On The Adaptive Coefficient Scanning of JPEG XR / HD Photo

Vanessa Testoni[†], Max H. M. Costa[†], Darko Kirovski[‡], and Henrique S. Malvar[‡]
[†] University of Campinas - Unicamp, Campinas, SP, Brazil
[‡] Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
{vtestoni,max}@decom.fee.unicamp.br, {darkok,malvar}@microsoft.com

Abstract

We explore several local and global strategies for adaptive scan ordering of transform coefficients in JPEG XR/HD Photo. This codec applies a global adaptive scan-order heuristic with respect to an effective localized predictor. The global ordering heuristic, although simple, performs as well as localized techniques that are computationally significantly more complex. We conclude that effective localized prediction not only minimizes but also essentially randomizes coefficient residuals, so that a global statistic is sufficient to deliver near-optimal compression performance.

1 Introduction

In a block transform image coder, coefficient scanning is the process of reordering transform coefficients into a linear array before the entropy coding step. As many modern codecs use various prediction methods to reduce the coefficient entropy, in this paper we are particularly concerned with scanning and then encoding the difference between the transform coefficients and their predicted values. Thus, in the remainder of this article, when we refer to a "coefficient," we actually mean its post-prediction residual. In order to increase the entropy coding efficiency, it is desirable that these coefficients be scanned in a "descending on the average" order; that means scanning the most probable nonzero coefficients first in an orderly fashion.

A common approach to do this is to scan coefficients by selecting one out of a collection of precomputed scan patterns and then encode the selection [1, 2, 3]. Modern high-performance codecs such as JPEG XR/HD Photo [4, 5, 6] deploy sophisticated coefficient prediction tools, which have a significant effect on scanning performance. In this paper we explore whether we can improve compression performance in such codecs by localized or hybrid methods for adaptive reordering of post-prediction coefficient remainders.

Our efforts were motivated by the computation of an optimistic loose bound on the performance of an adaptive scan order for HD Photo. We found that when all transform coefficients are scanned in an exact descending order and the reordering permutation is not encoded (thus disregarding its corresponding entropy), the compression rate in HD Photo improves from 24% for relatively smooth images to 10% for images rich in detail, for encoding rates from 0.5 to 4 bits per pixel (bpp), as shown in Fig. 1. Intuitively, for a given compression rate, the compression gain from reordering reduces as coefficient prediction efficiency increases. We note that peak signal-to-noise ratio (PSNR) is not affected, as coefficient scanning does not introduce any additional data loss beyond that from

