# WAVELENGTH AS TIME: HYPERSPECTRAL COMPRESSION WITH VIDEO CODECS

*Arthur Lorin*[1]*, Maxime Istasse*[1]*, Damien Vincke*[2]*, Antoine Deryck*[2]*,*
*Juan Antonio Fernandez Pierna*[2]*, Paul Vanabelle*[1]

[1]CETIC, DSIDE, Charleroi, Belgium
[2]CRA-W, Gembloux, Belgium

## ABSTRACT

A hyperspectral cube can be viewed as a sequence of single-band images along the spectral axis, in effect a monochrome video. We take this view literally: an unmodified video codec receives the near-duplicate bands as frames, and its inter-frame prediction replaces the spectral transform, with no training. The reference standard for onboard compression never exceeds about $10\times$ on 16-bit cubes. Across 16 cubes from three sensors (1040 runs), AV1, HEVC, and VVC extend the reachable operating range one to three orders of magnitude past that ceiling. Spectral fidelity (spectral angle) degrades gracefully across this range. In a companion segmentation study, the codecs retain 96–99% of the uncompressed task accuracy at roughly $90\times$ compression; the near-lossless standard cannot pass $16\times$. This questions how often bit-exact retention is worth its cost. Where bit-exactness is required, a recent per-band still-image codec is the best standard option, up to $6.2\times$, through entropy coding of the low raw dynamic range. The video recipe needs no dedicated graphics hardware, decodes with hardware and browser-native decoders, and a variant is deployed in a web-based hyperspectral annotation platform.

***Index Terms*—** hyperspectral imaging, video compression, AV1, HEVC, VVC, JPEG XL, CCSDS 123, BD-Rate

## 1. INTRODUCTION

Hyperspectral imaging (HSI) sensors sample hundreds of narrow spectral bands per scene: a single 16-bit cube can exceed a gigabyte (Table 1). Such volumes stress satellite downlink, ground storage, and web delivery; sharing cubes in web applications is a live use case, exemplified by the *SegLab-HSI* annotation platform, which integrates a variant of this approach (Sec. 5.1).

CCSDS 123.0-B-2 is the standardized answer for onboard HSI compression [1]. It provides lossless and near-lossless coding with a bounded per-sample error. By design it never operates below about 1.5 bpppb (bits per pixel per band, the average bits spent per sample). Because our cubes store 16-bit samples, this floor already gives at least a $10\times$ size reduction. Any comparison at lower rates probes a region the standard was not built for. Still-image standards (JPEG 2000, JPEG XL) exploit inter-band correlation weakly or not at all (per-band intra); learned methods reach low rates but need training and graphics processing units (GPUs) and lack deployable decoders (Sec. 2).

Our idea is simple: treat wavelength as time. A cube becomes a monochrome video, and inter-frame prediction performs spectral decorrelation without an explicit spectral transform and without training.

This paper tests that idea in both the lossy and lossless regimes: we evaluate off-the-shelf, unmodified video codecs (AV1, HEVC, VVC) in this band-as-frame mapping against CCSDS 123.0-B-2 and still-image baselines, across 16 cubes from three sensors (seven distinct scenes) and 1040 runs, with the following contributions.

- A config-driven benchmark of the band-as-frame approach, reproducible end to end on the two public cubes. It compares AV1, HEVC, and VVC against CCSDS 123.0-B-2, JPEG 2000 (a per-band intra baseline), JPEG XL, and FFV1, with rate-distortion (RD) curves, a two-anchor BD-Rate (Bjøntegaard-Delta rate) protocol anchored on CCSDS 123.0-B-2 and HEVC, and spectral metrics (spectral angle mapper, SAM).
- Quantified evidence that these video codecs extend the reachable rate-quality region one to three orders of magnitude below the roughly 1.5 bpppb CCSDS rate floor, to 0.0006–0.15 bpppb at 31–56 dB min-max-normalized peak signal-to-noise ratio (PSNR, Sec. 4). BD-Rate over the overlap region is a secondary quantification.
- A bit-exact reference point: JPEG XL, a per-band intra codec that exploits no inter-band correlation, is the best standard bit-exact codec on every cube. Its advantage comes from entropy coding of the low raw dynamic range. With the reference libaom toolchain, AV1 16-bit lossless is not bit-exact over multi-frame sequences; we document this and work around it with a bit-exact dual-stream split.
- A companion validation of the recommended recipe: a separate wheat-disease segmentation study (Sec. 5.1) finds that the band-as-frame codecs retain most of the uncompressed task accuracy at rates CCSDS cannot reach, and the codec ranking carries over to the task. This questions the need for lossless retention in this workflow.

