



TITLE: Report on Core Experiment: FLC versus VLC

SOURCE: Tor A. Ramstad, Andrew Perkis, Geir Øien and
Till Halbach

PROJECT: JPEG2000

STATUS: Contribution

**REQUESTED
ACTION:** Provide feed-back and comments

DISTRIBUTION: WG1 Los Angeles meeting and web-site

Contact:
ISO/IEC JTC 1/SC 29/WG 1 Convener - Dr. Daniel T. Lee
Hewlett-Packard Company, 11000 Wolfe Road, MS42U0, Cupertino, California 95014, USA
Tel: +1 408 447 4160, Fax: +1 408 447 2842, E-mail: Daniel_Lee@hp.com

1 Introduction

In the last document concerning this topic [1], fixed-length coding (FLC) has been presented as a good alternative to variable-length coding (VLC) when compressed still images are transmitted over wireless and mobile communication channels. The FLC codec is more robust against channel errors than the arithmetic bitplane codec (VLC), whereas a small degradation in PSNR has to be accepted in the error-free case.

In this paper the technology for the implementation of the FLC is submitted. The developed codec will be called Core Experiment (CE) in the following and has been directly integrated into the VM2.0. All modifications have been included with a definition parameter during compilation and can therefore easily be discarded. Since all the changes for the release of the 2.1 Verification Model (VM) are skipped, this CE can be compared in a straight forward manner to the current release, i.e. VM2.1.

2 Encoder Structure

Fixed-length non-uniform scalar quantizers (SCQ) and adaptive bit allocation have been introduced to improve the robustness against transmission errors, the TCQ and the succeeding arithmetic BP codec are skipped completely. To make clear where the SCQ has been inserted the structure of the encoder is depicted in Figure 2.1.

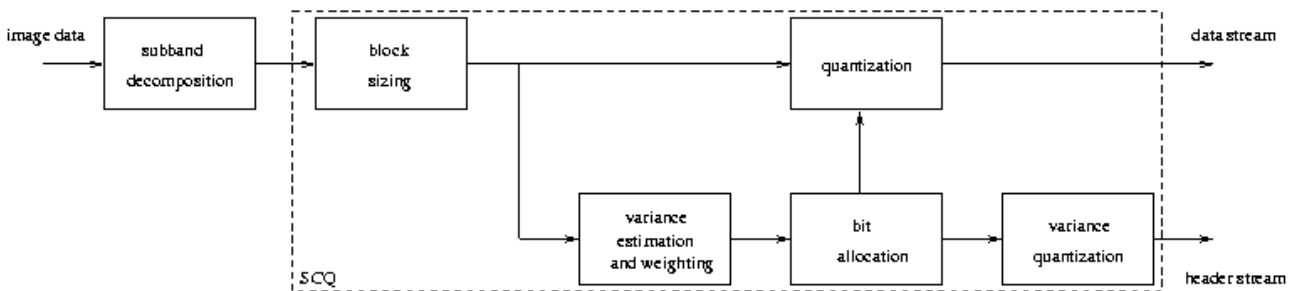


Fig. 2.1: Encoder structure with SCQ

2.1 Subband Decomposition

First the image information is decomposed by the discrete wavelet transform (DWT), thereby splitting it into a certain number of subbands which are called sequences in the following. Different filter coefficients (e.g. 9×7) and decomposition trees (e.g. Mallat) can be used for that purpose. For this paper only the Mallat tree option has been taken into account, the block classification (see later on) for other decomposition trees could easily be adapted.

With the application in mind, lossless coding is not relevant. Only the floating coefficient values (wavelet samples) have therefore been considered. No integer DWT is supported.

2.2 Quantizers

As depicted in Fig. 2.2, the probability density function (PDF) of the higher-frequency sequences tends towards a Laplacian distribution. The quantizers used have been optimized for Laplacian distribution with respect to minimum MSE which results in non-uniform decision levels. The MSE criterion denotes that the quantization error is minimized for a given code word (CW) length. This considers the CWs to be encoded with a bit rate constraint. Typically, the decision levels will be close to each other in regions with a high probability density and far apart in regions with a low one. Even though the resulting fixed-length CWs are

not equiprobable, a following entropy coding should not be employed as it would counteract the advantages of the FLC for robust transmission.

