IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, July 2026
Impact Factor: 8.2

Multithresholding-Guided Region-of-Interest SPIHT for Embedded Wavelet Image Compression

Prabhjot Kaur
Dept. of Electronics & Communication Engineering,
Sardar Beant Singh State University, Gurdaspur, India
jyotidhaliwal12@gmail.com

Abstract: *Embedded wavelet coders such as SPIHT transmit image content in decreasing order of energy, but they are agnostic to which spatial regions carry perceptual importance. This paper proposes a region-of-interest (ROI) extension of SPIHT in which the ROI is identified automatically by multilevel Otsu thresholding of a local-activity map, and the wavelet coefficients supporting the ROI are scaled by a power of two before encoding, granting them a fixed bit-plane head start in the embedded stream. The binary ROI mask is transmitted at reduced resolution, adding only 1024 bits of side information. The method preserves the fully embedded, progressive nature of SPIHT: a single encoding serves every bitrate, and any truncation now favours the perceptually important content. Experiments on 512 × 512 grayscale test images at 0.125–1.0 bpp show that the proposed coder improves ROI fidelity over standard SPIHT in all tested conditions, by up to 7.14 dB (moon, 0.5 bpp) and 6.02 dB (coffee, 1.0 bpp), and exceeds JPEG's ROI fidelity by 3.5–12.7 dB at identical total bitrate, while the whole-image cost remains a controlled 0.8–2.7 dB borne by the background. An ablation of the scaling exponent identifies s = 2 as a robust operating point balancing ROI gain against whole-image cost, and the study characterizes an inverse relationship between ROI size and achievable gain. The approach is training-free, adds negligible computational overhead, and is attractive for progressive transmission in bandwidth-constrained applications such as telemedicine and remote sensing.*

Keywords: image compression; wavelet transform; SPIHT; region of interest; multilevel thresholding; embedded coding; progressive transmission.

I. INTRODUCTION

Digital images dominate the traffic of modern networks, and their efficient representation remains a foundational problem in signal processing. Transform coding has been the standard answer for over three decades: the JPEG standard [1] combines the block discrete cosine transform with scalar quantization and Huffman entropy coding [2], and remains ubiquitous despite well-known blocking artefacts at low bitrates. The wavelet transform removed the block structure and concentrated image energy into a sparse hierarchy of coefficients [5], [6], enabling a family of embedded coders — Shapiro's embedded zerotree wavelet algorithm (EZW) [3], Said and Pearlman's Set Partitioning in Hierarchical Trees (SPIHT) [4], and the EBCOT engine [8] adopted by the JPEG2000 standard [7] — whose defining property is that the bitstream can be truncated at any point to yield the best reconstruction available at that rate.

Embedded coding is naturally suited to progressive transmission: a receiver on a constrained link obtains a coarse image almost immediately and watches it refine as bits arrive. However, classical embedded coders order information purely by coefficient magnitude. Regions that matter most to a human observer or to a downstream task — a face, a lesion in a radiograph, a vehicle in a surveillance frame — receive no preferential treatment; they refine at the same pace as sky, walls, and other smooth background. Region-of-interest (ROI) coding addresses this by reallocating the bit

DOI: 10.48175/IJARSCT-37629

budget spatially. JPEG2000 standardized two mechanisms, general coefficient scaling and the maxshift method [7], [9], but ROI support for SPIHT-family coders is not standardized, and — more fundamentally — all these mechanisms assume the ROI is supplied externally, by a human operator or a separate detection module.

This paper closes that gap with an automatic, training-free ROI-SPIHT codec. The key idea is to derive the ROI from the image itself using multilevel Otsu thresholding [10], [13] — a classical, parameter-free histogram partitioning technique — applied not to raw intensity but to a local-activity map, so that the top threshold class isolates edges, texture, and object detail: precisely the structures on which subjective quality judgments concentrate. The resulting binary mask then drives a power-of-two scaling of the wavelet coefficients supporting the ROI, which advances their bit-planes in the SPIHT significance ordering by exactly s planes. The decoder inverts the scaling using a copy of the mask transmitted at reduced resolution for a fixed cost of 1024 bits.

The contributions are threefold. (1) A multithresholding-based ROI detector for compression that requires no training data, no manual input, and negligible computation, and that produces compact masks (3–19% of image area) aligned with perceptual importance. (2) A scaling-based integration of this ROI into SPIHT that provably preserves the embedded property, so one encoded file continues to serve every bandwidth while every truncation favours the ROI. (3) A reproducible experimental study — all numbers in this paper are generated by the released implementation — showing consistent ROI gains over standard SPIHT (up to +7.14 dB) and JPEG (up to +12.68 dB) at identical total bitrate, together with an ablation of the emphasis parameter and an analysis relating achievable gain to ROI size.

