

# 글로벌 ICT 표준 컨퍼런스 2023

Global ICT Standards Conference 2023

(세션2) AI: 새로운 가능성을 열어가는 혁신

## AI를 위한 MPEG 부호화 표준 기술

정순흥 책임연구원, ETRI

주최



과학기술정보통신부  
Ministry of Science and ICT



특허청  
Korean Intellectual  
Property Office

주관



국립전파연구원  
National Radio Research Agency



IITP

KEA

kista

ETRI

# Index

01 MPEG 소개

02 MPEG-AI 개요

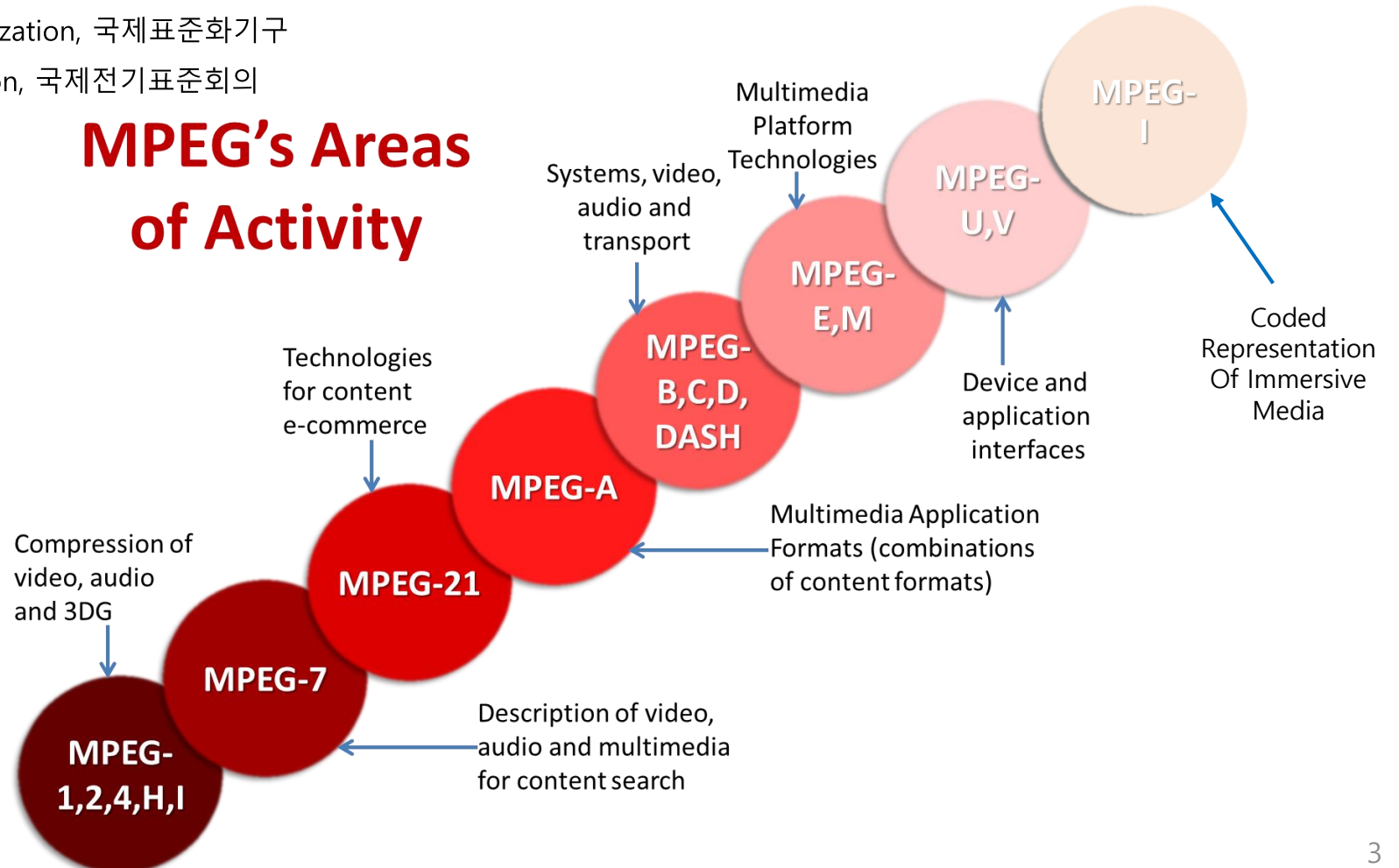
03 Video Coding for Machines 표준화 동향

## 01. MPEG 소개

### ISO/IEC JTC 1/SC 29 WG2~WG8에서 표준화 활동 수행

- ISO: International Organization for Standardization, 국제표준화기구
- IEC: International Electrotechnical Commission, 국제전기표준회의
- JTC 1: Joint Technical Committee 1, 정보기술
- SC 29: Subcommittee 29, 멀티미디어 부호화
- WG: Working Group

