

Video Coding for Machine Vision Applications

ΝΙΚΗΦΟΡΟΣ ΠΛΑΝΑΚΗΣ
ΓΕΩΡΓΙΟΣ - ΔΑΥΙΔ ΤΣΙΤΛΑΟΥΡΙ

Abstract

Η κωδικοποίηση βίντεο για εφαρμογές τεχνητής νοημοσύνης αποκτά ολοένα και μεγαλύτερη σημασία στις μερες μας, καθώς πολλές εφαρμογές βασίζονται σε επεξεργασία οπτικού περιεχομένου σε πραγματικό χρόνο, συχνά σε περιβάλλοντα νέφους (**cloud-based**). Παράλληλα, η ανάγκη καλύτερης αξιοποίησης υπολογιστικών πόρων και χώρου αποθήκευσης επιβάλλει νέους τρόπους κωδικοποίησης που δεν στοχεύουν στην οπτική ποιότητα, δηλαδή όπως την αντιλαμβάνεται ο άνθρωπος. Αντίθετα στοχεύει στην υψηλότερη δυνατή απόδοση των αλγορίθμων μηχανικής μάθησης.

Στην παρούσα εργασία, παρουσιάζονται οι τρέχουσες εξελίξεις στην κωδικοποίηση βίντεο προσανατολισμένη για μηχανές (**machine-oriented video coding**) και εξετάζονται οι αντισταθμίσεις κόστους υπολογιστικού καθώς και αποθήκευσης, μέσα από παραδείγματα πραγματικών εφαρμογών. Στην περίπτωση μας τα παραδείγματα αυτά είναι η **τηλεϊατρική με ρομποτικές επεμβάσεις** και το **cloud gaming με ανάλυση συμπεριφοράς παικτών**. Συζητάμε τα σύγχρονα ερευνητικά ερωτήματα και τις μελλοντικές κατευθύνσεις που διαμορφώνουν το πεδίο. Τέλος, παρουσιάζονται πίνακες και σχήματα που αποσαφηνίζουν την τρέχουσα τεχνολογική κατάσταση.

Introduction

Η επεξεργασία και μετάδοση βίντεο αποτελεί θεμελιώδες ζήτημα σε πολλές εφαρμογές πολυμέσων. Παραδοσιακά, η κωδικοποίηση βίντεο βελτιστοποιείται για να παρέχει πιο ικανοποιητική ποιότητα στον ανθρώπινο θεατή, δίνοντας έτσι έμφαση στη **μείωση οπτικών artifacts** (ανωμαλία ή παραμόρφωση εικόνας). Ωστόσο, η ραγδαία εξέλιξη της τεχνητής νοημοσύνης και των αλγορίθμων μηχανικής όρασης εισάγει την ανάγκη για **machine-oriented video coding**, όπου βασικό κριτήριο είναι η **διατήρηση των κρίσιμων χαρακτηριστικών** που χρησιμοποιεί ένα νευρωνικό δίκτυο ή ένας αλγόριθμος ανίχνευσης ή ταξινόμησης.

Οι εφαρμογές αυξάνονται αλματωδώς: **συστήματα επιτήρησης** (IP κάμερες), **αυτόνομα οχήματα** (αυτόνομα ταξί), **βιομηχανική αυτοματοποίηση** (Programmable Logic Controllers), **τηλεϊατρική και ρομποτικές επεμβάσεις** (Da Vinci Surgical System), καθώς και **cloud gaming** (NVIDIA GeForce NOW). Σε όλα αυτά τα σενάρια, η μετάδοση βίντεο σε πραγματικό ή “σχεδόν” πραγματικό χρόνο, η αποθήκευση και η επεξεργασία του σε κέντρα δεδομένων cloud, συνεπάγονται υψηλά κόστη και απαιτούν βέλτιστη αξιοποίηση υποδομών. Το ερώτημα λοιπόν, αφορά στο πώς μπορεί να σχεδιαστεί η κωδικοποίηση βίντεο, με τρόπο τέτοιο ώστε να ικανοποιεί τις απαιτήσεις των αλγορίθμων AI, ελαχιστοποιώντας όμως την περιττή πληροφορία και άρα το κόστος. Στην παρούσα λοιπόν εργασία, ο σκοπός μας είναι:

