Orderlessand Blurred Visual Tracking via Spatio-Temporal Context

Manna Dai[1], Peijie Lin[1], Lijun Wu[1], Zhicong Chen[1],Songlin Lai[1], Jie Zhang[1],Shuying Cheng[1*], Xiangjian He[2]

[1]Institute of Micro/Nano Devices and Solar Cells,College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China
[2]Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia
[*]Corresponding author email: sycheng@fzu.edu.cn

Abstract. In this paper,a novel and robust method which exploits the spatio-temporal context for orderless and blurred visual tracking is presented.This lets the tracker adapt to both rigid and deformable objects on-line even if the image is blurred.We observe that a RGB vectorof animage which is resizedinto a small fixed size can keep enough useful information.Based on this observation and computational reasons,we propose to resize the windows ofboth template and candidate target images into 2×2 and use Euclidean Distance to compute the similarity between these two RGB imagevectors for the preliminary screening.We then apply spatio-temporal context based on Bayesian framework to further compute a confidence map for obtaining the best target location.Experimental results on challenging video sequences in MATLAB without code optimization show the proposed tracking method outperforms eightstate-of-the-art methods.

Keywords:spatio-temporal-context, resize, Euclidean Distance, Bayesian framework

1 Introduction

Visual object tracking is to estimate location of a target in an image sequence. It has been a long standing research topic due to its wide range of applications such as video surveillance, human computer interaction, traffic control and so on[1]. However, visual tracking is challenging due to abrupt motion, illumination change, cluttered background and occlusion. Although a significant progress has been made to overcome these challenges, developing an efficient and robust visual tracking method is

still a crucial topic,particularly when a rigid or a deformable object moving disorderlyoccurs in a blurred image sequence.

The existing visual trackingapproaches can be categorized into generative [2-7]and discriminative[8-13] methods. The generative tracking methodssearch for image regions that are most similar to the template, whilediscriminative methodsaim at differentiating the target from the background.However, their main shortcomings are also remarkable as follows. Firstly, too many samples to be extracted make the computational load very heavy.Secondly, the effective searching algorithm and measured approach between template and candidate samples are difficult. Thirdly, it is hard to distinguish the target from complicated background because of the broadly varying background, the similarity between object and background, or the object which moves too fast to make the object itself and its surrounding blurred in an image.

In this paper, we propose a novel and robust tracking algorithmto exploit spatio-temporal local context information.Firstly, we use a simple and powerful workto search for object applying objectness scores. Our work is motivated by the fact that generic objects with well-defined closed boundaries [15-17] share strong correlation after resizing of their corresponding image to small fixed size. Therefore, the template and candidate model can be resized separately into 2×2 to efficiently quantify the objectness of an image.Euclidean Distanceis used to compute the similarity between the template and candidate model. We choose the maximum similarity as the best result to compute the promising center of the object.Thenspatio-temporal local context information is exploited to further determine the position of the object by using the previous promising center. We apply the max similarity of the template and candidate model to update the template in the next frame.The template-update in spatio-temporal context will also consider the several max confidence maps in the previous certain frames. These two update measures canbring in the current target information when the true template changes much or occlusion is occurred,andkeep the original information if occlusion is removed or previous tracking results are not really exact.

The main contributions of this paper are as follows.(1)Anovel and robust spatio-temporal context based orderlessand blurred visual tracking method is proposed. (2) An efficient search algorithm is adopted in each tracking round by resizing an image intosize 2×2 and usingEuclidean Distance to compute the similarity between template and candidate image for efficiently reducing the compute load. (3) Our method makes advantage of a strong spatio-temporal relationship between the local scenes containing the object in consecutive frames.(4) The experiments show that our method is robust to appearance variations introduced by abrupt motion, occlusion, pose variations, background clutter, and illumination variation, especially in blurred and disordered scene.

2 The STC Tracker

Our approachis based on the STC tracker presented inreference [14]. The STC tracker formulates the spatio-temporal relationships between the object of interest and its local context based on a Bayesian framework. It models the statistical correlation

between the low-level features in the target and its surrounding. Here, we provide a brief overview of this approach [14].