### MPEG's Areas of Activity



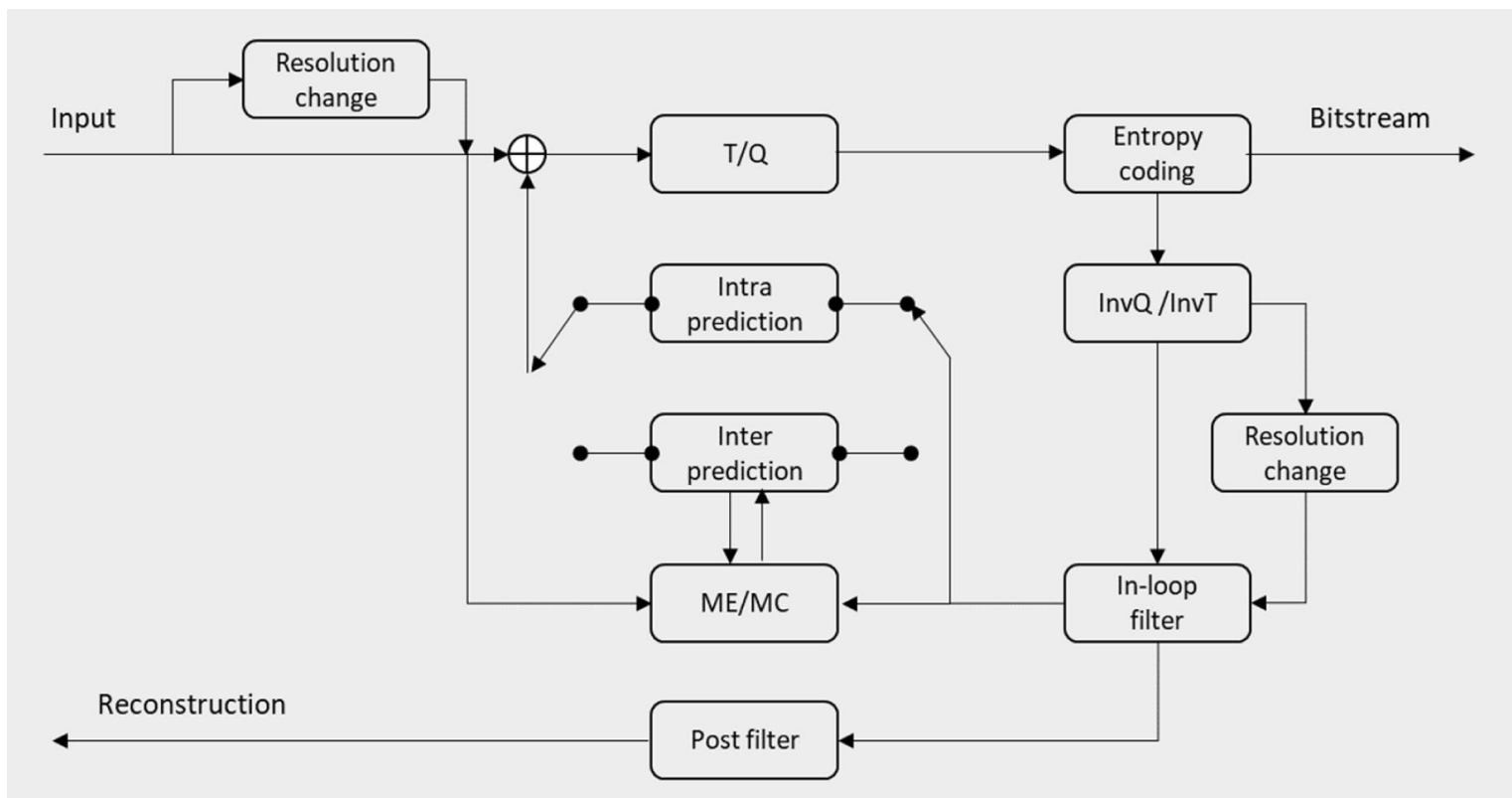
## 02. MPEG-AI 개요

AI를 위한 새로운 MPEG 프로젝트 수립 필요성 제기 – 2022.10.

- Draft TR. Artificial intelligence for multimedia. Part 1: Vision and scenarios – 2023.1. ~
- Two aspects of interaction between AI and multimedia
  1. AI as multimedia coding tool
  2. Multimedia for consumption by AI

