3-D MEAN-SEPARATION-TYPE SHORT-TIME DFT WITH ITS APPLICATION TO MOVING-IMAGE DENOISING

Takashi Komatsu, Ken Tyon, and Takahiro Saito

Dept. of Electrical, Electronics & Information Eng., Kanagawa University, Yokohama, Japan

ABSTRACT

Although for a still image the 2-D DFT and the 2-D DCT have similar properties to each other, for a moving-image sequence the 3-D DFT gets an advantage of representing the sequence more compactly over the 3-D DCT. Through the mathematical analysis of the 3-D DFT and the 3-D DCT based on a simple signal model of a moving-image sequence, this paper shows that the even symmetrization employed implicitly by the 3-D DCT may cause deterioration of representation efficiency and hence the 3-D DFT can achieve better representation efficiency than the 3-D DCT. In addition, to improve the suitability of the 3-D short-time DFT to processing of video signals which generally have significant local DC components carrying important structural information, this paper introduces a technique of local-mean-separation as a preprocess of the 3-D short-time DFT, thus to construct 3-D mean-separation-type ST-DFT; applies it to video denoising, and demonstrates its advantage over the existing 3-D transforms through experimental simulations of video denoising.

Index Terms— 3-D transform, short-time DFT, local-mean separation, phase-preserving-type shrinkage, video processing

1. INTRODUCTION

As a 3-D transform usable for processing of a color moving-image sequence, this paper takes up the 3-D DFT, and addresses the issue of enhancing its adaptability to processing of video signals, local DC components of which usually carry structural information relating directly to image contents.

The 2-D DCT has been regarded as the most useful tool for image processing, whereas up to the present 3-D transforms have not so widely utilized for video processing. In 1970's, the 3-D DCT was studied from the viewpoint of its application to video coding [1], [2], but it was considered as an impractical technique because of its necessity of a large amount of storage memories and computations; video coding with the 3-D DCT was excluded from candidate methods for the international video compression standarization at a relatively early stage. Since the 1980's, image and video coding methods with the 2-D DCT have been adopted as the international standarization methods of image and video compression, viz. JPEG, MPEG and so on; thus the 2D-DCT has been widely recognized as the most efficient standard 2-D transform for image signals, and its application to image processing has spread over the relevant fields. However, for the last two decades, with increasing high computational power and large storage memories, the 3-D DCT has been revitalized, and it has been judged to be a practical technique, which is useful especially for video compression [3] ~ [10]. Furthermore, recently the authors have experimentally revealed that 3-D redundant DCT (3-D RDCT), which is an extension of the 3-D DCT, performs well for video restoration [11] ~ [13].

Meanwhile, there has been not any remarkable extension in the utilization of the 2-D and/or 3-D DFT for image and video processing. The DFT is a classic standard tool to analyze signals, but the straightforward application of the DFT to signal processing does not necessarily achieve a high performance. If DFT coefficients of each subblock extracted from an input signal are processed for signal restoration, annoying artificial distortions referred to as the blocking artifacts occur along subblock boundaries on its restored signal. For a long time, the short-time DFT (ST-DFT) has been recognized as a practically effective technique to suppress the occurrence of the blocking artifacts [14] ~ [17]. In the ST-DFT, an input signal to be transformed is multiplied by a window function which is nonzero for only a narrow range in the signal domain and decays to zero near subblock boundaries, each windowed subblock is transformed with the DFT, and then the DFT coefficients are processed. In the field of speech and acoustic signal processing, the 1-D ST-DFT has been generally used as a handy efficient tool to date [14] ~ [16]. On the other hand, in the field of image processing, in 1980, J. S. Lim proposed a still-image denoising method with the 2-D ST-DFT, in which a window function such as the Hamming window was used [17], but there has been no further extension of the 2-D ST-DFT proposed by J. S. Lim to image and/or video processing.

