Target Salient Confidence for Visual Tracking

Hongkai Chen, Xiaoguang Zhao, and Min Tan
The State Key Lab. of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences
Beijing 100190, China
Email: {hongkai.chen, xiaoguang.zhao, min.tan}@ia.ac.cn

***Abstract*—In this paper, a novel visual tracking algorithm that uses the target salient confidence (TSC) is proposed. Two contributions are summarized as follows. First, we put forward a novel target salient confidence (TSC) model which combines the static saliency map (SSM) based on the selective visual attention model, motion attention map (MAP) and the target prior confidence (TPC). Second, we propose to use the target salient confidence for particle filter. It manipulates the distribution expressed by the particle cloud towards a better match with the target salient confidence model. In this way particle sampling can be locked on those regions with higher target salient confidence. Experiments in some video sequences show that the target salient confidence is useful for visual tracking and our algorithm is effective.**

***Index Terms*—Visual Attention; Target Salient Confidence; Visual Tracking;**

I. Introduction

As well known, visual tracking has been widely studied in recent years and has many successful applications such as human robot interaction, surveillance, and vision-based mobile robots [1]. Especially in human robot interaction, there is a great significance for tracking as a partner of human beings.

There are plenty of works published during the past decades. Such as interest points based tracking algorithm [2]–[4], Mean Shift algorithm [5] (or CamShift algorithm [6]), particle filter tracker [7] and online learning approaches [8]–[10]. However, visual tracking is still a difficulty task owing to some reasons, such as illumination variation, occlusions, deformations, scale and appearance changes, etc. Hence, visual tracking needs to be further studied in computer vision.

There are two factors influencing human visual attention: bottom-up (also called stimulus-driven) and top-down (also called task-relevant) influences [11]. Cognitive and psychological findings reveal that the human perception is selective and attentive [12]. The human visual system is able to quickly and accurately discriminate the salient areas of the image, and can find out the important elements in the scene. Besides, it also can properly make visual attention focused on the important areas which is important to human visual perception process. Image visual saliency reflects the saliency in different areas of the image. The selective attention mechanism ignores other salient objects by selecting a few salient objects to process priority. It has attracted many computer scientists to join in this field. Itti *et al.* proposed the first computation framework of bottom up attention based on [13], they pointed out that the location is different from its surrounding regions based on the center-surround mechanism; this model will produce a saliency map where each regions competes to get saliency, however, only one position will stand out [14]. This kind of computational model can be simulated as the human visual selective attention mechanism and can be applied to engineering easily. In addition, more and more visual attention model have been proposed in recent years [11], however, most of published works are focused on bottom-up models, while few published works about top-down (task-relevant) attention modelling.

Note that it is easy for human visual system to track interesting objects in certain environment. Recently, Wolfe and Horowitz revealed that tracking targets is an important top-down (stimulus-driven) factor for guiding human visual attention [15]. Mahadervan and Vasconcelos proposed the saliency hypothesis of tracking and proposed a discriminant saliency based framework for tracking [16]. Inspired by human visual perception system, Frintrop and Kessel proposed a most salient region tracker where a top-down guided visual search module is equipped to find the most salient region for tracking [17]. To overcome the influence of illumination changes, Zhang *et al.* proposed to use visual attention mechanism to detect the visual saliency of particles, and regard saliency as the weight of the particles [18].

In this paper, we discuss a visual tracking algorithm based on biological think and propose a novel visual tracking algorithm using target salient confidence. The main contributions of our work are summarized as follows. First, we propose a novel target salient confidence model. Second, we propose to use target salient confidence for particle filter. We can manipulate the distribution expressed by the particle cloud towards a better match with the target salient confidence model. In this way particle sampling can be locked on those regions with higher target salient confidence. Experiments in some video sequences show that the target salient confidence is useful for visual tracking and our algorithm is effective. Besides, it can be used in real-time applications.

The followings are the structure of the rest paper. Section II gives the details of our target salient confidence model. In Section III, we propose our tracker and we will show how to use the target salient confidence for particle sampling. Experimental results are given in Section IV. Finally, we draw some conclusions in Section V.