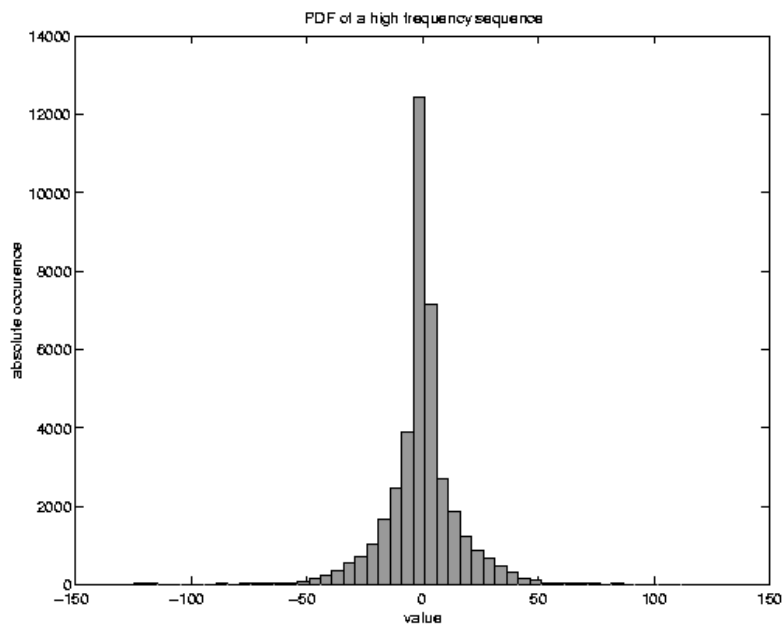


Fig. 2.2: PDF of a higher-frequency subband

In this paper six Lloyd-Max quantizers [3][4] of type mid-tread have been implemented, supporting CW lengths of $n=2$ to $n=7$. The mid-tread quantizers are only suboptimal since one CW is not used. Introducing a mid-rise quantizer in following versions should therefore improve the PSNR slightly but may cause image artifacts in low-activity regions (granular noise). The decision and reconstruction levels have been computed by the help of [5]. In the following figure the characteristics of a Lloyd-Max quantizer with a code word length of $n=4$ is shown.

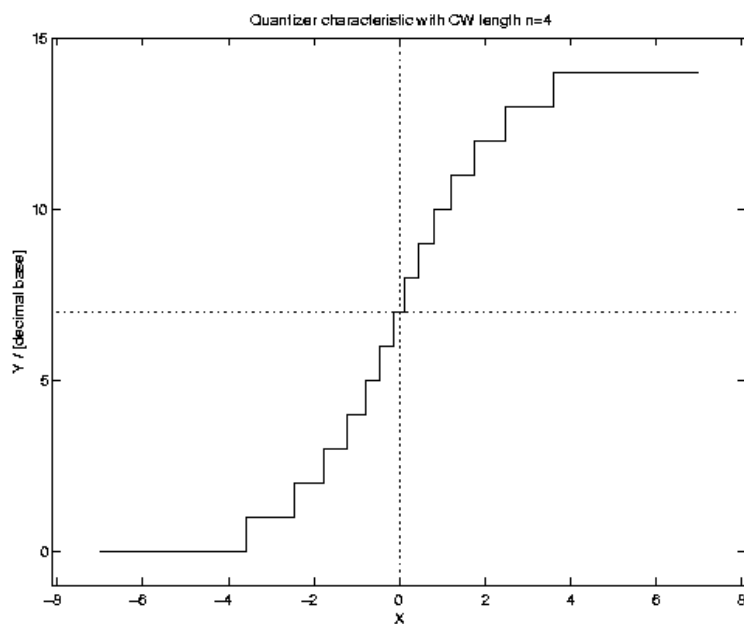


Fig. 2.3: Lloyd-Max quantizer with $n=4$

2.3 Block Classification

Following the approach made in [6], after signal analysis the samples of the sequences are partitioned into blocks. The size of the blocks has to be kept small to exploit local energy variations, tending to a Gaussian distribution of the samples for one block. At the same time the block size should be as large as possible to minimize side information, since the classification of every block (i.e. the bit allocation table) has to be transmitted separately. Thus, the total number of blocks should be small, which also greatly effects the computation time for the bit allocation algorithm. Larger block sizes also tend to the Laplacian distribution where the quantizers have been optimized for.

