–Supplementary Material–
Fast Kernelized Correlation Filter without Boundary Effect

Ming TANG, Linyu ZHENG, Bin YU, and Jinqiao WANG
National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing 100190, China
University of Chinese Academy of Sciences, Beijing 100049, China

Abstract

This supplementary material contains four parts. 1) The formal steps of ACSII and CCIM (Sec.1). 2) The implementation details of nBEKCF-D, its comparison to trackers with ImageNet pre-trained features, and comments on C-COT, ECO, and UPDT (Sec.2). 3) An illustration to explain how Algorithm 2, CCIM, works exactly (Sec.4). 4) Mathematical proof of the correctness of CCIM (Sec.5).

1. Algorithms

The formal steps of ACSII and CCIM are presented in Algs.1 and 2, respectively.

Algorithm 1 Autocorrelation with Square Integral Image (ACSII)

- **Input**: $\mathbf{X} \in \mathbb{R}^{M\times N\times D}$ with $\mathbf{x}_{i,j} \in \mathbb{R}^D$'s as its elements, m, and n, where $m \leq M$ and $n \leq N$.
- **Output**: $\mathcal{X} \circ \mathcal{X} = \mathbf{B} \in \mathbb{R}^{(M-m+1)\times(N-n+1)}$ with $B_{p,q}$'s as its elements.
- **Construct squared image**

  Set $M \times N$ matrix $\mathbf{A}$ with $A_{i,j}$'s as its elements, $A_{i,j} = \|\mathbf{x}_{i,j}\|_2^2$.
- **Construct squared integral image I**
  1. $I_{0,0} = A_{0,0}$.
  2. for $p = 1$ to $M-1$: $I_{p,0} = I_{p-1,0} + A_{p,0}$.
  3. for $q = 1$ to $N-1$: $I_{0,q} = I_{0,q-1} + A_{0,q}$.
  4. for $p = 1$ to $M-1$, $q = 1$ to $N-1$: $I_{p,q} = I_{p-1,q} + I_{p,q-1} - I_{p-1,q-1} + A_{p,q}$.
- **Calculate autocorrelation** $\mathcal{X} \circ \mathcal{X}$
  1. $B_{0,0} = I_{m-1,n-1}$.
  2. for $p = 1$ to $M-m$: $B_{p,0} = I_{p+m-1,n-1} - I_{p-1,n-1}$.
  3. for $q = 1$ to $N-n$: $B_{0,q} = I_{m-1,q+n-1} - I_{m-1,q-1}$.
  4. for $p = 1$ to $M-m$, $q = 1$ to $N-n$: $B_{p,q} = I_{p+m-1,q+n-1} - I_{p-1,q+n-1} - I_{p+m-1,q-1} + I_{p-1,q-1}$.

2. Experiments of nBEKCF-D

The latest development of SRDCF is UPDT [2]. UPDT applies the same way as SRDCF's to relax the boundary effect, and achieves the state-of-the-art accuracy through the following series of improvements, *i.e.*, (1) the interpolation of feature resolution which was designed to improve SRDCF and resulted in C-COT [4]; (2) the reduction of feature dimensionality, linear weighting of features, clustering samples, and sparse update which were developed to improve C-COT and resulted in ECO [3]; (3) the employment of deeper networks, ResNet50, the augmentation of training samples, and the adaptive fusion of models with shallow and deep features which were designed to improve ECO. Therefore, UPDT inherits various defects of SRDCF on the boundary effect.

In our experiments, for a fair comparison, the same type of state-of-the-arts trackers are among the list of compared trackers. The same type of state-of-the-arts trackers means the top ones with the similar motivation, similar features, and similar scale adaptation scheme to the nBEKCF's. Therefore, UPDT [2] from the trackers with ImageNet pre-trained features, which, like nBEKCF, both adopt the scale-pyramid scheme to decide proper scales of target.

2.1. Implementation Details

In nBEKCF-D, as in the top CF tracker with ImageNet pre-trained features, UPDT, the Conv-1 and Block-4 feature maps of ResNet50 are exploited, and the bilinear interpolation is applied to increase the resolution of the Block-4 feature maps 4×4 times. Linear kernel is used to construct the kernel correlation matrices. The learning rate $\gamma = 0.004$. The size of learning or search region is $M = N = 4.5\sqrt{mn}$. ACSII and CCIM are implemented in C++ with CUDA, and

Algorithm 2 Cyclic Correlation with Integral Matrix (CCIM)

- **Input**: $\mathbf{Z} \in \mathbb{R}^{m\times n\times D}$ with $\mathbf{z}_{s,t} \in \mathbb{R}^{D}$'s as its elements, $\mathbf{X} \in \mathbb{R}^{M\times N\times D}$, $m \leqslant M$, and $n \leqslant N$.
- **Output**: $\mathcal{X} \diamond \mathcal{Z} = [\mathrm{vec}(\mathbf{Z}^{0,0}\text{☆}\mathbf{X}), \cdots, \mathrm{vec}(\mathbf{Z}^{m-1,n-1}\text{☆}\mathbf{X})]$.
- **Construct fundamental matrices**
  /* Cyclically shift fundamental calculations $\mathbf{z}_{s,t}\text{☆}\mathbf{X}$'s to construct fundamental matrices $\mathbf{B}^{s,t}$'s so as to $\sum_{s,t} \mathbf{B}^{s,t}_{((0,0),(M-m,N-n))} = \mathbf{Z}\text{☆}\mathbf{X}$.
  for $s = 0$ to $m-1$, $t = 0$ to $n-1$:
  $$\mathbf{B}^{s,t} = \mathbf{P}_M^{-s}\left(\mathbf{z}_{s,t}\text{☆}\mathbf{X}\right)\mathbf{Q}_N^{-t}.$$
- **Construct integral matrix** $\mathbf{M} \equiv [\mathbf{M}^{s,t}]_{m\times n}$ with $\mathbf{M}^{s,t} \in \mathbb{R}^{M\times N}$ being its block element.
  1. $\mathbf{M}^{0,0} = \mathbf{B}^{0,0}$,
  2. for $s = 1$ to $m-1$: $\mathbf{M}^{s,0} = \mathbf{M}^{s-1,0} + \mathbf{B}^{s,0}$.
  3. for $t = 1$ to $n-1$: $\mathbf{M}^{0,t} = \mathbf{M}^{0,t-1} + \mathbf{B}^{0,t}$.
  4. for $s = 1$ to $m-1$, $t = 1$ to $n-1$:
     $\mathbf{M}^{s,t} = \mathbf{M}^{s-1,t} + \mathbf{M}^{s,t-1} - \mathbf{M}^{s-1,t-1} + \mathbf{B}^{s,t}$.
- **Calculate correlation** $\mathbf{Z}^{s,t}\text{☆}\mathbf{X}$.
  for $s = 0$ to $m-1$, $t = 0$ to $n-1$:
  /* Divide $\mathbf{Z}^{s,t}\text{☆}\mathbf{X}$ into four parts and calculate each of them one by one by means of $\mathbf{M}$.
  1. $\mathbf{L} = \mathbf{M}^{m-1,n-1} - \mathbf{M}^{m-1,n-t-1} - \mathbf{M}^{m-s-1,n-1} + \mathbf{M}^{m-s-1,n-t-1}$.
     If $s = 0$ or $t = 0$, $\mathbf{L} = \mathbf{0}_{M\times N}$.
  2. $\mathbf{G} = \mathbf{M}^{m-1,n-t-1} - \mathbf{M}^{m-s-1,n-t-1}$.
     If $s = 0$, $\mathbf{G} = \mathbf{0}_{M\times N}$.
  3. $\mathbf{K} = \mathbf{M}^{m-s-1,n-1} - \mathbf{M}^{m-s-1,n-t-1}$.
     If $t = 0$, $\mathbf{K} = \mathbf{0}_{M\times N}$.
  4. $\mathbf{J} = \mathbf{M}^{m-s-1,n-t-1}$.
  /* Align the results of four sub-correlations:
  5. $\mathbf{L} \leftarrow \mathbf{P}_M^{m-s}\mathbf{L}\mathbf{Q}_N^{n-t}$,
  6. $\mathbf{G} \leftarrow \mathbf{P}_M^{m-s}\mathbf{G}\mathbf{Q}_N^{-t}$,
  7. $\mathbf{K} \leftarrow \mathbf{P}_M^{-s}\mathbf{K}\mathbf{Q}_N^{n-t}$,
  8. $\mathbf{J} \leftarrow \mathbf{P}_M^{-s}\mathbf{J}\mathbf{Q}_N^{-t}$.
  /* Sum the aligned four sub-correlations:
  9. $\mathbf{Z}^{s,t}\text{☆}\mathbf{X} = (\mathbf{L}+\mathbf{G}+\mathbf{K}+\mathbf{J})_{((0,0),(M-m,N-n))}$.

other parts in Pytorch. The experiments are performed on a single NVIDIA TITAN X GPU.