This work has been partially supported by the National Natural Science Foundation of China (Grant 61271432, 61273337) and the national CNC machine tool special Foundation of China (Grant No. SK201201A35-02).

978-1-4799-5835-1/14/$31.00 ©2014 IEEE

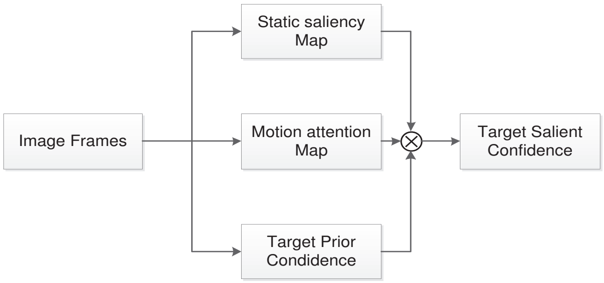

Fig. 1. The flow chart of the target salient confidence (TSC) model.

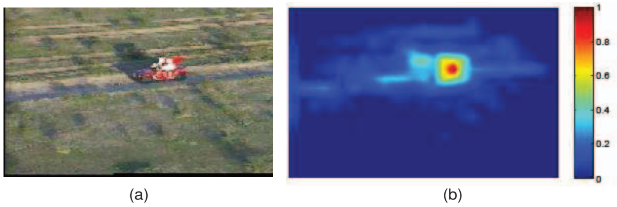

Fig. 2. (a) shows the original image and (b) represents the static saliency map (SSM) according the proposed selective visual attention model.

II. Target Salient Confidence Model

Inspired by some biological and psychological findings, we present a novel top-down visual attention computational model for tracking. In visual tracking, a target-relevant visual application, the region of the target is known in the 1st image frame. Meanwhile, the surrounding regions of target are regard as background. The full image is considered as the input of the bottom-up information, while the target region is an important target-relevant cue. We propose that the target is most likely to appear in those regions with high target salient confidence.

In this section, we will show how to learn a target salient confidence map. And the framework of proposed target salient confidence model is shown in Fig. 1.

A. Selective Visual Attention Model

Image regions which human eye focuses on tend to have higher visual saliency. That is to say, compared with the surrounding area, the salient region is much more unique (shown in Fig. 2). In order to predict human selective attention region, we present a visual attention model to calculate the saliency of an image. In this subsection, the proposed computational model of visual attention mainly has the following two key steps:

1. Image patches representation for image: an local image area will be assigned similar saliency values, and the entire objects can be highlighted uniformly;

2. Saliency based on the local color contrast: visual saliency of a local region depends mainly on its color contrasts to the nearby regions, while color contrasts to distant regions tend to be less significant.

The above two opinions are supported by biological and psychological findings: the human eye receptive field will produce strong reaction for visual information with high contrast (such as the central bright surrounding dark, or the central red surrounding green). Rajashekar *et al.* showed that the visual saliency of attention regions depend mainly on its contrasts to the nearby regions in response of some features [19]. Besides, the relevant research of Gestalt principles of perceptual organization [20] showed that it is more likely to be known as a part of the visual distinctiveness for a local area of an image, if it is similar to its nearby regions, and dissimilar to its far regions. And the similarity degree of internal regions in a salient area is greater than that of the salient area to the other regions in a same image.

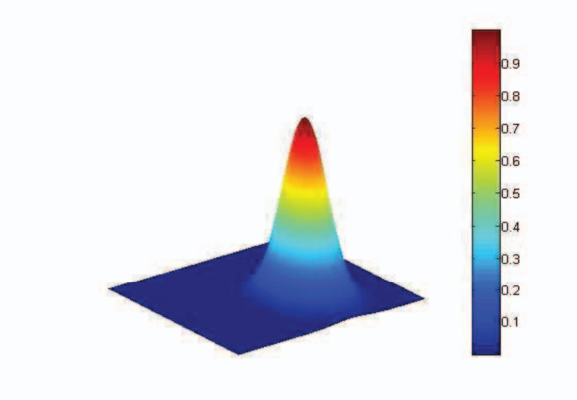

Fig. 3. This picture shows the target prior confidence according to the tracked target image position of the last image frame.

