# VQ-ENHANCED FRACTAL IMAGE COMPRESSION

*Raouf Hamzaoui, Martin Müller, Dietmar Saupe*

Universität Freiburg, Institut für Informatik, Am Flughafen 17, 79110 Freiburg, Germany

## ABSTRACT

A novel hybrid scheme combining fractal image compression with mean-removed shape-gain vector quantization is presented. The scheme uses a small set of VQ codebook blocks as a block classifier for the domain blocks and as an alternative means of coding when able to provide a satisfying distortion. Our scheme is shown to improve the performance of conventional fractal coding in all its aspects. The rate-distortion curve is ameliorated, and both the encoding and the decoding are faster.

## 1. INTRODUCTION AND PREVIOUS WORK

The analogy between fractal image compression and vector quantization (VQ) has already been stressed by Jacquin in his first publications [1]. In the encoding, both methods use a set of vectors known as codebook blocks to approximate the blocks of the original image. In VQ, the codebook blocks are typically generated from a set of training images. These blocks are needed by the decoder in order to reconstruct an approximation of the original image. In fractal image compression, the codebook is obtained from the original image. However, this codebook is neither present nor needed at the decoder. The encoder finds a contractive image operator whose fixed point is an approximation of the original image. The fixed point is produced in the decoding by iteratively applying the operator to *any* arbitrary initial image.

Though evoked by some authors, combining fractal image compression with VQ has not been deeply investigated. In [2] it is only suggested that fractal coding should be employed for sharp-edge blocks, whereas VQ would be more advantageous for the other blocks. Gharavi-Alkhansari and Huang [3] claim that VQ can be seen as a special case of their generalized transform. However, no emphasis is given to the VQ aspect of their coder.

It has been first pointed out by Lepsøy et al. [4, 5] that a special variant of product code VQ, known as mean-removed shape-gain VQ (MRSG-VQ) [6], is very similar to fractal image compression. In this paper we propose to merge the two methods yielding a hybrid scheme which has the potential to outperform both of them.

In a previous article [7] we implemented a distance based classification scheme that allowed to attain almost full search fidelity by considering only a small number of domain blocks from the whole domain pool. Classes were represented by cluster centers designed either adaptively from the test image or from a set of training images. Clusters were formed by grouping feature vectors of domains and ranges around their corresponding nearest neighbors in the set of cluster centers. The encoding consisted of finding matches inside the same cluster or in the neighboring ones. In this paper we propose to take profit of the precomputed cluster centers which can be considered as an integral part not only of the encoder but also of the decoder. If the least squares approximation of a range block by an affine transformation of the nearest cluster center is "good enough", then the cluster center will serve as a VQ codebook block in the MRSG way. Otherwise, the range block will be encoded by a domain block as in standard fractal image compression. In this way, the bit rates can be improved by a clever choice of the ratio of the number of cluster centers to the number of domain blocks used in the fractal code. For example, if we denote by $N_R$ the number of range blocks and by $N_1$ the number of range blocks VQ encoded, then the hybrid scheme will improve the rate of the fractal code if $N_1 > \frac{N_R}{p-k}$, where $2^p$ is the number of domain blocks and $2^k$ is the number of cluster centers. In the above computation, one bit per range has been included to specify the way a range block has been encoded.

## 2. NOTATIONS AND MATHEMATICAL BACKGROUND

Let us assume that a sampled image is partitioned into nonoverlapping square blocks of size $N \times N$ called range blocks. This is not a restriction since it will be clear how the principles described carry over to more general partitions. Depending on the situation, we consider each range block as a vector $R$ in the linear vector space $\mathbf{R}^n$ where $n = N \times N$ or as a 2-D array.

The domain pool is a collection of square blocks which are typically larger than the ranges and taken also from the image, called domain blocks. The domain pool may be enlarged by including blocks obtained after applying the eight isometries of the square to the domain blocks. By pixel averaging, the size of these blocks is reduced to the size of a range block. The resulting blocks are called codebook blocks. In the encoding process for a range block $R \in \mathbf{R}^n$ a search through the codebook blocks $D_1, \ldots, D_{N_D} \in \mathbf{R}^n$ is required. We let $E(D_i, R)$ denote the least squares error of an approximation of the range block $R$ by an affine transformation of the codebook block $D_i$, i.e., $E(D_i, R) = \min_{a,b \in \mathbf{R}} \|R - (aD_i + bC)\|$, where $C$ is the block of constant intensity $C = (1, \ldots, 1)^T$. In order to ensure the convergence of the decoding process, the scaling factor $a$ is clipped to, e.g., $(-1, 1)$. Then the codebook block $D_i$ which gives the smallest quantized error $E(D_i, R)$ is selected. Now let $O$ be the orthogonal projection operator which maps $\mathbf{R}^n$ onto the orthogonal complement $\mathcal{C}^{\perp}$, where $\mathcal{C}$ is the linear span of $C$. Keeping the same notations we have the theorem [8], which sets the mathematical basis for our clustering algorithm.

