DCT AND WAVELET BASED REPRESENTATIONS OF ARBITRARILY SHAPED IMAGE SEGMENTS

Ujjaval Y. Desai

Dept. of Electrical Engineering, MIT
Cambridge, MA 02139
email: uydesai@mit.edu

ABSTRACT

This paper describes two efficient algorithms for representing arbitrarily shaped objects. The first representation, called *cosine filling* [1], is a DCT-based approach for obtaining a compact description of objects, making it possible to achieve high compression. A wavelet-based approach is also proposed. This representation not only provides compact description, but also facilitates smooth extension, which can be useful in extrapolation and interpolation of objects.

1. INTRODUCTION

1.1. Arbitrarily shaped objects

Standard image sequence coding methods, such as MPEG, use block-based motion compensation and transform techniques to reduce temporal and spatial redundancies. Higher compression, however, can be achieved with object-based image sequence representation. In an object-based representation scheme, the image sequence is decomposed into objects on the basis of motion, texture, structure, etc. The block diagram of an object-based coder is shown in figure 1. For details on the implementation of an object-based coder, refer to [2]. By efficiently encoding the arbitrarily shaped objects produced by the segmentation module, significantly lower bit rates can be achieved.

It is important to note that arbitrarily shaped objects are not unique to object-based image sequence representation. They are also encountered in model-based image coding and in coding of still-images using segmentation based on intensity or texture [3]. It is widely believed that the next generation of image coders will utilize mid-level vision techniques that include segmentation and object representation. These image coders will inevitably require coding of arbitrarily shaped objects, making the algorithms described in this paper highly applicable.

1.2. Problem description

From the above discussion of the importance of object coding algorithms, it is quite clear that the desired representation must provide a compact description of objects. However, in some applications, one is interested in smooth extension of arbitrarily shaped signals. This problem is encountered in various transform techniques, like the Discrete Wavelet Transform and the DCT, where extension of a finite length signal is necessary for reducing distortions at block boundaries [4]. Smooth extension can also be useful in reconstruction from noisy and incomplete data. Thus, the desired representation must facilitate smooth interpolation an extrapolation of objects while maintaining its compactness. As mentioned in the abstract, our wavelet-based representation has both of these desired attributes.

1.3. Background

Consider the 8 x 8 arbitrarily shaped image shown in figure 2. The unsupported pixels are shown in black. Traditional compression schemes like JPEG can not compress this image efficiently because of the sharp edge present at the boundary between the supported and unsupported regions. In order to develop a better compression technique, let us look at the JPEG algorithm in some detail.

The JPEG algorithm performs the DCT of the input to obtain a set of coefficients which are then quantized and source coded to achieve high compression. The DCT, a frequency-domain transform, uses cosines of different frequencies as its basis functions. Let $B = \{b_1, b_2, ..., b_{64}\}$ be the 8 x 8 DCT basis. The DCT coefficients are obtained by projecting the input onto these basis functions. Since an edge in the spatial domain corresponds to a spread in the frequency domain, the JPEG algorithm provides poor compression of the image in figure 2. An improved compression algorithm is presented below.

2. COSINE FILLING / MATCHING PURSUIT

In the cosine filling algorithm, the input image is decomposed into a set of masked DCT basis functions. This set is obtained by deleting all the unsupported pixels in each of the DCT basis functions. Let $B' = \{b'_1, b'_2, ..., b'_{64}\}$ be the set of masked basis functions. Suppose the input image has M supported pixels, then each b'_i also has M pixels. It is easy to see that B' is an over-complete set. Thus, the main idea behind cosine filling is to decompose the supported region of the input image into an over-complete set. It is important to note that this decomposition is not unique. We have already seen that for high compression, the set of coefficients must have good energy compaction. The cosine

I would like to thank Ted Adelson, John Wang, and Stephane Mallat for their significant contributions. This research was sponsored in part by the Television of Tomorrow project at the MIT Media Laboratory.

filling algorithm, therefore, tries to obtain a decomposition that has near optimal compaction. It can be implemented as a greedy, iterative algorithm that, at each stage of iteration, selects the coefficient that minimizes the energy of the residue. The algorithm is as follows:
Let f be the input signal and Rf be the residue.

