ROBUST VISUAL TRACKING USING JOINT SCALE-SPATIAL CORRELATION FILTERS

Mengdan Zhang, Junliang Xing, Jin Gao, Weiming Hu

National Laboratory of Pattern Recognition, Institute of Automation, Beijing 100190, P. R. China
{mengdan.zhang, jlxing, jin.gao, wmhu}@nlpr.ia.ac.cn

ABSTRACT

Scale adaptation is crucial to object tracking as the visual size of the target changes continuously. Many existing tracking algorithms, however, simply ignore scale changes either for the consideration of tracking efficiency or the lack of principle ways to scale estimation. In this work, we present an efficient and effective scale adaptive tracking algorithm by proposing a correlation filter based tracker in the joint spatial and scale space. We find that the exhaustive template searching in this joint space can be well modeled by a block-circulant matrix. With the properties of the block-circulant matrices, we prove that the expensive template matching can be transformed to efficient dot product in frequency domain by fast Fourier Transform. Based on these findings, our new tracker significantly improves the robustness and adaptability of previous competitive spatial correlation trackers. On the latest single object tracking benchmark, our tracker advances the state-of-the-art tracking results with a very large margin.

Index Terms— Object tracking, scale adaptation, block-circulant matrix, Fast Fourier Transform (FFT)

1. INTRODUCTION

Visual object tracking is a difficult problem in computer vision and has been extensively studied in the last decades. In this problem, appearance model and tracking strategy are two crucial components, to which great efforts in previous work have been devoted. Early methods select object templates as the appearance model and perform tracking by matching object candidates exhaustively in the next frame with the stored templates [1, 2, 3]. Although very straightforward, this kind of tracking techniques is often very time-consuming and suffers from object deformations and occlusions.

To deal with these problems, discriminative learning based methods have become mainstream in this field [4, 5, 6, 7], which formulates object tracking as a classification problem and adopts online learning methods to learn a discriminative classifier between the object and backgrounds. The discriminative classifier increases the tracking robustness towards object deformations and occlusions. To avoid exhaustively evaluating all possible samples from the object and backgrounds, these methods usually select a small subset of samples to learn and update the classifier, as well as estimating the object state. The number of evaluated samples, therefore, has great impact on the tracking performance.

Recently, a new template matching based tracking technique [8, 9] has achieved excellent performances on two largest tracking benchmarks [10, 11] and re-attracts the attention in this field to template matching based tracking techniques. The key innovation of this technique is to approximate the spatial exhaustive searching by efficient dot product in the frequency domain by means of circulant matrices. Its main problem, however, is that scale variation can not be handled, which leads to lack of flexibility. This is due to its failure of formulating the circulant structure for multilevel template matching. On the other hand, if this algorithm is applied individually to several image layers as mentioned in [12], the comparison between layers can be ambiguous and tricky. To surmount this problem, we propose to perform the exhaustive template searching in the joint scale-spatial space, and find that this operation can be well modeled by a block-circulant matrix, which likewise, can transform the joint scale-spatial template matching operations to efficient multiplication operations in the Fourier domain. The final object scale and position are obtained simultaneously from a joint scale-spatial distribution by simply an inverse Discrete Fourier Transform. Based on these innovations, our new Joint Scale-Spatial Correlation (JSSC) tracker significantly outperforms previous Spatial Correlation trackers [13, 14] in the robustness and adaptability respects, and advances the best performance on the benchmark [10] with a very large margin.

Our contributions are as follows: (1) we find a new scale adaptation scheme which is compatible with the template matching based framework, significantly improving the performance of correlation filter based trackers; (2) we find that the dense template matching operations in the joint scale-spatial space imply the block-circulant structure, which helps transform the exhaustive matching to efficient multiplication in frequency domain by fast Fourier Transform; (3) the state-of-the-art results obtained according to recent benchmark [10] further prove our tracker's robustness and adaptability.