## 02. MPEG-AI 개요

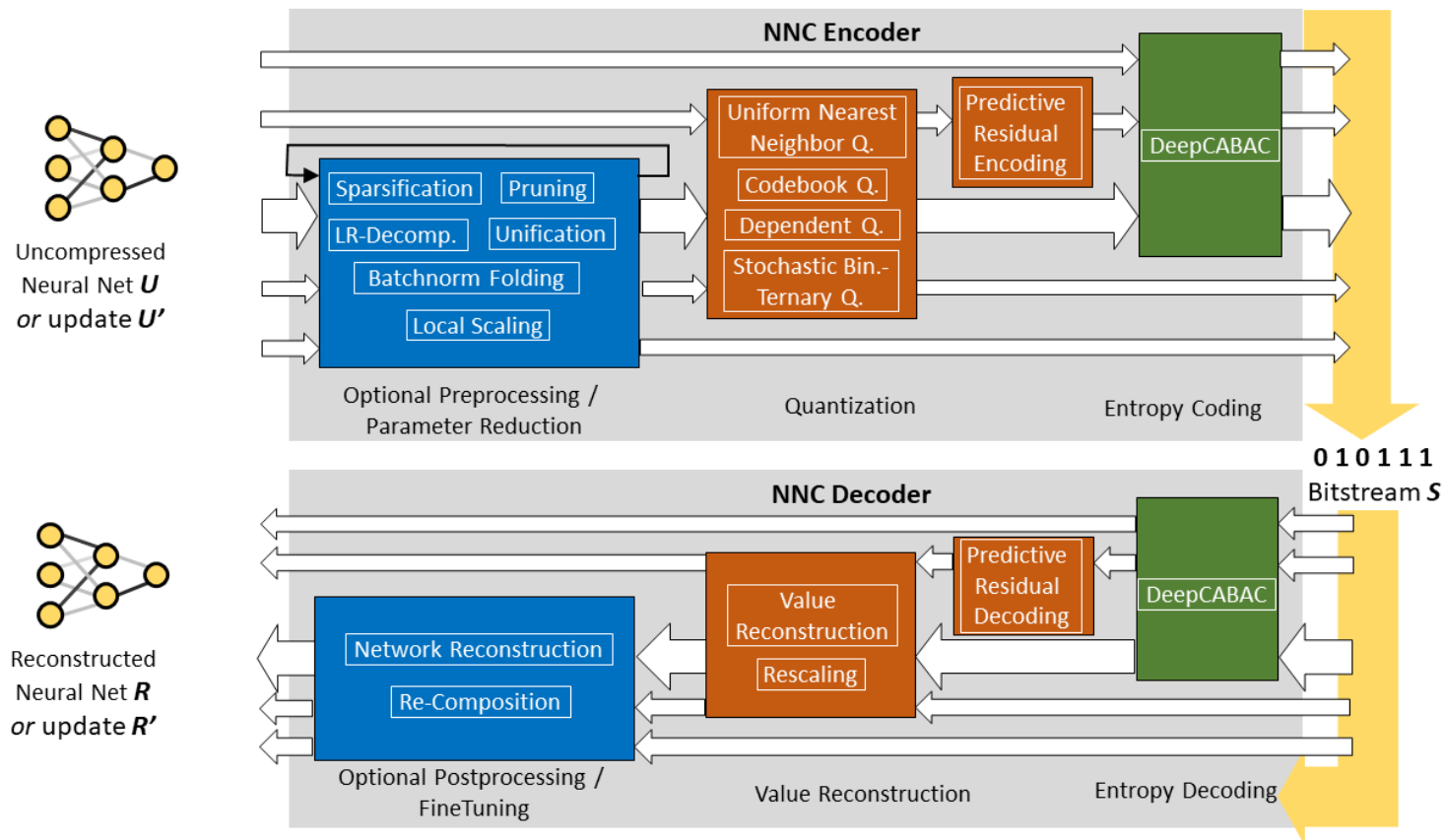
### NNVC: Neural Network-based Video Coding



하이브리드 비디오 코딩 프레임워크에서 신경망 기반 기술 개발을 위한 기능 예시

## 02. MPEG-AI 개요

### NNC: Neural Network Coding



NNC edition 2 Overview

## 02. MPEG-AI 개요

---

- AI based 3D Graphics Coding
- **Video Coding for Machines**
- **Feature Coding for Machines**
- Audio Coding for Machines

## 03. Video Coding for Machines 표준화 동향

### 필요성 및 서비스 시나리오 제시 – 2019.7.

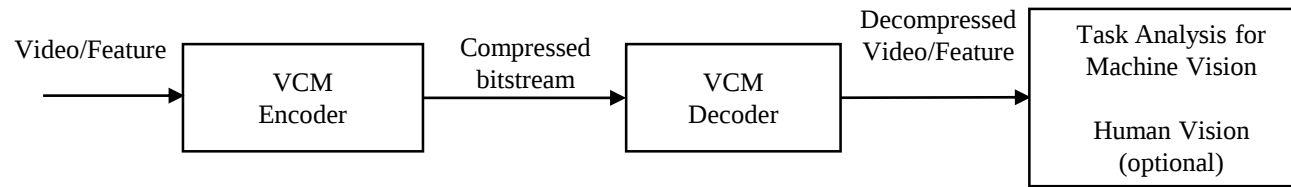
- 인공지능 기술의 발전으로 기계간 폭증하는 트래픽을 효율적으로 처리(압축)할 수 있는 표준 필요