**Theorem 1** *Let $n \geq 2$ and $X = \mathbf{R}^n \backslash \mathcal{C}$. Define the function $\Delta : X \times X \to [0, \sqrt{2}]$ by $\Delta(D, R) = \min(||\phi(R) + \phi(D)||, ||\phi(R) - \phi(D)||)$, where $\phi(Z) = \frac{OZ}{||OZ||}$. For $D, R \in X$ the least squares error $E(D, R)$ is given by $E(D, R) = \langle R, \phi(R) \rangle \, g(\Delta(D, R))$ where $g(\Delta) = \Delta\sqrt{1 - \frac{\Delta^2}{4}}$.*

It follows from the theorem that we may replace the computation and minimization of $N_D$ least squares errors $E(D_i, R)$ by the search for the nearest neighbor of $\phi(R) \in \mathbf{R}^n$ in the set of $2N_D$ vectors $\pm\phi(D_i) \in \mathbf{R}^n$.

## 3. THE ALGORITHM

The block average intensity classification [9] is an ingenious way to divide square blocks into three main classes. In the following we describe this classification, since it will be used in our algorithm. For every block $B \in \mathbf{R}^n$ there exists a unique isometry $I_B : \mathbf{R}^n \to \mathbf{R}^n$ (corresponding, when block $B$ is seen as a 2D-array, to one of the 8 isometries of the square) that transforms $B$ such that the average pixel intensities $B_i^*$, $i = 1, 2, 3, 4$ of the four quadrants (upper left, upper right, lower left and lower right) of the 2D-array representation of its transform $B^* = I_B(B)$ are ordered in one of the three canonical positions

$$\begin{aligned}
&\text{Major class 1: } B_1^* \geq B_2^* \geq B_3^* \geq B_4^*,\\
&\text{Major class 2: } B_1^* \geq B_2^* \geq B_4^* \geq B_3^*,\\
&\text{Major class 3: } B_1^* \geq B_4^* \geq B_2^* \geq B_3^*.
\end{aligned}$$

Using the procedure described in [7], a set of fixed cluster centers is designed from several training images. It can be proved that these cluster centers belong to the space $\mathcal{C}^\perp$, i.e., they have zero mean. These cluster centers are then normalized and denoted by $\{m_1, \ldots, m_{N_m}\}$. The clustering of the image codebook consists of mapping each feature vector $\phi(I_{D_i}(D_i))$ to its nearest normalized cluster center. To encode a range block $R$ and if we want to obtain only positive scaling factors we consider codebook blocks whose feature vectors lie in the cluster whose center is the nearest neighbor of $\phi(I_R(R))$. In our study [7] we demonstrated the superiority of this approach as compared to the previous state-of-the-art classification method of [9]. We now introduce our VQ hybrid scheme which provides a further enhancement. Let $c = \arg\min_i ||\phi(I_R(R)) - m_i||$. If

$$\frac{1}{\sqrt{n}} E(R, I_R^{-1}(m_c)) \leq \delta \tag{1}$$

or if the inequality

$$E(R, I_R^{-1}(m_c)) \leq (1 + \epsilon) E(R, I_R^{-1}(I_{D_i}(D_i))) \tag{2}$$

holds for all codebook blocks $D_i$ in the cluster with center $m_c$, then cluster center $m_c$ is retained for VQ encoding of the range block. Here $\epsilon$ and $\delta$ are parameters of our method. If neither condition (1) nor condition (2) are satisfied, then the range block will be encoded by the codebook block $D$ minimizing $E(R, I_R^{-1}(I_{D_i}(D_i)))$ in the cluster with center $m_c$. The code for a range block consists of the index $c$, the isometry $I_R^{-1}$ (or the address of $D$ and the isometry $I_R^{-1} \circ I_D$), and the corresponding scaling factor $a$ and offset $b$. In case the scaling factor is zero, the index of the cluster center (or the address of $D$) and the isometry are not stored. Note that for VQ encoded range blocks no contractivity restriction on the scaling factor is required and that the offset reduces to the mean of the block. Note also that Theorem 1 ensures that $m_c$ is the cluster center that can best approximate the range block in the least squares sense.[1] Clearly, our new scheme reduces the complexity of the already fast algorithm described in [7] since the search for a codebook block is started only if the cluster center was not able to provide an acceptable approximation. As explained in [7], the search for a matching codebook block can be extended to neighboring clusters. Also to include encodings with negative scaling factors we apply the same technique to the vector $\phi(I_{-R}(-R))$.