For a still image the 2-D DFT and the 2-D DCT have similar properties to each other, whereas for a moving-image sequence the 3-D DFT and the 3-D DCT differ with each other in efficiency of representation. The even symmetrization employed implicitly by the 3-D DCT may cause deterioration of representation efficiency, and accordingly the 3-D DFT can provide a performance advantage over the 3-D DCT. By analyzing the 3-D DFT and the 3-D DCT theoretically based a simple signal model of a moving-image sequence, this paper shows that the 3-D DFT has superiority in representation efficiency to the 3-D DCT and is more suitable to video processing. Moreover, to enhance the adaptability of the 3-D ST-DFT to processing of video signals which generally have significant local DC components carrying important structural information about image contents, this paper introduces the technique of local-mean-separation into the 3-D ST-DFT, thus to construct 3-D mean-separation-type ST-DFT (3-D MS^{2}T-DFT) and gives a perfect reconstruction formula to compensate for the local-mean-separation on the occasion of the inverse transform. Furthermore, this paper applies the 3-D MS^{2}T-DFT to video denoising, and demonstrates that the video denoising method with the 3-D MS^{2}T-DFT outperforms the video denoising method with the 3-D RDCT [11] ~ [13], through experimental simulations of video denoising.

2. 3-D FOURIER TRANSFORM VERSUS 3-D COSINE TRANSFORM OF VIDEO SIGNALS

Still images are well characterized by their spatial statistical properties such as spatial correlations, whereas video signals containing moving objects have more complex signal structures caused by object motion. The spatial property of video signals is different from their temporal property.

The cosine transform, widely utilized for image and/or video processing, is equivalent to the transform, in which signals to be transformed are first converted into even-symmetric signals about the origin and then the symmetric signals undergo the Fourier transform; thus the cosine transform produces only real transform coefficients, which correspond to coefficients of the cosines basis functions. As for sparse representation of still images, there is little difference between the 2-D cosine transform and the 2-D Fourier transform. On the other hand, as for sparse representation of moving-image sequences containing moving objects and/or camera works, there is a significant difference between the 3-D cosine transform and the 3-D Fourier transform. If the 3-D Fourier transform is applied to video signals, phases of specific Fourier transform coefficients will be affected by motions of moving objects. On the other hand, if the 3-D cosine transform is applied to video signals, many significant cosine transform coefficients will occur, which is caused by objects' motions and the even-symmetrization employed implicitly by the DCT; therefore for video signals representation efficiency of the 3-D DCT will possibly degrade.

2.1. 3-D transforms for video signals with parallel translation

This section takes up a simple moving-image sequence whose image frames move at a constant velocity, as an example. To model this sequence mathematically, firstly let a 2-D image, $s(x, y)$, be defined as a causal image, as follows:

$$s(x,y)=0, \quad \text{for } x<0 \text{ or } y<0. \tag{1}$$

And, let the Fourier transform of the image, $s(x, y)$, be denoted by

$$S(\omega_x,\omega_y)=\mathbb{F}_{x,y}\left[s(x,y)\right]. \tag{2}$$

Next, with $s(x, y)$, let a causal moving-image frame sequence, $f(x, y, t)$, undergoing parallel displacement at a constant velocity (v_x, v_y) for $t \geq 0$, be defined as follows:

$$f(x,y,t):=s(x-v_xt,\,y-v_yt)u(t), \tag{3}$$

$u(t)$: Unit step function.

The 3-D Fourier transform of the sequence, $f(x, y, t)$, is given by

$$\begin{aligned}\mathbb{F}_{x,y,t}\left[f(x,y,t)\right]&=\pi S(\omega_x,\omega_y)\delta(\omega_t+\omega_xv_x+\omega_yv_y)\\&\quad-jS(\omega_x,\omega_y)/(\omega_t+\omega_xv_x+\omega_yv_y).\end{aligned} \tag{4}$$

The first term of the above equation corresponds to intrinsic frequency components caused by the parallel displacement of the frame sequence, and generates nonzero significant transform coefficients only on the 2-D plane, $\omega_t + \omega_x v_x + \omega_y v_y = 0$, passing through the origin of the 3-D spatiotemporal frequency domain. The second term means that the truncation by the multiplication of $u(t)$ in (3) causes other frequency components, and the magnitude of the frequency components, corresponding to the discontinuity at $t = 0$, decays in inverse proportional to the distance from the 2-D plane, $\omega_t + \omega_x v_x + \omega_y v_y = 0$.