| item | Machine-oriented Analysis Use Cases         | Description                                                                                                                 | Techniques                                                        |
|------|---------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|
| 1    | Smart Glasses                               | Record daily activity log, Navigation (indoor/outdoor w/o GPS), Video recording wake up                                     | Object detection, Segmentation                                    |
| 2    | Unmanned store                              | tracking customer activity, check items in shopping cart                                                                    | Object detection, Pose estimation, Object tracking                |
| 3    | Unmanned Warehouse/Store Robot              | Robot navigation, stocking, inventory checking                                                                              | Detection, Segmentation, Classification                           |
| 4    | Smart Retailer                              | Shopping Center Customer Group Analysis, Detect hot spot inside store by customer age, gender. Customer Traffic information | Detection, Heat map, Activity analysis.                           |
| 5    | Industrial Production Line Detect Equipment | Detect failure elements in production line for quality control                                                              | Detection, Tracking                                               |
| 6    | Smart factory /Automatic Machinery          | Detect appearance of object, pick up and transfer to destination                                                            | Detection, Segmentation, Key point detection                      |
| 7    | Smart fishery/agriculture                   | Detect diseases                                                                                                             | Detection, Classification                                         |
| 8    | Drone                                       | Real time environment monitoring and automatic collision avoidance                                                          | Detection, Segmentation, tracking                                 |
| 9    | Automation ADAS, Autonomous Vehicles        | from CDVA use case list, details refer to [3]                                                                               | Object detection, Segmentation, Object tracking, Pose estimation. |

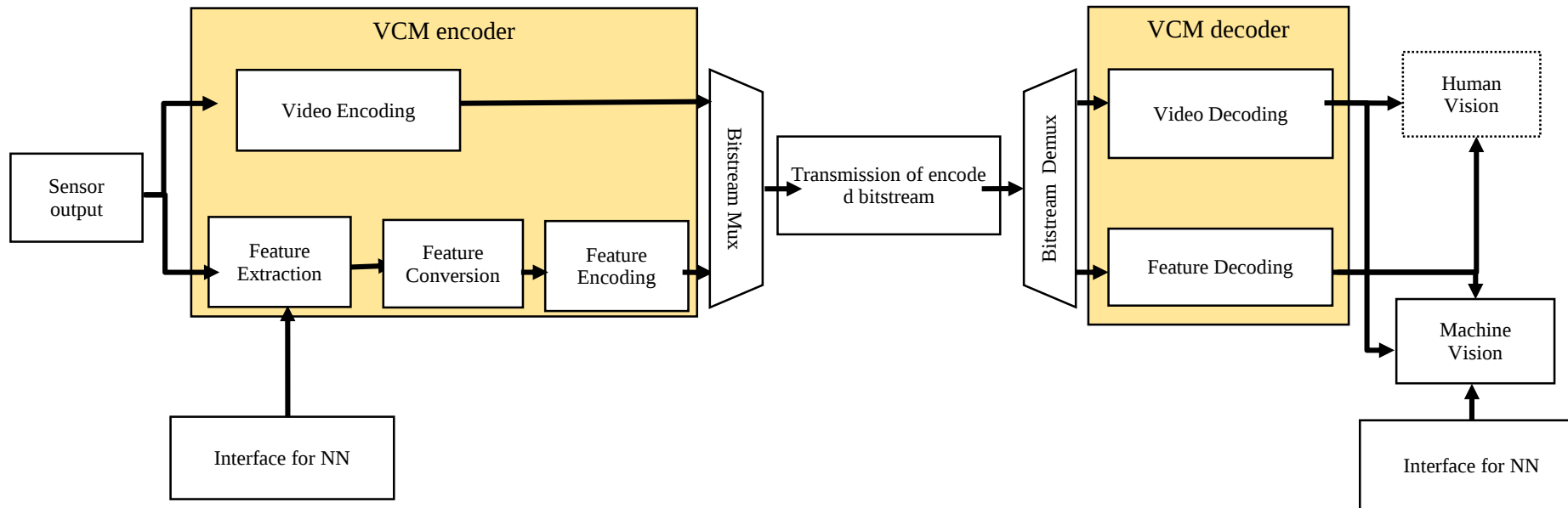
| item | Combined Machine and Human representation Use Cases | Description                                                                                                                                                                                                              | Techniques                                         |
|------|-----------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------|
| 1    | AR and Video Game Goggles                           | Capture live video and detect environment elements                                                                                                                                                                       | Detection, Segmentation, Pose Estimation, Tracking |
| 2    | Sports Game animation                               | From live game video, create animation                                                                                                                                                                                   | Detection, Segmentation, tracking                  |
| 3    | Surveillance                                        | from CDVA use case list, details refer to [3][4], Traffic monitoring, People counting, Crowd counting, Extraction of people flow direction and speed<br>Target Detection, Event Recognition, Intelligent video retrieval | Detection, Heat map, Tracking, classification      |

