Ukrainian Catholic University

Faculty of Applied Sciences

Data Science Master Programme

Discrete Wavelet Transform for Multiresolution Analysis

Linear Algebra final project report

Authors:
Denys Botuk
Yurii Voevidka
Markiian Mandzak

Monday 25th March, 2024

Abstract

This study explores and compares Discrete Wavelet Transform against various image compression techniques, including scalar quantization, vector quantization, 2D Fourier transform and 2D discrete cosine transform (DCT). These methods are implemented and evaluated using diverse datasets, assessing compression ratio, compression time, peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM). Scalar and vector quantization are lossy techniques that map input values to quantized levels or codewords. Fourier and DCT methods transform image data to the frequency domain for efficient encoding. The DWT approach decomposes images into wavelet coefficients across multiple resolutions for selective encoding. Through comprehensive evaluation, this study aims to provide insights into the trade-offs between compression efficiency and visual quality for each technique, guiding the selection of appropriate compression algorithms for various applications.

1 Introduction

Image compression is a critical technique in the field of digital image processing, serving as a fundamental tool for efficient storage and transmission of visual data. With the ever-increasing demand for handling and exchanging digital images across various domains, such as multimedia communication, image archiving, remote sensing, and medical imaging, the need for effective image compression techniques has become paramount. These techniques aim to reduce the amount of data required to represent an image while preserving its visual quality to the greatest extent possible, thereby enabling more efficient storage and faster transmission over networks.

The primary objective of image compression is to exploit the inherent redundancies and irrelevancies present in image data, allowing for a significant reduction in the overall data size. Redundancies in image data can arise from various sources, such as spatial redundancy (correlation between neighboring pixels), spectral redundancy (correlation between different color components), and psychovisual redundancy (information that is imperceptible to the human visual system). By identifying and removing these redundancies, image compression algorithms can substantially reduce the amount of data required to represent the original image.

Irrelevancies, on the other hand, refer to the information in the image data that may not be essential for human perception or the intended application. For instance, certain high-frequency components or fine details in an image may not be discernible to the human eye or may not significantly impact the overall visual quality. By selectively discarding or approximating these irrelevant components, image compression techniques can further reduce the data size while maintaining an acceptable level of visual quality.

Image compression techniques can be broadly classified into two categories: lossless and lossy compression. Lossless compression techniques aim to preserve the exact pixel values of the original image during the compression and decompression processes, ensuring that the reconstructed image is an exact replica of the original. These techniques are particularly useful in applications where any loss of information is unacceptable, such as medical imaging. However, lossless compression typically achieves lower compression ratios compared to lossy techniques.

Lossy compression techniques, on the other hand, allow for some controlled loss of information during the compression process, trading off a certain degree of visual quality for higher compression ratios. These techniques are suitable for applications where minor visual distortions are tolerable, such as multimedia communication, image archiving, and general-purpose imaging. By selectively discarding or approximating less perceptually relevant information, lossy compression techniques can achieve significantly higher compression ratios compared to lossless methods, enabling more efficient storage and transmission of image data. (see [1])

In this study, we explore Discrete wavelet transform and compare it with several other lossy image compression techniques, including scalar quantization, vector quantization and discrete cosine transform (DCT). Each of these techniques employs different strategies and linear algebra principles to achieve image compression, offering varying trade-offs between compression performance and visual quality.

By evaluating and comparing these techniques using various datasets and metrics, such as compression ratio, compression time, peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM), we aim to provide insights into their respective strengths and weaknesses. This analysis will aid in identifying the most suitable scenario of application of Discrete wavelet transform among other widely used methods, considering factors such as computational complexity, compression efficiency, and visual quality requirements.

Ultimately, the development and optimization of image compression techniques play a crucial role in enabling efficient storage, transmission, and processing of digital images across a wide range of domains, facilitating advancements in fields such as multimedia, computer vision, and digital imaging technologies.

