The Science Behind Video Forensics

A plain-language guide to digital video tampering detection, the forensic techniques that expose it, and the math that ties it together.

quevidkit

Open-Source Forensic Video Analysis Toolkit

Table of Contents

1. How Digital Video Actually Works
2. How People Tamper with Video
3. How Forensic Detection Works
4. What Each quevidkit Check Does and Why
5. When the Detector Gets It Wrong
6. How Weak Signals Become Strong Evidence
7. Chain of Custody and Legal Use
8. The State of the Field
9. Further Reading

Important: No automated tool can provide legal certainty. quevidkit produces **evidence-backed probability with explanation**, not proof. The tool identifies anomalies; a qualified human interprets what those anomalies mean in context.

1. How Digital Video Actually Works

The Two Layers: Container and Codec

Every video file has two distinct layers, like a shipping box and the item packed inside.

The **container** is the file format you see — .mp4, .mkv, .mov. It holds together the video stream, audio stream, metadata (timestamps, titles, GPS), and sometimes subtitles or chapters. The container does not determine quality — it is packaging.

The **codec** (coder-decoder) is the algorithm that compresses and decompresses the actual picture and sound data. Without compression, one minute of uncompressed 1080p video at 30 fps would be roughly **10 GB**. Codecs make this manageable by discarding information human eyes are unlikely to notice.

Codec	Notes
H.264 (AVC)	Dominant for 15+ years. Breaks frames into 16×16 pixel macroblocks with motion prediction. Universally supported.
H.265 (HEVC)	Successor to H.264. Flexible Coding Tree Units up to 64×64. Achieves 25–50% better compression at equal quality.
VP9 / AV1	Open-source alternatives. AV1 is increasingly popular for web streaming.

Frame Types: I, P, and B

Not every frame stores a complete picture. To save space, codecs use three types:

I-frames (Intra) — Complete, standalone images. Like a photograph. The largest frames because they contain everything needed to display themselves. They serve as anchor points in the video.

P-frames (Predicted) — Store only what *changed* since the previous frame. If a person walks across a static background, only the person's new position is encoded. Much smaller than I-frames.

B-frames (Bi-directional) — The most compressed. They reference *both* past and future frames, storing only the differences from both directions. Smallest frames, most processing to decode.

GOP: Group of Pictures

Frames are organized into **GOPs** — Groups of Pictures. Each GOP starts with an I-frame and is followed by P and B frames until the next I-frame:

```
I  B  B  P  B  B  P  B  B  P  B  B  I  B  B  P  ...
■■■■■■■■■■ one GOP ■■■■■■■■■■
```

The **GOP size** is the number of frames between I-frames (commonly 30, 60, 120, or 250). Longer GOPs mean better compression but less resilience.

> **Forensic insight:** A legitimate, singly-encoded video has a **consistent, regular GOP pattern**. When someone edits or splices video, the GOP pattern almost always becomes irregular.

How Compression Works

Video compression exploits two kinds of redundancy:

Spatial redundancy (within one frame): Large areas of similar color can be described efficiently. The frame is divided into blocks, and the **Discrete Cosine Transform (DCT)** converts pixel values into frequency coefficients. High-frequency detail is aggressively reduced; low-frequency information is preserved. This is why heavily compressed video looks blocky.

Temporal redundancy (between frames): Most of frame 100 is identical to frame 99. **Motion estimation** finds blocks that moved and encodes only the **motion vectors** (direction + distance) plus any residual difference.

The **quantization parameter (QP)** controls how aggressively detail is discarded. Higher QP = smaller file, more artifacts. This is forensically critical: **re-encoding changes the QP, and double compression leaves detectable mathematical traces**.

2. How People Tamper with Video

Type	What It Is	What It Disrupts
Splicing	Combining footage from different sources into one apparently continuous video.	Metadata, compression characteristics, noise patterns, lighting, audio, and GOP structure at edit points.
Frame Deletion	Removing frames to make events happen faster or erase critical moments.	Timestamps jump, motion appears unnaturally fast, GOP structure breaks.
Frame Insertion / Duplication	Adding frames to make events appear longer or insert content.	Duplicate frames create zero-difference sequences. Inserted frames have different noise and compression.
Re-encoding	Decoding and re-encoding with different settings, often to smooth over edits.	Double compression artifacts, changed metadata, predictable quality degradation patterns.
Deepfakes	AI-generated face swaps or synthetic video from neural networks.	Subtle blinking, lighting, skin texture, and frequency-domain artifacts.
Audio Replacement	Replacing original audio while keeping the video.	Lip sync mismatch, spectral discontinuities, background noise profile changes, metadata mismatches.
Speed Manipulation	Altering playback speed of portions.	Inconsistent motion vectors, audio pitch shifts, timestamp irregularities.

