ELSEVIER

Arbitrarily shaped image coding by using translation invariant wavelet transforms

Bing-Fei Wu*, Chorng-Yann Su

Department of Electrical and Control Engineering, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu 300, Taiwan

Abstract

This study presents two sets of translation invariant wavelet transforms for coding an arbitrarily shaped image. Each set can be viewed as a shape-adaptive discrete wavelet transform (SA-DWT) with the property of translation invariance. The proposed transform schemes have the following merits: (1) they are translation invariant, (2) no sharp transition appears at the edges of the image, (3) the number of pixels is maintained constant after transformation, (4) the correlation of pixels is fully exploited, and (5) the property of self-similarity across scales is preserved. © 1999 Elsevier Science B.V. All rights reserved.

Zusammenfassung

Diese Studie präsentiert zwei translationsinvariante Wavelet-Transformationen für die Kodierung eines beliebig geformten Bildes. Jede davon kann als eine formadaptive diskrete Wavelet-Transformation (SA-DWT) mit der Eigenschaft der Translationsinvarianz angesehen werden. Die vorgeschlagenen Transformationen haben folgende Vorteile: (1) sie sind translationsinvariant, (2) es erscheint kein scharfer Übergang an den Bildrändern, (3) die Pixelanzahl bleibt nach der Transformation konstant, (4) die Korrelation der Pixel wird vollständig ausgenutzt, und (5) die Eigenschaft der Selbstähnlichkeit wird für alle Skalen erhalten. © 1999 Elsevier Science B.V. All rights reserved.

Résumé

Cette étude présente deux ensembles de transformations en ondelettes invariantes par translation pour le codage d'images de forme arbitraire. Chaque ensemble peut être vu comme une transformation en ondelettes discrète adaptive en forme (SA-DWT) présentant la propriété d'invariance par translation. Les méthodes de transformation proposées ont les avantages suivants: (1) elles sont invariantes par translation, (2) aucune transition abrupte n'apparaît sur les contours de l'image, (3) le nombre de pixels est maintenu constant après transformation, (4) la corrélation des pixels est pleinement exploitée, et (5) la propriété d'auto-similarité au travers des échelles est préservée. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Arbitrarily shaped image coding; Translation invariant; Shape-adaptive discrete wavelet transform

* Corresponding author. Tel.: + 886-3-571-2121 ext. 54313; fax: + 886-3-571-2385.
E-mail addresses: bwu@cc.nctu.edu.tw (B.-F. Wu), u8312813@cc.nctu.edu.tw (C.-Y. Su)

Nomenclature

$a[n]$	input sequence	$B[k]$	high-pass transform coefficient
$\tilde{a}[n]$	extended version of input sequence	$B'[k]$	convolution result in high-pass channel
d	type of a downsampler	D	distortion measure
$g[n]$	analysis high-pass filter	Q	set of all admissible quantizers
$\tilde{g}[n]$	synthesis high-pass filter	V	set of all possible decomposition configuration
$h[n]$	analysis low-pass filter		
$\tilde{h}[n]$	synthesis low-pass filter	R_b	coding budget
$A[k]$	low-pass transform coefficient	λ	Lagrangian multiplier
$A'[k]$	convolution result in low-pass channel		

1. Introduction

Several algorithms based on the discrete wavelet transform (DWT) have been suggested to efficiently remove the redundancy of an object including object wavelet transform (OWT) [5], edge sensitive wavelet transform (ESWT) [3], region-based subband transform (RBST) [2], arbitrarily shaped wavelet packets (ASWP) [4], matching pursuit (MP) [8], and shape-adaptive DWT (SA-DWT) [6]. Among these algorithms, MP and SA-DWT have the best transform efficiency because they exploit the correlation of pixels completely. More specifically, in addition to its lower computational complexity than MP, SA-DWT has been applied to MPEG-4 still image coding. However, DWT is not a translation invariant transformation, neither is SA-DWT. A small shift of an object in the space domain may dramatically change the transform domain. In this study, we present two sets of translation invariant DWTs (TI-DWTs) to resolve the problem of translation sensitivity of SA-DWT coding. Each set 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. While the former is the extension of the transformation designed for rectangular images in [7], the latter is initially addressed herein.