2 Related Work

2.1 Scalar Quantization

Scalar Quantization is a lossy compression technique that reduces the number of bits required to represent each sample or coefficient in an image or signal. It involves mapping a range of input values to a single quantized value or output level. Scalar quantization quantizes each sample independently. (see [2])

2.2 Vector Quantization

Vector Quantization is a lossy compression technique that reduces the number of bits required to represent each sample in the signal. But contrary to scalar quantization, this technique takes into account the distribution of samples. In our implementation, we consider KMeans quantization by finding centroids of sample values and replacing each sample with the corresponding centroid. (see [2])

2.3 Discrete Cosine Transform (DCT)

DCT is a widely used method for image compression, particularly in standards like JPEG. It involves transforming image data from the spatial domain to the frequency domain using cosine functions. DCT concentrates most of the image energy into a small number of low-frequency coefficients, which can be quantized and encoded more efficiently. While DCT-based compression provides good compression ratios and is computationally efficient, it may not preserve fine image details as effectively as other methods, leading to loss of quality, especially at higher compression levels. (see [3])

3 Theory

3.1 Continuous Wavelet Transform (CWT)

The continuous wavelet transform of a signal $\mathbf{x}(t)$ with respect to a wavelet function $\psi(t) \in L^2(\mathbb{R})$ is defined as:

$$W_\psi[\mathbf{x}](a,b) = \langle \mathbf{x}, \psi_{a,b} \rangle = \int_{-\infty}^{\infty} \mathbf{x}(t)\overline{\psi}_{a,b}(t)dt \tag{1}$$

where $a \in \mathbb{R}/\{0\}$ scaling paramenter, $b \in \mathbb{R}$ translation parameter, and $\overline{\psi}_{a,b}(t)$ is the complex conjugate of a wavelet function derived from the mother wavelet $\psi(t)$ through scaling and translation:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{2}$$

The wavelet function $\psi(t)$ should satisfy the admissibility condition:

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \tag{3}$$

where $\hat{\psi}$ is the Fourier transform of ψ. (see [4])

This condition also implies that ψ is of zero mean:

$$\int \psi(t)dt = \hat{\psi}(0) = 0 \tag{4}$$

The wavelet function is usually normalized to have unit energy $\|\psi_{a,b}(t)\| = 1$

3.2 Inverse Continuous Wavelet Transform (iCWT)

When the admissibility condition (3) is met, it is possible to find inverse continuous transformation by Calderon's reproducing identity

$$\mathbf{x}(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{0}^{\infty} \frac{1}{a^2} W_\psi[\mathbf{x}](a,b)\psi_{a,b}(t)dadb \tag{5}$$

3.3 Discrete Wavelet Transform (DWT)

To discretize CWT, we use its redundancy and select minimal discrete sets of a and b such that transformation is still invertible. Such sets can be obtained by using *critical sampling*:

$$a = 2^{-j},\ b = k2^{-j},\ j, k \in \mathbb{Z} \tag{6}$$

This way we obtain a basis $\{\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k) \mid j, k \in \mathbb{Z}\}$ that is orthonormal in $L^2(\mathbb{R})$:

$$\langle \psi_{jk}, \psi_{lm} \rangle = \int_{-\infty}^{\infty} \psi_{jk}(t)\overline{\psi_{lm}}(t)dt = \delta_{jl}\delta_{km} \tag{7}$$

where δ_{ij} is Kronecker delta:

$$\delta_{ij} = \begin{cases} 1,\ i = j \\ 0,\ i \neq j \end{cases} \tag{8}$$

and we can expand any function $\mathbf{x}(t) \in L_2(\mathbb{R})$ as

$$\mathbf{x}(t) = \sum_{j,k=-\infty}^{\infty} c_{jk}\psi_{jk}(x) \tag{9}$$

where coefficients c_{jk} is given by equation (1),

$$c_{jk} = W_\psi[\mathbf{x}](a, b) \tag{10}$$

3.4 Multiresolution Analysis (MRA)

A multiresolution analysis is usualy defined as a sequence of closed subspaces $V_n,\ n \in \mathbb{Z}$ in $L_2(\mathbb{R})$ such that satisfy the following three properties

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \tag{11}$$

$$\bigcap_n V_j = \varnothing \tag{12}$$

$$\bigcup_n V_j = L_2(\mathbb{R}) \tag{13}$$

The sequence (11) is constructed in such a way that

$$f(2^j t) \in V_j \iff f(t) \in V_0 \tag{14}$$

and there exists *father wavelet* (also *scaling function*) $\phi \in V_0$ such that

$$V_0 = \{f \in L_2(\mathbb{R}) \mid f(t) = \sum_k c_k\phi(t - k)\} \tag{15}$$

and for which $\{\phi(t - k) \mid k \in \mathbb{Z}\}$ is an orthonormal basis.
The scaling functions should be of non-zero mean

$$\int \phi(t)dt \neq 0 \tag{16}$$

but usually it is equal to 1 (see [4, 5]).

