

KOMPRESJA WIZJI DLA MASZYN

VIDEO COMPRESSION FOR MACHINES

prof. dr hab. inż. Marek Domański, Politechnika Poznańska, Instytut Telekomunikacji Multimedialnej
(marek.domanski@put.poznan.pl)

STRESZCZENIE: Metody kompresji wizji są opracowywane tak, by przy zadanej małej bitowej szybkości transmisji dawało się w odbiornikach uzyskiwać możliwie dobrą jakość zdekodowanych obrazów. Dotychczas miarą tej jakości była uśredniona ocena grupy widzów. Jednak w ostatniej dekadzie coraz częstsze są zastosowania, w których obraz jest poddawany kompresji przed wykonaniem różnych operacji widzenia maszynowego. Powstaje więc pytanie, czy dotychczasowe metody kompresji zoptymalizowane pod względem subiektywnie ocenianej przez ludzi jakości zdekodowanych obrazów są także optymalne dla kompresji wizji przeznaczanej do wykorzystania w systemach widzenia maszynowego. Odpowiedź na to pytanie okazuje się być negatywna i dlatego w ostatnim dziesięcioleciu prowadzi się intensywne badania nad nowymi metodami kompresji odpowiednimi dla wizji używanej przez systemy widzenia maszynowego, czyli nad metodami kompresji wizji dla maszyn (*video compression/coding for machines*).

Praca skrótowo opisuje rezultaty badań dotyczących metod kompresji wizji dla maszyn. W szczególności podano najważniejsze cechy kodeka MPEG MVC, który został ujęty w projekcie części drugiej normy ISO/IEC MPEG-AI.

SŁOWA KLUCZOWE: kompresja wizji, widzenie maszynowe, MPEG, normy

ABSTRACT: Until recently, video coding standards were developed from the perspective of a human observer; that is, the main goal was to achieve a high subjective quality of the decoded video. This quality was measured by mean opinion score. However, over the past decade, the development of semi- and fully autonomous vehicles, AI-based video analysis, intelligent video surveillance, and video-based control in many new areas has led to a sharp increase in the amount of video data shared between computers. In these cases, direct human consumption of the decoded video is not the primary application. These factors have sparked research interest in video coding where the decoded data serves as input for machine vision tasks.

This paper briefly summarizes the results of research on video coding for machines. Specifically, it briefly describes the upcoming MPEG standard for machine video coding as part two of the ISO/IEC MPEG-AI standard.

KEYWORDS: video compression, machine vision, MPEG, standards

Wynalazek efektywnej kompresji wizji (obrazów ruchomych, ang. video) zrewolucjonizował technikę i umożliwił powstanie wielu nieznanych wcześniej usług i systemów, począwszy od telewizji cyfrowej poprzez cyfrowy zapis filmów, skomplikowane systemy dozoru wizyjnego i strumieniowanie wizji w internecie, aż po systemy wirtualnej rzeczywistości. Dotychczas opracowano wiele metod kompresji wizji: MPEG-2 [1,2], AVC (H.264) [3,4], HEVC (H.265) [5,6], VVC (H.266) [7,8] i inne. Wspomniane metody były opracowywane tak, by przy pewnej bitowej szybkości transmisji dawało się w odbiornikach uzyskiwać zdekodowaną wizję o możliwie dobrej jakości postrzeganej przez widzów. Miarą tej jakości była uśredniona ocena uzyskiwana od grupy widzów. Takie podejście do kompresji wizji odpowiadało potrzebom telewizji i innych usług wspomnianych wyżej.

W ostatniej dekadzie coraz częstsze są jednak zastosowania, w których obraz jest poddawany kompresji przed wykonaniem różnych operacji widzenia maszynowego. Powstaje więc pytanie, czy dotychczasowe metody kompresji zoptymalizowane pod względem subiektywnie ocenianej przez ludzi jakości zdekodowanej wizji są także optymalne dla kompresji wizji przeznaczanej do wykorzystania w systemach widzenia maszynowego. Odpowiedź na to pytanie okazuje się być negatywna i dlatego od kilkunastu lat prowadzi się badania nad nowymi metodami kompresji odpowiednimi dla wizji używanej przez systemy widzenia maszynowego, czyli nad metodami kompresji wizji dla maszyn (ang. *Video Compression/Coding for Machines* – VCM).