Pseudocode:

$Rf := f$	*(initialization)*
repeat	
$\tilde{Rf} := c(i)\frac{b_i'}{\|b_i'\|}$	*(Find the b_i' that best approx. Rf)*
$Rf := Rf - \tilde{Rf}$	*(Update Rf, get residue)*
until $(\|Rf\| < T)$	*(check energy with threshold)*

where

$$c(i) = \frac{Rf^T b_i'}{\|b_i'\|}, \tag{1}$$

and i is given by:

$$i = \arg \max_j \frac{Rf^T b_j'}{\|b_j'\|} \tag{2}$$

Once the set of coefficients is obtained, it is transmitted (after quantization and source encoding) to the receiver. The decoder performs inverse DCT on this set of coefficients to obtain an image which is a harmonic extension of the input signal. In order to reconstruct the input signal, the decoder needs to have the edge information. However, since there exist efficient techniques for encoding the edge information [1], it can be transmitted with the set of coefficients without increasing the bit-rate significantly. [1] presents some compression results for the cosine filling algorithm.

One drawback of this algorithm is that it is not suitable for smooth extension of objects. This is because the basis functions used in the algorithm are cosines, which only produce harmonic extensions. Also, if block-based DCT is used, the extended signal becomes discontinuous at block boundaries.

Before proposing the wavelet approach, we should point out that the greedy, iterative algorithm used in cosine filling is very similar to the *matching pursuit* algorithm [7] for decomposing a signal into an over-complete set. In the rest of the paper, we will use the term *matching pursuit* to refer to the iterative algorithm.

3. WAVELET-BASED REPRESENTATION

In the previous section, we saw that the DCT-based matching pursuit algorithm does not allow for smooth extension of objects, which could be useful in reconstruction from lossy data and in reducing distortions at block boundaries. This extension property, however, can be achieved if matching pursuit is performed using the discrete wavelet basis. This is possible because the wavelet basis functions are much smoother than the DCT basis functions [5], and it has been shown that the wavelet transform has good energy compaction [6]. By exploiting these two features of the wavelet basis, the desired representation can be obtained.

The algorithm is a minor modification of the cosine filling (DCT-based matching pursuit) algorithm. The basis functions b_i are now chosen to be the orthogonal discrete wavelet basis functions. These functions are specified by the QMF filters $h(n)$ and $g(n)$ [5], and J, the total number of scales. Thus $h(n)$ and J are the parameters that are used to design the wavelet basis functions. The wavelet basis functions are then masked using the mask of the input segment. After the masked basis functions have been obtained, they are used to perform matching pursuit on the arbitrarily shaped input image segment.

Figure 3(a) shows a 1-D test signal. Only the non-zero values belong to the supported region. The wavelet-based matching pursuit algorithm is performed on this signal, resulting in a compact set of coefficients. As described in the previous section, these coefficients can be used to reconstruct the signal. The reconstructed signal is shown in figure 3(b). In the supported region, this signal is almost equal to the input signal, however, it takes on very large values in the unsupported region. This is because the matching pursuit algorithm does not put any constraints on the values in the unsupported region. For compression applications, we do not care about the unsupported values since the mask (which is encoded and transmitted to the decoder) can be used to obtain the original arbitrarily shaped signal. However, our goal is to obtain a compact representation in which the reconstructed signal is a smooth extension of the original signal. This can be achieved by modifying the matching pursuit algorithm.

To see why we get spurious behavior in the unsupported region, let us focus on the matching pursuit algorithm. In each iteration of the algorithm, the input signal is projected on each of the masked basis functions to obtain a set of coefficients. The coefficient with the maximum magnitude is then selected from this set. If the selected coefficient belongs to a low-scale basis function, the reconstructed signal has very sharp edges because the low-scale wavelet basis functions correspond to very high frequency signals. On the other hand, if the selected coefficient belongs to a masked basis function with very little energy, the unsupported region of the reconstructed signal will be highly unstable.

In order to rectify this problem, we use a weighted matching pursuit algorithm. In this new algorithm, each masked basis function is assigned a weight proportional to its energy and scale. Hence, the low-scale and low energy masked basis functions are given very low priority, thereby producing a smooth extension of the original signal. Figure 3(c) shows the result of this weighted pursuit algorithm. This algorithm can be easily extended to 2-D. A detailed description of the 2-D discrete orthogonal wavelet basis can be found in [5]. The result of matching pursuit on a 2-D test signal is shown in figure 4.