2. Overview of proposed transform scheme

To reduce computational complexity, this work performs the wavelet decomposition separately to the rows and columns of an arbitrarily shaped image. Moreover, the decomposition is only applied to object regions. Restated, if a row or column of a region is split into M unconnected segments, each segment is filtered and downsampled separately.

Symmetric extension on both boundaries of each segment is used to reduce edge effects. Among the four symmetric structures, i.e. whole-point symmetry (WS), half-point symmetry (HS), whole-point anti-symmetry (WA), and half-point anti-symmetry (HA) – the WS structure is used for odd length filters and the HS structure for even length filters.

Without loss of generality, each segment can be relocated to a new starting position 0 or 1, depending on its original starting position being even or odd, respectively. According to the starting position (0 or 1) and the length (even or odd) of a segment, there are four cases of processing. An SA-DWT is designed to deal with the four cases. Because the filter length is either even or odd, there are two types of SA-DWT.

2.1. SA-DWT for both analysis filters being of odd length

In a two-band, perfect reconstruction filter bank, [1] indicated that the filters should satisfy

$$g[n] = (-1)^n \tilde{h}[1 - n] \text{ and } \tilde{g}[n] = (-1)^n h[1 - n],\tag{1}$$

where $g[n]$, $h[n]$, $\tilde{g}[n]$ and $\tilde{h}[n]$ represent the analysis high-pass filter, the analysis low-pass filter, the synthesis high-pass filter and the synthesis low-pass filter, respectively.

Assume that a pair of symmetric odd length filters $(h, \tilde{h})$ is given and defined in the following

intervals: $h[n]$ for $-p \leqslant n \leqslant p$ and $\tilde{h}[n]$ for $-q \leqslant n \leqslant q$. From (1), we can obtain the values: $g[n]$ for $-q + 1 \leqslant n \leqslant q + 1$ and $\tilde{g}[n]$ for $-p + 1 \leqslant n \leqslant p + 1$. According to the definition of discrete convolution, we have the following situation:

$$A'[k] = \sum_{n=-p}^{p} h[n]\tilde{a}[k-n] \quad \text{and}$$

$$B'[k] = \sum_{n=-q+1}^{q+1} g[n]\tilde{a}[k-n], \qquad (2)$$

where $A'[k]$ and $B'[k]$ represent the convolution results, and $\tilde{a}[n]$ the extended version of input sequence $a[n]$. Notably, $\tilde{a}[n]$ is of WSWS structure herein because the WS structure is applied to both the left and the right sides of $a[n]$.

After the convolution, a downsampler with a downsampling factor of 2, follows in order to remove redundancy. Depending on the retained index terms of a convolution result, whether it is even or odd, we have even downsampling (ED) and odd downsampling (OD), respectively. For simplicity, we define that

$$A[k] = A'[2k + d] \quad \text{and} \quad B[k] = B'[2k + d], \qquad (3)$$

where d represents the type of the downsampler. For ED, $d = 0$. For OD, $d = 1$. Fig. 1 illustrates sequences involved in computing a discrete convolution. The input sequence $a[n]$ is assumed to be of even length ($2K$). The extended samples of $a[n]$ are denoted with dashed lines. The symmetric axes of sequences are also shown.

Table 1 lists the symmetric structures of $A[k]$ and $B[k]$ when filtering the four cases. Each sequence ($A[k]$ or $B[k]$) has two symmetric axes, and we only need to retain the samples that fall into the axes. For example, if the symmetric axes of $A[k]$ are $(0, K - 1/2)$, we retain $A[k]$ for $0 \leqslant k \leqslant K - 1$. As can be easily verified, the total length of retained samples for each of the pair $A[k]$ and $B[k]$ is equal to the length of input sequence $a[n]$.

