Adaptive Probabilistic Tracking with Visual Saliency Selection Reliable Particles

Linshan Liu, Suiwu Zheng, and Hong Qiao
State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
Email: {linshan.liu, suiwu.zheng, hong.qiao}@ia.ac.cn

Abstract—It is very difficult to guarantee the stability and accuracy of object tracking in real-world scenarios. Although probabilistic tracking has become popular, there are two major problems: discriminative appearance modeling and reliable particles selection. To address these issues, this paper presents a visual saliency inspired approach that can capture the varying appearance characteristic of target and background, and select reliable particles during tracking. The global image signature descriptor, which can find the discriminant features for tracking, is used to model the appearances of target and background, as well each hypothetical observation, and can easily be embedded into the particle filter framework. The weight of each particle is estimated not only through saliency measurement implemented by Hamming distance between the target model and each hypothetical observation, but also through the background model and each hypothetical observation. The saliency selection reliable particles are used to estimate the target state. Finally, experimental results demonstrate the efficiency and effectiveness of the proposed method in presence of occlusion and large illumination variation, even non-target with similar features.

Index Terms—image signature, reliable particles, saliency selection, probabilistic tracking

I. INTRODUCTION

Object tracking has many practical applications in the field of computer vision, such as intelligent video surveillance, weapon guidance, and mobile robotics vision. Though numerous approaches have been proposed in the literatures [1], it is very difficult to guarantee the stability and accuracy of object tracking in practice due to appearance change and the complex real-world scenarios. Particle filter (PF) [2] tracking algorithm [3] has been successfully used to track the image sequences. However, it shows poor performances in the presence of large illumination variation, occlusion and non-target with similar features distributions.

Recently, many improved methods of PF have emerged. Different types of features and their combination are exploited to represent the tracked object in this framework. The selected features can be divided into two groups: low level attributes, such as color [3], edge [4] and high level ones such as saliency [5] features. However, color distribution becomes ineffective in the presence of illumination changes. Edges are sensitive to clutter and computationally expensive. Mahadevan and Vasconcelos [6] show that saliency feature played an important role in object tracking. For the better performances, one can combine saliency and other features as in [7]-[9]. In [7], an improved Itti model [5] is employed to measure the similarity between tracked object and candidate one, which was more accurate. In addition, target model was used as prior knowledge to calculate the weight set which was utilized to construct their saliency features adaptively [8]. D. Sidibé [9] introduces frequency tuned saliency features into particle filter. However, all of the approaches either suffer very large computational complexity in real applications or do not use the background information which is likely to improve tracking stability. Therefore, these methods are apt to be distracted by background with similar features which is likely to cause drift. Target model as prior knowledge may be ineffective when it changes drastically.

In this paper, we provide an effective and efficient adaptive probabilistic tracking approach with saliency selection reliable particles. We introduce a binary, simple and holistic image descriptor called the “image signature” [10] to extract saliency features of the object. The proposed method estimates the weight of each particle through the saliency measurement implemented by Hamming distance between target model and each hypothetical observation, but also through background model and each hypothetical observation. The saliency selection reliable particles are used to estimate the target state. Experimental results demonstrate the goodness of the proposed method under diverse conditions.

In the rest, in Section2, an overview of the PF based tracking is given. In Section3, we explain our method in details. Experimental results were presented in Section4, and finally, Section5 gives the conclusion and further works.

II. PARTICLE FILTER TRACKING FRAMEWORK

A. Particle Filtering Framework

As is shown in [2], [3], particle filtering tracking is viewed as Bayesian inference that is recursively obtained by state prediction and update with the following two rules:
\begin{align}
p(X_i | Z_{i:k}) &= \int p(X_i | X_{i:k}) p(X_{i:k} | Z_{i:k}) dX_{i:k} \quad (1) \\
p(X_i | Z_{i:k}) &= \frac{p(Z_i | X_i)p(X_{i:k} | Z_{i:k})}{p(Z_i | X_{i:k})} \quad (2)
\end{align}

where \( p(X_i | X_{i:k}) \) and \( p(X_{i:k} | Z_{i:k}) \) correspond to the transition model and the posterior density at time \( k - 1 \), \( p(Z_i | X_i) \) represents the observation model, and \( p(Z_{i:k} | X_{i:k}) \) is a normalization constant. Then, the object state \( X_i \) can be obtained through the posterior density, which is approximated by

\[
p(X_i | Z_{i:k}) = \sum_{s=1}^{S} \pi_s^i \delta \left( X_i - X_s^i \right) \quad (3)
\]