2. RELATED WORK

Scale variation is a very common problem in visual object tracking. Most previous tracking methods [4, 5], however, ignore the scale changes of the object during tracking. The

reason lies partly in the ambiguity in sample generation and labeling for neighboring image layers. Furthermore, with the incorporation of scale estimation, the workload of training and detection of the classifier increases dramatically. For generative methods, Danelljan *et al.*[15] proposed a two-stage algorithm, ranking first on the basis of benchmark [11]. They learn separate filters for position and scale estimations. Unfortunately, once the location ambiguity occurs when the target is undergoing large appearance variation, the unconvincing result of the position filter will adversely affect the performance of the scale filter. Thus, tracking error accumulates over time. In the framework of particle filtering, most tracking algorithms (e.g. VTD [2], SCM [16], and ASLA [17]) either neglect scale evaluation in the principal training process, or just treat samples from different scale levels equally. Over time, though, this can degrade the model and cause drift.

Compared to other scale adaptation methods, our algorithm has merits. Firstly, general in Bayesian framework, our motion model is based on the scale prior obtained from last frame. So the detection is executed on several scale layers. Besides, we modify the Bayesian framework by inserting the scale factor into the observation model. Consequently, the interaction between samples from different scale levels is carefully considered in the training process. In addition, the joint scale-spatial response is defined as a multivariate Gaussian distribution. This special structure of response weights training samples diversely and permits our algorithm to estimate the position and scale simultaneously. We use linear interpolation to ensure continuity of scale estimation.

3. PROPOSED ALGORITHM

To incorporate scale estimation into visual tracking, it is ideal to extract image patches continuously from the joint scale-spatial space. Fortunately, a block-circulant structure is dug out for this process and further works out the impact on the Kernelized Ridge Regression algorithm. It turns out that the kernel matrix of the joint space exhibits block-circulant structure. Considering the property of block-circulant matrices, we diagonalize the kernel matrix with the Discrete Fourier Transform (DFT) matrix and transform templates matching to efficient multiplication operations. We also prove the efficiency of our scale adaptation scheme under the Bayesian framework in a holistic view.

3.1. Dense Sampling and Circulant Matrices

The performance of sparse sampling based trackers is often limited by the number of samples. We thus intend to use all the samples in both training and detection process. Assume a $1D$ image and a single-channel feature. Then the base sample is defined as a particular image patch whose center is located at the estimated position of the target. When we sample continuously around the target, without considering the boundary effect, the translation of the search window can be approximately considered as the cyclic shift of the base sample. Thus, the universal set of samples for an image can be represented as an circulant matrix [18]:

$$C(\mathbf{c}) = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ c_n & c_1 & \cdots & c_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ c_2 & c_3 & \cdots & c_1 \end{bmatrix}, \tag{1}$$

where each row is a cyclic shift of the row above it, defined as $\mathrm{row}(i)=P^{i-1}\mathbf{c}$, P is a cyclic shift operator, base sample $\mathbf{c}$ is the first row. The structure can also be characterized by noting that the (k,j) entry of the circulant matrix is given by

$$C_{k,j} = c_{(j-k) \bmod n}. \tag{2}$$

A very good property of the Circulant matrices is that it can be diagonalized via the DFT matrix F:

$$C(\mathbf{c}) = F\,\mathrm{diag}(\hat{\mathbf{c}})F^H, \tag{3}$$

where $\hat{\mathbf{c}}$ denotes the DFT of the base sample, $\hat{\mathbf{c}}=F\mathbf{c}$. From now on, we use a hat as shorthand for the DFT of a vector. Since scale estimation is incorporated, we finally arrive at S circulant matrices defined as

$X=(X_1^T, X_2^T, \cdots, X_S^T)^T=(C(\mathbf{x}_1)^T, C(\mathbf{x}_2)^T, \cdots, C(\mathbf{x}_S)^T)^T$,

where S denotes the size of the scale space and $\mathbf{x}_i \in R^n (i=1,2,\ldots,S)$ represent base samples of different scale levels.

