

# Comparison of Wavelet and Cosine Basis for Representation of Arbitrarily Shaped Image Segments

Liu Ying\*, Surendra Ranganath\*\*, Xiaofang Zhou\*  
\*School of Information Technology and Electrical Engineering,  
University of Queensland, Queensland, 4072, Australia  
\*\*Department of Electrical and Computer Engineering,  
National University of Singapore, Singapore, 117576

## ABSTRACT

*The efficient representation of arbitrarily shaped image segments is an important issue in object-based image coding. This paper describes an algorithm by which the given image segment is successively approximated using 2-D shape-independent basis functions defined on a rectangle circumscribing the image segment. Discrete cosine and discrete wavelet basis were used in the system respectively, and performance comparison between them was made. Simulation results show that wavelet basis usually yields better approximations than the discrete cosine basis. However, using wavelet basis, the extrapolation outside the image segment is not smooth, due to the selection of basis functions with small overlap with the image segment. To solve this problem, we proposed a new basis selection method which gives priority to basis functions that have larger overlap with the image segment. The extrapolation outside the segment is improved without degrading approximation quality.*

## 1. INTRODUCTION

Object-based image coding can overcome the disadvantages of traditional block-based image coding techniques, such as blocking artifacts and mosquito noise. In addition, it offers an important functionality-object-based scalability. This feature plays a key role in advanced interactive video communication systems. Hence, object-based image coding is a promising technique.

In object-based image coding schemes, images are decomposed into objects (image segments) on the basis of motion, texture, structure, etc. [1]. The efficient description of image segments having arbitrary shapes is an important issue in object-based image coding.

There are many approaches to discrete linear approximation of an image segment, such as using polynomial basis functions [2] or orthogonal basis

functions [3]. However, they have the disadvantage that the coefficients of the approximation depend on the shape (contour) of the segment to be approximated. Thus, at the decoder side, if there is loss in the contour information, the decoded texture will be affected. To overcome this problem, Kaup described an algorithm[4] based on Matching Pursuits (MP) [5] to represent arbitrarily shaped image segments by selecting a small set of cosine basis functions from a larger basis set defined on a rectangle circumscribing the region[4]. The author segmented image segment into many homogeneous segments, this then required significant effort to encode the contours of all these sub-segments. We applied the MP-based basis selection idea to the whole image segment directly, and formulated MP-DWT and MP-DCT by using discrete wavelet basis [6][7] and discrete cosine basis, respectively. Comparison of the above two methods showed that approximation of the given segment by MP-DWT was usually better than MP-DCT. However, the extrapolation outside the image segment was not smooth. To overcome this problem, we proposed a new basis selection method by defining a coefficient selection mask (CSMask) which restricts the basis functions which can be chosen. Results show that with CSMask, the extrapolation result of MP-DWT is greatly improved.

## 2. SUCCESSIVE APPROXIMATION BASED ON MATCHING PURSUITS

Assume that we have an image segment  $f(m,n)$  containing  $M$  pixels, we aim to find an approximation  $g(m,n)$  that is as close as possible to  $f(m,n)$  and which can be compactly coded to achieve data compression.

Let's start from a set of basis functions  $\Phi$  defined on a rectangular region  $L$  circumscribing the given image segment  $A$ .  $L$  is of size  $N=M_0*N_0$ . The basis functions are assumed to be linearly independent with respect to  $L$ .

$$\Phi = \{ \mathbf{j}_{kl}(m,n) \mid 0 \leq k < M_0, 0 \leq l < N_0, (m,n) \in L \} \quad (1)$$

According to linear approximation theory,  $g(m,n)$  can be represented by a weighted sum of a set of arbitrary, linearly independent basis functions  $\mathbf{j}_{kl}(m,n)$  as

$$g(m,n) = \sum_{(k,l) \in \mathbf{k}} C_{kl} \mathbf{j}_{kl}(m,n) \quad (2)$$

The set  $\mathbf{k}$  comprises of index pairs  $(k,l)$  of all basis functions used in the expansion [4].

With a total of  $N=M_0*N_0$  basis functions available as in (1), we aim to approximate  $f(m,n)$  using only a small set of these basis functions, and we need to find the approximation coefficients  $c_{kl}$  for a best approximation  $g(m,n)$  to  $f(m,n)$ .

In MP-DWT (MP-DCT),  $f(m,n)$  is successively approximated using selected wavelet (cosine) basis. At each iteration, the basis function which best matches the residual signal is selected. Details are described below.

