SATFD

Lab. 6

May 24, 2024

Exercise 6: Image Processing. JPEG and JPEG 2000 Graphic Format

Introduction

The Fourier transform and the wavelet transform can be successfully used in algorithms for compressing images, sounds, and videos. We will examine this with the example of the JPEG and JPEG 2000 image compression standards. Both standards are based on the transformation from the original domain to the frequency domain. They begin by dividing the image into $N \times M$ blocks. This is not strictly required by the standards, as the algorithms can also handle images without block division (in practice: up to a size of $2^{32} \times 2^{32}$). The standard block size is 8×8. If the image size is not a multiple of $N \times M$, the image is padded to the nearest multiple in both the vertical and horizontal directions. The next step is to perform the appropriate transformation (cosine for the JPEG standard or wavelet for the JPEG 2000 standard), followed by quantization in the frequency domain, which is responsible for the actual image compression. Various quantization algorithms can be used - compressing the image more or less aggressively. If the image consists of multiple components, such as R, G, B channels, or Y, U, V, or in the simplest case of black and white images, grayscale levels. The quality of compression is measured by the PSNR (peak signal to noise ratio), which may be understood as the total error and is expressed in dB. PSNR is defined for the original image f and the compressed image f_0 as:

$$PSNR = 10 \cdot \log_{10}\left(\frac{k^2}{\text{MSE}}\right) \tag{1}$$

where $k =$ (number of colors - 1), and MSE is:

$$MSE = \frac{1}{N \cdot M}\sum_{i=1}^{N}\sum_{j=1}^{M}\left([f(i,j) - f_0(i,j)]^2\right) \tag{2}$$

As can be seen, MSE appears in the denominator, so the higher the PSNR value, the better. PSNR can be measured for different compression levels expressed in bits per pixel (bpp).

Standard JPEG

The transformation used in the JPEG standard is a two-dimensional cosine transform. It is composed of two successive one-dimensional cosine transforms (DCT - discrete cosine transform). First, each row of the image is replaced with coefficients of expansion in the cosine basis. Then, each column of the thus prepared image is replaced with coefficients of expansion in the cosine basis. Numerically, DCT can be implemented using FFT. Formally, the transformation is given by the formula (3) (according to [1]):

$$X_{\text{DCT}}(k,l) = \sum_{m=0}^{M-1}\left[\sum_{n=0}^{N-1} x(m,n)\cdot\alpha(l)\cdot\cos\left(\frac{\pi l}{N}\left(n+\frac{1}{2}\right)\right)\right]\cdot\beta(k)\cdot\cos\left(\frac{\pi k}{M}\left(m+\frac{1}{2}\right)\right) \tag{3}$$

where

$$\beta(k) = \begin{cases}\sqrt{\frac{1}{M}} & \text{if } k = 0\\ \sqrt{\frac{2}{M}} & \text{if } k = 1,\ldots,M-1\end{cases}, \quad \alpha(l) = \begin{cases}\sqrt{\frac{1}{N}} & \text{if } l = 0\\ \sqrt{\frac{2}{N}} & \text{if } l = 1,\ldots,N-1\end{cases} \tag{4}$$