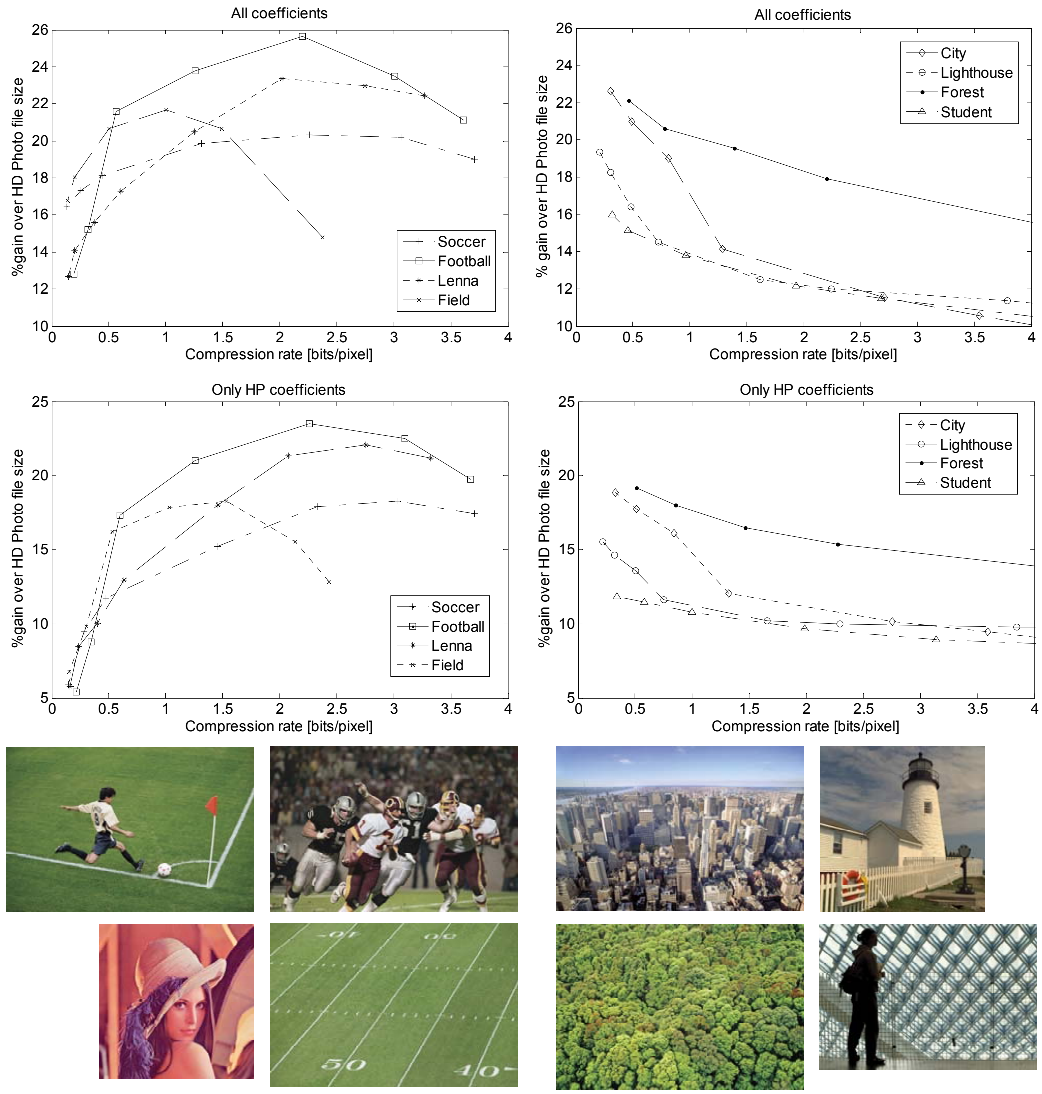

Figure 1: Improvement in compression ratio for specific compression performance enforced over the original HD Photo for the case when all or HP-only coefficients are sorted in descending order.

quantization. In this work we investigate wether we can achieve meaningful performance improvements by new practical coefficient scanning approaches. These approaches do not require changes in either the other HD Photo processing steps or the bitstream definition.

2 JPEG XR / HD Photo

JPEG XR (Joint Photographic Experts Group Extended Range) is a new international standard for image coding based on a Microsoft technology known as HD Photo [4, 5, 6]. HD Photo is a still image file format that offers PSNR performance comparable to JPEG 2000 with computational and memory performance more closely comparable to JPEG [7, 8]. Recent work has addressed improvements to the codec, especially to the lapped transforms and core transforms of JPEG XR [9, 10, 11, 12], but we are not aware of previous work on the efficiency of the coefficient scanning

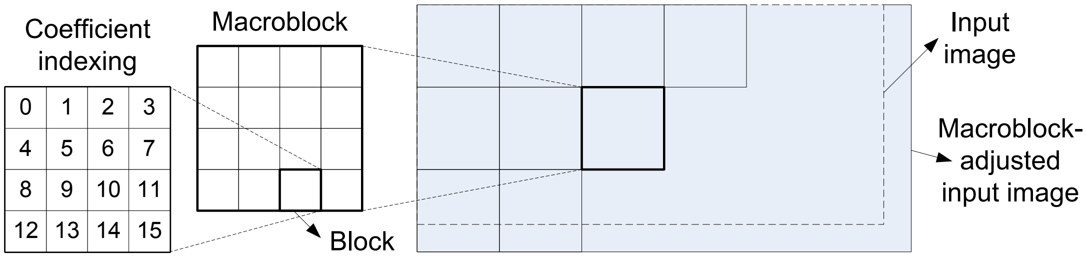

Figure 2: Illustration of the main constituents of the HD Photo image format, macroblocks, blocks, and intra-block coefficient indexing.

heuristic employed in the standard.

HD Photo tiles an adjusted input image into macroblocks, as shown in Fig. 2. Each macroblock is a matrix of 4x4 blocks, each of them a 4x4 pixel matrix. HD Photo then applies to the macroblock a two-stage hierarchical lapped biorthogonal transform, composed of an overlapping operator followed by the PCT (Photo Core Transform). In the first stage, it applies the lapped transform to individual blocks; within a macroblock, the 15 highpass coefficients of each one of the 16 blocks constitute the HP (highpass) subband. In the second stage, for each macroblock, HD Photo groups the 16 DC coefficients of the encompassed 4x4 blocks, and applies the same transform to this DC-only block. As a result, for each macroblock we have: ***i***) the main DC coefficient, ***ii***) the lowpass (LP) subband that consists of the non-DC frequency transform coefficients of the DC-only block, and ***iii***) 16x15 HP coefficients. The three subbands (DC, LP and HP) are then quantized and fed to a prediction stage. Prediction residuals then go through coefficient scanning prior to entropy coding.