Regularization parameter $\lambda = 0.01$. Gaussian response $\mathbf{y}$ is identical to that in KCF with variance 0.01.

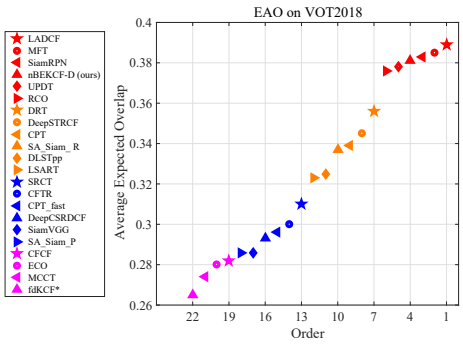


Figure 1: Comparison among nBEKCF-D and state-of-the-art trackers on VOT2018 challenge in terms of EAO.

Tracker	nBEKCF-D	HCF	ECO	CCOT	MDNet	SiamFC	CFNet	GOTURN
$SR_{0.50}$ (%)	44.8	29.7	30.9	32.8	30.3	35.3	40.4	37.5
$SR_{0.75}$ (%)	12.5	8.8	11.1	10.7	9.9	9.8	14.4	12.4
AO (%)	42.0	31.5	31.6	32.5	29.9	34.8	37.4	34.7

Table 1: Comparison among nBEKCF-D and state-of-the-art trackers on the GOT10k test set in terms of average overlap (AO) and success rates (SR) at overlap thresholds 0.5 and 0.75. The best three results are shown in red, green and blue, respectively.

Tracker	nBEKCF-D	UPDT	ECO	SiamFC	CFNet	MDNet
Prec. (%)	56.3	55.7	49.2	53.3	53.3	56.5
Norm.Prec. (%)	70.8	70.2	61.8	66.6	65.4	70.5
Success (AUC) (%)	61.6	61.1	55.4	57.1	57.8	60.6

Table 2: Comparison among nBEKCF-D and state-of-the-art trackers on the TrackingNet test set in terms of precision, normalized precision, and success. The best three results are shown in red, green and blue, respectively.

2.2. Evaluation on VOT, GOT10k, and TrackingNet

VOT2018 [8]. We evaluate our nBEKCF-D on Visual Object Tracking (VOT) 2018 challenge which consists of 60 sequences. On the VOT2018 experiment, we compare nBEKCF-D against the latest CF tracker fdKCF* [13] along with the top-20 trackers on VOT2018 challenge. Following the VOT challenge protocol, all trackers are evaluated by expected average overlap (EAO). Fig. 1 shows the result. It is seen from the figure that our nBEKCF-D achieves EAO 0.381, outperforming most state-of-the-art CF trackers, C-COT [4], ECO [3], CFCF [5], UPDT [2], and fdKCF*. In fact, the main reason that SiamRPN outperforms nBEKCF-D is that the former trains a deep network to regress the bounding boxes of targets finely on largest datasets, while the latter only employs a scale-pyramid scheme.

GOT10k [7]. We evaluate nBEKCF-D on the test set of GOT10k which is a large-scale tracking benchmark and contains 180 test videos. On the GOT10k experiment, we compare nBEKCF-D against seven state-of-the-art trackers, MDNet [11], HCF [9], ECO, CCOT, GOTURN [6], SiamFC [1], and CFNet [12], in which there is no similar step to SiamRPN to learn to regress the bounding boxes finely. Note that UPDT and fdKCF* are not publicly tested on GOT10k. Following the GOT10k challenge protocol, we evaluate all trackers by average overlap, and success rates at overlap thresholds 0.5 and 0.75. The results are shown in Table 1. It is seen from the table that nBEKCF-D outperforms other trackers with large margins, except for CFNet in terms of $\text{SR}_{0.75}$. In fact, CFNet learns its features in an end-to-end way on large datasets, whereas nBEKCF-D only applies off-the-shelf features.

TrackingNet [10]. We evaluate our nBEKCF-D on the test set of TrackingNet which is a large-scale tracking benchmark and provides 511 test videos in the wild to assess trackers. On the TrackingNet experiment, we compare nBEKCF-D against five state-of-the-art trackers, MDNet, ECO, SiamFC, CFNet, and UPDT. Identically, there is no similar step to that in SiamRPN to learn to regress the bounding boxes finely in all these trackers. Note that fdKCF* is not publicly tested on TrackingNet. The results are shown in Table 2. It is seen from the table that nBEKCF-D outperforms other trackers, except for MDNet on the precision. It is known that MDNet needs to be trained on large datasets, whereas our nBEKCF need not.

It is interesting to notice from Table 2 that our nBEKCF-D still outperforms UPDT, even if it only applies ResNet50 and data augmentation, and does not employ other improvements that UPDT adopted to develop SRDCF.

3. Notation

Let $\langle , \rangle$ be the dot product, $\langle \mathbf{H}_1, \mathbf{H}_2 \rangle = \langle \text{vec}(\mathbf{H}_1), \text{vec}(\mathbf{H}_2) \rangle$, where $\text{vec}(\mathbf{H})$ indicates the vectorization of the matrix $\mathbf{H}$, $\mathbf{H}_{((a_1,b_1),(a_2,b_2))}$ be the sub-matrix of matrix $\mathbf{H}$ with (a_1, b_1) and (a_2, b_2) as its top left and down right corners, respectively, $H_{a,b}$ or $\mathbf{H}(a, b)$ be an element of matrix $\mathbf{H}$, and $[\bullet]$ is a matrix with $\bullet$ as its element. $\mathbf{F}\text{☆}\mathbf{S} = [\langle \mathbf{F}, \mathbf{S}_{((a_1,b_1),(a_2,b_2))} \rangle]$,[1] where $((a_1, b_1), (a_2, b_2)) \in \mathbb{N}^2$ and $\mathbb{N}^2$ is a domain of spatial location. The matrix with a pair of superscripts is still a matrix.

[1] In fact, in this paper and its supplementary material, ☆ is exactly the same as the correlation without padding in convolutional neural networks.

4. Illustration of Algorithm 2 (CCIM)

Suppose

$$\mathbf{Z} \equiv \mathbf{Z}^{\mathbf{0,0}} = \begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{3} \\ \mathbf{4} & \mathbf{5} & \mathbf{6} \\ \mathbf{7} & \mathbf{8} & \mathbf{9} \end{bmatrix}$$

and

$$\mathbf{X} = \begin{bmatrix} \mathbf{a} & \mathbf{b} & \mathbf{c} & \mathbf{d} & \mathbf{e} \\ \mathbf{f} & \mathbf{g} & \mathbf{h} & \mathbf{i} & \mathbf{j} \\ \mathbf{k} & \mathbf{l} & \mathbf{m} & \mathbf{n} & \mathbf{o} \\ \mathbf{p} & \mathbf{q} & \mathbf{r} & \mathbf{s} & \mathbf{t} \\ \mathbf{u} & \mathbf{w} & \mathbf{x} & \mathbf{y} & \mathbf{z} \end{bmatrix},$$

where the numbers in $\mathbf{Z}$ indicate its elements, rather than the elements' real values. $\mathbf{z}_{s,t} \in \mathbb{R}^D$ and $\mathbf{x}_{s,t} \in \mathbb{R}^D$ are element of $\mathbf{Z}$ and $\mathbf{X}$, respectively. $\mathbf{Z}$ and $\mathbf{X}$ are called base patch and learning region, respectively.

