

A Novel Particle Filter based Object Tracking Framework via the Combination of State and Observation Optimization

Xudong Luo , Long Ye, Wei Zhong, Qin Zhang

Department of Electronic information engineering Communication University of China, Beijing, China
luoxudong343920710@163.com, yelong@cuc.edu.cn, zhangqin@cuc.edu.cn

Abstract - Using particle filter to figure visual object tracking, a key problem is to choose appropriate image features as the observation model. In this paper, we present a novel particle filter based object tracking framework via the combination of state and observation optimization. We apply the technique to articulated human movement tracking. Result demonstrates the effectiveness of our method in solving the tracking problem like self-occlusion and cluttered background.

Index Terms - particle filter, object tracking, feature optimization

1. Introduction

As a Bayesian estimation problem, visual object tracking have two main steps: prediction and update. Prediction uses Chapman-Kolmogorov (C-K) function to obtain target movement prior probability $p(x_n|y_{1:n-1})$, which can be seen as Eq.1 and n indicates time. And in update stage, the likelihood in Bayes' theorem is used to modify the priori PDF. Therefore posterior PDF can be approximated by Eq.2.

$$p(x_n|y_{1:n-1}) = \int p(x_n|x_{n-1}, y_{1:n-1})p(x_{n-1}|y_{1:n-1})dx_{n-1} \quad (1)$$

$$p(x_n|y_{1:n}) = c_n p(y_n|x_n)p(x_n|y_{1:n-1}) \quad (2)$$

Where c_n is the normalizing constant, $p(x_n|x_{n-1}, y_{1:n-1})$ represents the process equation of the conditional transition of an observed state variable x_n and $p(x_n|y_{1:n})$ represent the observation conditioned on x_n , which is need to be estimated.

Because of occlusion and cluttered background, the distributions during filtering are usually multi-modal, which highly restrict the use of Kalman filter [1]. Particle filter can provide an effective solution to the multi-modal filtering problem as approximating the posterior density by a finite set of normalized weighed particles, as:

$$p(x_n | y_{1:n}) \approx \sum_{i=1}^{N_s} \omega_n^{(i)} \delta(x_n - x_n^{(i)}) \quad (3)$$

Where $\omega_n^{(i)}$ is the normalized weight of particle and can be updated by inserting Eq.3 into Eq.2, as Eq.4 where $q(\cdot)$ is importance sampling function and often be chosen as $p(x_n|x_{n-1})$. Then Eq.4 becomes Eq.5.

$$\omega_n^{(i)} \propto \omega_{n-1}^{(i)} \frac{p(y_n | x_n^{(i)})p(x_n^{(i)} | x_{n-1}^{(i)})}{q(x_n | x_{n-1}^{(i)}, y_n)} \quad (4)$$

$$\omega_n^i \propto \omega_{n-1}^i p(y_n | \mathbf{x}_n^{(i)}) \quad (5)$$

In Eq.5, $p(y_n|x_n^{(i)})$ presents the observation model in Bayes' theorem to modify the priori PDF embodied as $\omega_n^{(i)}$. $p(y_n|x_n^{(i)})$ and $\omega_n^{(i)}$ are respectively corresponding the two processes of particle filter showed as Eq.1. The relationship between $\omega_n^{(i)}$ and $\omega_n^{(i-1)}$ depends on the selected state prediction model. For the prediction, Auto-Regressive is the most popular weak model for motion prediction. And if the movement activity is known, cubic B-spline [2], HMM [3], PCA [4] can make it strong. Unlike the motion model, another idea for effective prediction is to partitioned the parameter space, MacCormick *et al.* [5] partitioned particle filter (PPF) track parameters independently, but takes no account of the hierarchical dependency between parameters which is essential to the articulated body motion, and this defect motivated the present of hierarchical partitioned particle filter (HPPF) [6] which performs a stochastic search in a high dimensional space.