Due to the non-uniform subband splitting the block sizes have to vary among the sequences. This is because the decimation factors lead to a correspondance of a small number of samples of a lower-frequency block to a larger sample number of a high-frequency block as depicted in Fig. 2.4. It is shown that for the Mallat tree decomposition, the decimation factors for ascending levels are 4, 16, 64 etc. N samples in level K therefore correspond to $4*N$ samples in level $K-1$.

In [7], 16 was found to represent a good compromise between a high degree of local adaptivity and the requirements for bits to be spent on side information for equispaced subbands (in a 64 band split), i.e., subbands of the same size. For non-equispaced subband decomposition as the Mallat tree, a block size of 4 for the highest level, i.e. the lowest-frequency subband (LFS) has been derived as reasonable, leading to a block size of 16, 64, 256 etc. in the lower levels. This results in an average BR of 0.01 bpp for the side information.

The resulting block classification denotes which quantizer to use for the quantization of a certain block. A classification of 0 means that all samples within the block are set to zero, whereas coefficients of a block with a classification of 4 are quantized with CW length $n=4$. A classification of 1 for mid-tread quantizers is not possible. Because of this, occuring classifications are here 0 and [2..7].

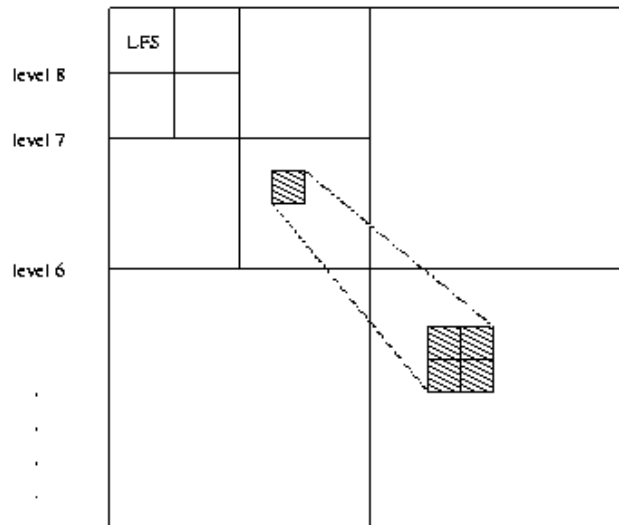


Fig. 2.4: Mallat tree decomposition

2.4 Bit Allocation

After block formation of the sequences the variance of every single block is estimated. Additionally, from the given bit rate (BR), the image size and the expected side information, the number of bits available for quantization, i.e. the bitpool, is found. We then follow the approach of [8], where a simple bit-allocation algorithm for equispaced subbands is derived from the discussion of the well-known formula in [9]. In [10], this algorithm has been extended to non-equispaced subbands. The algorithm is then as follows:

- 1) An estimation of the variance of each block is made.
- 2) The variance is then weighted with the inverse of the respective decimation factors to take the block size into account.
- 3) The largest product (weighted variance) is assigned one bit out of the bitpool (that is hereby reduced one bit per coefficient in the block).
- 4) The product of that respective block is replaced by the product divided by the division (or reduction) factor.
- 5) Continue with step 3 until all bits have been spent.

In [6], a division factor of 1.75 was found to yield the best compromise between image artefacts and PSNR performance. This has also been applied here.

2.5 Variance Quantization

For later transmission as side information, the variances of the blocks now have to be quantized. For this, all variances of a certain classification are collected and then the mean is computed, leading to six different variances. It is found that the quadratic and arithmetic mean show nearly the same performance for that purpose, whereas for the geometric mean the performance degrades approximately 0.289 dB.

Unfortunately, the weighting of the variances with the decimation factors for the bit allocation procedure increases the dynamic range of a variance interval such that the error, made when the wavelet samples are normalized with the quantized variances afterwards, grows. Thus, we collect the variances according to their classification within each decomposition level (s. **Fig. 2.4**), resulting in *number_of_levels* * *number_of_classifications* (here e.g.: $8*6=48$) different quantized variances. By this, we increase the side information slightly (in comparison to a codec where only six quantized variances -one for each classification- are regarded as overhead) but improve the performance approximately by 0.376 dB. The increase of side information is dependant on the number of bits representing one quantized variance. Here, 32 bits (4 bytes) have been used which results in a very high accuracy. The number of bits could therefore be reduced to e.g. 16 without large degradation in performance.