## 2. RELATED WORK

**Standards and classics.** CCSDS 123.0-B-2 is the reference on-board standard for spaceborne instruments [1]: a predictive coder for lossless and near-lossless compression whose near-lossless mode enforces a bounded per-sample error through an $L_\infty$ constraint. The CCSDS suite also standardizes transform-based lossy image coding (CCSDS 122.0-B, extended to spectral data by 122.1-B); we scope this study to 123.0-B-2, the predictive standard deployed as the HSI reference, and leave the transform-based path to future work. JPEG 2000 is a widely used still-image standard for hyperspectral data [2]; it supports spectral decorrelation via Part 2 multi-component transforms, but our baseline is per-band intra (the ffmpeg encoder applies no cross-band transform), so we do not benchmark Part 2. JPEG XL is the modern still-image standard [3], and FFV1 is our archival intra-frame lossless reference [4].

**Video codecs for hyperspectral data.** Standard video codecs have been applied to spectral cubes before: Santos et al. evaluate H.264/AVC with bands treated as frames [5]. Paul et al. build a modified HEVC pipeline that predicts a reference band with a Gaussian-mixture model [6]. Seltsam et al. apply VVC to multispectral images, adding principal component analysis (PCA) variants for decorrelation [7]. We differ in scope and protocol. We evaluate these codecs unmodified, with no added spectral transform, on a 16-cube grid from three sensors. We quantify the operating range they reach relative to CCSDS 123.0-B-2 under a two-anchor BD-Rate protocol, and audit lossless bit-exactness.

**Learned compression.** Spatio-spectral autoencoders learn a joint spatial-spectral latent [8], implicit neural representations fit a network to one cube and store its weights [9], and the HySpecNet-11k benchmark supports such learned methods at scale [10]. Our approach needs no training or fine-tuning and no GPU; standardized decoders already exist in hardware and web browsers, an orthogonal deployment advantage we exercise in Sec. 5.1.

## 3. METHOD

**Band as frame.** We map each cube onto a monochrome video sequence: each spectral band becomes one frame in native band order. Frame extraction writes every band as an uncompressed single-channel 16-bit TIFF, and this sequence feeds every codec unchanged. We apply no explicit spectral decorrelation before encoding (no principal component transform, Karhunen-Loève transform, or spectral wavelet); inter-band redundancy is left to each codec's own prediction, which for the video codecs is inter-frame prediction across the spectral axis. This is our central hypothesis: a standard video codec, given bands as frames, recovers spectral redundancy well enough to be useful. For the video codecs the entire spectral axis is one group of pictures (GOP), so every band is available for inter-frame prediction.

**Bit-depth policy.** We extract frames at 16 bits and set the encode bit depth per codec. For the lossy paths of AV1, HEVC, JPEG 2000, and JPEG XL we encode at 12 bits (`gray12le`). VVC is the exception: `vvencapp` exposes a 10-bit Main10 profile, so we right-shift to 10 bits. We state this neutrally: it is a tool limit, not a design choice. The two fewer bits cap the PSNR that VVC can reach near transparency, while the effect at low rates is indeterminate, because pre-quantized 10-bit input can even reduce rate. Every reconstruction is decoded back to 16 bits, and we compute all metrics against the 16-bit reference cube regardless of encode depth.

**Lossless dual-stream scheme.** Strict lossless AV1 over multi-frame sequences is not bit-exact on our reference toolchain. We probe this through the libavif AVIF front-end (`avifenc`/`avifdec` 1.3.0 on the aom 3.13.1 backend, the same libaom AV1 family as our ffmpeg `libaom-av1` lossy path). A 91-frame sequence under `--lossless` reconstructs with a maximum absolute difference of 117 digital numbers. We report this as a toolchain observation, without isolating the failing stage. The HEVC and VVC lossless modes cannot substitute for it: x265 and vvencapp cap at 12 and 10 bits, below the 16-bit samples. This motivates a split-stream scheme. We split each 16-bit sample into a high byte (`x >> 8`) and a low byte (`x & 0xFF`). We encode the high byte with x265 in lossless mode and the low byte with FFV1, then recombine as `(high << 8) | low`, bit-exact by construction.