3. How Forensic Detection Works

Each detection technique targets a different physical or mathematical property of video. Here is what they detect and *why* the technique works.

Metadata Analysis

Targets: *Inconsistencies between what the file claims and what it actually is.*

When a camera records, it writes consistent metadata. Editing software writes different metadata, creating mismatches: duration discrepancies, bitrate calculation errors, audio/video length gaps, and software encoder tags (Adobe, HandBrake, FFmpeg).

Compression Artifacts

Targets: *Evidence of double compression and spatial quality inconsistencies.*

When re-encoded, the interaction between old and new quantization grids creates periodic patterns. P-frame sizes show peaks at the original GOP interval. I-frame sizes become unusually variable because some land on former I-frame positions (high detail) and others on former P-frame positions.

Error Level Analysis

Targets: *Regions or segments with different compression histories.*

Re-compress at a fixed quality and measure the residual. Uniformly compressed content shows uniform error levels. Tampered regions from a different source show different residuals — like detecting which pages in a book were reprinted by photocopying all of them.

Temporal Analysis

Targets: *Disruptions in the continuous flow of time.*

Legitimate video has smoothly incrementing DTS/PTS timestamps. Frame deletion creates gaps. Frame insertion creates clusters. Non-monotonic timestamps indicate timeline disruption at splice points.

GOP Structure Analysis

Targets: *Irregular keyframe patterns.*

Encoders place keyframes at fixed intervals. Editing disrupts this regularity. The coefficient of variation of GOP intervals reveals inconsistency.

Scene + GOP Correlation

Targets: *Unnatural scene transition placement relative to keyframes.*

Encoders insert keyframes at scene changes. In spliced video, scene changes appear at non-keyframe positions — like chapter headings appearing mid-page instead of at page tops.

Noise Pattern Analysis

Targets: *Changes in video source (different camera/encoder).*

Every sensor has a unique noise fingerprint. Splicing from a different source shifts the noise profile. Measured via Laplacian std-dev and Sobel high-frequency energy.

Audio Spectral Analysis

Targets: *Audio splices and source changes.*

FFT decomposes sound into frequencies. Abrupt shifts in spectral centroid, RMS energy, or zero-crossing rate indicate environment changes. Short silence gaps suggest splice joints.

Perceptual Hashing

Targets: *Duplicate or near-duplicate frames.*

Perceptual hashes produce similar fingerprints for similar images. Hamming distance between consecutive frames detects freeze-frames, duplication, and copy-move regions.

Statistical Methods

Targets: *Deviations from expected mathematical distributions.*

Z-scores (MAD-based) flag outliers. Autocorrelation detects ghost GOP periodicity. Coefficient of variation measures regularity. Distribution comparison across temporal windows detects segment-level encoding changes.

4. What Each quevidkit Check Does

Standard Checks (all presets)

1. Metadata & Codec Consistency

Category: metadata | Function: metadata_codec_checks()

Compares container vs. stream duration, declared vs. observed bitrate, audio vs. video duration. Scans encoder tags for editing software markers.

Science: Genuine recordings have internally consistent metadata. Editing disrupts this.

■ *Like checking if the label on a package matches what's inside.*

2. Packet Timing Anomalies

Category: timing | Function: packet_timing_checks()

Examines DTS/PTS monotonicity and timestamp delta distribution. Flags non-monotonic DTS and spikes >3.5× or <0.25× the median.

Science: Legitimate video has smoothly incrementing timestamps. Splicing disrupts the timeline.

■ *Like finding a clock that skips from 2:00 to 2:05 — five minutes went missing.*

3. Frame Structure Anomalies

Category: codec | Function: frame_structure_checks()

Measures GOP interval coefficient of variation, detects resolution changes, counts color profile variants.

Science: Single-session recordings have regular GOP patterns. Editing disrupts regularity.

■ *Like finding a book where some chapters have 30 pages and others have 7 — inconsistent structure.*

4. Frame Quality Shift

Category: quality | Function: opencv_frame_quality_checks()

Computes dHash perceptual hashes, Laplacian blur, 8×8 blockiness, and pixel differences. Detects duplicates, gaps, and quality shifts via z-score.

Science: Quality should not change abruptly in a genuine recording. Duplicates are statistically improbable.

■ *Like noticing one page in a document is printed in higher resolution than the rest.*

Advanced Checks (deep preset)

5. Compression Consistency

Category: codec | Function: compression_consistency_checks()

Splits video into 6 temporal windows. Computes median packet size per frame type (I/P/B) in each window. Flags windows where the median shifts >40% from the global median.

Science: Single-pass encoding produces stationary distributions. Re-encoding a portion shifts the distribution because the new encoder's quantization decisions differ.

