com sur™

the missing piece of CCTV

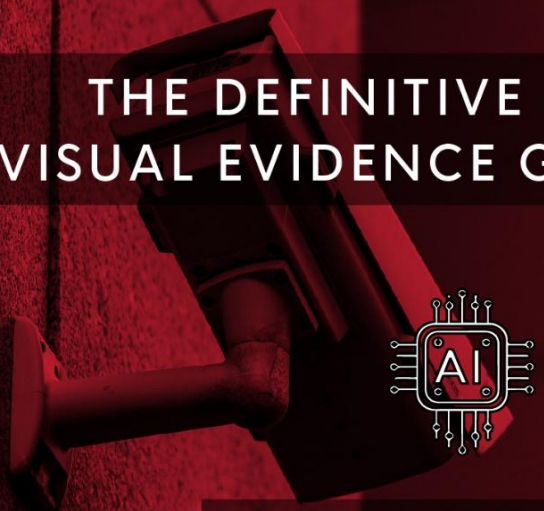
GET THE BOOK
INTERNATIONAL PRICING

"SEE WHAT THE CAMERA SAW"

IN 95% LESS TIME

KNOW WHAT YOU DON'T KNOW
TURN GARBAGE TO GOLD

THE DEFINITIVE GUIDE TO VISUAL EVIDENCE GOVERNANCE



BUILD 'BETTER' AI
POWERED BY DAILY AUDITS

100+ SECTORS - AIRPORTS TO ZOOS

GAUTAM D. GORADIA

THE EIGHTH APPROACH

Abstract

Surveillance cameras now generate more footage than any organisation can watch. A family of technologies, including video synopsis, forensic search, vision language summarization and time lapse compression, has grown up to reduce what a human being must sit through. Each of these technologies filters, detects, or interprets the video before a person ever sees it.

This comparison deliberately reaches beneath detection to the recording layer itself. Two of the seven approaches examined here, continuous DVR, NVR or VMS recording and motion-triggered recording, are not detection or summarization technologies at all. They are the baseline infrastructure that every detection and summarization technology was built to work around: one gives complete coverage with no reduction, the other decides what to keep before anyone looks, and both create the exact problem the other five technologies exist to solve.

COM-SUR does none of this. It does not detect, classify, filter or interpret anything. It samples the complete rendered image at a fixed interval, without judging what matters, and leaves that judgement to a human. This paper proposes a taxonomy of the principal approaches used to reduce the burden of reviewing recorded video and distinguishes review technologies from Visual Evidence Governance. Visual Evidence Governance is treated in full in a companion paper; here it is discussed only to the extent necessary to position COM-SUR correctly in relation to the proposed taxonomy.

1. Cameras See More Than Crime

Video surveillance is still commonly framed around crime, alarms and post-incident investigation. That framing is too narrow. Cameras also see unsafe practices, process failures, poor housekeeping, equipment misuse, compliance lapses, bottlenecks, near misses and

training gaps, most of which never trigger an alert or match a predefined search criterion.
A camera does not prevent anything by itself. Prevention becomes possible only when what the camera records is reviewed soon enough, interpreted correctly, and converted into action. That requires an organisation to be able to answer a short list of questions:

- Was the visual data reviewed systematically, or only after an incident?
- Who reviewed it, and what was found?
- What corrective or preventive action followed?
- Which visual evidence was preserved, and can it be retrieved?
- Can the organisation demonstrate that a defined process was followed?

2. From Post-Incident Review to Continuous Improvement

Post-incident investigation asks what happened after a failure has already occurred. Continuous visual auditing asks a broader question: what is going wrong, changing or being overlooked before the next failure? The reviewer does not begin with a known person, object or violation in mind. The task is to examine the record and discover what would otherwise stay unnoticed.

Observation → Discovery → Corrective Action → Preventive Action → Improvement

This is Kaizen applied to visual evidence.
A camera should not merely explain a failure after the fact. It should help an organisation find weaknesses and correct them early. Throughout this cycle, human judgement stays central: technology can reduce, order or describe visual material, but a person still has to interpret context, decide what requires action, and confirm the action worked.

