IW44 Wavelet Encoding

Abstract

The IW44 (Incremental Wavelet 44) encoding scheme is a wavelet-based compression method used in DjVu to encode image data. It employs a hierarchical set difference coding technique combined with arithmetic coding to achieve high compression ratios while preserving image quality.

1 Wavelet Transform

The forward wavelet transform decomposes the image into multiple resolution levels using a lifting scheme. For a given input image $I(x, y)$ with dimensions $W \times H$, the transform is applied iteratively across scales. The lifting steps for the vertical filter (`filter_fv`) are defined as follows:

1.1 Predict Step (1–Δ)

For each sample at position (x, y) with scale s, the prediction is computed as:

$$\Delta y \;=\; I(x,y) \;-\; \left\lfloor \frac{I(x,y-s) + I(x,y+s) + 1}{2} \right\rfloor,$$

where s is the current scale, and boundary conditions use mirroring.

1.2 Update Step (2–Σ)

The update for the coarse scale is:

$$I(x,y{-}3s) \;=\; I(x,y{-}3s) + \left\lfloor \frac{(I(x,y-s) + I(x,y+s)) \cdot 9 \;-\; (I(x,y-3s) + I(x,y+3s)) + 16}{32} \right\rfloor.$$

This process is repeated for horizontal filtering (`filter_fh`) with identical lifting steps, adjusting indices to x instead of y. The transform iterates over scales $s = 1, 2, 4, 8, 16, 32$, halving the resolution at each step.

2 Quantization

After wavelet decomposition, coefficients are quantized using a set of thresholds. The quantization threshold for band b at bit-plane k is derived from the initial threshold Q_b (from the `IW_QUANT` table) as:

$$T_b(k) \;=\; \frac{Q_b}{2^k},$$

where k ranges from 0 (most significant bit) to 15 (least significant bit) for 16-bit coefficients. The low-frequency band (band 0) uses a separate threshold array $Q_{\text{lo}}[i]$, while higher bands use $Q_{\text{hi}}[b]$.

3 Hierarchical Set Difference Coding

The encoded coefficients are represented using a hierarchical set difference coding scheme. For each bit-plane k, the encoder processes all bands sequentially. The significance of a coefficient c at position (x, y) in band b is determined by

$$\text{Significant}(c) \;=\; |c| \;\geq\; T_b(k).$$

3.1 State Transitions

- Initially, all coefficients are in state `UNK` (unknown).
- If $|c| \geq T_b(k)$ and the coefficient was not previously significant, it transitions to `NEW`.
- Once encoded as significant, it becomes `ACTIVE`.
- If no bits remain significant, it is `ZERO`.

The encoding process uses an arithmetic coder (Z-Encoder) with context models. The probability of a coefficient being significant is modeled based on neighboring coefficients and previous states.

4 Chunk Structure

The encoded data is organized into IW44 chunks (e.g., `BG44` or `PM44`). Each chunk includes:

- A header with version (1.2), dimensions $W \times H$, and delay bits.
- A series of slices, where each slice encodes one bit-plane across all bands.
- The payload: the arithmetic-coded bitstream.

The total number of slices is determined by a stopping condition, such as a target signal-to-noise ratio (SNR) in decibels:

$$\text{SNR} \;=\; 10 \log_{10}\left(\frac{\sum_{x,y}|I(x,y)|^2}{\sum_{x,y}|I(x,y) - \hat{I}(x,y)|^2}\right),$$

where $\hat{I}$ is the reconstructed image.

Notes:

- The equations follow the lifting scheme and quantization logic from the IW44 specification (DjVu 3.0, Sec. 5.2).
- Context-specific details (e.g., exact `IW_QUANT` values or Z-Encoder parameters) are omitted but may be added as needed.
- If other entropy-coding or mask-interpolation equations are required, supply them and they will be integrated.