次世代數位影音多用途光碟系統之光機電整合研究(1/3)---子計畫(四)數位視訊與音訊處理

The Integration in Optics, Mechanics, and Electronics of Digital Versatile Disc Systems (1/3) ---(IV) Digital Video and Audio Signal Processing

計畫編號 ： NSC87-2218-E-009-036
執行期間 ： 86 年 8 月 1 日 --- 87 年 7 月 31 日

主持人 ： 吳炳飛
執行機構及單位 ： 國立交通大學 電機與控制工程學系

摘要:

This study presents two sets of translation invariant wavelet transforms for coding of arbitrarily shaped image. Each of the sets can be viewed as a shape-adaptive discrete wavelet transform (SA-DWT) with the properties of translation invariant. The merits of the proposed transform schemes include as follows: 1) they are translation invariant; 2) no sharp transition appears at image edges; 3) the number of pixels maintains constant after transformation; 4) the correlation of pixels is fully exploited; and 5) the properties of self-similarity across scales are preserved. Experimental results show that the proposed transform schemes outperform other object-based transforms, including the SA-DWT.

計畫緣由與目的:

In addition to the conventional frame-based coding of MPEG-1 and MPEG-2 standards, the MPEG-4 video algorithm will also support the object-based coding within video scenes. Several algorithms have focused on the object-based coding. Except the algorithm in which uses the shape-adaptive discrete cosine transform (SA-DCT) to decompose an object region, the other algorithms are based on the discrete wavelet transform (DWT) in order to obtain a better transform efficiency. Depending on the different designs in filtering a one-dimensional (1-D) odd length segment, we can classify DWT-based transformations into three groups.

The algorithms fallen into the first group deal with the case of filtering an odd length segment by adding some extra pixels. These algorithms include object wavelet transform (OWT), edge sensitive wavelet transform (ESWT) and region-based subband transform (RBST). OWT adds many extra pixels into contour blocks and their neighboring blocks for the application of regular subband filtering operation. It requires a low-pass filter to smooth all of the extrapolated regions because extrapolating pixels between adjacent blocks are generally different. Because a long-tap filter affects the coefficients outside the region, the coding using OWT should always preserve the number of transformed coefficients to be more than the number of pixels contained in the original object region. In this way a fattened region is obtained in this scheme. The redundancy will lower the coding performance. On the other hand, ESWT and RBST add only one extra pixel to an odd length segment to form an even length segment for using regular DWT. The extra pixel is well designed so that the perfect reconstruction is available, even though one of the transformed coefficients is discarded to maintain the number of samples constant after transformation. In other words, ESWT and RBST can both achieve the requirements of the critical sampling. However, the added pixel may lower the correlation of itself and its neighbors, and thus reduces the transform efficiency.

The algorithms fallen into the second group deal with the odd length filtering by transforming the largest even length pixels and a remaining pixel. The remaining pixel is not processed by the filter bank and is put directly in the low-pass subband after having been scaled in gain of the low-pass analysis filter. Since these algorithms do not exploit the correlation of remaining pixel and its neighbors, it may lower the transform efficiency. The arbitrarily shaped wavelet packets (ASWP) is the algorithm of this group.

The algorithms fallen into the third group deal with the filtering of an odd length segment by applying the transform to the segment directly. No extra pixel is added and no pixel is remained. These algorithms include matching pursuit (MP) and shape-adaptive DWT (SA-DWT). MP utilizes an exhaustive search among the predefined basis pool to find the most suitable basis function, and results in high computing costs. In contrast to MP, the SA-DWT uses only a pair of biorthogonal wavelet filters and thus has less computational complexity. The SA-DWT is a generalized DWT. Its difference from regular DWT is the selection schemes for remaining coefficients. We will show this difference later. Since the correlation of pixels is completely exploited, the SA-DWT can obtain a better transform efficiency. In addition, the properties of self-similarity across scales is preserved. However, DWT is not a translation invariant transformation, neither is SA-DWT. A small shift of an object in the space domain may result in dramatic change in the transform domain. Thus, it is necessary to deal with the problem of translation sensitivity of the SA-DWT coding.

