Fast DCT-based image convolution algorithms and application to image resampling and hologram reconstruction

Leonid Bilevich*[a] and Leonid Yaroslavsky**[a]
[a]Department of Physical Electronics, Faculty of Engineering,
Tel Aviv University, 69978, Tel Aviv, Israel[1]

ABSTRACT

Convolution and correlation are very basic image processing operations with numerous applications ranging from image restoration to target detection to image resampling and geometrical transformation. In real time applications, the crucial issue is the processing speed, which implies mandatory use of algorithms with the lowest possible computational complexity. Fast image convolution and correlation with large convolution kernels are traditionally carried out in the domain of Discrete Fourier Transform computed using Fast Fourier Transform algorithms. However standard DFT based convolution implements cyclic convolution rather than linear one and, because of this, suffers from heavy boundary effects. We introduce a fast DCT based convolution algorithm, which is virtually free of boundary effects of the cyclic convolution. We show that this algorithm have the same or even lower computational complexity as DFT-based algorithm and demonstrate its advantages in application examples of image arbitrary translation and scaling with perfect discrete sinc-interpolation and for image scaled reconstruction from holograms digitally recorded in near and far diffraction zones. In geometrical resampling the scaling by arbitrary factor is implemented using the DFT domain scaling algorithm and DCT-based convolution. In scaled hologram reconstruction in far diffraction zones the Fourier reconstruction method with simultaneous scaling is implemented using DCT-based convolution. In scaled hologram reconstruction in near diffraction zones the convolutional reconstruction algorithm is implemented by the DCT-based convolution.

1. INTRODUCTION

A linear (aperiodic) convolution

$$\tilde{a}_k = a_k * h_k \equiv \sum_{n=0}^{N-1} h_n a_{k-n} \quad \left(k = 0, \ldots, N-1\right) \tag{1}$$

is a fundamental basic operation in digital image processing. It is implemented by fast algorithms in transform domain, usually in DFT domain through FFT algorithms. However, these algorithms compute a cyclic (circular) convolution:

$$\tilde{a}_k = a_k * h_k \equiv \sum_{n=0}^{N-1} h_n a_{(k-n)\bmod N} \quad \left(k = 0, \ldots, N-1\right) \tag{2}$$

rather than a linear one.

Because in digital processing signals are always given by a finite number of their samples, any digital convolution algorithm requires one or another method for determination of signals and convolution kernels outside their boundaries.

In order to cope with this problem, several DFT domain digital convolution methods were suggested. In particular, in Ref. [1] zero padding of both the signal and the kernel is recommended and in Ref. [2] zero padding of the signal and mirror reflection of the kernel is suggested. However these methods, widely used in digital signal processing, suffer from heavy boundary effects caused by discontinuities in signals due to zero padding. A method of digital convolution that is virtually free from such boundary effects was suggested in [3-5]. According to this method, signals are extended to

double length by their mirror reflection and the kernel is zero-padded to double length. This results in the following convolution algorithm:

$$\tilde{a}_k = \sqrt{2N}\left[\text{IDCT}\left(\text{DCT}[a_k]\Re\{\text{DFT}[\text{SH}(\text{ZP}_{2N}(h_k))]\}\right) + \text{IDcST}\left(\text{DCT}[a_k]\Im\{\text{DFT}[\text{SH}(\text{ZP}_{2N}(h_k))]\}\right)\right], \quad (3)$$

where DFT, DCT, IDCT and IDcST are the Discrete Fourier Transform, Discrete Cosine Transform, Inverse Discrete Cosine Transform and Inverse Discrete Cosine-Sine Transform defined by:

$$\alpha_r^{(DFT)} = \text{DFT}(a_k) \equiv \frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} a_k \exp\left(i2\pi\frac{kr}{N}\right), \quad (4)$$

$$\alpha_r^{(DCT)} = \text{DCT}(a_k) \equiv \sqrt{\frac{2}{N}}\sum_{k=0}^{N-1} a_k \cos\left(\pi\frac{(k+1/2)r}{N}\right), \quad (5)$$

