Adaptive Tracking with Patches and a New Particle Filter

Lin Ma, Weiming Hu
Institute of Automation, Chinese Academy of Sciences
Beijing 100190, China
E-mail: lin.ma@ia.ac.cn

Abstract— This paper presents a new tracking method based on color histogram model and patch appearance model. We propose a new particle filter which processes tracking concurrently in several routes, and each route creates particles separately. Each particle is optimized with a new regular iteration method. Then we evaluate each particle in two levels: the whole level and the patch level. The two levels use color histogram and patch images to evaluate the particles respectively. Different routes have different appearance models. We also maintain a combined appearance model. Experimental results demonstrated the effectiveness of our method.

Keywords— Multi-level; patch; particle filter; tracking

1. INTRODUCTION

Object tracking is widely used in many fields such as video surveillance and action understanding. Various features are adopted to perform object tracking and color histogram is an important one among these features. Bradski [1] proposed a color histogram based method: Camshift (Continuous Adaptive Meanshift) to perform face tracking. Comaniciu et al. [2] used a new meanshift based method in tracking with kernel method. Particle filter was also widely used in tracking. Gordon [3] raised particle filter to deal with non-Gaussian problems. Isard and Blake [4] introduced particle filter into object tracking. Wang et al. [5] involved spatial information in Mixture of Gaussians and proposed a new object appearance model: SMOG (Spatial-color Mixture of Gaussians). Ablavsky [6] used layered graphical models to track objects with particle filter. Zhang et al. [7] proposed a new particle filter using Sequential Particle Swarm Optimization. Some researchers also combined a particle filter with mean shift together to perform object tracking [8]. Zhang et al. [9] used a new SVD based Kalman particle filter to perform tracking. Adam, et al. [10] used patches to perform tracking and they also used integral image to improve the performance efficiency. Babenko et al. [11] used multiple instance learning to perform tracking which represented object and background with bags of instances. Tran and Davis [12] tracked objects at multiple levels of spatial resolution. However we constructed the relationship between different routes in a different way from [12]. Pan and Schonfeld [13] determined the optimal particle number and memory constraints using rate-distortion theory under the fixed particle number and memory constraints respectively. Some papers used multiple cues to perform tracking and achieved good results [14, 15].

In our paper, we propose a new particle filter which performs tracking in several routes concurrently. Each route maintains a route appearance model. We also maintain a combined appearance model. This particle filter can model different conditions of object state’s variances efficiently. In each route, we represent the object in two levels. In the whole level, we evaluate each particle based on color histogram, and in the patch level, we use patch images to detect occlusion and evaluate the particle. We propose a new way to update the patch model. The novelties of our paper mainly exist in:

1) We propose a new particle filter, concurrent particle filter with different variances. The new particle filter performs tracking in several routes. Each route maintains a route appearance model. We also maintain a combined appearance model for all the routes. This new particle filter can present different conditions of object state’s variances by using multi-routes.

2) We use a simple new regular iteration method to optimize the particles. This method is able to confine each iteration of searching in a predefined searching space which can make the searching more robust and efficient.

3) We present a new way to update the route patch model and the combined patch model. This new updating method updates the mean patch incrementally and can remain the object appearances in different conditions more efficiently.

The rest of the paper is organized as follows: In Section II, we present the new particle filter. We present the observation likelihood in Section III. Experiments are shown in Section IV and conclusions and future work are shown in Section V.

II. CONCURRENT PARTICLE FILTER WITH DIFFERENT VARIANCES

A. Concurrent Particle Filter with Different Variances

Particle filter is formed based on Bayesian formula. Given the observation sequence \( O_{t+1} \), the posteriori probability density of the object state is defined as

\[
p(X_{t+1}, O_{t+1}) \propto p(O_{t+1}, X_{t+1}) \int p(X_{t+1}|X_t)p(X_t|O_{1:t})dX_t
\]

where \( X_t \) (composed of \( tx, ty, sx, sy \)) represents the object state at time \( t \). Each particle is warped to a standard \( 32 \times 32 \) sub-image. We define \((tx, ty)\) as the coordinate of the bottom left point of the state rectangle in the frame image, and define
sx, sy as the ratio of the object’s width and height to the standard sub-image’s width and height respectively.

