Nick Clark · Published July 30, 2026 · PDF

When the Same Work Returns Wearing a New Hash

A rights administrator at a regional photography archive receives a file on a Tuesday and cannot answer a one-sentence question about it: is this ours? The picture is plainly the archive's shoreline photograph, but the delivered bytes were resized, re-encoded, and stripped of metadata somewhere inside a client's production pipeline, so the digest her catalog indexes on matches nothing. By Friday the invoice cycle closes and the print run has shipped. This article describes subject matter disclosed in PCT/US26/28630, which discloses deriving a content identifier from an artifact's internal variance structure rather than from its bytes, its storage path, or a sidecar metadata record.

A Tuesday in the rights office

The rights administrator at a mid-size regional photography archive opens a message from an institutional client she has billed every quarter for six years. The client's design agency has finished a printed exhibition guide. Before the invoice cycle closes on Friday, the client wants written confirmation that the photograph on page nine came from the archive's licensed collection and not from a free image site the agency also uses. Attached is the exact file the agency placed in the layout.

She drops it into her intake tool. The tool computes a SHA-256 digest over the delivered bytes and reads the embedded metadata block, which is how her catalog has recognized returning files since the collection was digitized. The digest matches nothing. The metadata block is empty. Reading the file properties, she can reconstruct roughly what happened: the agency scaled the image down to fit a layout grid, exported it at a reduced quality setting, and flattened it on the way out. None of that was done to hide anything. It is what her client's production pipeline does to a picture on its way to a press.

She can see the photograph on her screen. She recognizes the shoreline, the particular gray of that morning, the fence post at the left edge. Somewhere in her storage there is a master scan of the same negative. What she does not have is a way to put the question to her own catalog in the form she actually holds it: here is a picture, tell me which of my works this is. Her collection runs to roughly four hundred thousand scans. Visual review is not something she can complete before Friday, and she knows it by the time she has scrolled the third contact sheet.

The claim window that does not reopen

Friday arrives and she writes the honest answer, which is that she cannot confirm the file originated with the archive. The invoice goes out without the line item. The guide has already been printed and distributed at the exhibition opening, so the use is finished; there is no future moment at which she gets to observe that particular use again and decide differently. The file the agency sent her is all she will ever hold of it, and in her intake it is the one thing that does not identify itself.

That is the loss she can put a number on. The one she cannot is quieter. Eight months later, the same agency donates its working files back to the archive as part of a project handoff, and the resized export is ingested as a new accession under a new record, because in her catalog it is a new object: new digest, no metadata, no link to anything. Her collection now counts two works where the archive owns one. She has no record

that ties them, and she will not have one later, because nothing in her intake captured the relationship at the moment the second file arrived. Every downstream number she reports, holdings counts, per-work licensing history, the usage record she would need if the photographer's estate ever asks what became of this image, has drifted by one, silently, in a direction she cannot audit her way back out of.

The photographer died in 2009. Her archive is what remains of an accounting of his work, and its accuracy is the thing she is actually paid to protect.

Why her index keeps missing the same work

The shape of her problem is not that her tooling is careless. It is that the identifier her intake computes is a function of the delivered bytes, and every routine step her counterparties perform changes the bytes without changing the picture. In her workflow, the transformations that break a match are the same transformations that make a file usable: resizing for a layout, converting format for a web CMS, re-encoding at a lower quality for email. Her index is precise about a property her clients destroy in the ordinary course of doing their jobs.

The metadata path fails her for a related reason. The provenance block her archive writes on export rides alongside the delivered file rather than inside the structure her intake measures, so in her chain it survives for exactly as long as no tool along the way flattens, strips, or rewrites it, and past the point of delivery she controls none of those tools.

There is a third edge to it in her setup, and it is the one that keeps her caseload full. A digest gives her a binary answer. Match, or no match. Almost her entire working reality lives in between: a crop that keeps two thirds of the frame, a color-graded version, a composite where her shoreline is the background behind someone else's foreground. Were her index able to return a degree of correspondence rather than a verdict, most of

what lands on her desk would be triageable in an afternoon. As her catalog is configured today, all of it lands in the same undifferentiated pile marked no match, and she works it by memory.