Suppose  $g^{(v)}(m,n)$  is an approximation to  $f(m,n)$  in the  $v^{th}$  step of an iterative procedure. From (2),

$$g^{(v)}(m,n) = \sum_{(k,l) \in K_v} c_{kl}^{(v)} \mathbf{j}_{kl}(m,n) \quad (3)$$

where  $K_v$  denotes the set of basis function indices used in the expansion for  $g^{(v)}(m,n)$ . Then, the residual between the given image segment and the approximation is given by

$$r^{(v)}(m,n) = f(m,n) - g^{(v)}(m,n) \quad (4)$$

In order to concentrate most of the energy of  $f(m,n)$  into only a few approximation coefficients  $c_{kl}$ ,  $r^{(v)}(m,n)$  is approximated by selecting the best matching basis function  $\mathbf{j}_{pq}(m,n)$  to make the remaining residual error  $r^{(v+1)}(m,n)$  as small as possible:

$$r^{(v+1)}(m,n) = r^{(v)}(m,n) - \Delta c \mathbf{j}_{pq}(m,n) \quad (5)$$

According to [4],  $\mathbf{j}_{pq}(m,n)$  is chosen as the basis

function which maximizes  $\Delta E_A^n$  in (6),

$$\Delta E_A^{(v)} = \frac{\left[ \sum_{(m,n) \in A} r^{(v)}(m,n) \mathbf{j}_{pq}(m,n) \right]^2}{\sum_{(m,n) \in A} \mathbf{j}_{pq}^2(m,n)} \quad (6)$$

and the weight  $\Delta c$  is hence given by

$$\Delta c_{pq}^{(v)} = \frac{\left[ \sum_{(m,n) \in A} r^{(v)}(m,n) \mathbf{j}_{pq}(m,n) \right]}{\sum_{(m,n) \in A} \mathbf{j}_{pq}^2(m,n)} \quad (7)$$

Then, the corresponding coefficient  $c_{pq}$  of the approximation function (3) is updated by adding  $\Delta c$ , so that we have

$$c_{pq}^{(v+1)} = c_{pq}^{(v)} + \Delta c_{pq} \quad (8)$$

The just selected basis function  $\mathbf{j}_{pq}(m,n)$  is then added to the set of basis functions  $K_v$ , if this function is not yet included in the list. And, the new approximation function  $g^{(v+1)}(m,n)$  is determined by evaluating (3).

Repeating the above steps for  $v=1,2,3,\dots$ , until it reaches a certain number of iterations or the error energy drops below a pre-specified threshold. Then  $g^{(t)}(m,n), (m,n) \in A$  approximates the image segment, meanwhile, it provides an extrapolation for  $(m,n) \in L \setminus A$ .

Note that in order to use fast wavelet transform in the algorithm,  $M_0, N_0$  are selected as the smallest powers of two to just enclose the segment. In addition,  $f(m,n)$  is initially extended into  $r^{(0)}(m,n)$  of rectangular region  $L$  by zero padding [4].

### 3. COMPARISON OF MP-DCT AND MP-DWT

We measure the quality of the approximation results by PSNR (Peak Signal to Noise Ratio) as well as visually. We used eye segment and face segment as shown in Figure 1(a) and 1(b) (eye segment with  $M=524, N=M_0*N_0=32*32$ , face segment with  $M=2396, N=64*64$ ). Figure 1(c) gives a homogeneous texture image segment with  $M=955, N=64*64$ .

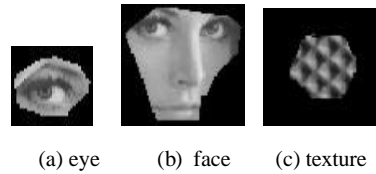
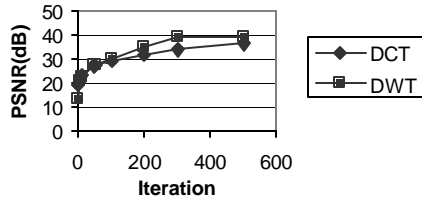


Figure 1 Three image segments used

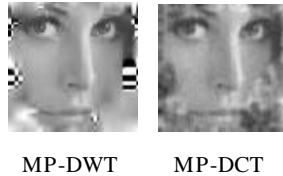
With both MP-DCT and MP-DWT, simulation results show that as iterations progress, the PSNR values increase, because  $f(m,n)$  is built up gradually. We compared the performances of MP-DCT and MP-DWT by the PSNR values and visual results of the reconstructed images.