We assume that each component, e.g. $tx$, of the particle state, is independent of the others, and its distribution conforms to the Gaussian distribution, that is, $X^{t}_{i}(t+1) \sim N(X_{i}(t); X^{t}_{i}, \sigma^{2}(t))$. Here, $X_{i}(t)$ represents the $i$th component, e.g. $tx$, of state $X^{t}_{i}$, $X^{t}_{i}$ is the optimal state at time $t$ (after the optimization with regular iteration). We compute the likelihood function $p(O_{t+1} | X^{t}_{i+1})$ with

$$p(O_{t+1} | X^{t}_{i+1}) \propto p(O_{h_{t+1}} | X^{t}_{i+1})p(O_{p_{t+1}} | X^{t}_{i+1})$$

(2)

where $p(O_{h_{t+1}} | X^{t}_{i+1})$ is the likelihood of $X^{t}_{i+1}$ based on the histogram model, $p(O_{p_{t+1}} | X^{t}_{i+1})$ is the likelihood of $X^{t}_{i+1}$ based on the patch model. The two likelihood functions are presented in detail in Section III. We use gray image to perform the tracking.

We perform the tracking in several routes concurrently and each route maintains a route model (histogram and patch). We also maintain a combined model to retain the relations between different routes. The combined model is defined as

$$T_{r,t} = \sum_{r} w_{r,t} \cdot T_{r,t}$$

(3)

where $T_{r,t}$ is the model of route $r$ at time $t$, $T_{c,t}$ is the combined model at time $t$, $m$ is the number of routes. $w_{r,t}$ is the weight of route $r$ at time $t$. The detailed presentation of obtaining the combined patch model is shown in Section III (C.2)). We compute $w_{r,t}$ in an accumulated way

$$w_{r,t} = w_{r,t-1} \cdot V^{t}_{r,t}$$

(4)

where $V^{t}_{r,t}$ is the evaluation of the optimal state $X^{t}_{r,t}$ of route $r$ at time $t$. $w_{r,t}$, $r=0, ..., m-1$ is normalized with the sum to 1 constraint.

We define $X^{t}_{r,R,t}$, $X^{t}_{r,C,t}$ are the selected optimal particles from route $r$’s particles according to $T_{r,t}$ and $T_{c,t}$ respectively. We define

$$X^{t}_{r,t} = \frac{w^{R}_{r,t}}{w^{R}_{r,t} + w^{C}_{r,t}} X^{t}_{r,R,t} + \frac{w^{C}_{r,t}}{w^{R}_{r,t} + w^{C}_{r,t}} X^{t}_{r,C,t}$$

(5)

as the optimal state of route $r$, where $w^{R}_{r,t} = p(O_{t} | X^{t}_{r,R,t}) \cdot p(X^{t}_{r,R,t} | X^{t}_{r,t-1})$ and $w^{C}_{r,t} = p(O_{t} | X^{t}_{r,C,t}) \cdot p(X^{t}_{r,C,t} | X^{t}_{r,t-1})$ are the evaluations of $X^{t}_{r,R,t}$ and $X^{t}_{r,C,t}$ according to $T_{r,t}$ and $T_{c,t}$ respectively. $X^{t}_{r,t}$’s evaluation is computed as

$$V^{t}_{r,t} = \frac{w^{R}_{r,t}}{w^{R}_{r,t} + w^{C}_{r,t}} w^{R}_{r,t} + \frac{w^{C}_{r,t}}{w^{R}_{r,t} + w^{C}_{r,t}} w^{C}_{r,t}$$

(6)

The particle $X^{t}_{r,t}$ from the route $r$ is chosen as the final optimal particle at time $t$ if $p_{w,r,t}$ is the largest among $p_{w,r,t}$, $r=0, ..., m-1$.