3.5 Multiresolution Analysis Procedure

The scaling function ϕ is the solution of the following two-scale equation:

$$\phi(x) = \sum_{n=0}^{L} h_n \sqrt{2}\phi(2x - n), \qquad x \in R \tag{17}$$

where h_n - a finite sequence of real numbers.

The following two-scale expression gives the wavelet function ψ:

$$\psi(x) = \sum_{n=0}^{L} g_n \sqrt{2}\phi(2x - n), \qquad g_n = (-1)^n h_{L-n} \tag{18}$$

Given scaling function $\phi(x)$ and wavelet function $\psi(x)$ the following two-dimensional functions are constructed:

- $\phi(x)\phi(y)$ - two-dimensional scaling function (corresponds to low-pass filter in the x and y directions)
- $\psi(x)\phi(y)$ - vertical wavelet function (corresponds to high-pass filter in the x and low-pass filter in the y directions)
- $\phi(x)\psi(y)$ - horizontal wavelet function (corresponds to low-pass filter in the x and high-pass filter in the y directions)
- $\psi(x)\psi(y)$ - diagonal wavelet function (corresponds to high-pass filter in the x and y directions)

The two-dimensional scaling function is called an *approximation* because it is smooth and has large values. The other three parts are called *details* because they emphasize horizontal, vertical, and diagonal edges, respectively. These three parts have small absolute values except for the edges (see [3]).

The Multiresolution Analysis procedure involves a multi-level decomposition, which is applied to the successive approximation on each level. When we apply this decomposition to approximations and details, we have a series of multi-level decompositions called *wavelet packet decompositions*.

4 Experimental Setup

4.1 Datasets

4.1.1 Set5

Set5 is a widely used benchmark dataset for super-resolution tasks. It consists of five high-resolution images: Baby, Bird, Butterfly, Head, and Woman (see [6]).

4.1.2 Set14

Set14 is another popular dataset for super-resolution tasks, containing 14 high-resolution images with a variety of scenes, including animals, plants, and architectural structures (see [7]).

4.1.3 Urban100

Urban100 is a dataset specifically designed for evaluating the performance of super-resolution algorithms on urban scenes. It contains 100 high-resolution images captured in various urban environments, such as buildings, streets, and cityscapes (see [8]).

4.1.4 BSDS100

The BSDS100 dataset is derived from the Berkeley Segmentation Dataset and Benchmark (BSDS). It consists of 100 natural images with ground truth segmentation masks, making it suitable for evaluating both super-resolution and segmentation tasks (see [9]).

4.1.5 Manga109

Manga109 is a dataset specifically designed for super-resolution of manga (Japanese comic) images. It contains 109 high-resolution manga images with various styles and complexity levels, allowing for evaluating the performance of super-resolution algorithms on this specific type of content (see [10]).

4.2 Metrics

4.2.1 Compression Time (CT)

Compression time is the time taken for compression process. This metric varies from machine to machine and depends on the performance of the hardware components. The compression time is also affected by the complexity of the time of a compression algorithm. In the context of image processing, it is important to consider both compression and decompression times.

4.2.2 Compression Factor (CF)

The compression factor is the ratio between the size of the original image and the size of the compressed image.

$$\text{CF} = \frac{\text{original image size}}{\text{compressed image size}} \tag{19}$$

4.2.3 Peak Signal to Noise Ratio (PSNR)

Peak Signal to Noise Ratio measures the ratio between the maximum possible *power* of a *signal* and the power of distorting noise, which affects the quality of its representation as a logarithmic decibel scale. In Image Processing scenario, the signal is an image and its power is intensity values of a pixels. The higher the image quality, the higher the PSNR value, and the lower the image quality corresponds to significant differences between the original and reconstructed images.

The PSNR is computed as follows:

$$\text{PSNR} = 20 \cdot \log_{10}\left(\frac{\text{MAX}}{\sqrt{MSE}}\right) \tag{20}$$

where MAX corresponds to the maximum intensity pixel value of an image and MSE is the mean-squared error between the original and the reconstructed image.

4.2.4 Structural Similarity Index Measure (SSIM)

Originaly described in [11] The Structural Similarity Index Measure is widely used to assess structural similarity between reference and distorted image. It considers such factors as contrast, luminance, and structure. The range of values is between -1 suggesting that images are dissimilar and 1 suggesting compete match.