The decoding proceeds as with a conventional fractal decoder, i.e., through iterations from any initial image with the advantage, however, that the reconstruction of the VQ encoded range regions is already achieved after the first iteration. Thus, in addition to a less complex decoder, a faster convergence is expected.

## 4. RESULTS AND DISCUSSION

With Kohonen's SOM [10] we designed a set of 256 fixed cluster centers corresponding to the nodes of a $16 \times 16$ rectangular array. The training sequence was generated from 9 images of size $512 \times 512$. In a first experiment, we compared the encoding of several $512 \times 512$ images that were not used in the training sequence by our distance classification based fractal coder [7] and by the hybrid scheme. For the two schemes both positive and negative scaling factors were considered (option "Both" in [7]). A fixed range size ($4 \times 4$) was selected to better analyze the performance of the new scheme. Thus we had 16384 nonoverlapping range blocks and 4096 nonoverlapping domain blocks having twice the range size. We opted for a 4-class search since it provides an acceptable trade-off between speed and fidelity. We spent respectively 5 and 7 bits for the scaling factor and the offset. All coefficients were uniformly quantized. The maximum value of the scaling factor was set to 1 for the fractal part.

Table 1 shows the results obtained when varying the tolerance bounds $\epsilon$ and $\delta$ for the $512 \times 512$ Boat image. The PSNR was computed without postprocessing. All the times reported were measured in seconds on an SGI Indigo2 running an MIPS R4400 150 MHz processor.

Figure 1 shows the bit plane indicating the way each range block was encoded. This bit plane corresponds to the sequence of 16384 bits that must be transmitted to the decoder accordingly.

We have also noticed that the convergence of the decoding was significantly faster for the hybrid scheme (see Table 2).

The results show that for a large range of the tolerance bounds $\epsilon$ and $\delta$, all aspects of fractal coding are improved.

[1] Of course quantization effects may make us choose a suboptimal cluster center.

| $\epsilon$ | $\delta$ | $N_1$ | Comp | PSNR | Time |
|---|---|---|---|---|---|
| 0.1 | 2 | 9075 | 5.67 | 36.72 | 49 |
| | 3 | 12662 | 6.27 | 36.54 | 39 |
| 0.15 | 2 | 9947 | 5.72 | 36.64 | 49 |
| | 3 | 13313 | 6.34 | 36.47 | 39 |
| 0.2 | 2 | 10786 | 5.78 | 36.53 | 49 |
| | 3 | 13937 | 6.40 | 36.38 | 39 |
| 0.25 | 2 | 11523 | 5.85 | 36.44 | 49 |
| | 3 | 14441 | 6.46 | 36.29 | 39 |

Table 1: Performance of the hybrid scheme for the $512 \times 512$ Boat image for different settings of the tolerance bounds $\epsilon$ and $\delta$. Here $N_1$ is the number of VQ encoded range blocks. The fourth column indicates the compression ratio. The extra bit per range specifying the encoding way (VQ or fractal) is taken into account. Lempel-Ziv coding (Gzip) of the corresponding bit plane is used. The last two columns show the PSNR in dB and encoding time in seconds. In comparison, the distance based fractal code necessitated 57 seconds for the encoding, had a compression ratio of 4.74 and provided a PSNR of 36.29 dB. A full search (256-class search) would increase the PSNR to 36.52 dB.

| Iteration | Hybrid | Fractal |
|---|---|---|
| 1 | 22.46 | 11.56 |
| 3 | 35.48 | 21.48 |
| 5 | 36.47 | 35.20 |
| 9 | - | 36.29 |

Table 2: Convergence of the decoding for the two methods with $\epsilon = 0.15$ and $\delta = 3$ for the $512 \times 512$ Boat image. The last two columns give the PSNR in dB for the hybrid scheme and the distance based fractal scheme. The last two rows correspond to the iteration step at which the convergence of the hybrid scheme and the pure fractal scheme occurred.

The compression ratio is higher, the PSNR is better, and both the encoding and the decoding are faster.

We point out that encoding all range blocks with only cluster centers gave a PSNR of 35.81 dB. Thus it seems that the domain pool was of vital importance for an accurate encoding of some parts of the image.