W tej pracy skrótowo zostaną omówione podstawowe zagadnienia związane z metodami kompresji wizji dla maszyn i ich normalizacją w Międzynarodowej Organizacji Normalizacyjnej (ISO).

PRZYKŁADY ZASTOSOWAŃ KOMPRESJI WIZJI DLA MASZYN

Grupa ekspertów MPEG (*Moving Pictures Experts Group*) działająca w ramach ISO/IEC JTC1/SC29 przygotowanie norm dotyczących kompresji wizji dla maszyn poprzedziła analizą potencjalnych zastosowań takiej kompresji [9]. Wstępna analiza potencjalnych rozwiązań doprowadziła do wniosku, że istnieją dwie podstawowe konfiguracje systemów związanych z kompresją wizji dla maszyn. W systemach pierwszej grupy (rys. 1) wizja jest poddawana kompresji, a w dekodерze odtwarza się wizję stanowiącą obiekt działania systemów widzenia maszynowego. Równolegle z analizą maszynową zdekodowanej wizji możliwe jest też jej oglądanie przez ludzi.

W systemach drugiej grupy już w koderze przeprowadza się analizę wizji prowadzącą do ekstrakcji cech, a właściwej kompresji poddaje się niekoniecznie całą wizję, ale często tylko wyznaczone cechy. Zdekodowane cechy są po stronie odbiorczej przekazywane do systemu widzenia maszynowego (rys. 2). Techniki odpowiadające drugiej grupie są na razie mniej rozwinięte i dlatego w dalszej części artykułu nasza uwaga będzie skupiona na systemach pierwszej grupy. Przeprowadzona analiza [9] wykazała bardzo dużą liczbę bardzo różnorodnych potencjalnych zastosowań. Jako potencjalne obszary zastosowań wymieniono:

- systemy dozoru wizyjnego,
- inteligentne systemy transportowe, w tym komunikacja V2V (*vehicle to vehicle*) oraz I2V (*infrastructure to vehicle*),
- inteligentne miasta,
- inteligentny przemysł,
- inteligentne treści wizualne,
- elektronika powszechnego użytku (ang. *consumer electronics*).

Z wymienionymi przykładowymi obszarami powiązano typowe zadania widzenia maszynowego realizowane po stronie odbiorczej systemu, w tym między innymi:

- wykrywanie obiektów powiązane z określaniem klasy obiektu, identyfikatora obiektu i jego lokalizacji,
- segmentację obrazów,
- poprawę jakości obrazów,
- wyszukiwanie obrazów,
- wyszukiwanie i rozpoznawanie zdarzeń,
- wykrywanie anomalii,
- przetwarzanie wizji wielowidokowej.

Naszkicowana wyżej różnorodność zastosowań i zadań okazała się istotną trudnością w rozwoju technik kodowania wizji dla maszyn, w tym dla procesów normalizacji w tym zakresie.

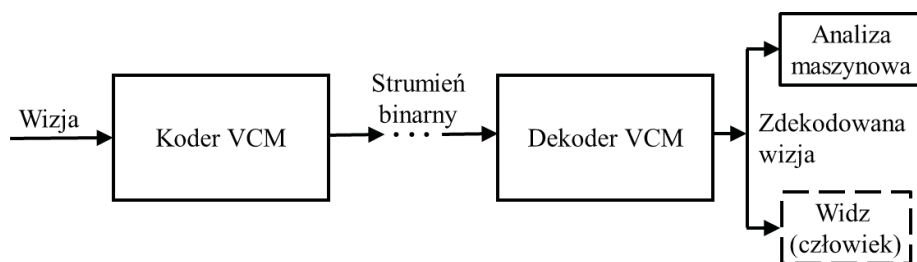
OCENA EFEKTYWNOŚCI METOD KOMPRESJI DLA MASZYN