The SSIM is defined as

$$\text{SSMI}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{21}$$

where μ_x, μ_y, σ_x and σ_y are pixel sample means and variances of x and y correspondingly, σ_{xy} is covariance between x and y and constants c_1 and c_2 are defined as following:

$$c_1 = (k_1 L)^2 \qquad c_2 = (k_2 L)^2 \tag{22}$$

where L is dynamic range of pixel values defined as

$$L = 2^{\text{bits per pixel}} - 1 \tag{23}$$

and k_1 and k_2 are constants set to 0.01 and 0.03 by default.

5 Numerical Experiments

In this section, we present the experimental results obtained by evaluating Scalar Quantization, Vector Quantization, DCT and DWT methods on Set5, Set14, Urban100, BSDS100 and Manga109 datasets.

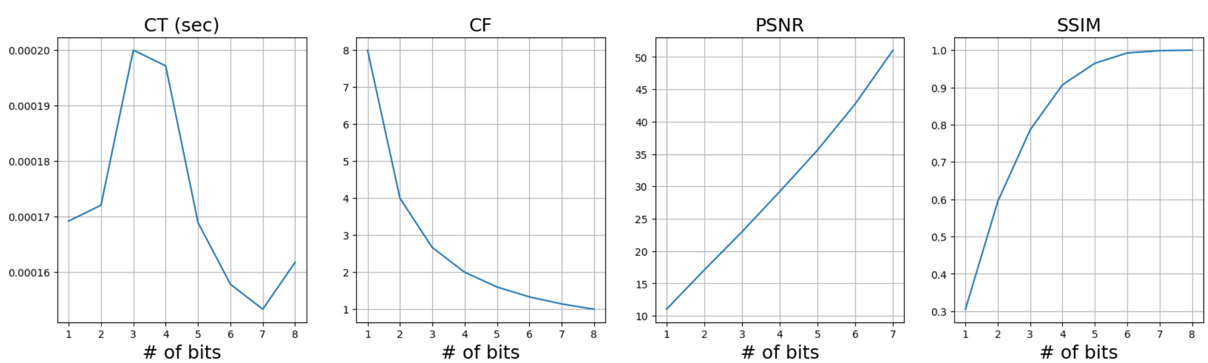
To mitigate bias toward a particular distribution, we undersampled each dataset, leaving 5 images per set (25 images in total).

For each method, we present Compression Time, Compression Factor, PSNR and SSIM averaged over all samples as well as examples of compression under different parameters of the corresponding method.

5.1 Scalar Quantization

Scalar quantization is one of the simplest lossy compression techniques. It quantizes each pixel by reducing the number of bits needed to represent each intensity value.

According to plots the quality degrades very rapidly without providing high compression. The eight-fold compression is the best possible on standard 8-bit images.



(a) int8 (CF: 1) (b) int4 (CF: 2) (c) int2 (CF: 4) (d) int1 (CF: 8)

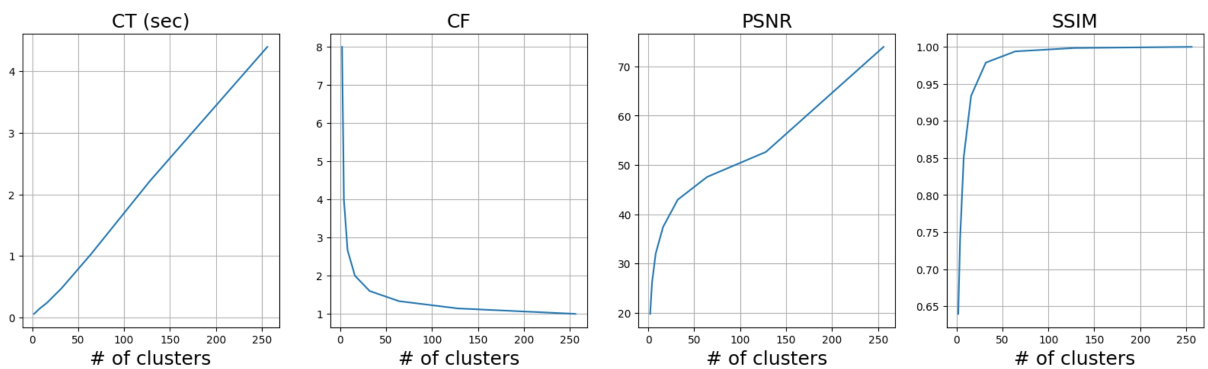
5.2 Vector Quantization

Vector quantization is similar to scalar quantization in that sense that it maps a range of pixel intensity values to its subset. But the main difference is that each value is mapped to the closest centroid determined by KMeans over the image.