$$a_k = \text{IDCT}\left(\alpha_r^{(DCT)}\right) \equiv \frac{1}{\sqrt{2N}}\left\{\alpha_0^{(DCT)} + 2\sum_{r=1}^{N-1}\alpha_r^{(DCT)}\cos\left(\pi\frac{(k+1/2)r}{N}\right)\right\}, \quad (6)$$

$$a_k = \text{IDcST}\left(\alpha_r^{(DST)}\right) \equiv \frac{1}{\sqrt{2N}}\left\{(-1)^k \alpha_N^{(DST)} + 2\sum_{r=1}^{N-1}\alpha_r^{(DST)}\sin\left(\pi\frac{(k+1/2)r}{N}\right)\right\}; \quad (7)$$

$\text{ZP}_{2N}(h_k)$ denotes a zero-padding of $\{h_k\}$ to double length and SH denotes a shift by $\lceil N/2 \rceil$, where $\lceil \bullet \rceil$ denotes integer part of the argument:

$$\text{SH}[\text{ZP}_{2N}(h_k)] \equiv [\text{ZP}_{2N}(h_k)]_{(k-N/2)\bmod 2N}, \quad (8)$$

$\Re$ denotes a real part and $\Im$ denotes an imaginary part. The algorithm given by Eq. (3) was used to implement a p-fractional translation algorithm [6].

2. BOUNDARY EFFECT SAFE DIGITAL CONVOLUTION IN DCT DOMAIN

Algorithm given by Eq. (3) requires computing the Inverse Discrete Cosine-Sine Transform IDcST which is not commonly available. However IDcST can be easily computed via IDCT [7], and the Eq. (3) can be converted into the following form:

$$\tilde{a}_k = \sqrt{2N}\,\text{IDCT}\left(\text{DCT}[a_k]\Re\{\text{DFT}[\text{SH}(\text{ZP}_{2N}(h_k))]\}\right)$$
$$+ \sqrt{2N}(-1)^k\left\{\text{IDcST}\left([\text{DCT}[a_k]\Im\{\text{DFT}[\text{SH}(\text{ZP}_{2N}(h_k))]\}]_{N-r}\right)\right\}_k. \quad (9)$$

This equation assumes computing DFT of the zero-padded convolution kernel of double length.

Furthermore, this equation can be converted into the following form that requires computing over original N samples:

$$\tilde{a}_k = \sqrt{\frac{N}{2}}\left\{\text{IDCT}\left[\text{DCT}(a_k)\text{DCT}^{\mathbf{I}}(\hat{h}_k)\right] + \text{IDcST}\left[\text{DCT}(a_k)\text{DcST}^{\mathbf{I}}(\hat{h}_k)\right]\right\}$$
$$= \sqrt{\frac{N}{2}}\left[\text{IDCT}\left[\text{DCT}(a_k)\text{DCT}^{\mathbf{I}}(\hat{h}_k)\right] + (-1)^k\left\{\text{IDCT}\left[\left(\text{DCT}(a_k)\text{DcST}^{\mathbf{I}}(\hat{h}_k)\right)_{N-r}\right]\right\}_k\right], \quad (10)$$

where $\text{DCT}^{\mathbf{I}}$ and $\text{DcST}^{\mathbf{I}}$ are the type-$\mathbf{I}$ Discrete Cosine Transform and type-$\mathbf{I}$ Discrete Cosine-Sine Transform defined by:

$$\alpha_r^{(DCT-I)} = \text{DCT}^{\mathbf{I}}(a_k) \equiv \frac{1}{\sqrt{2N}} \left[a_0 + (-1)^r a_N + 2\sum_{k=1}^{N-1} a_k \cos\left(\pi\frac{kr}{N}\right) \right] \quad (k=0,\dots,N, \quad r=0,\dots,N), \quad (11)$$