3.2. Ridge Regression with Kernel Trick

When we match the object template with the universal set of training samples from the joint scale-spatial space, assume the matching scores obey a multivariate Gaussian distribution. The goal of training is to get the template that minimizes the squared error over sample response and the defined matching scores. We use regularized Ridge Regression to achieve this and get the closed-form solution in the dual space for the Kernelized Ridge Regression [19]:

$$\alpha = (K + \lambda I_{Sn})^{-1} Y, \tag{4}$$

where the $(S\times n)\times 1$ vector Y is the universal set of matching scores. The $S\times S$ block matrix K is a collection of the kernel matrices generated between different scale layers. Concisely, a block $K_{ij} (i,j=1,2,\ldots,S)$ denotes the $n\times n$ kernel matrix calculated from the scale layers X_i and X_j . Additionally, the output of the kernel function for each pair of samples from the two scale layers can be given by:

$$K_{ij}(q,l) = \kappa(P^{q-1}\mathbf{x}_i, P^{l-1}\mathbf{x}_j), (q,l=1,2,\ldots,n). \tag{5}$$

3.3. Block-circulant Structure

It is hard to handle the inverse of the large non-sparse matrix in (4). However, if the block matrix K implies block-circulant structure, the matrix inversion can be significantly simplified.

Theorem 1 *Given circulant matrices X_i and X_j, the kernel matrix K_{ij} is circulant if the kernel function satisfies $\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j)=\kappa(M\boldsymbol{x}_i, M\boldsymbol{x}_j)$, for any permutation matrix M.*

We use the Gaussian kernel and obtain that each block of the matrix K is circulant [9]. Then, we select elements from the

same place of each block of K and store them in an $S \times S$ matrix. Finally, we will get an $n \times n$ block-circulant matrix $\bar{K}$. To further confirm the conclusion, we have

$$\bar{K}_{ql}=\{K_{ij}(q,l)\}_{i,j=1}^{S}=\{\kappa(\mathbf{x}_i, P^{(l-q) \bmod n}\mathbf{x}_j)\}_{i,j}. \quad (6)$$

Similarly, we refer to the first row of the block-circulant matrix as the base block sequence, denoted $[\Psi_1, \Psi_2, \cdots, \Psi_n]$. As in [20], the block-circulant matrix is diagonalized as:

$$\bar{K} = W \operatorname{diag}(g(u_0), g(u_1), \cdots, g(u_{n-1}))W^H, \quad (7)$$

$$g(x) = \Psi_1 + \Psi_2 x + \cdots + \Psi_n x^{n-1}, \quad (8)$$

$$W = F \otimes I_S, \quad (9)$$

$$u_k = \exp(-j\frac{2\pi k}{n}), \quad (10)$$

where $g(x)$ calculates the DFT of the base block sequence.

3.4. Model Training

Since we obtain a block-circulant matrix after the rearrangement of block matrix K, the JSSC solution in the Fourier domain is extended as

$$\hat{\bar{\alpha}}^*=(\operatorname{diag}(g(u_0), g(u_1), \cdots, g(u_{n-1})) + \lambda I_{Sn})^{-1}\hat{\bar{Y}}^*, \quad (11)$$

$$g(u_c) = \begin{bmatrix} \hat{k}_c^{\mathbf{x}_1\mathbf{x}_1} & \hat{k}_c^{\mathbf{x}_1\mathbf{x}_2} & \cdots & \hat{k}_c^{\mathbf{x}_1\mathbf{x}_S} \\ \hat{k}_c^{\mathbf{x}_2\mathbf{x}_1} & \hat{k}_c^{\mathbf{x}_2\mathbf{x}_2} & \cdots & \hat{k}_c^{\mathbf{x}_2\mathbf{x}_S} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{k}_c^{\mathbf{x}_S\mathbf{x}_1} & \hat{k}_c^{\mathbf{x}_S\mathbf{x}_2} & \cdots & \hat{k}_c^{\mathbf{x}_S\mathbf{x}_S} \end{bmatrix}, \quad (12)$$

