



PDF Download  
3746027.3761872.pdf  
09 February 2026  
Total Citations: 0  
Total Downloads: 65

 Latest updates: <https://dl.acm.org/doi/10.1145/3746027.3761872>

ABSTRACT

## Real-time GenAI Solutions for Video Streaming in Low-bandwidth Settings

CLAUDIO BAECCHI

MATTEO BRUNI

FABIO CLABOT

MARCO BERTINI, University of Florence, Florence, FI, Italy

Open Access Support provided by:

University of Florence

Published: 27 October 2025

[Citation in BibTeX format](#)

MM '25: The 33rd ACM International  
Conference on Multimedia  
October 27 - 31, 2025  
Dublin, Ireland

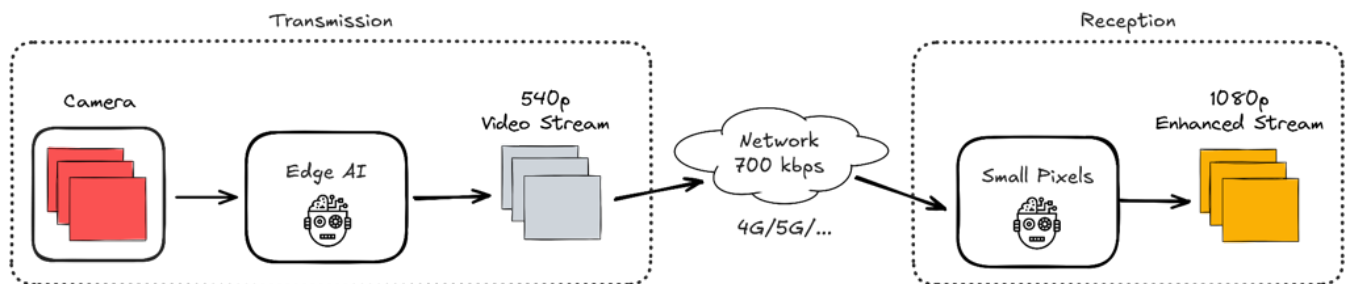
Conference Sponsors:  
SIGMM

# Real-time GenAI Solutions for Video Streaming in Low-bandwidth Settings

Claudio Baccchi\*  
Matteo Bruni\*  
claudio-baccchi@small-pixels.com  
matteo.bruni@small-pixels.com  
Small Pixels  
Firenze, Italy

Fabio Clabot  
fabio.clabot@small-pixels.com  
Small Pixels  
Firenze, Italy

Marco Bertini  
marco.bertini@unifi.it  
marco.bertini@small-pixels.com  
Università degli Studi di Firenze  
Small Pixels  
Firenze, Italy



**Figure 1: System schema: Small Pixels on receiver and on Edge. An AI-based pre-processing step can be applied on the edge device before transmission, enhancing video quality at low bitrates. This improves resilience to compression and helps preserve key visual details during streaming.**

## Abstract

Surveillance systems such as bodycams and drones often operate under bandwidth constraints that limit video quality and degrade both human monitoring and AI-based analytics. Traditional compression techniques introduce artifacts that obscure critical details, especially in high-motion scenarios. We present a generative AI-powered video compression framework developed by Small Pixels, a spin-off of the University of Florence, designed to deliver Full HD video at significantly reduced bitrates. The system combines edge-side preprocessing for compression resilience with real-time receiver-side super-resolution, enabling up to 50% bandwidth savings while preserving perceptual quality and detection accuracy. Objective evaluations on EgoSeg and VisDrone datasets show +6.2 VMAF improvement and stable YOLOv11 detection performance with 30% less bitrate. Live trials in Singapore, within the Singapore Hatch-X Global Innovation Program, validated real-time operation with minimal latency, demonstrating clearer faces and motion in challenging conditions. The solution integrates seamlessly into existing infrastructures without hardware upgrades, offering a practical path to reliable, high-quality video streaming in bandwidth-limited environments.

\*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MM '25, Dublin, Ireland

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2035-2/2025/10

<https://doi.org/10.1145/3746027.3761872>

## CCS Concepts

• **Computing methodologies** → *Object detection*; • **Computer vision tasks**; • **Information systems** → **Multimedia streaming**.

## Keywords

GenAI, video coding, video super resolution, video streaming, bodycam

## ACM Reference Format:

Claudio Baccchi, Matteo Bruni, Fabio Clabot, and Marco Bertini. 2025. Real-time GenAI Solutions for Video Streaming in Low-bandwidth Settings. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25)*, October 27–31, 2025, Dublin, Ireland. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3746027.3761872>

## 1 Introduction

Video surveillance systems increasingly rely on mobile devices such as body-worn cameras and drones to provide real-time situational awareness in public safety and security contexts. However, these systems are constrained by limited or unstable bandwidth, especially in field operations where wireless connectivity cannot guarantee high throughput. Traditional video compression schemes (e.g., H.264/H.265) face a fundamental trade-off: higher compression reduces bandwidth but introduces severe artifacts such as blocking, flickering, and motion blur. These artifacts hinder both human operators, who require clarity to identify persons and behaviors, and AI systems, which depend on visual fidelity for accurate detection and recognition. Recent advances in generative AI (GenAI) for video super-resolution and enhancement offer new opportunities

to overcome these limitations [3–5]. Instead of treating compression artifacts as unavoidable, GenAI models can learn to anticipate or reconstruct lost details, enabling high-quality streaming at significantly reduced bitrates. This paper presents a GenAI-powered solution developed by Small Pixels (University of Florence spin-off) and validated through the Singapore Hatch-X Global Innovation Program. The system is designed for real-time deployment on bodycams and drones, offering up to 50% bandwidth savings while maintaining Full HD perceptual quality and operational performance.