$$\alpha_r^{(DST-I)} = \text{DcST}^{\mathbf{I}}(a_k) \equiv \sqrt{\frac{2}{N}} \sum_{k=1}^{N-1} a_k \sin\left(\pi\frac{kr}{N}\right) \quad (k=1,\dots,N-1, \quad r=1,\dots,N-1) \quad (12)$$

and $\hat{h}_k$ denotes a normalized kernel:

$$\hat{h}_k \equiv \begin{cases} 2h_0 & \text{if} & k=0 \\ h_k & \text{if} & k=1,\dots,N-1. \\ 0 & \text{if} & \text{k}=N \end{cases} \quad (13)$$

Since the algorithm Eq. (10) operates on signals of the original length (and not double one) and fast FFT-type algorithms are available for computation of different types of DCT/DcST transforms [8], the DCT convolution Eq. (10) can be computed in efficient way suited for real-time applications.

3. APPLICATIONS

The convolution algorithm Eq. (10) can be applied to various applications including resampling and hologram reconstruction. As representative applications, we'll describe a method of scaling and two methods of scaled holographic reconstruction.

3.1 Scaling

The input signal a_k has to be scaled by arbitrary scaling factor σ. In order to map the center of the input image into the center of the scaled image, we have to introduce the translational "centering factor" Δ:

$$\Delta \equiv (\lceil N\sigma \rceil - 1) - (N-1)\sigma \quad (14)$$

(the distance between the center of the input image and the center of the scaled image is equal to $\Delta/2$). The scaled signal is computed in the following way [2, 9]:

$$\tilde{a}_k^{(\sigma)} = \frac{1}{\sqrt{N}} \exp\left(-i\pi\frac{k^2-(N-1)(k-\Delta/2)}{\sigma N}\right)$$
$$\times \sum_{r=0}^{N-1} \left[\text{SDFT}^{(0,(N-1)/2)}(a_k)\exp\left(-i\pi\frac{r^2-\Delta r}{\sigma N}\right) \right] \exp\left(i\pi\frac{(k-r)^2}{\sigma N}\right), \quad (15)$$

where SDFT is the Shifted Discrete Fourier Transform [4] defined by:

$$\alpha_r^{(u,v)} = \text{SDFT}^{(u,v)}(a_k) \equiv \exp\left(-i2\pi\frac{uv}{N}\right)\frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} a_k \exp\left(i2\pi\frac{(k-u)(r-v)}{N}\right). \quad (16)$$

In the scaling method Eq. (15) the problem of computation of scaled signal boils down to computation of convolution that is implemented in DCT domain using Eq. (10). The peculiar choice of phase factors in the Eq. (15) ensures that for the real input image the computed scaled image is real and centered correctly. For the case of zoom-out ($\sigma < 1$) the method Eq. (15) does not require the low-pass pre-filtering of the input signal (because this pre-filtering is built-in in the DCT convolution Eq. (10). The interpolation formula of the scaling algorithm Eq. (15) is given by:

$$\tilde{a}_k^{(\sigma)} = \sum_{n=0}^{N-1} a_n \operatorname{sincd}\left(N; N; \frac{k - \Delta/2}{\sigma} - n\right), \tag{17}$$

where sincd is a discrete sinc function defined by:

$$\operatorname{sincd}(M; N; x) \equiv \frac{\sin\left(\pi \dfrac{Mx}{N}\right)}{N \sin\left(\pi \dfrac{x}{N}\right)} . \tag{18}$$

Examples of images obtained by the scaling method Eq. (15) with scaling factor $\sigma = \sqrt{2}$ implemented by the DFT convolution [2] and by the DCT convolution Eq. (10) are shown in Figure 1. As one can see, boundary effect in form of heavy oscillations at the image borders are present in the result of DFT convolution and are absent in the result of DCT convolution.

Figure 1. Comparison of image scaling algorithms. Left: the original image. Middle: the scaled image computed using DFT-domain convolution ($\sigma = \sqrt{2}$). Right: the scaled image computed using DCT-domain convolution ($\sigma = \sqrt{2}$).