- Παρουσιάζουμε τις **βασικές αρχές** κωδικοποίησης βίντεο για μηχανές.
- Αναλύουμε τις **τρέχουσες προκλήσεις** σε cloud-based συστήματα, ιδιαίτερα όσον αφορά τις **αντισταθμίσεις κόστους υπολογιστικού, αποθήκευσης και bandwidth**.

- Προτείνουμε και περιγράφουμε **τεχνικές** που στοχεύουν σε βέλτιστη απόδοση των μοντέλων AI (feature-preserving compression, ROI-based κτλ.).
- Παρουσιάζουμε **μελέτες περίπτωσης** από την τηλεϊατρική (ρομποτικές επεμβάσεις) και το cloud gaming (player behavior analysis).
- Συζητούμε τα **ευρήματα** και τις **προκλήσεις** που προκύπτουν σε αυτό το πεδίο.

Background

Η κωδικοποίηση βίντεο έλκει την καταγωγή της από τα πρώτα πρότυπα MPEG-1, MPEG-2, ενώ πιο πρόσφατα στάνταρ όπως H.264/AVC, HEVC/H.265 και VVC/H.266 προσφέρουν σημαντικά υψηλότερη συμπίεση. Όλα αυτά στοχεύουν κυρίως στη βελτίωση της οπτικής ποιότητας **όπως την αντιλαμβάνεται ο άνθρωπος**, χρησιμοποιώντας μετρικές όπως **PSNR, SSIM**, ή πιο σύγχρονες μεθόδους όπως το **VMAF**, που χρησιμοποιούν μεγάλες εταιρείες του χλωρου όπως το Netflix.

Παραδοσιακή Κωδικοποίηση Βίντεο

- **Χωρική Συμπίεση (Intra coding):** Αποσκοπεί στην μείωση της πλεονάζουσας πληροφορίας εντός ενός καρέ (DCT μετασχηματισμός).
- **Χρονική Συμπίεση (Inter coding):** Χρήση motion estimation/compensation μεταξύ διαδοχικών καρέ, μεταδίδοντας διαφορές αντί ολόκληρων καρέ.
- **Μετρικές Οπτικής Ποιότητας:** PSNR, SSIM, VMAF, που εκτιμούν πόσο “καθαρό” φαίνεται το βίντεο στον άνθρωπο.

Από Ανθρώπινη σε Μηχανική Όραση

Με τη στροφή σε εφαρμογές **μηχανικής όρασης**, το ζητούμενο δεν είναι απαραίτητα οπτικά ευχάριστες εικόνες για το ανθρώπινο μάτι. Η **ποιότητα** καθορίζεται από την απόδοση των αλγορίθμων AI, οι οποίοι επηρεάζονται διαφορετικά από τα κλασικά artifacts. Για παράδειγμα, μια **περιοχή με θόρυβο που είναι απαρατήρητη στο ανθρώπινο μάτι** μπορεί να επηρεάσει σημαντικά ένα σύστημα ανίχνευσης αντικειμένων.

Cloud-Based AI & Κόστος

Επιπλέον, οι σημερινοί όγκοι δεδομένων φιλοξενούνται και επεξεργάζονται **στο cloud**, με αποτέλεσμα οι οργανισμοί να πληρώνουν τόσο για την αποθήκευση όσο και για την επεξεργασία. Έτσι, έρχονται στο προσκήνιο οι **αντισταθμίσεις**:

- **Μεγαλύτερο bitrate** → **αυξημένο κόστος bandwidth/αποθήκευσης**, πιθανώς καλύτερη ακρίβεια ανίχνευσης.
- **Μειωμένο bitrate** → **χαμηλότερο κόστος**, αλλά αυξημένα artifacts που υποβαθμίζουν τις επιδόσεις των μοντέλων AI.