2.1 HP/LP Coefficient Energy Distribution

A global scheme for adaptive scanning of coefficients should consider the average energy of coefficients at each index. Fig. 3 shows the average energy of HP subband coefficients depending upon their index. We see that coefficients with indices 10, 5, 12, 1, 2, and 8 are dominant, and thus should be scanned first. The energy distribution is relatively insensitive to the level of detail present in the image.

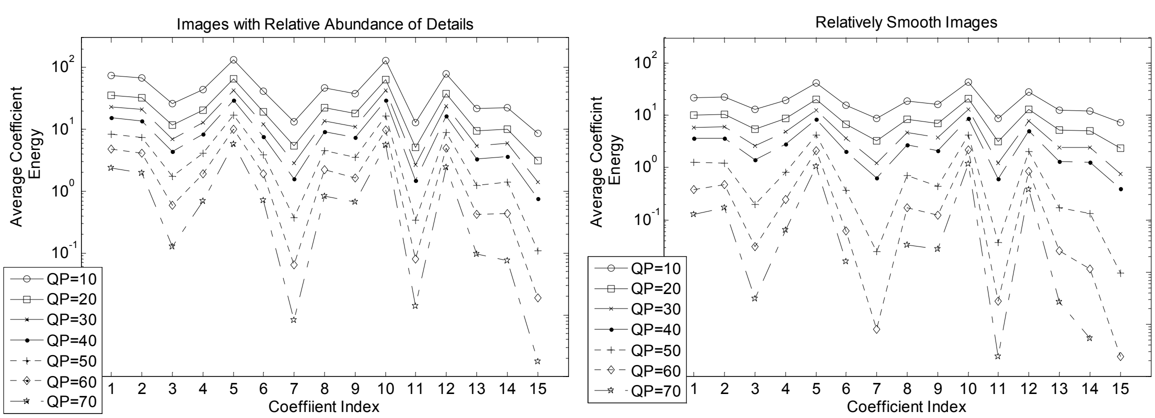

Figure 3: The average HP subband coefficient energy for the two groups of images classified in Figure 1, with HD Photo's Quantization Parameter (QP) varying from 10 to 70.

2.2 Coefficient Scanning

HD Photo scans macroblocks in raster order and their DC coefficients are first scanned across all color planes. The LP coefficients are scanned per color plane using an adaptive heuristic. Before scanning each color plane of the HP subband, a macroblock is divided into 4 8x8 pixel sub-macroblocks. Within each sub-macroblock, the encompassed 4 blocks are scanned in raster order and their HP coefficients are scanned using an adaptive heuristic [6].

The causal adaptation rule in the HD Photo scan order is based on the latest global probability of nonzero scanned coefficients. Two arrays are employed:

- ***order***$[i], i = 1 \ldots 15$, contains the current scan order, i.e., ***order***[3] = 5 means that the coefficient indexed 5 should be scanned third, and
- ***totals***$[i], i = 1 \ldots 15$, contains the number of nonzero coefficients scanned prior to the current block, i.e., ***totals***[3] = 24 means that prior to scanning this block 24 nonzero coefficients have been found at the coefficient indexed 5 (note: ***order***[3] = 5).

When scanning the ***order***[i]-th coefficient, if and only if it is nonzero, the associated ***totals***[i] is incremented by one. After scanning a full block of coefficients, ***totals*** is sorted in decreasing order using a single-pass bubble-sort, and the elements of ***order*** are correspondingly reordered to reflect this sorting. Since, there exist two sets of coefficients scanned, HP and LP, there exist two pairs of corresponding $\{$***order***$_x$,***totals***$_x\}$ arrays, where $x \in \{\mathrm{HP}, \mathrm{LP}\}$.

The arrays are initialized at the start (top left macroblock) of the image. The ***totals*** arrays, both for the HP and LP subband, are always initialized with a constant array of descending values $\mathbf{t}_0 \equiv \{28, 26, 24, \ldots, 0\}$. Three predefined scan patterns are used for the ***order*** arrays. The ***order***$_{HP}$ is set depending on current macroblock's dominant orientation computed as vertical, horizontal, or neutral at the prediction step. The HP scan pattern for a vertically dominant macroblock is $\mathbf{o}_{\mathbb{V}} \equiv \{10, 2, 12, 5, 9, 4, 8, 1, 13, 6, 15, 14, 3, 11, 7\}$, where the indexing is performed according to the schedule shown in Figure 2, and the HP scan pattern for the other two macroblocks (horizontally dominant and with no dominant orientation) is $\mathbf{o}_{\mathbb{H}} \equiv \{5, 10, 12, 1, 2, 8, 4, 6, 9, 3, 14, 13, 7, 11, 15\}$. The ***order***$_{LP}$ array is also always initiated with $\mathbf{o}_{\mathbb{LP}} \equiv \mathbf{o}_{\mathbb{H}}$. One can observe that the default ordering used in HD Photo is in-line with our experiments that quantified average coefficient energies as shown in Fig. 3.