**Reproducibility.** Every experiment is one self-contained YAML config generated from a sweep grid; each run embeds the dataset, codec, quality, and a config hash, with durable per-run manifests. The grid ran to completion (1040 runs, 0 missing), and the configs regenerate every RD plot; the pipeline, configs, and per-run manifests will be released upon acceptance.

## 4. EXPERIMENTAL SETUP

### 4.1. Datasets

We evaluate 16 cubes covering seven distinct scenes from three sensors (AVIRIS, Specim FX10, Specim FX17), organized by scene in Table 1. Only the two publicly reproducible cubes are named: the AVIRIS Indian Pines site 3 cube [11] and the FX10 DA-9 reflectance cube. The 14 proprietary cubes get short IDs: Refl-16 and Refl-12 are one FX17 wheat reflectance scene at 16-bit and at a 12-bit requantization, and the raw digital number (DN) industrial cubes are FX10-1 to FX10-4 and FX17-1 to FX17-8. All cubes are ENVI uint16, and we compress the full band count with no band filtering.

The scene grouping is what keeps the 16 cubes from being treated as 16 independent scenes. The four industrial specimens (iron, manganese, and their polluted variants) are each

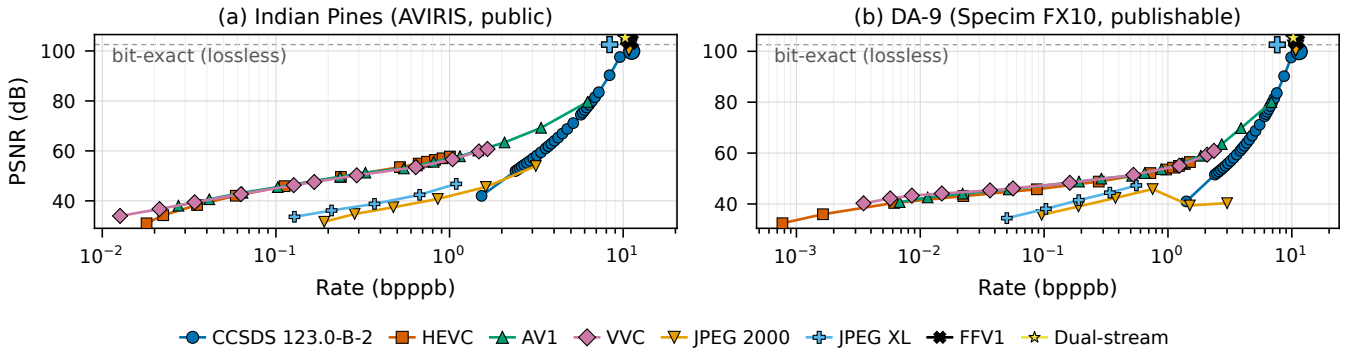



**Fig. 1**. Rate-distortion (bpppb vs PSNR) on the public Indian Pines (a) and publishable FX10 DA-9 (b) cubes; the two panels span different rate ranges. Bit-exact runs (zero error, PSNR undefined) sit on the dashed strip, drawn at an arbitrary height.

**Table 1**. The 16 evaluated cubes (ENVI uint16, full band count), grouped by their seven distinct scenes. Industrial specimens are imaged by both sensors. Cube sizes (H×W×bands): Indian Pines 145×145×220; reflectance cubes 640×2468×112; DA-9 1024×1785×224; FX17 raw DN 640×790–815×224; FX10 raw DN 1024×1445–1557×448. Indian Pines and DA-9 are publicly reproducible.

| Scene | Cube ID(s) | Radiom. |
|---|---|---|
| Indian Pines | Indian Pines | raw DN |
| Wheat | Refl-16, Refl-12 | refl. |
| DA-9 | DA-9 | refl. |
| Iron | FX10-1, FX17-1–3 | raw DN |
| Manganese | FX10-2, FX17-4–6 | raw DN |
| Iron, polluted | FX10-3, FX17-7 | raw DN |
| Manganese, polluted | FX10-4, FX17-8 | raw DN |

imaged by both sensors, and the six FX17 clean-iron and clean-manganese cubes (FX17-1 to FX17-6) are two specimens each scanned under three reference tiles (20% grey, 50% grey, and white); Refl-16 and Refl-12 are one cube at two quantizations. For the polluted specimens we keep the white-reference variant only and exclude the grey-reference variants to avoid pseudo-replication.

