com sur™

the missing piece of CCTV

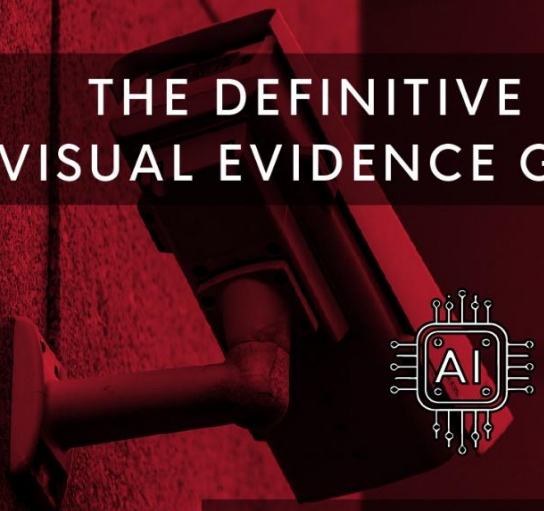
GET THE BOOK
INTERNATIONAL PRICING

"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

'PLAYBACK'
IS THE VILLAIN.
WHO IS THE HERO?

Life itself is full of playback. We replay memories, revise what we have studied, replay music, movies, photographs, conversations, our own thoughts, and even as you think about this paper, you are engaging in 'playback'. Most of this feels effortless. But video surveillance is different. The moment we try to replay and meaningfully review what multiple cameras recorded over hours or days, playback becomes cumbersome, computationally demanding and difficult for humans to consume.

That is where the villain enters the story.

For decades, the video surveillance industry has improved almost everything. Cameras have moved from analogue to digital, from standard definition to 4K and beyond. Storage has become cheaper. Compression has become increasingly sophisticated. Video Management Systems have become powerful. Artificial Intelligence can now detect, classify and search for an expanding range of objects, behavior and events. Yet one of the most fundamental problems of video surveillance remains largely untouched.

That problem is Conventional playback.

<u>Conventional Playback: The Villain of Video Surveillance. Why?</u>

Playback forces us to consume recorded events largely in the same dimension in which they occurred: *time*.

Record one hour and you have one hour of video.

Record 24 hours and you have 24 hours.

Now multiply that by the number of cameras. Consider 16 cameras recording at 25 frames per second.

Each camera produces: 2,160,000 frames every 24 hours.

Across 16 cameras: 34,560,000 frames.

Increase the number of cameras and the scale quickly becomes extraordinary:

Cameras	Frames in 24 hours at 25 fps
16	34.56 million
25	54 million
32	69.12 million
64	138.24 million

We have become extraordinarily good at recording all of this. The problem begins when somebody wants to look at it. That is where conventional playback becomes the villain.

The Hidden Burden of Conventional Playback

Digital CCTV commonly uses video compression standards such as H.264 and H.265. Compression is indispensable because it substantially reduces the bandwidth and storage required for modern digital video.[1][2]

But compression principally addresses transmission and storage. When recorded compressed video has to be viewed, the relevant information must be retrieved, decoded, and rendered into pictures that humans can see.

Now consider simultaneous playback.

Sixteen cameras at 25 fps represent 400 frames every second across those streams.

Twenty-five cameras represent 625 fps.

Thirty-two cameras represent 800 fps.

Sixty-four cameras represent 1,600 fps.

Frame count alone does not determine computing workload. Resolution, bitrate, codec, GOP structure, storage performance, client server architecture, hardware acceleration and VMS implementation all matter. But the computational burden is real.

Axis states that hardware decoding can improve performance and reduce CPU usage, particularly for high-resolution streams. Milestone similarly uses GPU hardware acceleration to improve decoding capability and notes that CPU load becomes high when a view contains many cameras.[3][4]

This is important because it separates two different problems:

The computer must process the visual information.

The human must meaningfully examine it.

Increasing computing power can help the first. It cannot eliminate the second.

Why Conventional Faster Playback cannot become the Hero

Increasing playback speed does not solve the underlying problem.

Conventional playback still requires multiple compressed video streams to be retrieved, decoded, synchronized and continuously rendered. As the number of cameras being played simultaneously increases, so does the processing, memory, storage and graphics workload. Even with modern techniques such as sub-streaming, lower-resolution secondary streams and dynamic stream switching, aggregate multi-camera decoding and rendering remain practical system considerations.

Modern VMS platforms have made substantial advances in dealing with this. Hardware-accelerated decoding, adaptive streaming, thumbnails, event search, motion search,

metadata and analytics can dramatically improve investigation and retrieval.[3][4][5]

These are important advances and should not be understated. But there is also a human problem. Reviewing recorded footage is rarely a simple linear journey from start to finish.