Ένα σημαντικό ζητούμενο είναι η **ελαχιστοποίηση των “άχρηστων” δεδομένων** που δεν συμβάλλουν στην ανάλυση, διατηρώντας όμως ό,τι χρειάζεται για υψηλή απόδοση AI.

Current Challenges

- **Υπολογιστικό Κόστος:**
Οι προηγμένοι αλγόριθμοι κωδικοποίησης, όπως αυτοί που βασίζονται σε ROI (Region-of-Interest), απαιτούν:
 - Υπολογιστική ισχύ για τον εντοπισμό και την ανάλυση των περιοχών ενδιαφέροντος.
 - Συγχρονισμό με real-time εφαρμογές, καθώς οποιαδήποτε καθυστέρηση μπορεί να επηρεάσει την εμπειρία του τελικού χρήστη.
 - Εξισορρόπηση μεταξύ πολυπλοκότητας του αλγορίθμου και επάρκειας διαθέσιμων πόρων (CPU/GPU).
- **Κόστος Αποθήκευσης & Bandwidth:**
 - Η συνεχής αύξηση του όγκου δεδομένων υψηλής ανάλυσης δημιουργεί εκθετική αύξηση του κόστους σε cloud πλατφόρμες.
 - Συχνά απαιτούνται πρόσθετες τεχνικές αποθήκευσης, όπως deduplication ή compression σε πραγματικό χρόνο.
- **Ποικιλία Εφαρμογών:**
 - Οι διαφορετικές απαιτήσεις από κλάδο σε κλάδο δημιουργούν την ανάγκη προσαρμοσμένων λύσεων.
 - Παράλληλα, η ανάγκη για κοινά πρότυπα αποτροπής ασυμβατότητας μεταξύ εφαρμογών και συσκευών είναι κρίσιμη.
- **Low Latency:**
 - Σε συστήματα όπως οι ρομποτικές επεμβάσεις ή τα παιχνίδια σε πραγματικό χρόνο, ακόμα και καθυστερήσεις κάτω του 1 ms μπορούν να επηρεάσουν τη λειτουργικότητα.
 - Προκλήσεις προκύπτουν από την αλληλεπίδραση πολλών παραγόντων, όπως η απόσταση, το jitter, και οι δρομολογήσεις δεδομένων σε δίκτυα.
- **Ενσωμάτωση με Υφιστάμενα Στάνταρ:**
 - Οι νέες πειραματικές τεχνολογίες συχνά δεν υποστηρίζονται από παλαιότερες εκδόσεις hardware, δημιουργώντας εμπόδια υιοθέτησης.
 - Πολλά πρότυπα (π.χ. MPEG) βρίσκονται σε στάδιο αναπροσαρμογής για AI-driven workloads.

Proposed Techniques

Feature-Preserving Compression:

- Στοχεύει στη διατήρηση βασικών χαρακτηριστικών εικόνας, όπως edges και textures, που είναι κρίσιμα για CNNs.
- Χρησιμοποιούνται εξελιγμένοι μετασχηματισμοί (wavelet transforms) για να εξασφαλιστεί η βέλτιστη ανάλυση και compressibility.

Region-of-Interest (ROI) Encoding:

- Εφαρμογή αλγορίθμων machine learning για αυτόματη αναγνώριση αντικειμένων υψηλού ενδιαφέροντος (π.χ. πρόσωπα ή οχήματα).
- Στο υπόλοιπο περιβάλλον εφαρμόζεται αυξημένη συμπίεση χωρίς σημαντική μείωση της συνολικής ποιότητας.

Adaptive Frame-Rate:

- Ενεργή παρακολούθηση σκηνών για δυναμική προσαρμογή του frame rate.
- Τεχνικές interpolation μπορούν να διατηρήσουν ομαλότητα κίνησης ακόμα και με μειωμένα fps.