As expected, neither of the scan patterns include the DC coefficient. The scan patterns are invariant across macroblock's color planes. The ***totals*** arrays are reset at every eight macroblocks and at the start of every tile, while the ***order*** arrays are reset only at the start of tiles. Tiles are regular structures grouping macroblocks in arbitrary multiples of 16. Each tile is coded independently. If the image is untiled, the ***order*** arrays are never reset. It is important to note that the periodic resets of ***totals*** are the only action taken in HD Photo's coefficient scanning process to address the content locality for the encoded image.

3 Examples of Localized and Averaging Ordering Heuristics

The HD Photo coefficient scanning process defined in Section 2.2 employs a simple adaptive scan order for all blocks in the image, and an adaptation rule that classifies coefficients simply as zero/nonzero. So, one can argue that such a heuristic is global in its construction, and thus might not handle efficiently abrupt changes in the image content.

Our primary objective is to analyze scan order heuristics that are locally adaptable to the image content. Towards that goal, we created several per-block scan order heuristics that explore the

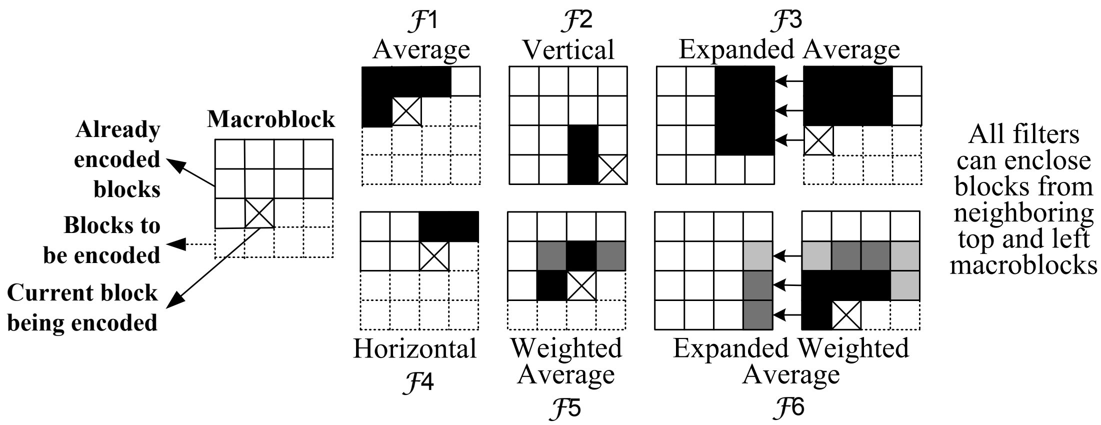

Figure 4: A collection of six localized averaging filters used for scanning coefficients in HD Photo's HP subband.

inter-block correlation, for block sizes as small as 4x4. The main guideline for all heuristics we considered was the premise that the coefficient magnitudes in one block are likely to be similar to the coefficient magnitudes at the same position in neighboring blocks. In other words, we assume that the energy spreading patterns of neighboring blocks are similar.

We experimented with several kinds of averaging filters to affect the scan order based upon localized heuristics. Each filter was used to reorder blocks as follows. In general, a filter $\mathcal{F}$ was defined for each coefficient $c_i(b)$ of the current block b as:

$$\mathcal{F} \equiv \left\{ f_i(b) = \frac{1}{|\mathbb{N}(b)|} \sum_{x \in \mathbb{N}(b)} w(x) c_i(x), i = 1, \ldots, 15 \right\}, \tag{1}$$

where $\mathbb{N}(b)$ was a set of blocks neighboring to b considered for averaging and $w(x), x \in \mathbb{N}(b)$ was an arbitrary real scalar applied as weight to each coefficient individually. We allowed that coefficients coming from each block x from b's neighborhood were weighted distinctly. In order to be able to recover the ordering at the decoder, only blocks that were already parsed by HD Photo's raster order block scanning, were considered for $\mathbb{N}(b)$. After a specific filter was applied, the resulting vector $\mathcal{F}(b) \equiv \{f_1(b), \ldots, f_{15}(b)\}$ was sorted in decreasing order with a resulting permutation π_b. This permutation was applied to sorting the coefficients of b and the result, $\pi_b(b)$, was passed to HD Photo's run-length encoder. Clearly, in order to recover π_b, the decoder would have to compute $\mathcal{F}(b)$ from already decoded blocks, and then establish the correct order of coefficients as $b = \pi_b^{-1}(\pi_b(b))$.