By converting the sequence, $f(x, y, t)$, which has causality along each axis of the 3-D coordinates (x, y, t), into even-symmetric signals about the origin of the 3-D spatiotemporal signal domain, anew this section defines an even-symmetric moving-image frame sequence, $g(x, y, t)$, as follows:

$$\begin{aligned}g(x,y,t):&=f(x,y,t)+f(-x,y,t)+f(x,-y,t)\\&\quad+f(-x,-y,t)+f(x,y,-t)+f(-x,y,-t)\\&\quad+f(x,-y,-t)+f(-x,-y,-t).\end{aligned} \tag{5}$$

The 3-D Fourier transform of $g(x, y, t)$ is equivalent to the 3-D cosine transform of $f(x, y, t)$, and is given by

$$G_{x,y,t}(\omega_x,\omega_y,\omega_t)=G^{(1)}_{x,y,t}(\omega_x,\omega_y,\omega_t)+G^{(2)}_{x,y,t}(\omega_x,\omega_y,\omega_t), \tag{6}$$

$$\begin{aligned}G^{(1)}_{x,y,t}(\omega_x,\omega_y,\omega_t)&=\pi\Big[\left\{S(\omega_x,\omega_y)+\overline{S(\omega_x,\omega_y)}\right\}\\&\quad\times\left\{\delta(\omega_t+\omega_xv_x+\omega_yv_y)+\delta(\omega_t-\omega_xv_x-\omega_yv_y)\right\}\\&\quad+\left\{S(\omega_x,-\omega_y)+\overline{S(\omega_x,-\omega_y)}\right\}\\&\quad\times\left\{\delta(\omega_t+\omega_xv_x-\omega_yv_y)+\delta(\omega_t-\omega_xv_x+\omega_yv_y)\right\}\Big],\\G^{(2)}_{x,y,t}(\omega_x,\omega_y,\omega_t)&=-j\left\{S(\omega_x,\omega_y)-\overline{S(\omega_x,\omega_y)}\right\}\\&\quad\times\left\{\frac{1}{\omega_t+\omega_xv_x+\omega_yv_y}-\frac{1}{\omega_t-\omega_xv_x-\omega_yv_y}\right\}\\&\quad-j\left\{S(\omega_x,-\omega_y)-\overline{S(\omega_x,-\omega_y)}\right\}\\&\quad\times\left\{\frac{1}{\omega_t+\omega_xv_x-\omega_yv_y}-\frac{1}{\omega_t-\omega_xv_x+\omega_yv_y}\right\}.\end{aligned}$$

The first term of the above equation means intrinsic frequency components caused by the parallel displacement of the frame sequence, and generates nonzero significant transform coefficients on the four different 2-D planes, defined as follows:

$$\left\{\begin{array}{ll}\omega_t+\omega_xv_x+\omega_yv_y=0, & \omega_t-\omega_xv_x+\omega_yv_y=0\\ \omega_t+\omega_xv_x-\omega_yv_y=0, & \omega_t-\omega_xv_x-\omega_yv_y=0\end{array}\right\}. \tag{7}$$

All the 2-D planes are passing through the origin of the 3-D spatiotemporal frequency domain. In the first quadrant of the 3-D frequency space, the first term of (6) actually causes significant transform coefficients on the three 2-D planes except the first plane of (7).

The above analysis based on the simple moving-image sequence model reveals that for the sequence the 3-D Fourier transform does not increase the number of nonzero significant transform coefficients much in comparison with the still-image case. As for the 3-D cosine transform of the sequence the number of nonzero significant transform coefficients clearly increases in comparison with the still-image case. Therefore, the analysis suggests that the 3-D Fourier transform will possibly provide more efficient representation than the 3-D cosine transform in that the 3-D Fourier transform produces a smaller number of nonzero significant transform coefficients than the 3-D cosine transform.

3. 3-D MEAN-SEPARATION-TYPE SHORT-TIME DFT FOR VIDEO PROCESSING

As a practical method of the 3-D DFT usable for video processing, this paper presents 3-D mean-separation-type ST-DFT (3-D MS^2T-DFT), in which after separating a local mean from each 3-D subblock to be transformed, the 3-D ST-DFT with a proper

slidable window function is applied to the local-mean-separated 3-D subblock.