3. Why Millions of Frames Is the Real Problem

A single camera recording at 25 frames per second produces roughly 90,000 frames an hour and 2.16 million frames a day. A modest site of 16 cameras arranged in a 4x4 grid produces roughly 34.5 million frames a day. These figures are illustrative rather than exact; actual volume depends on frame rate, resolution and camera count, and they are given here once because they explain every technology described in the next section.

The problem is not simply storage. It is human attention. Every technology examined below exists to reduce that same volume of frames down to something a person can realistically review, and each one does it in a different way.

4. A Proposed Taxonomy of Video Review Technologies: Seven Established Approaches, and Where COM-SUR Diverges

Please refer to the diagram on page no. 9 for the proposed taxonomy.

The first two entries below are not usually discussed as summarization technologies, but they define the baseline problem that every technology after them, including COM-SUR, exists to solve.

i. Continuous DVR, NVR or VMS recording**:** keeps the full, unfiltered stream. It gives complete coverage but no reduction: reviewing a day of sixteen cameras still means scrubbing through many hours of playback, and multi-camera decoding strains most systems badly enough that daily review is rarely attempted in practice. This constraint, not a lack of information in the footage, is why every summarization technology below was built.

ii. Motion-triggered recording: decides what to keep at the point of capture. Anything below the motion threshold, or anything that involves no motion at all, is never recorded: a spark too small to trigger the algorithm, a crack slowly widening on a wall, an insect near food. None of this is a tuning problem that a lower threshold would fix, because the system is built to answer whether something moved, not what the scene currently looks like.

iii. Object-based video synopsis: such as BriefCam analyses recorded video, detects people and vehicles, and displays activity from different times together against a common background. The result is a constructed summary built for fast investigation, not a recording of any single moment.

iv. Metadata-based forensic search: used by platforms such as Avigilon, Genetec and Axis, converts video into searchable attributes: object type, colour, direction, time in scene, licence plate. A query is matched against this index. It is built for retrieving footage an investigator can already describe, not for surfacing what nobody thought to search for.

v. Vision-language model summarization: as used in Milestone XProtect, analyses a clip and generates a natural-language description of the scene. Its output is an interpretation in words, useful for incident logs and rapid orientation, but an interpretation nonetheless.

vi. Event-based clip extraction and alerting: surfaces only the clips tied to a configured rule, alarm or detected activity. Anything that falls outside that rule is never shown to a reviewer at all.

vii. Construction time-lapse: platforms such as EarthCam and OpenSpace capture images at intervals measured in minutes and compile them into an accelerated sequence. Their purpose is to show gradual change over months, not to support second-by-second auditing.

Within the proposed taxonomy, six of these approaches decide, at some point before a human looks at the footage, what is worth keeping, detecting, extracting, describing or presenting. Continuous recording is the exception: it preserves the complete stream without reducing review burden. COM-SUR makes no such decision. It samples the complete image displayed on screen at a fixed interval, commonly one image per second, regardless of camera brand, VMS or recording format. It does not classify a person, recognise a vehicle, detect motion or judge whether a moment matters. It simply preserves what a human would have seen, had they been watching at that instant, and leaves every judgement about significance to the person reviewing it afterward.

In one line: *every other approach answers what to look at by filtering, detecting or describing.*

COM-SUR answers it by preserving the complete rendered scene at deterministic intervals and handing the decision to a human reviewer.

5. What Interval Sampling Does Not Catch, and Why It Still Wins

Sampling the fully decoded, rendered visual state at a fixed interval has an honest limit. An event whose entire duration fits inside the gap between two captures, and leaves no trace at the sample before or after it, will not appear in the record. A fast hand-off, a flash of movement, a brief gesture. This is a genuine property of interval sampling, not an implementation defect, and it should be stated plainly rather than argued away.