> ■ *Like checking if paragraphs in a document were printed by different printers — each produces slightly different ink density.*

6. Scene Cut Forensics

Category: timing | Function: scene_cut_forensics_checks()

Runs ffmpeg scene detection, checks alignment with keyframes (tolerance: 15% of GOP + 1 frame). Detects suspicious clustering (<150ms apart).

Science: Encoders place keyframes at scene changes. Spliced content shows scene changes at non-keyframe positions.

> ■ *Like checking chapter breaks in a book — a natural publisher puts chapter headings at the top of new pages; a cut-and-paste job has them mid-page.*

7. Audio Spectral Continuity

Category: audio | Function: audio_spectral_checks()

Extracts audio as 16kHz mono PCM. Computes RMS energy, spectral centroid, and ZCR over 100ms windows. Z-score analysis (threshold >5) detects discontinuities. Flags short silence gaps.

Science: Genuine audio has continuous spectral characteristics. Splicing creates abrupt frequency-domain breaks.

> ■ *Like noticing someone changed radio stations mid-sentence — even if both play music, the color of the sound is different.*

8. Temporal Noise Consistency

Category: quality | Function: temporal_noise_consistency_checks()

Samples frames at ~2fps. Computes Laplacian std-dev (noise floor) and Sobel magnitude (high-frequency energy). Detects shifts via z-score (>4.5). Compares half-distributions.

Science: Every camera has a unique noise fingerprint. Splicing from a different source changes the profile.

> ■ *Like hearing a change in background hum when two phone recordings from different rooms are stitched.*

9. Double Compression Detection

Category: codec | Function: double_compression_detection()

Computes autocorrelation of P-frame sizes. Searches for periodic peaks not aligned with the current GOP. Checks I-frame size CV (>0.45 is suspicious).

Science: Re-encoding leaves a ghost of the original GOP cadence — a periodic fingerprint detectable in the autocorrelation function.

> ■ *Like seeing faint ruled-paper lines through a photocopy — the old grid is still visible under the new one.*

10. Error Level Analysis (ELA)

Category: quality | Function: ela_frame_analysis()

Re-compresses frames at JPEG quality 75 and measures the residual. Detects ELA shifts via z-score (>4) and high temporal variance (CV >0.35).

Science: Uniformly compressed content produces uniform residuals. Mixed-compression produces inconsistencies.

> ■ *Like testing if all pages of a book were printed at the same time — photocopy every page and compare. A replacement page looks different under the same copying process.*

11. Bitstream Structure

Category: codec | Function: bitstream_structure_checks()

Inspects mid-stream color parameter changes, interlaced/progressive switches, frame-size outlier rates (>5%), and B-frame declaration consistency.

Science: A single encoding session produces constant codec parameters. Mid-stream changes indicate content assembled from different sessions.

> ■ *Like checking if all bricks in a wall are the same type — half red clay, half gray concrete means the wall was built in two phases.*

5. When the Detector Gets It Wrong

No single forensic check is definitive. Many completely legitimate processes produce artifacts that look like tampering. Understanding these is critical for responsible forensic analysis.

Source	What Happens	Impact
Transcoding	Format conversion (MOV→MP4) re-encodes the entire video. Changes encoder metadata, introduces double compression, may alter GOP.	This is the #1 source of false positives.
Social Media Upload	Platforms re-encode to their own settings. Downloaded videos have double compression, platform metadata, and possible cropping.	Indistinguishable from tampering at codec level.
Screen Recording	Creates an entirely new recording with new metadata, different frame rates, and different compression.	Cannot verify original integrity from a screen recording.
Variable Bitrate	VBR allocates more bits to complex scenes. Natural frame-size variation can resemble compression inconsistency.	Common in all modern video.
Adaptive GOP	Some encoders insert keyframes at scene changes rather than fixed intervals, producing irregular patterns.	Looks similar to editing artifacts.
Multi-Camera Editing	Professional productions cut between cameras with different color profiles and noise characteristics.	Triggers scene, noise, and bitstream checks.
Audio Processing	Legitimate noise reduction and normalization change spectral characteristics.	Can trigger audio spectral checks.
CCTV Concatenation	Security systems create files at intervals and concatenate them, introducing genuine discontinuities.	Timestamp resets at each junction.

*A forensic tool finding "suspicious" means: "this video has characteristics inconsistent with a single unmodified recording." It does **not** mean: "this video was maliciously tampered with." Expert interpretation is always required.*

6. How Weak Signals Become Strong Evidence

No single check is reliable enough alone. The power comes from **combining multiple independent signals**. If metadata AND timing AND compression AND noise AND audio ALL show anomalies at the same timestamp, the probability of coincidental false positives drops dramatically.