The formula corresponding to (2) for reconstruction is

$$a[n] = \sum_{\substack{k=-q \\ k+n-d:\text{even}}}^{q} \tilde{h}[k]A\left[\frac{k+n-d}{2}\right]$$

$$+ \sum_{\substack{k=-p+1 \\ k+n-d:\text{even}}}^{p+1} \tilde{g}[k]B\left[\frac{k+n-d}{2}\right]. \qquad (4)$$
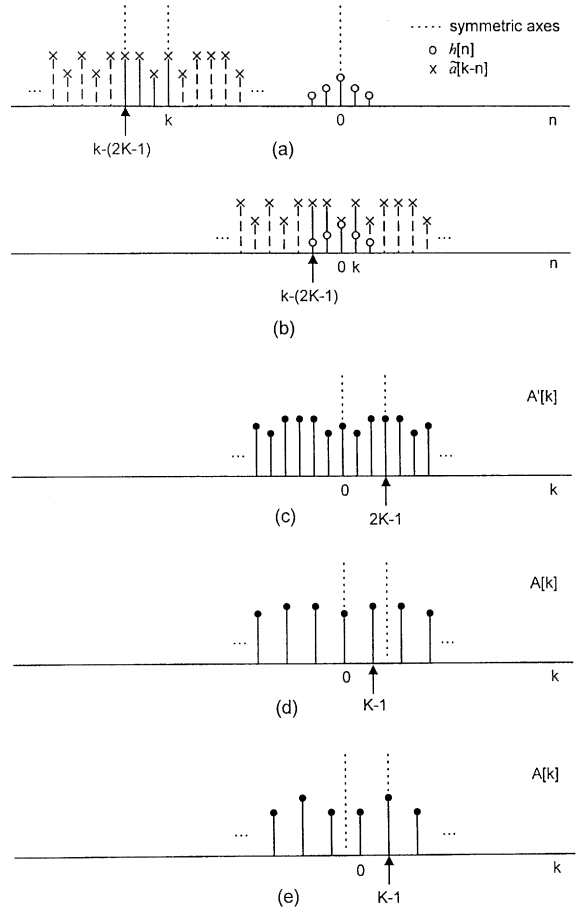

Fig. 1. Sequences involved in computing a discrete convolution. (a) and (b) The sequences $\tilde{a}[k-n]$ and $h[n]$ as a function of n for different values of k. (c) Corresponding output sequence as a function of k. (d) Even downsampling output. (e) Odd downsampling output.

2.2. SA-DWT for both analysis filters being of even length

In this case, assume that the given symmetric filter pair $(h, \tilde{h})$ is defined at $h[n]$ for $-p \leqslant n \leqslant p + 1$, and $\tilde{h}[n]$ for $-q \leqslant n \leqslant q + 1$. From (1), we have the following values: $g[n]$ for $-q \leqslant n \leqslant q + 1$, and $\tilde{g}[n]$ for $-p \leqslant n \leqslant p + 1$. Herein, both $g[n]$ and $\tilde{g}[n]$ are of HA structure. Replacing the boundaries of summation index n in (2) and using HSHS structure to $\tilde{a}[n]$ allow us to obtain the output symmetric structures of $A[k]$ and $B[k]$ as listed in Table 2.