$$k^{\mathbf{x}\tilde{\mathbf{x}}} = \exp(-\frac{1}{\sigma^2}(\|\mathbf{x}\|^2 + \|\tilde{\mathbf{x}}\|^2 - 2\mathcal{F}^{-1}(\hat{\mathbf{x}} \odot \hat{\tilde{\mathbf{x}}}^*))), \quad (13)$$

where $k_c^{\mathbf{x}_i\mathbf{x}_j}$ is the c-th element of the base sample of the Gaussian kernel matrix K_{ij}, the horizontal bars represent the rearrangement, $\mathcal{F}^{-1}$ denotes the Inverse DFT and $\odot$ denotes element-wise product.

3.5. Modeling Testing

Similar to the training section, we wish to match the object template with the universal set of candidates from the joint scale-spatial space.The template matching scores for all the candidate patches can be computed as:

$$f(Z) = K^{ZX}\alpha. \quad (14)$$

The asymmetric kernel matrices between all candidate patches and training patches are stored in the block matrix $K^{ZX}=(C(k^{\mathbf{z}_i\mathbf{x}_j}))_{i,j=1}^{S}$. Considering the block-circulant matrix properties, the full detection response is given by

$$\hat{\bar{f}}(Z)=\operatorname{diag}(h^*(u_0), h^*(u_1), \cdots, h^*(u_{n-1}))\hat{\bar{\alpha}}, \quad (15)$$

$$h(u_c) = \begin{bmatrix} \hat{k}_c^{\mathbf{z}_1\mathbf{x}_1} & \hat{k}_c^{\mathbf{z}_1\mathbf{x}_2} & \cdots & \hat{k}_c^{\mathbf{z}_1\mathbf{x}_S} \\ \hat{k}_c^{\mathbf{z}_2\mathbf{x}_1} & \hat{k}_c^{\mathbf{z}_2\mathbf{x}_2} & \cdots & \hat{k}_c^{\mathbf{z}_2\mathbf{x}_S} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{k}_c^{\mathbf{z}_S\mathbf{x}_1} & \hat{k}_c^{\mathbf{z}_S\mathbf{x}_2} & \cdots & \hat{k}_c^{\mathbf{z}_S\mathbf{x}_S} \end{bmatrix}, \quad (16)$$

Intuitively, the detection process can be considered as a spatial filtering over the kernel values $(k_c^{\mathbf{z}_i\mathbf{x}_j})_{c=1}^{n}$. Moreover, a filtering operation over the DFT of kernel values $(\hat{k}_c^{\mathbf{z}_i\mathbf{x}_j})_{j=1}^{S}$ deals with the correlation between different scale levels.

3.6. Scale Adaptation Scheme

Our scale adaptation scheme is based on the Bayesian framework with some modification of observation model. We insert the scale factor into observation by modeling it as a rectangular cuboid of the feature pyramid. If the size of the target patch is $M \times N$, the cuboid is then of size $M \times N \times S$ and is centered at the target's estimated location and scale. The Bayesian formulation is expressed as

$$p(X_t|Y_t^{1:S}) \propto p(Y_t^{1:S}|X_t) \int p(X_t|X_{t-1}) p(X_{t-1}|Y_{1:t-1}^{1:S}) dX_{t-1}, \quad (17)$$

where $Y_t^{1:S}$ is the observation at time t and X_t denotes corresponding state. Apparently, several scale levels are compared with each other in the training process, which contributes to the sensitiveness to the scale variation. The posteriori probability is expected to be a multivariate Gaussian distribution. It is used as scale-spatial sample weighting during model training. In further detection, we search on different scale levels and adopt linear interpolation according to the posteriori probability, which ensures continuity of the scale estimation.