In this study, we present two sets of translation invariant DWTs (TI-DWTs) for coding of arbitrarily shaped image. Each of the sets can be viewed as a translation invariant SA-DWT (TI-SA-DWT). One set is for odd length biorthogonal wavelet filters and the other is for even length biorthogonal wavelet filters. The proposed transformations are the extensions of the transformation designed for rectangular image in to an arbitrarily shaped image. The symmetric extension at image boundaries is used to reduce edge effects. Experimental results are shown and compared with other DWT-based coders published in the literature. In addition, the coding results of the SA-DCT coder in are also displayed for comparison.

結果與討論:

In this section, some experimental results are shown for comparison. The first frame of the MISS AMERICA sequence and the first frame of the AKIYO sequence are used for demonstration. Only the luminance component of each image is used in our experiment. The original format of the MISS AMERICA image is in a common intermediate format (CIF): 352×288. We use a reference block with size 512×512 to cover the extracted object, i.e. the header-and-shoulder image (see Fig. 1(a)). The total number of pixels in the object region is 50,865. The binary alpha plane for the object is depicted in Fig. 1(b). Herein, the white region represents the object region while the black region is the background. As for the original size of the AKIYO image, it is 720×486 while the reference block size for the extracted object (see Fig. 1(c)) is 1024×1024. There are 128,343 pixels in this object region. For these two images, each pixel consists of 8 bits. A three-stage decomposition is performed for each image. The corresponding binary alpha plane of the AKIYO image after a three-stage decomposition is shown in Fig. 1(d). The 9/7 taps filter pair is used. All of the experimental results are calculated in peak signal-to-noise ratio (PSNR),

$$\mathrm{PSNR} = 10\log_{10}(\frac{255^2}{\mathrm{MSE}})\ \mathrm{dB}.$$

Table I tabulates the results of the proposed transform scheme in the rate-distortion coding stated in Section III. The results of using SA-DWT [13] are also shown for comparison. The SA-DWT in [13], which uses the configuration of ED at all decompositions, is a special case in our searching space V. In the rate-distortion coding, all the transformed coefficients are quantized by the optimal quantizer q^*, and then the quantized values are arithmetically coded. The MSE is calculated from the reconstructed image and the original image. From Table III, we see that our proposed transform scheme, TI-SA-DWT, can gain about 0.1 dB improvement in the PSNR. This is especially obvious for the cases of lower bit rates.

Figure 2 displays the results of the proposed transform scheme in zerotree coding. To show the transform efficiency of the proposed scheme, some other object-based transforms are also simulated for comparison. These transforms include ASWP [7], ESWT [6], SA-DCT [21], and SA-DWT [13]. Notice that except SA-DCT, all of the transforms use the zerotree coding algorithm [19] to encode the transformed coefficients and the action of sending out bits is only necessary for the coefficients in the transformed object region (white region in Fig. 1(d)). As for SA-DCT, it uses the block size 8×8 in our experiments. The lowest frequency coefficient is differentially coded for a SA-DCT contour block or a regular DCT block, while the other coefficients are run length coded. In addition, the informative JPEG Huffman lookup tables are used to elevate its coding performance.

The experimental results show that the SA-DWT outperforms the other object-based approaches over a wide range of bit rates. For example, the average goes beyond ESWT coding by 0.66 dB for MISS AMERICA, and surpasses SA-DCT coding by 1.7 dB for AKIYO. Our proposed scheme promotes the performance of SA-DWT further by the average about 0.1 dB.

Note that the coding results shown herein do not contain the shape coding. In other words, we assume that the shape information is known at the receiver or decoder side. The shape information can be coded by differential chain codes [10].

計畫成果自評:

本子計畫主要在研究小波轉換應用於視訊壓縮的成效。子計劃第一年重點在於開發比 MPEG 2 更有效率的視訊壓縮技術，以因應於公元 2000 年，壓縮技術的重大改革－小波轉換的引進。研究成果與計劃書的預期成果相似，個人覺得評量不錯。非常謝謝國科會經費支援使得研究能進行。

參考文獻:

[1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, April 1992, pp. 205-220.