Eventually, the wavelet samples are normalized by the respective standard deviations taking the subband level and block classification into consideration. The blocks then have a variance equal to one in accordance with the optimized quantizers.

2.6 Quantization

Finally, every sample belonging to a block with non-zero bit-classification is quantized according to block classification. The samples of the block for which the bit pool ran out of bits is quantized systematically until all bits are spent, achieving exactly the target bit rate (i.e., the target file size). The remaining samples are set to zero. The classification value (here from 0 to 7) is written in the header bitstream spending 3 bits for every value. A typical header size (including normalization factors and bit allocation table) is 421 bytes. The CWs are added to the data bitstream. Finally, header and data bit stream are merged to the stream written as a file.

3 Decoder Structure

The decoder uses the same quantizer characteristics as the encoder. Because of the low complexity of the decoder in comparison to the encoder the codec could be denoted as asymmetric. First, the bit stream is split into header and data information. Then the normalization factors are read from the bit stream. Block formation is done with respect to the decomposition tree used at the encoder side as well.

3.1 Dequantization

During dequantization, the CWs are replaced by the representation levels of the respective quantizer according to block classification. After that, the samples are denormalized, for the LFS the mean is added, and they are then passed to the inverse wavelet transform.

3.2 Error Treatment

There is no error treatment with a scalar code representation. The introduction of the Gray code representation improves the PSNR about 2.121 dB when the images are transmitted over a BSC with a BER of 0.001, whereas the rise in complexity of the algorithm stays small. In future releases there might be smarter ways of error resilience methods than the Gray code since its performance is as bad as for the binary code when the MSB is affected.

4 Testing

In the following the VM2.1 with standard parameters has to compete with the newly developed CE. The parameters for the VM2.1 are:

```
-r <BR> -a -e 0.002 0 -s <image_width> <image_height> -u8 -mallat -w7x9 -tcq -pbr -ER_mode  
sequence
```

This means using a 7x9 wavelet filter bank, a Mallat decomposition tree and a following classification of samples into sequences of similar statistics. Error resynchronization markers have been inserted in the bit stream for error resilience. For further information the reader is referred to [2].

The parameters for the CE are exactly the same, but some of them (e.g. *-tcq* and *-ER*) are not taken into consideration during the program run since the TCQ is skipped as mentioned before.

4.1 Objective Evaluation

The results have been obtained following the requirements given in the document *Common testing conditions for error resilience core experiments* as agreed by the *Ad-Hoc Group on Error Resilience*. They are from the average of 100 runs of simulation. The channels have been simulated using the software package distributed to the JPEG2000 community by NTT-DoCoMo [11][12]. The header bits have not been affected by the channel errors. In real transmission systems the header must be protected separately with higher levels of protection, e.g. FEC codes.

It can be observed that, in the error free case and in the average, the CE performs 3,94 dB worse than the VM2.1. However, when the image is transmitted over noisy channels such as fading channels and BSC, the CE yields an average improvement in PSNR of 4,27 dB when compared to the current VM.

bike			0.125 bpp			0.5 bpp			
	ber	length		VM2.1	CE	gain	VM2.1	CE	gain
burst	0.001	1	max	19,80	22,39	2,59	17,41	26,83	9,42
	0.001	1	mean	16,47	21,74	5,27	15,63	25,48	9,85
	0.001	1	min	12,50	17,32	4,82	11,97	18,28	6,31
	0.001	1	stddev	1,03	1,05	0,02	1,10	1,91	0,81
	0.001	10	max	25,07	22,46	-2,61	24,79	27,12	2,33
	0.001	10	mean	21,98	22,13	0,15	20,82	26,25	5,43
	0.001	10	min	18,07	13,00	-5,07	17,61	13,32	-4,29
	0.001	10	stddev	1,51	1,04	-0,47	1,33	1,61	0,28
	0.01	1	max	14,82	20,75	5,93	14,88	22,50	7,62
	0.01	1	mean	13,04	18,32	5,28	12,28	19,45	7,17
	0.01	1	min	11,02	14,76	3,74	9,55	15,20	5,65
	0.01	1	stddev	1,14	1,53	0,39	1,22	1,92	0,70
0.01	10	max	18,43	21,89	3,46	17,41	24,10	6,69	
0.01	10	mean	16,26	19,47	3,21	15,46	21,14	5,68	
0.01	10	min	12,24	12,04	-0,20	11,78	12,26	0,48	
0.01	10	stddev	1,06	2,28	1,22	11,18	2,87	-8,31	
bsc	0.001		max	14,54	22,26	7,72	14,28	26,33	12,05
	0.001		mean	12,74	21,48	8,74	12,28	24,69	12,41
	0.001		min	10,97	18,36	7,39	10,97	19,50	8,53
	0.001		stddev	1,08	0,88	-0,20	1,11	1,62	0,51
	0.01		max	12,92	19,73	6,81	10,97	20,84	9,87
	0.01		mean	10,99	17,41	6,42	10,97	10,08	-0,89
	0.01		min	10,97	13,71	2,74	10,97	13,96	2,99
	0.01		stddev	0,19	1,15	0,96	0,00	1,32	1,32
	0.1		max	10,97	11,49	0,52	10,97	11,20	0,23
	0.1		mean	10,97	10,42	-0,55	10,97	10,30	-0,67
	0.1		min	10,97	9,24	-1,73	10,97	9,23	-1,74
	0.1		stddev	0,00	0,39	0,39	0,00	0,37	0,37
efc				25,65	22,46	-3,19	32,81	27,18	-5,63