We use 2 routes in our paper. At frame $t$, we first optimize the optimal particle $X^{t}_{r,t-1}$ of route $r$ with regular iteration (see Section II (B)) in the new frame, and then sample respectively some particles around $X^{t}_{r,t-1}$ for route $r$, $r=0,1$. At frame $t$, each of the two routes only saves its optimal particle $X^{t}_{r,t}$, $r=0,1$. We select the optimal particle from the route with the larger value between $p_{w_{0,t}}$ and $p_{w_{1,t}}$ as the final optimal state at time $t$.

We define the state components’ (e.g. $tx$) variances of different routes in different ways. We define functions

$$g_{0}(x) = 0.7 + 0.3 \times e^{-x}$$

(7)

$$g_{1}(x) = 0.7 + 0.3 \times e^{-x}$$

(8)

With $g_{r}(x)$, $r=0,1$, we define

$$\sigma^{2}_{r,t}(i) = \sigma^{2}_{r,t}(i) \cdot g_{r}(E_{r})$$

(9)

where $\sigma^{2}_{r,t}(i)$, $r=0,1; i=0, ..., 3$ is the initial value of $\sigma^{2}_{r,t}(i)$, $E_{r}$ represents the mean of the occlusion extents (Equation (15)) of the recent $m$ frames. There can be other ways to define $g_{r}$. In our paper, we define $g_{r}$ as (7) and (8).

When the difference between $X^{t}_{r,t-1}$, $r=0,1$ is small, we consider that most of the particles converge around the ground truth (the real state of the object), and then we need a smaller number of particles; or else we need a larger number of particles. So we define the particle number at time $t$ as

$$N_{p,t} = N_{p} \cdot \left[ \frac{1}{2} + \frac{1}{2} \exp \left( -1 \cdot \frac{1}{\sigma^{2}_{r,t}(0)} \right) \left| X^{t}_{0,t-1} - X^{t}_{1,t-1} \right| \right]$$

(10)

1/2 in (10) is used to confine $N_{p,t}$ within $(1/2N_{p,0}, N_{p})$.

The algorithm of the new particle filter is shown in Algorithm 1 (take $r=2$ as an example).

**Algorithm 1. Concurrent particle filter with different variances**

1) At $t$, optimize $X^{t}_{r,t-1}$, $r=0,1$ with regular iteration.
2) Sample $N_{p,t}$ particles for route 0 and 1 around $X^{t}_{r,t-1}$, $r=0,1$ respectively.
3) Compute $w^{R}_{r,t}$ and $w^{C}_{r,t}$, $r=0,1$.
4) Compute $X^{t}_{r,t}$ and $V^{t}_{r,t}$ with Equation (5) and (6).
5) Choose $X^{t}_{r,t}$ from the route with the larger value between $p_{w_{0,t}}$ and $p_{w_{1,t}}$ as the final optimal state at time $t$.
6) If not the last frame, update system and back to 1).

B. Optimize the Particle with Regular Iteration

When frame $t$ comes, we first optimize $X^{t}_{r,t-1}$, $r=0,1$ based on the mean pixel density iteratively (Figure 1). The pixel density is the histogram value of the bin which the pixel gray

![Figure 1: Optimize the state with regular iteration. Black point represents the initial center of the particle. * represents the optimal center in the first iteration. Red point represents the beginning center point in the second iteration.](image-url)
value belongs to according to the histogram model (Figure 4(a)). The mean pixel density is the average of the pixel densities of the pixels within a particle. In each iteration, we first define the beginning point, i.e., the point where the center of the new optimal particle initially lies. Then we make the center of new optimal particle walk in the $5 \times 5$ neighborhood around the beginning point and find the particle with the largest mean pixel density. When the optimal point is selected out, we compute the intersection point of the boundary lines of the $5 \times 5$ area and the line passing the beginning point and the optimal point. We choose the intersection point at the side of the new optimal point as the new beginning point (the red point in Figure 1), and then begin the next iteration. We use the optimal point in the last iteration as the center of the new particle around which we sample new particles. In our paper, we run 2 iterations. In each iteration, we can also first sample some points among the $5 \times 5$ points, and select the optimal one among the sampled points. Figure 2 shows the result of utilizing the regular iteration to optimize the initial particle.