[2] H. J. Barnard, J. H. Weber, and J. Biemond, "Efficient signal extension for subband/wavelet decomposition of arbitrary length signals," in *Proc. SPIE VCIP'93*, vol. 2094, Nov. 1993, pp. 966-975.

[3] M. Bi, S. H. Ong, and Y. H. Ang, “Comment on “Shape-adaptive DCT for generic coding of video,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, no. 6, Dec. 1996, pp. 686-688.

[4] J. R. Casas, L. Torres, “A region-based subband coding scheme,” *Signal Processing: Image*

Communication 10, 1997, pp. 173-200.

[5] K. Cinkler and A. Mertins, "Edge sensitive subband and coding of images," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, ICASSP'96*, Atlanta, GA, May 1996, pp. 2363-2366.

[6] K. Cinkler, "Very low bit-rate wavelet video coding," *IEEE Jour. Selected Areas in Communications*, vol. 16, no. 1, Jan. 1998, pp. 4-11.

[7] O. Egger, T. Ebrahimi, and M. Kunt, "Arbitrarily-shaped wavelet packets for zerotree coding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, ICASSP'96*, Atlanta, GA, May 1996, pp. 2335-2338.

[8] H. Jafarkhani and N. Farvardin, "Adaptive image coding using spectral classification," *IEEE Trans. Image Processing*, vol. 7, no. 4, April, 1998, pp. 605-610.

[9] H. Katata, N. Ito, T. Aono and H. Kusao, "Object wavelet transform for coding of arbitrarily shaped image segments", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, no. 1, Feb. 1997, pp. 234-237.

[10] H. Katata, N. Ito, and H. Kusao, "Temporal-scalable coding based on image content," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, no. 1, Feb. 1997, pp. 52-59.

[11] R. Koenen Ed., "Overview of the MPEG-4 Version 1 Standard," Doc. ISO/IEC JTC1/SC29/WG11 N1909, MPEG97, Oct. 1997.

[12] Final committee draft of MPEG-4 Visual Standard, MPEG document number N2202, ISO/IEC 14496-2, May 1998, annex A.

[13] S. Li, W. Li, Z. Wu, and H. Sun "Shape Adaptive Wavelet Coding," *ISCAS98 Proceedings*, Monterey, California.

[14] J. Liang and T. W. Parks, "Image coding using translation invariant wavelet transforms with symmetric extensions," *IEEE Trans. Image Processing*, vol. 7, no. 5, May 1998, pp. 762-769.

[15] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, Dec. 1993, pp. 3397-3415.

[16] S. A. Martucci, "Symmetric convolution and the discrete sine and cosine transforms," *IEEE Trans. Signal Processing*, vol. 42, no. 5, May 1994, pp. 1038-1051.

[17] R. Neff and A. Zakhor, "Very low bit-rate video coding based on matching pursuits," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, no. 1, Feb. 1997, pp. 158-171.

[18] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. Image Processing*, vol. 2, no. 2, April 1993, pp. 160-175.

[19] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, no. 3, June. 1996, pp.243-250.