Oscillatory signals such as speech and acoustic signals do not have any meaningful DC components, whereas image and video signals generally have significant local DC components carrying important structural information about image contents, which are meaningful to human observers. If the ST-DFT is directly applied to image signals, the multiplication of their local DC components by a window function decaying to zero near subblock boundaries must cause false AC components, whose bad effect spreads over a relatively wide range in the frequency domain, and must hamper discrimination between significant signal components and nuisance noise components contaminating the original image signals. To alleviate this bad effect, a local mean within each subblock should be subtracted from input image signals in advance of the application of the ST-DFT. On the other hand, to restore an image from processed ST-DFT coefficients, it is indispensable to compensate for the local-mean separation.

In the following, the introduction of the local-mean separation into the ST-DFT and the reconstruction formula satisfying a perfect reconstruction condition are presented, but for ease of understanding, the following descriptions are limited to the 1-D case, that is to say the 1-D MS^{2}T-DFT.

Firstly, let a slidable window function, which is used for extracting a data subblock within a range of $0 \le n < N$ from a 1-D input signal $f(n)$, be denoted by $w(n)$. Moreover, this section assumes that the ST-DFT multiplies the 1-D signal, $f(n)$, by the window function displaced for every L samples and it applies the N-point DFT to each segmented data subblock of length N. The displaced windowed function is represented by

$$w_i(n) := w(n - iL), \quad \text{for } i \in \mathbb{Z}, \tag{8}$$

where the subscript 'i' in $w_i(n)$ means shifting $w(n)$ to the location of $i \cdot L$. A subblock segmented by the displaced window function, $w_i(n)$, corresponds to a closed interval, $[i \cdot L, i \cdot L+N-1]$, of length N. Within the i-th subblock, its local-mean value, $\mu_{[i]}$, is computed by

$$\mu_{[i]} = \frac{1}{N} \sum_{n=iL}^{iL+N-1} f(n), \tag{9}$$

where the subscript '$[i]$' in $\mu_{[i]}$ means computing its relevant value within the i-th subblock. Then, the local mean value, $\mu_{[i]}$, is subtracted from the input signals within the i-th subblock, and the local-mean-separated signal is multiplied by the displaced window function, $w_i(n)$, as follows:

$$h_{[i]}(n) = w_i(n)\left(f(n) - \mu_{[i]}\right). \tag{10}$$

Lastly, the N-point DFT is applied to the local-mean-separated and windowed signal, $h_{[i]}(n)$.

In the mean-separation-type ST-DFT (MS^{2}T-DFT), if the window function, $w(n)$, satisfies the following condition:

$$\sum_{i=-\infty}^{\infty} w_i(n) = \sum_{i=-\infty}^{\infty} w(n - iL) = \mathrm{C}, \tag{11}$$

C: Constant irrespective of n,

then the 1-D original signal $f(n)$ can be perfectly reconstructed from the MS^{2}T-DFT coefficients. The condition of (11) is referred to as the perfect reconstruction condition. The formula of the perfect reconstruction is given by

$$f(n) = \frac{1}{C} \sum_{i=-\infty}^{\infty} h_{[i]}(n) + \frac{1}{C} \sum_{i=-\infty}^{\infty} w_i(n)\, \mu_{[i]}. \tag{12}$$

The second term of the above perfect-reconstruction formula means making compensation for the local-mean-separation at the analysis stage.

4. APPLICATION TO COLOR MOVING-IMAGE DENOISING

As video processing with the 3-D MS^{2}T-DFT, this paper takes up color moving-image denoising, and evaluates efficiency of the 3-D MS^{2}T-DFT by conducting denoising simulations on test color moving image sequences. For the denoising simulations, this paper uses the ITE standard color moving-image sequences, the image contents of which are summarized in Table 1, and certain image frames of which are shown in Fig. 1.

The specifications of the color moving-image denoising method with 3-D MS^{2}T-DFT are as follows.

1) The subblock size in the 3-D MS^{2}T-DFT is set to $16 \times 16 \times 16$.