Klasyczne metody kodowania wizji są oceniane przede wszystkim pod kątem efektywności kompresji rozumianej jako uzyskanie możliwie wysokiej jakości obrazu zdekodowanego ze strumienia binarnego o możliwie małej szybkości bitowej. Jakość mierzy się przede wszystkim subiektywną miarą wyznaczaną średnią ocen grupy obserwatorów (ang. *Mean Opinion Score* – MOS). Ze względu na uciążliwość i wysokie koszty takiej oceny jakości powszechnie stosuje się prostsze do wyznaczenia miary takie, jak PSNR (ang. *Peak Signal To Noise Ratio*) lub bardziej zaawansowane miary jak SSIM (*Structural Similarity Index Measure*) [10] oraz VMAF (*Video Multimethod Assessment Fusion*) [11].

Nawet przy ograniczeniu analizy do układu z rys. 1, zdefiniowanie metod oceny efektywności kompresji dla maszyn stwarza istotne trudności wynikające ze zróżnicowania metod przetwarzania zdekodowanej wizji. W grupie ekspertów MPEG przeprowadzono długie dyskusje, które ostatecznie doprowadziły do kompromisowych ustaleń mówiących, że podczas wyboru metod stanowiących podstawę przyszłej normy badana będzie efektywność kompresji tylko dla detekcji obiektów, segmentacji obiektów oraz śledzenia obiektów [12].

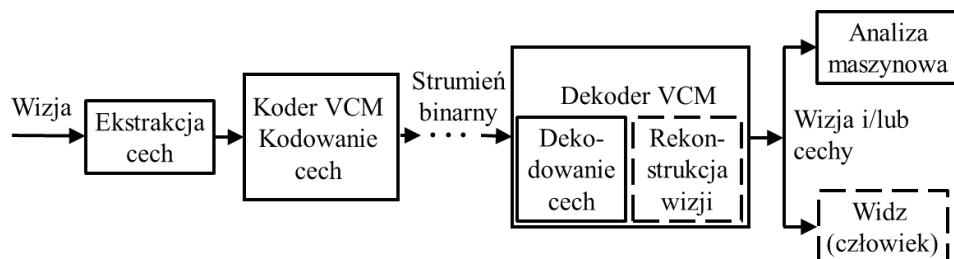
Dla detekcji i segmentacji obiektów przyjęto uśrednioną przeciętną precyzję mAP (*Mean Average Precision*) [12].

Dla danej klasy obiektów liczby decyzji prawdziwie pozytywnych (ang. *true positive*) oraz fałszywie pozytywnych (ang. *false positive*) są określone w odniesieniu do wielkości części wspólnych obiektów IoU (ang. *Intersection over Union*), przekraczających wartość progową T_{IoU} określoną



» Rys. 1. Schemat blokowy systemu, w którym zadania widzenia maszynowego są wykonywane na zdekodowanej wizji

» Fig. 1. Block scheme of a system using decoded video for machine vision tasks



» Rys. 2. Schemat blokowy systemu, w którym zadania widzenia maszynowego są wykonywane na zdekodowanych cechach

» Fig. 2. Block scheme of a system using decoded features for machine vision tasks

dla tej klasy obiektów. Wspomniane wartości prawdziwe i fałszywe dotyczą wyjścia sieci neuronowej służącej do analizy wizji zdekodowanej w odbiorniku.

Wartość precyzji określa się jako:

$$\text{precision}(T_{IoU}) = \frac{TP(T_{IoU})}{TP(T_{IoU}) + FP(T_{IoU})} \quad (1)$$

Dla poszczególnych klas obiektów wyznacza się przeciętne wartości precyzji *AP* (ang. *Average Precision*). Uśredniona przeciętna precyzja *mAP* (ang. *Mean Average Precision*) jest wartością *AP* uśrednioną po wszystkich klasach obiektów w zakresie ustalonych wartości progowych dla IoU.