Let 5×5 fundamental calculation matrices

$$\mathbf{A}^{s,t} = \mathbf{z}_{s,t}\text{☆}\mathbf{X}$$

and 5×5 fundamental matrices

$$\mathbf{B}^{s,t} = \mathbf{P}_5^{-s}\mathbf{A}^{s,t}\mathbf{Q}_5^{-t} = \mathbf{P}_5^{-s}(\mathbf{z}_{s,t}\text{☆}\mathbf{X})\mathbf{Q}_5^{-t},$$

where $\mathbf{P}_5$ and $\mathbf{Q}_5$ are defined in the paper with $m = n = 5$, $s = 0, 1, 2$, and $t = 0, 1, 2$. $\mathbf{A}^{s,t}$'s and $\mathbf{B}^{s,t}$'s are shown in Fig.2.[2] Intuitively, $\mathbf{B}^{s,t}$ is generated through cyclically shifting $\mathbf{A}^{s,t}$.

In the rest of this section, we will explain how CCIM works exactly with $\mathbf{Z}\text{☆}\mathbf{X}$ (*i.e.*, $\mathbf{Z}^{0,0}\text{☆}\mathbf{X}$) and $\mathbf{Z}^{1,1}\text{☆}\mathbf{X}$.

4.1. Correlation of Z and X

It is seen that $(\mathbf{Z}\text{☆}\mathbf{X})_{(0,0)} = \langle \mathbf{Z}, \mathbf{X}_{((0,0),(2,2))} \rangle$. According to the construction of $\mathbf{A}^{s,t}$'s and $\mathbf{B}^{s,t}$'s,

$$\langle \mathbf{z}_{s,t}, \mathbf{x}_{s,t} \rangle = B_{0,0}^{s,t},$$

where $s = 0, 1, 2$ and $t = 0, 1, 2$, $B_{0,0}^{s,t}$ is the element of $\mathbf{B}^{s,t}$ at the top left corner $(0, 0)$. $\langle \mathbf{z}_{s,t}, \mathbf{x}_{s,t} \rangle$'s are marked by the blue blocks in Fig.2(b). Therefore,

$$(\mathbf{Z}\text{☆}\mathbf{X})_{(0,0)} = \sum_{s=0}^{2}\sum_{t=0}^{2}\langle \mathbf{z}_{s,t}, \mathbf{x}_{s,t} \rangle = \sum_{s=0}^{2}\sum_{t=0}^{2} B_{0,0}^{s,t}.$$

Generally,

$$(\mathbf{Z}\text{☆}\mathbf{X})_{(u,v)} = \langle \mathbf{Z}, \mathbf{X}_{((u,v),(u+2,v+2))} \rangle,$$

where $u = 0, 1, 2$ and $v = 0, 1, 2$. Because

$$\langle \mathbf{z}_{s,t}, \mathbf{x}_{u+s,v+t} \rangle = B_{u,v}^{s,t},$$

[2] All the numbers of equations and figures refer to the equations and figures of this supplementary material.

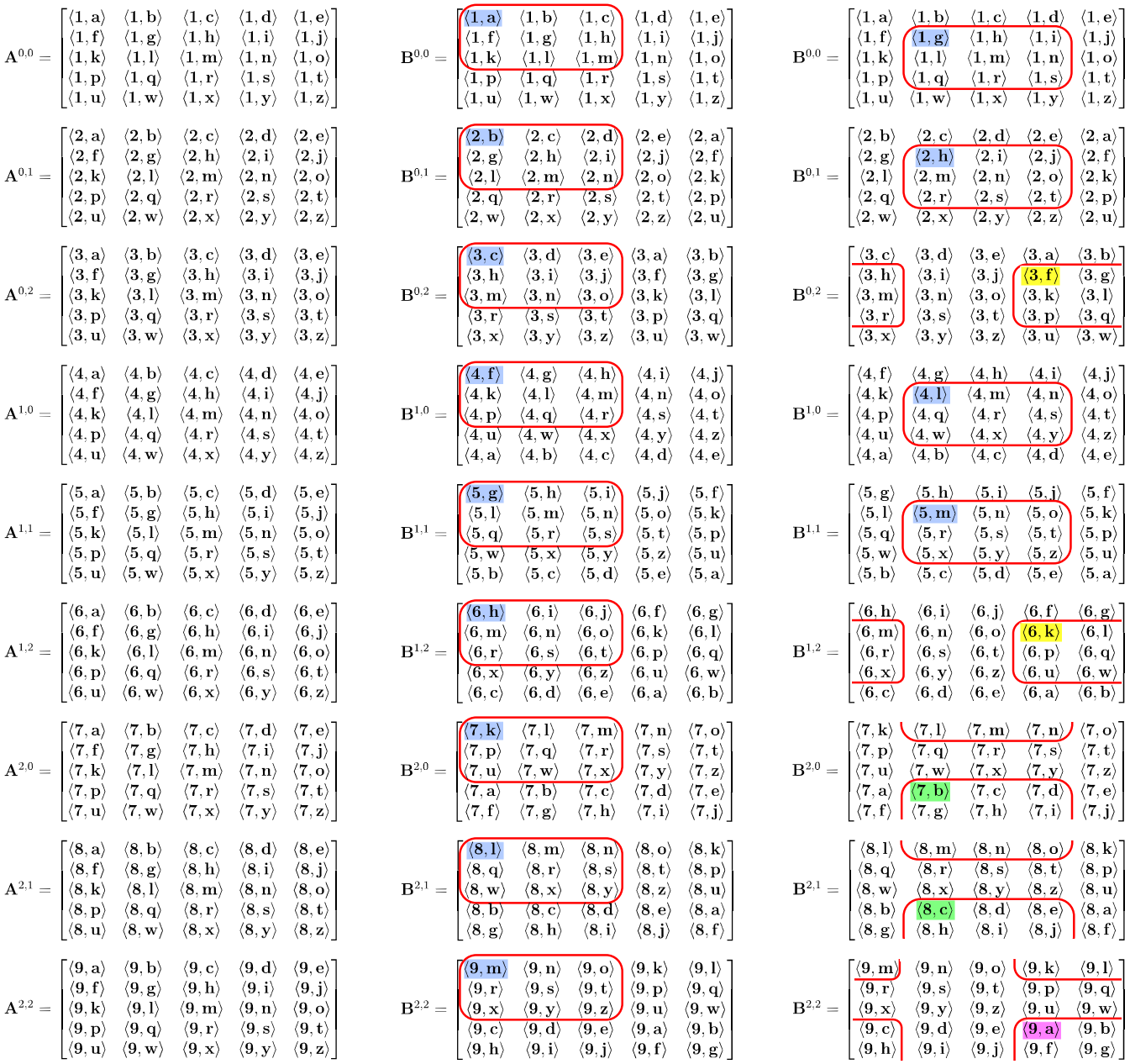
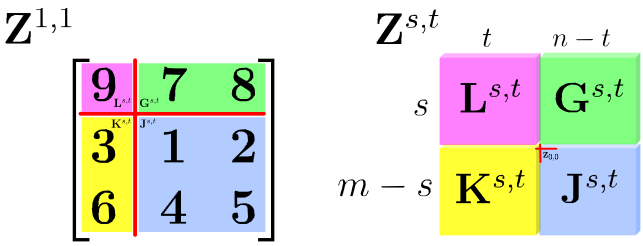


Figure 2: (a) Fundamental calculation matrices $\mathbf{A}^{s,t}$'s. (b) Fundamental matrices, $\mathbf{B}^{s,t}$'s, with sub-matrices marked by red bounding boxes. These sub-matrices are used while calculating $\mathbf{Z}\text{☆}\mathbf{X}$. See Sec.4.1 for details. (c) Fundamental matrices, $\mathbf{B}^{s,t}$'s, with sub-matrices marked by red bounding boxes. These sub-matrices are used while calculating $\mathbf{Z}^{1,1}\text{☆}\mathbf{X}$. See Sec.4.2 for details.