2) As a 3-D window function, the 3-D MS^{2}T-DFT employs the separable Butterworth-type window function, $w(l, m, n)$, given by

$$w(l, m, n) := \rho(l)\rho(m)\rho(n), \tag{13}$$

$$\rho(k) := \begin{cases} \dfrac{1}{1 + (|k - 8| / 4)^6}, & 0 \le k < 16, \\ 0, & \text{otherwise}, \end{cases}$$

whose half-value width is $8 \times 8 \times 8$. This 3-D window function, $w(l, m, n)$, decays very rapidly near subblock boundaries.

3) The skip parameter, L, in (8) is set to 1, that is to say the 3-D window function, $w(l, m, n)$, is displaced for every sample along each axis of the 3-D coordinates, (x, y, t).

4) Complex 3-D MS^{2}T-DFT coefficients of the three-primary color channels are processed by the phase-preserving-type hard color-shrinkage (PP-HCS) with the redundant color-transform [18]. The PP-HCS utilizes interdependence among the three-primary color channels to suppress color artifacts produced by the shrinkage. Firstly, the PP-HCS maps three complex 3-D MS^{2}T-DFT coefficients, r, g, and b, of the three-primary color channels with the redundant color-transform:

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} r \\ g \\ b \end{pmatrix}, \tag{14}$$

applies the hard shrinkage to the magnitude of each complex coefficient, p_i (i = 1, 2, 3, 4), thus to yield its shrunken magnitude λ_i [19], and then produces its shrunken complex coefficient, q_i (i = 1, 2, 3, 4), as follows:

$$q_i = \lambda_i \frac{p_i}{|p_i|}, \quad \text{for } i = 1, 2, 3, 4. \tag{15}$$

The phase of each complex coefficient, p_i, is preserved as it is in its shrunken complex coefficient, q_i. Lastly, the PP-HCS applies to the shrunken complex coefficients, q_1, q_2, q_3, and q_4, the least-squares-type generalized inverse of the redundant color-transform:

$$\begin{pmatrix} \hat{r} \\ \hat{g} \\ \hat{b} \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 & 1 & 0 & -1 \\ 1 & -1 & 1 & 0 \\ 1 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix}. \tag{16}$$

In the redundant color-transform of (14), the statistical property of p_1 is different from the statistical properties of the other components, p_2, p_3, and p_4. Accordingly this paper sets the threshold parameter, η_+, of the hard shrinkage for p_1 and the

threshold parameter, η_{-}, for p_2, p_3, and p_4, separately.

In the denoising simulations, noisy test color moving-image sequences, each of which is composed of 96 image frames, are generated by adding white Gaussian noise with zero mean and variance, σ^2, to original color moving-image sequences, pixel values of which are expressed in the range, [0, 1]. In the denoising simulations, the variance, σ^2, of the additive noise is set to 0.012, and the PSNR*'s* of the noisy test sequences are about 19.2 [dB]. This paper compares denoising performance of the following five different color moving-image denoising methods:

1) Our denoising method with the 3-D MS^{2}T-DFT,

2) Denoising method with the 3-D ST-DFT of the 16 × 16 × 16 subblock size, in which the local-mean-separation is not adopted and uses the Butterworth-type window function,

3) Denoising method with the 3-D redundant DFT (3-D RDFT) of the 8 × 8 × 8 subblock size, which does not employ the local-mean-separation and is equivalent to the 3-D ST-DFT with the boxcar-type window function,

4) Our previously proposed denoising method with the 3-D redundant DCT (3-D RDCT) of the 8 × 8 × 8 subblock size [11] ~ [13], in which DCT coefficients are processed by the hard color-shrinkage [18], and

5) CVBM3D denoising method [20], [21], which is one of the state-of-the-art color moving-image denoising methods.

Table 2 shows PSNR*'s* of denoised color moving-image sequences. Figure 2 shows PSNR*'s* of every frame in denoised image sequences, for the test sequence, "Walk through the Square". As shown in Fig. 2, at the beginning and the last parts of the sequence, all the denoising methods but the CVBM3D denoising method deteriorate in PSNR considerably, which is caused by the cyclic shift employed by the denoising simulations for the temporal subblock partition, but in the case of an actual long sequence it hardly matters. Taking account of this point, Table 2 shows PSNR*'s* not only over all frames of each sequence but also over 82 frames in the middle part of each sequence.