We keep the industrial cubes as raw DN deliberately: the shipped grey references are darker than the scene, so reflectance calibration saturates and clips about 94% of the samples. Raw DN preserves the raw dynamic range of roughly 9 to 11 bits and matches the onboard compression use case that CCSDS 123.0-B-2 [1] targets. Of the 16 cubes, 14 are proprietary; Indian Pines (CC0, Purdue PURR DOI 10.4231/R7RX991C) and DA-9 (freely publishable) support external reproduction. The proprietary cubes may be shared on reasonable request, subject to the partner agreement (CRA-W, QualiSpectra).

### 4.2. Codecs and operating points

We compare seven codecs on the same band sequences. AV1 uses ffmpeg libaom-av1 at 7 CRF (constant rate factor) points from 0 to 50, with `cpu-used 6` and a single keyframe. HEVC uses x265 at 8 CRF points from 2 to 45, preset faster, Main12 profile. VVC uses vvencapp 1.13.1 at 5 QP (quantization parameter) points from 22 to 42, Main10 profile, preset faster. CCSDS 123.0-B-2 uses the ccsds123 Issue 2 encoder at 1 lossless plus 29 near-lossless points, the densest curve and our rate-distortion anchor. Our JPEG 2000 baseline is a per-band intra still-image ffmpeg configuration (implementation-limited) at 6 rates, with lossless coding via libopenjpeg reversible 5/3. JPEG XL uses cjxl per band at distances 1 to 16 and `d=0` for lossless, effort 7. FFV1 is an intra 16-bit lossless codec. These presets favor tractable software runtime, not best rate-distortion. For the three figure cubes the RD curves add high-quality points near the CCSDS operating zone (AV1 CRF 1–8, HEVC CRF 0–1 and 3–4, VVC QP 0–20); BD-Rate uses the common base grid (1040 runs) only, so these points refine the plotted curves but not the reported medians.

### 4.3. Metrics

We compute PSNR as $10\log_{10}(1/\mathrm{MSE})$ on cubes min-max normalized to $[0, 1]$, i.e. PSNR with peak 1: the peak is each cube's actual dynamic range, not the 16-bit container maximum, so dB values stay comparable across effective bit depths. We compute the spectral angle mapper (SAM) per pixel in degrees, reported as the 99th percentile (SAM p99, lower is better). We report BD-Rate [12] against two anchors: CCSDS 123.0-B-2 (domain reference) and HEVC (best-video reference), where a negative value means less rate at equal quality. For lossless runs we verify bit-exactness (maximum absolute difference 0). CCSDS near-lossless enforces a per-sample maximum-error ($L_\infty$) bound, whereas the video codecs optimize average distortion and give no per-sample guarantee, a distinction the maximum absolute difference makes observable.

For scenes with a large dark background we mask the foreground with an Otsu threshold on mean intensity; the mask self-disables when no dark cluster is present. Figures and BD-Rate tables use full-frame SAM p99; we quote foreground-masked values in the text where relevant. We also compute mean structural similarity (SSIM), spatial correlation, normalized RMSE, and spectral information divergence, not tabulated for space.