In STC tracker, a tracking problem is formulated by computing a confidence map that estimates the object location likelihood. In the current frame, we get the object location x^* and define the feature set as $X^c = \{c(z) = (I(z), z)|z \in \Omega_c(x^*))\}$ where $I(z)$ represents image intensity at location z and $\Omega_c(x^*)$ is the neighborhood of locationx^*. We can compute the object location likelihood by

$$\mathrm{m(x)} = \mathrm{P(x|o)} = \sum_{c(z)\in X^c} P(x, c(z)|o) = \sum_{c(z)\in X^c} P(x|c(z), o)\, P(c(z)|o). \quad (1)$$

The spatial context model is a conditional probability function, which is defined as

$$\mathrm{P(x|c(z), o)} = h^{sc}(x - z). \quad (2)$$

$h^{sc}(x - z)$is a function regarding the relative distance and direction between object location x and its local context location z, and it encodes the spatial context relationship of the target and its spatial relation. However, $h^{sc}(x - z)$ is not a radially symmetricfunction.

We model the context prior probability in (2) as

$$\mathrm{P(c(z)|o)} = \mathrm{I(z)}w_\sigma(z - x^*), \quad (3)$$

where$\mathrm{I}(\cdot)$ is the image intensity that represents appearance of the context.

Inspired by the biological visual system, a focus of attention function is used as a weighted function defined by

$$w_\sigma(z - x^*) = ae^{-\frac{|z-x^*|^2}{\sigma^2}}, \quad (4)$$

where$a$ is a normalization constant whichranges from 0 to 1 to satisfy the definition of probability and σ is a scale parameter. In this weighted function,the closer the location is to the object center, the more context locations are considered.

The confidence map of an object location is in formula (1) defined as

$$\mathrm{m(x)} = \mathrm{b}e^{-\left|\frac{x-x^*}{a}\right|^\beta} = \sum_{z\in\Omega_c(x^*)} h^{sc}(x - z)I(z)w_\sigma(z - x^*) = h^{sc}(x)\otimes\big(I(x)w_\sigma(x - x^*)\big). \quad (5)$$

Here,bis a normalization constant, α and β are a scale parameter and a shape parameterrespectively. Eq. (5) can be transformed to the frequency domain for fast convolution:

$$\mathrm{F}\left(\mathrm{b}e^{-\left|\frac{x-x^*}{\alpha}\right|^\beta}\right) = \mathrm{F}(h^{sc}(\mathrm{x}))\odot\mathrm{F}(\mathrm{I(x)}w_\sigma(\mathrm{x} - x^*)). \quad (6)$$

Here,F denotes the Fast Fourier Transform (FFT) function and $\odot$ denotes the element-wise product. The $h^{sc}(x)$ is defined as

$$h^{sc}(x) = F^{-1}\left(\frac{F(be^{-\left|\frac{x-x^*}{\alpha}\right|^{\beta}})}{F(I(x)w_{\sigma}(x-x^*))}\right). \quad (7)$$

Here,F^{-1} denotes the inverse FFT function. For more details, we refer to [14].

3 Orderlessand Blurred Visual Tracking via Spatio-Temporal Context

3.1 Framework

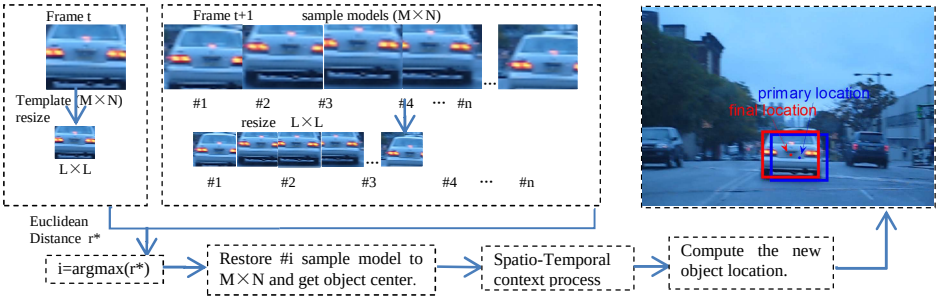


Fig.1. Basic flow of our tracking algorithm is as shown in figure. We resize template and sample models into L×L (e.g. 2×2). Euclidean Distance is used to simply compute the similarity between template and each sample models to get the primary location. We use the best image center(center of #i image) which has the max similarity to conduct the spatio-temporal context process for obtaining the final object location.