4. EXPERIMENTS

4.1. Experimental Settings

Our tracker is evaluated by a recent benchmark [10] that includes 50 video sequences. It competes with KCF, SSITDT [21], SAMF [12] and other 29 trackers evaluated in the benchmark. We choose HOG descriptors [22]. Let J_t denote the scale coefficient of the last frame and S be the number of scale layers. We resize the current image with scale factors $J_t a^l (l \in \{\lfloor -\frac{S-1}{2} \rfloor, \ldots, \lfloor \frac{S-1}{2} \rfloor\})$. Here, $a = 1.02$ restricts the sampling granularity in the scale space. The scale variance of the multivariate Gaussian distribution is $\sigma_s^2 = \frac{\sigma_x^2}{0.0142}$. Rest parameters are similar to KCF. We use distance precision (D-P) at a threshold of 20 pixels and overlap precision (OP) at a threshold of 0.5 to evaluate trackers' performance. Besides a precision plot with DP scores, we also exhibit an extra success plot using the area under the curve (AUC) criteria.

4.2. Parameter Analyses

The scale layers S is the most important parameter. We evaluate its effects on the tracking performance when its value varies as $S = 3, 5, 7$, thus with corresponding $a = 1.0404, 1.02, 1.0133$. The scale variance increases with the number of scale layers. These three variants are denoted as JSSC_3, JSSC_5 and JSSC_7 respectively. Fig. 1 shows the tracking results of these three variants using different data subsets and evaluation metrics as defined in [10]. Obviously, the 3-layers design is enough for continuous scale estimation. However, 5-layers design is better for overall performance evaluation. Since the maximal scale ratio is 1.0404 for consecutive frames, the 7-layers design may be too meticulous for model training and scale change capture, thus degrades the performance. Based on these analysis, we therefore choose $S = 5$ in all the following experiments, while operating at real-time.

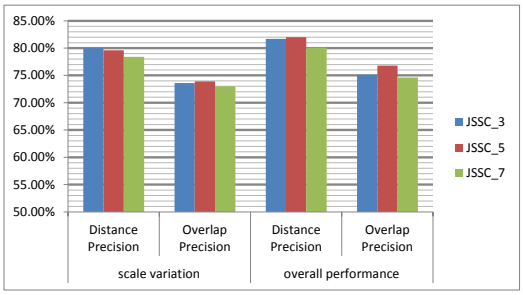

Fig. 1: A precision comparison for scale variation and overall performance evaluation among three variants of JSSC with different number of scale layers.

4.3. Scale Adaptation Evaluation

In this section, we focus on all the 28 sequences annotated with the scale variation attribute in [10]. We use KCF tracker as a baseline. Our method improves the baseline by 11.8% DP and 35.6% OP respectively. Table 1 provides a per-video OP comparison with the top 5 existing trackers and the baseline tracker. Obviously, the performance of KCF is much poorer in complex sequences. Since scale evaluation is involved in the training and detection process, our tracker can be highly adaptive. The scale levels are updated continuously based on recent scale estimation, so our tracker follows the target closely with higher OP scores. Moreover, our tracker performs better than other 5 trackers on 18 out of the 28 sequences. This shows that our scale adaptation scheme is compatible with the template matching based framework.

