A new transform for document image compression

Aye Aung; Ng, Boon Poh; Shwe, Cherry Tin

2009

Aye, A., Ng, B. P., & Shwe, C. T. (2009). A New Transform For Document Image Compression. 2009 7th IEEE International Conference on Information, Communications and Signal Processing.

<https://hdl.handle.net/10356/94133>

<https://doi.org/10.1109/ICICS.2009.5397739>

© 2009 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at:
<http://dx.doi.org/10.1109/ICICS.2009.5397739>.

Downloaded on 26 Apr 2023 21:55:16 SGT

A New Transform For Document Image Compression

Aye Aung

School of Electrical Engineering and Computer Science
The University of Newcastle, Callaghan, Australia
Email: aye.aung@newcastle.edu.au

Boon Poh Ng, Cherry Tin Shwe

School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore
Emails: ebpng@ntu.edu.sg, cher0015@ntu.edu.sg

Abstract—In this paper, we propose a new real transform (rCS-SCHT) derived from the conjugate symmetric sequency-ordered complex Hadamard transform (CS-SCHT) for document image compression. Like the CS-SCHT, the proposed transform is also an orthogonal transform. The transform computation is very simple and it requires only addition/subtractions with the use of a fast algorithm. This transform possesses good energy compaction property which makes it suitable to be employed in document image compression. Simulations are conducted in order to demonstrate the energy compaction capability of the transform as compared to the discrete cosine transform (DCT) and the Walsh Hadamard transform (WHT), particularly for scanned document images. Further discussions on the bit rate vs peak signal-to-noise (PSNR) performances of the proposed transform, the DCT and the WHT are presented using the default JPEG setting for simulations.

I. INTRODUCTION

In 21st century, the world is digitally connected via the World-Wide Web (WWW) which has been used as a platform for the world's universal libraries. Most of the world's major libraries are converting their collections into digital forms due to the fact that it is less costly to keep the documents digitally than to provide for the buildings and shelves to house them in paper form. Besides, storing the documents in digital forms is very handy for people to carry around with them for easy access even outside offices or homes. In some countries, online submission of personal particulars has been given a priority consideration due to fast communication and easy management of scanned digital documents especially for visa application. However, the distribution of these digital documents over the web or saving them in hard disks or portable storage devices demands the sizes of those scanned copies to be as small as possible while maintaining at an acceptable level of readability.

Document image compression provides an attractive option to store large amounts of data efficiently. Documents can be scanned and compressed in order to reduce redundancies using various methods [1]–[6]. Transform coding is one of the commonly used methods to compress the scanned document images. JPEG [1] is a popular method which uses the discrete cosine transform (DCT) to obtain the compact representations of 2-dimensional (2-D) images. The DCT is a well-known orthogonal transform in image compression since it possesses very good energy compaction property [7]. After the transform has been performed on the images, suitable techniques such as Shannon-Fano coding, Huffman coding, and run-length coding

are used in order to obtain the efficient encoded data [4]. The JPEG encoding is widely used for both color and grey-scaled images. In JPEG, the DCT is performed on the 8×8 image blocks in order to obtain the transform coefficients which are then quantized and coded. But, the computational complexity of the DCT computation is very high, hence, which may not be preferred in low-end applications and/or real-time applications. The Walsh Hadamard transform (WHT) [1], [8] is another suitable option which compromises between the computational complexity and the energy packing capability.

In this paper, we propose a new transform which is derived from the recently introduced conjugate symmetric sequency-ordered complex Hadamard transform (CS-SCHT) [9]. This transform exhibits relatively good energy compaction property and it is promising to be employed especially for document image compression with very low computational complexity. The organization of the paper is as follows. Section II describes the derivation of the proposed transform from the CS-SCHT and its definition. The energy compaction capability of the proposed transform for grey-scaled document images is illustrated in Section III. The simulation results are presented and discussed in Section IV. Finally, this paper is concluded in Section V.