$$(\mathbf{Z}\text{☆}\mathbf{X})_{(u,v)} = \sum_{s=0}^{2}\sum_{t=0}^{2}\langle \mathbf{z}_{s,t}, \mathbf{x}_{u+s,v+t}\rangle = \sum_{s=0}^{2}\sum_{t=0}^{2} B^{s,t}_{u,v}.$$

Consequently,

$$\mathbf{Z}\text{☆}\mathbf{X} = \left(\sum_{s=0}^{2}\sum_{t=0}^{2}\mathbf{B}^{s,t}\right)_{((0,0),(2,2))}.$$



(a) (b)

Figure 3: (a) Based on the position of $\mathbf{z}_{0,0} = \mathbf{1}$, $\mathbf{Z}^{1,1}$ is divided into four parts, $\mathbf{L}^{1,1} = (9)$, $\mathbf{G}^{1,1} = (7, 8)$, $\mathbf{K}^{1,1} = (3; 6)$, and $\mathbf{J}^{1,1} = ((1, 2); (4, 5))$. (b) The division of a general cyclic base $\mathbf{Z}^{s,t}$ into four parts, $\mathbf{L}^{s,t}$, $\mathbf{G}^{s,t}$, $\mathbf{K}^{s,t}$, and $\mathbf{J}^{s,t}$ by means of the position of $\mathbf{z}_{0,0}$.

4.2. Correlation of $\mathbf{Z}^{1,1}$ and $\mathbf{X}$

$\mathbf{Z}^{1,1}$, as shown in Fig.3(a), is generated by means of Eq.(1) defined in Sec. 5 with $m = n = 5$ and $s = t = 1$. It can be divided into four parts, $\mathbf{L}^{1,1}$, $\mathbf{G}^{1,1}$, $\mathbf{K}^{1,1}$, and $\mathbf{J}^{1,1}$. Then, we have

$$\begin{aligned}\mathbf{Z}^{1,1}\text{☆}\mathbf{X} = &(\mathbf{L}^{1,1}\text{☆}\mathbf{X})_{((0,0),(2,2))} + (\mathbf{G}^{1,1}\text{☆}\mathbf{X})_{((0,1),(2,3))} \\ &+ (\mathbf{K}^{1,1}\text{☆}\mathbf{X})_{((1,0),(3,2))} + (\mathbf{J}^{1,1}\text{☆}\mathbf{X})_{((1,1),(3,3))}.\end{aligned}$$

Suppose the 0^{th} and 4^{th} rows are adjacent, and the 0^{th} and 4^{th} columns are adjacent too in $\mathbf{B}^{s,t}$'s. Then, $(\mathbf{L}^{1,1}\text{☆}\mathbf{X})_{((0,0),(2,2))}$ equals to the sub-matrix of $\mathbf{B}^{2,2}$ marked by a red bounding box in $\mathbf{B}^{2,2}$ of Fig.2(c). Cyclically shifting the marked sub-matrix so as to make the element marked by the pink block in $\mathbf{B}^{2,2}$ of Fig.2(c) be at the top left corner, and letting $\mathbf{L}$ be the resulting matrix, we have

$$(\mathbf{L}^{1,1}\text{☆}\mathbf{X})_{((0,0),(2,2))} = \mathbf{L}_{((0,0),(2,2))}.$$

Similarly, $(\mathbf{G}^{1,1}\text{☆}\mathbf{X})_{((0,1),(2,3))}$ equals to the summation of two sub-matrices of $\mathbf{B}^{2,0}$ and $\mathbf{B}^{2,1}$ marked by two red bounding boxes in $\mathbf{B}^{2,0}$ and $\mathbf{B}^{2,1}$ of Fig.2(c), respectively. Cyclically shifting $\mathbf{B}^{2,0} + \mathbf{B}^{2,1}$ so as to make the summation of two elements marked by the green blocks in $\mathbf{B}^{2,0}$ and $\mathbf{B}^{2,1}$ of Fig.2(c) be at the top left corner, and letting $\mathbf{G}$ be the resulting matrix, we have

$$(\mathbf{G}^{1,1}\text{☆}\mathbf{X})_{((0,1),(2,3))} = \mathbf{G}_{((0,0),(2,2))}.$$

$(\mathbf{K}^{1,1}\text{☆}\mathbf{X})_{((1,0),(3,2))}$ equals to the summation of two sub-matrices of $\mathbf{B}^{0,2}$ and $\mathbf{B}^{1,2}$ marked by two red bounding boxes in $\mathbf{B}^{0,2}$ and $\mathbf{B}^{1,2}$ of Fig.2(c), respectively. Cyclically shifting $\mathbf{B}^{0,2} + \mathbf{B}^{1,2}$ so as to make the summation of two elements marked by the yellow blocks in $\mathbf{B}^{0,2}$ and $\mathbf{B}^{1,2}$ of Fig.2(c) be at the top left corner, and letting $\mathbf{K}$ be the resulting matrix, we have

$$(\mathbf{K}^{1,1}\text{☆}\mathbf{X})_{((1,0),(3,2))} = \mathbf{K}_{((0,0),(2,2))}.$$

Finally, $(\mathbf{J}^{1,1}\text{☆}\mathbf{X})_{((1,1),(3,3))}$ equals to the summation of four sub-matrices of $\mathbf{B}^{0,0}$, $\mathbf{B}^{0,1}$, $\mathbf{B}^{1,0}$, and $\mathbf{B}^{1,1}$ marked by four red bounding boxes in $\mathbf{B}^{0,0}$, $\mathbf{B}^{0,1}$, $\mathbf{B}^{1,0}$, and $\mathbf{B}^{1,1}$ of Fig.2(c), respectively. Cyclically shifting $\mathbf{B}^{0,0} + \mathbf{B}^{0,1} + \mathbf{B}^{1,0} + \mathbf{B}^{1,1}$ so as to make the summation of four elements marked by the blue blocks in $\mathbf{B}^{0,0}$, $\mathbf{B}^{0,1}$, $\mathbf{B}^{1,0}$, and $\mathbf{B}^{1,1}$ of Fig.2(c) be at the top left corner, and letting $\mathbf{J}$ be the resulting matrix, we have

$$(\mathbf{J}^{1,1}\text{☆}\mathbf{X})_{((1,1),(3,3))} = \mathbf{J}_{((0,0),(2,2))}.$$

Consequently, let $\mathbf{C}^{1,1} = \sum_{\mathbf{D}\in\{\mathbf{L},\mathbf{G},\mathbf{K},\mathbf{J}\}} \mathbf{D}$. Then,

$$\mathbf{Z}^{1,1}\text{☆}\mathbf{X} = \mathbf{C}^{1,1}_{((0,0),(2,2))}.$$

It is easy to see that $\mathbf{Z}^{s,t}\text{☆}\mathbf{X}$'s can also be obtained with similar procedures, where $s \in \{0, 1, 2\}$, $t \in \{0, 1, 2\}$.

4.3. Design Integral Matrix to Accelerate Correlation

Given $\mathbf{Z}$, according to the construction of $\mathbf{B}^{s,t}$, where $s = 0, 1, 2$ and $t = 0, 1, 2$, $\mathbf{z}_{s,t}$ is the common factor of all elements of the $\mathbf{B}^{s,t}$. Replacing $\mathbf{x}_{s,t}$ by $\mathbf{B}^{s,t}$ in $\mathbf{Z}$, we have the block matrix

$$\mathbf{T} = \begin{bmatrix} \mathbf{B}^{0,0} & \mathbf{B}^{0,1} & \mathbf{B}^{0,2} \\ \mathbf{B}^{1,0} & \mathbf{B}^{1,1} & \mathbf{B}^{1,2} \\ \mathbf{B}^{2,0} & \mathbf{B}^{2,1} & \mathbf{B}^{2,2} \end{bmatrix}.$$

We call $\mathbf{B}^{s,t}$ a $\mathbf{T}$'s element in the rest of this section.