	SCM	SSITDT	ASLA	TLD	SAMF	KCF	**Proposed**
Car4	*0.97*	0.30	**1**	0.79	**1**	0.36	**1**
David	0.91	0.70	0.96	*0.97*	0.96	0.62	**1**
Trellis	0.85	0.71	0.86	0.47	**1**	0.84	*0.96*
Soccer	0.24	0.18	0.12	0.12	0.19	*0.39*	**0.48**
Matrix	0.30	0.01	0.02	0.07	*0.34*	0.13	**0.45**
Ironman	*0.13*	0.07	*0.13*	0.07	0.11	**0.15**	*0.13*
Skating1	0.42	0.30	*0.69*	0.23	0.49	0.36	**0.80**
Shaking	**0.90**	*0.52*	0.38	0.4	0.01	0.02	0.02
Singer1	**1**	0.80	**1**	*0.99*	0.58	0.28	**1**
Boy	0.44	*0.99*	0.44	0.94	**1**	*0.99*	**1**
Dudek	*0.98*	0.82	0.90	0.84	*0.98*	0.00	**1**
Crossing	**1**	**1**	**1**	0.52	**1**	0.98	*0.99*
Couple	0.11	*0.62*	0.08	**1**	0.48	0.24	0.54
Doll	**0.99**	0.84	*0.92*	0.62	0.67	0.55	0.75
Girl	0.88	0.79	0.91	0.76	*0.98*	0.72	**1**
Walking2	**1**	0.89	0.40	0.34	*0.96*	0.38	**1**
Walking	*0.96*	0.88	**1**	0.38	**1**	0.51	**1**
Fleetface	*0.70*	0.68	0.61	0.57	*0.70*	0.00	**0.75**
Freeman1	*0.81*	0.60	0.31	0.21	0.30	0.16	**0.84**
Freeman3	*0.93*	0.83	**0.94**	0.58	0.28	0.29	0.77
Freeman4	0.24	0.15	0.17	*0.27*	0.17	0.17	**0.44**
CarScale	0.65	0.58	*0.69*	0.44	0.62	0.44	**0.87**
Skiing	*0.09*	0.06	**0.11**	0.07	0.05	0.07	0.05
Dog1	0.85	0.43	*0.92*	0.67	0.70	0.65	**1**
Liquor	0.32	0.40	0.24	*0.58*	0.41	0.00	**0.99**
Lemming	0.17	0.32	0.17	0.59	*0.90*	0.45	**0.94**
MotorRolling	0.07	0.06	*0.10*	**0.17**	0.08	0.01	0.08
Woman	0.86	0.73	0.19	0.16	*0.92*	**0.93**	0.85
Average	*0.635*	0.545	0.544	0.494	0.603	0.383	**0.739**

Table 1: Per-video overlap precision (OP) on the 28 benchmark sequences for scale variation evaluation.

4.4. Overall Performance Evaluation

We evaluate the overall performance on all the 50 sequences. Our method provides a DP of 82.0% compared to 77.1% obtained by the best existing method SAMF. Meanwhile, an OP

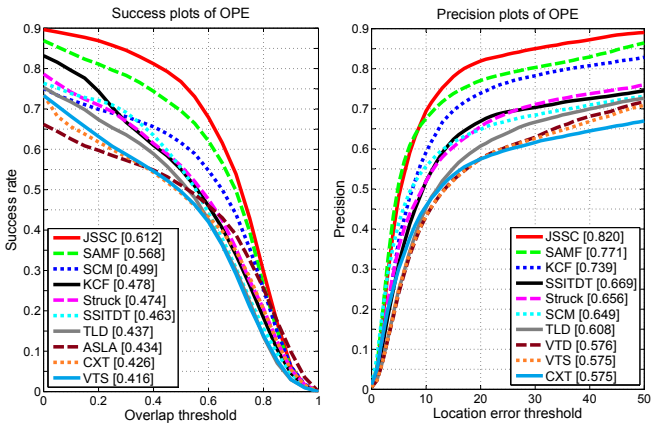

Fig. 2: Success plots and precision plots over all the 50 sequences.

gain of 7.3% is also achieved compared to 69.5% obtained by SAMF ranking the first previously. Fig. 2 exhibits the success plot using AUC criteria and the precision plot containing DP scores. It illustrates vividly that our tracker performs much better than other 32 trackers. According to the AUC scores of the success plots for 11 attributes, our tracker always comes first except the deformation attribute. We improve the best results by more than 5% for scale variation, in-plane rotation, illumination variation, background clutter, fast motion and motion blur. Since the template matching score is expected to obey a multivariate Gaussian distribution in the joint scale-spatial space, the relationship among dense samples from the joint space is learned carefully. Hence, our tracker handles background clutter and abrupt motion firmly. As the updating rate varies with the joint scale-spatial response, heavy occlusion does not affect the appearance model very much. Interestingly, compared to TLD [23], a significant gain of 17.5% success ratio is obtained when tracking out of view targets. Although there is no re-detection and failure recovery mechanism, our tracker is still highly adaptable to complex situations.