Fig.1 shows the basic flow of our proposed tracking algorithm. The tracking process has three steps.

Step I. We resize the candidate models setX $= \{x_1, x_2, \dots, x_n\}$ and template to L× L.In the image sequences which we used, theminimum of the lengthand the width of the objects is51. Therefore,the value of L is from 2 to 51 with the interval as 7. We use the different L to compute the average CLE and average FPS of our research datasets. We find that the average CLE does not highly increase and the average FPS reduces highly as the value of L increases.So we set L as 2 in order to get a good balance between CLE and FPS.

Step II. The suitable image (#i) center which is most similar to the template is selected. The similarity set$r^* = \{r_1^*, r_2^*, \dots, r_n^*\}$ can be obtained by Euclidean Distance algorithm. Then, we compute the #i image center as a prior center in current frame.

Step III. We use the prior center to conduct the spatio-temporal context process to get the result image location in this frame. After separately updating the template which is used to compute by Euclidean Distance algorithm in the first step and the one which is to conduct the spatio-temporal context process in the last step, we continue to sample a new candidate models set to circulate from the first step until the end of the image sequence. The details in each step will be introduced in the following content.

3.2 Image Resizing Measure

Usually,we depend on the previous object locationstopredict the promising location in next frame. However,this measure will fail if the object moves fast and disorderly.Some researchers use a sliding window fashion to search for the object [18-19] which will increase the computation burden.Therefore, we propose a local search algorithm with image resizing to help us search for the promising object location.

Objects are stand-alone things with well-defined closed boundaries and centers [15-17]. If we resize the image to a small fixed size, the little variation that closed boundaries could present in such abstracted view. In this paper, in order to process larger scale and reduce the computation burden at the same time, we decide to resize template and candidate models into 2×2 size (seen in Fig.1.).

3.3 Similarity By Euclidean Distance

We aim at narrowing the scope of our search for reducing the complexity in the following work. We choose the Euclidean Distance to calculate the max similarityφ of resizing template and resizing candidate models in order to provide us a promising object location and center. The Euclidean Distance is defined as follow.

$$\mathrm{dist}(\mathrm{X},\mathrm{Y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{8}$$

Here, X denotes the RGB vector of template, and Y represents the RGB vector of each candidate model. The Euclidean Distance not only reflects whether two vectors are relevant, but also can calculate the level of similarity between them.

We sample the candidate modelsset Xas

$$\mathrm{X} = \{x: \|x - (x_t^* + d)\| < d\}, \tag{9}$$

where x is the location of the sample model, x_t^*denotes the location of the object in the *t*-th frame, and d represents the distance in which the center of the object moves between the (*t*-1)-th frame and (*t*-2)-th frame.Note that d can be positive and negative.

We choose RGB as the feature in this step because we will exploit the gray-scale map in the following step. We hope to introduce more information of an image into our work so thatit can have a better result.Because RGB can be affected by the illumination, we use zero mean treatment to eliminate the effects of illumination and shadow.

The template in the next frame will be updated by the following function

$$M_{t+1} = (1 - \varphi^*)M_t + \varphi^*\mathrm{s} \tag{10}$$

Here,$M_{t+1}$is the template in the next frame, M_t is the template in current frame, φ^* is the max similarity which has been normalized in this frame, and s is the model image whose similarity is φ^*in this frame. Note that all parameters areunder the condition of 2×2 image.

3.4 Further Compute with Spatio-Temporal Context

We apply the spatio-temporal context in reference[14] to the following process.After computing the similarity between template and candidate models, we get a primary location of the object. Then, we will use this promising center to calculate the final location. Note that, we use the image in the original size from now on, not the 2 ×2 size.

In this process, assume that we initialize the target location in the first frame by some object detection algorithms. We learn the spatial context model $h_t^{sc}(x)$ (7) in the *t*-th frame, which is used to update the spatio-temporal context model $H_{t+1}^{stc}(x)$ (13) and detect the object location and center in the (*t*+1)-th frame.In reference [14], the object location x_{t+1}^* in the (*t*+1)-th frame is calculated by maximizing the new confidence map:

$$x_{t+1}^* = arg \max_{x \in \Omega_c(x_t^*)} \mathrm{m}_{t+1}(x), \tag{11}$$

where$\Omega_c(x_t^*)$ is the local context region which is based on the tracked location x_t^* in the *t*-th frame, and we construct the corresponding context feature set $x_{t+1}^c = \{c(z) = (I_{t+1}(z), z) | z \in \Omega_c(x_t^*)\}$.