![Figure 2: Initial particle optimization of route 0. The red rectangle represents the initial particle, The yellow rectangle represents the particle optimized by the regular iteration.]

III. OBSERVATION LIKELIHOOD

We evaluate the particles with a multi-level structure. In the whole level, we refine a histogram model incrementally and compare it with the particle's histogram. In the patch level, we divide each particle into $h^4$ patches, and use the patches to detect occlusion and match between the particle and the patch models.

A. Warping the Particle

We present the corresponding relationship between the standard sub image and the particle with

$$ P' = A_x A_y P, \quad A_x = \begin{bmatrix} sx & 0 & 0 \\ 0 & sy & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_y = \begin{bmatrix} 1 & 0 & tx \\ 0 & 1 & ty \end{bmatrix} $$

where $P'$ is a point in the standard sub image, $P'$ is its corresponding point in the frame image.

B. Color Histogram Matching

We adopt the weighted integral image [10] to compute the histogram of each particle. We divide each particle into 4 layers, and define the weight of layer $i=0,...,3$ as

$$ u_i(x) \propto \exp\left(-i^2/(2\sigma^2)\right), \quad \text{where } \sigma^i \text{ is a parameter. Each particle histogram is unified (Figure 4(a)).} $$

Every 5 frames, if $E_r$ is smaller than a threshold, we update the histogram model $q_{r,t}^*$ of route $r$ at time $t$ with

$$ q_{r,t+5}^* = s \cdot q_{r,t}^* + (1-s) q_{r,t}^\circ $$  \hspace{1cm} (12) $$

where $q_{r,t}^\circ$ is the mean of the histograms corresponding to $X_r^{t+1}, ..., X_r^{t+5}$, $s$ is a constant.

We compute the similarity between particle histogram $q_{r,t}$ and $q_{r,t}^*$ as

$$ p(q_{r,t}, q_{r,t}^*) = \sum_{k=1}^N \sqrt{q_{r,t}(k)q_{r,t}^*(k)} $$

We define

$$ p(O_h,l, X_{r,R}) = \exp\left\{ p(q_{r,t}, q_{r,t}^*) / N_h \right\} $$

where $N_h$ is the histogram bins’ number. We set $N_h = 32$. $p(O_h,l, X_{r,C})$ is computed in a similar way.

C. Patch Matching

1) Patch Similarity: We divide the particle into $4 \times 4$ patches to consider spatial information and detect occlusion at the beginning of dealing with each particle. We evaluate the patch according to the patch image model $M_{r,i,j}^k$ (route) and $M_{r,i,j}^{k,\text{combined}}$ (combined), $i,j=0,...,3$; $r=0,1$; $k=0,...,4$. We take route 0 based on $M_{r,i,j}^k$ as an example. For route 0 based on $M_{r,i,j}^{k,\text{combined}}$ for route 1, the likelihoods are computed in a similar way.

We define $p_{i,j}^k$ as the patch $(i,j)$'s image of route 0. Similar to 1-NN (1-nearest neighbor) method, we find the model image $M_{r,i,j}^{0,k}$ among $M_{r,i,j}^k$ which has the smallest L1 distance from $p_{i,j}^k$. If the distance is larger than a threshold, we consider the patch is occluded, i.e. $a_{i,j}^0 = 1$. The occlusion extent of the particle is computed as

$$ \hat{O}_{i,j} = \sum_{k=0}^3 w(i,j) a_{i,j}^k $$

where $w(i,j)$ is the weight of patch $(i,j)$. We define that each of the 4 center patches has weight 1/8, and each of the boundary ones has weight 1/24. We increase the importance of the likelihood about $O_{i,j}$ along with the occlusion extent.