The six averaging filters that performed well in experiments are shown in Fig. 4. Each filter was developed to address blocks with specific correlation patterns. The filters differ in the encompassed "block neighborhood" and in the coefficients weighting. The gray levels in Fig. 4 indicate the coefficient weights. The darker the color, the higher the weight; each increase in gray level indicates a doubling in weight. The "block neighborhood" encloses up to four blocks for the average filters and up to 12 blocks for the expanded average filters $\mathcal{F}_3$ and $\mathcal{F}_6$. That depends on the block position, as the macroblocks and blocks are encoded in raster order. We applied the averaging filters only to the HP subband for two reasons: 1) Fig. 1 shows that we can expect that approximately 90% of the potential gains in the compression ratio are due only to the HP subband, and 2) the coefficients of the HP subband span over smaller localities, which raises the conjecture that localized scan-order heuristics may be more applicable to this subband. Finally, note that the proposed coefficient ordering techniques would marginally increase the computational complexity of the overall codec, both at encoding and decoding.

3.1 Hybrid Techniques

We note that blocks located in highly correlated neighborhoods (relative to the imposed quantization step) are usually well predicted and result in coefficient residuals of low energy. As a consequence, these coefficients can usually be considered too randomized, and thus HD Photo's original scan order with its global adaptation rule performs relatively well for such blocks. On the other hand, blocks that still contain high energy coefficients even after the prediction step, are usually located in areas of the image with high frequency content. For such blocks, localized scan order heuristics often outperform global ones. We now consider that trade-off in more detail.

First, let us define an order-difference metric as follows. For two orders, $\mathbf{x}$ and $\mathbf{y}$, of the same set of elements $\mathbb{Z}$, $\mathbf{x} = \pi_{\mathbf{x}}(\mathbb{Z})$ and $\mathbf{y} = \pi_{\mathbf{y}}(\mathbb{Z})$, we define a distance metric:

$$\Delta(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{|\mathbb{Z}|} \omega(i)|y_i - x_i| \tag{2}$$

where x_i and y_i denote the i-th element of $\mathbf{x}$ and $\mathbf{y}$ respectively and $\omega(i)$ denotes a scaling factor designed to emphasize the importance of ordering high-energy coefficients correctly. We chose:

$$\{\omega(i), i = 1, \ldots, 15\} = \{16, 16, 8, 8, 4, 4, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1\} \tag{3}$$

to reflect upon the exponential rate of coefficient attenuation in most sorted sources. The number of relevant high energy coefficients changes depending upon the enforced bit rate. One can observe from Fig. 3 that the signal energy is concentrated in fewer coefficients for increasing values of HD Photo's quantization parameter (QP). For QPs smaller than 40 (i.e., bit rates higher than 1 bpp), the energy is basically all concentrated in the top eight highest-energy coefficients and for QPs higher than 60 (bit rates smaller than 0.5 bpp), the energy is almost all concentrated only in the top 2 highest-energy coefficients. In fact, in this scenario, optimal scan of these two coefficients in the top two order positions is usually enough to reach the performance obtained when all coefficients are scanned in an exact descending order (shown in Fig. 1). The configuration of the scaling factor $\omega(i)$ reflects these considerations and improves compression especially at low bit rates.

Consider the following experiment: for each block b in a test image, we compute the perfect descending order of its coefficients, $\pi_S(b)$, the orders resulting from the application of all six filters introduced in Fig. 4, $\pi_1(b), \ldots, \pi_6(b)$, and the existing HD Photo scan order $\pi_7(b)$, then announce the filter indexed:

$$j(b) = \arg\min_{i=1\ldots7} \Delta(\pi_i(b), \pi_S(b)) \tag{4}$$

as the "best performing filter." Fig. 5 illustrates for two different QP values of 30 and 70 applied to the "Lighthouse" image from our benchmark:

a) the block energy profile of the resulting coefficients and

b) the indexes of "winning" filters for each specific block – in this case, we considered the global (original HD Photo) as well as all six localized filters defined in Fig. 4.

