

An Adaptive Transform for Compression of Mixed Con-tone and Graphics

Michael J. Gormish[†], Nekka Matsuura*, and Takahiro Yagashita*

[†]Ricoh Silicon Valley, Inc., 2882 Sand Hill Road, Suite 115, Menlo Park, CA 94025

*Imaging Technology Development Center, Ricoh Company, Ltd.

email: gormish@rsv.ricoh.com, {yagi,matsuura}@gkn.rp.ricoh.co.jp, www.crc.ricoh.com/CREW/

Abstract: *This paper introduces an adaptive or “text transform” which can be used with a wavelet compression system to provide greater compression/fidelity for images containing a mix of text, line-art, and con-tone regions. The most common source of such images are “page images” from an electronic publishing source. The adaptive transform can provide up to a factor of two compression advantage at similar quality levels for these images, while still considering all regions of the images as images (and not doing an explicit division in to text objects or line-art regions). The advantage of resolution and SNR scalability of modern wavelet compression systems is maintained.*

1 Introduction

Wavelet transforms have become common place for both lossy and lossless image compression. The submissions to the ISO’s JPEG 2000 work item included several wavelet coders capable of lossless compression[1]. For “natural” images wavelet transforms provide the best known lossy compression and very good lossless compression. However, for text, graphics, or other image sources with perfectly sharp edges or regions of constant color the continuous nature of wavelet transforms causes a loss of possible compression (especially compared to spatial domain prediction methods, like the JPEG-LS standard)[2]. The current paper presents a method for using a wavelet transform for continuous portions of the image and a four point “text transform” for graphical parts of the image. The “text transform” provides a “low pass coefficient” which can be used for a reduced resolution version of the image, thus maintaining attractive features of the wavelet transform. The decision about which transform to use can be made on every 2x2 pixel region of the image, it can even change from one scale of the transform to the next.

2 An Adaptive Compression System

This paper describes the changes to the original CREW algorithm[3] which result in improved performance for document images. The changes could be easily incorporated in many wavelet systems.

2.1 Original CREW

Crew contains a wavelet transform, a

sophisticated context model, quantizer, and binary entropy coder[4][5]. The particular binary entropy coder and context model as well as complete detail sufficient for implementation are provided in [6].

CREW uses a reversible integer 2-10 filter which allows lossless compression if no quantization is performed on the coefficients. This wavelet has a two tap low pass analysis filter. The high pass analysis filter for the 2-10 wavelet operates on 4 low pass coefficients and the central two pixels. Let $x(0)$, $x(1)$, $x(2)$,... be the input signal, and let $s(0)$, $s(1)$, and $d(0)$, $d(1)$, be the smooth and the detail outputs respectively. The smooth output and the detail output are the results of applying the low-pass and the high-pass filters, respectively. The 2-10 transform filters are:

$$\begin{cases} s(n) = \left\lfloor \frac{x(2n) + x(2n+1)}{2} \right\rfloor \\ d(n) = x(2n) - x(2n+1) + p(n) \end{cases} \quad (1)$$

where $p(n)$ is defined by

$$\left\lfloor \frac{3s(n-2) - 22s(n-1) + 22s(n+1) - 3s(n+2) + 32}{64} \right\rfloor \quad (2)$$

The inverse of the 2-10 transform filters are,

$$\begin{cases} x(2n) = s(n) + \left\lfloor \frac{d(n) - p(n) + 1}{2} \right\rfloor \\ x(2n+1) = s(n) - \left\lfloor \frac{d(n) - p(n)}{2} \right\rfloor \end{cases} \quad (3)$$

As in all Mallat decomposition image compression systems the transform is applied horizontally and vertically and repeated on the low pass output coefficients.