Dla pomiarów efektywności śledzenia obiektów przyjęto [12] dokładność śledzenia wielu obiektów *MOTA* (ang. *Multiple Object Tracking Accuracy*) [13].

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + mme_t)}{\sum_t g_t} \quad (2)$$

gdzie FN_t , FP_t , mme_t and g_t są odpowiednio liczbami decyzji fałszywie negatywnych oraz fałszywie pozytywnych, a także liczbą błędów niedopasowania i rzeczywistą liczbą obiektów w chwili.

W grupie ekspertów MPEG podczas prac zmierzających do ustanowienia normy dla systemów zgodnych ze schematem z rys. 1 przyjęto, że wartości *mAP* i *MOTA* będą wyznaczane dla ujednoliconych implementacji systemów widzenia maszynowego przetwarzających zdekodowaną wizję w sztucznych sieciach neuronowych opartych na strukturze sieci Detectron [14, 15].

Oczywiście wszystkie badania efektywności kompresji wymagały zdefiniowania zbiorów obrazów i sekwencji wizyjnych, co zostało przedstawione na przykład w dokumencie [12].

ROZWÓJ METOD KOMPRESJI WIZJI DLA MASZYN

W ostatnim dwudziestoleciu nastąpił szybki rozwój internetu rzeczy, analizy wizji wykorzystującej sztuczną inteligencję, inteligentnych systemów dozoru wizyjnego oraz inteligentnych systemów sterownia. Stymulowało to gwałtowny wzrost ilości danych, w tym danych wizyjnych, wymienianych pomiędzy komputerami. Wymienione zjawiska pobudziły rozwój badań dotyczących kompresji wizji w zastosowaniach, w których obserwacja zdekodowanej wizji przez człowieka nie jest już najważniejszym zadaniem. W ostatnim dwudziestoleciu zostało zaproponowanych wiele odpowiednich metod. Wśród pionierskich metod można wymienić te przedstawione na przykład w pracach [16-21].

W drugiej dekadzie XXI wieku powszechnie uznano znaczenie problemu kompresji wizji dla maszyn, a termin „kodowanie/kompresja wizji dla maszyn” wszedł do szerokiego stosowania. Ponieważ telekomunikacja wizualna zawsze wymaga normalizacji, szybko dostrzeżono tę potrzebę także dla kompresji wizji dla maszyn. Ta potrzeba została też dostrzeżona przez grupę ekspertów MPEG, która już w roku 2019 utworzyła „grupę ad hoc” o nazwie „Video Coding for Machines”, której przewodniczącymi początkowo byli Yang

Zhang (China Telecom) oraz Patric Dong (*Gerfalcon Technologies*) [22]. Zadaniem wspomnianego zespołu było przygotowanie materiałów i metod niezbędnych do dalszych prac, prowadzących do ustalenia techniki, która mogłaby zostać ujęta w przyszłej normie kompresji wizji dla maszyn. Przyjęto, że najpierw zostaną podjęte badania związane z systemami z rys. 1, a dopiero później prace dotyczące systemów z rys. 2.

W roku 2022 MPEG ogłosił wezwanie do zgłaszania propozycji technik kompresji wizji dla maszyn (*Call for Proposals on Video Coding for Machines*) [23, 24]. Wezwanie było ograniczone do propozycji systemów o strukturze pokazanej na rys. 1. W odpowiedzi na to wezwanie wpłynęło 14 propozycji pochodzących od różnych zespołów badawczych z całego świata, w tym od zespołu z Politechniki Poznańskiej kierowanego przez prof. Marka Domańskiego i dr. hab. Olgierda Stankiewicza. Nadesłane propozycje były dosyć zróżnicowane pod względem technicznym, jednak większość z nich wykorzystywała najbardziej obecnie zaawansowaną ogólną technikę kompresji wizji, czyli technikę VVC [7, 8]. Również propozycja Politechniki Poznańskiej wykorzystywała technikę VVC uzupełnioną o różne nowe narzędzia opracowane w Poznaniu.