4. ANALYSIS

The results in the previous section show that the wavelet representation allows for efficient extrapolation and interpolation of objects. We will now consider the compactness of the representation. Figure 5 shows the compaction curve for wavelet-based matching pursuit, wavelet-based weighted matching pursuit, and (8 x 8) DCT-based matching pursuit. The x-axis is the number of (sorted) coefficients used for reconstruction, and the y-axis is the RMSE(Root Mean Square Error) between the original and reconstructed signals over the region of support. The curves suggest that wavelet-based matching pursuit provides better compaction over wavelet-based weighted matching pursuit. This is not surprising since weighted matching pursuit has to give up some compaction in order to achieve smooth extension. However, since the difference in compaction is not considerable, we can conclude that waveled-based weighted matching pursuit provides a very good trade-off between compaction and smooth extension.

On the other hand, (8 x 8) DCT-based matching pursuit performs quite poorly at very low bit rates. This poor performance is due to the fact that the 8 x 8 DCT is a non-scalable block-based transform. However, if the DCT is performed over bigger blocks, its compaction can be improved greatly. But as the block size increases, the computational complexity of DCT-based matching pursuit grows at a very high rate. It can be shown that if a full-image DCT-based matching pursuit is used, the number of basis in the pursuit algorithm is equal to the number of pixels in the image, while for most images, the number of basis needed for wavelet-based pursuit is less than $(2^J)NP$, where J = total number of scales, N = size of the mother wavelet, P = perimeter of the object. For most large images, since the area is much bigger than the perimeter, the wavelet method is much more computationally efficient.

4.1. Future work

More work is needed for implementing an object compression technique based on our representation. The design of an efficient quantizer and source coder for the matching pursuit algorithm turns out to be a challenging problem. Since matching pursuit achieves very high compaction by increasing the entropy in the location of high energy coefficients, better compaction yields poor source coding gain. This trade-off can be optimized by incorporating the quantization process in the matching pursuit algorithm.

5. CONCLUSION

This paper proposes DCT and wavelet based descriptions of arbitrarily shaped objects. The DCT-based description uses matching pursuit (cosine filling), while the wavelet-based representation uses a weighted matching pursuit algorithm. The compaction and smooth extension results for various test signals are presented. We show that the wavelet-based description provides a compact representation that facilitates smooth extension of objects, which is useful in reconstructing signals from incomplete data and in reducing distortions at block boundaries. The results obtained so far show that our weighted matching pursuit algorithm can potentially be useful in compressing arbitrarily shaped image segments.

6. REFERENCES

[1] U. Desai, *Coding of segmented image sequences,* Master of Engineering Thesis, MIT, June 1994.

[2] J. Wang and E. Adelson, *Layered Representation for image sequence coding,* IEEE ICASSP, vol. 5, April 1993, pp. 221-224.

[3] M. Gilge, T. Engelhardt, and R. Mehlan, *Coding of arbitrarily shaped image segments based on a generalized orthogonal transform,* Signal Processing: Image Comm., 1989, pp. 153-180.

[4] H. Barnard, J. Weber, and J. Biemond, *Efficient signal extension for subband/wavelet decomp. of arbitrary length signals,* Proc. SPIE: Visual Comm. and Image Proc., 1993, pp. 966-975.

[5] S. Mallat, *A theory for multiresolution signal decomposition: the wavelet representation,* IEEE Trans. on Pattern Analysis and Machine Intelligence, 1989, pp. 674-693.

[6] J. Shapiro, *Embedded image coding using zerotrees of wavelet coefficients,* IEEE Trans. on Signal Proc.: 41(12), 1993, pp. 3445-3462.

[7] G. Davis, S. Mallat, and Z. Zhang, *Adaptive time-frequency decompositions,* Optical Engineering: 33(7), 1994, pp. 2183-2191.