2.2 CREW with the Text Transform

Figure 1 shows one stage of the decomposition with the wavelet transform augmented by the text transform, and a decision between the text and wavelet transform. For each 2x2 block of input pixels (or prior stage wavelet coefficients), a decision is made to use the wavelet transform or the text transform. On blocks which use the text transform

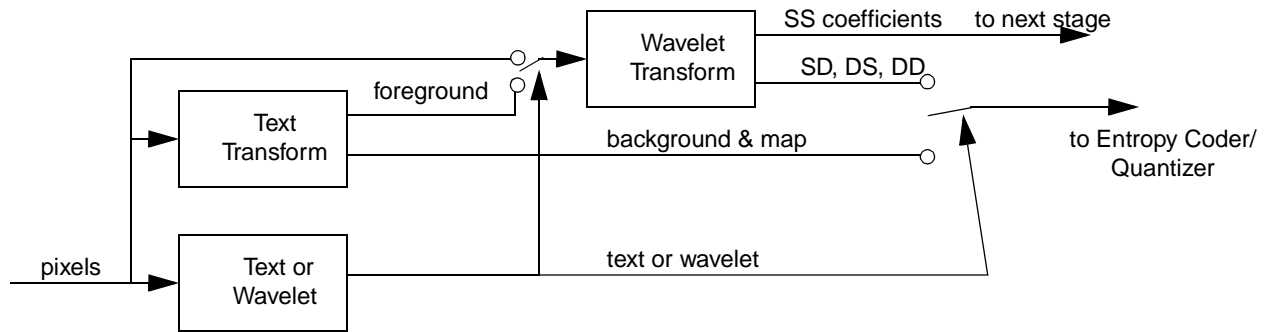


Fig. 1 One Stage of Wavelet and Text Transform

the pixels provided to the wavelet transform are changed to be the “foreground” value resulting from the text transform. The wavelet transform operates normally on the adjusted pixel values. However, the outputs for the SD, DS, and DD coefficients are discarded and replaced with outputs from the text transform (for the filter outputs centered on the 2x2 text transform block only). The process for making this decision is in Section 2.4.

The inverse (not diagramed) involves decoding the bits indicating the text and wavelet blocks (2x2), insuring that the SD, and DD coefficients are zero for text transformed blocks, performing the normal inverse 2-10 transform, and replacing the text blocks with the 2x2 data from the text transform. This process is completely invertible because the inverse 2-10 transform depends only on the smooth (low-pass) coefficients which are exactly the same as used in the forward transform except for the 2x2 block the filter is centered on. This 2x2 block will be replaced by the text transform data (which is also reconstructed losslessly).

2.3 The Text Transform

The text transform takes the values of four samples in a two by two region as input and outputs a “foreground” color (of the same depth as the original pixels), a background color (1 bit), and four map bits.

A text transform decoder reads the map bits and uses the foreground color when the map bit is “1” and the background color when the map bit is “0.” The background color is taken to be the maximum possible value if the background bit is “1” and zero if the background bit is “0.”

Example inputs and outputs of the text transform are shown in Figure 2. At the left are blocks of two by two pixels, on the right side in the “coefficient” domain eight bits are used for the foreground value (assuming an eight bit image), one bit to indicate if

the background is minimum (0) or maximum (1), and four bits to indicate if each pixel is a background (0) or foreground (1) pixel. The second and third rows of the figure show the same block of pixels being transformed in two different ways. These two potential representations allow the encoder to select the image provided by foreground coefficients (which might be produced if only a reduced resolution version of the image was decoded).

2.4 When to apply the Text Transform

The current system uses the text transform only when it provides lossless coding on a 2x2 block. For lossless compression the text transform can only be used on blocks where all four pixels are identical or where there are only two values and one of the two values is either black or white (minimum or maximum value).

In one sense the text transform is a very simple type of a “block truncation code” or BTC, and like