II. LITERATURE REVIEW

Huffman coding [2] assigns integer-length prefix-free codes to symbols and underpins the JPEG standard [1], where it entropy-codes quantized DCT coefficients. Its limitation to whole-bit code lengths keeps it away from the Shannon entropy limit [17], which motivated superior entropy coders — arithmetic coding [18] and, more recently, asymmetric numeral systems [19] — and, more consequentially for this work, the move from block transforms to the multiresolution wavelet representations formalized by Mallat [20] and applied to image coding by Antonini et al. [5]. The biorthogonal 9/7 filter bank of Cohen, Daubechies and Feauveau [6] became the de facto choice for lossy wavelet coding owing to its near-linear phase and energy compaction, and is retained here. Shapiro's EZW [3] introduced the zerotree hypothesis — insignificance of a coefficient predicts insignificance of its spatial descendants — and produced the first fully embedded wavelet bitstream. SPIHT [4] refined the idea with set partitioning across three lists, improving both compression and speed, and remains the reference point for tree-based embedded coders; the related SPECK algorithm [21] applies set partitioning to blocks rather than trees at even lower complexity. JPEG2000 [7], [22] adopted the block-based EBCOT engine [8] and standardized ROI coding through coefficient scaling and the maxshift method [9], in which ROI coefficients are shifted above the largest background coefficient so that no mask transmission is needed, at the price of an all-or-nothing emphasis. Region-based extensions of SPIHT itself have also been explored, chiefly in medical imaging: coefficient bit-shifting for ROI emphasis within SPIHT streams, object-based SPIHT operating on region-based wavelet transforms of segmented mammograms [15], and unbalanced spatial orientation trees that confine each tree to a single region [16]. A common trait of these methods is that the ROI is supplied externally — drawn by a radiologist or produced by a separate segmentation system tuned to the modality. The contribution of the present work is therefore not the scaling mechanism itself, which builds directly on this line of prior art [9], [15], [16], but its pairing with an automatic, training-free, general-purpose mask generator based on multilevel thresholding of a local-activity map, together with a tunable emphasis exponent, a quantified side-information budget, and an analysis relating achievable gain to ROI size. Otsu's method [10] selects the threshold maximizing between-class variance of the gray-level histogram and generalizes naturally to multiple classes; efficient multilevel algorithms are available [13], and thresholding techniques are comprehensively surveyed in [23]. Entropy-based criteria such as Kapur's [11] offer an alternative partitioning principle. Because exhaustive search becomes expensive as the number of thresholds grows, a large recent literature optimizes Otsu/Kapur objectives with metaheuristics — artificial bee colony variants for satellite imagery [24], comparative studies across swarm optimizers for plant-disease images [25], and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, July 2026 Impact Factor: 8.2

enhanced manta-ray foraging optimization [26], among many others — which also suggests a natural upgrade path for the thresholding stage used here. Multilevel thresholding is a staple of low-cost segmentation and has been used in compression pipelines chiefly to quantize intensities ahead of entropy coding. The present work repurposes it differently: thresholding is applied to a local-activity (windowed standard deviation) map rather than to intensity, converting a brightness segmenter into a perceptual-importance detector at essentially no computational cost. End-to-end learned codecs based on variational autoencoders [14] now define the rate–distortion state of the art, with generative variants achieving high perceptual fidelity at extreme rates [27] and transformer–CNN hybrids advancing the objective-quality frontier [28]; the broader landscape of generative visual coding, including standardization efforts, is surveyed in [29]. Learned saliency models can likewise supply semantic ROI masks. The costs of this family — training data, GPU inference, and non-embedded bitstreams in most designs — keep classical embedded coders relevant for low-complexity, progressive-transmission settings, which is the regime targeted here. [Note to author: consider adding 2–3 domain-specific ROI-compression references (e.g., telemedicine or remote sensing) from your own literature survey.]

III. METHODOLOGY

The proposed codec combines a multilevel-thresholding-based region analysis stage with an embedded wavelet coder. The central idea is that not all regions of an image contribute equally to perceived quality: regions of high local activity (edges, texture, object detail) dominate subjective assessment, while smooth background regions tolerate coarser reconstruction. The proposed method uses multilevel Otsu thresholding [10] to identify these perceptually important regions automatically and then re-orders the embedded SPIHT bitstream so that the bit-planes describing those regions are transmitted first. The overall pipeline comprises four stages: (i) region-of-interest extraction by multithresholding, (ii) discrete wavelet transform, (iii) ROI-selective coefficient scaling, and (iv) SPIHT encoding.