What the filed architecture computes instead

PCT/US26/28630 discloses deriving an identifier, referred to in the filing as a UID, deterministically from a digital artifact's internal variance and structural features rather than from its storage location, file name, cryptographic key, or transmission metadata.

In an embodiment described in the filing, a raster image is first converted to a normalized grayscale floating-point representation under perceptual luminance weighting of approximately 0.299, 0.587, and 0.114 for the red, green, and blue channels. A multi-scale variance flow analyzer then subdivides the normalized scalar field into three nested grid resolutions, a coarse 8x8 grid, a medium 16x16 grid, and a fine 32x32 grid, computing a variance-based proxy per cell and aggregating a mean and standard deviation of variance at each scale.

From those aggregates the disclosure derives a nine-dimensional variance vector on three axes. The X-axis energy behavior vector encodes the slope of mean variance from coarse to fine resolution, its curvature at the medium scale, and the asymptotic fine-scale energy value. The Y-axis frequency compaction vector encodes the rate of change of variance standard deviation across scales, the spread factor between per-scale extremes, and a variance floor convergence value. The Z-axis structural phase persistence vector is derived from an eight-bin gradient magnitude histogram spanning zero to pi radians, canonicalized by rotating the dominant bin to index zero, yielding a horizontal-vertical orientation bias, a diagonal-axial bias, and a stability coefficient.

The filing further discloses a spatial decomposition path. A canonical normalization stage rescales the artifact to a 256 by 256 square canvas by uniform scaling against the longest source edge, centered on a black fill, with image smoothing disabled so that anti-aliasing does not introduce artificial variance along rescaled edges. An orientation canonicalization module rotates the artifact when its dominant gradient orientation exceeds approximately 0.1 radians. Four non-overlapping quadrants are then extracted, each run through the same nine-dimensional pipeline and hashed under a coarsened quantization scheme, with X and Y components quantized at a step of 1/32 and Z components at a step of 1/8, which the filing describes as absorbing JPEG compression noise and format conversion artifacts at the sub-image level. The four quadrant hashes are sorted in lexicographic order and assigned to canonical positions q0 through q3, which the filing describes as letting a rotated or mirrored version of an artifact produce the same set of sorted quadrant hashes. A multi-segment FNV-64 combiner applies five distinct initialization seeds to the global hash and the sorted quadrant hashes, producing five 64-bit segments concatenated into a 320-bit UID, of which the first 16 hexadecimal characters serve as a short-form identifier.

Two properties in the filing speak to what the rights administrator is trying to do. The disclosure describes the resulting representation as designed to be stable under format conversion, resolution rescaling within a defined canonical size, and lossy compression at moderate quality levels, while varying predictably with semantic-content-altering transformations such as object insertion, removal, significant cropping, style transfer, or compositional remixing. The filing further describes the UID as encoding a position in a continuous variance space, such that cosine similarity between two UIDs is directly computable without decoding a fixed binary digest. It also describes per-quadrant comparison producing localized scores, so that a derivative modifying one spatial region exhibits quadrant similarity near 1.0 for unchanged regions and reduced scores for the modified region.

Around that identifier the filing discloses a governance layer: classification of each UID into one of five variance bands by global variance value, with thresholds at 0.02, 0.06, 0.12, and 0.22; anchor nodes that declare governance responsibility for the UIDs falling in their bands; multi-root lineage graphs in which a derivative may be linked to more than one parent, with edge weights assigned proportional to cosine similarity; and alias registration at band-local, zone-local, or global scope under cryptographically signed policy objects validated by anchor quorum.

Where this leaves her still exposed

The filing states that lineage contribution weights are recorded as structural signals and do not constitute legal determinations of authorship or ownership. For her purposes that boundary matters: a weighted edge would give her something to bring to a licensing conversation, not a conclusion that ends one.

Several described behaviors are conditioned on declared bounds rather than absolute. Stability is stated for rescaling within a defined canonical size and for lossy compression at moderate quality levels, and lineage edges are established where slope proximity falls within a configured semantic continuity threshold. Whether a given file from her client's pipeline lands inside those bounds would depend on how her deployment declared them, and the filing describes the configuration rather than fixing an answer for her collection.

The optional signatures carry their own conditions. The constellation signature described in the filing calls for at least three detected saliency hotspots, and where fewer are detected the filing describes the signature as not emitted and the corresponding field in the UID record as null, which would be a real consideration for the flat studio backdrops and sparse documentation shots in her holdings.