We define the likelihood based on route patch model as

$$ p(O_h,l, X_{r,C}) = \exp\left\{ 1/h \hat{O}_{i,j} + \sum_{i,j=0}^3 \frac{1}{54} \left\| F_{r,i,j} - M_{r,i,j}^k \right\|_{1} \left(1 - a_{i,j}^k \right) \right\} $$

2) Updating Patch Models: For each route, we update the patch models every 5 frames. To reduce the influence of the occlusions, we perform the updating of route $r$ when $E_r$ is below a threshold. We update each patch’s model separately. For patch $(i,j)$ of route $r$, we form a patch set $S_{r,i,j}$ containing $9$ patch images $F_{r,i,j}$. 
we select 15 patch images of $S_{r,i,j}$ which are the 4 previous model images (except $M^0_{i,j}$) and the 5 recent patch images. To form the new patch model $M^r_{i,j} = \sum_{\text{dist}}^5 p_{i,j}^{(r)}$, $k = 0, \ldots, 4$, we select 4 patch images from $S_{r,i,j}$. The selection process contains the following 2 steps:

Step 1: Remove from $S_{r,i,j}$ 3 patch images with 3 largest L1 distances from $M^0_{i,j}$. This step aims to reduce the influence of noises.

Step 2: Remove 2 patch images from the remained 6 patch images of $S_{r,i,j}$ after Step 1. For each of the 6 patch images $P_{r,i,j}^{(k)}$, $k = 0, \ldots, 5$, we compute the sum $D_k$ of the L1 distances between $P_{r,i,j}^{(k)}$ and $P_{r,i,j}^{(k')}$, $k_1 = 0, \ldots, 5$, $k_1 / k$ and between $P_{r,i,j}^{(k)}$ and $M^0_{i,j}$. We remove the 2 patch images with the 2 smallest values among $D_k$, $k = 0, \ldots, 5$. This step aims to remain the diversity of patch models, and remain patch images appearing in different conditions.

To our knowledge, there is no simple way to remove m images from n images while having largest sum of L1 distances between the remained n-m images. Thus, we use the method as presented in Step 2 to achieve an approximate result. We use the remained 4 patch images of $S_{r,i,j}$ to form $M^r_{i,j} = \sum_{\text{dist}}^4 p_{i,j}^{(r)}$, $k = 1, \ldots, 4$. And we obtain $M^r_{i,j,t+5}$ through

$$M^r_{i,j,t+5} = sM^0_{i,j,t} + (1-s)\frac{1}{4} \sum_{k=1}^4 M^r_{i,j,t+5}$$

The combined patch model $M^r_{i,j,t+5}$, $i,j = 0, \ldots, 4$; $k = 0, \ldots, 4$ is updated with $M^r_{i,j,t+5}$; $r = 0, 1$; $i,j = 0, \ldots, 4$; $k = 0, \ldots, 4$. We define (Figure 4(b))

$$M^r_{i,j,t+5} = p_{uv,t+5} \cdot M^0_{i,j,t+5} + p_{uv,t+5} \cdot M^r_{i,j,t+5}$$

We select 4 patch images from $M^r_{i,j,t+5}$, $k = 1, \ldots, 4$ to form $M^r_{i,j,t+5}$, $k = 1, \ldots, 4$. The 4 patch images are corresponding to the largest 4 distances between $M^r_{i,j,t+5}$, $r=0,1$; $k = 1, \ldots, 4$ and $M^r_{i,j,t+5}$. These 4 patch images are chosen to remain the diversity of the combined patch model.

IV. EXPERIMENTS

We tested our method on 6 videos using VC++6.0 on a computer with CPU at 2.53GHz, on the Win XP OS. We defined $N_p = 100$ and $s = 0.95\sim0.99$ (in (12)(17)). We used the first 4 frames to train the system. The running time was around 0.09 sec/frame. We used the sum of distances between the 4 state rectangle edges and the ground truth to represent the errors. The test videos we used were from EC Funded projects.