Table 1: PSNR [dB] results for *bike*

cafe			0.125 bpp			0.5 bpp			
	ber	length		VM2.1	CE	gain	VM2.1	CE	gain
burst	0.001	1	max	16,27	18,23	1,96	15,80	21,58	5,78
	0.001	1	mean	14,69	17,67	2,98	14,17	21,11	6,94
	0.001	1	min	12,06	14,67	2,61	11,88	16,40	4,52
	0.001	1	stddev	0,71	0,69	-0,02	0,70	0,70	0,00
	0.001	10	max	20,59	17,99	-2,60	20,73	21,71	0,98
	0.001	10	mean	18,57	17,26	-1,31	18,18	21,27	3,09
	0.001	10	min	16,47	12,80	-3,67	15,55	13,13	-2,42
	0.001	10	stddev	0,94	1,22	0,28	0,90	0,98	0,08
	0.01	1	max	13,23	17,64	4,41	12,96	19,37	6,41
	0.01	1	mean	12,11	15,59	3,48	11,74	17,64	5,90
	0.01	1	min	10,94	13,44	2,50	10,89	14,49	3,60
	0.01	1	stddev	0,51	0,89	0,38	0,61	1,09	0,48
0.01	10	max	16,18	16,70	0,52	15,58	20,41	4,83	
0.01	10	mean	14,56	15,36	0,80	13,96	18,26	4,30	
0.01	10	min	13,19	13,12	-0,07	11,72	12,55	0,83	
0.01	10	stddev	0,69	0,72	0,03	0,80	1,77	0,97	
bsc	0.001		max	12,76	18,01	5,25	12,33	21,36	9,03
	0.001		mean	11,85	16,98	5,13	11,56	20,88	9,32
	0.001		min	10,89	16,09	5,20	10,89	19,12	8,23
	0.001		stddev	0,49	0,38	-0,11	0,51	0,38	-0,13
	0.01		max	11,84	16,45	4,61	10,89	18,11	7,22
	0.01		mean	10,94	15,32	4,38	10,86	17,12	6,26
	0.01		min	10,06	13,80	3,74	8,16	15,40	7,24
	0.01		stddev	0,24	0,57	0,33	0,27	0,55	0,28
	0.1		max	10,89	10,77	-0,12	10,89	10,58	-0,31
	0.1		mean	10,89	10,25	-0,64	10,89	10,15	-0,74
	0.1		min	10,89	10,13	-0,76	10,89	9,76	-1,13
	0.1		stddev	0,00	0,09	0,09	0,00	0,19	0,19
efc				20,65	16,98	-3,67	26,64	21,74	-4,90