Table 1. Image contents of the ITE standard color moving-image sequences with 720 × 480 pixels.

Sequence number	Sequence title	Object motion	Camera work
1	Woman with Bird Cage	Moving woman with a bird cage	Slow panning
2	Whale Show	Jumping whale	Panning
3	Walk through the Square	Waking girl	Horizontal dolly
4	Harbor Scene	Walking crowds	Slow panning

(a) Woman with Bird Cage

(b) Whale Show

(c) Walk through the Square

(d) Harbor Scene

Fig. 1. Certain image frames of the standard ITE color moving-image sequences.

Table 2. PSNR*'s* of denoised color moving-image sequences in the case of $\sigma^2 = 0.012$.

Denoising methods	Sequence number			
	1	2	3	4
	all frames	all frames	all frames	all frames
	82 frames	82 frames	82 frames	82 frames
3-D MS^{2}T-DFT	34.018	**30.189**	**31.786**	**31.979**
	34.279	**30.247**	**32.011**	**32.251**
3-D ST-DFT	33.328	29.835	31.335	31.562
	33.509	29.879	31.518	31.785
3-D RDFT	33.125	28.992	30.000	30.027
	33.238	28.991	30.093	30.112
3-D RDCT	32.723	29.108	30.184	30.204
	32.822	29.091	30.281	30.306
CVBM3D	**34.279**	29.968	31.451	31.548
	34.278	29.890	31.469	31.567

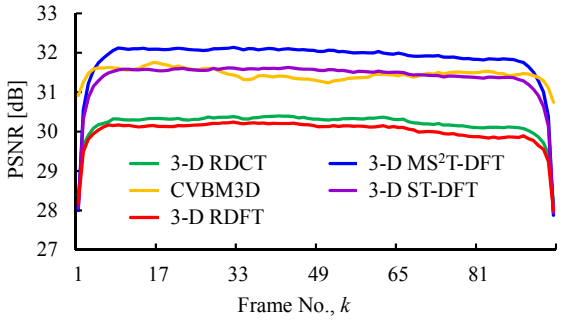

Fig. 2. PSNR*'s* of every frame in denoised color moving-image sequences, for the test sequence, "Walk through the Square".

In Table 2, the 3-D RDFT method does not necessarily gain a performance advantage over the 3-D RDCT method, which does not accord with the analysis results described in Section 2, and this unsatisfying behavior of the 3-D RDFT method is owing to discontinuities of subblock boundaries. The introduction of the 3-D window function into the 3-D RDFT method achieves an improvement of 0.2 ~ 1.6 dB, and the introduction of the local-mean-separation into the 3-D ST-DFT method further improves denoising performance by 0.3 ~ 0.7 dB. The 3-D MS^{2}T-DFT method is superior to the 3-D RDCT method by 1.0 ~ 1.9 dB, and generally outperforms the CVBM3D method in denoising performance.

5. CONCLUSIONS

Through the mathematical analysis based on a simple signal model of a moving-image sequence, this paper shows that the representation efficiency of the 3-D DFT is better than the 3-D DCT. To enhance the adaptability of the 3-D short-time DFT to video processing, this paper introduces the local-mean-separation into the 3-D short-time DFT, applies the 3-D mean-separation-type short-time DFT to denoising of a color moving-image sequence contaminated by additive white Gaussian noise, and demonstrates its superiority to the 3-D redundant DCT, through experimental simulations of video denoising.

6. REFERENCES

[1] T. Natarajan and N. Ahmed, "On interframe transform coding," *IEEE Trans. Commun.*, vol.25, no.11, pp.1323-1329, Nov. 1977.

[2] J. A. Roese, W. K. Pratt, and G. S. Robinson, "Interframe cosine transform image coding," *IEEE Trans. Commun.*, vol.25, no.11, pp.1329-1339, Nov. 1977.

[3] B. Bauer and K. Sayood, "Video coding using 3 dimensional DCT and dynamic code selection," *Proc. Data Compression Conf.*, p.415, March 1995.