According to the examples in Secs.4.1 and 4.2, it is necessary to calculate $\sum_{s=0}^{2}\sum_{t=0}^{2} \mathbf{B}^{s,t}$ for obtaining $\mathbf{Z}\text{☆}\mathbf{X}$, and to calculate $\mathbf{B}^{2,2}$, $\mathbf{B}^{2,0} + \mathbf{B}^{2,1}$, $\mathbf{B}^{0,2} + \mathbf{B}^{1,2}$, and $\mathbf{B}^{0,0} + \mathbf{B}^{0,1} + \mathbf{B}^{1,0} + \mathbf{B}^{1,1}$ for obtaining $\mathbf{Z}^{1,1}\text{☆}\mathbf{X}$. In order to eliminate repeated additions, the three summations of matrices, which are necessary to calculate both $\mathbf{Z}\text{☆}\mathbf{X}$ and $\mathbf{Z}^{1,1}\text{☆}\mathbf{X}$, should be shared. Notice that each above summation of $\mathbf{B}^{s,t}$'s is involved with the summation of $\mathbf{T}$'s elements. Constructing the block integral matrix $\mathbf{M}$, we have

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}^{0,0} & \mathbf{M}^{0,1} & \mathbf{M}^{0,2} \\ \mathbf{M}^{1,0} & \mathbf{M}^{1,1} & \mathbf{M}^{1,2} \\ \mathbf{M}^{2,0} & \mathbf{M}^{2,1} & \mathbf{M}^{2,2} \end{bmatrix}$$

with $\mathbf{M}^{i,j} = \sum_{0\leqslant s\leqslant i, 0\leqslant t\leqslant j} \mathbf{B}^{s,t}$ as its block elements.[3] Then,

$$\begin{aligned}&\mathbf{M}^{2,2} = \sum_{s=0}^{2}\sum_{t=0}^{2} \mathbf{B}^{s,t}, \\ &\mathbf{M}^{2,2} - \mathbf{M}^{1,2} - \mathbf{M}^{2,1} + \mathbf{M}^{1,1} = \mathbf{B}^{2,2}, \\ &\mathbf{M}^{2,1} - \mathbf{M}^{1,1} = \mathbf{B}^{2,0} + \mathbf{B}^{2,1}, \\ &\mathbf{M}^{1,2} - \mathbf{M}^{1,1} = \mathbf{B}^{0,2} + \mathbf{B}^{1,2}, \\ &\mathbf{M}^{1,1} = \mathbf{B}^{0,0} + \mathbf{B}^{0,1} + \mathbf{B}^{1,0} + \mathbf{B}^{1,1},\end{aligned}$$

[3] In the paper, we denote a block element of $\mathbf{M}$ with $\mathbf{M}_{i,j}$.

It is seen that the five summations can be achieved within a constant duration by means of $\mathbf{M}$. This way, the repeated additions of $\mathbf{B}^{s,t}$'s are eliminated.

In general, because $\mathbf{D}^{s,t}$ is a sub-matrix of $\mathbf{Z}$, where $\mathbf{D}^{s,t} \in \{\mathbf{L}^{s,t}, \mathbf{G}^{s,t}, \mathbf{K}^{s,t}, \mathbf{J}^{s,t}\}$, the $\mathbf{B}^{i,j}$'s with the elements of $\mathbf{D}^{s,t}$ as their common factors also constitute a $\mathbf{T}$'s elements. While calculating $\mathbf{D}_L^{s,t}$ of Eq. (2) which is defined in Sec. 5, similar to the examples in Secs.4.1 and 4.2, it is necessary to calculate the summation of the $\mathbf{T}$'s elements. For a given $\mathbf{Z}$, while (s,t) traverses all possible pairs, *i.e.*, $\mathbf{Z}^{s,t}$ traverses all possible cyclic base patches, the brute-force calculation of the summations will arise a great deal of repeated summations, reducing the efficiency of correlations greatly especially when the size of base patches is large. Therefore, it is a natural choice to construct the integral matrix, as described in CCIM, to obtain such summations without repeated additions.

5. Correctness of Algorithm 2 (CCIM)

Suppose $\mathbf{Z} \in \mathbb{R}^{m\times n\times D}$ is the base patch with $\mathbf{z}_{s,t} \in \mathbb{R}^D$'s as its elements, cyclic base patches $\mathbf{Z}^{s,t}$'s are generated by using

$$\mathbf{Z}^{s,t}(d) = \mathbf{P}_m^s \mathbf{Z}(d) \mathbf{Q}_n^t, \tag{1}$$

where $d = 0, \ldots, D-1$, $s = 0, \ldots, m-1$, and $t = 0, \ldots, n-1$, $\mathbf{Z}(d)$ is the d-th channel of $\mathbf{Z}$, $\mathbf{P}_{l_1}$, $l_1 \in \{m, M\}$, and $\mathbf{Q}_{l_2}$, $l_2 \in \{n, N\}$, are the permutation matrices of $l_1 \times l_1$ and $l_2 \times l_2$, respectively.[4] $\mathbf{X} \in \mathbb{R}^{M\times N\times D}$ is the learning region with $\mathbf{x}_{p,q} \in \mathbb{R}^D$'s as its elements.

Let $M \times N$ fundamental calculation matrix

$$\mathbf{A}^{s,t} = \mathbf{z}_{s,t} \text{☆} \mathbf{X},$$

where $s = 0, \ldots, m-1, t = 0, \ldots, n-1$, and $M \times N$ fundamental matrix

$$\mathbf{B}^{s,t} = \mathbf{P}_M^{-s} \mathbf{A}^{s,t} \mathbf{Q}_N^{-t} = \mathbf{P}_5^{-s} (\mathbf{z}_{s,t} \text{☆} \mathbf{X}) \mathbf{Q}_5^{-t}.$$

$\mathbf{B}^{s,t}$'s are generated through cyclic-shifts of $\mathbf{A}^{s,t}$'s. Then, $\mathbf{z}_{s,t}$ only appears in $\mathbf{A}^{s,t}$ and $\mathbf{B}^{s,t}$ as a common factor of their elements. The sets of whole $\mathbf{A}^{s,t}$'s and $\mathbf{B}^{s,t}$'s are denoted as $\mathfrak{A}$ and $\mathfrak{B}$, respectively.

In general, as shown in Fig.3(b), given s and t, $\mathbf{Z}^{s,t}$ can always be divided into four parts, $\mathbf{L}^{s,t}$, $\mathbf{G}^{s,t}$, $\mathbf{K}^{s,t}$, and $\mathbf{J}^{s,t}$ by the position of $\mathbf{z}_{0,0}$. The sizes of $\mathbf{L}^{s,t} \equiv [\mathbf{l}_{i,j}^{s,t}]$, $\mathbf{G}^{s,t} \equiv [\mathbf{g}_{i,j}^{s,t}]$, $\mathbf{K}^{s,t} \equiv [\mathbf{k}_{i,j}^{s,t}]$, and $\mathbf{J}^{s,t} \equiv [\mathbf{j}_{i,j}^{s,t}]$ are $s \times t$, $s \times (n-t)$, $(m-s) \times t$, and $(m-s) \times (n-t)$, respectively, and their top left elements are $\mathbf{l}_{0,0}^{s,t} = \mathbf{z}_{m-s,n-t}$, $\mathbf{g}_{0,0}^{s,t} = \mathbf{z}_{m-s,0}$, $\mathbf{k}_{0,0}^{s,t} = \mathbf{z}_{0,n-t}$, and $\mathbf{j}_{0,0}^{s,t} = \mathbf{z}_{0,0}$, respectively. Note that $\mathbf{L}^{s,t}$ and $\mathbf{G}^{s,t}$ or $\mathbf{L}^{s,t}$ and $\mathbf{K}^{s,t}$ will not exist if $s = 0$ or $t = 0$. $\mathbf{Z}^{s,t} \text{☆} \mathbf{X}$ can then be divided into four parts, *i.e.*,

$$\mathbf{Z}^{s,t} \text{☆} \mathbf{X} = \mathbf{D}_L^{s,t} + \mathbf{D}_G^{s,t} + \mathbf{D}_K^{s,t} + \mathbf{D}_J^{s,t}, \tag{2}$$