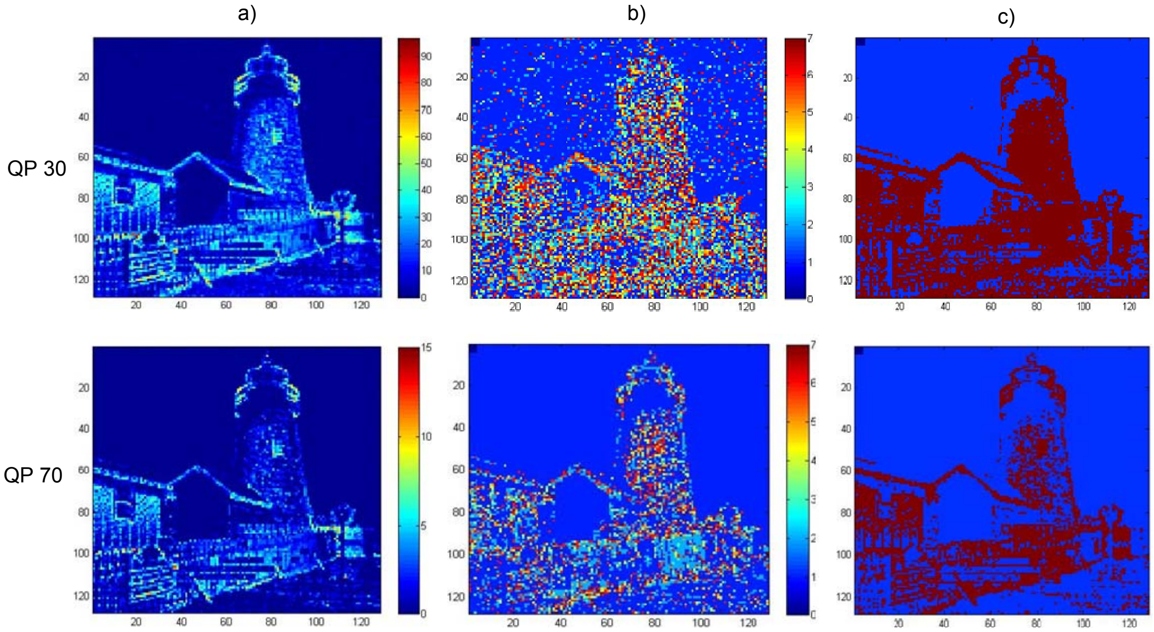

Figure 5: a) Energy amount per 4x4 block areas for the image "Lighthouse" from our benchmark, b) Best performing per-block filters denoted in figures' colorbars and indexed using the following color schedule: 1 - global, original HD Photo, 2 - Average, 3 - Weighted Average, 4 - Horizontal, 5 - Vertical, 6 - Expanded Average and 7 - Weighted Expanded Average, and c) Best performing per-block filters, where blue is the global filter and red is the expanded average filter.

In support of the data presented in Fig. 5(b), Table 1 presents the ratio of blocks where a specific filter performed the "best" (see Eqns.2–4) among the proposed set of filters over all images in our benchmark presented in Fig. 1. Localized filters in the table are marked 1–6 according to Fig. 4.

We see in Fig. 5 that for blocks with low energy, the global filter typically performs well, while the localized filters usually outperform the global filter in all other cases. Based upon the results presented in Table 1 we heuristically select the expanded average filter $\mathcal{F}_3$ as the "best" among the localized filters as it captures the majority of high-energy blocks as the best coefficient scanning tool. When compressing images from Fig. 1 at high bit rates (QPs around 30), localized averaging filters outperform the global heuristic roughly half the time. As expected, this ratio is reduced in low

QP	Original	$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$	$\mathcal{F}_4$	$\mathcal{F}_5$	$\mathcal{F}_6$
30	52.9%	11.4%	5.7%	13.5%	6.3%	5.8%	4.4%
40	66.4%	8.3%	3.8%	9.8%	4.2%	4.4%	3.1%
50	71.7%	6.9%	3.0%	9.7%	3.3%	3.1%	2.3%
60	77.0%	5.4%	2.5%	9.4%	2.4%	2.1%	1.2%
70	84.5%	3.5%	1.3%	8.3%	1.2%	0.9%	0.3%

Table 1: Ratio of blocks where a specific filter performed as the "best" (see Eqns.2–4) among the proposed set of filters over all images in our benchmark presented in Fig. 1. Localized filters are marked 1–6 according to Fig. 4.

bit rate scenarios, but even at QPs around 70, approximately 15% of all blocks are scanned more efficiently with localized filters.

3.2 A Candidate Hybrid Coefficient Scanning Technique

Motivated by the result illustrated in Fig. 5, we unify the global and local scan orders using a simple thresholding technique. Essentially, in our construction, if the energy of a specific block is lower than E_T, then we select the original HD Photo scan order as the "winner" for this block and use it to scan its coefficients. In the alternate case, we use the expanded average filter identified in Table 1 as the "best" localized filter. This is an example of a hybrid strategy that can be reproduced at the decoder. Note that the decision metric defined by Eqn. 4 requires information that is not available at the decoder while decoding a block. The energy threshold E_T could be taken as a constant for all images and established as part of the codec design. Alternatively, it could be adaptively adjusted per image and per quantization parameter at encoding time. The encoder can randomly select a few block samples, empirically compute an adequate value for E_T, and encode it as part of the header of the resulting image file.