Figure 2(a) compares the PSNR of MP-DWT and MP-DCT for the face segment. During the initial iterations, MP-DCT has higher PSNR, but as iteration increases, the PSNR of MP-DWT becomes higher (6-level DWT, Daubechies filter of length 6 used). Note that for homogeneous image segments, e.g. the texture in Figure 1(c), MP-DCT has higher PSNR than MP-DWT throughout the iterations, showing that MP-DCT works very well for homogeneous regions.

Figure 2(b) compares the approximation and extrapolation results of MP-DWT and MP-DCT for face segment after 300 iterations. It showed that with MP-DWT, there are unusual blocks in the extrapolated area. This is because in MP-DWT, sizes of the selected basis are not all the same, and basis functions with little energy over the image segment may be selected to yield unstable reconstructed images. In contrast, in MP-DCT each basis function has the same size and contributes to the entire rectangle  $L$ .



(a) Comparison of PSNR of MP-DCT and MP-DWT



(b) Extrapolation results of MP-DCT and MP-DWT

**Figure 2 Comparison of MP-DWT and MP-DCT**

Thus, we have the following conclusions: 1) MP-DWT usually provides better approximation results possibly because of its multi-resolution nature, and 2) extrapolation results of MP-DWT are usually less smooth than results of MP-DCT. To rectify this problem, we proposed a new basis selection method by using CSMask, the details are given in Section 4.

## 4 COEFFICIENT SELECTION MASK (CSMask)

### 4.1 Description of CSMask

In MP-DWT, most of the basis functions have support of size less than  $N$ . If a basis function has a very small overlap region with  $A$ , the energy of the

basis function inside  $A$  will be very small, resulting in very high values of (7). If such a basis is selected, the extrapolated region of the reconstructed image segment  $g^{(v)}(m, n)$  will be highly unstable. To rectify this problem, we propose a new basis selection method. This method gives priority to the basis functions that have larger overlap with  $A$  by selecting basis only from a smaller region defined by coefficient selection mask (CSMask). Considering the segment in Figure 1(a), the specification of CSMask is illustrated below.

Firstly, we define a binary 'mask' image which has value '1' in  $A$ , and '0' in  $L \setminus A$ . This defines the shape of the image segment, as shown in Figure 3(b). This is used to specify a set of masks at resolutions corresponding to various subbands in the wavelet decomposition. For example, a sub-mask of size  $N/4$  is obtained for wavelet level 1 with one pixel for every  $2 \times 2$  block in the 'mask'. If all 4 pixels in a block are '0', then the corresponding pixel in sub-mask is set to '0'; for blocks containing one or more nonzero pixels, the corresponding pixel in sub-mask is set to '1'. This yields the binary sub-mask as shown in Figure 3(c). Four of these sub-masks are used together to describe the segment shape in the 4 subbands from one level of wavelet decomposition, as shown in Figure 3(d).

This process is repeated iteratively on the submask in the upper left hand corner of Figure 3(d) to generate a set of masks describing the segment shape in different subbands. An example is shown in Figure 3(e) to describe segment shape for 3 level wavelet decomposition. All the masks as in Figure 3(e) are collectively called CSMask. Thus, CSMask is related to the shape of the image segment and to the wavelet decomposition depth as well.



(a) eye (b) Mask (c) SubMask (d) CSMask1 (e) CSMask

**Figure 3 Procedure to get CSMask**

The number of nonzero values in CSMask  $N^1$  is usually somewhat larger than  $M$ , but smaller than  $N^0$  (the number of basis functions which overlap with  $A$ ). For the eye segment in Figure 1(a), with 5-level DWT, and Daubeches filter of length 4 (Db4), we obtained  $N=1024, M=524, N^0=724, N^1=613$ .

By considering only the basis functions giving rise to wavelet coefficients inside the CSMask, those basis that have very little overlap with  $A$  will not be chosen because the corresponding coefficients does not fall into CSMask. And, at each iteration, we only need to compute (6)  $N^1$  times rather than  $N$  times to select the best matching basis function.

## 4.2 Simulation Results with CSMask

From the simulation results, we found that with CSMask, the unusual blocks in the extrapolated part of the reconstructed image are removed, this means the quality of the reconstructed image is improved.