## 03. Video Coding for Machines 표준화 동향

코덱 구조, 서비스 시나리오, 요구사항, 공통 실험 조건, 성능 평가 방법 등 논의 - 2019.10.~



Pipeline for VCM

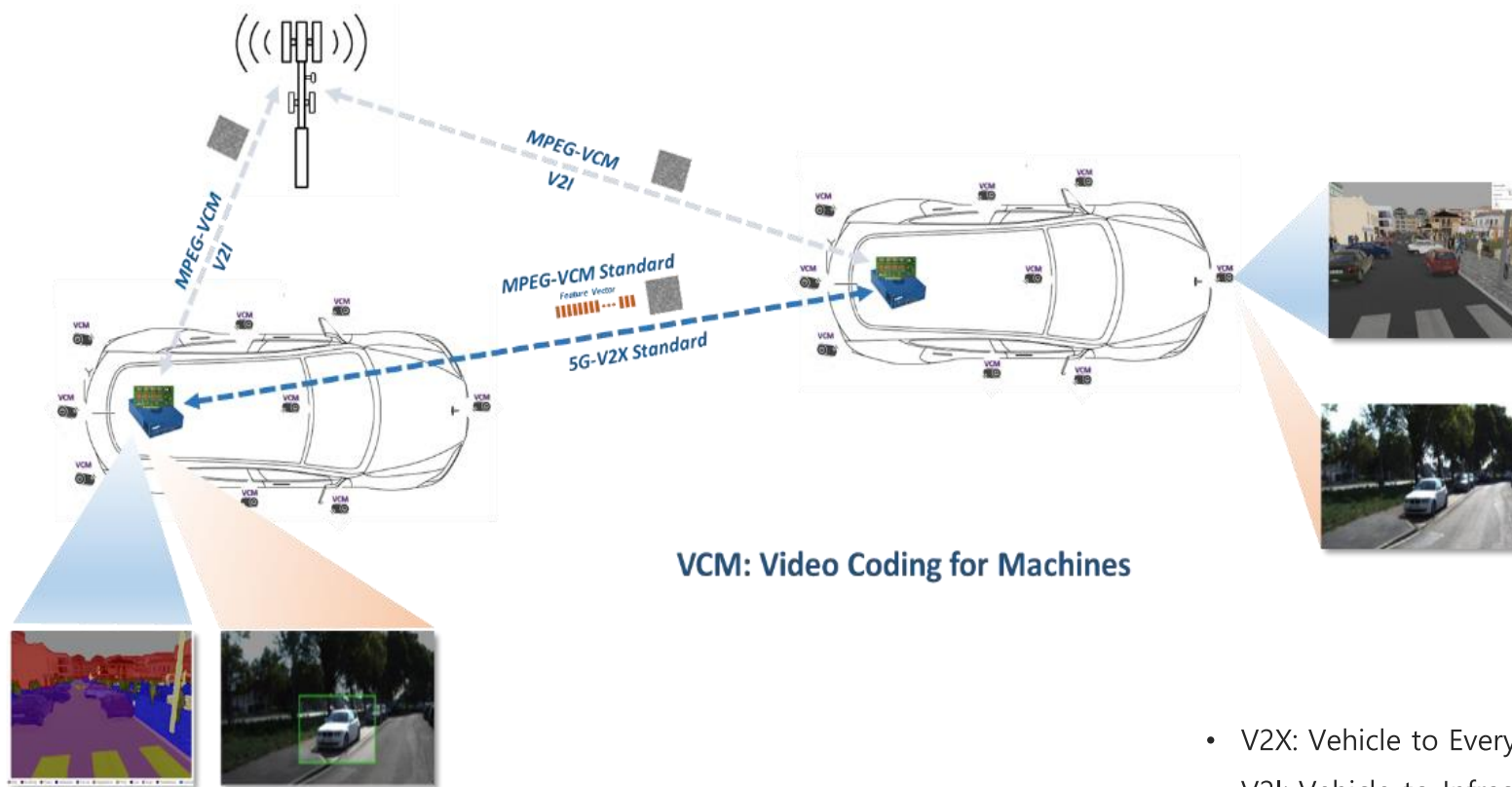


An example of potential VCM architecture

## 03. Video Coding for Machines 표준화 동향

### 서비스 시나리오의 예

- Cooperative V2X-VCM

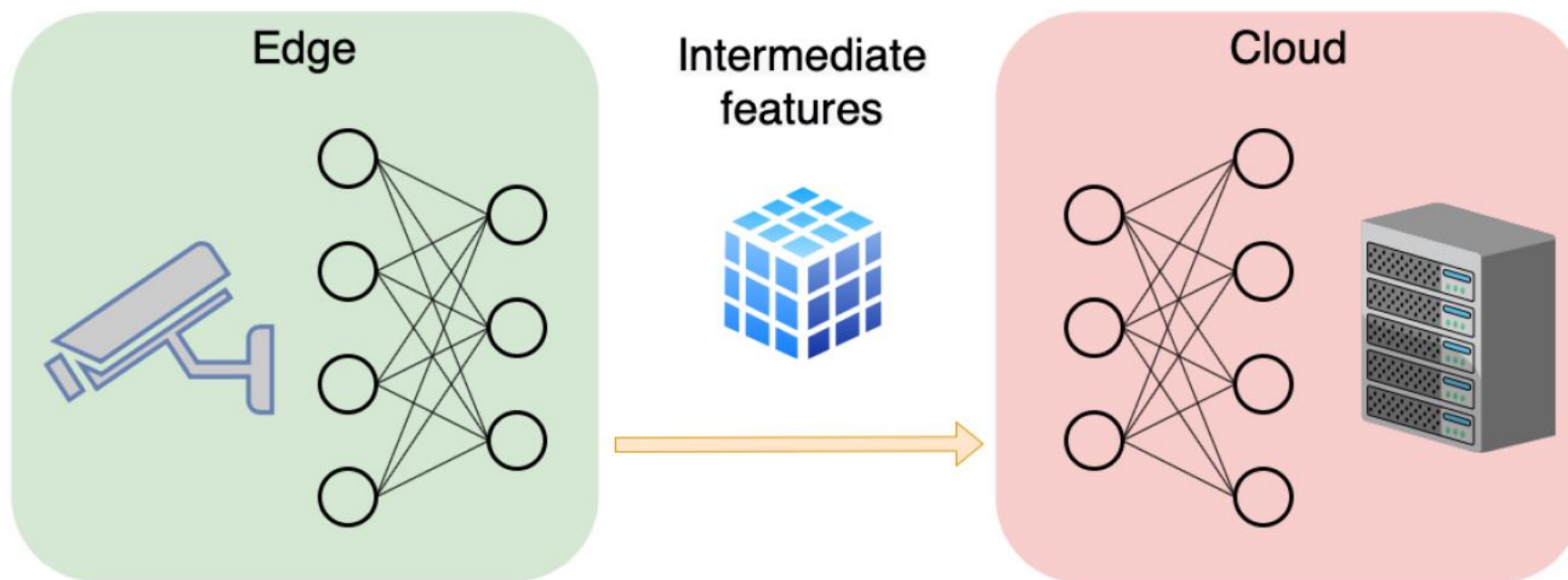


- V2X: Vehicle to Everything
- V2I: Vehicle-to-Infrastructure

## 03. Video Coding for Machines 표준화 동향

### 서비스 시나리오의 예

- Collaborative intelligence: A framework in which AI inference is shared between the edge devices and the cloud



Basic collaborative intelligence system

## 03. Video Coding for Machines 표준화 동향

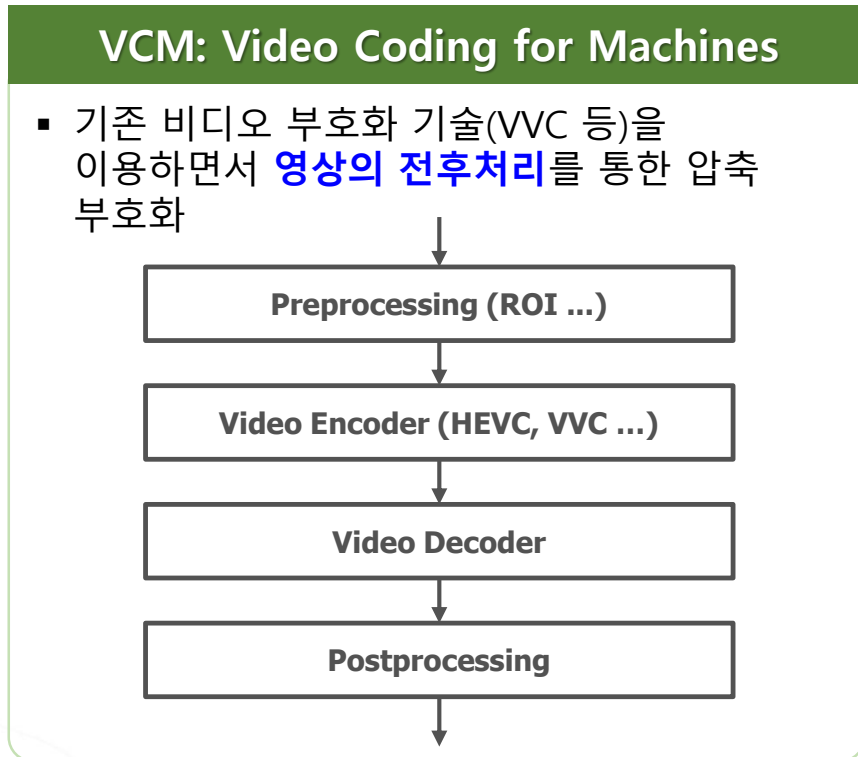
### 요구사항 수립

- a) **VCM shall support video coding for machine task consumption purposes.**
- b) VCM shall support feature coding. (Feature coding)
- c) VCM shall support a coding efficiency improvement for **at least 30% BD-rate over the VVC standard on machine vision tasks. (video coding)**
- d) VCM shall support a broad spectrum of encoding rates.
- e) VCM shall support various degrees of delay configuration.
- f) **VCM shall be agnostic to network models. (Video/Feature coding)**
- g) **VCM shall be agnostic to machine task types. (Video/Feature coding)**
- h) VCM shall provide description of the meaning or the recommended way of using the decoded data. (Feature coding)
- i) VCM should support the use and inclusion of information such as descriptors in its bitstream. (Feature/Video coding)
- j) A single VCM bitstream shall support any number of instances of machine tasks. (video coding/feature coding)
- k) VCM shall support at least the following colour formats; monochrome, RGB, and YUV (YCbCr). (Video coding)
- l) VCM shall support at least the following input bit depths: 8-bit and 10-bit. (Video coding)
- m) VCM complexity shall allow for feasible implementation within the constraints of the available technology at the expected time of usage.
- n) VCM shall support rectangular picture format up to 7680x4320 pixels (8K).
- o) VCM shall support fixed and variable rational frame rates for video inputs.
- p) VCM shall support any input source from video or image.
- q) VCM shall support privacy and security (Mandated by ISO)