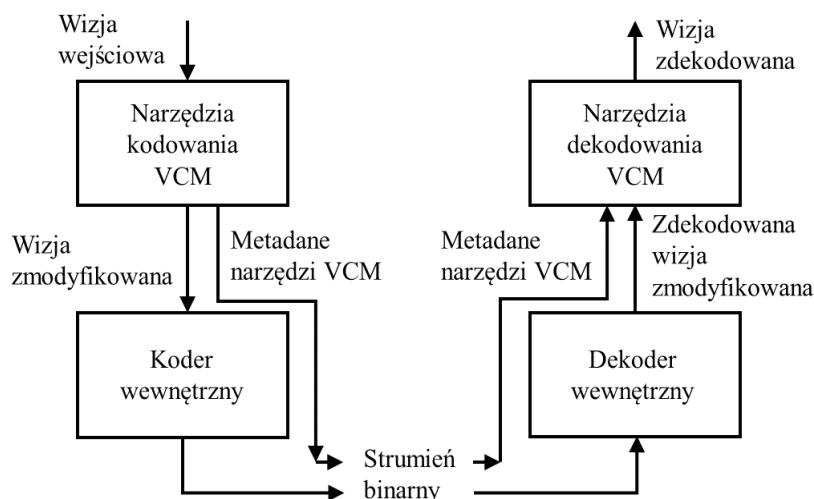
Ponieważ żadna ze zgłoszonych technik nie wyróżniła się zdecydowanie, postanowiono, że do eksperymentów będzie się używać oprogramowania modelowego zawierającego techniki zaproponowane przez różne ośrodki, w tym te pochodzące z Politechniki Poznańskiej. Oprogramowanie modelowe realizuje technikę, która w najbliższej przyszłości ma stanowić normę, co może nastąpić po pozytywnej ocenie w organizacji ISO.

Od pewnego czasu podgrupa zajmująca się przyszłą normą kompresji wizji dla maszyn jest kierowana przez Shan Liu (Tencent) oraz Chrisa Rosewarne (Canon). Koordynatorem oprogramowania VCM-RS (*Video Coding for Machines – Reference Software*), a więc zarządzającym konfiguracjami różnych technik jest Olgierd Stankiewicz (Politechnika Poznańska). Obecnie dla kompresji wizji dla maszyn (według schematu z rys. 1) gotowy jest projekt tekstu normy [25], który wkrótce ma się stać drugą częścią międzynarodowej normy MPEG-AI (ISO/IEC 23888). Ta technika bardzo skrótowo zostanie omówiona w następnej części pracy.

Prace badawcze mające na celu ustanowienie norm dla technik nadających się do kodowania wizji dla maszyn są podejmowane także poza grupą MPEG. Związany z grupą MPEG zespół JVET (*Joint Video Experts Team*) stanowiący wspólną grupę roboczą ISO i ITU (Międzynarodowy Związek Telekomunikacyjny) prowadzi prace mające na celu uwzględnienie kodowania wizji dla maszyn w przyszłej ogólnej normie kompresji wizji, która ma wejść w życie za kilka lat i oferować szybkości bitowe dwukrotnie mniejsze niż kodeki zgodne z obecnie najnowszą normą ISO/IEC/ITU, czyli VVC.

Także działająca w ramach ISO grupa robocza JPEG (ISO/IEC JTC1/SC29/WG1) prowadzi prace dotyczące kompresji obrazów statycznych dla maszyn.

Równolegle w grupie MPEG są prowadzone badania mające doprowadzić do opracowania techniki kompre-



» Rys. 3. Schemat blokowy kodeka MPEG VCM

» Fig. 3. Block scheme of MPEG VCM codec

sji działającej w systemach pokazanych na schemacie z rys. 2. W roku 2023 ogłoszono wezwanie do zgłaszania propozycji odpowiednich technik [26]. W odpowiedzi uzyskano pewną liczbę propozycji technicznych i obecnie trwa faza współpracy mająca zakończyć się zaproponowaniem odpowiedniej normy, dla której już opracowano wstępne szkice tekstu [27].