[4]The detailed explanation of Eq. (1) can be found in the paper.

where

$$\begin{aligned}
\mathbf{D}_L^{s,t} &= (\mathbf{L}^{s,t} \text{☆} \mathbf{X})_{((0,0),(M-m,N-n))},\\
\mathbf{D}_G^{s,t} &= (\mathbf{G}^{s,t} \text{☆} \mathbf{X})_{((0,t),(M-m,N-n+t))},\\
\mathbf{D}_K^{s,t} &= (\mathbf{K}^{s,t} \text{☆} \mathbf{X})_{((s,0),(M-m+s,N-n))},\\
\mathbf{D}_J^{s,t} &= (\mathbf{J}^{s,t} \text{☆} \mathbf{X})_{((s,t),(M-m+s,N-n+t))},
\end{aligned}$$

$\mathbf{D}_L^{s,t} = [\mathbf{D}_L^{s,t}(u,v)]$, $\mathbf{D}_G^{s,t} = [\mathbf{D}_G^{s,t}(u,v)]$, $\mathbf{D}_K^{s,t} = [\mathbf{D}_K^{s,t}(u,v)]$, $\mathbf{D}_J^{s,t} = [\mathbf{D}_J^{s,t}(u,v)]$, $u = 0, \ldots, M-m$ and $v = 0, \ldots, N-n$.

5.1. Calculation of $\mathbf{D}_L^{s,t}(u,v)$

It is clear that

$$\mathbf{D}_L^{s,t}(u,v) = \sum_{i=0}^{s-1} \sum_{j=0}^{t-1} \langle \mathbf{l}_{i,j}^{s,t}, \mathbf{x}_{u+i,v+j} \rangle$$

and $\mathbf{l}_{i,j}^{s,t} = \mathbf{z}_{m-s+i,n-t+j}$ which only appears in $\mathbf{B}^{m-s+i,n-t+j}$. We have

$$\langle \mathbf{l}_{i,j}^{s,t}, \mathbf{x}_{u+i,v+j} \rangle = A_{u+i,v+j}^{m-s+i,n-t+j}$$

and

$$\begin{aligned}
\mathbf{B}^{m-s+i,n-t+j} &= \mathbf{P}_M^{-(m-s+i)} \mathbf{A}^{m-s+i,n-t+j} \mathbf{Q}_N^{-(n-t+j)}\\
&= \mathbf{P}_M^{-(m-s)} \mathbf{P}_M^{-i} \mathbf{A}^{m-s+i,n-t+j} \mathbf{Q}_N^{-j} \mathbf{Q}_N^{-(n-t)}.
\end{aligned}$$

Let $\mathbf{A}^{(L)} = \mathbf{P}_M^{-i} \mathbf{A}^{m-s+i,n-t+j} \mathbf{Q}_N^{-j}$. Then,

$$A_{u,v}^{(L)} = A_{u+i,v+j}^{m-s+i,n-t+j}.$$

That is,

$$\langle \mathbf{l}_{i,j}^{s,t}, \mathbf{x}_{u+i,v+j} \rangle = A_{u,v}^{(L)}.$$

$$\begin{aligned}
&\because \mathbf{B}^{m-s+i,n-t+j} = \mathbf{P}_M^{-(m-s)} \mathbf{A}^{(L)} \mathbf{Q}_N^{-(n-t)},\\
&\therefore \langle \mathbf{l}_{i,j}^{s,t}, \mathbf{x}_{u+i,v+j} \rangle = B_{(u-(m-s))\bmod M,(v-(n-t))\bmod N}^{m-s+i,n-t+j}.
\end{aligned}$$

Therefore,

$$\mathbf{D}_L^{s,t}(u,v) = \sum_{i=0}^{s-1} \sum_{j=0}^{t-1} B_{(u-(m-s))\bmod M,(v-(n-t))\bmod N}^{m-s+i,n-t+j}. \tag{3}$$

It is seen that the subscripts of every item are irrelative to (i,j) in the right hand side of Eq. (3). That is, all of the items possess the same subscripts, although they belong to different fundamental matrices.

It is seen from Fig.3(b) that the top left and bottom right elements of $\mathbf{L}^{s,t}$ are $\mathbf{z}_{m-s,n-t}$ and $\mathbf{z}_{m-1,n-1}$, respectively. Let $\mathbf{S}^{L,s,t} = \sum_{i=m-s}^{m-1} \sum_{j=n-t}^{n-1} \mathbf{B}^{i,j}$. According to the

construction of integral matrix $\mathbf{M}$, $\mathbf{S}^{L,s,t}$ can be calculated within a constant time as

$$\begin{aligned}\mathbf{S}^{L,s,t} &= \mathbf{M}^{m-1,n-1} - \mathbf{M}^{m-1,n-t-1} \\ &\quad - \mathbf{M}^{m-s-1,n-1} + \mathbf{M}^{m-s-1,n-t-1}.\end{aligned}$$

Then, we have

$$\mathbf{D}_L^{s,t}(u,v) = S^{L,s,t}_{(u-(m-s))\bmod M,(v-(n-t))\bmod N}.$$

5.2. Calculation of $\mathbf{D}_G^{s,t}(u,v)$

It is clear that

$$\mathbf{D}_G^{s,t}(u,v) = \sum_{i=0}^{s-1}\sum_{j=0}^{n-t-1}\langle \mathbf{g}_{i,j}^{s,t}, \mathbf{x}_{u+i,v+t+j}\rangle$$

and $\mathbf{g}_{i,j}^{s,t} = \mathbf{z}_{m-s+i,j}$ which only appears in $\mathbf{B}^{m-s+i,j}$. We have

$$\langle \mathbf{g}_{i,j}^{s,t}, \mathbf{x}_{u+i,v+t+j}\rangle = A^{m-s+i,j}_{u+i,v+t+j}$$

and

$$\begin{aligned}\mathbf{B}^{m-s+i,j} &= \mathbf{P}_M^{-(m-s+i)}\mathbf{A}^{m-s+i,j}\mathbf{Q}_N^{-j} \\ &= \mathbf{P}_M^{-(m-s)}\mathbf{P}_M^{-i}\mathbf{A}^{m-s+i,j}\mathbf{Q}_N^{-j}.\end{aligned}$$

Let $\mathbf{A}^{(G)} = \mathbf{P}_M^{-i}\mathbf{A}^{m-s+i,j}\mathbf{Q}_N^{-j}$. Then,

$$A^{(G)}_{u,v+t} = A^{m-s+i,j}_{u+i,v+t+j}.$$

That is,

$$\langle \mathbf{g}_{i,j}^{s,t}, \mathbf{x}_{u+i,v+t+j}\rangle = A^{(G)}_{u,v+t}.$$

$$\begin{aligned}&\because \mathbf{B}^{m-s+i,j} = \mathbf{P}_M^{-(m-s)}\mathbf{A}^{(G)}, \\ &\therefore \langle \mathbf{g}_{i,j}^{s,t}, \mathbf{x}_{u+i,v+t+j}\rangle = B^{m-s+i,j}_{(u-(m-s))\bmod M,v+t}.\end{aligned}$$

Therefore,

$$\mathbf{D}_G^{s,t}(u,v) = \sum_{i=0}^{s-1}\sum_{j=0}^{n-t-1} B^{m-s+i,j}_{(u-(m-s))\bmod M,v+t}. \quad (4)$$

It is seen that the subscripts of every item are irrelative to (i,j) in the right hand side of Eq. (4). That is, all of the items possess the same subscripts, although they belong to different fundamental matrices.