Table 2: PSNR [dB] results for *café*

woman			0.125 bpp			0.5 bpp			
	ber	length		VM2.1	CE	gain	VM2.1	CE	gain
burst	0.001	1	max	22,96	25,57	2,61	22,65	29,19	6,54
	0.001	1	mean	20,98	24,51	3,53	20,05	27,42	7,37
	0.001	1	min	12,77	15,81	3,04	12,84	16,09	3,25
	0.001	1	stddev	1,71	1,95	0,24	2,32	2,89	0,57
	0.001	10	max	27,02	25,60	-1,42	26,10	29,36	3,26
	0.001	10	mean	24,52	25,20	0,68	23,77	28,60	4,83
	0.001	10	min	19,71	11,76	-7,95	18,82	11,87	-6,95
	0.001	10	stddev	1,30	1,56	0,26	1,24	2,09	0,85
	0.01	1	max	20,30	24,29	3,99	19,96	26,00	6,04
	0.01	1	mean	15,91	19,71	3,80	14,46	20,37	5,91
0.01	1	min	12,09	14,16	2,07	12,09	14,30	2,21	
0.01	1	stddev	2,34	2,64	0,30	2,48	3,02	0,54	
0.01	10	max	22,92	25,38	2,46	22,40	27,74	5,34	
0.01	10	mean	20,64	22,11	1,47	19,87	23,50	3,63	
0.01	10	min	12,77	10,84	-1,93	12,60	10,93	-1,67	
0.01	10	stddev	1,90	3,71	1,81	2,40	4,37	1,97	
bsc	0.001		max	18,76	25,49	6,73	17,97	28,96	10,99
	0.001		mean	15,67	23,80	8,13	14,62	26,10	11,48
	0.001		min	12,09	19,40	7,31	12,09	20,02	7,93
	0.001		stddev	1,99	1,64	-0,35	2,17	2,52	0,35
	0.01		max	15,96	22,16	6,20	12,09	23,05	10,96
	0.01		mean	12,18	18,15	5,97	12,09	18,53	6,44
	0.01		min	10,25	14,79	4,54	12,09	14,93	2,84
	0.01		stddev	0,69	1,53	0,84	0,00	1,67	1,67
	0.1		max	12,09	11,94	-0,15	12,09	11,88	-0,21
	0.1		mean	12,09	10,55	-1,54	12,09	10,51	-1,58
0.1		min	12,09	9,13	-2,96	12,09	9,07	-3,02	
0.1		stddev	0,00	0,51	0,51	0,00	0,50	0,50	
efc				27,50	25,60	-1,90	33,74	29,39	-4,35

Table 3: PSNR [dB] results for *woman*

Additionally, according to the new testing conditions, the image *gold* has been included for overhead estimation. Its size is 720x576 pel, resulting in 414,720 pixels and the same file size (in bytes) with a 1 byte per pel representation. The compressed file size with a BR of 0.125 bpp will then be 6,480 bytes. For this image the VM2.1 generates a header (side information) size of 704 bytes which corresponds to a BR of 0.0017 bpp, whereas the CE generates 754 header bytes which means a BR of 0.0018 bpp. The amount of overhead can therefore be denoted as very low for both codecs. For larger images such as *bike*, the side information is 0.0006 bpp for both codecs. So the size of the side information can be considered as the same and thus, the number of errors affecting the transmitted bit stream is comparable.

4.2 Subjective Evaluation

Here, two remarkable cases are picked out to visualize the different performances of the VM2.1 and the CE. The image *bike* has been compressed with a bit rate of 0.5 bpp. The header sizes (VM2.1: 3061 bytes and CE: 3111 bytes) are nearly the same.

Considering first the error-free case, the VM2.1 leads to a PSNR of 32.811 dB, whereas we have to accept a PSNR of 27.185 dB (i.e., a degradation of 5.626 PSNR) for the CE, still yielding a subjective good image quality.

Then, second case, the image is ‘transmitted’ over a BSC with a BER of 0.001, using the channel models mentioned in Chapter 4.1, and decompressed afterwards. The VM2.1 performs poorly, resulting in a PSNR of 12.917 dB, 9.894 dB worse than in the error-free case. The CE, however, performs very satisfactory with a PSNR of 26.023 dB which means a degradation of 1.162 dB compared to the error-free case. The gain of the CE in comparison to the VM2.1 is 13.106 dB. This is because in the VM2.1 most of the sequences are concealed or replaced by the mean (LFS). We therefore lose all high-frequency information, i.e., the details of the image. In the CE, one sees very well that only a single wavelet sample is affected by the channel errors, leading to a ‘smeared’ but finite error area. Most remarkable are the errors in the lower-frequency subbands because they result in larger disturbed image areas. The errors in the higher-frequency subbands are almost negligible. Thus, we receive a highly detailed image in contrast to the VM2.1.