3.2 Hologram reconstruction with scaling

In numerical reconstruction of holograms it is frequently required to reconstruct images with different scale factor commensurable with illumination wave length. The two most important methods of numerical reconstruction of Fresnel holograms are the Fourier reconstruction algorithm (for far diffraction zones) and the convolutional reconstruction algorithm (for near diffraction zones) [5, 10]. The suggested above digital convolution algorithm can naturally be employed for solving this task. For the Fourier reconstruction algorithm, the reconstruction and scaling can be performed simultaneously in one step while for the convolutional reconstruction algorithm the reconstruction and scaling have to be performed in two steps.

Fourier reconstruction algorithm with scaling

For Fourier reconstruction algorithm, σ-scaled reconstructed samples of the object wave front are computed from the input hologram samples $\{\alpha_r\}$ by the following formula [5, 10]:

$$\tilde{a}_k^{(\sigma)} = \frac{1}{\sqrt{N}} \exp\left(-i\pi \frac{\left[(k\mu/\sigma + w)^2 - k^2/\sigma\right]}{N}\right)$$

$$\times \sum_{r=0}^{N-1} \left[\alpha_r \exp\left(-i\pi \frac{(1/\mu^2 - 1/\sigma)r^2 - 2wr/\mu}{N}\right)\right] \exp\left(-i\pi \frac{(k-r)^2}{\sigma N}\right). \qquad (19)$$

Therefore the problem of computation of scaled hologram reconstruction boils down to the problem of computation of convolution that can be implemented in DCT domain using Eq. (10). Examples of Fourier hologram reconstruction algorithm without scaling ($\sigma = 1$) and with scaling ($\sigma = \sqrt{2}$) are shown in Figure 2.
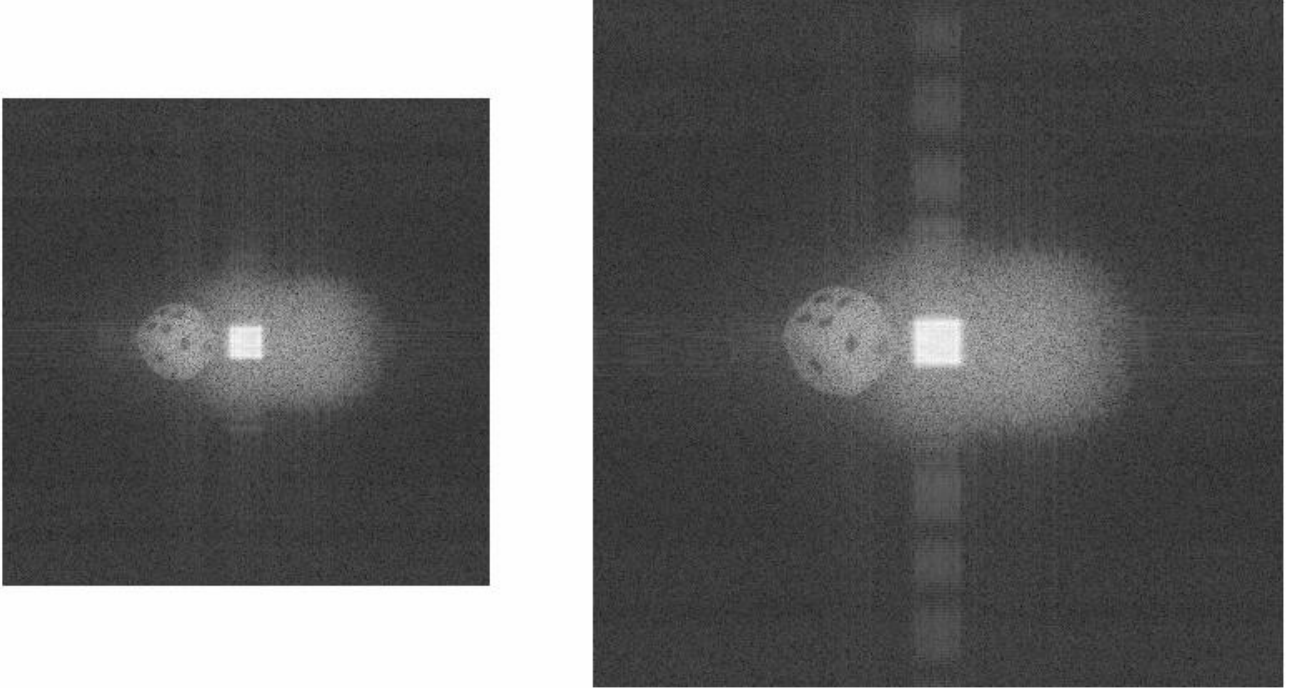