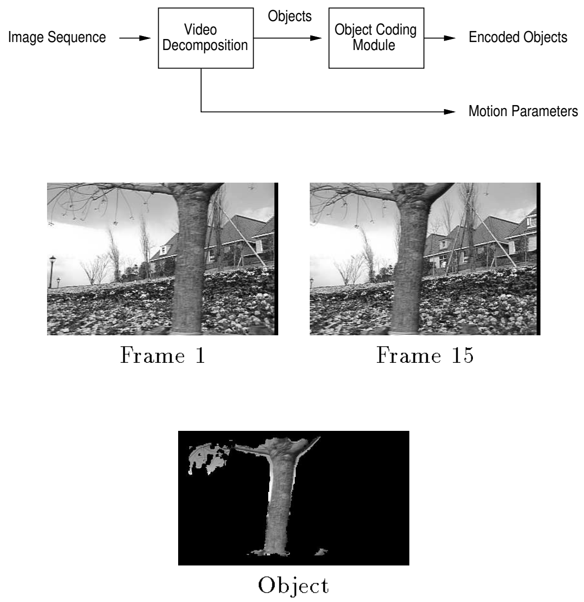

Figure 1: A general block diagram of an object-based coder

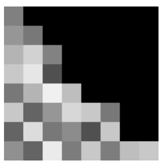

Figure 2: An 8 x 8 image; unsupported pixels are shown in black

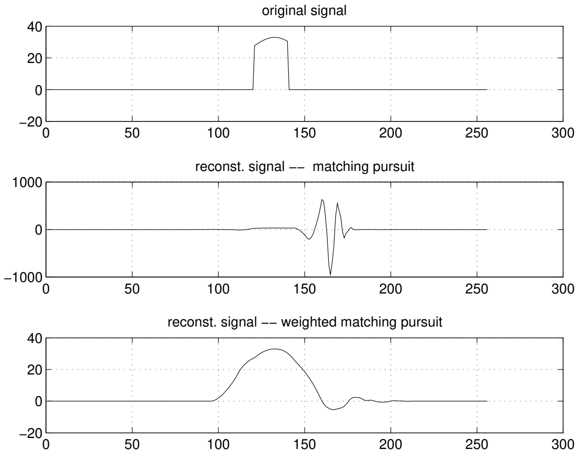

Figure 3: Wavelet-based matching pursuit on a 1-D test signal.

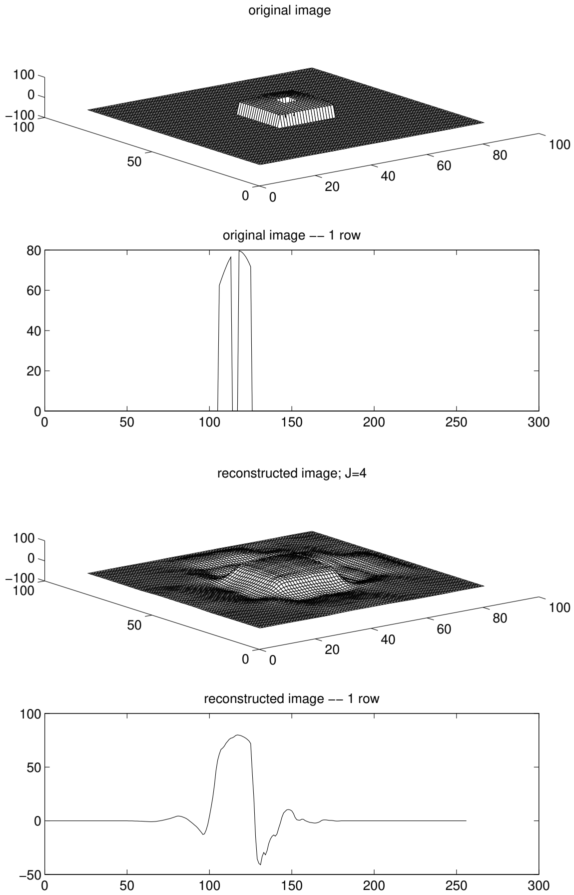

Figure 4: Wavelet-based matching pursuit on a 2-D test signal.

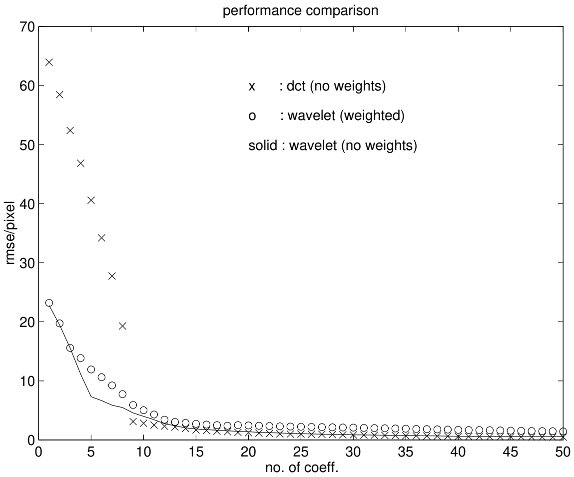

Figure 5: Compaction curve