KODEK MPEG VCM

Jak już wspomniano, w grupie ekspertów MPEG przeprowadzono w latach 2022-25 badania mające na celu opracowanie techniki kompresji wizji dla maszyn. Dla systemów działających w strukturze z rys. 1 ten etap został już zasadzie zakończony publikacją projektu normy [25], która będzie określać składnię i sposób dekodowania strumienia binarnego. Techniczny opis kodeka można znaleźć w dokumencie [28].

Architektura kodeka MPEG MVC (rys. 3) wykorzystuje „kodek wewnętrzny”, który w zasadzie może być dowolnym kodekiem, nawet takim, który dopiero zostanie opracowany w przyszłości. W trakcie badań realizowanych przez grupę MPEG wykorzystywano kodek VVC [7, 8].

Zaznaczone na rys. 3 narzędzia kodowania MVC modyfikują kodowaną wizję w taki sposób, by minimalizować szybkość bitową transmisji na wyjściu koder wewnętrzny, ale by jednocześnie zapewnić dobrą efektywność zadań widzenia maszynowego wykonywanych z wykorzystaniem zdekodowanej wizji uzyskiwanej z narzędzi dekodera VCM.

Narzędzia koder VCM obejmują między innymi:

- próbkowanie w dziedzinie czasu,
- próbkowanie przestrzenne i retargeting,
- upraszczanie tła,
- przetwarzanie w regionach zainteresowania,
- dekoloryzację.

W dekodерze przeprowadza się operacje analogiczne do operacji w koderze.

Na koniec można podkreślić bardzo aktywny udział grupy z Politechniki Poznańskiej w opracowaniu tego kodeka. Ten udział dotyczył przede wszystkim upraszczania tła [29], narzędzi przetwarzania w regionach zainteresowania [30] oraz koloryzacji [31].

PODSUMOWANIE

Przedstawiony przegląd wspiera tezę, że kompresja wizji dla maszyn stanowi bardzo ważny obszar badań. Należy się spodziewać, że wyniki badań będą w coraz większym stopniu wykorzystywane w rozmaitych usługach i w przemyśle. W ciągu następnych niewielu lat tematyka kompresji wizji dla maszyn prawdopodobnie będzie musiała znaleźć swoje miejsce w nauczaniu przyszłych inżynierów telekomunikacji, teleinformatyki i informatyki.

LITERATURA

- [1] ISO/IEC 13818-1 | ITU-T Rec. H.262, Generic coding of moving pictures and associated audio information: Video.
- [2] Mitchell J., Pennebaker W., Fogg Ch., LeGall D.: *MPEG video compression standard*, Kluwer, Dordrecht, 1996.
- [3] ISO/IEC IS 14496-10 | ITU-T Rec. H.264, Advanced video coding.
- [4] Richardson I.: *The H.264 advanced video compression standard*, 2nd ed., John Wiley & Sons, Chichester, 2010.
- [5] ISO/IEC IS 23008-2 | ITU-T Rec. H.265, High efficiency video coding.
- [6] Wien M.: *High Efficiency Video Coding – Coding Tools and Specification*, Springer, Berlin, Heidelberg, 2014.
- [7] ISO/IEC IS 23090-3 / ITU-T Recommendation H.266, Versatile video coding.
- [8] Saldanha M., Sanchez G., Marcon C., Agostini L., *Versatile Video Coding (VVC)*, Springer, Berlin, Heidelberg, 2022.
- [9] ISO/IEC JTC1/SC29/WG2, Doc. N161, Use cases and requirements for Video Coding for Machines, January 2022.
- [10] Wang Zhou, Bovik, A. C., Sheikh, H. R., Simoncelli, E. P., *Image quality assessment: from error visibility to structural similarity*, in IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, April 2004.
- [11] Rassool, R., VMAF reproducibility: *Validating a perceptual practical video quality metric*, in Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB), Cagliari, Italy, Jun. 2017, pp. 1–2.
- [12] ISO/IEC JTC1/SC29/WG2, Doc. N379, Common Test Conditions for Video Coding for Machines, Geneva, July 2023.
- [13] Bernardin, K., Elbs, A., Stiefelhofen, A., *Multiple Object Tracking Performance Metrics and Evaluation in a Smart Room Environment*, in Sixth IEEE International Workshop on Visual Surveillance in Conjunction with ECCV, May 2008.
- [14] Wang, Z., Zheng, L. et al., *Towards real-time multi-object tracking*, in European Conf. Computer Vision (ECCV), 2020, pp. 107-122.
- [15] Wu, Y., Kirillov, A., Massa, F., et al. *Detectron2*, <https://github.com/facebookresearch/detectron2>.
- [16] Duan, L., Liu, J., Yang, W., Huang, T., Gao, W., *Video Coding for Machines: A Paradigm of Collaborative Compression and Intelligent*