Besides state prediction model, another way to improve the visual tracking performance is to choose appropriate image features as the observation model. D. Comaniciu *et al.* [7] use color feature to represent appearance of rigid or non-rigid objects during visual tracking. The Histogram of oriented Gradients (HoG) representation has been successfully applied to pedestrian detection and human movement tracking. Xinyu Xu *et al.* [8] use edges and silhouettes to compute the likelihood for tracking articulated human body. Zhaozheng Yin *et al.* [9] adopt intensity, texture, motion, saliency and template matching features, and then propose a generic likelihood map fusion framework to combine these heterogeneous features into a fused soft segmentation suitable for mean-shift tracking. In a word, aimed to suiting for different visual tracking conditions, the observation model should be set with the most effective image likelihood function. Meanwhile, because of the variety in the appearance of the tracking objects between image frames, the observation model also should adjust with the object moving. But in the existed particle filter based visual tracking framework, people only focused on the optimization of particle state, ignored the effect

from the fitness of observation model. In this paper, we present a novel particle filter based object tracking framework via the combination of state and observation optimization and then apply the technique to articulated human movement tracking.

2. Problem Formulation

A. Image Likelihood

Using particle filter to figure visual object tracking, the image likelihood or observation model is used to measure the fitness of the particles. It is built by calculating the feature difference between the model and current frame. Four image features used in our research are as followed:

Color Histogram: it is a statistic about each color value of all pixels in one image, as:

$$CH(i) = \frac{n_i}{N} \quad i = 0, 1, \dots, K \quad (6)$$

in which n_i is the number of pixels with color index i , N is the total number of pixels, and K is maximum value of color. The Color Histogram is calculated by the Eq.6 to form a K -dimension eigenvector, and $K=64$ (According to the people's perception about color, the color description is firstly converted into the HSV domain, then the Hue is unevenly divided into 16 parts, each part contains 2-dimensional Saturation parameters and 2-dimensional Value parameters). Color Histogram is a statistical description of the color distribution, and cannot describe the certain object or target in the image.

Color Moments: This feature is based on then-order moments of the color values of every pixel, $n=1, 2$ and 3 in this paper. The calculation of the three color moments is listed as Eq.7 and Eq.8

$$\begin{aligned} \mu(i) &= \frac{1}{N} \sum_{i=1}^N P_{ij} \\ \sigma(i) &= \left(\frac{1}{N} \sum_{j=1}^N (P_{ij} - \mu_i)^2 \right)^{1/2} \end{aligned} \quad (7)$$

$$s(i) = \left(\frac{1}{N} \sum_{j=1}^N (P_{ij} - \mu_i)^3 \right)^{1/3} \quad (8)$$

in which N is the pixel number, P_{ij} is the value for the i -rows, j -columns pixel. The computation complexity of Color Moments is very small, and the color moments feature is 9-dimensional vector, as:

$$CM = [\mu_R, \delta_R, s_R, \mu_G, \delta_G, s_G, \mu_B, \delta_B, s_B]$$

Texture: The texture feature is exacted by Gabor filterbank. Gabor filters are directly related to Gabor wavelets, since they can be designed for a number of dilations and rotations. Therefore, a filter bank consisting of Gabor filters with various scales and rotations is created.

The kernel function of 2-Dimension Gabor filter is:

$$g_{mn} = a^{-m} \times \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x'^2 + y'^2}{2\sigma^2}\right) \quad a > 1 \quad (9)$$

where $x' = a^m(x \cos \theta + y \sin \theta)$, $y' = a^{-m}(-x \sin \theta + y \cos \theta)$, $\theta = n\pi/k$, ($n \in [0, k]$), and k is the number of directions.

In this paper, we choose 4 scales and 6 directions to configure 24 Gabor filters as multipath filter group. After filtering, the mean and variance value in each filters are chosen as texture eigenvector by Eq.10 and 11.

$$\mu_{mn} = \int |w_{mn}(x, y)| dx dy \quad (10)$$

$$\sigma_{mn} = \sqrt{\int (|w_{mn}(x, y) - \mu_{mn}|)^2 dx dy} \quad (11)$$

Where μ_{mn} and σ_{mn} are mean and variance components, thena 48-dimensional texture eigenvector can be built as:

$$T = [\omega_{0,0}, \psi_{0,0}, \omega_{0,1}, \psi_{0,1}, \dots, \omega_{4,6}, \psi_{4,5}]$$

Shape: The direct expression of shape features is sensitive to the details so that it is easily influenced by slight changes of object or noise; even a little change in the coordinates of boundary points will lead to a great variation in shape feature, so its querying performance is usually poor. However, Fourier transformation can obtain general features of an object through switching from sensitive direct expression to frequency domain, which is not easily affected by slight change or noise. For a given image, the *canny* detector is firstly used to get the edge image, and then the FFT is implemented for the edge part. The shape vector is the former 10 parameters that take the most energy in the frequency domain, as Eq.12 and 13.