Here \( \delta (\cdot) \) is Dirac function, and \( \pi_s^i \) which satisfies \( \sum_{s=1}^{S} \pi_s^i = 1 \) is the weights associated with the particles, which is calculated by

\[
\pi'_i \propto \hat{\pi}'_i = \frac{p(Z_i | X'_i)p(X'_i | X_{i:k})}{q(X'_i | X_{i:k}, Z_i)} \quad (4)
\]

where \( q(\cdot) \) is the proposal distribution used to sample the particles, \( \hat{\pi}'_i \) represents the particle weight after the resampling step that draws \( \{ X'_i \}_{i=1}^{N} \) from the set \( \{ X_i \}_{i=1}^{N} \) according to resampling function \( \{ \pi_s^i \}_{i=1}^{N} \). Usually, \( \pi'_i = \pi_s^i \) hence, the weight of each particle can be measured by

\[
\pi'_i \propto p(Z_i | X'_i) \quad (5)
\]

The state of target can be estimated by

\[
E(X_i) = \sum_{s=1}^{S} \pi'_i X'_i \quad (6)
\]

**B. Appearance Modeling**

Within a probabilistic tracking framework, the object is usually represented by a rectangle, in which, different types of features can be used to measure the observation likelihood of the samples. The object state is defined by vector \( X_i = [x_i, y_i, \sigma_i] \), where \( (x_i, y_i) \) is the coordinate of the region center, and \( \sigma_i \) represents the scale of the object. As is shown in [4], through incorporation of the background information, the model of target and background, as well the observation of each particle can be represented by M bin color histograms \( \{ H_{\sigma_i}^{\text{bf}} \}_{\sigma_i} \), \( \{ H_{\sigma_i}^{\text{bf}} \}_{\sigma_i} \) and \( \{ H_{\sigma_i}^{\text{bf}} \}_{\sigma_i} \), here \( \sum_{\sigma_i} H_{\sigma_i}^{\text{bf}} = 1 \), \( \sum_{\sigma_i} H_{\sigma_i}^{\text{bf}} = 1 \) and \( \sum_{\sigma_i} H_{\sigma_i}^{\text{bf}} = 1 \). Then, the weight of particle \( X'_i \) is defined by:

\[
\pi'_{\sigma_i} \propto \exp \left\{ -\lambda \cdot \left( d[H'_i, H_s] - d^2[H'_i, H_s] \right) \right\} \quad (7)
\]

where \( \lambda \) is a constant determined in practice. The distance measure \( d \) is derived from the Bhattacharyya coefficient.

**III. ADAPTIVE PROBABILISTIC TRACKING WITH SALIENCY SELECTION RELIABLE PARTICLES**

It should be noted that there was several limitations in particle filtering framework [4]. First, only color features used may not be discriminative enough under certain circumstances. Second, the weight of particles has poor discriminative ability. The random particles scatter around the real state of the target, far from which some particles contribute little to estimate the target state. Different from [7]-[9], we use a simple, binary and holistic image descriptor called the “image signature” [10] as appearance model. The salient region is considered as the global proposal distributions where particles are sampled. The weight of each reliable particle is adaptively selected by saliency information.

**A. Extraction Saliency Features**

In comparison to [5], the image signature derived from human vision attention [10, 11] is a simple, very compact and global descriptor, which possesses important properties related to the foreground of an image. Given a video frame which exhibits the following structure:

\[
\text{Img} = \text{Fog Bag, Img, Fog, Bag} \in \mathbb{IR}^N \quad (8)
\]

where \( \text{Fog Bag} \) represent the foreground and the background respectively. The image signature is defined as

\[
\text{image signature}(\text{Img}) = \text{sign}(\text{DCT(Img)}) \quad (9)
\]

It is assumed that a foreground target is visually conspicuous relative to its background, and then saliency features can be extracted by smoothing the squared reconstructed image signature

\[
\text{Sal} = g * (\text{IDCT(image signature)} \circ \text{IDCT(image signature)}) \quad (10)
\]

where \( g \) is a Gaussian kernel, \( \circ \) is the Hadamard (entry wise) product operator.

