Designing a Custom Wavelet Packet Image Compression Scheme, with Applications to Fingerprints and Seismic Data*

Mladen Victor Wickerhauser†

December 31, 1997

1 Introduction

No single image compression algorithm can be expected to work well for all images, and designing a transform coding image compression algorithm for a given application is itself a meta-algorithm. Sampling rates, frequency content, and pixel quantization all influence the compressibility of the original data. Subsequent machine or human analyses of the compressed data, or its presentation at various magnifications, all influence the nature and visibility of distortion and artifacts. Thus, algorithms like JPEG [1], established for a "natural" images intended to be viewed by humans, do not satisfy the requirements for compressing fingerprint images intended to be scanned by machines. In that particular example, it was necessary to develop a new algorithm *WSQ* [2].

One procedure focuses on the transform portion of the compression algorithm: the *best basis method* automatically finds a transform which provides the best average compression of a representative set of images, selected from a set of "fast" transforms. A version of this method was used to design the WSQ fingerprint image compression algorithm, while another was used to design compression algorithms for various types of seismic exploration data.

2 Transform coding image compression

The generic transform coding compression scheme is depicted in Figure 1. It consists of three pieces:

- *Transform:* Apply a function, invertible in exact arithmetic, to decorrelate nearby pixels in the image. Do this by decomposing the image into a superposition of independent patterns, producing a sequence of floating-point amplitudes of the new components.
- *Quantize:* Replace the transform amplitudes with (small) integer approximations. This is *lossy*, or non-invertible; all distortion is introduced here.
- *Code:* Rewrite the integer stream into a more efficient alphabet, so as to approach the information-theoretic minimum bit rate. This operation is invertible.

These three steps are depicted at the left of Figure 1; to recover an image from the coded, stored data, they are inverted.

Compression and decompression are judged together by their *rate-distortion curve*. The *Unquantize* block does not in general produce the same amplitudes that were given to the *Quantize* block during compression, but the errors thus introduced can be reduced in exchange for a lower *compression ratio*, which is computed by dividing the size of the input image file by the size of the stored file. This must take into account all of the side information that is needed for reconstruction.

*Research partially supported by NSF, AFOSR, and Southwestern Bell Corporation
†Department of Mathematics, Washington University, St. Louis, Missouri, 63130 USA

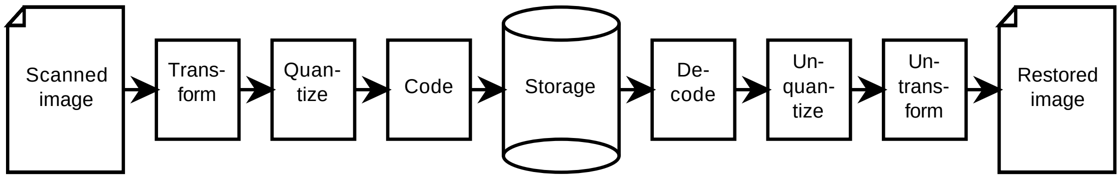

Figure 1: Generic transform coding image compression device and its inverse.

If the coding step is perfectly efficient, the compression ratio is maximized for a given distortion when the transform and quantize steps produce a sequence with minimal entropy. However, since minimal entropy is hard to characterize and harder to achieve, it is better to aim at a broader target: a sequence with almost all of the values being zero. Any such sequence will have low entropy, since its value distribution with be highly peaked at zero, and a best one can be chosen from any set by maximizing the number of zeroes. To produce these sequences, there are large families of wavelet, wavelet packet, and local trigonometric transforms, all of which have low-complexity implementations. All of them are orthogonal or nearly orthogonal, so that their condition number is close to 1. The best is the one which produces the largest fraction of negligibly small amplitudes.

3 Custom transforms

Two fast ways to implement transforms are: splitting into small blocks of pixels and then applying some fast transform to the blocks, or splitting the whole image into frequency subbands by convolving with short filters. Both methods cost $O(P \log P)$ operations for an P-pixel image. Detailed formulas and a proof of the complexity statement is omitted here; they can be found in Reference [6].