$$\xi(x) = \sum_{k=0}^{n-1} \text{canny}(I(x)) \exp\left(-\frac{2\pi i}{n} j_k\right) \quad (12)$$

$$S = [\xi_1, \xi_2, \xi_3, \dots, \xi_{10}] \quad (13)$$

In our work, we extract the features directly to eliminate the effect of the segmentation. When we use color histogram, color moments, texture and shape to track the car showed in figure 1(a) separately, the normalized image likelihood calculated by Eq.14 as:

$$p(x_n | y_{1:n}) = \exp\left(\frac{-(f - \text{reference})^2}{\sigma^2}\right) \quad (14)$$

Where f denoted the chosen feature, and δ is the covariance.

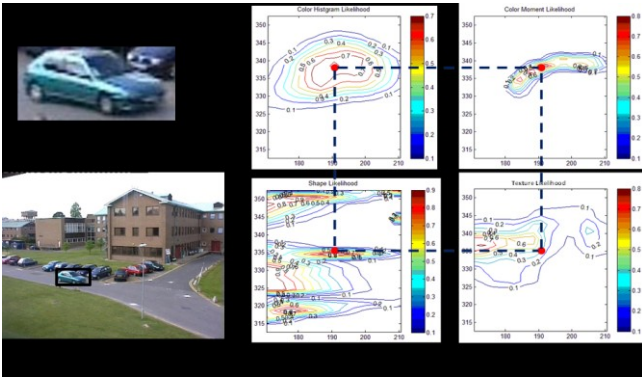


Figure 1: (a) The tracking object whose features are extracted as the reference; (b) Visual tracking frame, the position of black rectangular box corresponded the position of red points in (c); (c) The distribution of image likelihood using the four features separately.

From Figure 1, we can see that the tracking performance influenced by different image features is also different. Meanwhile, because of the variety in the shape or texture of the tracking objects between image frames, the influences from confused background and occlusion phenomenon, the observation model should adapt to the performance of different image features. In the existed particle filter based visual tracking framework, people only focused on the optimization of particle state, ignored the effect from the fitness of observation model. With this consideration, the paper introduces a novel particle filter based visual tracking scheme with the combination of state and observation optimization.

B. The combination of state and observation optimization

The goal of the traditional Bayesian framework is to estimate the hidden state X with the observation Y , as Eq.2. When we taking consider of the combination of state and observation optimization, the problem became the estimation of $p(X, f | f_r, Y)$, and can be extended as Eq. 15.

$$p(X, f | f_r, Y) \propto p(Y | f_r, X, f) p(f_r | f, X) p(X) \quad (15)$$

Where f denotes the chosen features style and f_r is the reference features.

As $p(X)$ is the prior distribution of particle state determined by prediction model, the optimization of $p(X, f | f_r, Y)$ is up to the distributions of $p(Y | f_r, X, f)$ and $p(f_r | f, X)$. Take the car tracking in figure 1 for example, the tracking state is $X = [x_1, x_2]$, and the image likelihood is set to Eq.16:

$$p(X, f | f_r, Y) = \sum_{m=1}^4 \alpha_m E^m \quad (16)$$

where $E^m = \exp\left(\frac{-(f^m - f_{reference}^m)^2}{(\sigma^m)^2}\right)$, and $\sum_{m=1}^4 \alpha_m = 1$, then the combination of state and observation optimization can be formulized as:

$$[x_1, x_2, \alpha_1, \alpha_2, \alpha_3, \alpha_4]_{opt} = \operatorname{argmax}_X p(X, f | f_r, Y) \quad (17)$$

In this paper, we use Simulated Annealing [4] to find the optima. Simulated annealing (SA) is a generic probabilistic metaheuristic for the global optimization problem of locating a good approximation to the global optimum of a given function in a large search space. The fundamental theory of Simulated Annealing algorithm starts with selecting the initial solution, produces a series of Markov chain with the reduction of controlling function, uses a random-result proposal and an accepting principle to repeat a process and iterates the current solution constantly to optimize the target function. The process is as flowed:

- 1) Producing new solution;
- 2) Calculating the difference of the target function;
- 3) Determining whether the new solution is acceptable;
- 4) Accepting or discarding the new solution,

3. Tracking Framework

A. Initialization

Before the tracking, we need to initialize the sample states. Initialization plays an important role in tracking as it may apply more useful prior information for tracking and draw many researches on [10]. To make it easy and accurate, we adopt a manual initialization process. With the guide of ‘marker’, the size of object model can be fixed as the prior information and the reference features are also be extracted in the segmented image parts.