Our method was also successfully implemented with a three-level quadtree scheme using two additional sets of 256 VQ cluster centers of size $8 \times 8$ and $16 \times 16$. The classification was carried out at the lowest level of the quadtree, i.e., for the 16-dimensional feature vectors $\phi(\widehat{I_B(B)})^2$ but computations of the least squares errors were done for blocks at the real size. Here $\hat{B}$ denotes the block obtained from $B$ after reducing its size to $4 \times 4$ by pixel averaging. The core of the quadtree algorithm for the hybrid scheme can be described as follows. For a given range block $R \in \mathbf{R}^n$

[2]For the sake of simplicity, we discuss the case of positive scaling factors only.

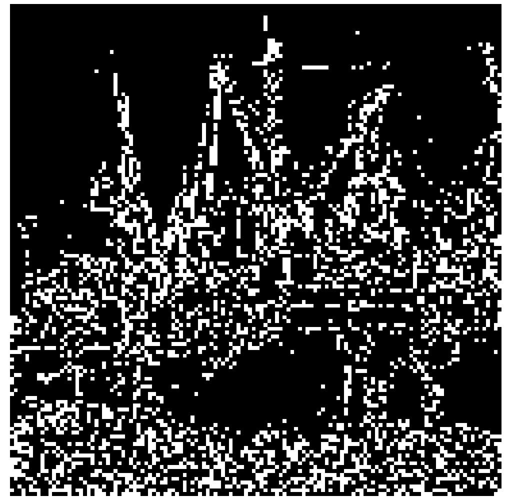

Figure 1: Bit plane for the $512 \times 512$ Boat image with $\epsilon = 0.15$ and $\delta = 3$. VQ encoded range blocks are shown in black while fractal encoded range blocks are shown in white.

and root mean square tolerance levels $\delta$ and $t$, if

$$\frac{1}{\sqrt{n}} E(R, I_R^{-1}(m_c)) \leq \delta \tag{3}$$

then range block $R$ was encoded with cluster center $m_c$. For range blocks having a size larger than $4 \times 4$, $m_c$ was the cluster center that yielded the best root mean square approximation from the $s$ cluster centers whose feature vectors $\phi(\widehat{m_i})$, $i = 1, \ldots, s$ were the $s$ nearest neighbors of $\phi(\widehat{I_R(R)})$ in 16-dimensional feature space. If condition (3) was not satisfied, then, letting $D$ denote the best candidate codebook block in all the classes searched, if

$$\frac{1}{\sqrt{n}} E(R, I_R^{-1}(I_D(D))) > t \tag{4}$$

then the range block was not accepted and partitioned into 4 smaller blocks. In the other case, if

$$E(R, I_R^{-1}(m_c)) \leq (1+\epsilon) E(R, I_R^{-1}(I_D(D)))$$

then range block $R$ was encoded with cluster center $m_c$ and otherwise it was encoded with codebook block $D$. Of course, at the lowest level of the quadtree, test (4) is not needed.

Figure 2 compares the rate-distortion performance of the $512 \times 512$ Lenna image for the following schemes.

1. Fisher's code where 24 classes from a total of 72 were searched.
2. Our hybrid scheme where 4 classes out of 256 were searched and $s = 4$ cluster centers were considered.
3. A MRSG-VQ quadtree scheme with three shape codebooks of size 256 each. The results were obtained directly from the hybrid scheme by fixing $t$ to a negative value in (4), $\epsilon$ to a large value and $s$ to 20. In this manner only cluster centers were used. The extra bit per range was of course not used here.

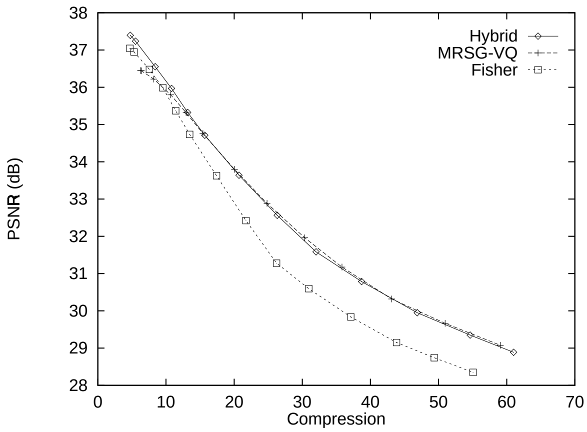


Figure 2: PSNR vs. compression ratio for the 512 × 512 Lenna image.