The $\mathrm{m}_{t+1}(x)$ in (11) in [14] is defined as

$$\mathrm{m}_{t+1}(\mathrm{x}) = F^{-1}(F(H_{t+1}^{stc}(x)) \odot F(I_{t+1}(x) w_{\sigma_t}(x - x_t^*))), \tag{12}$$

which is deduced from (6).

We update the spatio-temporal context model in [14] by

$$H_{t+1}^{stc} = (1 - \rho) H_t^{stc} + \rho h_t^{sc}, \tag{13}$$

Here,ρis considered as a learning parameter, and h_t^{sc} is the spatial context model which can be obtained from (7) in the *t*-th frame.

Note that we will use zero mean treatment to every frame in order to remove the effect from the illumination change. In addition, the intensity in the context region exploitsa Hamming window to reduce the frequency influence from the image boundary on the FFT [20-21]. Hamming window is defined as

$$w(\mathrm{t}) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi}{\tau} t\right), & |t| \le \frac{\tau}{2} \\ 0, & |t| > \frac{\tau}{2} \end{cases} \tag{14}$$

The tracking procedure is summarized in Algorithm 1.

Algorithm 1. The proposed tracking method

Input: Video frame f=1:F

1. For f=1:F
2. If f==1
3. Select the tracking object.

4. Compute the h^{sc} ,then construct the template in spatio-temporal context as$H_1^{stc} = h^{sc}$.
5. Obtain the locationx_1^* of the tracking object.
6. Resize the image of tracking object in the location x_1^* to 2×2 size, namelys.
7. The template in RGB similarity is initialized as $M_1 =$ s.
8. Else
9. Calculate the distance d between the (t-1)-th frame and (t-2)-th frame.
10. Sample the candidate models setX $= \{x_1, x_2, \dots, x_n\}$by X $= \{x: \|x - (x_t^* + d)\| < d\}$.
11. Resize the template M_t and candidate models set X^* to 2×2 and calculate the similarity r^*,as the max similarity is defined as i $=$ argmax r^*, i $= 1,2, \dots,$ n.
12. Normalize the max similarity as φ^*.
13. Restore the i-th sample model to original size and compute its center C_t.
14. Update the template M_t.
15. Use the center C_t to conduct the spatio-temporal context process to get h_t^{sc}.
16. Update the templateH_{t+1}^{stc}.
17. The object location $x_{t+1}^*$is defined as $x_{t+1}^* = arg \max_{x \in \Omega_c(x_t^*)} c_{t+1}(x)$.
18. End if
19. End for

Output: Tracking results $\{x_1^*, x_2^*, \dots, x_F^*\}$.

4 Experimental Results and Analysis

4.1 Experimental Setup

In order to make our result more reasonable, we compare our method with other methods in the same experiment environment and equipment. Our approach is implemented in Matlab. The experiments are performed on an Intel(R) Core(TM) i5-2410M 2.30 GHz CPU with 2 GB RAM.In our experiments, the parameters are used in our algorithm as follows: the parameters of the map function are set to $\alpha = 1.8$and $\beta = 1$. The learning parameter $\rho = 0.075$. Here β and ρ are set as same as inreference [14]. But in [14], $\alpha = 2.25$.In Eq.(5),the greater α is, the larger region around the center of object will be considered. On the other words, as our work focuses on the scene where image is blurred and object moves disorderly, we will pay more attention to the region around the center to avoid the blurred false image being introduced into our procedure as a noise.Therefore,we choose a smaller α.

Datasets: We use 8 color sequences namely: body, car2, car4, face, dollar, deer, shaking, david. The sequences used in our experiments pose challenging situations such as motion blur, abrupt movement, illumination changes, scale variation, occlusions, rotation, background clutter, and pose variation. Especially, we can tackle the tracking problem in image blur and fast disorder motion.