Input: 2x2 Pixel Block	Output: Foreground, Background Bit, Map Bits												
<table border="1"> <tr><td>255</td><td>255</td></tr> <tr><td>28</td><td>28</td></tr> </table>	255	255	28	28	<table border="1"> <tr><td>28</td><td>1</td><td>0</td><td>0</td></tr> <tr><td></td><td></td><td>1</td><td>1</td></tr> </table>	28	1	0	0			1	1
255	255												
28	28												
28	1	0	0										
		1	1										
<table border="1"> <tr><td>255</td><td>0</td></tr> <tr><td>0</td><td>255</td></tr> </table>	255	0	0	255	<table border="1"> <tr><td>0</td><td>1</td><td>0</td><td>1</td></tr> <tr><td></td><td></td><td>1</td><td>0</td></tr> </table>	0	1	0	1			1	0
255	0												
0	255												
0	1	0	1										
		1	0										
<table border="1"> <tr><td>255</td><td>0</td></tr> <tr><td>0</td><td>255</td></tr> </table>	255	0	0	255	<table border="1"> <tr><td>255</td><td>0</td><td>1</td><td>0</td></tr> <tr><td></td><td></td><td>0</td><td>1</td></tr> </table>	255	0	1	0			0	1
255	0												
0	255												
255	0	1	0										
		0	1										
<table border="1"> <tr><td>63</td><td>0</td></tr> <tr><td>63</td><td>0</td></tr> </table>	63	0	63	0	<table border="1"> <tr><td>63</td><td>0</td><td>1</td><td>0</td></tr> <tr><td></td><td></td><td>1</td><td>0</td></tr> </table>	63	0	1	0			1	0
63	0												
63	0												
63	0	1	0										
		1	0										

Fig. 2 Examples of the Text Transform

BTC, the text transform maintains high quality for text-like blocks. While most BTC methods provide lossy compression at a fixed bitrate, the text transform is designed to be used primarily when lossless compression is possible for a 2x2 block. Furthermore the text transform can be repeated in a hierarchical manner like wavelet transforms. Outputs of the text transform are processed by a variable length entropy coder thus providing better compression than most BTC algorithms. Relative to BTC, the variable length code does sacrifice the random access benefits of a fixed length code. However, most modern compression systems provide the needed level of random access with pointers, markers or tiles.

3 Encoding of Wavelet and Text Transform

In the CREW compression system all coefficients are arranged into importance levels, which are essentially bit-planes of coefficients. Each bit-plane is coded by a binary context model and entropy coder. The context model is designed specifically for coding bit-planes of signed wavelet coefficients, but uses the style of context bins and probability estimates from JBIG[7]. This allows maximal compression of the large areas of small coefficients.

The first text transform encoder used exactly the same context model as the wavelet data. One bit of the DD coefficient which could not be “1” under the regular wavelet transform was set to 1 to indicate the two-by-two block was coded with the text transform. The background bit, and the map bits were stored in the remaining bits of the DD coefficient. The DS and DD coefficients were set to zero (which was coded efficiently by the wavelet context model). Finally the SS coefficient was passed to the next scale to be transformed again (or just stored for the final scale).

The current system codes the map of text transform blocks and wavelet blocks prior to other coding. This provides greater compression. The map and background bits are coded as large spatial bitplanes using a small template of nearby pixels similar to JBIG. The use of spatial template for this spatial data provides improved compression with a slight increase in the working memory required.

4 Performance Comparisons

Figure 3 shows portions of the “cmpnd1” image (from the JPEG 2000 test set) compressed to roughly 0.4 bpp both with and without the text transform. The portion selected shows both text and natural image,



Fig. 3 Image Comparison (text transform and wavelet top, wavelet only bottom)

while the image as a whole is two-thirds text. PSNR (which is admittedly an exceedingly poor measure for these images) is 9 dB better when using the text transform at this bit-rate. More importantly the text portions show no perceptual differences compared to the original when the text transform is used and substantial degradation when a wavelet transform is applied. For this image the lossless compression ratio is 44% higher with the text transform than without.

Table 1 contains results for two compound images from the JPEG 2000 test set and 4 compression systems. Each system was used to compress the images losslessly and this produced one bitrate for each image (note that the popular JPEG-baseline with the DCT [8][9] is not even capable of lossless compression and has substantially inferior results to any of the systems presented here). Once the lossless compressed image was created the compression systems were used to decode only a fraction of the compressed image to obtain a particular quality (measured in peak SNR). The JPEG-LS compressor [2] while providing the best lossless rate in all cases cannot provide different quality images by discarding compressed data. A bitrate in bits per pixel is shown for the CREW system without the text transform and with the text transform. Finally results for the most recent trail system for the JPEG 2000 effort [11] are presented. This system based on “EBCOT” collects the latest ideas for coding wavelet coefficients, but could

probably benefit substantially from the adaptive transform presented in this paper to obtain the best overall performance.