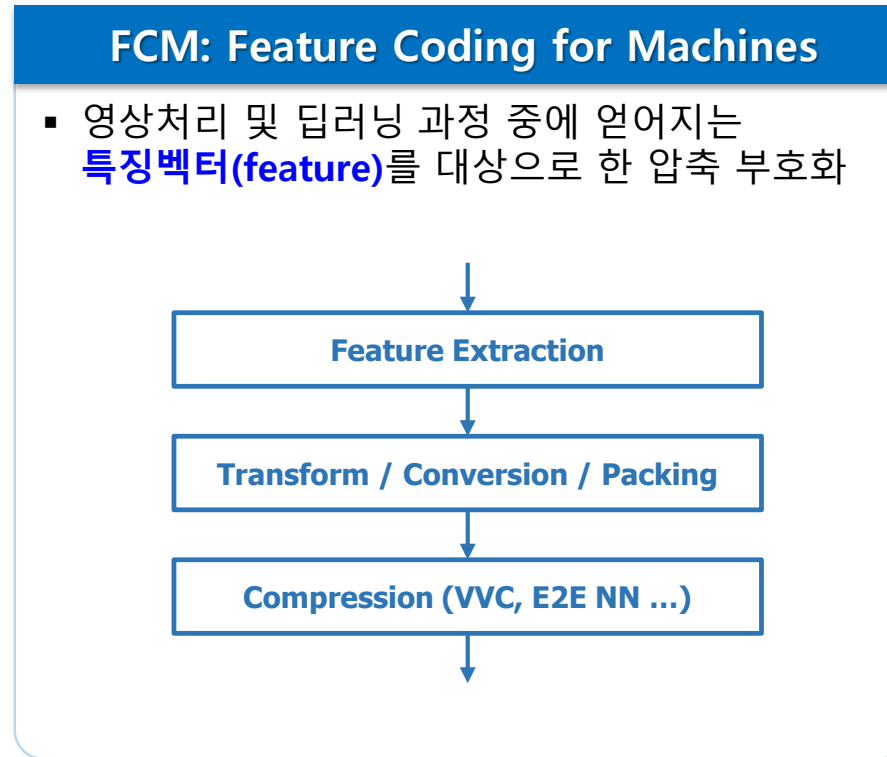
## 03. Video Coding for Machines 표준화 동향

### VCM, FCM 분리 진행 결정 – 2021.10.

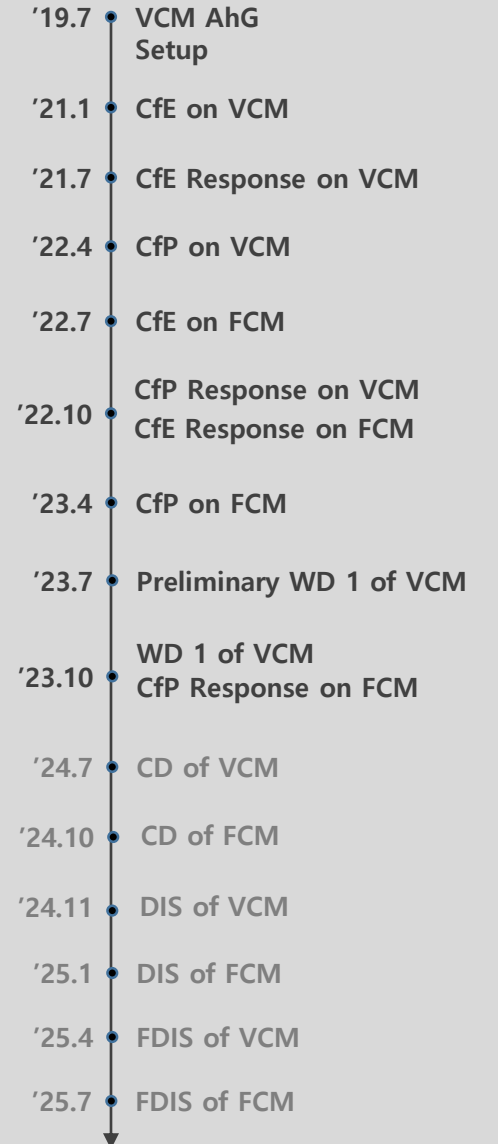
- VCM의 진행이 빨라 VCM, FCM, 두개의 track으로 나누어서 진행



[VCM CfP 참여기관] Alibaba, CAS-ICT, China Telecom, City University of Hong Kong, Ericsson, ETRI, FAU, Konkuk University, Myongji University, Nokia, OP Solutions, PUT, Tencent, V-Nova, Wuhan University, Zhejiang University