Two things narrow that gap without closing it. First, the original source video remains fully intact and available, so any question raised by the sampled record can be resolved by going back to the original footage at that exact timestamp. Second, a reviewer can raise the capture density on a specific viewport, up to five captures a second, wherever a location is judged critical enough to warrant it. Both measures reduce risk. Neither is a claim of completeness, and running every camera at that density would recreate the multi-camera decoding load that this approach exists to avoid in the first place.

The residual gap matters less in practice than it sounds in theory. Most of what an organisation needs to catch, a safety violation, a process failure, a person or vehicle present at a location, unfolds over seconds, not fractions of a second. At a one-second interval, the samples immediately before and after a missed instant will very likely still show the person, vehicle or condition in question. This is a probabilistic argument, not a completeness claim, and it is presented here as one.

The more consequential limitation belongs to every technology built around detecting or searching for something specific, not to interval sampling. Object-based synopsis and forensic search only surface what somebody told them to look for. If an investigator flags a man in a blue jacket, both technologies find him efficiently. Neither has any mechanism for noticing that his supervisor, wearing a different colour each day, had been present at the same location on each of the three preceding days, because nobody wrote a query for a recurring person in changing clothing. That pattern is invisible to any technology that filters before a human looks at the footage. It survives review only where nothing was filtered out to begin with.

Interval sampling can miss an instant. Search and synopsis can miss a pattern nobody thought to ask about. The second failure is the more expensive one, and it is the one Visual Evidence Governance is built to catch.

6. From Review Technology to Visual Evidence Governance

The taxonomy proposed in this paper distinguishes review technologies from the broader discipline of Visual Evidence Governance (VEG). Introduced by COM-SUR, VEG addresses the systematic auditing, reporting and preservation of visual evidence. VEG does not compete with cameras, recorders, VMS platforms, analytics or any of the seven approaches above; it is the framework within which those technologies deliver accountable, measurable outcomes.

A companion paper, *Visual Evidence Governance in the Realm of Video Surveillance* [6], sets out that framework in full, including its governing principles and the Daily CCTV Video Footage Auditing Chart. What matters here is narrower: where a camera or an analytics engine operates on frames, clips and alarms, governance operates on the observation, something a human has reviewed, classified and acted on, and COM-SUR is built specifically to feed that observation-level record rather than to compete with any of the review technologies described in the proposed taxonomy.

India's National Occupational Standard and National Qualifications Register already recognise footage auditing as a distinct function through the CCTV Video Footage Auditor qualification, independent support for treating this as a discipline in its own right rather than a by-product of any single technology.

7. How COM-SUR Puts Governance Into Practice

Patented COM-SUR is the result of more than a decade of work aimed at making systematic visual auditing practical at scale. It works across live and recorded sources, including CCTV, drones, body-worn cameras, dashcams and mobile video, without depending on a particular manufacturer, VMS or AI platform.

7.1 Multi-Camera Review in Minutes, Not Hours

Where a 16-camera site generates roughly 34.5 million frames a day, as set out in Section 3, COM-SUR reduces that same volume to 86,400 composite review points, one per second across a full day, while the original video remains available for detailed examination. This is a comparison of review units rather than a storage ratio; actual storage still depends on compression, resolution and retention policy.

7.2 A Media Player Built for Daily Auditing

COM-SUR's media player allows a reviewer to move quickly through recorded material, examine synchronised multi-camera grids, pan and zoom, and compare visual conditions without dragging through long conventional timelines. Detailed examination is supported by brightness, contrast, black-and-white and false-colour views. These are analytical tools applied to the source image; they create no new evidence, and the original recording always remains available alongside any enhanced view.

7.3 Turning Observations Into Institutional Memory

Findings are tagged and logged in the Daily CCTV Video Footage Auditing Chart described in the companion governance paper, turning each review into a structured record rather than a note left behind in raw footage.

7.4 One Platform, Many Industries

Because the underlying method makes no assumption about what a scene contains, the same platform supports multiple sectors, from airports to zoos. In highly regulated settings such as healthcare, COM-SUR functions as a review and documentation aid, and it does not replace certified diagnostic systems or clinical judgement.