We have considered the later approach and have conducted experiments to evaluate its effectiveness. Fig. 5(c) shows the "winning heuristics" when we restrict the filter choice to the original (global) filter or the expanded average filter $\mathcal{F}_3$ and decide among them based upon a precomputed and optimized energy threshold. One can observe that the expanded average filter is selected for many of the blocks coded with localized filters in Fig. 5(b) which suggests that the employed energy thresholds are appropriate. Even though the expanded average filter produces better performance than the original filter for these blocks, it does not cover satisfactorily all localized correlations.

3.3 Empirical Performance Evaluation

We evaluated the hybrid coefficient scanning technique described in Subsections 3.1 and 3.2 using the Microsoft HD Photo Device Porting Kit [4] as a basis for implementation and several images selected from a large database of perceptually diverse content [13]. For the experiments we parameterized HD Photo as follows: no tiling, spatial mode, one-level of overlap in the transformation stage, no skipped subbands and no chroma sub-sampling. Because at relatively low bit rates, most coefficients have low magnitudes, we deemed improving the scan order of chrominance planes as not worthy. Thus, we applied the hybrid scanning technique only to images' luminance planes and scanned the chrominance planes using the original HD Photo scanning order.

Fig. 6 summarizes the results. The visual effects of the compression suite are visible in the top row of plots in the figure, which present the Y-PSNR performance for all images in the benchmark as the compression bit rate is increased to 4 bpp, and two sample images from the database compressed at QP=70. The bottom row of four plots shows for both smooth (first and third) and high-frequency (second and fourth) images in the benchmark. The percentage improvement in compression rate is shown for two schemes. The left two sets of curves refer to an optimistic bound where for each block we applied the "best-of-7" filter to scan coefficients without imposing the overhead to encode the filter selection in HD Photo's bitstream. The right two sets of curves refer to the constructive hybrid scheme described above. In the example of the "Lighthouse" image, the "winning" filters for both tools and QP $\in \{30, 70\}$ are illustrated in Fig. 5. In the "best-of-7" case, the obtained optimistic bound shows that images could be compressed up to 10% better, but we have not been able to develop an efficient encoding of the overhead to capitalize on this potential. The proposed operational (constructive) hybrid scheme resulted in about 1% improvement in the effective compression rate. Therefore, the Y-PSNR performance of the hybrid scheme is highly similar to the HD

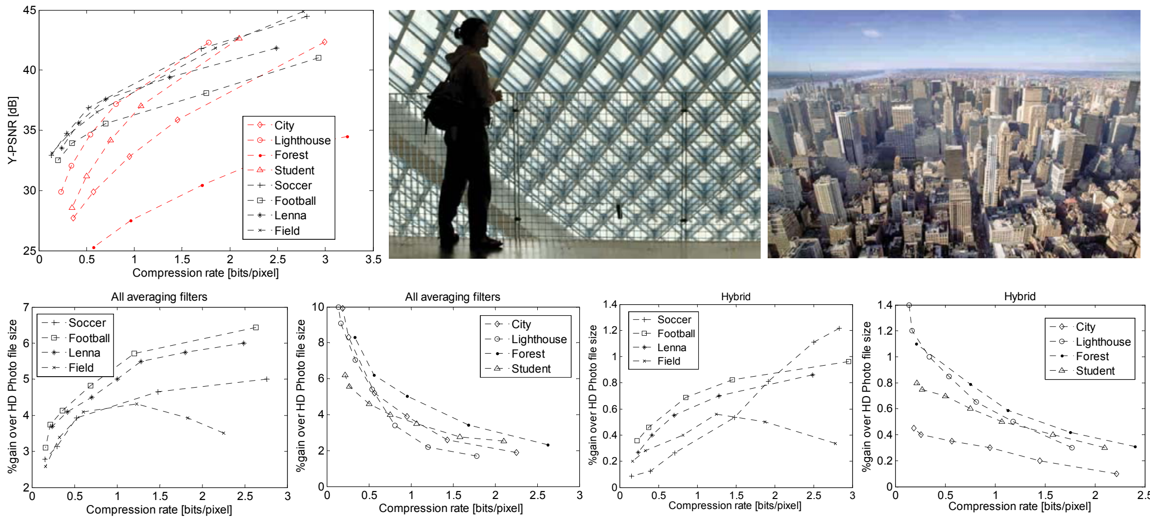

Figure 6: *top left*) Y-PSNR for different HD Photo compression bit rates for our benchmark, *top center and right*) resulting images compressed with HD Photo at QP=70, *bottom row*) percentage of file size reduction due to the bound computed using best-of-7 filter and the proposed constructive hybrid technique at each block; results reported for smooth and high-frequency images from our benchmark.