It is seen from Fig.3(b) that the top left and bottom right elements of $\mathbf{G}^{s,t}$ are $\mathbf{z}_{m-s,0}$ and $\mathbf{z}_{m-1,n-t-1}$, respectively. Let $\mathbf{S}^{G,s,t} = \sum_{i=m-s}^{m-1}\sum_{j=0}^{n-t-1}\mathbf{B}^{i,j}$. According to the construction of integral matrix, $\mathbf{S}^{G,s,t}$ can be calculated within a constant time as

$$\mathbf{S}^{G,s,t} = \mathbf{M}^{m-1,n-t-1} - \mathbf{M}^{m-s-1,n-t-1}.$$

Then, we have

$$\mathbf{D}_G^{s,t}(u,v) = S^{G,s,t}_{(u-(m-s))\bmod M,v+t}.$$

5.3. Calculation of $\mathbf{D}_K^{s,t}(u,v)$

It is clear that

$$\mathbf{D}_K^{s,t}(u,v) = \sum_{i=0}^{m-s-1}\sum_{j=0}^{t-1}\langle \mathbf{k}_{i,j}^{s,t}, \mathbf{x}_{u+s+i,v+j}\rangle$$

and $\mathbf{k}_{i,j}^{s,t} = \mathbf{z}_{i,n-t+j}$ which only appears in $\mathbf{B}^{i,n-t+j}$. We have

$$\langle \mathbf{k}_{i,j}^{s,t}, \mathbf{x}_{u+s+i,v+j}\rangle = A^{i,n-t+j}_{u+s+i,v+j}$$

and

$$\begin{aligned}\mathbf{B}^{i,n-t+j} &= \mathbf{P}_M^{-i}\mathbf{A}^{i,n-t+j}\mathbf{Q}_N^{-(n-t+j)} \\ &= \mathbf{P}_M^{-i}\mathbf{A}^{i,n-t+j}\mathbf{Q}_N^{-j}\mathbf{Q}_N^{-(n-t)}.\end{aligned}$$

Let $\mathbf{A}^{(K)} = \mathbf{P}_M^{-i}\mathbf{A}^{i,n-t+j}\mathbf{Q}_N^{-j}$. Then,

$$A^{(K)}_{u+s,v} = A^{i,n-t+j}_{u+s+i,v+j},$$

That is,

$$\langle \mathbf{k}_{i,j}^{s,t}, \mathbf{x}_{u+s+i,v+j}\rangle = A^{(K)}_{u+s,v}.$$

$$\begin{aligned}&\because \mathbf{B}^{i,n-t+j} = \mathbf{A}^{(K)}\mathbf{Q}_N^{-(n-t)}, \\ &\therefore \langle \mathbf{k}_{i,j}^{s,t}, \mathbf{x}_{u+s+i,v+j}\rangle = B^{i,n-t+j}_{u+s,(v-(n-t))\bmod N}.\end{aligned}$$

Therefore,

$$\mathbf{D}_K^{s,t}(u,v) = \sum_{i=0}^{m-s-1}\sum_{j=0}^{t-1} B^{i,n-t+j}_{u+s,(v-(n-t))\bmod N}. \quad (5)$$

It is seen that the subscripts of every item are irrelative to (i,j) in the right hand side of Eq. (5). That is, all of the items possess the same subscripts, although they belong to different fundamental matrices.

It is seen from Fig.3(b) that the top left and bottom right elements of $\mathbf{K}^{s,t}$ are $\mathbf{z}_{0,n-t}$ and $\mathbf{z}_{m-s-1,n-1}$, respectively. Let $\mathbf{S}^{K,s,t} = \sum_{i=0}^{m-s-1}\sum_{j=n-t}^{n-1}\mathbf{B}^{i,j}$. According to the construction of integral matrix, $\mathbf{S}^{K,s,t}$ can be calculated within a constant time as

$$\mathbf{S}^{K,s,t} = \mathbf{M}^{m-s-1,n-1} - \mathbf{M}^{m-s-1,n-t-1}.$$

Then ,we have

$$\mathbf{D}_K^{s,t}(u,v) = S^{K,s,t}_{u+s,(v-(n-t))\bmod N}.$$

5.4. Calculation of $\mathbf{D}_J^{s,t}(u,v)$

It is clear that

$$\mathbf{D}_J^{s,t}(u,v) = \sum_{i=0}^{m-s-1}\sum_{j=0}^{n-t-1}\langle \mathbf{j}_{i,j}^{s,t}, \mathbf{x}_{u+s+i,v+t+j}\rangle$$

and $\mathbf{j}_{i,j}^{s,t} = \mathbf{z}_{i,j}$ which only appears in $\mathbf{B}^{i,j}$. We have

$$\langle \mathbf{j}_{i,j}^{s,t}, \mathbf{x}_{u+s+i,v+t+j} \rangle = A_{u+s+i,v+t+j}^{i,j}$$

and

$$\mathbf{B}^{i,j} = \mathbf{P}_M^{-i} \mathbf{A}^{i,j} \mathbf{Q}_N^{-j}.$$

Then, $B_{u+s,v+t}^{i,j} = A_{u+s+i,v+t+j}^{i,j}$. Therefore

$$\langle \mathbf{j}_{i,j}^{s,t}, \mathbf{x}_{u+s+i,v+t+j} \rangle = B_{u+s,v+t}^{i,j}$$

and

$$\mathbf{D}_J^{s,t}(u,v) = \sum_{i=0}^{m-s-1} \sum_{j=0}^{n-t-1} B_{u+s,v+t}^{i,j}. \tag{6}$$

It is seen that the subscripts of every item are irrelative to (i,j) in the right hand side of Eq. (6). That is, all of the items possess the same subscripts, although they belong to different fundamental matrices.

It is seen from Fig.3(b) that the top left and bottom right elements of $\mathbf{J}^{s,t}$ are $\mathbf{z}_{0,0}$ and $\mathbf{z}_{m-s-1,n-t-1}$, respectively. Let $\mathbf{S}^{J,s,t} = \sum_{i=0}^{m-s-1} \sum_{j=0}^{n-t-1} \mathbf{B}^{i,j}$. According to the construction of integral matrix, $\mathbf{S}^{J,s,t}$ can be calculated within constant time as

$$\mathbf{S}^{J,s,t} = \mathbf{M}^{m-s-1,n-t-1}.$$

Then, we have

$$\mathbf{D}_J^{s,t}(u,v) = S_{u+s,v+t}^{J,s,t}.$$

5.5. Correlation of $\mathbf{Z}^{s,t}$ and $\mathbf{X}$

According to the proofs in Secs.5.1, 5.2, 5.3, and 5.4, we have got $\mathbf{D}_L^{s,t}$, $\mathbf{D}_G^{s,t}$, $\mathbf{D}_K^{s,t}$, and $\mathbf{D}_J^{s,t}$. Set

$$\begin{aligned}
r_L(u,v) &= (u-(m-s)) \text{mod} M - u, \\
c_L(u,v) &= (v-(n-t)) \text{mod} N - v, \\
r_G(u,v) &= (u-(m-s)) \text{mod} M - u, \\
c_G(u,v) &= t, \\
r_K(u,v) &= s, \\
c_K(u,v) &= (v-(n-t)) \text{mod} N - v, \\
r_J(u,v) &= s, \\
c_J(u,v) &= t,
\end{aligned} \tag{7}$$

and

$$\begin{aligned}
\mathbf{S}_L^{s,t} &= \mathbf{P}_M^{-r_L(u,v)} \mathbf{S}^{L,s,t} \mathbf{Q}_N^{-c_L(u,v)}, \\
\mathbf{S}_G^{s,t} &= \mathbf{P}_M^{-r_G(u,v)} \mathbf{S}^{G,s,t} \mathbf{Q}_N^{-c_G(u,v)}, \\
\mathbf{S}_K^{s,t} &= \mathbf{P}_M^{-r_K(u,v)} \mathbf{S}^{K,s,t} \mathbf{Q}_N^{-c_K(u,v)}, \\
\mathbf{S}_J^{s,t} &= \mathbf{P}_M^{-r_J(u,v)} \mathbf{S}^{J,s,t} \mathbf{Q}_N^{-c_J(u,v)}.
\end{aligned}$$

Then, $\mathbf{D}_L^{s,t}(u,v)$, $\mathbf{D}_G^{s,t}(u,v)$, $\mathbf{D}_K^{s,t}(u,v)$, and $\mathbf{D}_J^{s,t}(u,v)$ are shifted to $\mathbf{S}_L^{s,t}(u,v)$, $\mathbf{S}_G^{s,t}(u,v)$, $\mathbf{S}_K^{s,t}(u,v)$, and $\mathbf{S}_J^{s,t}(u,v)$, respectively.