In the pixel splitting scheme, the image is cut into blocks small enough so that the intensities of all pixels contained within a block are correlated. This cutting is depicted in the left half of Figure 2. Decorrelation is performed by applying the two-dimensional discrete cosine transform (DCT) to the blocks, as in JPEG [4], or by Malvar transform (LCT) as in the Amoco seismic data compression algorithm [3]. The resulting amplitudes represent spatial frequency components in the blocks. Digitized images are limited in their spectral content, so most of the amplitudes in each block will be negligible. To maximize the proportion of negligible amplitudes, the blocks are chosen as large as possible subject to the constraints that (1) only a few spatial frequencies are present in each block, and (2) describing the block boundaries does not create too much side information.

In the subband splitting scheme, a low-pass and a high-pass filter are used along rows and columns to split the image into four subimages characterized by restricted frequency content. This process is repeated on the subimages, down to some maximum depth of decomposition, resulting in a segmentation of frequency space into subbands. The segmentation used for WSQ [2] is depicted in the right half of Figure 2. The resulting amplitudes again represent spatial frequency components. Again, for images of limited spectral content, most of these amplitudes will be negligible.

4 The joint best basis

Both splitting schemes can be organized as quadtrees to a specified depth, with the selected transform determined by the leaves of a subtree like the one depicted in Figure 3. To choose the subtree and thus the transform, each member of a representative training set of images is decomposed into the complete quadtree of amplitudes. Then the squares of these amplitudes are summed into a sum-of-squares quadtree. Using an information cost function such as "fraction of non-negligible amplitudes", the sum-of-squares quadtree is searched for its *best basis*, which is the one that minimizes this cost ([6], p. 282). Figure 4 depicts this algorithm. The best basis for the Σ quadtree is the *joint best basis* for the training set of images $1, 2, \ldots, N$. That is the transform which produces, on average, the largest number of negligible output coefficients.

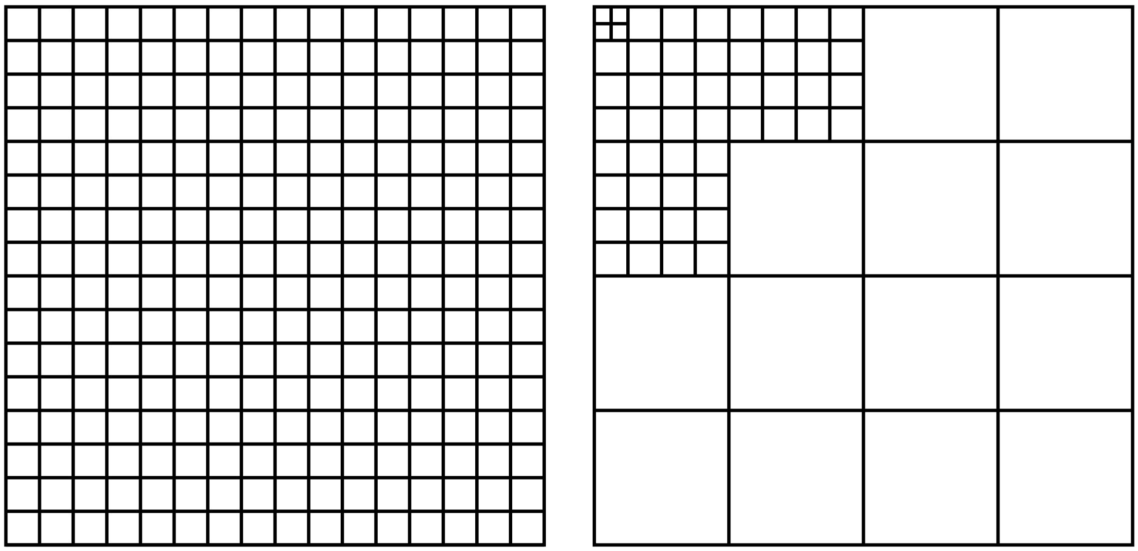

Figure 2: Left: division of a 128×128 pixel image into 8×8 blocks, as in JPEG. Right: Division of an image into orthogonal subbands, as in WSQ.