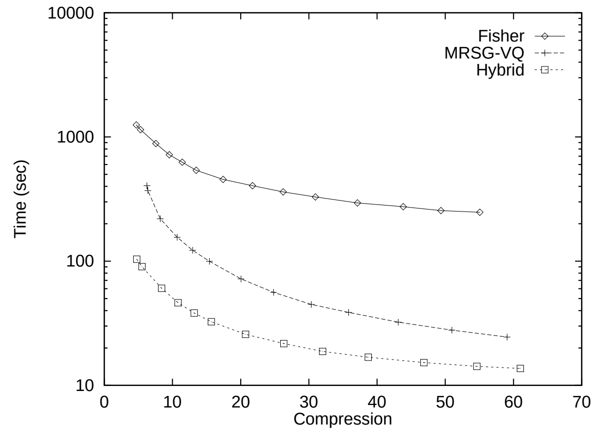


Figure 3: Time vs. compression ratio for the 512×512 Lenna image.

The compression ratio was varied by letting the tolerance level $\delta$ take the values 20, 18, 16, 14, 12, 10, 8, 6, 5, 4, 3, 2 and 1. For the hybrid scheme, the parameters $t$ and $\epsilon$ were set to $\delta$ and 0.15 respectively. For each quadtree level the domain pool consisted of blocks having twice the range size and centered on a grid with vertical and horizontal spacing of 8 pixels. Thus we had 3969 overlapping domains of size $16 \times 16$ and 3721 overlapping domains of size $32 \times 32$. Figure 3 shows the time as a function of compression ratio for the same series of tests. Note that due to the nature of our MRSG-VQ scheme its encoding times are not really representative. An optimized version would be evidently much faster. Clearly our hybrid scheme had a better performance than Fisher's scheme with a gain of more than 1 dB over the range of high compression ratios. The comparison to the MRSG-VQ scheme is more delicate. The hybrid scheme was superior only at low compression ratios. Moreover, it is possible to improve the rate-distortion results of the MRSG-VQ scheme by using codebooks of larger size. However, this comes at the expense of a severe increase in both time and space complexity. Even though the first difficulty can be handled by using, e.g., fast nearest neighbor search techniques, it is hard to imagine how the second issue could be solved.

## 5. CONCLUSION

We have proposed a new hybrid scheme which combines fractal image compression and product code vector quantization. First, a set of VQ codebook blocks is employed to classify the domain blocks. Then, for each range block, the algorithm decides automatically whether to use a domain block or a VQ codebook block for the encoding. Experimental results show that our hybrid scheme yields superior performance over conventional fractal coding.

## 6. REFERENCES

[1] Jacquin, A. E., *Image coding based on a fractal theory of iterated contractive image transformations,* IEEE Trans. Image Processing 1 (1992) 18–30.

[2] Jacquin, A. E., *Fractal image coding: A review,* Proceedings of the IEEE 81,10 (1993) 1451–1465.

[3] Gharavi-Alkhansari, M., Huang, T., *Fractal based techniques for a generalized image coding method,* in: *Proc. IEEE ICIP-94,* Austin, Texas, Nov. 1994.

[4] Lepsøy, S., *Attractor Image Compression: Fast Algorithms and Comparisons to Related Techniques,* PhD Thesis, The Norwegian Institute of Technology, Trondheim, Norway, June 1993.

[5] Ramstad, T. A., Lepsøy, S., *Block-based attractor coding: Potential and comparison to vector quantization,* in: *Conference Record of the 27th Asilomar Conference on Signals, Systems and Computers,* pp. 1504–1508, 1993.

[6] Murakami, T., Asai, K., Yamazaki, E., *Vector quantizer of video signals,* Electronics Letters 7 (1982) 1005–1006.

[7] Hamzaoui, R., *Codebook clustering by self-organizing maps for fractal image compression,* in: *NATO ASI Fractal Image Encoding and Analysis,* Trondheim, July 1995, to appear in a special issue of the journal Fractals. Also available by anonymous ftp in ftp.informatik.uni-freiburg.de in /documents/papers/fractal/Hamz95.ps.gz.

[8] Saupe, D., *From classification to multi-dimensional keys,* in: *Fractal Image Compression — Theory and Application,* Y. Fisher (ed.), Springer-Verlag, New York, 1994.

[9] Fisher, Y., *Fractal Image Compression — Theory and Application,* Springer-Verlag, New York, 1994.

[10] The Self-Organizing Map, Version 3.1, Helsinki University of Technology, Laboratory of Computer and Information Science, April 7, 1995.