Viewed functionally rather than technically, COM-SUR serves as a visual productivity platform for video and imagery. In much the same way

that Microsoft Office provides a common environment for creating, reviewing, organising and communicating documents, spreadsheets and presentations across diverse industries, COM-SUR provides a common environment for reviewing, comparing, documenting, organising, preserving and reporting on visual data. It is not intended to replace cameras, VMS platforms, analytics or artificial intelligence. Rather, it complements them by making visual data more usable, actionable and accountable across a wide range of operational and investigative workflows.

8. Better AI Needs Governed Human Feedback

A generic AI model rarely reflects one site's exact conditions: camera angle, lighting, local behaviour and risk patterns all differ from its training data. Feeding it real, site-specific, human-validated observations closes that gap, and building that feedback loop is exactly what VEG is for, covered in depth in the companion governance paper. COM-SUR's own role stays narrow: it documents the observation and does not train, retrain or interact with third-party AI models itself, leaving model development to the AI vendor and the organisation's own governance process.

9. Why This Is Everyone's Requirement

Every camera and imaging system produces visual data that may become evidence, operational intelligence or institutional knowledge. Video synopsis, forensic search and vision-language summarization earn their place where an organisation has a defined need for that specific capability and the budget to deploy it. COM-SUR addresses the requirement common to every organisation that expects an accountable outcome from its cameras, regardless of sector, camera brand or dependence on artificial intelligence.

10. Conclusion

Cameras record far more than crime and incidents. They capture the everyday conditions that shape safety, compliance, efficiency and future risk, and most of that value is lost when footage is examined only after something has already gone wrong.

The proposed taxonomy presented in this paper identifies seven established approaches to reducing review burden. These approaches compress, retrieve, extract or describe visual material, and each is genuinely useful within its intended purpose. None of them removes the need for human verification and accountable action, and none of them is Visual Evidence Governance.

COM-SUR is not object-based synopsis, not forensic search, not vision-language summarization, not event-based extraction, not time-lapse, and not a substitute for continuous recording or motion-triggered alerting. It is the deterministic, unfiltered layer that makes daily, multi-camera visual auditing practical for the first time, and Visual Evidence Governance is the discipline that makes that practice repeatable and accountable.

The question that matters is not which technology is superior. It is whether an organisation governs what its cameras and imaging systems actually produce.

NO GOVERNANCE = NO VALUE

References

1. A. Rav-Acha, Y. Pritch and S. Peleg, “Making a Long Video Short: Dynamic Video Synopsis,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 435-441, 2006.

2. E. Apostolidis, E. Adamantidou, A. I. Metsai, V. Mezaris and I. Patras, “Video Summarization Using Deep Neural Networks: A Survey,” Proceedings of the IEEE, vol. 109, no. 11, pp. 1838-1863, 2021. DOI: 10.1109/JPROC.2021.3117472.

3. National Skill Development Corporation, “CCTV Video Footage Auditor,” Qualification Code MEP/Q7204, NSQF Level 4.

4.National Qualifications Register, “CCTV Video Footage Auditor,” NCVET-NQR Code 2022/SEC/MEPSC/06149.

5. E. Tabassi, “Artificial Intelligence Risk Management Framework (AI RMF 1.0),” NIST AI 100-1, National Institute of Standards and Technology, 2023. DOI: 10.6028/NIST.AI.100-1.