An operator may need to pause, reverse, move forward, slow down, accelerate, zoom, inspect a particular moment across cameras, enhance brightness or contrast, apply forensic filters, compare cameras, preserve relevant evidence and immediately create a report.

Conventional VMS playback was primarily designed around retrieving, investigating and interacting with recorded video, rather than the systematic, routine auditing of large quantities of footage as an organizational process.

So, the issue is not that faster playback is bad. The issue is that making conventional playback faster does not remove the computational burden or make large-scale human review practical.

Research into video summarization has recognized this broader consumption problem for years. The objective of summarization is precisely to enable users to grasp large quantities of video content more efficiently, while multi-view research specifically addresses the additional complexity and redundancy created by multiple camera views.[6][7]

Human-factors research identifies another constraint. CCTV monitoring involves visual search, cognitive and perceptual overload, attentional failures and decision-making across dynamically changing environments. Research comparing single-screen and multiple-screen CCTV monitoring also demonstrates that display configuration affects operator performance. [8][9] UK NPSA guidance similarly treats CCTV as a human-machine system in which operator capability, workload and control-room design materially affect performance.[10]

No amount of technology makes human attention infinite. Modern search certainly helps. Motion search, object search, event metadata, analytics and increasingly sophisticated AI can reduce the amount of footage an investigator needs to examine. In a separate paper, a proposed taxonomy of video review technologies examines these approaches and distinguishes them from deterministic, human-led visual auditing.[16]

But search has an inherent prerequisite:
There must be something to search for, detect, classify or infer. That distinction becomes important later. And what about the normal CCTV user?

It is also important not to confuse the capabilities of sophisticated enterprise VMS platforms with the everyday reality of the entire video surveillance market. Not every CCTV installation operates Milestone, Genetec, Avigilon or another premium enterprise VMS.

Major camera manufacturers themselves provide free or lightweight applications for everyday users. Hikvision, for example, describes iVMS-4200 as free software compatible with most Hikvision hardware and providing functions including live view and playback. Dahua positions SmartPSS Lite as a lightweight platform aimed at small and medium-sized businesses.[11][12]

Many CCTV environments therefore operate through manufacturer-supplied DVR/NVR software, desktop clients, web browsers, mobile applications or comparatively basic viewing interfaces. For these users, the question is even more fundamental:

How do they practically review large volumes of footage from multiple cameras, preserve what

matters, document what they find and act upon it? The playback problem is not confined to sophisticated control rooms. It exists wherever recorded footage exceeds the practical ability of people to review it.

The Real Limitation of Conventional Playback

Knowing when something happened helps locate the relevant footage. It does not solve the playback problem. If an incident occurred between 14:00 and 14:30, reviewing one camera may be manageable. But what if that same 30-minute period has to be examined across 16, 25, 64 or 100 cameras?

The system must still retrieve, decode, synchronize and render the relevant streams. The operator must still navigate them, pause, reverse, change speed, compare cameras and inspect relevant moments. Even a relatively short time window can therefore become cumbersome when multiplied across cameras.

And when there is no known incident, no alarm, no identified time and nothing already defined to search for, the challenge becomes substantially greater. The workload is determined not only by how much time must be reviewed, but by how many cameras, streams and frames must be processed and meaningfully examined.

The problem remained. So, we approached it differently. Read on…

The Typewriter Analogy

Consider the typewriter and Microsoft Office. The problem with the typewriter was not text. Nor was it the human being writing. The limitation was the mechanism through which humans were required to work with information.

Word processing changed the working model. Humans could create, edit, search, copy, organize and communicate information with radically greater efficiency. It did not replace human intelligence. It made human effort far more productive. The analogy is not that video and text are technically equivalent. They clearly are not. The analogy is about productivity.

Video Surveillance needs a similar Rethink

Video surveillance has created what may be one of the world's largest pools of dark visual data. Cameras record continuously, but how much of what they record is ever meaningfully reviewed?

Why are we not asking:

1. How can we see what the cameras saw, across an entire day and multiple cameras, in minutes?

2. How can we manage the enormous data burden?

3. How can we create cost-effective disaster recovery for important visual evidence?

4. How can we standardise incident reporting?

5. Why can CCTV signage not itself become part of a governance framework?

6. Could routine review create an entirely new professional skill, the CCTV Video Footage Auditor?

Most importantly:

Why are we not auditing CCTV footage? And if we are not auditing it: Where is the governance?