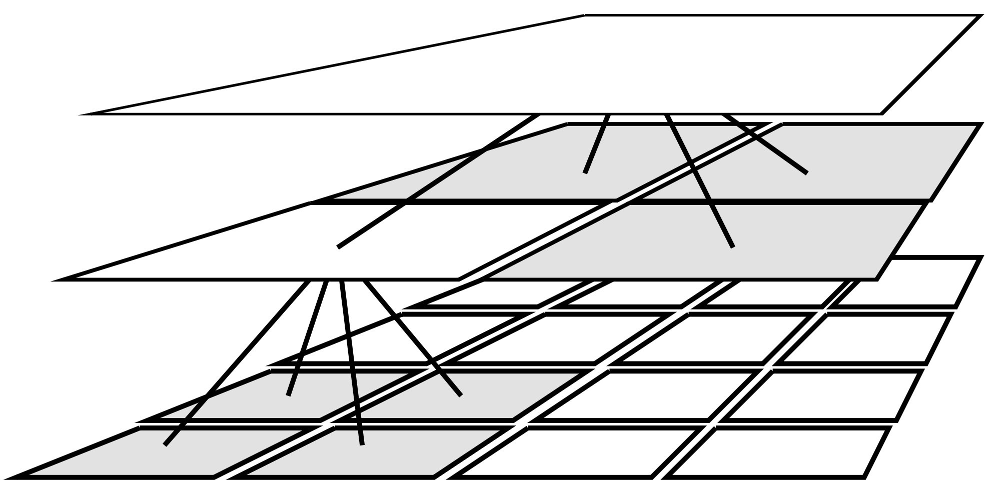

Figure 3: Splitting schemes produces quadtrees; custom bases are determined by the leaves of a subtree such as the one shown here, shaded for emphasis.

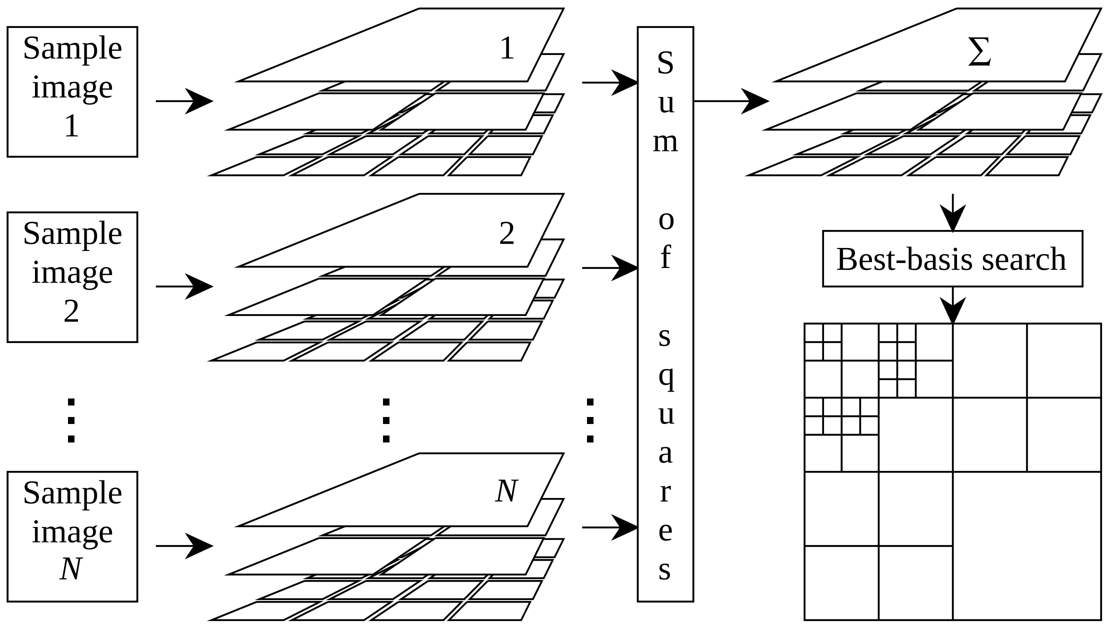

Figure 4: A joint best basis from a class of splitting algorithms is determined by a sample set of N images.

To find the best basis requires examining each coefficient in the quadtree and examining each subband or pixel block at most twice, which means that the complexity is $O(P \log P)$ for P-pixel images. To find the joint best basis requires building the sum-of-squares tree first, which dominates the total complexity with its $O(NP \log P)$ cost for a training set of N P-pixel images.