## 2 System design

The solution combines two complementary approaches:

- **Edge-side preprocessing** – High-resolution video is processed directly on the capture device or nearby edge node to enhance compression resilience, reducing artifact propagation during transmission.
- **Receiver-side upscaling** – Alternatively, video is captured at lower resolution (e.g., 540p) and transmitted at reduced bitrate. On the receiver side, a GenAI super-resolution model reconstructs the stream to Full HD in real time.

This design achieves two key objectives:

- **Bandwidth optimization:** transmission at 700 kbps compared to >1 Mbps baselines.
- **Quality preservation:** sharper facial features, clearer textures, and more stable motion rendering.

The models were trained on a curated dataset comprising bodycam-like and drone-like scenarios collected in-house, complemented by publicly available datasets such as EgoSeg [1] and VisDrone [6]. Training emphasized high-motion and dynamic backgrounds, where compression artifacts are most severe.

## 3 Methodology and evaluation

The intended real-world use case involves video stream analysis by a human operator and AI-based crowd counting algorithms. Two simulated applications were tested:

- **Bodycam:** chest-mounted camera simulating police usage.
- **Drone:** elevated camera simulating aerial surveillance.

### Experimental setup:

- **Baseline (A):** 1080p at 1000 kbps, standard encoding, without using Small Pixels technology.
- **Small Pixels AI (B):** 540p at 700 kbps, upscaled to 1080p using Small Pixels GenAI super-resolution.
- **Small Pixels AI + Edge (C):** 540p at 700 kbps with Small Pixels preprocessing + Small Pixels upscaling on receiver.

Evaluation used VMAF [2] for perceptual quality, YOLOv11 mean Average Precision (mAP) for downstream AI tasks, and subjective assessments during live demos. The EgoSeg dataset was used for visual quality evaluation, while the VisDrone dataset was used for AI algorithm processing evaluation.

## 4 Results

The expected goal, that was met, was to obtain improved or same quality and detection with ~30% bitrate. Table 1 summarizes results, showing how the goals were met in all the use cases.



Figure 2: *top)* frames from the EgoSeg dataset; *bottom)* frames from the VisDrone dataset.

Table 1: Performance comparison of baseline vs. Small Pixels AI solutions.

Scenario	Bitrate	Resolution	VMAF ↑	mAP ↑
Baseline (A)	1000 kbps	1080p	69.57	31.02%
SP AI (B)	700 kbps	540p→1080p	75.81	31.07%
SP AI+Edge (C)	700 kbps	540p→1080p	<b>77.69</b>	<b>31.48%</b>

Results show ~30% bandwidth savings, +6.2 VMAF improvement, and stable or improved detection performance. Live demos in Singapore confirmed subjective gains in clarity, particularly for faces and motion. In particular setup (C) that combines Small Pixels GenAI both on camera and on receiver results in the best performance.

## 5 Conclusion

We presented a GenAI-powered video compression system enabling high-quality surveillance at reduced bitrates. The solution demonstrated improvements in both perceptual quality and AI analytics, with seamless integration into existing infrastructures.

The results highlight the potential of GenAI to redefine the compression-quality trade-off in constrained video streaming. By learning compression dynamics, the Small Pixels models either optimize video before encoding or reconstruct detail after decoding. Importantly, the solution requires no hardware upgrades to existing cameras, easing deployment for security agencies.

At the conference, we will demonstrate real-time operation on police bodycams and drones, illustrating how the system supports law enforcement in bandwidth-constrained environments.

## References

- [1] 2020. EgoSeg Dataset. <https://vision.huji.ac.il/egoseg/videos/dataset.html>.
- [2] 2021. VMAF: The Journey Continues. <https://netflixtechblog.com/vmaf-the-journey-continues-44b51e9ed12>.
- [3] Leonardo Galteri, Marco Bertini, Lorenzo Seidenari, Tiberio Uricchio, and Alberto Del Bimbo. 2020. Increasing Video Perceptual Quality with GANs and Semantic Coding. In *Proc. of ACM International Conference on Multimedia (ACM MM)* (Seattle, WA, USA) (MM '20). Association for Computing Machinery, New York, NY, USA, 862–870. doi:10.1145/3394171.3413508
- [4] Leonardo Galteri, Lorenzo Seidenari, Marco Bertini, Tiberio Uricchio, and Alberto Del Bimbo. 2019. Fast Video Quality Enhancement using GANs. In *Proc. of ACM International Conference on Multimedia (ACM MM)*. 1065–1067. doi:10.1145/3343031.3350592
- [5] X. Wang et al. 2020. Deep learning for video compression: A review. *IEEE Transactions on Circuits and Systems for Video Technology* 30, 9 (2020), 3442–3458.
- [6] Pengfei Zhu et al. 2018. VisDrone-DET2018: The Vision Meets Drone Object Detection in Image Challenge Results. In *Proc. of the ECCV Workshops*.