ISO 400



Fig. 4.1: VM2.1 (VLC), EFC, PSNR = 32.811 dB

ISO 400



Fig. 4.2: CE (FLC), EFC, PSNR = 27.185 dB

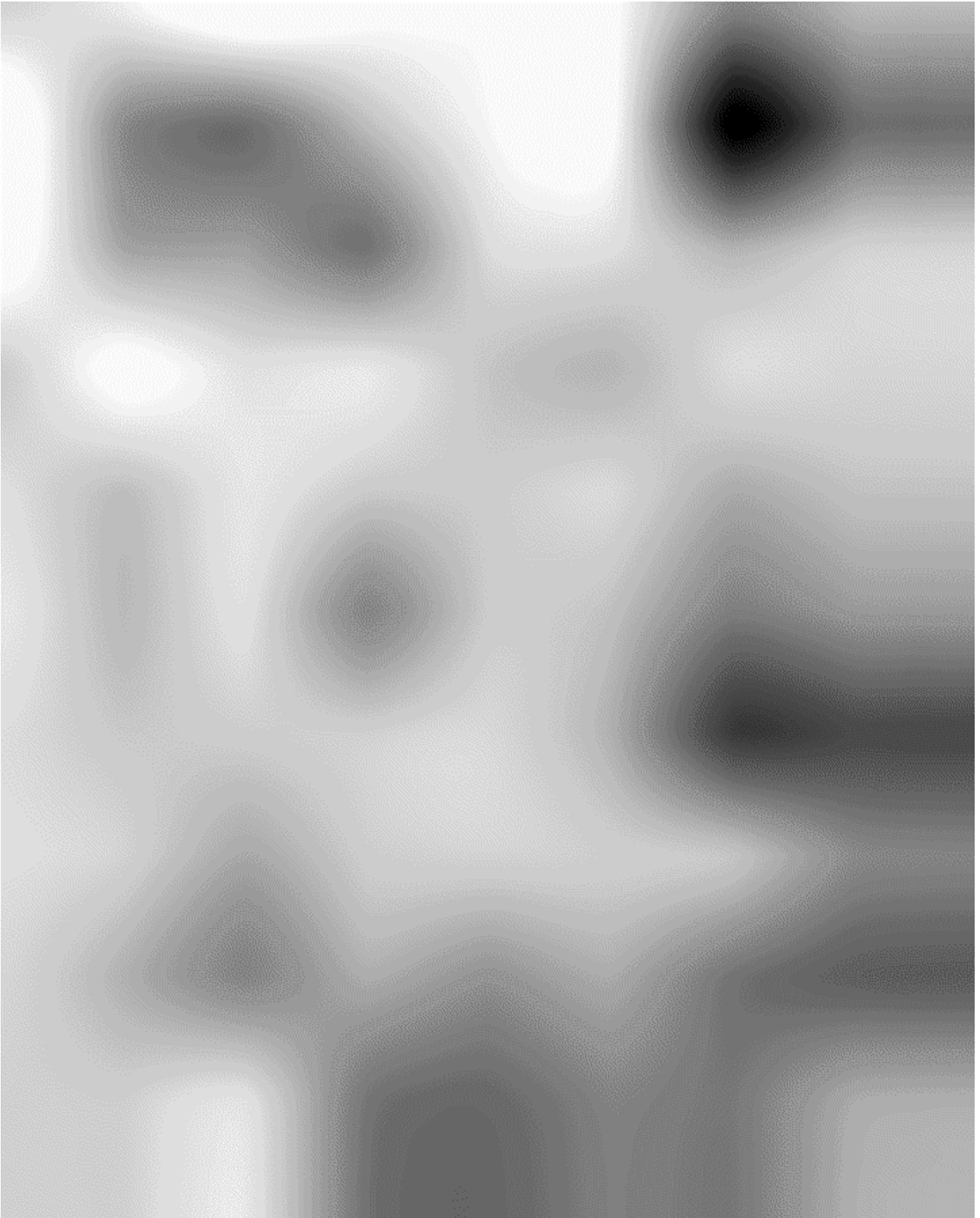


Fig. 4.3:VM2.1 (VLC), BSC, PSNR = 12.971 dB

ISO 400



Fig. 4.4: CE (FLC), BSC, PSNR = 26.023 dB

5 Conclusions and Outlook

It has been shown that when compressed still images are transmitted over noisy channels, a non-uniform scalar (Lloyd-Max) quantizer results in a fixed-length code which is more robust against channel errors than a variable-length code. As expected, the results show that only a single wavelet sample is affected by the channel errors when a FLC is used, which leads to a only limited error area in the image. Introduction of the Gray code representation as error resilience improves the codec performance in noisy environments further. In the error-free case, however, a (significant) degradation in PSNR has to be accepted for the FLC when compared to the VLC which, in the VM2.1, is generated by the arithmetic BP encoder. Despite of the use of resynchronization markers in the VM2.1 synchronization is lost for one sequence if the bit stream is affected by bit errors. Furthermore, all coefficients of a sequence are then set to zero for error resilience which leads to the loss of the total sequence.