Photo's performance shown at the top row of plots in Fig. 6.

The level of detail (high frequency content) in images impacts the compression gains significantly. With smooth images the gains tend to rise as the bit rate increases. The opposite behavior is verified in images with relatively abundant detail. A possible explanation for this is that with smooth images and higher bit rates, the filter that is based on localized statistics tends to perform better than the original HD Photo filter, a product of global statistics. Thus, improved coefficient scanning schemes can produce higher gains in the resulting compression rates. Nevertheless, the relative gains obtained by scanning coefficients using the proposed hybrid heuristic are not significant. Therefore, we conclude that the global heuristic present in HD Photo proves to be highly efficient, considering its low computational complexity and the performance obtained using the investigated localized methods.

Conclusion

Coefficient scanning is typically the last stage of processing a compressed signal in a transform coder, before it is fed to the final entropy encoding stage. In modern high-efficiency coders, coefficients are transformed by sophisticated prediction techniques, which decorrelate (whiten) and reduce the variance of prediction residuals, thus reducing the opportunity for specialized scanning to improve compression performance. Our investigation has shown that although optimistic bounds may indicate a potential for performance gains, in a constructive scanning method that balanced localized averaging filters with global statistics, we achieved only about 1% improvement in compression rates, across a wide range of effective bit rates. This helps in validating the high efficiency of the approach to coefficient scanning used in HD Photo.

Acknowledgment

This work was partially supported by a grant from CAPES, Ministry of Education, Brazil.

References

[1] L. Zhang, W. Gao, Q. Wang, and D. Zhao, "Macroblock-Level Adaptive Scan Scheme for Discrete Cosine Transform Coefficients," *Proc. IEEE Int. on Symp. Circuits and Systems*, pp.537–540, 2007.

[2] Y. Ye and M. Karczewicz, "Improved H.264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning," *Proc. IEEE Int. Conf. on Image Processing*, pp.2116–2119, 2008.

[3] Y. Tao, Y. Peng, and Z. Liu, "More scanning patterns for entropy coding for H.264," *Proc. Int. Symp. on Intelligent Signal Processing and Communication Systems*, pp.490–493, 2007.

[4] Microsoft Corp., HD Photo Device Porting Kit. Available: http://www.microsoft.com/whdc/xps/wmphoto.mspx

[5] S. Srinivasan, C. Tu, S.L. Regunathan, and G.J. Sullivan, "HD Photo: a new image coding technology for digital photography," *Proc. SPIE*, vol. 6696, 2007.

[6] ITU-T Rec. T.832 and ISO/IEC 29199-2, *JPEG XR Image Coding System - Part 2: Image Coding Specification*, 2009.

[7] T. Tran, L. Liu, and P. Topiwala, "Performance comparison of leading image codecs: H.264/AVC intra, JPEG 2000, and Microsoft HD Photo," *Proc. SPIE*, vol. 6696, 2007.

[8] F. de Simone, L. Goldmann, V. Baroncini, and T. Ebrahimi, "Subjective Evaluation of JPEG XR Image Compression," *Proc. SPIE*, vol. 7443, 2009.

[9] T. Richter, "Visual Quality Improvement Techniques of HDPhoto/JPEG-XR," *Proc. IEEE Int. Conf. on Image Processing*, pp.2888–2891, 2008.

[10] D. Schonberg, G.J. Sullivan, S. Sun, and Z. Zhou, "Perceptual encoding optimization for JPEG XR image coding using spatially adaptive quantization step size control," *Proc. SPIE*, vol. 7443, 2009.

[11] H. Koichi, T. Hiroshi, O. Hiroyuki; and N. Yukihiro, "An architecture of photo core transform in HD photo coding system for embedded systems of various bandwidths," *Proc. IEEE Asia Pacific Conf. on Circuits and Systems*, pp. 1592–1595, 2008.

[12] C. Tu, S. Srinivasan, G.J. Sullivan, S. Regunathan, and H.S. Malvar, "Low-complexity hierarchical lapped transform for lossy-to-lossless image coding in JPEG XR/HD Photo," *Proc. SPIE*, vol. 7073, 2008.

[13] S. He and D. Kirovski, "A novel visual perceptual model with an application to high-fidelity image annotation," *Proc. IEEE Workshop on Multimedia Signal Processing*, pp.92–97, 2006.