II. DEFINITION OF THE TRANSFORM

In this section, we shall briefly mention about the proposed transform which is derived from the CS-SCHT. The CS-SCHT is a complex orthogonal transform which consists of the elements $\{\pm 1, \pm j\}$ and whose row vectors are arranged in an ascending order of sequencies. Sequency is analogous to frequency in the discrete Fourier transform (DFT). Sequency ordering of complex Hadamard transform is shown to be useful in spectrum estimation and image signal processing [9]. But the CS-SCHT coefficients are the complex numbers comprised of real and imaginary terms. Since an image can be represented as a matrix consisting of real numbers, real-valued transforms are advisable to be used to transform the image in order to get real-valued coefficients. A new real transform matrix, which is

denoted as $\mathbf{C}_N$, can be derived from the CS-SCHT as follows:

$$\mathbf{C}_N = \begin{bmatrix} \mathbf{H}_N(0, k) \\ \frac{1}{2}(\Im\{\mathbf{H}_N(1, k)\} - \Im\{\mathbf{H}_N(N-1, k)\}) \\ \frac{1}{2}(\Re\{\mathbf{H}_N(1, k)\} + \Re\{\mathbf{H}_N(N-1, k)\}) \\ \vdots \\ \frac{1}{2}(\Im\{\mathbf{H}_N(\frac{N}{2}-1, k)\} - \Im\{\mathbf{H}_N(\frac{N}{2}+1, k)\}) \\ \frac{1}{2}(\Re\{\mathbf{H}_N(\frac{N}{2}-1, k)\} + \Re\{\mathbf{H}_N(\frac{N}{2}+1, k)\}) \\ \mathbf{H}_N(\frac{N}{2}, k) \end{bmatrix} \quad (1)$$

where $\mathbf{H}_N$ is the CS-SCHT matrix of dimension $N \times N$. Since we consider the 8×8 image blocks for transformation, $\mathbf{C}_8(N=8)$ is derived particularly as follows:

$$\mathbf{C}_8 = \begin{bmatrix} \mathbf{H}_8(0, k) \\ \frac{1}{2}(\Im\{\mathbf{H}_8(1, k)\} - \Im\{\mathbf{H}_8(7, k)\}) \\ \frac{1}{2}(\Re\{\mathbf{H}_8(1, k)\} + \Re\{\mathbf{H}_8(7, k)\}) \\ \frac{1}{2}(\Im\{\mathbf{H}_8(2, k)\} - \Im\{\mathbf{H}_8(6, k)\}) \\ \frac{1}{2}(\Re\{\mathbf{H}_8(2, k)\} + \Re\{\mathbf{H}_8(6, k)\}) \\ \frac{1}{2}(\Im\{\mathbf{H}_8(3, k)\} - \Im\{\mathbf{H}_8(5, k)\}) \\ \frac{1}{2}(\Re\{\mathbf{H}_8(3, k)\} + \Re\{\mathbf{H}_8(5, k)\}) \\ \mathbf{H}_8(4, k) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 & -1 \\ 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix} \quad (2)$$

which is well suited to be employed in document image compression. Since, the CS-SCHT is a complex orthogonal transform, its real-valued transform is also an orthogonal transform.

Definition: The input signal can be represented as a linear combination of the row vectors of $\mathbf{C}_8$, and the transform is defined as follows:

$$\mathbf{X}_8 = \mathbf{D}_8 \mathbf{C}_8 \mathbf{x}_8 \quad (3)$$

where $\mathbf{D}_8 = \text{diag}\left\{\sqrt{\frac{1}{8}}, \sqrt{\frac{1}{4}}, \sqrt{\frac{1}{4}}, \sqrt{\frac{1}{4}}, \sqrt{\frac{1}{4}}, \sqrt{\frac{1}{4}}, \sqrt{\frac{1}{4}}, \sqrt{\frac{1}{8}}\right\}$ is the diagonal matrix which consists of the normalization factors, $\mathbf{X}_8 = [X(0), X(1), \dots, X(7)]^T$ is the transformed column vector and $\mathbf{x}_8 = [x(0), x(1), \dots, x(7)]^T$ represents the input data vector. The data sequence can be recovered uniquely from the inverse transform, that is,

$$\mathbf{x}_8 = \mathbf{C}_8^T \mathbf{D}_8^T \mathbf{X}_8 \quad (4)$$

since it is an orthogonal transform. The 2-D transform for an 8×8 real image matrix can be obtained by applying 1-D transform on the rows followed by the columns. The resultant transform coefficients are real numbers, which are suitable for document image compression.

