles. In the general condensation algorithm for visual tracking, the prior transition probability is used as the proposal distribution, the main shortcoming of which is that the most recent observation is neglected, so it frequently results in the poor performance. In this paper a new face tracking algorithm based on a particle filter with mean shift importance sampling is proposed, which uses a mean shift tracker to construct the proposal distribution. The essential idea is to make the proposal distribution embody the most recent observation via the mean shift tracker. The experimental results of face tracking demonstrate the efficiency of this algorithm.

References

[1] Maskell S, Gordon S, Clapp N. A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking [J]. *IEEE Transactions on Signal Processing*, 2002, **50**(2): 174 – 188.

[2] Merwe R, Doucet A, Freitas N, et al. The unscented particle filter, CUED/FINFENG/TR 380 [R]. Cambridge: Cambridge University, 2000.

[3] Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering [J]. *Statistics and Computing*, 2000, **10**(3): 197 – 208.

[4] Haykin S, Huber K, Chen Z. Bayesian sequential state esti-

mation for MIMO wireless communication [J]. *Proceedings of the IEEE*, 2004, **92**(3): 439 − 454.
[5] Kotecha J H, Djuric P M. Gaussian particle filtering [J]. *IEEE Transactions on Signal Processing*, 2003, **51** (10): 2592 − 2601.
[6] Kotecha J H, Djuric P M. Gaussian sum particle filtering [J]. *IEEE Transactions on Signal Processing*, 2003, **51** (10): 2602 − 2612.
[7] Isard M, Blake A. Condensation-conditional density propagation for visual tracking [J]. *International Journal of Computer Vision*, 1998, **29**(1): 5 − 28.
[8] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(5): 564 − 577.
[9] Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(5): 603 − 619.
[10] Isard M, Blake A. ICONDENSATION: unifying low-level and high-level tracking in a stochastic framework [C]// *Proceeding of the 5th European Conference on Computer Vision*. Freiburg, Germany, 1998, **1**: 893 − 908.

基于均值移动重要性采样的粒子滤波人脸跟踪算法

高建坡　杨　浩　安国成　吴镇扬

（东南大学信息科学与工程学院,南京 210096）

摘要: 针对 condensation 目标跟踪算法中用先验转移概率作建议分布函数时没有充分考虑最新观测信息的缺点,提出了一种基于均值移动重要性采样的粒子滤波人脸跟踪算法. 算法首先利用均值移动跟踪器粗略定位人脸目标,然后再用此跟踪结果去构造建议分布函数进行粒子传播. 由于通过该方法所构造的建议分布函数中包含了最新的观测信息,所以它可以使大多数粒子点都能分布在真实状态区域周围,进而提高了粒子传播的准确性. 人脸跟踪结果表明,该算法的跟踪性能明显优于标准 condensation 方法.

关键词: 人脸跟踪;粒子滤波;重要性采样;condensation;均值移动

中图分类号: TP391