The specific calculation process of visual saliency for an image is following:

1) Patch representation: inspired by [21], we divide an $W \times H$ image I into $L = \lfloor W/k \rfloor \times \lfloor H/k \rfloor$ blocks (the size of each image patch is $k \times k$), and patches are denoted by $p_i (i = 1, 2, \cdots, L)$. Represent each image patch as a column vector $f_i (f_i \in R^{3k^2})$ according to the RGB color channels. Finally, the image can be represented as sample matrix expressed by $A = [f_1, f_2, \cdots, f_L]$.

2) Dimensionality reduction: for processing speed, we use PCA [22] to reduce dimension of matrix A. Then the image I can be approximately represented by $U = [U_1, U_2, \cdots, U_L]$ where $U_i = [u_{1i}, u_{1i}, \cdots, u_{di},]^T (d \ll L, i = 1, 2, \cdots, L)$.

3) Visual Saliency: the proposed image visual saliency of a local region depends mainly on its color contrasts to the nearby regions, while color contrasts to distant regions tend to be less significant. Specifically, the saliency of an image patch p_i is defined as:

$$S_i = \sum_{j \neq i} \omega(z_{ij}) \cdot D(p_i, p_j) \tag{1}$$

where $D(p_i, p_j)$ is the color space Euclidean distance of image patch p_i and p_j. z_{ij} is the Euclidean space distance of image patch p_i and p_j. $\omega(x)$ is a weighted function defined by $\omega(x) = \alpha \cdot e^{-\frac{|x|^2}{\sigma^2}}$. α is a normalization constant and σ controls the size

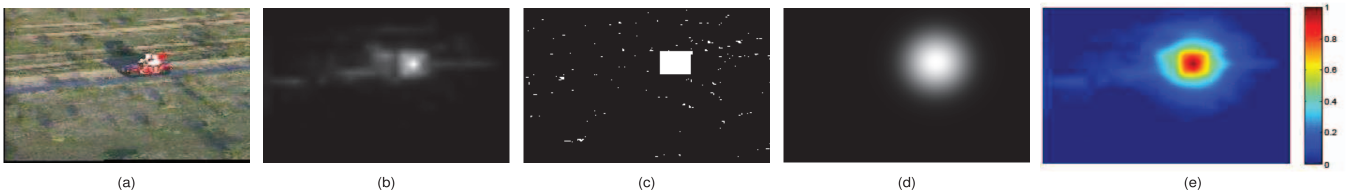

Fig. 4. This figure shows the basic flow of the proposed target salient confidence model. In figure, (a) is the source image. (b) shows the static saliency map (SSM) according to the proposed selective visual attention model. (c) is the motion attention map (MAP). (d) is the target prior confidence map according to the tracked target image position of the last image frame. (e) is the final target saliency map via incorporating the static saliency map (SSM), motion attention map (MAP) and the target prior confidence (TPC).

of the image local region which will influence the image visual saliency.

4) Static saliency map: given an $W \times H$ input image I, it will produce a $\lfloor W/k \rfloor \times \lfloor H/k \rfloor$ static saliency map after completing the above steps. To get a full size saliency map, we should resize it to a $W \times H$ static saliency map $S_s^{'}$. Finally, use the following rule to produce the final static saliency map:

$$S_s(x,y) = max\{0, S_s^{'}(x,y) - S_{su}^{'}\} \tag{2}$$

where the arithmetic mean value of the $S_s^{'}$ is $S_{su}^{'}$.

B. Motion Attention Map

Human attention is more sensitive to moving object than static object. There are many ways to obtain motion feature such as optical flow method [23], block matching algorithm [24] and frame difference [25]. To meet the real-time applications, the frame difference is a simply way to get the motion salient region.

Given a successive image sequence $\{\ldots, f_{i-1}, f_i, f_{i+1}, \ldots\}$, it will produce the difference of images using the formula of

$$Diff_{i,j}(x,y) = |f_i(x,y) - f_j(x,y)|(j = i+2) \tag{3}$$

where $Diff_{i,j}(x,y)$ is the difference image that is between the i-th and the j-th image frame.