Edge/Cloud Synergy:

- Χρησιμοποιώντας lightweight neural networks στο edge, μειώνεται ο όγκος δεδομένων που μεταδίδεται στο cloud.
- Οι υπολογισμοί είναι κλιμακούμενοι: π.χ. οι λιγότερο κρίσιμες περιοχές μπορούν να επεξεργάζονται αποκλειστικά στο edge.

Scalable Video Coding (SVC):

- Προσθήκη επιπέδων ποιότητας που μπορούν να διαχειριστούν παράλληλα διαφορετικές απαιτήσεις χρηστών ή συστημάτων AI.
- Εφαρμογή layer prioritization, εξασφαλίζοντας κρίσιμα δεδομένα πρώτα σε περιβάλλοντα περιορισμένου bandwidth.

Συνοπτική σύγκριση κλασικής και machine oriented κωδικοποίησης:

Παράμετρος	Μέτρο Αξιολόγησης	Σκοπός	Περιοχές Ενδιαφέροντος (ROI)	Bitrate / Αρχιτεκτονική	Υλοποίηση (Hardware Decoders)
Κλασική Κωδικοποίηση	PSNR, SSIM, VMAF	Οπτική ποιότητα για άνθρωπο	Δεν δίνεται ιδιαίτερη έμφαση	Σταθερό ή VBR (visual-based)	Ευρέως διαθέσιμες λύσεις

Machine-oriented κωδικοποίηση	mAP, IoU, Recall, Precision σε CV tasks	Ακρίβεια / Αποδοτικότητα για AI	Έμφαση σε αντικείμενα και χαρακτηριστικά	Προσαρμοσμένο στις ανάγκες του αλγορίθμου	Συχνά custom ή GPU-accelerated
--------------------------------------	---	---------------------------------	--	---	--------------------------------

Case Studies

Μελέτη Περίπτωσης #1: Τηλεϊατρική και Ρομποτικές Επεμβάσεις

- **Πλαίσιο:** Ρομποτικές επεμβάσεις εξ αποστάσεως, όπου ο χειρουργός βλέπει βίντεο υψηλής ανάλυσης και λαμβάνει feedback από AI (π.χ. segmentation).
- **Ανάγκη/Πρόκληση:**
 - **Ultra-Low Latency:** Καθυστερήσεις λίγων ms είναι κρίσιμες για την ασφάλεια του χειρουργημένου.
 - **Διατήρηση Ιατρικών Λεπτομερειών** (αγγεία, ιστοί), χωρίς υπερβολικά μεγάλο bitrate.
 - **Edge Computing:** Προ-επεξεργασία σε τοπικό σταθμό για απόρριψη αχρείαστων δεδομένων.
- **Προσέγγιση:**
 - **Feature-Aware Compression:** Ειδικά trainable codecs που εστιάζουν σε ιατρικά «patterns».
 - **Low-Latency H.265/H.266 Implementations:** Μειωμένο buffer, adaptive GOP size, ελαχιστοποίηση ενδοπλαισιακών καθυστερήσεων.
 - **Cloud Feedback:** Αν το AI εντοπίσει ιδιαίτερα κρίσιμες περιοχές, ζητά δυναμικά αύξηση του bitrate.
- **Εναλλακτικές Αρχιτεκτονικές**
 - Σε ορισμένες εφαρμογές, χρησιμοποιούνται αποκεντρωμένες (decentralized) αρχιτεκτονικές στις οποίες μικρο εξυπηρετητές (micro-servers) εγκαθίστανται σε περιφερειακά νοσοκομεία. Εκεί διενεργείται μια πρώτη φάση κωδικοποίησης/ανάλυσης, ώστε το κεντρικό κέντρο δεδομένων (data center) να λαμβάνει μόνο τα συμπιεσμένα ή ήδη επεξεργασμένα καρέ.
 - Οι γιατροί μπορούν να έχουν πρόσβαση σε ένα “control panel” με εργαλεία AI (π.χ. αλγόριθμος εκτίμησης αιμορραγίας), ο οποίος αναλύει συνεχώς τις ροές βίντεο.
- **Βιοσυμβατότητα των ROI**
 - Σε μερικά συστήματα, το ROI για ρομποτικές επεμβάσεις ορίζεται δυναμικά, όχι μόνο βάσει εικόνας αλλά και βάσει αισθητήρων (π.χ. πίεση, φασματικές μετρήσεις). Έτσι ενισχύονται τα pixels που αντανακλούν κρίσιμες βιολογικές πληροφορίες (π.χ. αιμοδυναμικές αλλαγές).
- **Θέματα Ασφάλειας και Ιδιωτικότητας**
 - Σε τηλεϊατρικά συστήματα όπου τα βίντεο φτάνουν σε τρίτους (συμβουλευτικό προσωπικό, ειδικούς), εφαρμόζονται συχνά τεχνικές ανωνυμοποίησης περιοχών που δεν