Now we prove $\mathbf{P}_M^{-r_L(u,v)} = \mathbf{P}_M^{-r_L(0,0)}$.

$\because 0 \leq s \leqslant m-1, \therefore -m \leqslant -(m-s) \leqslant -1 \leqslant 0. \because M \geqslant m, \therefore 0 \leqslant M-m \leq M-(m-s) \leqslant M-1. \therefore \mathbf{P}_M^{-r_L(0,0)} = \mathbf{P}_M^{-(M-(m-s))}$.

Now we split the set $\{0,\ldots,M-m\}$ into two parts, $S_l = \{0,\ldots,m-s-1\}$ and $S_t = \{m-s,\ldots,M-m\}$.

If $u \in S_l$, then $0 \leqslant u \leqslant m-s-1. \therefore -M \leqslant -(m-s) \leqslant u-(m-s) \leqslant -1$. In this case, $r_L(u,v) = (u-(m-s)) \text{mod} M - u = M+u-(m-s)-u = M-(m-s)$.

If $u \in S_t$, then $m-s \leqslant u \leqslant M-m. \therefore 0 \leqslant u-(m-s) \leqslant M-m-(m-s) < M$. In this case, $r^L(u,v) = (u-(m-s)) \text{mod} M - u = u-(m-s)-u = -(m-s)$.

$\because \mathbf{P}_M^{-(M-(m-s))} = \mathbf{P}_M^{m-s}, \therefore \mathbf{P}_M^{-r_L(u,v)} = \mathbf{P}_M^{-(M-(m-s))}, \therefore \mathbf{P}_M^{-r_L(u,v)} = \mathbf{P}_M^{-r_L(0,0)}$.

Similarly, we can prove

$$\begin{aligned}
&\mathbf{P}_M^{-r_G(u,v)} = \mathbf{P}_M^{-r_G(0,0)}, \\
&\mathbf{Q}_N^{-c_L(u,v)} = \mathbf{Q}_N^{-c_L(0,0)}, \text{ and} \\
&\mathbf{Q}_N^{-c_K(u,v)} = \mathbf{Q}_N^{-c_K(0,0)}.
\end{aligned}$$

Consequently, given (s,t), the right hands of all equations in Eq.(7) are constant and not related to (u,v). All elements of $\mathbf{S}^{L,s,t}$ are shifted by the same number of rows and the same number of columns, respectively. So do $\mathbf{S}^{G,s,t}$, $\mathbf{S}^{K,s,t}$, and $\mathbf{S}^{J,s,t}$. $\mathbf{D}_L^{s,t}(0,0)$, $\mathbf{D}_G^{s,t}(0,0)$, $\mathbf{D}_K^{s,t}(0,0)$, and $\mathbf{D}_J^{s,t}(0,0)$ are shifted to $\mathbf{S}_L^{s,t}(0,0)$, $\mathbf{S}_G^{s,t}(0,0)$, $\mathbf{S}_K^{s,t}(0,0)$, and $\mathbf{S}_J^{s,t}(0,0)$, respectively. Specifically,

$$\begin{aligned}
&(\mathbf{S}_L^{s,t})_{((0,0),(M-m-1,N-n-1))} = \\
&\mathbf{S}^{L,s,t}_{((M-m+s,N-n+t),((M-2m+s) \text{mod} M,(N-2n+t) \text{mod} N))}, \\
&(\mathbf{S}_G^{s,t})_{((0,0),(M-m-1,N-n-1))} = \\
&\mathbf{S}^{G,s,t}_{((M-m+s,t),((M-2m+s) \text{mod} M,N-n+t))}, \\
&(\mathbf{S}_K^{s,t})_{((0,0),(M-m-1,N-n-1))} = \\
&\mathbf{S}^{K,s,t}_{((s,N-n+t),(M-m+s,(N-2n+t) \text{mod} N))}, \\
&(\mathbf{S}_J^{s,t})_{((0,0),(M-m-1,N-n-1))} = \\
&\mathbf{S}^{J,s,t}_{((s,t),(M-m+s,N-n+t))}.
\end{aligned}$$

That is, $\mathbf{D}_L^{s,t}$, $\mathbf{D}_G^{s,t}$, $\mathbf{D}_K^{s,t}$, and $\mathbf{D}_J^{s,t}$ are aligned. Let

$$\mathbf{C}^{s,t} = \mathbf{S}_L^{s,t} + \mathbf{S}_G^{s,t} + \mathbf{S}_K^{s,t} + \mathbf{S}_J^{s,t}.$$

According to Eq.(2),

$$\mathbf{Z}^{s,t} \bigstar \mathbf{X} = \mathbf{C}^{s,t}_{((0,0),(M-m-1,N-n-1))}.$$

Q.E.D.

References

[1] Luca Bertinetto, Jack Valmadre, Joao F Henriques, Andrea Vedaldi, and Philip HS Torr. Fully-convolutional siamese networks for object tracking. In *European conference on computer vision*, pages 850–865. Springer, 2016.

[2] Goutam Bhat, Joakim Johnander, Martin Danelljan, Fahad Shahbaz Khan, and Michael Felsberg. Unveiling the power of deep tracking. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 483–498, 2018.

[3] Martin Danelljan, Goutam Bhat, Fahad Shahbaz Khan, and Michael Felsberg. Eco: Efficient convolution operators for tracking. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6638–6646, 2017.

[4] Martin Danelljan, Andreas Robinson, Fahad Shahbaz Khan, and Michael Felsberg. Beyond correlation filters: Learning continuous convolution operators for visual tracking. In *European Conference on Computer Vision*, pages 472–488. Springer, 2016.

[5] Erhan Gundogdu and A Aydın Alatan. Good features to correlate for visual tracking. *IEEE Transactions on Image Processing*, 27(5):2526–2540, 2018.

[6] David Held, Sebastian Thrun, and Silvio Savarese. Learning to track at 100 fps with deep regression networks. In *European Conference on Computer Vision*, pages 749–765. Springer, 2016.

[7] Lianghua Huang, Xin Zhao, and Kaiqi Huang. Got-10k: A large high-diversity benchmark for generic object tracking in the wild. *arXiv preprint arXiv:1810.11981*, 2018.

[8] Matej Kristan, Ales Leonardis, Jiri Matas, Michael Felsberg, Roman Pflugfelder, Luka Cehovin Zajc, Tomas Vojir, Goutam Bhat, Alan Lukezic, Abdelrahman Eldesokey, et al. The sixth visual object tracking vot2018 challenge results. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018.

[9] Chao Ma, Jia-Bin Huang, Xiaokang Yang, and Ming-Hsuan Yang. Hierarchical convolutional features for visual tracking. In *Proceedings of the IEEE international conference on computer vision*, pages 3074–3082, 2015.

[10] Matthias Muller, Adel Bibi, Silvio Giancola, Salman Alsubaihi, and Bernard Ghanem. Trackingnet: A large-scale dataset and benchmark for object tracking in the wild. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 300–317, 2018.

[11] Hyeonseob Nam and Bohyung Han. Learning multi-domain convolutional neural networks for visual tracking. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4293–4302, 2016.

[12] Jack Valmadre, Luca Bertinetto, João Henriques, Andrea Vedaldi, and Philip HS Torr. End-to-end representation learning for correlation filter based tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2805–2813, 2017.

[13] Linyu Zheng, Ming Tang, Yingying Chen, Jinqiao Wang, and Hanqing Lu. Fast-deepkcf without boundary effect. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.