5. CONCLUSION

In this work, we dig up block-circulant structure to model the exhaustive template matching in the joint scale-spatial space, and prove that expensive templates matching can be transformed to efficient multiplication operations in the frequency domain. The proposed tracker based on these findings achieves high performance on a large tracking benchmark. In future, we plan to further explore the potential of our tracker to other tracking difficulties, e.g., object rotation changes.

6. ACKNOWLEDGEMENTS

This work is partly supported by the 973 basic research program of China (Grant No. 2014CB349303), the Natural Science Foundation of China (Grant No. 61303178 and 61472421), the National 863 High-Tech R&D Program of China (Grant No. 2012AA012504), and the Project Supported by Guangdong Natural Science Foundation (Grant No. S2012020011081).

7. REFERENCES

[1] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 25, no. 5, pp. 564–577, May 2003.

[2] J. Kwon and K. Lee, "Visual tracking decomposition," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogni.*, Jun. 2010, pp. 1269–1276.

[3] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogni.*, Jun. 2006, pp. 798–805.

[4] K. Zhang, L. Zhang, and M. Yang, "Real-time compressive tracking," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2012, pp. 864–877.

[5] B. Babenko, M. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogni.*, Jun. 2009, pp. 983–990.

[6] X. Li, C. Shen, A. Dick, and A. van den Hengel, "Learning compact binary codes for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogni.*, Jun. 2013, pp. 2419–2426.

[7] S. Avidan, "Support vector tracking," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 26, no. 8, pp. 1064–1072, Aug. 2004.

[8] J. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2012, pp. 702–715.

[9] J. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. PP, no. 99, pp. 1, Aug. 2014.

[10] Y. Wu, J. Lim, and M. Yang, "Online object tracking: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogni.*, Jun. 2013, pp. 2411–2418.

[11] M. Kristan and R. Pugfelder *et al*., "The visual object tracking VOT2014 challenge results," Sep. 2014.

[12] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Eur. Conf. Comput. Vis. Workshop*, Oct. 2014.

[13] D. Bolme, J. Beveridge, B. Draper, and Y. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogni.*, Jun. 2010, pp. 2544–2550.

[14] K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M. Yang, "Fast visual tracking via dense spatio-temporal context learning," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 127–141.

[15] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proc. British Mach. Vis. Conf.*, Jul. 2014.

[16] W. Zhong, H. Lu, and M. Yang, "Robust object tracking via sparsity-based collaborative model," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogni.*, Jun. 2012, pp. 1838–1845.

[17] X. Jia, H. Lu, and M. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogni.*, Jun. 2012, pp. 1822–1829.

[18] P. Davis, *Circulant matrices*, American Mathematical Society, New York, 1979.

[19] K. Murphy, *Machine learning: a probabilistic perspective*, MIT press, London, 2012.

[20] T. De Mazancourt and D. Gerlic, "The inverse of a block-circulant matrix," *IEEE Trans. Antennas Propag.*, vol. 31, no. 5, pp. 808–810, Sep. 1983.

[21] J. Gao, J. Xing, W. Hu, and S. Maybank, "Discriminant tracking using tensor representation with semi-supervised improvement," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1569–1576.

[22] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 32, no. 9, pp. 1627–1645, Jul. 2010.

[23] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 34, no. 7, pp. 1409–1422, Dec. 2012.