Enter the Hero: COM-SUR™

Protected by multiple granted patents, COM-SUR is the world's only CCTV Video Footage Auditing, Smart Backup, and Standardized Intelligent Incident Reporting platform. It works with visual

footage from CCTV, drones, body-worn cameras, dashcams, satellites, underwater ROVs and mobile cameras.

COM-SUR approached the problem differently. Instead of asking humans and computers to struggle with millions of conventional video frames, it makes an entire day's footage from multiple cameras practical to review rapidly.

Return to our 16 cameras

At 25 fps, they collectively represent 34.56 million frames over 24 hours. At a one-second interval, or another cadence defined by the user, COM-SUR transforms that day into: 86,400 time-stamped multi-camera images, each showing all 16 cameras. This is an important distinction.

The 86,400 images are not claimed to be equivalent to the original 34.56 million frames, nor do they replace the original recorded video. They provide a radically more manageable auditing representation of it.

The original video remains available whenever a visual observation requires closer examination, validation, investigation or evidentiary preservation. The capture cadence can also be selected according to the environment, risk and purpose of review. A one-second cadence is therefore a productivity choice, not a claim that events occurring between sampled moments cease to matter.

Then COM-SUR's purpose-built Media Player changes the review experience completely. The user can move rapidly through the time-stamped multi-camera images, forward or reverse, accelerate or slow down, pause, move frame-by-frame, zoom, inspect a particular moment across cameras, enhance brightness and contrast, apply forensic filters, preserve relevant evidence and move immediately into reporting.

COM-SUR therefore absolutely does use playback. The difference is what is being played back, how efficiently it can be navigated, and what the user can do while reviewing it. Instead of continuously wrestling with millions of frames distributed across multiple compressed video streams, the operator works with a radically more manageable visual representation of the day.

The objective is not merely to play video faster. It is to make an entire day's footage across multiple cameras practically auditable, with the time required depending on the audit cadence, camera layout, scene activity, audit criteria and depth of examination. That changes the purpose of playback. It is no longer merely a means of revisiting a known event. It becomes part of a structured human review process.

Where does AI fit?

AI has an important role, and its capabilities will continue to improve dramatically. Modern VMS and analytics platforms already demonstrate how effectively metadata, motion, objects, events and other defined attributes can accelerate forensic investigation.[5] But AI does not eliminate the need for systematic auditing.

AI is particularly powerful when detecting objects, behavior, patterns or anomalies that a system has been designed, configured or trained to identify. There is, however, another category of knowledge:

What don't we know that we should be looking for? You cannot necessarily write an analytics rule for something whose significance has not yet been recognized. Systematic human auditing can reveal recurring behavior, unusual conditions, operational exceptions and previously unrecognized patterns. Those observations can then identify entirely new opportunities for automation.

NIST's AI Risk Management Framework recognizes the importance of deployment context, domain

expertise, human oversight and feedback, including the identification of emergent scenarios and risks.[13][14]

But there is another side to this. Video surveillance is overwhelmingly associated with finding out what went wrong.

Why not apply the same discipline to identify what went right?

Cameras also record employees following procedures correctly, security personnel remaining alert, housekeeping teams maintaining standards, workers using PPE properly, staff assisting customers, drivers following safety protocols, teams responding correctly to emerging situations, and countless other examples of good practice. Applying this inverse logic allows organizations not only to detect failures, but also to identify, recognize and replicate good practice.

This must, of course, be governed appropriately. Systematic review involving employees, customers or members of the public must operate within the applicable legal purpose, privacy, access-control and data-protection framework. Data-protection authorities specifically emphasize purpose limitation and safeguards against unexpected secondary uses of surveillance footage.[15]

Auditing therefore becomes more than a control mechanism. It becomes a source of organizational learning, positive reinforcement and new automation opportunities. And this is where something else begins to emerge: New, site-specific data about what is actually happening in the organization, including patterns, behaviour, exceptions and good practices that nobody may previously have thought to look for.

We call this Frontier Data. Frontier Data is fresh, site-specific visual data from the organization's own operating environment, including observations that existing systems may not have been configured or trained to identify.

COM-SUR's Daily CCTV Video Footage Auditing Chart adds another critical component: structured human observations and feedback.

Together, the visual data and human observations can create a feedback loop:

Audit → Observe → Document → Learn → Identify patterns → Automate where appropriate → Audit again

This does not mean that human observations automatically become AI training data. Where Frontier Data is to be used for AI development, it must undergo the appropriate selection, labelling, annotation, governance, privacy, quality-control, validation and model-development processes.