4.2 Comparison with State-of-the-art

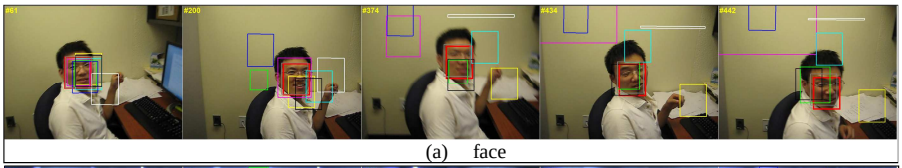
We compare our method with 8 different state-of-the-art trackers shown to provide excellent results in literature. The trackers used for comparison are:STC[14], WMIL[22], MIL[23], CT[8], L1[24], L1-APG[25], LOT[7] and Color Tracking[26].

In this paper, we follow the protocol used in [27] to validate our work. We will use three evaluation metrics: center location error (CLE), distance precision (DP) and overlap precision (OP). CLE is valued by the average Euclidean Distance between the estimated center location of the object and the ground-truth. DP is the relative number of frames in the sequence where CLE is smaller than a certain threshold. Here, the threshold is set as 20 pixels. OP is defined as the percentage of frames where the bounding box overlap exceeds a threshold $t \in [0,1]$. The trackers are ranked using DP scores at 20 pixels. We also present the speed of the trackers in average frames per second (FPS).

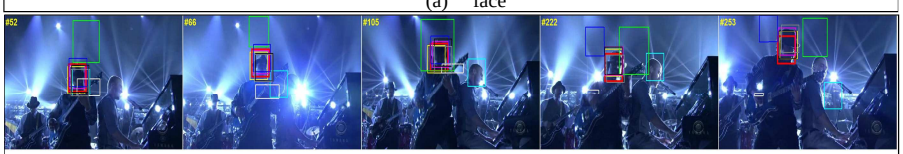
Table1 shows CLE, where smaller CLE means more accurate tracking results. From Table1, we can know that the quantitative results in which our tracking algorithm achieves the better performance. Fig.2 shows part of tracking results by different tracking methods. Table2 shows a comparison with the mentioned state-of-the-art methods on 8 challenging sequences. We also present the speed in average frames per second (FPS). The best three results are shown in red, blue and green fonts respectively. Our method carries outwell both in terms of speed and accuracy.

Table 1. Center location error (CLE) (in pixels). Red fonts indicate the best performance while blue fonts indicate the second best ones,and green fonts indicate the third best ones.

sequence	WMIL	STC	MIL	LOT	L1	L1-APG	CT	Colortracking	ours
face	127	113	123	33.4	149	183	55.8	7.5	3.9
shaking	12	8.2	145	73.6	29.1	104	11.2	13.2	10.7
dollar	12.1	20.4	70.5	71.6	20.5	71.5	9.3	17.1	7.1
deer	15.6	401	202	63.7	78.1	78.9	211	5.1	5.4
car4	95.6	2610	140	31.3	61.6	15.3	115	8.2	15.6
car2	163	5.41	73.9	26.2	49.9	156	104	86.9	5.1
body	54.4	148	128	84.5	131	31.6	122	36.5	18.1
david	34.3	43.4	38.1	108	76.1	39.9	40.2	24.1	10.9
Average CLE	64.3	418.7	115.1	61.5	74.4	84.9	83.6	24.8	9.6



(a) face



(b) shaking

Fig.2. Comparison of our approach with state-of-the-art trackers in challenging situations such as motion blur, abrupt movement, illumination changes, scale variation, partial occlusions, rotation, background clutter, and pose variation. Especially we can tackle the tracking problem in image blur and fast disorder motion.

Table 2. Quantitative comparison of our trackers with 8 state-of-the-art methods on 8 challenging sequences. The results are presented in average center location error (CLE) (in pixels), average distance precision (DP) (%) and average overlap precision (OP) (%). We also provide the average frames per second (FPS). The best three results are shown in red, blue and green fonts. Note that our method is best in average CLE, average DP and average OP,and the second best in terms of speed.