Figure 2. Fourier reconstruction algorithm with scaling. Left: hologram reconstruction without scaling ($\sigma = 1$). Right: scaled hologram reconstruction ($\sigma = \sqrt{2}$).

Convolutional reconstruction algorithm with scaling

For the convolutional algorithm, reconstructed samples of the object wave are computed from the input hologram samples $\{\alpha_r\}$ by the following formula [5, 10]:

$$a_k \equiv \sum_{r=0}^{N-1} \alpha_r \, \text{frincd}\left(N; \mu^2; k - r + w\right), \qquad (20)$$

where frincd is a discrete frinc function defined by:

$$\mathrm{frincd}(N;q;x) \equiv \frac{1}{N}\sum_{r=0}^{N-1}\exp\left(i\pi\frac{qr^2}{N}\right)\exp\left(-i2\pi\frac{xr}{N}\right). \tag{21}$$

In this case, the problem of hologram reconstruction boils down to the problem of computation of convolution that can be implemented in DCT domain using Eq. (10). The function frincd is computed from Eq. (21) using IDFT. The scaled hologram reconstruction $\tilde{a}_k^{(\sigma)}$ is computed from the hologram reconstruction a_k using scaling algorithm Eq. (15).

Examples of convolutional hologram reconstruction algorithm without scaling ($\sigma = 1$) and with scaling ($\sigma = \sqrt{2}$) are shown in Figure 3.
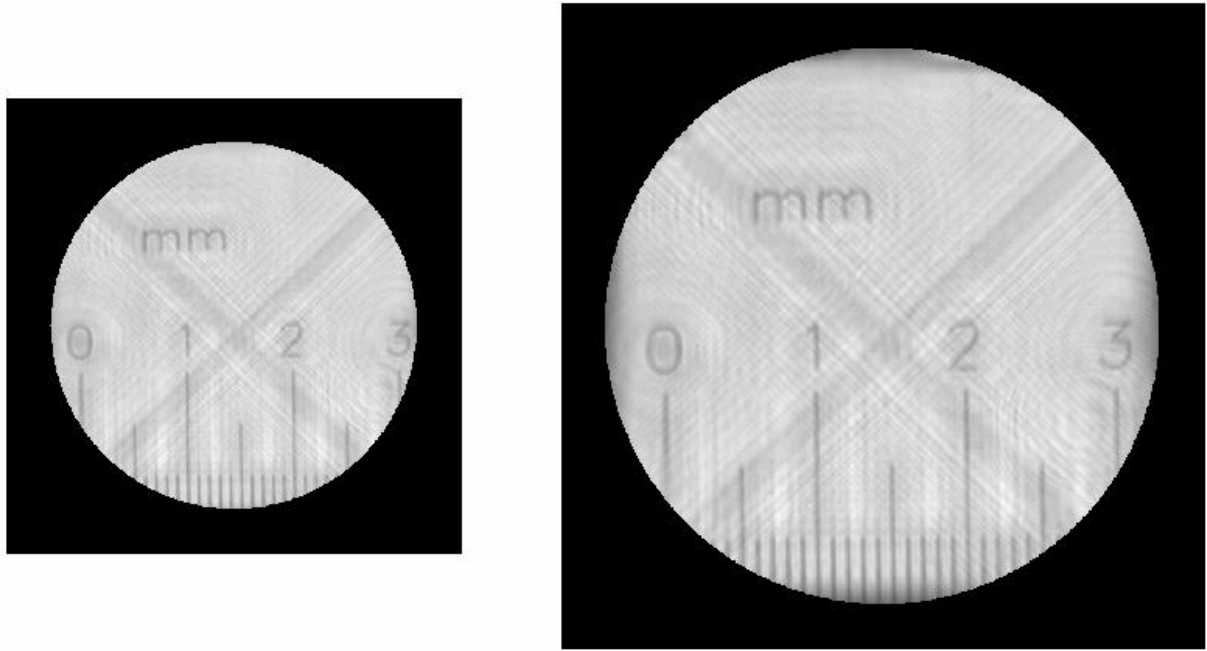