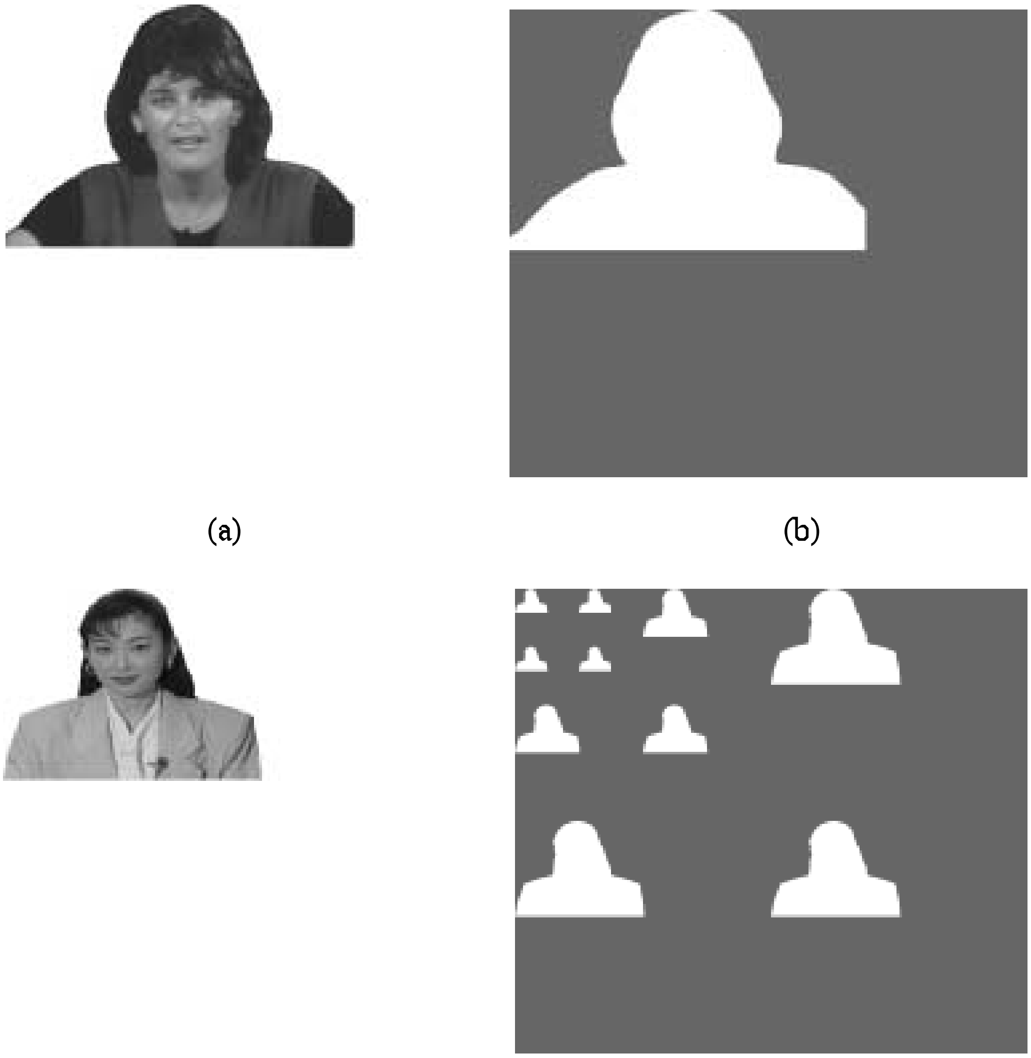
[20] J. M. Shapiro, "Embedded image coding using zerotrees of wavelets coefficients," *IEEE Trans. Signal Processing*, vol. 41, Dec. 1993, pp. 3445-3462.

[21] T. Sikora, "Low complexity shape-adaptive DCT for coding of arbitrarily shaped image segments," *Signal Processing: Image commun.*, vol. 7, 1995, pp. 381-395.

[22] T. Sikora, "The MPEG-4 Video Standard Verification Model," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, no. 1, Feb. 1997, pp. 19-31.

[23] M. Vetterli and C. Herley, "Wavelets and filter banks: theory and design,"