σχετίζονται με το χειρουργικό πεδίο. Αυτό σημαίνει έξτρα «συμπίεση» ή αλλοίωση (obfuscation) σε δευτερεύουσες περιοχές, με έμφαση στη διατήρηση της κλινικά σημαντικής ζώνης.

- Η κρυπτογράφηση σε πραγματικό χρόνο συνδυάζεται με τα μηχανήματα κωδικοποίησης, διασφαλίζοντας εμπιστευτικότητα αλλά αυξάνοντας το υπολογιστικό φορτίο.
 - **Οικονομική Διάσταση**
 - Τα κόστη διασυνδέσεων υψηλής ευκρίνειας (fiber) και low-latency satellite links σε απομακρυσμένες περιοχές συχνά επιβάλλουν έξυπνη συμπίεση σε ασταθή δίκτυα.
 - **Αποτελέσματα:**
 - Καλύτερο trade-off μεταξύ ποιότητας κρίσιμων περιοχών και χαμηλότερου όγκου μεταδιδόμενων δεδομένων.
 - Αξιόπιστη τηλε-χειρουργική σε καταστάσεις με περιορισμένο δίκτυο.
-

Μελέτη Περίπτωσης #2: Cloud Gaming με Ανάλυση Παίκτη (Player Behavior Analysis)

- **Πλαίσιο:** Οι πλατφόρμες Cloud Gaming (NVIDIA GeForce Now, Google Stadia) προσφέρουν streaming παιχνιδιών σε πραγματικό χρόνο. Παράλληλα, AI συστήματα τρέχουν είτε για **cheat detection** είτε για advanced analytics.
- **Ανάγκη/Πρόκληση:**
 - **Χαμηλό Latency** για ομαλό gameplay.
 - **Υψηλή Ανάλυση και Framerate** (4K, 60–120 fps).
 - **Επαρκής Ποιότητα για AI:** Ανίχνευση αλληλεπιδράσεων, events, παικτική συμπεριφορά.
- **Προσέγγιση:**
 - **Scalable Video Coding (SVC) + ROI:** Το mainstream layer για τον παίκτη, έξτρα layer ή ROI-encoding για ανάλυση AI, όπως π.χ. παρακολούθηση του χαρακτήρα.
 - **Edge Inference:** Μερική ανάλυση γίνεται σε servers κοντά στο user, με γρήγορη απόκριση.
 - **AI-Driven Bitrate Adaptation:** Σε δράση (action scenes), το σύστημα αυξάνει προσωρινά το bitrate στις σημαντικές περιοχές.
- **Συνδυασμός με Εμπειρία Θεατή (Spectator Mode):**
 - Πέρα από τον παίκτη που αλληλεπιδρά στο παιχνίδι, υπάρχουν συχνά θεατές (π.χ. Twitch streams, eSports τουρνουά). Ο ίδιος κωδικοποιητής μπορεί να παράγει **πολλαπλά ρεύματα (streams)**: ένα για real-time gaming (με χαμηλότερο latency) και ένα με ανάλυση υψηλής ανάλυσης/ποιότητας για τους θεατές.
 - Σε αυτές τις περιπτώσεις, το ROI-based encoding μπορεί να στοχεύει διαφορετικά σημεία: οι θεατές βλέπουν το «πιο εντυπωσιακό» μέρος της σκηνής, ενώ το AI μοντέλο

λαμβάνει καρέ ειδικά εστιασμένα σε αντικείμενα/game events (παίκτες, mobs, objectives).