B. PF for Tracking

Combining all the ideas above, the visual tracking framework based on particle filter and the combination of state and observation optimization can be summarized as follows.

- Manual initialization, then draw the N particles $\{X_0, \omega_0\}_{i=1:N}$ around the initialized state, set $V_0=0$;
- For $k = 1$ to the frame number
 - ✓ Prediction: $X_k = X_{k-1} + V_{k-1}$
 - ✓ Measure ω_k and do optimization with the combination of state and observation;
 - ✓ MAP output $X_{out,k}$ and set $V_k = X_{out,k} - X_{out,k-1}$.

C. Experiments

In this paper, we applied our method to figure the articulated human movement tracking. Articulated human movement tracking problem has three properties (High-dimensional, self-occlusion and cluttered background.) which drew many attentions on. We chose ‘inside’ and ‘outside’ to test the performance of our algorithm. In the ‘inside’ tracking scene, the similarity between the color of the elevator’s door and the trousers of human, and the occlusion phenomenon exists between the two legs make the tracking difficulty increasing; and in the ‘outside’ scene, we also need to consider the change of the human’s depth. By setting $[x_1, x_2, x_3, \theta_1, \theta_2, \theta_3]$ as the degrees of freedom of human body, $[\theta_4, \theta_5, \theta_6, \theta_7]$ as the degrees of freedom of thigh and $[\theta_8, \theta_9]$ as the degrees of freedom of crus, the tracking result can be seen as figure 2 and figure 3.



Figure 2: Articulated human movement tracking result (Inside)



Figure 3: Articulated human movement tracking result (Outside)

4. Conclusion

In this paper, we present a novel particle filter based object tracking framework via the combination of state and observation optimization. In the experiments, we applied our method to articulated human movement tracking. With the weights adjustment for every choosing image feature, we can improve the tracking performance by setting observation model. According to making every frame can find its more effective feature expression, our method can solve the tracking problem like self-occlusion and cluttered background

Acknowledgements:

This paper is supported by National Natural Science Foundation of China (61201236 and 61101166)

References

- [1] A.J. Huang, "A Tutorial on Bayesian Estimation and Tracking Techniques Applicable to Non-linear and Non-Gaussian Process", MITRE Technique Report, Jan 2005.
- [2] M.U.Ramos Sánchez, J.Matas, J.Kittler, "Statistical chromaticity models for lip tracking with B-splines", First International Conference, AVBPA'97 Crans-Montana, Switzerland, March 12–14, 1997.
- [3] Yunqiang Chen, Yong Rui, Huang T.S, "Multicue HMM-UKF for real-time contour tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 28, Issue: 9, Page(s): 1525 – 1529, 2006.
- [4] Yang F, Lu H, Zhang W, Yang G, "Visual tracking via bag of features", IET Image Processing, Volume: 6, Issue: 2, Page(s): 115 – 128, 2012
- [5] J.P. MacCormick, M. Isard, "Partitioned sampling, articulated objects, and interface-quality hand tracking", in: Eur. Conf. on Computer Vision II, pp. 3–19, 2000.
- [6] Long Ye, Qin Zhang, Ling Guan, "Use hierarchical genetic particle filter to figure articulated human tracking," Proc. IEEE Int. Conf. on Multimedia and Expo, Hanover, Germany, June 2008.
- [7] Comaniciu D, Meer. P, "Robust analysis of feature spaces: color image segmentation", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997
- [8] Zhixu Zhao, Shiqi Yu, Xinyu Wu, "A Multi-Target Tracking Algorithm Using Texture for Real-Time Surveillance", IEEE International Conference on
- [9] Z. Yin, F. Porikli and R. Collins, "Likelihood Map Fusion for Visual Object Tracking," IEEE Workshop on Application of Computer Vision (WACV), Copper Mountain, CO, Jan. 2008