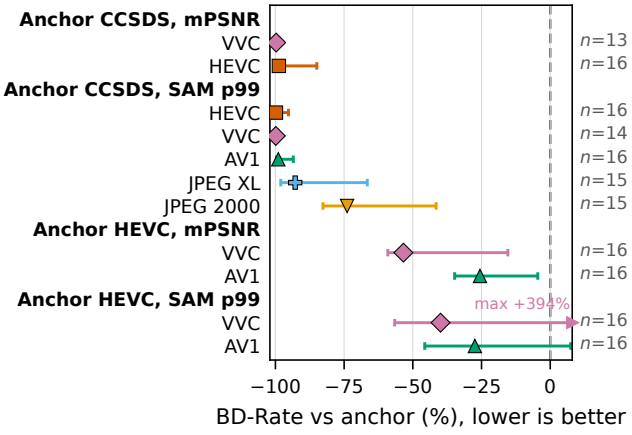



**Fig. 2**. BD-Rate over the overlap region across the 16 cubes: median marker, min/max whiskers (arrow: off-scale maximum), $n$ cubes per row (negative = less rate at equal quality). Cells whose anchor overlap needs extrapolation are excluded; on the CCSDS/PSNR anchor this drops AV1, JPEG XL and JPEG 2000.

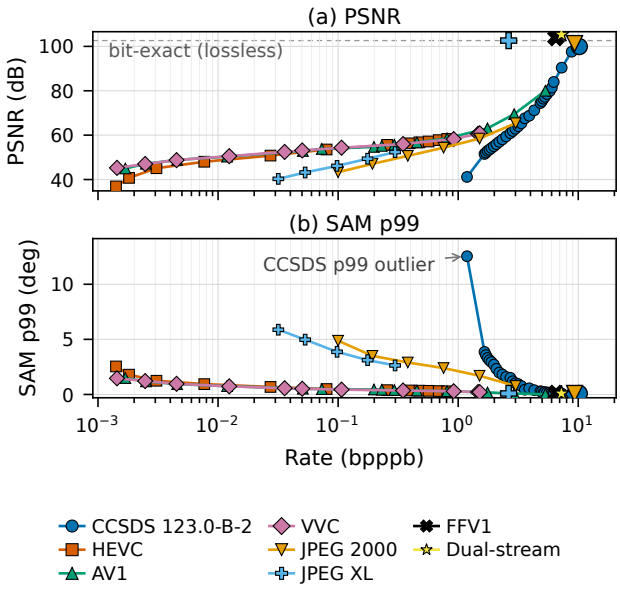



**Fig. 3**. Rate-distortion on cube FX17-1 (raw DN FX17, Sec. 4): (a) PSNR and (b) spectral fidelity (SAM 99th percentile, degrees, lower is better). The video-codec advantage holds on raw DN and for spectral angle, not only PSNR.

## 5. RESULTS

**Video codecs extend the operating range** (Fig. 1). CCSDS 123.0-B-2 never operates below a rate floor (the lowest bitrate it will produce) of about 1.5 bpppb, where it delivers roughly 41 to 42 dB PSNR. The video codecs continue one to three orders of magnitude below that floor, down to about 0.0006 bpppb; HEVC and VVC reach the lowest rates, with AV1 close behind. The per-band intra JPEG 2000 baseline sits between CCSDS and the video codecs, with a non-monotonic anomaly on DA-9 (Fig. 1b). CCSDS remains the only codec with a bounded per-sample error: at ∼42 dB PSNR on Indian Pines it bounds the worst case to 132 DN versus 1385 DN for HEVC, so the codecs serve different regimes. The absence of a bound does not imply harmful loss: the video codecs' SAM p99 stays low across their range (Fig. 3b) and the task absorbs the residual error (Sec. 5.1). The operating points are stable across the 16 cubes (PSNR at AV1 CRF 0 spans only $\pm 2.5$ dB around 80.5 dB); because the cubes cover seven distinct scenes, this spread understates scene-level variability.

**BD-Rate confirms the ranking** (Fig. 2). Against CCSDS on PSNR, the two reported video codecs need a small fraction of the anchor rate at equal quality (median $-99.7\%$ for VVC, $-98.7\%$ for HEVC). Against CCSDS on SAM p99, all five test codecs are strongly negative, so the rate advantage does not come at the cost of spectral angle; Fig. 3 shows both metrics on a raw DN FX17 cube. Among video codecs (anchor HEVC), VVC and AV1 improve on HEVC by a median $-53\%$ and $-26\%$ on PSNR ($-53\%$/$-22\%$ foreground-masked); both also improve on SAM p99 (median $-40\%$ for VVC, $-27\%$ for AV1). Neither foreground masking nor per-scene aggregation (the pseudo-replication of Sec. 4) changes any ranking; cited medians shift by only a few points. VVC achieves its PSNR lead while encoding at 10 bits (Sec. 3), whereas AV1 encodes at 12 bits and reaches the highest fidelity of any video codec here. AV1's absence from the CCSDS/PSNR cell (Fig. 2) is an overlap-extrapolation artifact, not a deficit: against HEVC it is second only to VVC. BD-Rate against CCSDS spans only a narrow shared range (CCSDS has no low-rate points), so the operating-range extension of Fig. 1 is the primary evidence.