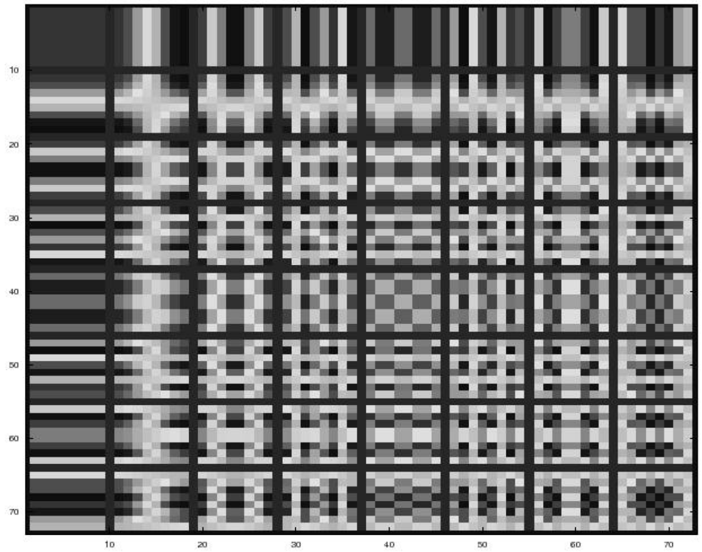

Figure 1: Spatial Frequencies

These transformations can also be written in matrix form (5), which is very convenient for numerical applications:

$$Y_{\mathrm{DCT}}(k,l) = \sum_{m=0}^{M-1} \left[\sum_{n=0}^{N-1} x(m,n) \cdot b_N(l,n) \right] \cdot b_M(k,m) \tag{5}$$

$$Y_{M\times N} = B_M \cdot X_{M\times N} \cdot B_N^T \tag{6}$$

where $b_P(k,m),\ k,m \in [0, P-1]$ - the basis functions - are written in successive rows of the matrix B. For example:

$$b_P(k,m) = \sqrt{\frac{1}{P}} \exp\left(j\frac{2\pi k}{P} m \right) \quad \text{for FFT} \tag{7}$$

$$b_P(k,m) = \beta(k) \cdot \cos\left(\frac{\pi k}{P}(m + \frac{1}{2}) \right) \quad \text{for DCT} \tag{8}$$

Figure 1 presents increasingly higher spatial frequencies shown as matrices. The analyzed matrix can be represented as the sum of the matrices from Figure 1 multiplied by the expansion coefficients (see [1], formulas 22.24-22.26).

Note that the lowest frequencies (lowest values of coefficients k, l in formula (3)) are located in the upper left corner. Frequency increases to the right and downward. Transformation (5) allows for spatial filtering, which involves multiplying the matrix of spectral expansion coefficients by the filter transmittance matrix. Extensive material on this topic is provided by Zielinski [1].

Due to the location of spatial frequencies, converting the matrix into a vector according to the formula shown in Figure 2 arranges the spectral coefficients in increasing order of frequency (or rather the sum of frequencies).

The next stage of compression is quantization, which involves dividing the matrix Y by a matrix of the same dimension, containing numbers that increase according to a zigzag pattern, asymmetric (the so-called standard luminance matrix - obtained experimentally) or symmetric (the so-called Hilbert matrix) with respect to the main diagonal. The elements of the matrix obtained after division are rounded down to the nearest integer. The measure of the degree of compression is the number of zeros that appear in the resulting matrix, and its coefficients, after conversion to a data vector (according to the zigzag from Figure 2), are subjected to Huffman-type compression, ensuring minimal entropy of the transmitted sequence of symbols (another form of compression) aimed at assigning labels from a dictionary to sequences of numbers occurring in the examined data sequence, with the label being shorter the more frequently a given symbol or word appears. Details are provided by Zielinski [1] in chapter 11, they are also described in algorithmic literature.