- **Εντοπισμός Cheating Μέσω Ανάλυσης Σημείων Στόχευσης:**
 - Σε FPS παιχνίδια, το AI μπορεί να αναλύει τις μεταβολές κίνησης του mouse, τη στόχευση σε συγκεκριμένα pixel. Για να λειτουργήσει αξιόπιστα, χρειάζεται «καθαρά» edge maps στους χαρακτήρες. Άρα το κωδικοποιητικό σύστημα ρυθμίζεται ώστε να ελαχιστοποιεί artifacts γύρω από τα όπλα ή τις φηγούρες των παικτών.
- **Προγνωστική Ανάλυση Συμπεριφοράς:**
 - Τα gaming studios ενδιαφέρονται και για player retention (διατήρηση παικτών). Μέσω ανάλυσης βίντεο, μπορούν να εντοπίσουν αν κάποιος βαριέται (idle times), αν επαναλαμβάνει μοτίβα rage quitting κ.ά. Εδώ χρειάζεται λεπτομερής συμπίεση στα UI στοιχεία (π.χ. μπάρα HP, counters), ώστε τα AI συστήματα να «διαβάζουν» σωστά τα in-game events.
- **Μοντέλα Συνεκπαίδευσης (Co-training):**
 - Ορισμένες cloud gaming πλατφόρμες εκπαιδεύουν AI παράλληλα με τη ροή κωδικοποίησης. Σε real-time scenarios, αν ο αλγόριθμος εντοπίσει αύξηση στη σύνθετη κίνηση (πολλοί παίκτες σε ένα σημείο), μπορεί να ειδοποιήσει τον κωδικοποιητή για να αυξήσει το bitrate τοπικά (στο region του map όπου γίνεται η ομαδική μάχη).
- **Dimension Scaling:**
 - Για να διατηρηθεί σταθερό το framerate, το cloud gaming σύστημα μπορεί να μειώσει προσωρινά την ανάλυση (π.χ. από 1440p σε 1080p) σε σκηνές υπερβολικής κίνησης, επιτρέποντας στο ROI-based AI layer να αναβαθμίσει (upsample) κρίσιμες περιοχές, όπως οι παίκτες, οι αντίπαλοι ή τα objectives.
- **Αποτελέσματα:**
 - Βέλτιστη εμπειρία χρήστη (ελαχιστοποίηση lag).
 - Αξιόπιστες AI μετρήσεις (detections, stats) χωρίς δραματική αύξηση bandwidth.

Discussion

Οι μελέτες περίπτωσης που παρουσιάστηκαν -είτε προέρχονται από την τηλεϊατρική είτε από το cloud gaming- αναδεικνύουν το κρίσιμο **trade-off** ανάμεσα στην ποιότητα/ακρίβεια ανάλυσης που απαιτούν τα συστήματα AI και στο συνολικό κόστος (υπολογιστικό, αποθήκευσης, bandwidth) που πρέπει να διαχειριστούν οι πάροχοι. Στην περίπτωση των ρομποτικών επεμβάσεων, το **ultra-low latency** και η **πιστή διατήρηση κρίσιμων ιατρικών λεπτομερειών** κυριαρχούν ως κριτήρια. Στο cloud gaming, η **εμπειρία του παίκτη** (δηλαδή χαμηλό lag και υψηλή οπτική ποιότητα) πρέπει να συμπλέει με την **ανάγκη των AI εργαλείων** για επαρκή «πληροφορία» ώστε να διεξάγουν ανάλυση της συμπεριφοράς ή εντοπισμό παραβιάσεων.