6. G. D. Goradia, “Visual Evidence Governance in the Realm of Video Surveillance,” COM-SUR Video Surveillance and Imaging AI Technologies Pvt. Ltd., Mumbai. Available at www.comsur.biz

DIAGRAM OF THE PROPOSED TAXONOMY

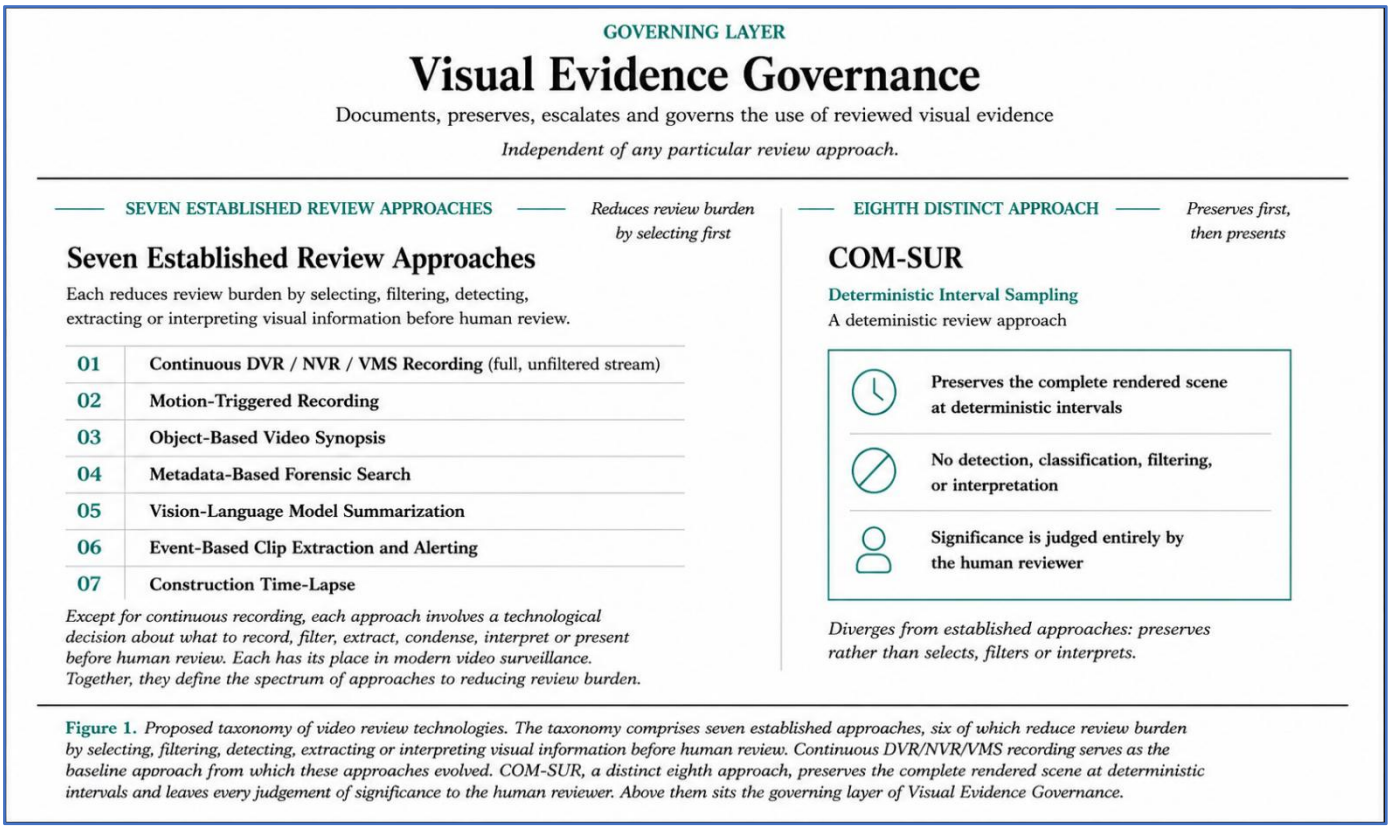


Figure 1. *Proposed taxonomy of video review technologies. The taxonomy comprises seven established approaches, six of which reduce review burden by selecting, filtering, detecting, extracting or interpreting visual information before human review. Continuous DVR/NVR/VMS recording serves as the baseline approach from which these approaches evolved. COM-SUR, a distinct eighth approach, preserves the complete rendered scene at deterministic intervals and leaves every judgement of significance to the human reviewer. Above them sits the governing layer of Visual Evidence Governance.*