Figure 3. Convolution reconstruction algorithm with scaling. Left: hologram reconstruction without scaling ($\sigma = 1$). Right: scaled hologram reconstruction ($\sigma = \sqrt{2}$).

4. CONCLUSIONS

The boundary effect safe DCT-domain convolution algorithm is presented and applications of this algorithm to image re-scaling and scaled holographic reconstruction are provided. Thanks to the availability of fast FFT-type algorithms for computing transforms involved in the algorithm, the suggested DCT-domain convolution represents a valuable alternative to DFT-domain convolution in real-time video processing applications.

REFERENCES

[1]   Stockham, T. G., "High-speed convolution and correlation with applications to digital filtering," in [*Digital Processing of Signals*], Gold, B. and Rader, C. M., eds., 203–232, McGraw-Hill, Inc. (1969).
[2]   Rabiner, L. R., Schafer, R. W., and Rader, C. M., "The chirp *z*-transform algorithm and its application," *Bell Syst. Tech. J.* **48**(5), 1249–1292 (May-June 1969).

[3]   Yaroslavsky, L., "Boundary effect free and adaptive discrete signal sinc-interpolation algorithms for signal and image resampling," *Appl. Opt.* **42**(20), 4166–4175 (July 2003).

[4]   Yaroslavsky, L., [*Digital Holography and Digital Image Processing: Principles, Methods, Algorithms*], Kluwer Academic Publishers (2004).

[5]   Yaroslavsky, L., "Discrete transforms, fast algorithms, and point spread functions of numerical reconstruction of digitally recorded holograms," in [*Advances in Signal Transforms: Theory and Applications*], Astola, J. and Yaroslavsky, L., eds., *EURASIP Book Series on Signal Processing and Communications* **7**, 93–141, Hindawi (2007).

[6]   Yaroslavsky, L., "Fast discrete sinc-interpolation: A gold standard for image resampling," in [*Advances in Signal Transforms: Theory and Applications*], Astola, J. and Yaroslavsky, L., eds., *EURASIP Book Series on Signal Processing and Communications* **7**, 337–405, Hindawi (2007).

[7]   Wang, Z., "A fast algorithm for the discrete sine transform implemented by the fast cosine transform," *IEEE Trans. Acoust., Speech, Signal Processing* **30**(5), 814–815 (Oct. 1982).

[8]   Britanak, V., Yip, P., and Rao, K. R., [*Discrete Cosine and Sine Transforms: General Properties, Fast Algorithms and Integer Approximations*], Elsevier Ltd. (2007).

[9]   Deng, X., Bihari, B., Gan, J., Zhao, F., and Chen, R. T., "Fast algorithm for chirp transforms with zooming-in ability and its applications," *J. Opt. Soc. Am. A* **17**(4), 762–771 (Apr. 2000).

[10]   Yaroslavsky, L., "Introduction to digital holography," in [*Digital Signal Processing in Experimental Research*], Yaroslavsky, L. and Astola, J., eds., *Bentham E-book Series*, 1–187 (2009).