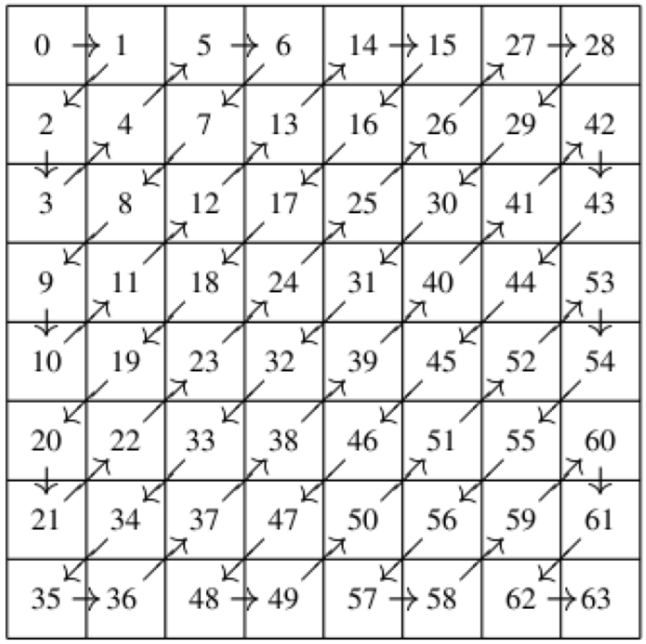

Figure 2: Zigzag Pattern for Spectral Coefficients

Standard JPEG 2000

A drawback of JPEG is the pixelation that occurs due to aggressive quantization, revealing the block size (e.g., 8×8). This flaw is absent in the JPEG 2000 method, which provides very good compression ratios without visible degradation in image quality. The downside of this method is the significant computational power requirement and unresolved issues related to existing patents. For the JPEG 2000 standard (unlike JPEG), the larger the block size, the better the image quality. The JPEG 2000 standard differs from its predecessor both in structure and in the achieved compression ratio, expressed by high PSNR values.

JPEG 2000 compression also begins by dividing the image into blocks. Then, for each block, a standard one-dimensional filtering is performed using a cascade of quadrature mirror filters (QMF), similar to a standard wavelet transform. Filtering is performed first for each row of the original image and then for the columns. The result of each wavelet transform is two images: one is the result of high-pass filtering (HP) – this signal can be described as details (D) (see [1], Fig. 22.32), and the other – the so-called approximation (A) – the result of low-pass filtering (LP). Since filtering is performed in two perpendicular directions, from one image we get after the first direction LP and HP images (A and D), and after the second direction up to four, as each of the images is again processed by a pair of QMF filters. We thus have four images. From LP we get LP-LP and LP-HP, and from HP: HP-HP and HP-LP. Decimation is an element of the transform. After row filtering, decimation leads to selecting every other coefficient, i.e., every other column, making the image half as wide. Then, after column filtering, decimation causes the four images to have half the height and width. The D-D (HP-HP), D-A (HP-LP), and A-D (LP-HP) images are stored "on the side," while the A-A (LP-LP) image is processed again by the next stage QMF filters. Again, due to decimation, the same filters can be used.

The wavelet expansion coefficients of subsequent stages can be recorded in specific regions instead of the original image. The scheme is shown in Figure 4, and the result of such decomposition (analysis) in Figure 5. The detail images do not show the blurred valley due to their low-frequency nature. The number of coefficients describing the image does not change – compression is achieved through quantization, similar to the JPEG format. The exact compression algorithm is described by Jin Li [8].

The choice of wavelet determines the filter coefficients G and H. The most commonly used is the CDF 5/7 (Cohen Daubechies Feauveau) wavelet for lossless compression or CDF 7/9 for lossy compression. The use of biorthogonal wavelets has the advantage that compression and decompression can be performed using the same filters.

Regarding compression quality, let's give the floor to Jin Li [8]. The graph discussed by him, Figure 22, is Figure 7 (14).

We compare the JPEG 2000 image compression standard with the traditional JPEG standard. The test image is the Bike standard image (gray, 2048x2560). Three modes of JPEG 2000 are tested, and compared against two modes of the JPEG standard. The JPEG modes are progressive (P-DCT) and sequential (S-DCT), both with optimized Huffman tables. The JPEG-2000 modes are single layer with the (9,7) wavelet (S-9,7), six layer progressive with the (9,7) wavelet (P6-9,7), and seven layer progressive with the (3,5) wavelet (P7-3,5). The JPEG-2000 progressive modes have been optimized for 0.0625, 0.125, 0.25, 0.5, 1.0, 2.0 bpp and lossless for the 5x3 wavelet. The JPEG progressive mode uses a combination of spectral refinement and successive approximation. We show the performance comparison in Figure 22. The JPEG-2000 results are significantly better than the JPEG results for all modes and all bitrates on this image. Typically, JPEG-2000 provides only a few dB improvement from 0.5 to 1.0 bpp but substantial improvement below 0.25 bpp and above 1.5 bpp. Also, JPEG-2000 achieves scalability at almost no additional cost. The progressive performance is almost as good as the single layer JPEG-2000 without the progressive capability. The slight difference is due solely to the increased signaling cost for the additional layers (which changes the packet headers). It is possible to provide "generic rate scalability" by using upwards of fifty layers. In this case, the "scallops" in the progressive curve disappear, but the overhead may be slightly increased.