Table 1
Output symmetric structures when both analysis filters are of odd length and symmetric

Input sequence $a[n]$		Convolution results (symmetric axes)		Even downsampling output (symmetric axes)		Odd downsampling output (symmetric axes)	
Length	Starting point	$A'[k]$	$B'[k]$	$A[k]$	$B[k]$	$A[k]$	$B[k]$
$2K$	0	WSWS $(0, 2K-1)$	WSWS $(1, 2K)$	WSHS $(0, K-1/2)$	HSWS $(1/2, K)$	HSWS $(-1/2, K-1)$	WSHS $(0, K-1/2)$
$2K$	1	WSWS $(1, 2K)$	WSWS $(2, 2K+1)$	HSWS $(1/2, K)$	WSHS $(1, K+1/2)$	WSHS $(0, K-1/2)$	HSWS $(1/2, K)$
$2K-1$	0	WSWS $(0, 2K-2)$	WSWS $(1, 2K-1)$	WSWS $(0, K-1)$	HSHS $(1/2, K-1/2)$	HSHS $(-1/2, K-3/2)$	WSWS $(0, K-1)$
$2K-1$	1	WSWS $(1, 2K-1)$	WSWS $(2, 2K)$	HSHS $(1/2, K-1/2)$	WSWS $(1, K)$	WSWS $(0, K-1)$	HSHS $(1/2, K-1/2)$

Table 2
Output symmetric structures when both analysis filters are of even length; one is symmetric and the other is antisymmetric

Input sequence $a[n]$		Convolution results (symmetric axes)		Even downsampling output (symmetric axes)		Odd downsampling output (symmetric axes)	
Length	Starting point	$A'[k]$	$B'[k]$	$A[k]$	$B[k]$	$A[k]$	$B[k]$
$2K$	0	WSWS $(0, 2K)$	WAWA $(0, 2K)$	WSWS $(0, K)$	WAWA $(0, K)$	HSHS $(-1/2, K-1/2)$	HAHA $(-1/2, K-1/2)$
$2K$	1	WSWS $(1, 2K+1)$	WAWA $(1, 2K+1)$	HSHS $(1/2, K+1/2)$	HAHA $(1/2, K+1/2)$	WSWS $(0, K)$	WAWA $(0, K)$
$2K-1$	0	WSWS $(0, 2K-1)$	WAWA $(0, 2K-1)$	WSHS $(0, K-1/2)$	WAHS $(0, K-1/2)$	HSWS $(-1/2, K-1)$	HAWA $(-1/2, 2K-1)$
$2K-1$	1	WSWS $(1, 2K)$	WAWA $(1, 2K)$	HSWS $(1/2, K)$	HAWA $(1/2, K)$	WSHS $(0, K-1/2)$	WAHS $(0, K-1/2)$

Of particular interest are some pairs of $A[k]$ and $B[k]$ in Table 2, preserving all the samples that fall into the symmetric axes would violate critical sampling. Fortunately, the sample value at the symmetric axis of a WA structure is always 0. This sample can be omitted to achieve critical sampling. At decoder side, as long as this zero value is added to the sequence, perfect reconstruction is still available.

3. Design of TI-SA-DWT for arbitrarily shaped image coding

Owing to that a two-band filter bank downsampled by 2 is shift-invariant with respect to even shifts and an SA-DWT is based on such a filter bank, an SA-DWT is also shift-invariant with respect to even shifts. In addition, the role of an ED-type SA-DWT is equivalent to the role of an OD-type SA-DWT with odd shifts (see Fig. 1). Therefore, search for the

best combination of EDs and ODs for an S-stage decomposition would allow us to obtain the most efficient representation of an object region regardless of how the object shifts. Briefly, each combination of EDs and ODs is denoted as a decomposition configuration (DC). In an S-stage decomposition, the number of all possible DCs is 4^S because each stage contains two SA-DWTs: one is for the rows of an image and the other is for the columns. In addition, each SA-DWT can use ED or OD. For a typical application, S is often limited to be 4 or 5. Thus, the searching space is not very large.