This technique provides the same range of compression factors as scalar quantization, but the structure preserved much better according to PSNR and SSIM plots.

The main downside is that such compression takes much longer, and linearly depends on number of clusters.

This can be mitigated to some degree by considering smaller number of iterations, larger tolerance or different centroids' initialization in KMeans algorithm.



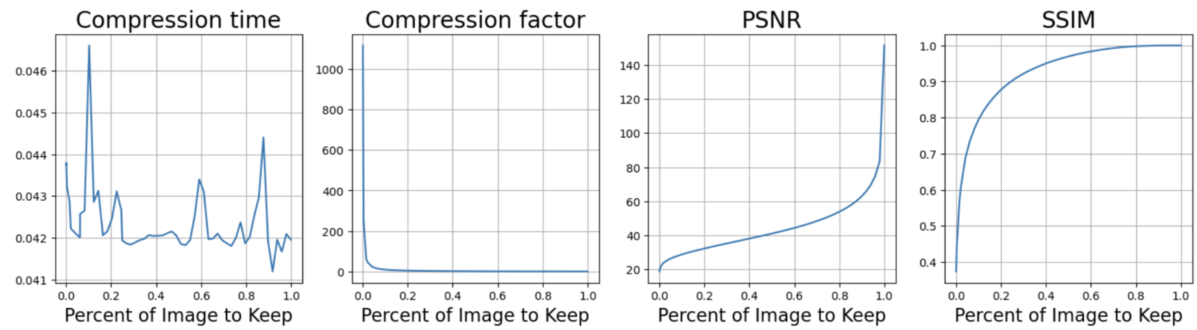
(a) K: 256 (CF: 1) (b) K: 16 (CF: 2) (c) K: 4 (CF: 4) (d) K: 2 (CF: 8)

5.3 Discrete Cosine Transform

The discrete cosine transform (DCT) expresses an image's pixel values in terms of cosine basis functions of varying frequencies.

In our implementation we use absolute value coefficients filtering which is part of JPEG standard, though other approaches such as frequency filtering or coefficients quantization are widely used.

As we can see, discrete cosine transform allows much higher compression rates than simple quantization techniques, while running faster than vector quantization. Also, it is much more flexible in terms of choosing compression factor as it operates on continuous threshold.



(a) 100% (CF: 1) (b) 25% (CF: 4) (c) 15.62% (CF: 64) (d) 0.39% (CF: 256)

5.4 Discrete Wavelet Transform

The discrete wavelet transform (DWT) represents images as a sum of localized wavelet functions at multiple scales. Wavelet coefficients across different subbands can be selectively encoded to achieve compression.

In our implementation the compression is achieved by discarding coefficients that account for horizontal, vertical and diagonal details of an image.

The proposed approach is evaluated with different wavelet functions and examples are provided for best and least performing ones according to SSIM score.

The PSNR and SSIM scores for the proposed approach degrade quicker than for DCT, but it is compensated by a steeper increase in compression factor.

The advantage of DWT over DCT is that it provides ability to even higher compression rates than DCT if one consider level 7 and above, while running faster than DCT and having logarithmic time complexity wrt. level.

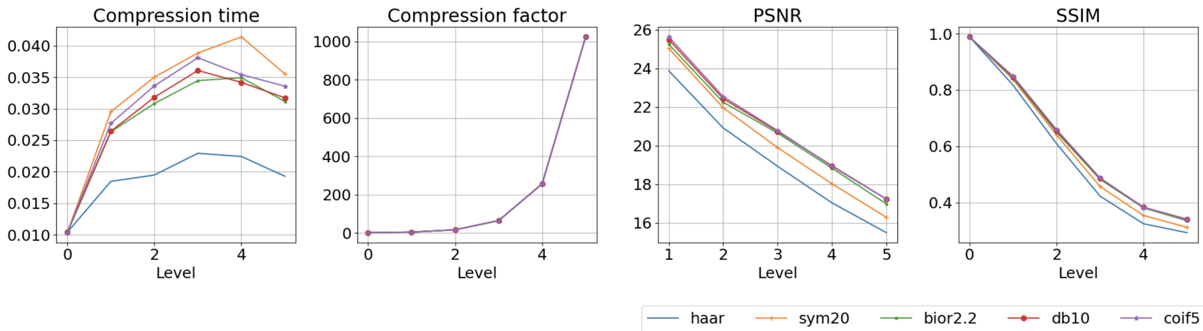


Figure 4: Experimental Results of DWT Image Compression


(a) Level 0 (CF: 1)