The progressive algorithm described by Jin Li is also discussed by Zielinski [1] (Chapter 22.4.2).

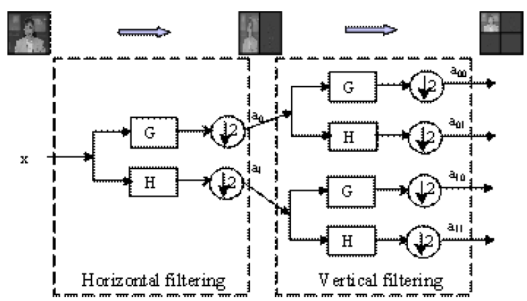

Figure 3: Two-dimensional wavelet transform [8].

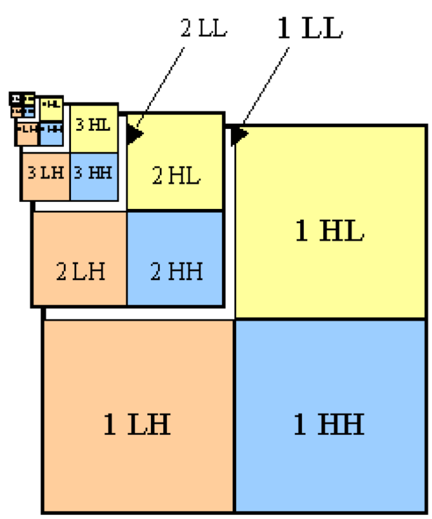

Figure 4: Wavelet decomposition coefficients at successive levels. As can be seen, they do not provide compression and in one version of the JPEG algorithm (integer), they are reversible [5].

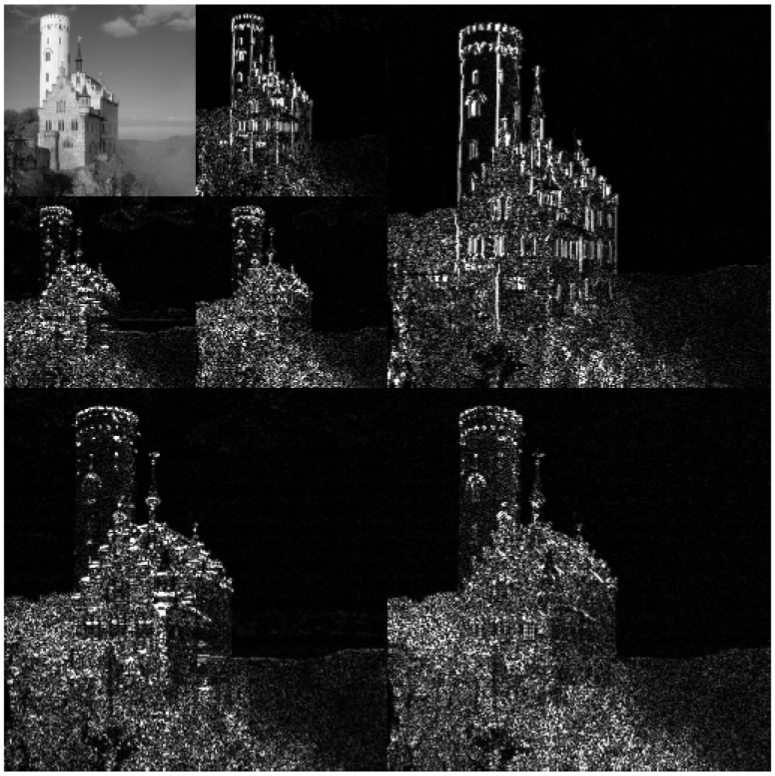

Figure 5: Wavelet decomposition coefficients after the first stage of two-dimensional wavelet transform. [6].