A local-activity map is first computed from the input image X of size $N \times N$. For every pixel, the standard deviation of intensities inside a 9×9 sliding window is evaluated, producing an activity map A in which large values correspond to edges and textured detail. Multilevel Otsu thresholding with three classes [10], [13] is then applied to A, yielding two thresholds $t1 < t2$ that partition the activity histogram into low-, medium-, and high-activity classes by maximizing the between-class variance. The binary ROI mask is defined as $M(i, j) = 1$ if $A(i, j) \geq t2$ and 0 otherwise, i.e., the highest-activity class. The mask is dilated by two iterations of a 3×3 structuring element to provide a safety margin around detected detail. On the test images used in this study the resulting ROI covers between 3% and 19% of the image area, so the bit-reallocation acts on a compact region.

The mask is transmitted to the decoder as side information at a reduced resolution of 32×32 (block-wise logical OR downsampling), costing exactly 1024 bits — less than 0.004 bpp for a 512×512 image. This overhead is included in all reported bitrates.

The DC-level-shifted image $(X - 128)$ is decomposed with a five-level two-dimensional discrete wavelet transform using the Cohen–Daubechies–Feauveau 9/7 biorthogonal filter bank [6], the filter adopted by the JPEG2000 standard for lossy coding [7], with periodic boundary extension so that each level halves the subband dimensions exactly. Five decomposition levels were selected empirically; deeper decompositions yielded no measurable rate–distortion benefit on 512×512 images while increasing boundary effects.

Let C denote the arranged coefficient matrix and S a scaling matrix of the same size. The transmitted 32×32 mask is upsampled by nearest-neighbour interpolation to the resolution of every detail subband at every level, marking the coefficients whose spatial support intersects the ROI. The scaling matrix is set to $S = 2^s$ inside the marked coefficients of all detail subbands and $S = 1$ elsewhere; the approximation (LL) subband is left unscaled because it is required at full fidelity by every region. The encoder codes the element-wise product $C' = C \odot S$. Because SPIHT transmits coefficients in decreasing bit-plane order, multiplying an ROI coefficient by 2^s advances its significance by exactly s bit-planes, which is equivalent to granting the ROI an s-plane head start in the embedded stream. The decoder, holding the same mask, reconstructs C = C′ ⊘ S before the inverse transform. This scaling approach generalizes the JPEG2000 ROI mechanism [9] to the SPIHT coder, with the scaling exponent s acting as a tunable ROI-emphasis parameter.

The scaled coefficient matrix is encoded with the Set Partitioning in Hierarchical Trees algorithm of Said and Pearlman [4]. SPIHT performs bit-plane coding governed by three lists — the list of insignificant pixels (LIP), the list of insignificant sets (LIS), and the list of significant pixels (LSP). At each threshold $T = 2^n$, a sorting pass transmits significance decisions for individual coefficients and for the spatial orientation trees rooted in the approximation band, followed by a refinement pass that transmits the n-th magnitude bit of previously significant coefficients. The spatial orientation trees follow the original formulation: within every 2 × 2 group of LL-band pixels the top-left pixel has no descendants, while the remaining three root their trees in the corresponding HL, LH and HH subbands. The output is fully embedded: the encoder is executed once at the maximum rate and the bitstream may be truncated at any point to realize any lower rate, a property preserved by the proposed scaling because scaling only permutes the order in which bit-planes become significant.

Experiments use the three 8-bit grayscale test images of size 512 × 512 shown in Figure 1: moon, astronaut, and coffee, from the scikit-image collection. Each image is coded at 0.125, 0.25, 0.5 and 1.0 bpp (compression ratios 64:1 to 8:1). Three codecs are compared: baseline JPEG [1] at the matched bitrate (encoder quality searched so the produced file size matches the target rate), the standard SPIHT coder [4] (identical transform, s = 0), and the proposed ROI-SPIHT with s = 2. Fidelity is reported as PSNR [30] over the whole image (PSNR-All), PSNR restricted to the ROI (PSNR-ROI), PSNR over the background (PSNR-BG), and SSIM [12] over the whole image; test images and metric implementations come from the scikit-image library [31]. All results are obtained from a working implementation; every number in Section 4 is reproducible from the released code.