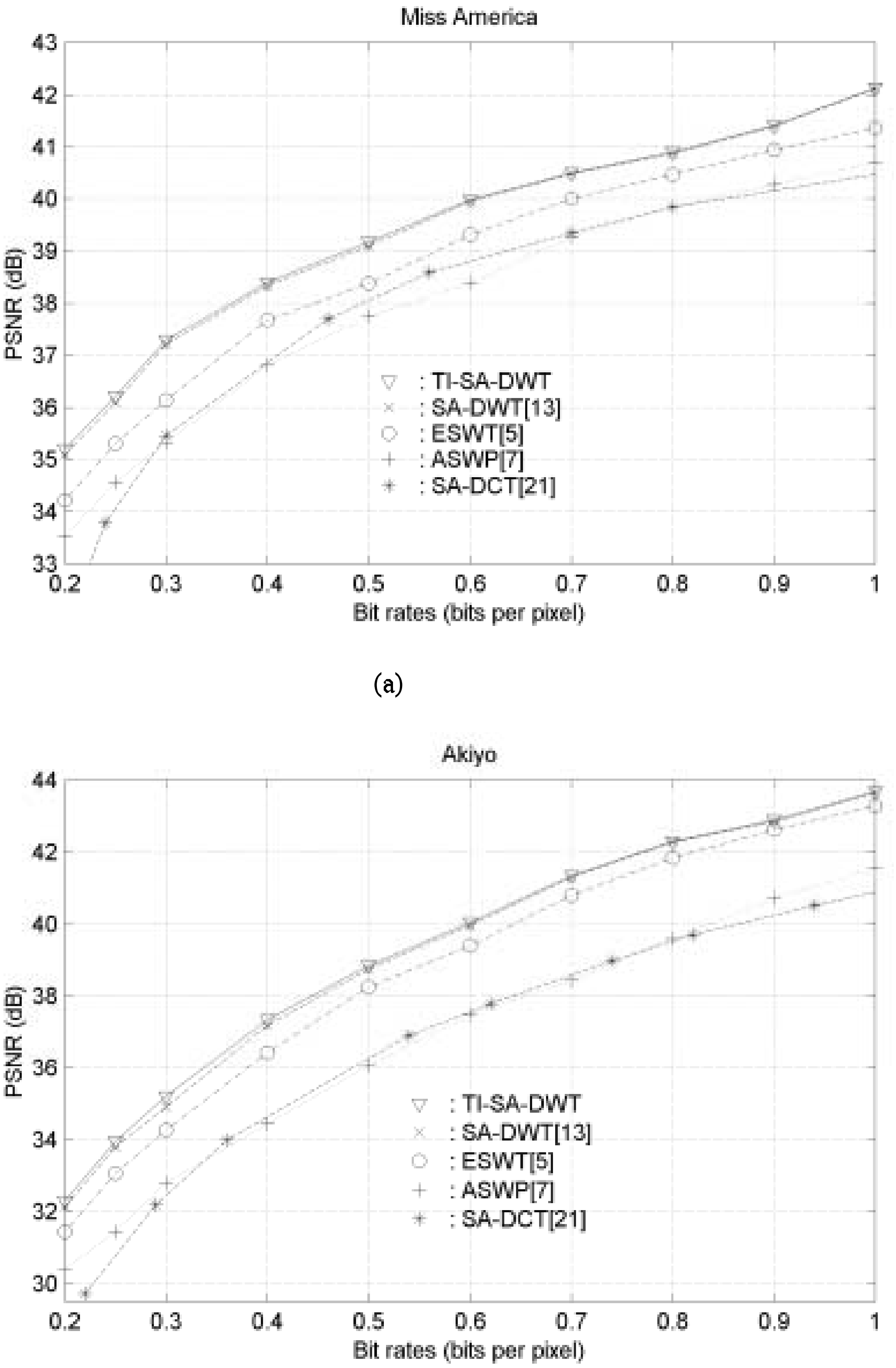
IEEE Trans. Signal Processing, vol. 40, no. 9, Sep. 1992, pp. 2207-2232.



(a) (b)

(c) (d)

Figure 1 The original arbitrarily shaped images and their own corresponding binary alpha plane;



(a)

(b)

Figure 2 Comparison of the rate-distortion performance; (a) for MISS AMERICA image; (b) for AKIYO image.

Table I

Comparison of experimental results of the SA-DWT in [13] and our translation invariant SA-DWT

SA-DWT				Translation invariant SA-DWT			
Miss America		Akiyo		Miss America		Akiyo	
Bit rate	PSNR	Bit rate	PSNR	Bit rate	PSNR	Bit rate	PSNR
0.28	34.78	0.29	32.42	0.28	34.87	0.29	32.53
0.37	36.37	0.31	32.85	0.38	36.43	0.31	32.98
0.46	37.51	0.48	36.16	0.48	37.65	0.47	36.11
0.58	38.5	0.52	36.82	0.58	38.54	0.49	36.47
0.67	38.96	0.61	38.12	0.7	39.15	0.61	38.12
0.71	39.14	0.7	39.07	0.88	39.98	0.72	39.42
0.87	39.88	0.8	40.18	0.9	40.09	0.8	40.33
0.92	40.15	0.89	41.05	1.08	40.91	0.9	41.24