**Robustness checks.** Splitting two cubes (Refl-16 and FX17-1) into six spatial tiles and re-running HEVC, JPEG XL and CCSDS near-lossless keeps the 95% confidence interval on the six-tile mean PSNR within $\pm 2.4$ dB of the full-cube value, so results are not driven by a single region. Full-frame SAM p99 is dominated by low-radiance background pixels, where the angle is noise-dominated: it exceeds the foreground-masked p99 by a median 2.5–10× per cube (worst on DA-9), so absolute SAM values on dark-background scenes should be read as foreground values.

**Where bit-exactness is required, a per-band still-image codec serves it best.** Two findings frame the bit-exact regime (Fig. 4). First, strict lossless AV1 is not bit-exact on our reference toolchain, which forces the dual-stream byte-split workaround (Sec. 3). The split is bit-exact by construction and beats CCSDS on every cube, but plain FFV1 beats it on most cubes (12 of 16), so the split itself costs efficiency. Second, even where coding is bit-exact, inter-band prediction does not pay: JPEG XL, an intra codec applied per band, exploits no inter-band correlation, yet gives the best standard-codec ratio on every cube, up to 6.2× on FX17 raw DN. Its advantage comes from entropy coding of the low raw dynamic range (∼ 9–11 bits) of these cubes, not from spectral modeling; in the lossy regime it sits far behind the video

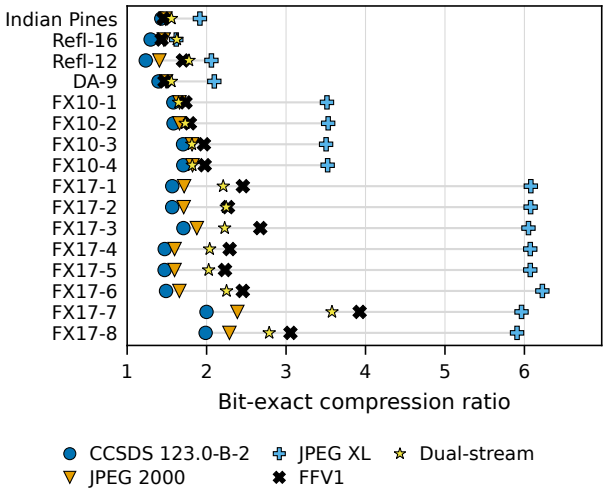



**Fig. 4**. Bit-exact compression ratios (source over compressed bytes) for the 16 cubes; video codecs are absent (Sec. 3).

codecs (Fig. 1). Only the non-standard dual-stream split beats it overall, on a single cube. Because spectral fidelity survives rates orders of magnitude lower, and task accuracy holds well below any bit-exact ratio (Sec. 5.1), bit-exact retention is a policy choice, not a technical necessity for the uses we tested.

### 5.1. Deployment and Downstream Impact

A variant of the recommended lossy recipe is deployed in *SegLab-HSI*, an HSI annotation platform built by CETIC with CRA-W (QualiSpectra project). It quantizes the working cube to 8-bit luma and encodes it client-side with the WebCodecs API, using our single-keyframe GOP and hardware HEVC, with H.264 and server-side decode as fallbacks. Only the bitstream and a JSON sidecar are stored, content-addressed, and reopened cubes are never re-encoded, so no generation loss accumulates. Cross-OS decode validation and in-application ratios are pending; the 8-bit point is a deployment choice.