(b) Level 1 (CF: 4)


(c) Level 3 (CF: 64)

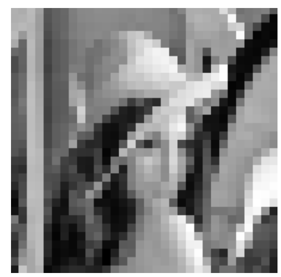

(d) Level 4 (CF: 256)

Figure 5: Examples of DWT image compression with Haar wavelet


(a) Level 0 (CF: 1)


(b) Level 1 (CF: 4)


(c) Level 3 (CF: 64)


(d) Level 4 (CF: 256)

Figure 6: Examples of DWT image compression with Symlet20

6 Conclusions

In this study, we explored and compared various image compression techniques, including scalar quantization, vector quantization, discrete cosine transform (DCT) against discrete wavelet transform (DWT). Through comprehensive experiments across multiple datasets, we evaluated the trade-offs between compression performance and visual quality for each method.

Scalar quantization and vector quantization, being relatively simple techniques, achieved lower compression ratios but provided fastest compression times and decreased SSIM reduction respectively. While scalar quantization suffered from rapid quality degradation, vector quantization preserved image structure more effectively, as evidenced by higher PSNR and SSIM scores.

The DCT method, widely used in standards like JPEG, demonstrated superior compression capabilities compared to quantization techniques, enabling higher compression factors while maintaining reasonable visual quality. Its performance was notable in terms of both compression efficiency and computational complexity.

The DWT approach, based on wavelet decomposition and selective encoding of coefficients across multiple resolutions, exhibited comparable or even better compression ratios than DCT, particularly at higher compression levels. DWT's advantage lies in its ability to achieve very high compression factors while maintaining acceptable visual quality, albeit its discretized compression steps.

It is worth noting that the choice of the optimal compression technique depends on the specific application requirements and constraints. For applications with stringent quality demands and low compression needs, scalar or vector quantization may suffice. However, when higher compression ratios are desired, DCT and DWT emerge as more suitable choices, with DWT providing an edge for extreme compression scenarios where visual quality can be traded for significant data reduction while DCT providing more granular control over compression.

In summary, this study contributes to a better understanding of the strengths and limitations of various image compression techniques, guiding the selection process based on factors such as compression efficiency, visual quality, and computational complexity. The results highlight the potential of DWT as a powerful tool for image compression, particularly in scenarios where high compression ratios are prioritized over acceptable quality losses.

References

[1] W. K. Ng, S. Choi, and C. V. Ravishankar, "Lossless and Lossy Data Compression," June 1998, doi: 10.1109/32.663423.

[2] M. A. T. Figueiredo, "Scalar and Vector Quantization," Departamento de Engenharia Electrotécnica e de Computadores, Instituto Superior Técnico, Lisboa, Portugal, Nov. 2008.

[3] T. S. Hasan, "Image Compression Using Discrete Wavelet Transform And Discrete Cosine Transform," Journal of Applied Sciences Research, vol. 13, no. 3, pp. 1-8, Mar. 2017.

[4] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image Coding Using Wavelet Transform," IEEE Transactions on Image Processing, vol. 1, no. 2, pp. 205, Apr. 1992.

[5] P. Cifuentes, K. S. Kazarian, and A. San Antolín, "Characterization of Scaling Functions in a Multiresolution Analysis," Proceedings of the American Mathematical Society, vol. 133, no. 4, pp. 1013-1023, Nov. 2004, doi: 10.1090/S0002-9939(04)07786-X.

[6] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," BMVA Press, 2012.

[7] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in International Conference on Curves and Surfaces, Springer, 2010, pp. 711-730.

[8] J.-B. Huang, A. Singh, and N. Ahuja, "Single Image Super-Resolution From Transformed Self-Exemplars," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.

[9] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, vol. 2, pp. 416-423, 2001.

[10] K. Aizawa, A. Fujimoto, A. Otsubo, T. Ogawa, Y. Matsui, K. Tsubota, and H. Ikuta, "Building a Manga Dataset 'Manga109' With Annotations for Multimedia Applications," IEEE MultiMedia, vol. 27, no. 2, pp. 8–18, Apr. 2020. doi: 10.1109/MMUL.2020.2987895.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, pp. 600-612, 2004.