Hence, we propose fixed-length non-uniform scalar quantization as the method of choice for applications where the error-free case is of no practical interest, such as wireless and mobile image communications. It should therefore be integrated into the current VM (2.1) as an option. The additional complexity is low compared with the implemented TCQ and, furthermore, the use of non-uniform scalar quantizers does not mean a radical departure from the current standard.

However, further research could still improve the performance of the implemented SCQ. Investigations should be made to validate that the adaptation of the division factor of 1.75 is optimal for the specific purpose. It would also be interesting to know if other block sizes determined by the block formation procedure during the encoding process would lead to better results. As mentioned before, the number of bits spent for quantization of the variances could be optimized as well. And finally, RLC of both quantized variances and bit allocation table could reduce the amount of header overhead further.

List of Abbreviations

BER	Bit Error Rate
BP	Bit Plane
BR	Bit Rate
BSC	Binary Symmetric Channel
CE	Core Experiment
CW	Code Word
DWT	Discrete Wavelet Transform
EFC	Error-Free Case
FEC	Forward Error Correction
FLC	Fixed Length Coding
IEEE	Institute of Electrical and Eletronics Engineers
ISO	International Standardization Organisation
JPEG	Joint Photographic Experts Group
LFS	Lowest-Frequency Subband
MSB	Most Significant Bit
MSE	Mean Square Error
PCM	Pulse Code Modulation
PDF	Probability Density Function
pel	Pixel
PSNR	Peak Signal to Noise Ratio
RLC	Run-Length Coding
SCQ	Scalar Quantization
TCQ	Trellis Coded Quantization
VLC	Variable Length Coding
VM	Verification Model
WG	Working Group

6 References

- [1] Andrew Perkis, Tor A. Ramstad, Geir Øien, Helge Coward, and Till Halbach. *Fixed length coding by nonuniform scalar quantization in JPEG2000*. ISO/IEC JTC 1/SC 29/WG 1 N851. June 1998
- [2] Charilaos Christopoulos. *JPEG2000 Verification model version 2.0/2.1*. ISO/IEC JTC 1/SC 29/WG 1. October 1998
- [3] S.P. Lloyd. *Least squares quantization in PCM*. IEEE Trans. Inf. Theory, IT-28, 129-137 (1982)
- [4] J. Max. *Quantizing for minimum distortion*. IRE Trans. Inf. Theory, IT-6(1), 7-12 (1960)
- [5] Hirohisa Yamaguchi. *Optimum quantization of Laplace density signal and its characteristics*. Electronics and Communications in Japan. Vol. 67-B, No. 5 (1994)
- [6] John H. Husøy. *Subband coding of still images and video*. Ph.D. thesis. The Norwegian Institute of Technology (1991)
- [7] J. B. Nørstebøen. *Subband coding of images*. Master's Thesis. The Norwegian Institute of Technology (1987), in Norwegian
- [8] T. A. Ramstad. *Considerations on quantization and dynamic bit allocation in subband coders*. Proc. ICASSP, pp. 841-844 (1986)
- [9] N. S. Jayant and P. Noll. *Digital Coding of Waveforms*. Englewood Cliffs. Prentice Hall (1984)
- [10] J. Katto and Y. Yasuda. *Performance evaluation of subband coding and optimization of its filters*. Proc. SPIE's Visual Communications and Image Processing, pp. 95-106 (Nov. 1991)
- [11] Toshio Miki, Toshiro Kawahara and Tomoyuki Ohya. *Error pattern generation programs for core experiments on error resilience*. ISO/IEC JTC 1/SC 29/WG 11/MPEG 96/1492
- [12] Toshio Miki, Toshiro Kawahara and Tomoyuki Ohya. ISO/IEC JTC 1/SC 29/WG 1 – WG 1726