Within the same project, a companion wheat-disease segmentation study (in preparation) quantifies the downstream cost. It confirms what the SAM curves suggest: without model adaptation, video codecs retain roughly 96–99% of the uncompressed task accuracy[1] up to about $90\times$ (77% at about $740\times$). At a matched $\sim$1.2 bpppb ($\sim 13\times$), near-lossless CCSDS keeps only about 25% and cannot compress past about $16\times$, because its bounded per-sample error must then admit thousands of DN the task cannot absorb. Fine-tuning on compressed cubes matches or slightly exceeds the uncompressed baseline, and this transfers across AV1, HEVC, and JPEG XL; the ranking carries over to the task.

## 6. CONCLUSION

We evaluated standard video codecs for HSI compression by mapping each band to a video frame; this wavelength-as-time coding extends the operating range far below the CCSDS 123.0-B-2 rate floor. Where bit-exact archival is required, we recommend the best standard codec, per-band JPEG XL: its advantage comes from entropy coding of the low raw dynamic range, not inter-band prediction. Our spectral and downstream evidence suggests that, for the uses we tested, this requirement is less common than assumed. CCSDS 123.0-B-2 remains the only bounded-error option, so the codecs are complementary, not competitors.

Four limitations bound these results: fast software presets not tuned for best rate-distortion; the encoder profile limits VVC to 10-bit; our data cover three sensors and seven distinct scenes; and timing is throughput only, not rate-matched.

A companion study already links these results to downstream analysis, measuring wheat-disease segmentation accuracy versus rate (Sec. 5.1); a low-rate comparison against the transform-based CCSDS 122 standards remains future work.

[1] Instance-level F1 at 10% intersection-over-union, over three seeds; the 90–740$\times$ range is measured on AV1.

## 7. REFERENCES


[1] Consultative Committee for Space Data Systems (CCSDS), "Low-Complexity Lossless and Near-Lossless Multispectral and Hyperspectral Image Compression," Blue Book, Recommended Standard CCSDS 123.0-B-2, CCSDS, 2019.

[2] Majid Rabbani and Rajan Joshi, "An overview of the JPEG 2000 still image compression standard," *Signal Processing: Image Communication*, vol. 17, no. 1, pp. 3–48, 2002.

[3] Jyrki Alakuijala, Ruud van Asseldonk, Sami Boukortt, et al., "JPEG XL next-generation image compression architecture and coding tools," in *Applications of Digital Image Processing XLII*, 2019, vol. 11137 of *Proc. SPIE*, p. 111370K.

[4] Michael Niedermayer, Dave Rice, and Jerome Martinez, "FFV1 video coding format versions 0, 1, and 3," RFC 9043, IETF, 2021.

[5] Lucana Santos, Sebastián López, Gustavo M. Callicó, et al., "Performance evaluation of the H.264/AVC video coding standard for lossy hyperspectral image compression," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 451–461, 2012.

[6] Manoranjan Paul, Rui Xiao, Junbin Gao, and Terry Bossomaier, "Reflectance prediction modelling for residual-based hyperspectral image coding," *PLOS ONE*, vol. 11, no. 10, pp. e0161212, 2016.

[7] Philipp Seltsam, Priyanka Das, and Mathias Wien, "Adaptive and scalable compression of multispectral images using VVC," in *Data Compression Conf. (DCC)*, 2023.

[8] Riccardo La Grassa, Cristina Re, Gabriele Cremonese, and Ignazio Gallo, "Hyperspectral data compression using fully convolutional autoencoder," *Remote Sensing*, vol. 14, no. 10, pp. 2472, 2022.

[9] Junqi Shi, Mingyi Jiang, Ming Lu, et al., "HINER: Neural representation for hyperspectral image," in *ACM Multimedia (MM '24)*, 2024.

[10] Martin Hermann Paul Fuchs and Begüm Demir, "HySpecNet-11k: A large-scale hyperspectral dataset for benchmarking learning-based hyperspectral image compression methods," in *IEEE IGARSS*, 2023, pp. 1779–1782.

[11] Marion F. Baumgardner, Larry L. Biehl, and David A. Landgrebe, "220 Band AVIRIS Hyperspectral Image Data Set: Indian Pine Test Site 3," Purdue Univ. Research Repository, 2015.

[12] Gisle Bjøntegaard, "Calculation of average PSNR differences between RD-curves," Doc. VCEG-M33, ITU-T VCEG, 2001.