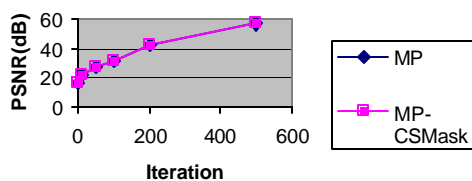
Figure 4 compares the results of MP-DWT with/without CSMask for eye segment after 113 iterations. Without CSMask, at iteration 113, the numerator in (7) is -0.04, while in the denominator, the energy of the corresponding basis function inside the eye segment is 0.000031. The newly selected coefficient is  $-0.04/0.000031 = -1141.17$ , which is very large. Thus, some pixel values in the reconstructed image become very high. With CSMask, at iteration 113, the numerator in (7) is 12.77, and the (denominator) energy of the corresponding basis function inside the eye segment is 1.0, to yield a newly selected coefficient of 12.77, which is well-behaved.



(a)without CSMask (b)with CSMask

**Figure 4 Results of MP-DWT with/without CSMask**

The results also show that with CSMask, there is almost no change in PSNR values (Figure 5). For example, for eye segment, using 5 levels decomposition, with filter Db4, after 100 iterations, without CSMask, the PSNR was 31.06dB, with CSMask, the value was 31.15dB, a change of 0.09dB only. This means that there is almost no change in the quality of the approximation results.



**Figure 5 Comparison of PSNR of MP-DWT with/without CSMask (eye segment, 5-level DWT with filter Db4)**

Basis functions having very little overlap energy inside the image segment give rise to large magnitude coefficients that can consume many bits for very small gain in approximation. Using CSMask, such kind of coefficients can be avoided. The improvement brought about by the use of CSMask has significant beneficial implication for coding the segment.

With the simulations results in Section 3 and Section 4.2 above, we can say that: 1) MP-DWT with CSMask

usually provides better approximation results than MP-DCT (refer to Figure 2(a) and Figure 5); 2) extrapolation results of MP-DWT are usually less smooth than results of MP-DCT. However, with CSMask, this disadvantage can be overcome (refer to Figure 6)



MP-DWT MP-DCT MP-DWT+CSMask

**Figure 6. Comparison of MP-DWT and MP-DCT**

In summary, the CSMask we proposed gives smoother extrapolation results, while keeping the quality of the approximation result of the image segment almost unchanged. In addition, as explained, this could bring significant coding gains.

## 5 CONCLUSION

We compared representation algorithms for arbitrarily shaped image segments using wavelet (MP-DWT) and cosine basis (MP-DCT) which are independent of the boundary of the image segment. Simulation results showed that MP-DWT exhibits high energy compactness, and provides better approximation of the given image segment than MP-DCT for non-homogeneous segments. The disadvantage of MP-DWT is that unusual blocks appear in the extrapolation outside the segment due to the selection of basis with small overlap with the image segment. To overcome this, we defined CSMask to select only those basis functions having larger overlap with the image segment, thereby greatly improving the extrapolation result. We also found that with CSMask, there is almost no change in the approximation results. The improvement brought about by the use of CSMask has significant beneficial implication for coding the segment.

## 6 REFERENCES

- [1] Raj Talluri, et.al., "A Robust, Scalable, Object-Based Video Compression Technique for Very Low Bit-Rate Coding". IEEE Trans. on Circuits and Systems for Video Technology, Vol 7, No.1, pp221-233, Feb. 1997
- [2] M.Kocher and M. Kunt, "A contour-texture approach to picture coding". in Proc. Int. Conf. Acoustics, Speech, Signal Processing, Paris, France, pp436-440, May 1982.
- [3] M.Gilge, T. Engelhardt, and R.Mehlan, "Coding of arbitrarily shaped image segments based on a

- generalized orthogonal transform”, Signal Processing: SPIE-1001, Cambridge, MA, pp178-186, Nov. 1998.
- [4] A. Kaup, T.Aach, “Coding of Segmented Images Using Shape-Independent Basis Functions” Trans. on Image Processing, vol 7, No. 7, pp937-947, July 1998.
- [5] S.G. Mallat, “ Matching pursuits with time-frequency dictionaries” . IEEE Trans. Signal processing, vol 41, pp3397-3415, Dec. 1993.
- [6] S.G. Mallat, “ A Theory for Multiresolution Signal Decomposition: The Wavelet Representation” , IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 11, No.7, July 1989.
- [7] I.Daubechies, “ Image Coding Using Wavelet Transform”, Vol. 1, No. 2, IEEE Trans. On Image Processing, pp205-220, April, 1992.