**B. Adaptive Saliency Selection Reliable Particles**

It is easy for saliency features (10) to be represented by histogram. For the same consideration as (4), the background, target and the hypothetical observation of each particle can be represented by M bin saliency histograms \( \{ H_{\sigma_i}^{\text{bf}} \}_{\sigma_i} \), \( \{ H_{\sigma_i}^{\text{bf}} \}_{\sigma_i} \) and \( \{ H_{\sigma_i}^{\text{bf}} \}_{\sigma_i} \), here \( \sum_{\sigma_i} H_{\sigma_i}^{\text{bf}} = 1 \), \( \sum_{\sigma_i} H_{\sigma_i}^{\text{bf}} = 1 \) and \( \sum_{\sigma_i} H_{\sigma_i}^{\text{bf}} = 1 \). Then, the weight of particle \( X'_i \) can be computed as

\[
\pi'_{\sigma_i} \propto \exp \left\{ -\lambda_{\sigma_i} \left( d_{\sigma_i}^{\text{bf}}[H'_i, H_s] - d_{\sigma_i}^2[H'_i, H_s] \right) \right\} \quad (11)
\]

where \( \lambda_{\sigma_i} \) is a constant determined in practice; the distance measure \( d_{\sigma_i} \) is derived from the Bhattacharyya coefficient. Moreover, reliable particles are selected by saliency features, the weight of each particle can be obtained by

\[
\pi'_{\sigma_i} \propto \exp \left\{ -\lambda_{\sigma_i} \left( d_{\sigma_i}^{\text{bf}}[\text{Sal}_i, \text{Sal}_s] - d_{\sigma_i}^2[\text{Sal}_i, \text{Sal}_s] \right) \right\} \quad (12)
\]
where \( \lambda \) is a constant determined in practice. The distance measure \( d_{sad} \) is derived from the \( t \) distance which is defined by

\[
d_{sad} (Sal_1, Sal_2) = \| \text{sign}(\text{imagesignature}_1) - \text{sign}(\text{imagesignature}_2) \|_0
\]

(13)

where \( \| \) is the Hamming distance. The combined weight of each particle can be measured by

\[
\pi_k^{(s)} = \pi_k^{(H^j)} \cdot \pi_k^{(Sal)}
\]

(14)

More precisely, despite of the distinctiveness of saliency feature, in some situation, the tracked target might be detected as being less salient than background. Therefore, we have taken advantage of the saliency feature by adaptively combining it with color cue to measure the weight of each particle. The combination needs to be carried out carefully.

Intuitively, automatically weighting feature respective contribution to the likelihood function, the all likelihood is generated by linear combination of the single one as the likelihood mixture

\[
\pi_i^c = \alpha_c \pi_i^{(c)} + (1 - \alpha_c) \pi_i^{(s)}
\]

(15)

The weight parameter \( \alpha_c \) is evaluated by using the following formula:

\[
\alpha_c = \frac{\pi_i^{(c)}}{\pi_i^{(c)} + \pi_i^{(s)}}
\]

(16)

where \( \pi_i^{(c)} \) and \( \pi_i^{(s)} \) are the mean value of \( \{ \pi_i^{(c)} \}_{i=1}^N \) and \( \{ \pi_i^{(s)} \}_{i=1}^N \) respectively.

It is shown that only the particles which tightly scatter around the target contribute more to estimate the target state. These particles are selected to estimate target state using (15) by saliency feature. The higher the weight \( \pi_i \), the more reliable the selected particles are. Following [12], the reliable particles are obtained by

\[
\pi_i^{new} = \begin{cases} C_s \pi_i^c & \text{if } \pi_i^c \geq \mu + \delta, \\ 0 & \text{otherwise} \end{cases}
\]

(17)

where \( \pi_i^{new} \) is the new weight, and \( C_s \) is normalization constant, \( \mu \) and \( \delta \) are the mean values and the standard deviation of all the particles.

C. Tracking

Since object tracking can viewed as a problem of estimating the target state from current image sequences, through (6), (17), the state are can be estimated by

\[
E(X_t) = \sum_{i=1}^N \pi_i^{new} X_i^t
\]

(18)

In the update stage, only when the observation likelihood is reliable enough, the new target model can be updated with saliency selection using the following formula:

\[
H_t^{new} = \begin{cases} \alpha H_t + (1 - \alpha) H_{t-1} & \text{if } \max_{i} \pi_i^c < \epsilon \\ H_t & \text{otherwise} \end{cases}
\]

(19)

where \( \alpha \) is the linear combination between \( H_t \) model and \( H_{t-1} \) model; \( H_{t-1} \) is the linear combination between \( H_t \) model and \( H_{t-1}^{new} \) model; \( \alpha \) is the learning factor, and \( \epsilon \) is threshold.