In Figure 5, we tracked a case (with parts of persons’ bodies) carried by two persons and compared our method HM+PM (Histogram model + patch model) with HM. The case was occluded by another walking man. The histogram was robust to distortions, but it lost much spatial information. Also the surroundings around the case were similar to the case in gray image (Figure 4(a)). Thus HM drifted away from the ground truth gradually. However in our method, we combined the histogram information and the patch information together to represent the object appearance, and we also considered the occlusion extent. So our method was able to track the object more accurately. Figure 6(a) showed the occlusion extents of the two routes in Figure 5 respectively. From Figure 6(a), we could see that generally when the occluding man stayed around the middle of the case, the occlusion extents were larger than around other positions of the case. That was because the middle part of the object area had larger probability of belonging to foreground. Figure 6(b) showed the choice of routes in Figure 5. We could see that when the case was occluded, the program mainly chose the optimal particle of route 0 as the final optimal particle. That was because in Figure 5, when the man occluded the case, route 0 made the particles spread more widely by making the state variances larger. Thus, the program was able to be prevented from being trapped in a local optimal solution. Figure 6(c) showed the particle number (same for each route) at each frame in Figure 5. The nearer the two optimal particles of the two routes were, the smaller the particle number was; and vice versa. The error map of Figure 5 was shown in Figure 7.

![Figure 4](image4.png)

(a) Pixel density. The lighter means the larger pixel density value. (b) The 4×4 grid of patch models. We only show the mean patch of the 5 patch model images. Left: route 0; middle: route 1; right: combined model.

![Figure 5](image5.png)

Figure 5: A case carried by two persons. Red rectangle: HM. White rectangle: HM+PM.

CAVIAR project/IST 2001 37540 [16] and [17].

In Figure 5, we tracked a case (with parts of persons’ bodies) carried by two persons and compared our method HM+PM (Histogram model + patch model) with HM. The case was occluded by another walking man. The histogram was robust to distortions, but it lost much spatial information. Also the surroundings around the case were similar to the case in gray image (Figure 4(a)). Thus HM drifted away from the ground truth gradually. However in our method, we combined the histogram information and the patch information together to represent the object appearance, and we also considered the occlusion extent. So our method was able to track the object more accurately. Figure 6(a) showed the occlusion extents of the two routes in Figure 5 respectively. From Figure 6(a), we could see that generally when the occluding man stayed around the middle of the case, the occlusion extents were larger than around other positions of the case. That was because the middle part of the object area had larger probability of belonging to foreground. Figure 6(b) showed the choice of routes in Figure 5. We could see that when the case was occluded, the program mainly chose the optimal particle of route 0 as the final optimal particle. That was because in Figure 5, when the man occluded the case, route 0 made the particles spread more widely by making the state variances larger. Thus, the program was able to be prevented from being trapped in a local optimal solution. Figure 6(c) showed the particle number (same for each route) at each frame in Figure 5. The nearer the two optimal particles of the two routes were, the smaller the particle number was; and vice versa. The error map of Figure 5 was shown in Figure 7.

![Figure 6](image6.png)

Figure 6: Corresponding to Figure 5. (a) Occlusion extent. (b) Route selection. (c) Number of Particles.

![Figure 7](image7.png)

Figure 7: Error map corresponding to Figure 5.
classical method which remained an appearance model of subspace. Our method also considered the pixel’s color distribution and we used patch images to obtain the patch model. We compared our method with SMOG and PCA in our paper. In Figure 8(a), the woman’s head was occluded for some frames, SMOG and PCA departed largely from the ground truth. Our method divided the sub-image to patches and we used patch images to obtain the patch subspace. Our method also considered the pixel’s color distribution and we used patch images to obtain the patch model. The error maps corresponding to Figure 8 were shown in Table 1.

In Figure 10, we showed the results of 3 other experiments. In the first row, we tracked a woman walking from near to far. In the second row, we tracked a woman sitting in a car. And in the third row, though the object was with complex background, by combining the patch model and the histogram model, we could track the object for a long sequence.

V. CONCLUSION AND FUTURE WORK

In this paper, we have presented a new method to track object with or without occlusion. We proposed a new particle filter, and performed tracking with multi-levels in our program. We’ll continue the research in making different routes of the new particle filter communicate more efficiently.

ACKNOWLEDGEMENT

This work is partly supported by NSFC (Grant No. 60825204 and 60935002) and the National 863 High-Tech R&D Program of China (Grant No. 2009AA01Z318).

REFERENCES