Neri *et al.* pointed out that we can reduce output noise by setting those values of difference image to zero which are less than an appropriate threshold ε [26]. In the view of the fixed camera, the motion attention map can be computed as

$$S_m(x,y) \quad = \quad \begin{cases} 1, & Diff_{i,j}(x,y) > \varepsilon \\ 0, & others \end{cases} \tag{4}$$

where $S_m(x,y)$ is the MAP of an image frame.

However, it seems to be unreliable under the dynamic scene. Hence, we give an approximate method to form a MAP. Firstly, using the following rule to produce the approximate image difference:

$$Diff_{i,j}(x,y) \quad = \quad \begin{cases} 1, & (x,y) \in I_{region} \\ 0, & others \end{cases} \tag{5}$$

where I_{region} region of the target at the previous image frame.

Then we use the morphology operation (opening operation and closing operation) to reduce the noise and find out the maximum external rectangle. Finally, it will produce the approximate final motion attention map S_m.

C. Target Prior Confidence

The tracking problem can be viewed as an estimation problem of object position. On one hand target prior confidence can be viewed as the simulation of the focus of human visual attention. On the other hand, there is likely to be several areas with high visual saliency in the static saliency map, then it is necessary to use target prior confidence to tune the static saliency map. After that, the saliency of the task-relevant regions will be strengthened, meanwhile, the saliency of other areas will fade.

In the current frame f_t, it is very easy to get the target position x^* (the tracked target center). Let the notation x be the object position in the next frame f_{t+1}. We can make such a weaker assumption that object location x in the next frame is closely to x^*. This assumption is a task-relevant, top-down cue. If x is close to x^*, it is more likely to be the position of the target in the coming image frame. Under the very weaker assumption, the prior of the object location x is

$$p(x) = \gamma \cdot e^{-\frac{(x-x^*)^2}{\sigma^2}} \tag{6}$$

where σ is a scale parameter. γ is a normalization coefficient that convert the prior probability $p(x)$ to the range from 0 to 1.

So, the target prior confidence shown in Fig. 3 can be formulated as

$$S_p(i,j) = p(x_{ij}, x^*) \tag{7}$$

where $S_p(x,y)$ is the prior confidence value in the j-th column and i-th row of the prior confidence map. And x_{ij} is the target position in the image frame f_{t+1}.

D. Target Salient Confidence Map

After having got the SSM, the MAP and the PCM, we can easily obtain the final full target salient confidence map by

$$S_f = \alpha_1 \cdot S_s + \alpha_2 \cdot S_m + \alpha_3 \cdot S_p \tag{8}$$

where $\alpha_i (i = 1,2,3; \alpha_i \in [0,1])$ is a weight coefficient and $\sum_i \alpha_i = 1$.

Fig. 4 illustrates the basic flow of our target salient confidence model which is full use of the task-relevant cue.

III. Proposed Object Tracking Framework

Our tracking algorithm is based on the particle filter proposed in [7]. In this section, we propose to use target salient confidence (TSC) for particle filter tracking. Based on the model proposed in the section II, we can manipulate the distribution expressed by the particle cloud towards a better match with the target salient confidence model.

In the first image frame, the initial position of an target is known. In the prediction stage of the particle filter, the samples are propagated through a dynamic model and target salient confidence map. In the update stage, the observation likelihood for each sample is defined by the object description (color histogram: $10 \times 10 \times 10$ bins in HSV color space).

At the time t, we put to use J $(here : J = 300)$ particles. The form os each particle is:

$$\vec{\phi}_t^j = (\vec{s}_t^j, \vec{q}_t^j, \pi_t^j) j = 1, 2, \cdots, J \tag{9}$$

where $\vec{s}_t^j$ is a particle state vector $\vec{s}_t^j = [x, y, w, h]^T$. (x, y) is the position of the target region image center, w is the width and h represents the height of each particle sample, respectively. $\vec{q}_t^j$ is the appearance model (color histogram: $10 \times 10 \times 10$ bins in HSV color space) that represents the appearance model of the particle region; π_t^j is an observation likelihood which determines the similarity of the particle with the target [27].

In the prediction stage of the particle filter, the samples are propagated through a second order auto regressive process model. The dynamic model is

$$s_{t+1} = A s_t + B s_{t-1} + C v_t \tag{10}$$

where the multivariate Gaussian random variable is denoted by v_t; A,B,C are the coefficient matrix, respectively [27].