In [7], some cost functions are listed for finding the best DC for an object. In this study, we use the separable rate-distortion cost function. Our problem can be stated simply as

$$\min_{v \in V} \min_{q \in Q} D(v, q) \text{ subject to } R(v, q) \leqslant R_b, \qquad (5)$$

where D is an unweighted mean-squared error distortion measure, V is the set of all possible DCs at an S-stage decomposition, Q is the set of all admissible quantizers, and R_b is the coding budget. Introducing the Lagrangian cost $L(v, q) = D(v, q) + \lambda R(v, q)$, the
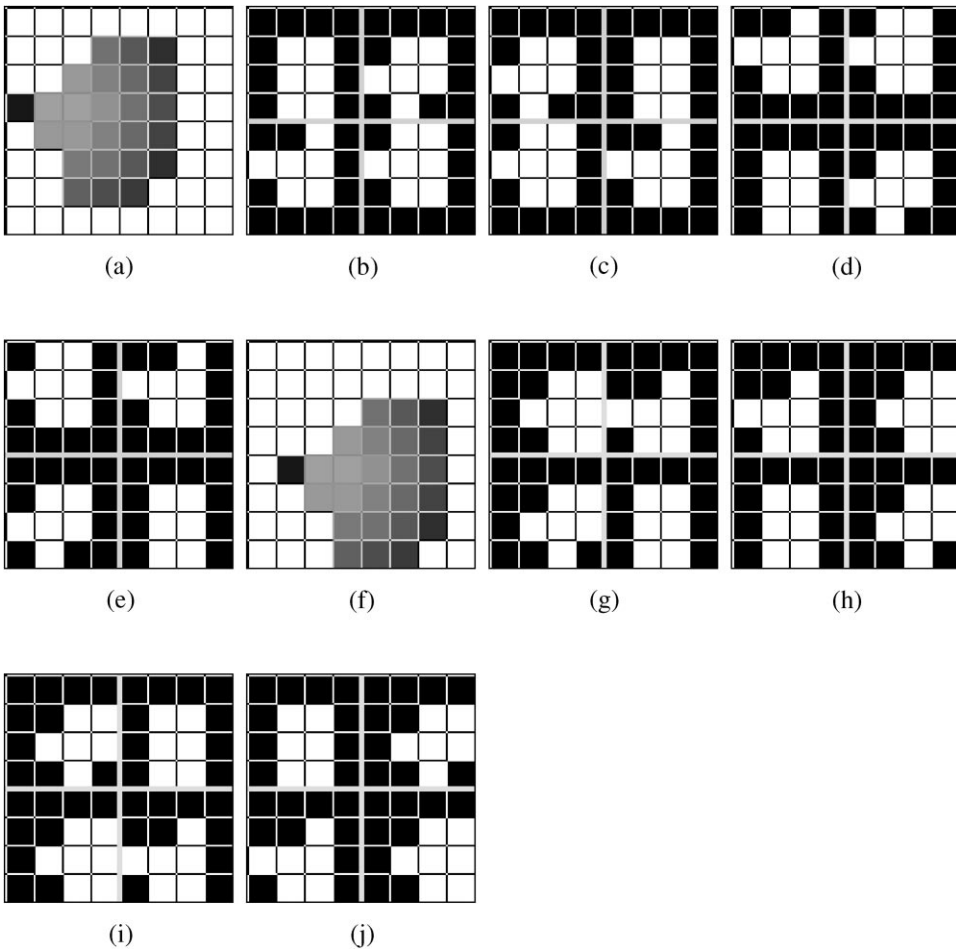

Fig. 2. A simple example used to explain the translation invariant property of the proposed transform. (a) An image with an object; (b)–(e) the binary alpha planes of the four possible DCs in a one-stage decomposition for (a); (f) the object shifted image; (g)–(j) the alpha planes of the four possible DCs in a one-stage decomposition for (f).

constrained optimization problem of (5) can be converted to

$$\max_{\lambda} \; \min_{v \in V} \; \min_{q \in Q} \; [D(v, q) + \lambda R(v, q)]. \tag{6}$$

To solve this equation, the first step is to search the optimal quantizer q^* for a fixed v and a fixed λ. The next step is to find the best v^* that minimizes the Lagrangian cost. Finally, a fast convex recursion in λ using the bisection algorithm is applied to obtain the optimal constant slope value λ^* [9].