IV. EXPERIMENTAL RESULTS

The proposed approach is implemented and tested on 7 challenging sequences which are publicly available. The “WalkByshop1cor” and “Meet_Split_3rdGuy” sequence come from the [13]. The “David indoor” sequence is provided in [14]. The “Tiger1” sequence is from [15]. The “Occluded face2” sequence is from [16]. The “Soccer” and “Shaking” has been used in [17]. These present large illumination changes, similar color object nearby, and partial occlusions. The tracking results are compared with color based tracker (CPF) [4], improved Itti tracker (ITPF) [8], the frequency based method (FFP) [9]. The HSV color histogram is a 16×16×16 dimension, and saliency histogram is 16 bins. In all experiments, the tracker object is manually initialized in the first frame, followed by continuous tracking to the remaining part of the sequence, and 100 particles are used for tracking. All experiments are implemented in MATLAB, which runs at average 20 frames per second on a Pentium Dual Core 3.3 GHz CPU with 4GB RAM. In the following the results of the examples are illuminated, and the analysis of the results is given by qualitative analysis and quantitative analysis.

A. Qualitative Analysis

Background with similar color: The “WalkByshop1cor” contains similar color and some other pedestrians in the background which can be considered as a disturbance. The tracked target is a human face from far to near. Tracking results are depicted in Fig. 1. The top row shows CF tracker; the second row is ITPF tracker; the third row is FFP tracker, the bottom row is our method. We can see that the tracker successfully follows the target with a reduced number of particles as few as only 75 without gradual drifting, which significantly reduces the computation load, because a saliency selection is implemented from saliency region of the current frame in presence of the background color distribution is similar to that of the face. However, this makes CPF tracker lose the target completely, and other trackers show poor results. Although they can give the goodness throughout the tracking process, it is unstable and apt to deviate to other wrong locations.
Slow illumination, large scale and pose variation: The sequence “David indoor” varied from dark to bright areas, and the target head size ranges from 88×105 pixels to 44×52 pixels gradually. The comparison results are shown in Fig. 2. It is shown that the CPF can fail to track the object reliably; ITPF and FPF tracker achieve better performance than CPF tracker. This could be because they are using salient feature, which can handle these difficulties and improve the tracking performance. However, their limitation is, in some frames, the tracker totally loses the target. The reason for the poor performance is that improved Itti model only throws away a lot of local information, and copes with abrupt motion. It is observed that the proposed method is able to track the head steadily throughout the sequence. The generative saliency features by image signature has been demonstrated to handle these troubles. In addition, only a little number of particles with saliency selection is used throughout the tracking. So our algorithm performs well.

Drastically lighting and short occlusion: The target in “WalkByshop1cor” enters the drastically lighting region, and leaves to dark region. This makes it particularly challenging to all tracker. Noting that the CPF, ITPF, and FPF tracker exhibits poor result, the performance of the proposed tracker is still acceptable. Two points are made out: First, saliency features from image signature is more resistant to these disturbances, since it obtained a much cleaner segmented target than all other methods. Second, the proposed algorithm obtains more reliable particles, resulting in more accurate results.

$$e_i = \frac{1}{K} \sum_{k=1}^{K} \left(1 - \frac{\sum_{q} w_{ij}^q b_{ij}^q b_{ij}^q}{\sum_{q} w_{ij}^q + \sum_{q} w_{ij}^q - \sum_{q} w_{ij}^q b_{ij}^q} \right)$$
where $e_i$ was used as the measurement of tracker performance. The object in frame $k$ is accurately estimated if $e_i > \text{Threshold}$. The tracking performance is shown in Table I. For the most sequences, we labeled the ground truth target bounding box. For the “Occluded face2” sequence, the authors of [15] provided the ground truth. Table I gives the success rate of the four approaches on above test sequences. Obviously, our tracker shows superior tracking performance in dealing with various challenges over other three trackers.

Table I. Average Tracking Error of the Four Tracking Compared

<table>
<thead>
<tr>
<th>Sequence</th>
<th>CPF</th>
<th>ITPF</th>
<th>FPF</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>WalkByshop1cor</td>
<td>36</td>
<td>45</td>
<td>68</td>
<td>80</td>
</tr>
<tr>
<td>David indoor</td>
<td>40</td>
<td>56</td>
<td>68</td>
<td>85</td>
</tr>
<tr>
<td>Meet_spli_3rdGuy</td>
<td>47</td>
<td>54</td>
<td>33</td>
<td>66</td>
</tr>
<tr>
<td>Tiger1</td>
<td>53</td>
<td>55</td>
<td>72</td>
<td>81</td>
</tr>
<tr>
<td>Occluded face2</td>
<td>44</td>
<td>58</td>
<td>66</td>
<td>67</td>
</tr>
<tr>
<td>Soccer</td>
<td>36</td>
<td>29</td>
<td>47</td>
<td>63</td>
</tr>
<tr>
<td>Shaking</td>
<td>29</td>
<td>37</td>
<td>57</td>
<td>74</td>
</tr>
</tbody>
</table>