The compensation routing embodiment would not reach her hardest cases as her relationships are structured today. The filing describes three pre-conditions for it: a registered UID in the anchor network, an alias registered with a compensation routing field specifying a payment address and compensation schedule, and a generative model operating within a governed execution environment that logs consultation events. Her agency counterparties run their own environments, and the filing does not describe a route that would place her inside them.

Finally, her non-image holdings, the oral history recordings and the typed accession ledgers, would reach the same pipeline only through the modality-specific normalization the filing describes, such as spectrogram representation for audio or token frequency distributions mapped onto a positional grid for text. That work would sit with whoever configured her deployment.

Disclosure Scope

This article describes subject matter disclosed in PCT/US26/28630, "Structural Content Identity and Rights-Grade Admissibility for Digital Artifacts." It is a technical description of embodiments appearing in that filing. Nothing in this article characterizes the scope of any claim, and nothing here is an admission regarding the state of the art. The scenario described is illustrative and does not refer to any actual party, organization, or product.

Content Anchoring (/content-anchoring) All 49 steps → (/inventive-steps)

Computable identity for media. Provenance from structural variance.

PCT/US26/28630 (/patents/pct-us26-28630)

PRIMARY TECHNICAL DISCLOSURE

- Content Anchoring: Computable Identity for Media That Changes (/articles/content-anchoring-computable-identity-for-media-that-changes)

SECONDARY TECHNICAL

- Multi-Axis Variance Vector Extraction: Nine Dimensions of Structural Content Identity (/articles/content-anchoring/variance-vector)
- Quadrant Decomposition: Spatial Sub-Region Fingerprinting for Partial Similarity Detection (/articles/content-anchoring/quadrant-decomposition)
- 320-Bit UID Construction: Multi-Segment Hashing for Negligible Collision Probability (/articles/content-anchoring/uid-construction)
- Structure Signature: Background-Invariant Matching Through Gradient-Only Descriptors (/articles/content-anchoring/structure-signature)
- Constellation Signature: Geometry-Invariant Matching Across Crop, Scale, and Occlusion (/articles/content-anchoring/constellation-signature)
- Five-Band Variance Classification: Content Routing by Structural Complexity (/articles/content-anchoring/variance-classification)
- Variance Saturation-Governed Cache Eviction: UID Density Replacing Static TTL (/articles/content-anchoring/cache-eviction)
- Multi-Root Composite Lineage Graphs: Provenance Through Variance Vector Similarity (/articles/content-anchoring/composite-lineage)
- Multi-Modal Content Identity: Unified Pipeline Across Image, Audio, Text, and Video (/articles/content-anchoring/multi-modal-identity)
- Rights-Grade Pre-Release Admissibility: Policy Evaluation Before Content Commitment (/articles/content-anchoring/pre-release-admissibility)
- Training Corpus Governance: Verifiable Lineage From Training Data to Model (/articles/content-anchoring/training-corpus-governance)
- Consultation Event Logging: Deterministic Records of Every Generation Reference (/articles/content-anchoring/consultation-logging)
- Model Output Provenance Fingerprint: Structural Proximity Without Model Access (/articles/content-anchoring/output-provenance)
- Creator Attribution and Compensation Routing: Payment From Consultation Lineage (/articles/content-anchoring/creator-attribution)
- Adversarial Robustness and Deepfake Detection: Content Identity as Detection Substrate (/articles/content-anchoring/adversarial-robustness)
- Client-Side Execution Architecture: Privacy-Preserving Variance Computation on Device (/articles/content-anchoring/client-side-execution)

- UID Resolution Query Protocol: Distributed Lookup Across Anchor Node Networks (/articles/content-anchoring/uid-resolution)
- Orientation Canonicalization: Rotation-Invariant Processing Through Gradient Normalization (/articles/content-anchoring/orientation-canonicalization)
- Cross-Band Resolution Pathfinding: Traversal Between Variance Bands Under Mutation (/articles/content-anchoring/cross-band-resolution)
- Identity by Position: Media as a Third Navigable Space (/articles/content-anchoring/identity-by-position)