Realization of the task

For the realization of the exercise, we will use the JPEGtool package, authored by Darrell Hankerson and Greg A. Harris.

For the chosen image in pgm format (portable greymap), display the following images sequentially:

1. Original image.
2. Result of applying DCT to the image.
3. Result of standard quantization.
4. Result of dequantization.
5. Result of inverse DCT.

Then repeat the above set for:

1. Another block size (standard is $N \times N$, where $N = 8$).
2. Another quantizer.

Compare and comment on the obtained results.

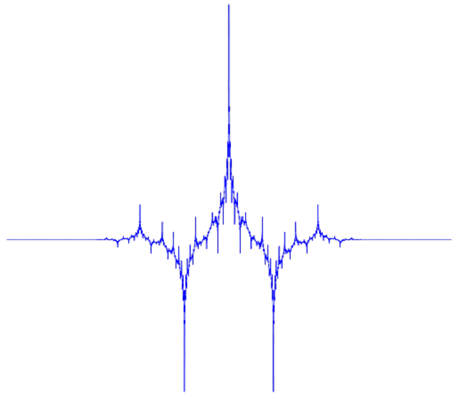

Figure 6: CDF 5/7 wavelet [4].

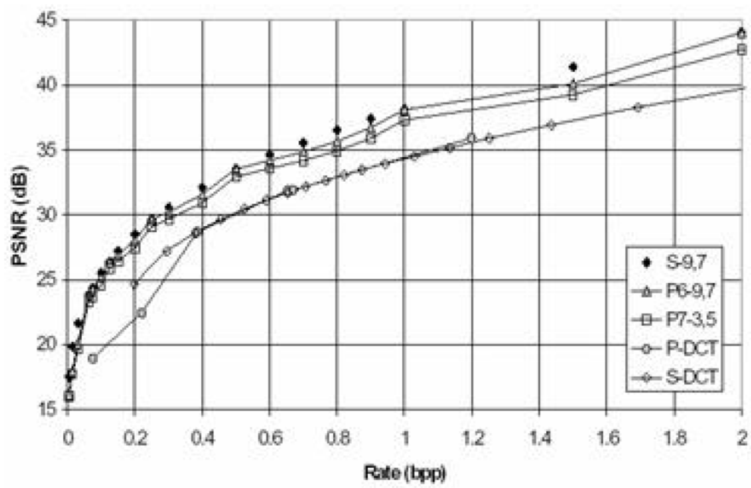

Figure 7: Performance comparison: JPEG 2000 versus JPEG (courtesy of Prof. Marcellin, et. al, [9]). [7].

Bibliography

1. T. Zielinski, Cyfrowe przetwarzanie sygnałów, od teorii do zastosowań, WKŁ Warszawa, 2009.
2. PSNR (Peak Signal-to-Noise Ratio)
3. MATLAB Central File Exchange - PSNR
4. JPEG 2000 - Answers.com
5. IP Cores for Accelerating JPEG2000
6. Jpeg2000 2-level wavelet transform - Lichtenstein
7. Jin Li, Image Compression - the Mechanics of the JPEG 2000
8. Jin Li, Image Compression - the Mechanics of the JPEG 2000
9. M. W. Marcellin, M. Gormish, A. Bilgin, M. P. Boliek, An Overview of JPEG2000, Proc. of the Data Compression Conference, Snowbird, Utah, March 2000, pp. 523-544.