The point is more fundamental: Auditing can reveal what may be worth automating in the first place. The relationship between humans and AI therefore need not be competitive. It can be iterative. AI helps us find more of what we know to look for. Auditing can help us discover what we don't know we should be looking for.

<u>From Playback to Visual Evidence Governance</u>

The real breakthrough is not simply faster review. It is what becomes possible once large volumes of recorded footage can actually be examined as a matter of routine. Observations can be documented. Relevant evidence can be preserved. Concerns can be escalated. Corrective and Preventive Action can be quicker. Good practices can be identified and replicated. New risks and patterns can be discovered. Frontier Data can be created. New opportunities for automation can emerge. And structured human observations can contribute to better-informed AI development.

That is Visual Evidence Governance: A disciplined process for systematically auditing what cameras recorded and ensuring that resulting visual evidence is reviewed, documented, preserved, escalated, acted upon and used for continuous improvement. CCTV can therefore move beyond being primarily a system for recording, live monitoring, alarms, analytics and post-incident investigation.

It can become a routinely governed source of visual evidence for security, safety, compliance, operations, training, risk management, organizational learning and better AI.

The Villain and the Hero

So, who is the villain? Not the camera. Not the human operator. Not video itself. The villain is conventional playback when it is expected to carry the burden of reviewing enormous quantities of recorded video across multiple cameras. The industry has fought that villain intelligently.

Faster playback, better codecs, hardware acceleration, smart search, metadata, thumbnails and AI analytics have all made conventional systems substantially more capable. But the villain remained.

Because even after the technology has retrieved, searched, decoded and displayed the footage, somebody still has to answer the most basic question: What did the cameras actually see that we need to know about?

Then came the Hero – COM-SUR

COM-SUR did not eliminate playback. It conquered its limitations. By transforming millions of frames from multiple cameras into a radically more manageable set of time-stamped multi-camera images, while retaining the original video as the underlying source, and combining this with a purpose-built Media Player and structured auditing workflow, COM-SUR changes the economics and practicality of human access to recorded visual information.

Playback stops being the bottleneck. It becomes the enabler. And once large volumes of footage across multiple cameras can be systematically reviewed as a routine activity, something much bigger becomes possible:

We can audit what the cameras saw.

We can document it.

We can preserve it.

We can learn from it.

We can act upon it.

And we can use what humans discover to determine what technology should automate next.

That is how the Hero conquers the Villain.

And that is where Visual Evidence Governance begins.

References

1. International Telecommunication Union, ITU-T Recommendation H.264: Advanced Video Coding for Generic Audiovisual Services.

2. International Telecommunication Union, ITU-T Recommendation H.265: High Efficiency Video Coding, current version approved January 2026.

3. Axis Communications, AXIS Camera Station Pro User Manual, hardware decoding and performance guidance.

4. Milestone Systems, XProtect Smart Client: Adaptive Streaming and Hardware Acceleration

and advanced settings documentation.

5. Milestone Systems, Finding Video Evidence with XProtect VMS, Centralized Search, metadata and forensic investigation capabilities.

6. Otani, M., Song, Y. and Wang, Y., Video Summarization Overview, 2022.

7. Hussain, T. et al., A Comprehensive Survey of Multi-View Video Summarization, Pattern Recognition, 2021.

8. Hodgetts, H.M. et al., See No Evil: Cognitive Challenges of Security Surveillance and Monitoring, Journal of Applied Research in Memory and Cognition, 2017.

9. Stainer, M.J. et al., Decision-making in Single and Multiple-Screen CCTV Surveillance, Applied Ergonomics, 2021.

10. UK National Protective Security Authority, Human Factors in CCTV Control Rooms: A Best Practice Guide.

11. Hikvision, iVMS-4200, manufacturer-provided free video-management software.

12. Dahua Technology, SmartPSS Lite, lightweight video-management platform for small and medium-sized businesses.

13. National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), 2023.

14. NIST AI Resource Center, AI RMF Core and Playbook, including deployment-context, domain-expert and feedback considerations.

15. European Data Protection Board, Guidelines 3/2019 on Processing of Personal Data Through Video Devices, and UK Information Commissioner's Office, Guidance on Video Surveillance.

16. G. D. Goradia, "A Proposed Taxonomy of Video Review Technologies," COM-SUR Video Surveillance and Imaging AI Technologies Pvt. Ltd., Mumbai, 2026.

COM-SUR product capabilities, Visual Evidence Governance, Frontier Data, the Daily CCTV Video Footage Auditing Chart and other COM-SUR-specific terminology in this paper are based on COM-SUR product documentation and operating methodology.