[4] R. Westwater and B. Furht, "Three-dimensional DCT video compression technique based on adaptive quantizers," *Proc. 2nd IEEE Int. Conf. on Engineering of Complex Computer Syst. (ICECCS '96)*, pp.189-198, Oct. 1996.

[5] M. Servais and G. de Jager, "Video compression using the three dimensional discrete cosine transform (3D-DCT)," *Proc. 1997 South African Symp. on Commun. & Signal Process. (COMSIG '97)*, pp.27-32, Sept. 1997.

[6] R. K. W. Chan and M. C. Lee, "3D-DCT quantization as a compression technique for video sequences," *Proc. Int. Conf. on Virtual Sys. & Multimedia*, pp.188-196, Sept. 1997.

[7] S. Boussakta and H. O. Alshibami, "Fast Algorithm for the 3-D DCT-II," *IEEE Trans. Signal Process.*, vol.52, no.4, pp.992-1001, April 2004.

[8] N. Božinović and J. Konrad, "Motion analysis in 3D DCT domain and its application to video coding," *Signal Process.: Image Commun.*, vol.20, no.6, pp.510-528, July 2005.

[9] M. C. Lee, R. K. W. Chan, and D. A. Adjeroh, "Fast three-dimensional discrete cosine transform," *SIAM J. Sci. Comput.*, vol.30, no.6, pp.3087–3107, Oct. 2008.

[10] M. Bhaskaranand and J. D. Gibson, "Distributions of 3D DCT coefficients for video," *Proc. 2009 IEEE Int. Conf. on Acoust., Speech & Signal Process. (ICASSP 2009)*, pp.793-796, April 2009.

[11] S. Kondou, T. Komatsu, and T. Saito, "A note on a video denoising method with the redundant 3D-DCT," *Proc. 2014 IEICE Gen. Conf.*, D-11-53, March 2014.

[12] T. Komatsu, S. Kondou, and T. Saito, "Restoration of a Poissonian-Gaussian color moving-image sequence with virtual multiplex imaging and super-resolution deblurring," *Proc. of 2016 IEEE Int. Conf. on Image Process. (ICIP 2016)*, pp.1963-1967, Sept. 2016.

[13] T. Komatsu, S. Kondou, and T. Saito, "3D redundant DCT restoration method for MPEG-compressed video," *Proc. IEEE 2016 Region 10 Conf. (TENCON 2016)*, pp.3075-3078, Nov. 2016.

[14] A. V. Oppenheim, "Speech spectrograms using the fast Fourier transform," *IEEE Spectrum*, vol.7, no.8, pp.57-62, Aug. 1970.

[15] S. H. Nowab and T. F. Quantieri, "Chap. 6: Short-time Fourier transform," in *Advanced Topics in Signal Processing*, J. S. Lim and A. V. Oppenheim, eds., Prentice Hall, Englewood Cliffs, NJ, USA, Oct. 1987.

[16] T. F. Quantieri, "Chap. 7: Short-time Fourier transform analysis and synthesis," in *Discrete-Time Speech Signal Processing*, Prentice Hall Inc., Englewood Cliffs, NJ, USA, 2002.

[17] J. S. Lim, "Image restoration by short space spectral subtraction", *IEEE Trans. Acoust., Speech, & Signal Process.*, vol.ASSP-28, no.2, pp.191-197, April 1980.

[18] T. Saito, Y. Ueda, and T. Komatsu, "Color shrinkage for sparse coding of color images," *Proc. 18th European Signal Process. Conf. (EUSIPCO 2010)*, pp.1023-1027, Aug. 2010.

[19] P. Kovesi, "Phase preserving denoising of images," *Proc. Fifth International/National Biennial Conference on Digital Image Computing, Techniques & Applications* (*DICTA 99*), pp.212-217, Perth, Australia, Dec. 1999.

[20] K. Dabov, A. Foi, and K. Egiazarian, "Image restoration by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol.16, no.8, pp.2080-2095, Aug. 2007.

[21] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3D transform-domain collaborative filtering," *Proc. 15th European Signal Process. Conf. (EUSIPCO 2007)*, pp.145-149, Poznań, Poland, Sept. 2007.