The Fusion Pipeline

```
Individual checks → Weighted combination → Logistic function → Verdict
```

	Description
Step 1	Each check produces a **score** (anomaly strength, 0–1) and a **confidence** (measurement reliability, 0–1).
Step 2	Scores are combined via **confidence-weighted mean**. Checks with more data have more influence.
Step 3	A **quality gate** reduces trust when quality checks show extreme anomalies (may indicate low quality, not tampering).
Step 4	The combined signal passes through a **logistic (sigmoid) function** that maps it onto 0–1 probability. The sensitivity parameter shifts the decision boundary.
Step 5	**Overall confidence** is estimated from coverage (how many checks ran: 70%), quality gate (20%), and agreement (10%).

Decision Thresholds

Condition	Label
Quality gate < 0.3 or confidence < 0.35	inconclusive
Probability ≥ 0.6	tampered
Probability ≥ 0.35	suspicious
Otherwise	authentic

The Bayesian Intuition

Each piece of evidence updates our belief about whether tampering occurred. We start with a prior (the bias term), and each check shifts the conclusion. Multiple weak signals pointing the same direction compound into strong evidence — the mathematical equivalent of the “wisdom of crowds.” The key is **independence**: the checks must measure genuinely different things. quevidkit achieves this by spanning metadata, timing, codec structure, visual quality, audio, and noise domains.

7. Chain of Custody and Legal Use

Chain of custody is the documented history of who handled evidence, what they did with it, when, and why. It creates an unbroken trail from collection to courtroom.

Step	What It Means
Acquisition	Document how the video was obtained, date/time, who performed it, and tools used.
Hashing	Compute SHA-256 immediately upon acquisition. This fingerprint proves the file has not changed. quevidkit does this automatically.
Secure Storage	Maintain evidence in write-protected, access-controlled storage with logging.
Analysis Docs	Record every tool, action, and result. quevidkit generates comprehensive JSON and HTML reports.
Transfer Docs	Log every handoff between individuals or systems.

Legal Admissibility

Authenticity — Proof the evidence is what it claims to be (hash verification).

Integrity — Proof it has not been altered since collection.

Reliability — Proof the forensic tools are scientifically sound. Open-source tools with readable code support this.

Reproducibility — Another examiner using the same tools should reach the same findings.

> **"If you did not write it down, it did not happen."**
> The guiding principle of digital forensics. Every action, every result, every handoff must be documented.

8. The State of the Field

The AI Arms Race

Video forensics is in an arms race. Generation quality is improving faster than detection. A landmark CSIRO study assessed 16 leading deepfake detectors and found **none performed reliably** across a broad range of manipulation techniques. Anti-forensics is real — adversaries specifically design manipulations to evade detection.

Why Traditional Forensics Still Matters

Property	Why It Matters
Interpretable	A court can understand “the GOP structure changed at 45.2s” in a way it cannot understand “the neural network output 0.87.”
Reproducible	Same input + same algorithm = identical result, every time. No non-deterministic behavior.
Catches what AI misses	Frame deletion, splicing, re-encoding are often invisible to deepfake detectors but clear to codec analysis.
No training data needed	The physics and mathematics *are* the training. Double compression detection works because of quantization math, not because a model was shown examples.

The Fundamental Truth

The most robust forensic analysis combines multiple independent detection techniques, expert interpretation of results, consideration of alternative explanations, proper chain of custody documentation, and reproducible methodology.

> No automated tool can provide legal certainty. quevidkit produces **evidence-backed probability with explanation**, not proof. The tool identifies anomalies; a qualified human interprets what they mean in context.

9. Further Reading

Academic Papers & Standards

- An Overview of Video Tampering Detection Techniques — IEEE, 2023
- Systematic Analysis of Video Tampering and Detection Techniques — Cogent Engineering, 2024
- Temporal Tampering Detection in Automotive Dashcam Videos — PMC, 2026
- Deepfake Media Forensics: Status and Future Challenges — PMC, 2025
- Double Compression Detection for H.264 with Adaptive GOP — Springer, 2019
- SWGDE Best Practices for Digital Forensic Video Analysis — SWGDE.org

Accessible Explanations

- How to Check Video Integrity by Detecting Double Encoding — Forensic Focus
- Error Level Analysis Tutorial — FotoForensics.com
- Screen Capture: It’s Not the Evidence — Amped Software Blog
- How to Maintain Chain of Custody for Digital Evidence — AMU/APUS
- How to Make Digital Evidence Admissible in Court — TrueScreen, 2026

Tools

- **quevidkit** — Open-source forensic video analysis (this project)
- **FFmpeg / FFprobe** — The foundational video processing toolkit quevidkit builds on
- **Amped FIVE** — Commercial forensic video analysis suite
- **MediaInfo** — Open-source metadata inspection tool



quevidkit

Open-Source Forensic Video Analysis

github.com/thumpersecure/quevidkit