	WMIL	STC	MIL	LOT	L1	L1-APG	CT	Colortracking	ours
Average CLE	64.3	418.7	115.1	61.5	74.4	84.9	83.6	24.8	9.6
Average DP	40.6	42.5	12.7	26.8	25.5	21.7	29.8	72.9	88.3

Average OP	44.2	32.1	15.8	30.3	36.8	27.8	33.3	80.2	93.0
Average FPS	12.8	14.6	1.1	0.3	0.2	7.8	12.1	21.1	14.1

Motion blur. Figs.2 (a), (e),(f) and (g) have the motion blur. Only our method can deal with all the four sequences. In (a) and (e), our approach and Colortracking perform well in terms of CLE. L1-APG also has a good performance in (e), but it is not better than ours and Colortracking at frame #122 and #198. In (f) and (g), only our method can achieve successful tracking from beginning to end.

Abrupt movement. Figs.2 (a), (d), (e) and (f) suffer from abrupt movement in whole sequences. In (a), (d) and (e), our approach and Colortracking can succeed to track the object. However,in (f), only our method still track the object successfully from frame #236 as we show.

Illumination changes. In Figs.2 (b) and (h), the illumination often changes strong, some of the trackers completely fail to track. In (b) at frame #66, when the illumination changes fast and strong, only STC, Colortracking and our method still perform well as these three method all use zero mean measure to tackle the effect of illumination. In (h) at frame #83, #231,and #374, zero mean measureplays an important role when the light changes from dark to bright.

Scale variation. In (h) at frame #231 to #496, we want to show the ability of our proposed method in disposing the scale variation. Our approach, STC and Colortrackingcan adapt to the scale variation of the object. Moreover, our approach has the best CLE among all 9 approaches. LOT, L1 and L1-APG completely fail to track the object while WMIL, MIL and CT suffer from sever drift.

Partial occlusions. The object in Fig.2(b) shaking demonstrates that the proposed method performs well in terms of position and rotation when the target undergoes partial occlusion. Our method and STC perform better than other methods at frame #105and #253, while other methods suffer from sever drift and some of them fail totrack. Thus, our method can handle occlusion and it is not sensitive to partial occlusion.

Rotation. Sequences of (g) and (h) emerge rotation of the object. In (g) from #140 to #310, our algorithm is the only one which can dispose the rotation of body of the walking man. All8 other state-of-the-art algorithmsfail. The man turns the bodyin (h) from #496 to #579, the trackers of STC, Colortracking and ours can deal with the rotation in this scene. In conclusion, our algorithm is the last winner in the tracking problem in the way of rotation.

Pose variation. Fig.2 (c) at frame #70, the part of dollar is folded so that MIL,L1 and CT present their sensitivity in a certain aspect of pose variation. At frame #130, a pile of dollar is being divided into two piles, at the same time WMIL, STC and our method still exactly distinguish the right location of the object. MIL, LOT and L1-APG have total failure when the two piles of dollar are completely separated at frame #204. As showed at #290, if the two piles come close to each other, it will also have some impact on the trackers. Therefore, only our method and Colortracking can continue to develop their ability. To sum up, only our proposed method can keep accurate performance in the whole sequence.

Background clutter. The trackers are easily confused if the object is very similar to the background.Figs.2 (b), (d) and (e) have the background clutter. Only our method can deal with this tracking problem. Other trackers drift or else completely fail to track.

4.3 Discussion

As shown in our experiments, our method can address the factors such as motion blur, abrupt movement, illumination changes, scale variation, heavy occlusions, rotation, background clutter, and pose variation. Especially we can tackle the tracking problem in image blur and fast disorder motion. The reasons are as follows. (1) Our method exploits temporal and spatial context information for tracking, which is very insensitive to multiple factors. (2) A simple but useful preliminary screening by Euclidean Distance is introduced between object template and candidate samples. (3) The measure of resizing the object template and candidate samples to a small size like 2×2 can keep enough information we need. In addition, it can abandon the redundant information to reduce the amount of calculation rapidly and can be realized easily.

5 Concluding Remarks

In this paper, a novel and robust method named orderless and blurred visual tracking is proposed. Firstly,template and candidate image are resized to small size to reduce the computation load. Then,Euclidean Distance is used to compute the similarity between these two RGBvectors from the resized template and candidate image for the preliminary screening. Finally,in order to address the shortcomings of current approaches for blurred images and orderless motion, we adopt the spatio-temporal contextbased on Bayesian framework to compute a confidence map for obtaining the best target location. Therefore, our method is very insensitive to appearance change. Experiments on some challenging video sequences have demonstrated the superiority of the proposed approach to 8 existing state-of-the-art ones in terms of accuracy and robustness.