Table 1 Compression System Comparison

Image	Quality	CREW (bpp)	Adapt+ CREW (bpp)	JPEG-LS (bpp)	VM3.2a (bpp)
cmpnd1	lossless	2.375	1.329	1.242	2.124
cmpnd1	40dB	0.855	0.380	--	0.848
cmpnd1	35dB	0.601	0.261	--	0.580
cmpnd1	30dB	0.373	0.181	--	0.378
cmpnd2	lossless	2.796	1.911	1.439	2.544
cmpnd2	40dB	0.829	0.627	--	0.706
cmpnd2	35dB	0.503	0.437	--	0.430
cmpnd2	30dB	0.269	0.299	--	0.227

4.1 Alternative approach

Another approach to the problem of mixed con-tone and graphics has been to separate the material into different “component planes.” One plane can store just con-tone information while another plane stores graphical content. Typically there is an additional plane which is a mask indicating how to composite the 2 content planes. This approach has been standardized with the name “Mixed Raster Content” or MRC[10].

MRC is complementary with the present approach. Indeed the CREW system allows the storage of several components so MRC like coding can be performed. MRC provides clear advantages when the source material is available because an entire image can be compressed with overlays coded separately. This allows the wavelet coder to deal with completely smooth data.

Of course effective use of the MRC framework requires both a con-tone and a graphic compressor, both of which could be the CREW system. Also, in many cases the source is not available and there is nothing underneath the graphic overlay, so for simple generation of a compressed image the one-pass adaptive compression is simpler. It may be desirable e.g. with visible water-marks, to not separate the text and graphics into separate components.

5 Conclusions and Future Directions

The adaptive wavelet transform (a combination of the “text transform” and a wavelet filter bank), provides substantially improved quality or compression for a class of mixed images. The described method is available in the CREW system and fully documented at a level sufficient for

implementation[6]. The adaptive style has shown application to a large class of “document” images. The technique could be made even more useful by providing a “text transform” which allows more colors, or more variation in the gray levels currently allowed. Since the decision to use the adaptive style is made at encode time, smart algorithms could be developed to determine when to encode a particular block with the “text transform.”

6 References

- [1] JPEG committee, “JPEG 2000 Algorithm Submissions,” *ISO/IEC JTC1/SC29/WG1 N705*, November 1997.
- [2] JPEG-LS, ISO-14495 *ISO/IEC DIS 14495-1 Information technology -- Lossless and near-lossless compression of continuous-tone still images -- Part 1: Baseline*.
- [3] A. Zandi, M. Boliek, E.L. Schwartz, M.J. Gormish, J.D. Allen, “*CREW Lossless/Lossy Image Compression Contribution to ISO/IEC JTC 1.29.12*,” proposal submitted to the ISO/IEC JTC1/SC29/WG1 (JPEG/JBIG), CRC-TR-9524, 30 June 1995.
- [4] M. Boliek, M. Gormish, E. L. Schwartz, and A. Keith, “Decoding compression with reversible embedded wavelets (CREW) codestreams,” *Journal of Electronic Imaging*, Vol. 7. No. 3, pp. 402-409, July 1998.
- [5] M. J. Gormish, E. L. Schwartz, A. F. Keith, M. P. Boliek, and A. Zandi, “Lossless and nearly lossless compression of high-quality images,” in *Proc. of SPIE*, San Jose, CA, USA, Mar. 1997, vol. 3025, pp. 62-70.
- [6] “Ricoh CREW Image Compression Standard Version 0.13,” 14 July 1998, see <http://www.crc.ricoh.com/CREW/>
- [7] JBIG, *Information technology - Coded representation of picture and audio information - Progressive bi-level image compression*, ISO/IEC IS 11544, 1993.
- [8] W. B. Pennebaker and J. L. Mitchell, *JPEG still image data compression standard*, Van Nostrand Reinhold, New York, NY, 1992.
- [9] JPEG, *Digital Compression and Coding of Continuous-tone Still Images, Part 1: Requirements and guidelines*, ISO/IEC DIS 10918-1, 1992.
- [10] “MRC. Mixed raster content mode,” ITU Recommendation T.44 (1997).
- [11] D. Taubman, “VM3.2A,” ISO/IEC JTC1/SC29/WG1 N1143, February 1999.