APPLICATIONS · GENERAL

- Forbidden-Content Blocking at Upload and Generation Time: Pre-Release Exclusion Against Signed Policy (/articles/content-anchoring/forbidden-content-blocking)
- Structural Provenance for Software Supply Chains: Binary and Firmware Identity Independent of SBOM Metadata (/articles/content-anchoring/software-supply-chain-provenance)
- Rights-Grade Generative AI: How to Pay Creators, Exclude Forbidden Content, and Prevent Infringement Before Release (/articles/content-anchoring/rights-grade-generative-ai)
- Deepfake Detection by Structural Provenance: Verifying Synthetic Media Without Watermarks (/articles/content-anchoring/deepfake-provenance)
- Creator Economy Attribution Without Platform Intermediaries (/articles/content-anchoring/creator-attribution-economy)
- Verifying Source Photos and Video in the Newsroom: Content Anchoring for Journalism (/articles/content-anchoring/journalism-verification)
- Detecting Image Manipulation and Proving Figure Provenance in Research Publications (/articles/content-anchoring/academic-research-integrity)
- Content Anchoring for Legal Evidence Chains (/articles/content-anchoring/legal-evidence-chain)
- Content Anchoring for Insurance Claims Evidence (/articles/content-anchoring/insurance-claims-evidence)
- Content Anchoring for Real Estate Documentation (/articles/content-anchoring/real-estate-documentation)
- Art Authentication and Provenance Verification with Content Anchoring (/articles/content-anchoring/art-authentication)
- Detecting Screenshot and Recapture Fraud in Identity-Document KYC With Structural Content Identity (/articles/content-anchoring/identity-document-kyc-recapture)
- **When the Same Work Returns Wearing a New Hash (/articles/content-anchoring/content-anchoring-stakes)**

APPLICATIONS · SPECIFIC

- C2PA vs Content Anchoring: Attached Provenance or Content-Intrinsic Identity? (/articles/content-anchoring/c2pa)
- Google SynthID Alternative: Content-Intrinsic Identity Beyond Watermarking (/articles/content-anchoring/google-synthid)
- Beyond Shutterstock: Content-Intrinsic Identity That Survives Re-Encoding and Cropping (/articles/content-anchoring/shutterstock)
- Spotify Alternative for Music Provenance: Structural Content Identity Beyond the ISRC Database (/articles/content-anchoring/spotify)
- Getty Images Alternative for Provenance: Structural Content Identity Beyond Metadata (/articles/content-anchoring/getty-images)
- Adobe Stock vs Structural Content Identity: Licensing Records Are Not Content Identity (/articles/content-anchoring/adobe-stock)
- YouTube Content ID vs Content Anchoring: Matching Against a Database, or Identity in the Content Itself (/articles/content-anchoring/youtube-content-id)
- Audible Magic Alternative: Structural Content Identity Beyond Database-Matched Fingerprinting (/articles/content-anchoring/audible-magic)
- Digimarc vs Structural Content Identity: Watermarks Are Added, Not Intrinsic (/articles/content-anchoring/digimarc)
- Irdeto vs Structural Content Identity: DRM Protects the Channel, Not the Payload (/articles/content-anchoring/irdeto)
- Truepic alternative: capture-time provenance versus structural identity derived from the artifact itself (/articles/content-anchoring/truepic)
- Microsoft PhotoDNA vs structural content identity: hash-matching known images versus screening artifacts before release (/articles/content-anchoring/microsoft-photodna)
- Pex alternative: structural content identity vs enrolled fingerprint matching (/articles/content-anchoring/pex)

HOW-TO GUIDES

- How to Attribute AI-Generated Content to Its Source for Licensing (/articles/guides/attribute-ai-generated-content-for-licensing)
- How to Detect a Screenshotted or Re-Photographed Document (/articles/guides/detect-recaptured-document)
- How to Detect AI-Generated Images Without a Watermark (/articles/guides/detect-ai-images-without-watermark)
- How to Verify Content Authenticity Without a Central Registry (/articles/guides/verify-content-authenticity-without-registry)

TERMINOLOGY

- constellation signature (/glossary#constellation-signature)
- fork adjudication (/glossary#fork-adjudication)
- governed corpus (/glossary#governed-corpus)
- memorization proximity score (/glossary#memorization-proximity-score)
- slope band (/glossary#slope-band)

Content Anchoring overview → (/content-anchoring)