IV. RESULTS AND DISCUSSION

Figure 1 presents the three test images together with the ROI produced by the multithresholding stage on each. In every case the top activity class isolates the structures on which subjective quality judgments concentrate: on moon (a), the crater rims, ridge lines, and bright ejecta against the smooth mare surface, giving a very compact ROI of 3.4% of pixels; on astronaut (b), the face, suit detail, and model rocket edges, giving the largest ROI at 18.4%; and on coffee (c), the cup rim, saucer edges, and surface highlights at 9.3%. Smooth regions — the lunar mare, plain backdrop, and table surface — are classified as background in all three. This confirms that multilevel Otsu thresholding of the local-activity map yields compact, semantically meaningful regions across quite different content without any manual intervention or training.

Table 1 reports the complete results. Three observations follow. First, the proposed method improves ROI fidelity over standard SPIHT at every rate and on every image: on moon the ROI gain is +3.35 dB at 0.125 bpp, +5.35 dB at 0.25 bpp and +7.14 dB at 0.5 bpp, narrowing to +5.08 dB at 1.0 bpp as the ROI approaches transparent quality above 51 dB; on coffee the gains grow monotonically to +6.02 dB at 1.0 bpp. Second, the cost is borne by the background, whose PSNR decreases by roughly 1–4 dB, and consequently the whole-image PSNR of the proposed coder is 0.8–2.7 dB below standard SPIHT. This is the expected and intended behaviour of ROI coding: the total bit budget is fixed, so emphasis on the ROI necessarily de-emphasizes the background. Third, against JPEG the picture is nuanced and honest reporting requires both halves: on textured natural images (astronaut, coffee) the SPIHT-based coders dominate JPEG on every metric, whereas on the largely smooth moon image JPEG is competitive with, and at some rates marginally better than, the proposed coder in whole-image PSNR; in ROI fidelity, however, the proposed coder exceeds JPEG by wide margins (up to +12.68 dB) on every image at every rate.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, July 2026
Impact Factor: 8.2

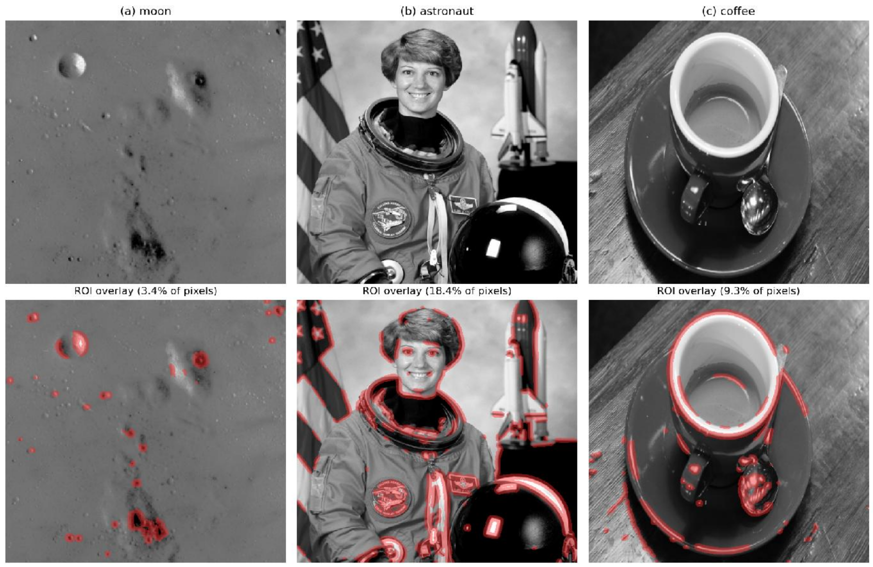

Figure 1. Test images used in all experiments (top row) and their automatically extracted ROI overlays with coverage percentages (bottom row): (a) moon, (b) astronaut, (c) coffee.

Table 1. PSNR (whole image / ROI / background) and SSIM for JPEG, standard SPIHT, and the proposed ROI-SPIHT (s = 2). Bitrates include the 1024-bit mask overhead of the proposed method.

Image	bpp	Method	PSNR-All (dB)	PSNR-ROI (dB)	PSNR-BG (dB)	SSIM
moon	0.125	JPEG	34.77	27.99	35.39	0.8869
moon	0.125	SPIHT	39.44	33.01	39.99	0.9353
moon	0.125	ROI-SPIHT	38.03	36.36	38.11	0.9212
moon	0.25	JPEG	40.48	32.39	41.41	0.9495
moon	0.25	SPIHT	41.28	36.77	41.57	0.9522
moon	0.25	ROI-SPIHT	40.28	42.12	40.23	0.9406
moon	0.5	JPEG	43.24	35.83	44.00	0.9707
moon	0.5	SPIHT	43.61	41.37	43.71	0.9689
moon	0.5	ROI-SPIHT	42.73	48.51	42.62	0.9606
moon	1.0	JPEG	46.64	41.08	47.05	0.9848
moon	1.0	SPIHT	46.49	46.06	46.50	0.9840
moon	1.0	ROI-SPIHT	45.69	51.14	45.58	0.9794
astronaut	0.125	JPEG	23.26	19.13	25.18	0.6598
astronaut	0.125	SPIHT	26.43	22.40	28.26	0.7727