In the coming image frame f_{t+1}, after propagating through an above dynamic model, actually there are some useless particles in traditional particle filter tracker. We use the target salient confidence proposed in Section II to tune sampling.

Firstly, sort the particles by descending according to their corresponding target salient confidence (now the particle set is $S_{t+1} = \{\vec{s}_{t+1}^j, j = 1, 2, \cdots, J\}$). In particle, we can choose some most salient particles via setting an appropriate target salient confidence threshold θ (suppose there are $K (K \leq J)$ particles in the chosen subset) and put them into a subset $S_{t+1}^{'} = \{\vec{s}_{t+1}^j, j = 1, 2, \cdots, K\}$. Then we added $J - K$ particles $\vec{s}_{t+1}^1$ into the subset and obtain a new particle set S_{t+1}. It can be formulate as

$$S_{t+1} = \{\vec{s}_{t+1}^j, j = 1, 2, \cdots, K\} \cup \{\underbrace{\vec{s}_{t+1}^1, \vec{s}_{t+1}^1, \cdots, \vec{s}_{t+1}^1}_{J-K}\} \tag{11}$$

Finally, some least salient particles are discarded and some new particles at salient ones are cerated. Then the target salient confidence is used to manipulate the distribution expressed by the particle cloud towards a better match.

Over the given time t, the color distribution $\vec{q}_t^j = \{\vec{q}_t^j(u), u = 1, 2, \cdots, N\}$ at location d_j is given by

$$\vec{q}_t^j(u) = K \sum_{i=1}^{I} k(\frac{d_j - l_i}{a}) \delta[b_t(l_i) - u] \tag{12}$$

where the total number of pixels is denote by I in particle $\vec{s}_t^j$, δ is the Kronecker delta function; K is a normalization coefficient which ensures $\sum_{u=1}^{N} \vec{q}_t^j(u) = 1$; $b_t(l_i)$ is used to produce the histograms and $k(\cdot)$ is a weighted function; the function of $k(\cdot)$ is that it assigns larger weights to the pixels that are closer to the target center [27].

Every time we should assign new weights for all particles according to the following rule:

$$\pi_t^j = c \cdot e^{-\lambda \cdot D^2(\vec{q}_t^j, \vec{q}^*)} \tag{13}$$

where $D(\vec{q}_t^j, \vec{q}^*) = \sqrt{1 - \sum_{u=1}^{N} \sqrt{\vec{q}_t^j(u) \vec{q}^*(u)}}$ is the Bhattacharyya coefficient which measures the distance between two distributions [10]. $\vec{q}^*$ is the reference target model. c is a normalization coefficient. λ is a scale constant.

Then according to the weights of particles, we do the resample operation. Finally, the current target state $\vec{s}_t^*$ is produced by:

$$\vec{s}_t^* = \sum_{j=1}^{J} \pi_t^j \cdot \vec{s}_t^j \tag{14}$$

To adapt to object appearance changes, we use the following rule to update the target model:

$$\vec{q}^* = \mu \vec{q}^* + (1 - \mu) \vec{q}_t^* \tag{15}$$

where $\mu > 0$ is a parameter that denote learning rate. In the t-th image frame, $\vec{q}_t^*$ represents the target model.

IV. Experiments

In this section, to verify the validity of the proposed algorithm, we implement our tracker using MATLAB on a 3.1GHz PC (runs in 10 fps). The experimental test videos are from VIVID [28], BoBoT [29] and PaFiSS [30] to make it has the potential to apply on a moving platform such as UAV (Unmanned Aerial Vehicle) and mobile robot. First, we test proposed algorithm working on these video sequences named redteam, ball, Exit1. Then, we also compare our algorithm with some other algorithms which are Mean Shift tracker [5], the color-based particle filter (CPF) algorithm [10] and one recent approach MIL [9].

To quantitatively evaluate the proposed method, we use two metric: the precision rate and success rate in this paper. The center location error (CLE) that is defined as the average Euclidean distance between the ground truths; the CLE is for precision and bounding box overlap as evaluation metric for success rate [31].