The transformation by $\mathbf{C}_8$ can be done efficiently with minimum computational cost by using the fast algorithm. The

Fig. 1: Signal flow graph of the fast forward algorithm for the real transform derived from the CS-SCHT ($N=8$).

fast algorithm to compute this transform is derived for the purpose of implementation for image compression. The signal flow graph for fast forward 8-point transform is illustrated in Fig. 1. For simplicity, the normalization factors are omitted in the figure, which in fact can be computed after or before the transformation. It can be seen from the figure that only the first stage needs eight addition/subtractions whereas the second and third stages merely require four and six addition/subtractions, respectively. Therefore, a total of $(8 + 4 + 6) = 18$ addition/subtractions is needed to compute an 8-point 1-D transform. On the other hand, 24 addition/subtractions are required to compute an 8-point traditional fast WHT using the fast algorithm [10]. The same number of multiplications (which is eight) is required for both transforms when the signal is scaled by the normalization factors, which is $\mathbf{D}_8$ mentioned in (3). Hence, an additional savings of $(24 - 18) = 6$ addition/subtractions (which is $\frac{6}{24} \times 100\% = 25\%$) can be gained for one transformation by using the proposed transform as compared to the fast WHT.

III. ENERGY COMPACTION CAPABILITY

Most orthogonal transforms tend to pack a large portion of the average energy of the images into a relatively few transform coefficients. Some of the well-known orthogonal transforms are the DFT, the discrete Wavelet transform (DWT), the Karhunen-Loeve transform (KLT), the DCT and the WHT. The fundamental goal of image compression is to reduce the bit rate (bits/pixel) which implies less transform coefficients to efficiently reconstruct the encoded images. Therefore, the transforms having high energy packing capabilities are desirable in document image compression. It is known that the KLT is an optimal transform in an information packing sense [7], but it is data dependent and obtaining the KLT coefficients for each image is a nontrivial costly computational task. The DWT is popular for its multiresolutional analysis for image signal processing [1]. As the number of decomposition levels increases, the information packing ability does as well, but the number of operations in the computations of the forward and inverse transforms becomes increased. The DCT is known

to be the most popular transform for its highly information packing capability as well as the availability of various efficient fast computational algorithms. The WHT is the simplest non-sinusoidal orthogonal transform which provides comparable performance in image compression as compared to the DCT. In this section, the computer simulation is conducted in order to evaluate the mean square error (MSE) performance of the proposed transform for energy compaction property.

In this paper, a document of A4 size (extracted from one of the conference proceedings) is scanned at 300 dpi (dots per inch) using the HP Deskjet F2120 scanner, and the resultant 8-bit grey-scaled scan of (2480×3504) pixels in bitmap (bmp) format is considered for simulation, which is shown in Fig. 2. The text and figure regions are extracted from the resultant scanned document image for simulation purpose since these kinds of portions are the most frequently found in most of the text documents. They are shown in Fig. 3 and Fig. 4, respectively. The size of the image blocks used in the simulation is 8×8 . Firstly, the image is divided into a number of 8×8 non-overlapping blocks and the respective transforms are performed on these blocks. The resultant transform coefficients are rounded before reconstructing the image. The coefficients with largest amplitudes are selected based on the compression ratio on each 8×8 transformed image block for decompressing the image. Fig. 3 and Fig. 4 show the original images and the corresponding reconstructed images using 8 transform coefficients for the proposed transform (rCS-SCHT),

Fig. 2: An example of scanned document image.

Fig. 3: The extracted text portion of (360×376) pixels: (a) The original, the decompressed images using (b) the proposed transform, (c) the DCT, and (d) the WHT for 8 transform coefficients.

Fig. 4: The extracted figure portion of (512×544) pixels: (a) The original, the decompressed images using (b) the proposed transform, (c) the DCT, and (d) the WHT for 8 transform coefficients.

the DCT and the WHT, respectively. The corresponding peak signal-to-noise ratios (PSNRs) are listed in Table I. It is found that PSNR of the DCT reconstructed image is the highest for Fig. 3 whereas the proposed transform provides higher PSNR value than the others for Fig. 4.

Figs. 5 and 6 show the mean square error (MSE) comparisons among the WHT, the proposed transform and the

TABLE I: Comparison of PSNRs (dB) for the reconstructed images using the introduced transform, the DCT and the WHT using 8 transform coefficients.