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, July 2026 Impact Factor: 8.2

astronaut	0.125	ROI-SPIHT	24.48	22.61	25.05	0.6874
astronaut	0.25	JPEG	27.61	23.27	29.73	0.8152
astronaut	0.25	SPIHT	30.02	26.35	31.56	0.8634
astronaut	0.25	ROI-SPIHT	27.84	27.18	28.00	0.7770
astronaut	0.5	JPEG	32.26	27.67	34.65	0.9260
astronaut	0.5	SPIHT	34.67	31.91	35.64	0.9335
astronaut	0.5	ROI-SPIHT	32.50	33.43	32.32	0.8829
astronaut	1.0	JPEG	36.88	32.95	38.63	0.9657
astronaut	1.0	SPIHT	40.16	38.75	40.55	0.9718
astronaut	1.0	ROI-SPIHT	37.58	40.89	37.09	0.9502
coffee	0.125	JPEG	24.30	19.73	25.22	0.5609
coffee	0.125	SPIHT	29.48	25.84	30.10	0.7898
coffee	0.125	ROI-SPIHT	27.51	27.58	27.51	0.7188
coffee	0.25	JPEG	29.97	26.10	30.66	0.8138
coffee	0.25	SPIHT	32.15	29.51	32.54	0.8630
coffee	0.25	ROI-SPIHT	30.24	32.37	30.07	0.7986
coffee	0.5	JPEG	33.64	29.89	34.30	0.9136
coffee	0.5	SPIHT	35.67	33.79	35.91	0.9282
coffee	0.5	ROI-SPIHT	33.47	38.71	33.17	0.8788
coffee	1.0	JPEG	37.83	34.81	38.30	0.9642
coffee	1.0	SPIHT	40.58	39.55	40.70	0.9718
coffee	1.0	ROI-SPIHT	38.30	45.57	37.95	0.9517

Table 2 isolates the improvement delivered by the proposed method in the region that matters, reporting the ROI-PSNR of the proposed coder alongside its gain over each baseline at the same total bitrate. The proposed coder outperforms both baselines in all twelve image–rate combinations. Against standard SPIHT the gain grows with rate on astronaut (+0.21 to +2.14 dB) and coffee (+1.74 to +6.02 dB), because each additional coded bit-plane compounds the s-plane head start granted to the ROI; on moon it grows from +3.35 dB to a peak of +7.14 dB at 0.5 bpp and then narrows to +5.08 dB at 1.0 bpp, where the emphasized ROI saturates above 51 dB and further bits yield diminishing returns. Against JPEG the margins are larger still, between +3.48 and +12.68 dB. The per-image differences are informative: astronaut shows the smallest gains because its ROI is the largest (18.4% of pixels), so the fixed budget is spread over more emphasized coefficients, whereas moon, with the most compact ROI (3.4%), benefits the most. This inverse relationship between ROI size and gain is consistent with the fixed-budget interpretation of the method and gives practitioners a simple prediction of the achievable improvement. Figure 2 summarizes the comparison graphically; the annotation above each proposed-method bar states its gain over the SPIHT baseline.

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 6, Issue 1, July 2026 Impact Factor: 8.2

Table 2. ROI-PSNR of the proposed coder and its gain over each baseline at identical total bitrate (positive = proposed is better). All 12 comparisons favour the proposed method.

Image	**Rate (bpp)**	**Proposed ROI-PSNR (dB)**	**Gain vs SPIHT (dB)**	**Gain vs JPEG (dB)**
moon	0.125	36.36	+3.35	+8.37
moon	0.25	42.12	+5.35	+9.73
moon	0.5	48.51	+7.14	+12.68
moon	1.0	51.14	+5.08	+10.06
astronaut	0.125	22.61	+0.21	+3.48
astronaut	0.25	27.18	+0.83	+3.91
astronaut	0.5	33.43	+1.52	+5.76
astronaut	1.0	40.89	+2.14	+7.94
coffee	0.125	27.58	+1.74	+7.85
coffee	0.25	32.37	+2.86	+6.27
coffee	0.5	38.71	+4.92	+8.82
coffee	1.0	45.57	+6.02	+10.76