The overlap score is defined as

$$S = \frac{|r_t \cup r_a|}{|r_t \cap r_a|} \tag{16}$$

where r_a is the bounding box of the ground truth of the test data and r_t is the bounding box of the tracked results; $\cap$ is the intersection of two image regions while $\cup$ are is union of two image regions; $|\cdot|$ represents the total number of pixels in an image region [31].

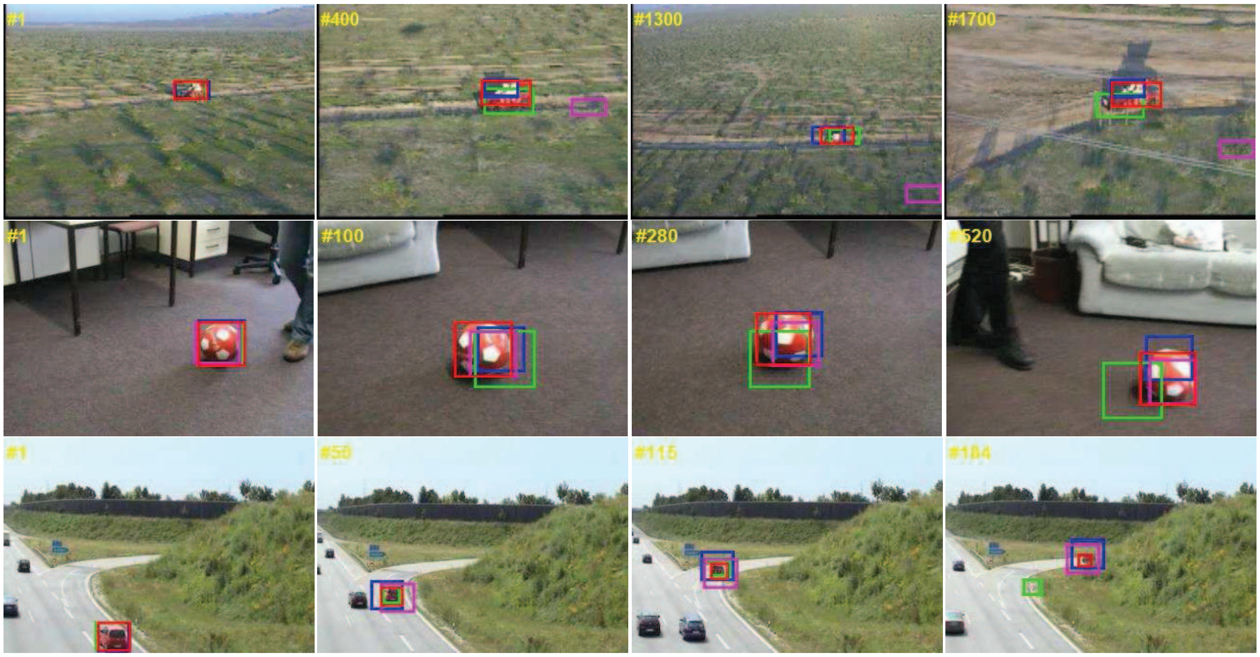

Fig. 5. The first, second, third row are the screenshots of tracking result of redteam, ball, Exit1, respectively. The results of proposed tracker, color-based particle filter tracker, MIL tracker and Mean Shift tracker are indicated by red, green, blue and pink boxes, respectively. Table 1. includes quantitative results for all trackers we evaluated.