Of course, the joint best basis transform is only optimal within its own class, and the class is determined by the technical details and mathematical properties of the splitting algorithm. If these constraints were removed and the search performed over all orthonormal transforms, then the joint best basis will be the *Karhunen–Loève* (KL) or *principal orthogonal* basis [5], which is known to be the minimizer of the fraction of non-negligible amplitudes. With the constraints, whose purpose is to speed things up, the chosen transform is just an approximation to KL.

5 Choosing the best transform from multiple classes

There is another meta-algorithm for relaxing the constraints a bit while preserving the speed. Namely, a custom transform can be chosen by checking many classes of splitting algorithms in order to further increase the expected number of negligible coefficients. This scheme was first proposed by Yves Meyer, and is depicted in Figure 5. At the end of each path is a cost figure, the expected fraction of non-negligible coefficients for the training set of images. The path that leads to the lowest cost determines which algorithm should be used to find the custom transform for compressing the images represented by the training set.

Examples of different classes are the different subband splitting schemes associated to different conjugate quadrature filters ([6], Chapter 5 and Appendix C), or the adapted local trigonometric bases determined by different windows ([6], Chapters 3 and 4).

6 Conclusion

Given a training set of images, a transform coding image compression algorithm may be rationally chosen from a class of fast splitting algorithms. The choice criterion is a cost function that, when low, yields high compression ratios for transform coding image compression. The method works for wavelet packet and local trigonometric transforms and thus produces well-conditioned compression and decompression methods of

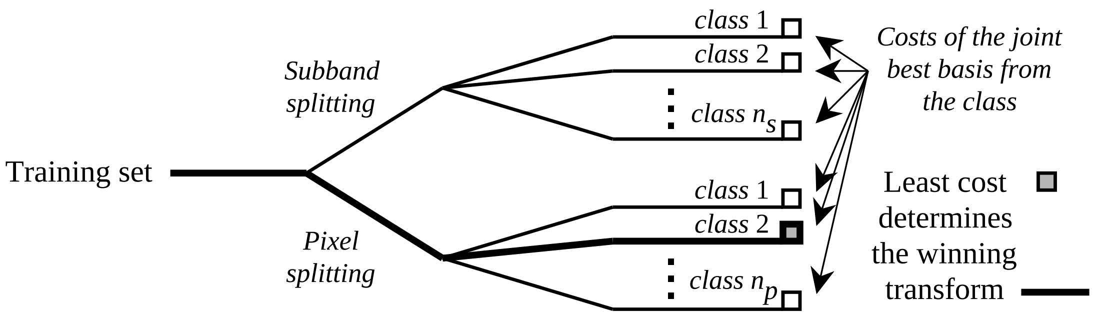

Figure 5: A meta-algorithm for deciding which splitting algorithm to use with a particular class of images.

complexity $O(P \log P)$ for P-pixel images. Searching for the best choice itself costs $O(NP \log P)$, where N is the number of training images.

References

[1] ISO/IEC JTC1 Draft International Standard 10918-1. Digital compression and coding of continuous-tone still images, part 1: Requirements and guidelines. Available from ANSI Sales, (212)642-4900, November 1991. ISO/IEC CD 10918-1 (alternate number SC2 N2215).

[2] IAFIS-IC-0110v2. WSQ gray-scale fingerprint image compression specification. Version 2, US Department of Justice, Federal Bureau of Investigation, 16 February 1993.

[3] Anthony Vassiliou and Mladen Victor Wickerhauser. Comparison of wavelet image coding schemes for seismic data compression. In Akram Aldroubi, Andrew F. Laine, and Michael A. Unser, editors, *Wavelet Applications in Signal and Image Processing V*, volume 3169, page 9. SPIE, SPIE, 27 February 1997.

[4] Gregory K. Wallace. The JPEG still picture compression standard. *Communications of the ACM*, 34:30–44, April 1991.

[5] Mladen Victor Wickerhauser. Fast approximate factor analysis. In Martine J. Silbermann and Hemant D. Tagare, editors, *Curves and Surfaces in Computer Vision and Graphics II*, volume 1610 of *SPIE Proceedings*, pages 23–32, Boston, October 1991. SPIE.

[6] Mladen Victor Wickerhauser. *Adapted Wavelet Analysis from Theory to Software.* AK Peters, Ltd., Wellesley, Massachusetts, 1994.