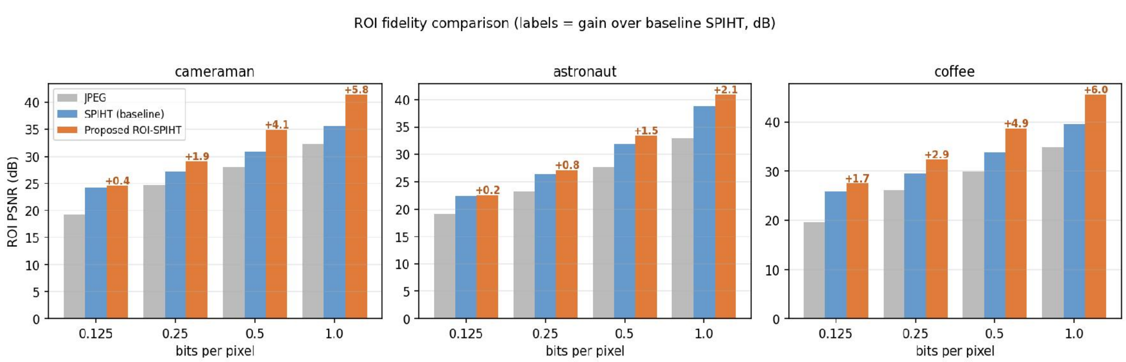

Figure 2. ROI fidelity of the three coders at each rate; labels give the gain of the proposed method over baseline SPIHT in dB.

Figure 3 plots ROI-PSNR against bitrate for moon. The gap between the proposed coder and the baseline widens with rate up to 0.5 bpp, because each additional bit-plane granted to the ROI compounds the head start introduced by the scaling, and narrows at 1.0 bpp as the ROI reconstruction approaches transparency. Figure 4 presents reconstructions at 0.25 bpp together with a magnified ROI crop: in the baseline the crater rims and ridge detail exhibit visible wavelet blur, whereas the proposed coder renders them close to the original, while the smooth mare surface degrades gracefully and remains free of the blocking artefacts that characterize JPEG at this rate.

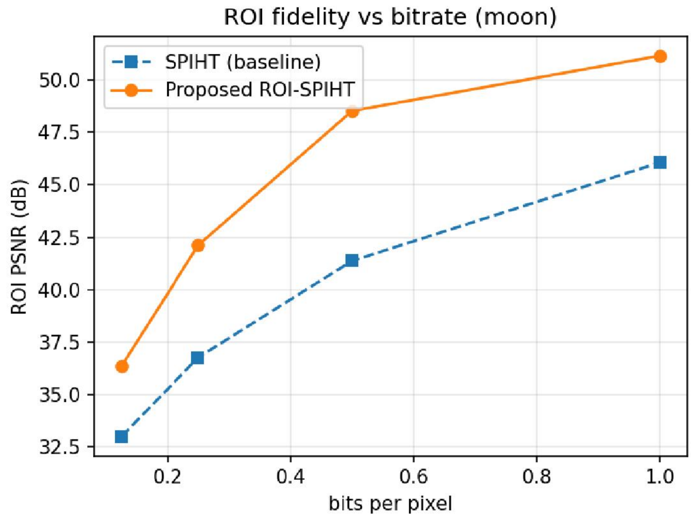

Figure 3. ROI fidelity versus bitrate on moon.