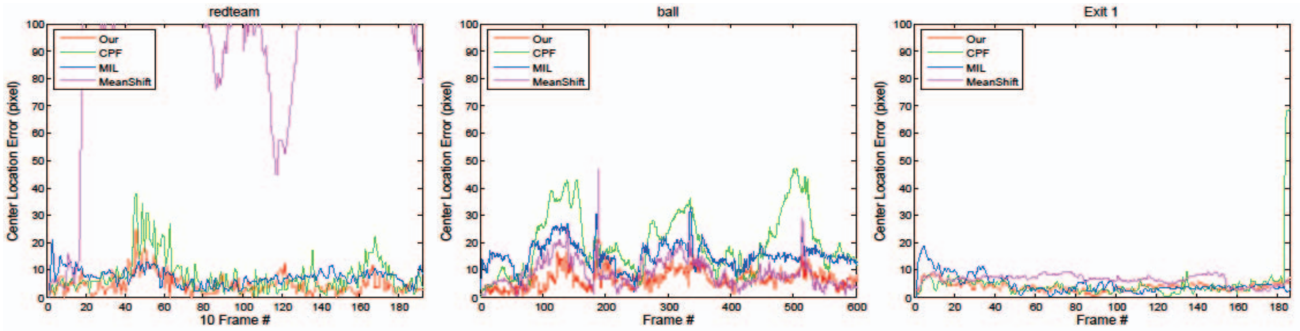

Fig. 6. The center location error of tracking results: the center location error results from left to right are redteam, ball, Exit1, respectively.

Fig. 5 is the screenshots of tracking result of test video sequences. The CLE is shown in Fig. 6. Then we count the number of successful image frames whose overlap score S is larger than the given threshold S_{thr} (we set $S_{thr} = 0.4$). To measure the accuracy of different algorithms in test video sequences, we define the accuracy as the ratio between the successful frames and the total frames. Table I gives the quantitative tracking result.

The first video is from VIVID set which is collected on a helicopter. In this test video sequence, a small jeep drives on straight road through the desert and turns a corner at the end. It undergoes the illumination variation, scale changes and long shadows. Fig. 5 and Fig. 6 clearly show that the Mean Shift tracker drifts. The CPF, MIL and the proposed tracker can successfully track the target due to the relative simple background and uniform target movement. Note that the CPF and MIL tracker produces partial drift. However, our approach has the minimum CLE.

The second video is from BoBoT data set which is collected by a moving cam. In this video, a red ball with white spots is kicked back and forth. The moving target undergoes the scale changes, rotation, fast direction changes and target template changes. From the Fig. 5 and Fig. 6 we can clearly see that the movement direction of the target produces three obvious mutations in the test video. We can see that our method on average has the minimum CLE.

The third video is from PaFiSS which comes from a collection of vehicle captures in the low level. In this video, a red car with significant scale and appearance changes appears in the field of view, and then gradually fled. Besides, there

TABLE I
TRACKING RESULTS OF TEST VIDEO SEQUENCES

Sequence	Frames	Mean Shift	CPF	MIL	Our
redtaem	1918	12.5%	84.9%	60.94%	**91.15**%
ball	602	30.73%	54.98%	78.57%	**99.36**%
Exit1	186	6.99%	88.71%	14.52%	**93.55**%
Average		16.44%	80%	39.33%	**95.52**%

are some other interferential targets in this video. The Mean Shift and MIL tracker can not handle the significant scale changes problem, while the CPF method drifts in the end of the video. However, our approach is always able to track the target. Besides, we on average get the minimum CLE.

Besides, compared with the CPF tracker, our algorithm achieves better result via using the target salient confidence.

V. CONCLUSION

In this paper, we come up with a novel visual tracking approach. The core of our tracker is the target salient confidence modeling. First, under the biological and psychological findings, we propose our selective visual attention using saliency based on local color contrast. Then, motion attention map and target prior confidence are incorporated to produce a target salient confidence map in which the target will be highlighted. The target salient confidence is used to manipulate the distribution expressed by the particle cloud towards a better match. It can deal with some difficult situations with illumination variation, fast moving direction changes, scale and appearance changes. Experiments in some video sequences show that the target salient confidence is useful for visual tracking and our algorithm is effective. Besides, it can be used in real-time applications such as UAV and mobile robot.

REFERENCES

[1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *Acm Computing Surveys (CSUR)*, vol. 38, no. 4, p. 13, 2006.

[2] S. Battiato, G. Gallo, G. Puglisi, and S. Scellato, "Sift features tracking for video stabilization," in *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*. IEEE, Conference Proceedings, pp. 825–830.

[3] T. Mathes and J. H. Piater, *Robust non-rigid object tracking using point distribution manifolds*. Springer, 2006, pp. 515–524.

[4] W. He, T. Yamashita, H. Lu, and S. Lao, "Surf tracking," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, Conference Proceedings, pp. 1586–1592.

[5] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 5, pp. 564–577, 2003.