TABLE II. Average Center Location Error of Four Tracking Compared

<table>
<thead>
<tr>
<th>Sequence</th>
<th>CPF</th>
<th>ITPF</th>
<th>FPF</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>WalkByshop1cor</td>
<td>47</td>
<td>34</td>
<td>31</td>
<td>25</td>
</tr>
<tr>
<td>David indoor</td>
<td>53</td>
<td>42</td>
<td>29</td>
<td>12</td>
</tr>
<tr>
<td>Meet_spli_3rdGuy</td>
<td>62</td>
<td>38</td>
<td>24</td>
<td>13</td>
</tr>
<tr>
<td>Tiger1</td>
<td>43</td>
<td>35</td>
<td>28</td>
<td>20</td>
</tr>
<tr>
<td>Occluded face2</td>
<td>39</td>
<td>33</td>
<td>24</td>
<td>17</td>
</tr>
<tr>
<td>Soccer</td>
<td>57</td>
<td>39</td>
<td>29</td>
<td>23</td>
</tr>
<tr>
<td>Shaking</td>
<td>45</td>
<td>34</td>
<td>27</td>
<td>15</td>
</tr>
</tbody>
</table>

Center error is a pixel-wise Euclidean distance from the center of estimated rectangle to the center of ground truth. Table II gives the average center location error of the four approaches on above test sequences. Obviously, our tracker is very stable and robust, while other trackers are unstable.

From foreseen experimental results, it is seen that our tracker is able to track target effectively under various challenging conditions, such as background with similar color, drastically lighting and short occlusion. The reasons for this are as follows: First, in contrast to other tracking approaches, saliency from image signature inspired approach can provide discriminative features that more accurately extract target from background, and are more resistant to diverse disturbances. Second, reliable particles with saliency selection can not only reduce the numbers of particles throughout the tracking processing but also are assigned to discriminative weights which correspond to the likelihood that is the true location of the object, which gives the success of the proposed tracker.

V. CONCLUSION

In this work, an adaptive probabilistic tracking method with saliency selection reliable particle is proposed. It is based on the integration of saliency features for describing the appearance model from image signature into the particle filter framework. The new weight of particle is evaluated through saliency measurement using Hamming distance between the target model and each hypothetical observation, but also through background model and each hypothetical observation. The updating model and the estimated state of target are implemented by saliency selection. Experimental results with different sequences demonstrate the proposed method can get good tracking results, while only requiring fewer particles.

REFERENCES

Linshan Liu received the B.Eng. degree in industrial electric automation from Northeastern University, Shenyang, China, in 1996, the M.Eng. degree in control theory and control engineering from Tianjin University, Tianjin, China, in 2005. He is currently working toward the Ph.D. degree in control theory and control engineering in state key laboratory of management and control complexity systems, in the Institute of Automation, Chinese Academy of Sciences, Beijing, China. His current research interests include computer vision, visual attention model, image processing and object tracking.

Suiwu Zheng received the B.Eng. degree in from Harbin Engineering University, Harbin, China, in 2005, the M.Eng. and Ph.D. degrees from the Institute of Automation, Beijing, Chinese, in 2010. He is currently an Assistant Professor with the Institute of Automation, Chinese Academy of Sciences. His current research interests include manifold learning methods for vision tracking, visual attention, robotics vision and intelligent agents, and pattern recognition.

Hong Qiao received the B.Eng. degree in hydraulics and control, the M.Eng. degree in robotics from Xian Jiaotong University, Xian, China, the M.Phil. degree in robotics control from the Industrial Control Center, University of Strathclyde, Strathclyde, U.K., and the Ph.D. degree in robotics and artificial intelligence from De Montfort University, Leicester, U.K., in 1995. She is currently a Professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. She first proposed the concept of the attractive region in strategy investigation that she successfully applied for robot assembly, robot grasping, and part recognition, which is reported in Advanced Manufacturing Alert (Wiley, 1999). Her current research interests include information based strategy investigation, robotics and intelligent agents, animation, machine learning, and pattern recognition. Dr. Qiao was a member of the Program Committee of the IEEE International Conference on Robotics and Automation from 2001 to 2004. She is currently an Associate Editor of the IEEE Transactions on Systems, Man, and Cybernetics, PartBand the IEEE Transactions on Automation Science and Engineering.