Type	rCS-SCHT	DCT	WHT
Text portion (Fig. 3)	20.77 dB	22.70 dB	20.52 dB
Figure portion (Fig. 4)	25.01 dB	24.74 dB	24.72 dB

Fig. 5: The MSE comparison among the transforms for the text portion shown in Fig. 3.

Fig. 6: The MSE comparison among the transforms for the figure portion shown in Fig. 4.

DCT with respect to the number of selected coefficients for the extracted document images shown in Fig. 3 and Fig. 4, respectively. As shown in Fig. 5, the DCT outperforms the other transforms for Fig. 3 which is the text portion of the scanned document. On the other hand, it can be seen from Fig. 6 that the MSE performance of the proposed transform is better than that of the others for any particular number of chosen coefficients for the figure region of the scanned document (shown in Fig. 4). It can be concluded that the proposed transform is expected to outperform the DCT and the WHT especially for the scanned document images which contain the contrasting texture (i.e., high frequency components as in Fig. 4).

As such the proposed transform possesses good energy compaction property like the DCT and the WHT, and it is

TABLE II: Comparisons of BR and PSNR among the DCT, the proposed transform, and the WHT for the text region shown in Fig. 3.

	rCS-SCHT		DCT		WHT	
M	BR	PSNR	BR	PSNR	BR	PSNR
8	0.56	16.73	0.61	18.77	0.59	17.09
12	0.75	18.07	0.79	21.12	0.77	18.58
16	0.90	19.98	0.93	23.79	0.92	20.33
24	1.20	20.02	1.21	27.09	1.21	22.33
32	1.45	25.05	1.40	29.39	1.46	25.43

M - The number of coefficients taken for reconstruction.

BR - Bit rate (bits/pixel).

PSNR - Peak signal-to-noise ratio (dB).

TABLE III: Comparisons among the transforms for the figure region shown in Fig. 4.

	rCS-SCHT		DCT		WHT	
M	BR	PSNR	BR	PSNR	BR	PSNR
8	0.48	19.75	0.49	20.58	0.48	19.87
12	0.63	20.48	0.65	21.16	0.64	20.65
16	0.76	21.08	0.77	21.55	0.76	21.05
24	0.99	21.89	0.99	22.19	0.99	21.82
32	1.18	22.95	1.18	22.91	1.19	22.96

well suited to be employed in document image compression. Another obvious advantage of the proposed transform in such application is the reduced computational complexity (which is to be explained in detail in Section IV) of the transform computation which requires only addition/subtractions.

IV. SIMULATION RESULTS

The computer simulation is conducted in order to evaluate the bit rate (BR) vs PSNR performance of the proposed transform as compared to the DCT and the WHT. The default setting for JPEG [1] is used. After the transform coefficients have been obtained, the Huffman coding is used in order to remove coding redundancy. The testing is focused on the text and figure regions as shown in Figs. 3 and 4. We fix the number of transform coefficients in order to make comparisons of bit rates and PSNRs for different transforms. Table II lists the BR and PSNR values at different values of M (number of selected coefficients for reconstruction) using the proposed transform, the DCT and the WHT for the text region shown in Fig. 3. It is shown in the table that when the value of M increases, the bit rate and PSNR become larger. The lower value of BR is required for higher compression whereas the larger value of PSNR is desired for better quality of the decompressed image. As shown in the table, the DCT offers the best quality of the image (i.e., higher PSNR) at any particular value of M with higher BR. But the proposed transform can provide lower BR which implies higher compression ratio with relatively low PSNR as compared to the DCT and the WHT. Table III summarizes the results for the figure region shown in Fig. 4. The same conclusions can also be derived from Table III.

But the obvious advantage of the proposed transform over the DCT and the WHT is simpler implementation and re-