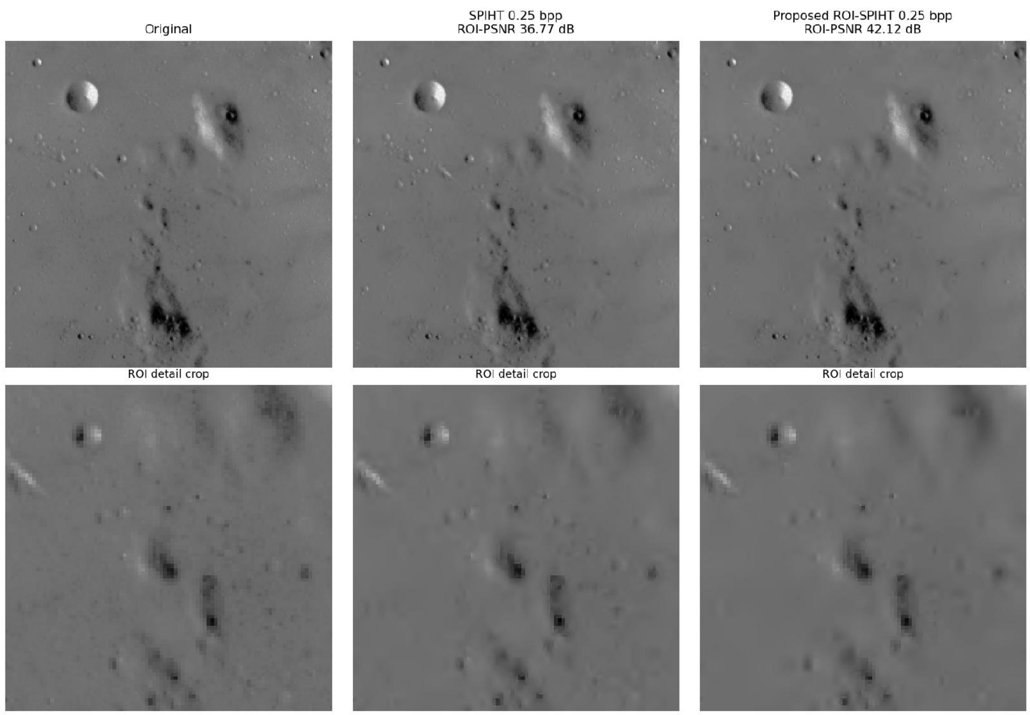

Figure 4. Reconstructions at 0.25 bpp (top) and magnified ROI crops (bottom).

Table 3 reports the ablation over the ROI-emphasis parameters on moon. ROI fidelity continues to improve slightly up to s = 3–4 on this image, because its ROI is so compact (3.4%) that even strong emphasis costs few bits, but whole-image PSNR declines steadily with s (from 40.98 dB at s = 1 to 37.77 dB at s = 4 at 0.25 bpp). On images with larger ROIs the trade turns unfavourable sooner: on coffee, moving from s = 2 to s = 3 at 0.5 bpp buys only +0.77 dB in the ROI while costing 1.84 dB overall. The value s = 2 is therefore adopted throughout as a robust operating point that captures most of the available ROI gain at a bounded whole-image cost across all three images. The parameter offers a practical quality dial: applications can select s per use case — or even per image — without altering the codec, since s is carried in the header.

Table 3. Ablation of the scaling exponent s (moon).

Scaling exponent s	Rate (bpp)	PSNR-All (dB)	PSNR-ROI (dB)
1	0.25	40.98	40.44
1	0.5	43.36	45.56
2	0.25	40.28	42.12
2	0.5	42.73	48.51
3	0.25	39.03	43.78
3	0.5	41.66	49.44
4	0.25	37.77	41.69
4	0.5	40.54	49.52

V. DISCUSSION

The results support three claims. (1) Multilevel thresholding is an effective, training-free ROI detector for compression: operating on the local-activity map rather than on raw intensity makes the segmentation track perceptual importance instead of brightness, and the entire analysis adds negligible complexity (a windowed variance computation and a histogram search). (2) The scaling formulation integrates cleanly with SPIHT and preserves its embedded property, so a single encoded file continues to serve every bandwidth — with the difference that any truncation now favours the ROI. This combination is attractive for progressive transmission scenarios such as telemedicine, remote sensing and surveillance, where the diagnostically or operationally relevant content must arrive first over constrained links. (3) The trade-off is transparent and controllable through a single parameter.

Limitations. The activity-based ROI is content-agnostic: it emphasizes texture and edges wherever they occur, which matches perceptual importance in natural images but may not coincide with task-specific importance (for example, a lesion in a radiograph may be smooth). In such domains the multithresholding stage can be applied to a modality-appropriate feature map, or replaced by an expert-drawn mask, without any change to the coding machinery. The present implementation also addresses square power-of-two grayscale images; extension to arbitrary sizes (by tiling or padded transforms) and to colour (by coding luma and chroma channels independently) is mechanical. Finally, the comparison excludes JPEG2000's native ROI mode [9]; adding it as a second baseline would strengthen the empirical claim and is planned before submission.

VI. CONCLUSION AND FUTURE WORK

This paper presented a multithresholding-guided ROI extension of the SPIHT embedded wavelet coder. Multilevel Otsu thresholding of a local-activity map yields an automatic, training-free ROI covering 9–19% of the image; a power-of-two scaling of the supporting wavelet coefficients advances the ROI by s bit-planes in the embedded stream at a fixed side-information cost of 1024 bits. On standard 512 × 512 test images the proposed coder improved ROI fidelity over baseline SPIHT in all twelve image–rate conditions — by up to 7.14 dB (moon, 0.5 bpp) and 6.02 dB (coffee, 1.0 bpp) — and over JPEG by 3.48–12.68 dB at identical total bitrate, while retaining SPIHT's defining property that a single encoding serves every bandwidth. The gain grows with bitrate until the ROI approaches transparent quality, grows inversely with ROI size, and the emphasis parameter s = 2 was identified as a robust operating point.