- Analytics*, in IEEE Transactions on Image Processing, vol. 29, pp. 8680-8695, 2020.
- [17] Fischer, K., Brand, F., Herglotz, C., Kaup, A., *Video Coding for Machines with Feature-Based Rate-Distortion Optimization*, 2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP), 2020.
- [18] Galteri, L., Bertini, M., Seidenari, L., Del Bimbo, A., *Video compression for object detection algorithms*, in Proc. International Conference on Pattern Recognition (ICPR), Aug. 2018, pp. 3007–3012.
- [19] Choi, H., Bajic, I. V., *High efficiency compression for object detection*, in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Apr. 2018, pp. 1792–1796.
- [20] Chao, J., Steinbach, E., *Keypoint encoding for improved feature extraction from compressed video at low bitrates*, in IEEE Trans. Multimedia, vol. 18, no. 1, pp. 25–39, Jan. 2016.
- [21] Ma, S., Zhang, X., Wang, S., Zhang, X., Jia, C., Wang, S., *Joint feature and texture coding: Toward smart video representation via frontend intelligence*, IEEE Trans. Circuits Syst. Video Technol., vol. 29, no. 10, pp. 3095–3105, Oct. 2019.
- [22] ISO/IEC JTC1/SC29/WG11, Doc. N1850, Conclusions of 127th meeting, Gothenburg, July 2019.
- [23] ISO/IEC JTC1/SC29/WG2, Doc. N191, Call for Proposals on Video Coding for Machines, April 2022.
- [24] ISO/IEC JTC1/SC29/WG2, Doc. N193, Evaluation Framework for Video Coding for Machines, April 2022.
- [25] ISO/IEC JTC1/SC29/WG4, Doc. N662, Text of ISO/IEC CD 23888-2 Video coding for machines, May 2025.
- [26] ISO/IEC JTC1/SC29/WG2, Doc. N282, Call for Proposals on Feature Compression for Video Coding for Machines, April 2023.
- [27] ISO/IEC JTC1/SC29/WG4, Doc. N669, PWD of feature coding for machines, May 2025.
- [28] ISO/IEC JTC1/SC29/WG4, Doc. N664, Algorithm description of tools in VCM reference software, June 2025.
- [29] Rózek, S., Stankiewicz, O., Maćkowiak, S., Domański, M., *Video coding for machines using object analysis and standard video codecs*, in 2023 IEEE International Conference on Visual Communications and Image Processing (VCIP), Jeju, Republic of Korea, December 2023.
- [30] Stankiewicz, O., Grajek, T., Maćkowiak, S., Stankowski, J., Rózek, S., Lorkiewicz, M., Wawrzyniak, M., Domański, M., *Region-of-interest-based video coding for machines*, in 2024 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), Niagara Falls, ON, Canada, 2024.
- [31] Lorkiewicz, M., Rózek, S., Stankiewicz, O., Grajek, T., Maćkowiak, S., Domański, M., *Video coding for machines with neural-network-based chroma synthesis*, in IEEE Access, Early access, 2025.