duced computational cost. As an example, let us consider the scanned document of (2408×3504) pixels shown in Fig. 2 for compression. The original uncompressed size is $2408 \times 3504 = 8,437,632$ bytes as the bit depth is 8 bits (grey-scaled image). If the proposed transform with 12 coefficients is used, the bit rate is 0.75 bits/pixel as shown in Table II and the compressed size will be $(8,437,632 \times 0.75) \div 8 = 791,028$ bytes which is approximately 11 times reduced. Now let us focus on the computational complexity. The scanned document image contains a total of (8×8) blocks, that is, $(2408 \times 3504) \div (8 \times 8) = 131838$ blocks to be transformed. Each 8×8 block is required 16 transformations (8 rows and 8 columns) by using the 8-point transform. In order to compute the 8-point DCT, the fast algorithm requires 12 multiplications and 29 addition/subtractions [11]. Therefore, a total of $(131838 \times 16 \times 12) = 25,313,088$ multiplications and $(131838 \times 16 \times 29) = 61,172,832$ addition/subtractions is needed to transform the scanned image to the DCT domain. On the other hand, the proposed transform and the WHT do not require any multiplication to compute the transform, and they only require to perform addition/subtractions for computation. As mentioned in Section II, the fast algorithm of the 8-point proposed transform (shown in Fig. 1) only requires 18 addition/subtractions, which is actually 25% less than that of fast WHT algorithm [10] which needs 24 addition/subtractions in order to compute an 8-point WHT. Hence, an additional saving of $(131838 \times 16 \times 24 \times 0.25) = 12,656,448$ addition/subtractions can be obtained by using the proposed transform in compression. Therefore, the proposed transform has the lowest computational cost among the transforms and it is a better choice to be employed for a high-speed document scanner for real-time applications.

V. CONCLUSION

This paper presents a new transform which is useful to be employed in document image compression. The derivation of the proposed transform from the CS-SCHT is presented. The fast algorithm to compute the transform is described especially for $N = 8$, and its computational complexity is examined. It has been found that the introduced transform is computationally efficient and it only requires 18 addition/subtractions in order to compute the 8-point transform. Energy compaction capability of the transform for scanned document images is demonstrated by evaluating the MSE performance of the reconstructed image with respect to the original image. It has been shown that the proposed transform possesses good energy compaction ability like the DCT and the WHT. Subsequently, the transform is applied in document image compression and the computer simulation is performed in order to evaluate the bit rate vs PSNR performance of the proposed transform as compared to the DCT and the WHT. It has been found that the proposed transform is able to achieve lower bit rate (bits/pixel) than the others but with relatively low PSNR at any particular value of M (number of selected coefficients for decompressing the image). Moreover, the proposed transform has been shown to be a good choice for high-speed document scanner for low-

end applications due to its simple implementation and very low computational cost.

REFERENCES

- [1] R. C. Gonzalez and R. E. Woods, *Digital image processing*, International ed. Pearson Prentice Hall, 2002.
- [2] P. Haffner, L. Bottou, P. G. Howard, P. Simard, Y. Bengio, and Y. LeCun, "High quality document image compression with DjVu," *Journal of Electronic Imaging*, vol. 7, no. 3, pp. 410–25, 1998.
- [3] W. Elloumi, M. Chakroun, M. Charfi, and M. A. Alimi, "Compression of the images of ancient arab manuscript documents based on segmentation," in *6th IEEE International Conference on Computer Systems and Applications*, Doha, Qatar, 2008, pp. 879–883.
- [4] O. E. Kia and D. S. Doermann, "Residual coding in document image compression," *IEEE Transactions on Image Processing*, vol. 9, no. 6, pp. 961–969, 2000.
- [5] I. Sharpe, L. H. and B. Manns, "JPEG 2000 options for document image compression," in *Proc. SPIE - Int. Soc. Opt. Eng. (USA)*, vol. 4670, USA, 2002, pp. 167–173.
- [6] E. I. Ageenko and P. Franti, "Enhanced JBIG-based compression for satisfying objectives of engineering document management system," *Optical Engineering*, vol. 37, no. 5, pp. 1530–1538, 1998.
- [7] J. G. Proakis and D. G. Manolakis, *Digital signal processing: Principles, algorithms, and applications*, 4th ed. Pearson Prentice Hall, 2007.
- [8] I. Valova and Y. Kosugi, "Hadamard-based image decomposition and compression," *IEEE Transactions on Information Technology in Biomedicine*, vol. 4, no. 4, pp. 306–319, 2000.
- [9] A. Aung, B. P. Ng, and S. Rahardja, "Conjugate symmetric sequency-ordered complex Hadamard transform," *IEEE Trans. Signal Process.*, to be published, 2009.
- [10] K. G. Beauchamp, *Applications of Walsh and related functions, with an introduction to sequency theory*. Academic Press, 1984.
- [11] C. W. Kok, "Fast algorithm for computing discrete cosine transform," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 757–760, 1997.