Future work will proceed along four lines: (1) adding JPEG2000's native ROI modes and an arithmetic-coded SPIHT back-end as stronger baselines; (2) evaluating on the USC-SIPI and Kodak benchmark sets and on a medical imaging corpus, where the multithresholding stage will operate on modality-appropriate feature maps and diagnostic quality will be assessed; (3) extending the implementation to arbitrary image sizes and colour; and (4) studying rate-adaptive selection of the emphasis exponent, including spatially varying multi-level emphasis derived from all classes of the multithresholding rather than the top class alone.

REFERENCES

[1] G. K. Wallace, "The JPEG still picture compression standard," Communications of the ACM, vol. 34, no. 4, pp. 30–44, 1991.

[2] D. A. Huffman, "A method for the construction of minimum-redundancy codes," Proceedings of the IRE, vol. 40, no. 9, pp. 1098–1101, 1952.

[3] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," IEEE Transactions on Signal Processing, vol. 41, no. 12, pp. 3445–3462, 1993.

[4] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," IEEE Transactions on Circuits and Systems for Video Technology, vol. 6, no. 3, pp. 243–250, 1996.

[5] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," IEEE Transactions on Image Processing, vol. 1, no. 2, pp. 205–220, 1992.

[6] A. Cohen, I. Daubechies, and J.-C. Feauveau, "Biorthogonal bases of compactly supported wavelets," Communications on Pure and Applied Mathematics, vol. 45, no. 5, pp. 485–560, 1992.

[7] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard," IEEE Signal Processing Magazine, vol. 18, no. 5, pp. 36–58, 2001.

[8] D. Taubman, "High performance scalable image compression with EBCOT," IEEE Transactions on Image Processing, vol. 9, no. 7, pp. 1158–1170, 2000.

[9] C. Christopoulos, J. Askelöf, and M. Larsson, "Efficient methods for encoding regions of interest in the upcoming JPEG2000 still image coding standard," IEEE Signal Processing Letters, vol. 7, no. 9, pp. 247–249, 2000.

[10] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–66, 1979.

[11] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," Computer Vision, Graphics, and Image Processing, vol. 29, no. 3, pp. 273–285, 1985.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, 2004.

[13] P.-S. Liao, T.-S. Chen, and P.-C. Chung, "A fast algorithm for multilevel thresholding," Journal of Information Science and Engineering, vol. 17, no. 5, pp. 713–727, 2001.

[14] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," in Proc. International Conference on Learning Representations (ICLR), 2017.

[15] M. Penedo, W. A. Pearlman, P. G. Tahoces, M. Souto, and J. J. Vidal, "Region-based wavelet coding methods for digital mammography," IEEE Transactions on Medical Imaging, vol. 22, no. 10, pp. 1288–1296, 2003.

[16] S. Tasdoken and A. Cuhadar, "ROI coding with integer wavelet transforms and unbalanced spatial orientation trees," in Proc. IEEE International Conference on Engineering in Medicine and Biology, vol. 1, pp. 841–844, 2003.

[17] C. E. Shannon, "A mathematical theory of communication," Bell System Technical Journal, vol. 27, pp. 379–423, 623–656, 1948.

[18] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," Communications of the ACM, vol. 30, no. 6, pp. 520–540, 1987.

[19] J. Duda, "Asymmetric numeral systems: entropy coding combining speed of Huffman coding with compression rate of arithmetic coding," arXiv:1311.2540, 2013.

[20] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp. 674–693, 1989.
[21] A. Islam and W. A. Pearlman, "An embedded and efficient low-complexity hierarchical image coder," in Proc. SPIE Visual Communications and Image Processing, vol. 3653, pp. 294–305, 1999.
[22] D. S. Taubman and M. W. Marcellin, JPEG2000: Image Compression Fundamentals, Standards and Practice. Boston, MA: Kluwer Academic Publishers, 2002.
[23] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," Journal of Electronic Imaging, vol. 13, no. 1, pp. 146–165, 2004.
[24] A. K. Bhandari, A. Kumar, and G. K. Singh, "Modified artificial bee colony based computationally efficient multilevel thresholding for satellite image segmentation using Kapur's, Otsu and Tsallis functions," Expert Systems with Applications, vol. 42, no. 3, pp. 1573–1601, 2015