4. Example

A simple example used to explain the translation invariant property of the proposed transform is shown in Fig. 2. Fig. 2(a) is an image with an object in which the total number of pixels is 25. We use 9/3 tapes filter pair [1] to filter the object. In a one-stage decomposition, the number of all possible DCs is 4. Figs. 2(b)–(e) depict the binary alpha planes for the four DCs. Herein, the white region denotes the object region and the black region represents the background. Each of the figures consists of four object regions. The object regions are separately located in different frequency bands. Fig. 2(f) contains the same object as shown in Fig. 2(a) but the object is shifted towards southeast by one pixel. The corresponding alpha planes of Fig. 2(f) are shown in Fig. 2(g)–(j). Comparing the shapes of object regions in Fig. 2(g)–(j) with those in Fig. 2(b)–(e), it is found that Fig. 2(b) is related to Fig. 2(j), Fig. 2(c) to Fig. 2(i), Fig. 2(d) to Fig. 2(h), and Fig. 2(e) to Fig. 2(g). In each pair of related figures, all the transform coefficients on the same object region are equivalent. Therefore, the searching space for Fig. 2(a) is identical to that for Fig. 2(f). Restated, if the alpha plane of the most efficient representation for Fig. 2(a) is as shown in Fig. 2(e), Fig. 2(g) will be the alpha plane of the most efficient representation for Fig. 2(f). Therefore, the transform coefficients we obtained are equal regardless of how the object shifts. This accounts for why the proposed transforms are translation invariant.

5. Conclusion

This study develops two sets of translation invariant wavelet transforms for coding an arbitrarily shaped image. Each set can be viewed as a translation invariant SA-DWT. The proposed scheme is efficient and outperforms other relevant object-based transforms, but it is limited only in that it takes more encoding time than the SA-DWT [6]. However, this drawback is acceptable since DWT is quite fast. Moreover, the computational complexity of our proposed scheme in decoding time is comparable to the SA-DWT. Thus, the proposed scheme is highly promising for browsing images from a database and MPEG-4 video player.

References

[1] M. Antonini, M. Barlaud, P. Mathieu, I. Daubechies, Image coding using wavelet transform, IEEE Trans. Image Process. 1 (2) (April 1992) 205–220.

[2] J.R. Casas, L. Torres, A region-based subband coding scheme, Signal Processing: Image Communication 10 (1997) 173–200.

[3] K. Cinker, Very low bit-rate wavelet video coding, IEEE J. Selected Areas Commun. 16 (1) (January 1998) 4–11.

[4] O. Egger, T. Ebrahimi, M. Kunt, Arbitrarily-shaped wavelet packets for zerotree coding, in: Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing, ICASSP'96, Atlanta, GA, May 1996, pp. 2335–2338.

[5] H. Katata, N. Ito, T. Aono, H. Kusao, Object wavelet transform for coding of arbitrarily shaped image segments, IEEE Trans. Circuits Systems Video Technol. 7 (1) (Februvary 1997) 234–237.

[6] S. Li, W. Li, Z. Wu, H. Sun, Shape adaptive wavelet coding, in: ISCAS'98 Proceedings, Vol. 5, Monterey, California, 1998, pp. 281–284.

[7] J. Liang, T. W. Parks, Image coding using translation invariant wavelet transforms with symmetric extensions, IEEE Trans. Image Process. 7 (5) (May 1998) 762–769.

[8] S. Mallat, Z. Zhang, Matching pursuits with time–frequency dictionaries, IEEE Trans. Signal Process. 41 (December 1993) 3397–3415.

[9] K. Ramchandran, M. Vetterli, Best wavelet packet based in a rate-distortion sense, IEEE Trans. Image Process. 2 (2) (April 1993) 160–175.