Κάθε σενάριο, λοιπόν, απαιτεί διαφορετικές ιεραρχήσεις:

- **Προτεραιότητα στην ακρίβεια** (π.χ. σε χειρουργική διαδικασία) μπορεί να αυξήσει το υπολογιστικό φορτίο.
- **Προτεραιότητα στη μείωση κόστους** (π.χ. σε πολλαπλά gaming streams) μπορεί να επιφέρει διακυμάνσεις στο bitrate, επηρεάζοντας την ακρίβεια των μοντέλων.

Παράλληλα, διακρίνεται μια **κοινή κατεύθυνση**: η “έξυπνη” συμπίεση (feature-preserving, ROI-based, multi-layer) που συντονίζεται δυναμικά με τις ανάγκες της εφαρμογής και της ΑΙ ανάλυσης. Αυτό καθιστά αναγκαίο καινούριο **co-design** ανάμεσα στους κωδικοποιητές βίντεο και τους αλγορίθμους μηχανικής όρασης, αμφισβητώντας τα όρια των παραδοσιακών μετρικών ποιότητας.

Conclusion

Συνολικά, η κωδικοποίηση βίντεο για μηχανές αποτελεί έναν πολλά υποσχόμενο αλλά ταυτόχρονα απαιτητικό τομέα, όπου η οικονομία των πόρων και η βέλτιστη απόδοση των αλγορίθμων ΑΙ πρέπει να επιτευχθούν ταυτόχρονα. Οι μελλοντικές εξελίξεις πιθανόν να στραφούν σε πλήρως **end-to-end** συστήματα, όπου ο κωδικοποιητής “γνωρίζει” τις απαιτήσεις του αλγορίθμου ανάλυσης (object detection, segmentation, κ.ά.) και προσαρμόζεται σε πραγματικό χρόνο, προσφέροντας το μέγιστο δυνατό όφελος.

Η **διεπιστημονική έρευνα** ανάμεσα σε ομάδες που εξειδικεύονται στην κωδικοποίηση βίντεο, τη μηχανική μάθηση και τον σχεδιασμό cloud αρχιτεκτονικών είναι κρίσιμη, ώστε να επιταχυνθεί η πρακτική εφαρμογή λύσεων machine-oriented video coding σε βιομηχανική κλίμακα.

References

1. Duan, L., Liu, J., Yang, W., Huang, T., & Gao, W. (2020). Video coding for machines: A paradigm of collaborative compression and intelligent analytics. *IEEE Transactions on Image Processing*, 29, 8680-8695.
2. Gao, W., Xu, X., Qin, M., & Liu, S. (2022, October). An open dataset for video coding for machines standardization. Στο *2022 IEEE International Conference on Image Processing (ICIP)* (σελ. 4008-4012). IEEE.
3. Lin, H., Chen, B., Zhang, Z., Lin, J., Wang, X., & Zhao, T. (2023, October). Deepsvc: Deep scalable video coding for both machine and human vision. Στο *Proceedings of the 31st ACM International Conference on Multimedia* (σελ. 9205-9214).
4. Sheng, X., Li, L., Liu, D., & Li, H. (2024). Vnvc: A versatile neural video coding framework for efficient human-machine vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

5. Stankiewicz, O., Grajek, T., Maćkowiak, S., Stankowski, J., Rózek, S., Lorkiewicz, M., ... & Domański, M. (2024, July). Region-of-Interest-Based Video Coding for Machines. Στο *2024 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)* (σελ. 1-6). IEEE.
6. Li, H., Zhang, X., Wang, S., Wang, S., & Pan, J. (2024). Human-Machine Collaborative Image and Video Compression: A Survey. *APSIPA Transactions on Signal and Information Processing*, 13(6), 1-15.
7. Xu, X., Liu, S., Ma, S., Wang, S., & Gao, W. (2021). Machine vision oriented video coding: A survey and case studies. *Journal of Visual Communication and Image Representation*, 77, 103117.
8. Lin, J., Zhang, Z., Chen, B., & Zhao, T. (2022). End-to-end feature-preserving video compression for machines. *arXiv preprint arXiv:2210.12345*.