[6] J. G. Allen, R. Y. Xu, and J. S. Jin, "Object tracking using camshift algorithm and multiple quantized feature spaces," in *Proceedings of the Pan-Sydney area workshop on Visual information processing*. Australian Computer Society, Inc., 2004, pp. 3–7.

[7] P. Prez, C. Hue, J. Vermaak, and M. Gangnet, *Color-based probabilistic tracking*. Springer, 2002, pp. 661–675.

[8] H. Grabner, M. Grabner, and H. Bischof, "Real-time tracking via on-line boosting," in *BMVC*, vol. 1, Conference Proceedings, p. 6.

[9] B. Babenko, M.-H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 8, pp. 1619–1632, 2011.

[10] Z. Kalal, J. Matas, and K. Mikolajczyk, "Pn learning: Bootstrapping binary classifiers by structural constraints," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, Conference Proceedings, pp. 49–56.

[11] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 1, pp. 185–207, 2013.

[12] R. L. Solso, O. H. MacLin, and M. K. Maclin, *Cognitive psychology*. Allyn and Bacon Boston, 1995.

[13] C. Koch and S. Ullman, "Shifts in selective visual-attention - towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.

[14] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 11, pp. 1254–1259, 1998.

[15] J. M. Wolfe and T. S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it?" *Nature Reviews Neuroscience*, vol. 5, no. 6, pp. 495–501, 2004.

[16] V. Mahadevan and N. Vasconcelos, "Saliency-based discriminant tracking," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, Conference Proceedings, pp. 1007–1013.

[17] S. Frintrop and M. Kessel, "Most salient region tracking," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, Conference Proceedings, pp. 1869–1874.

[18] G. Zhang, Z. Yuan, N. Zheng, X. Sheng, and T. Liu, *Visual saliency based object tracking*. Springer, 2010, pp. 193–203.

[19] U. Rajashekar, I. Van der Linde, A. C. Bovik, and L. K. Cormack, "Foveated analysis of image features at fixations," *Vision research*, vol. 47, no. 25, pp. 3160–3172, 2007.

[20] J. Ponce, D. Forsyth, E.-p. Willow, S. Antipolis-Mditerrane, R. d'activit RAweb, L. Inria, and I. Alumni, "Computer vision: a modern approach," *Computer*, vol. 16, p. 11, 2011.

[21] L. Duan, C. Wu, J. Miao, L. Qing, and Y. Fu, "Visual saliency detection by spatially weighted dissimilarity," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, Conference Proceedings, pp. 473–480.

[22] I. Jolliffe, *Principal Component Analysis*. Springer-Verlag, 1986.

[23] S. Sun, D. Haynor, and Y. Kim, "Motion estimation based on optical flow with adaptive gradients," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, vol. 1. IEEE, Conference Proceedings, pp. 852–855.

[24] J. Ribas-Corbera and D. L. Neuhoff, "Optimal block size for block-based motion-compensated video coders," in *Electronic Imaging'97*. International Society for Optics and Photonics, Conference Proceedings, pp. 1132–1143.

[25] A. M. Tekalp, *Digital video processing*. Prentice-Hall, Inc., 1995.

[26] A. Neri, S. Colonnese, G. Russo, and P. Talone, "Automatic moving object and background separation," *Signal Processing*, vol. 66, no. 2, pp. 219–232, 1998.

[27] A. Borji, S. Frintrop, D. N. Sihite, and L. Itti, "Adaptive object tracking by learning background context," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. IEEE, Conference Proceedings, pp. 23–30.

[28] R. Collins, X. Zhou, and S. K. Teh, "An open source tracking testbed and evaluation web site," in *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, Conference Proceedings, pp. 17–24.

[29] D. A. Klein, D. Schulz, S. Frintrop, and A. B. Cremers, "Adaptive real-time video-tracking for arbitrary objects," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, Conference Proceedings, pp. 772–777.

[30] V. Belagiannis, F. Schubert, N. Navab, and S. Ilic, *Segmentation based particle filtering for real-time 2d object tracking*. Springer, 2012, pp. 842–855.

[31] W. Yi, L. Jongwoo, and Y. Ming-Hsuan, "Online object tracking: A benchmark," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, Conference Proceedings, pp. 2411–2418.