6 References

1. A.Yilmaz, O. Javed, and M. Shah. : Object tracking: A survey. ACM Computing Surveys (CSUR), vol. 38, no. 4(2006).
2. J. Kwon and K. M. Lee.:Visual tracking decomposition. In: CVPR, pp. 1269-1276 (2010).
3. J. Kwon and K. M. Lee. : Tracking by sampling trackers. In: ICCV, pp. 1195-1202, 2011.
4. H. Li, C. Shen, and Q. Shi,“Real-time visual tracking using compressive sensing,” in CVPR, pp. 1305-1312 (2011).
5. R. T. Collins. : Mean-shift blob tracking through scale space. In: CVPR, vol. 2, pp. Ⅱ-234 (2003).

6. H. Li, C. Shen, and Q. Shi. : Real-time visual tracking using compressive sensing. In: CVPR, pp. 1305-1312 (2011).
7. S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan. : Locally orderless tracking. In: CVPR, pp. 1940-1947 (2012).
8. K. Zhang, L. Zhang, and M.-H. Yang. :Real-time compressive tracking. In: ECCV, pp. 864-877 (2012).
9. S. Hare, A. Saffari, and P. H. Torr. :Struck: Structured output tracking with kernels.In: ICCV, pp. 263-270 (2011).
10. K. Zhang and H. Song. : Real-time visual tracking via online weighted multiple instance learning.Pattern Recognition (2012).
11. K. Fu, C. Gong, Y.Qiao, J. Yang, and I. Y.-H. Gu. :One-class support vector machine-assisted robust tracking. Journal of Electronic Imaging, vol. 22, no.2, pp. 023002-023002 (2013).
12. T. Zhang, B. Ghanem, S. Liu, and N. Ahuja. : Robust visual tracking via multi-task sparse learning. In: CVPR, pp. 2042-2049 (2012).
13. J. Henriques, R, Caseiro, P. Martins, and J. Batista. : Exploiting the circulant structure of tracking-by-detection with kernels.In: ECCV, pp. 702-715(2012).
14. K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M.-H. Yang, :Fast visual tracking via dense spatio-temporal context learning. In: ECCV (2014).
15. B. Alexe, T. Deselaers, and V. Ferrari. : Measuring the objectness of image windows. IEEE TPAMI, 34(11) (2012).
16. G. Heitz and D. Koller. : Learning spatial context: Using stuff to find things. In: ECCV, pp. 30-43(2008).
17. M.-M. Cheng, Z. Zhang, W.-Y. Lin, P. Torr. : Bing: Binarized Normed Gradients for objectness estimation at 300fps. In: CVPR (2004).
18. N. Dalal and B. Triggs. :Histograms of oriented gradients for human detection. In: CVPR,volume 1, pp. 886-893 (2005).
19. P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. :Object detection with discriminatively trained partbased models. IEEE TPAMI, 32(9) : 1672-1645 (2010).
20. A. V. Oppenheim, A. S. Willsky, and S. H. Nawab. :Signals and systems, vol. 2. Prentice-Hall Englewood Cliffs, NJ (1983).
21. D. S. Bolme, B. A. Draper, and J. R. Beveridge. : Average of synthetic exact filters. In: CVPR, pp. 2105-2112 (2009).
22. K. Zhang and H. Song. : Real-time visual tracking via online weighted multiple instance learning. Pattern Recognition (2012).
23. D. A Ross, J. Lim, R.-S. Lin, and M.-H. Yang. : Incremental learning for robust visual tracking. International Journal of Computer Vision, vol. 77, no. 1-3, pp. 125-141 (2008).
24. X. Mei and H. Ling. : Robust visual tracking using 1 minimization. In: Computer Vision, 2009 IEEE 12th International Conference on. IEEE, pp. 1436-1443 (2009).
25. C. Bao, Y. Wu, H. Ling and H. Ji. : Real time robust L1 tracker using accelerated proximal gradient approach. In: CVPR (2012).
26. M. Danelljan, F. S. Khan, M. Felsberg, and J. van de Weijer. : Adaptive color attributes for real-time visual tracking. In: CVPR, pp. 1090-1097 (2014).
27. Y. Wu, J. Lim, and M.-H. Yang. : Online object tracking: A benchmark. In: CVPR (2013).