[FCM CfP 참여기관] Beihang University, Canon, China Telecom, Digital Insights Inc., ETRI, Florida Atlantic University, Hanbat National University, InterDigital, Korea Aerospace University, Kyung Hee University, OP Solutions, Sharp, vivo, Wuhan University, Zhejiang University



### 03. Video Coding for Machines 표준화 동향

VCM: 공통 실험 조건

Test sequences in SFU-HW dataset for object detection

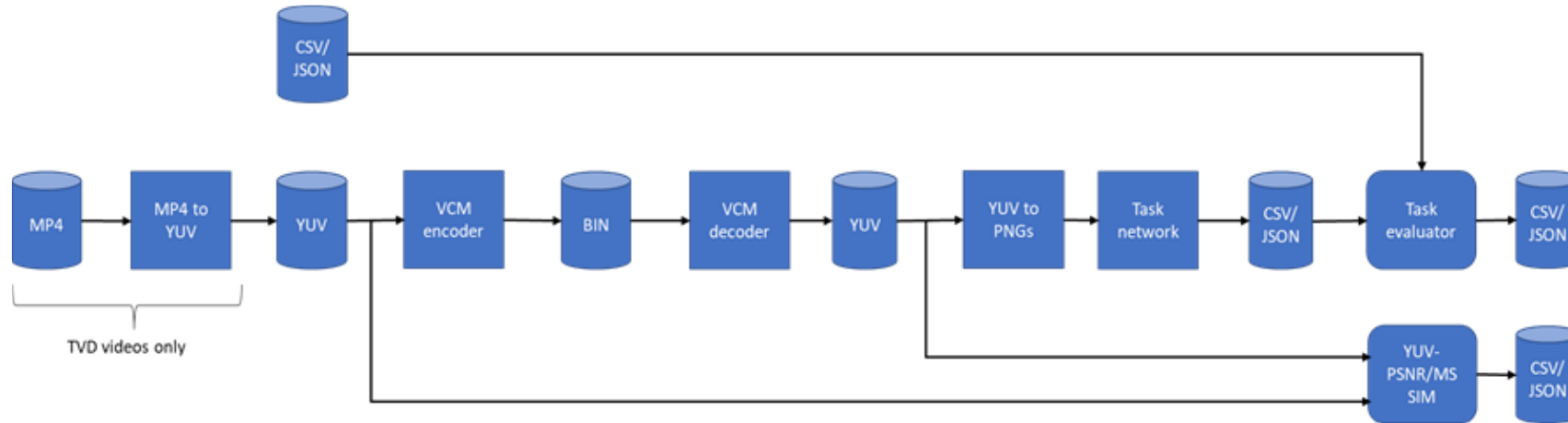
| Class | Sequence name   | Frame count | Frame rate | Bit depth | Frames skipped | Frames coded | AI | RA | LD |
|-------|-----------------|-------------|------------|-----------|----------------|--------------|----|----|----|
| A     | Traffic         | 150         | 30         | 8         | 117            | 33           | M  | M  | M  |
| B     | ParkScene       | 240         | 24         | 8         | 207            | 33           | M  | M  | M  |
| B     | Cactus          | 500         | 50         | 8         | 403            | 97           | M  | M  | M  |
| B     | BasketballDrive | 500         | 50         | 8         | 403            | 97           | M  | M  | M  |
| B     | BQTerrace       | 600         | 60         | 8         | 471            | 129          | M  | M  | M  |
| C     | RaceHorsesC     | 300         | 30         | 8         | 235            | 65           | M  | M  | M  |
| C     | BQMall          | 600         | 60         | 8         | 471            | 129          | M  | M  | M  |
| C     | PartyScene      | 500         | 50         | 8         | 403            | 97           | M  | M  | M  |
| C     | BasketballDrill | 500         | 50         | 8         | 403            | 97           | M  | M  | M  |
| D     | RaceHorsesD     | 300         | 30         | 8         | 235            | 65           | M  | M  | M  |
| D     | BQSquare        | 600         | 60         | 8         | 471            | 129          | M  | M  | M  |
| D     | BlowingBubbles  | 500         | 50         | 8         | 403            | 97           | M  | M  | M  |
| D     | BasketballPass  | 500         | 50         | 8         | 403            | 97           | M  | M  | M  |
| O     | Kimono          | 240         | 24         | 8         | 207            | 33           | O  | O  | O  |

Test sequences in TVD for object tracking

| Sequence name | Frame count | Frame rate | Bit depth | Frames skipped | Frames coded | AI | RA | LD |
|---------------|-------------|------------|-----------|----------------|--------------|----|----|----|
| TVD-01-1      | 3000        | 50         | 8         | 1500           | 500          | M  | M  | M  |
| TVD-01-2      | 3000        | 50         | 8         | 2000           | 500          | M  | M  | M  |
| TVD-01-3      | 3000        | 50         | 8         | 2500           | 500          | M  | M  | M  |
| TVD-02-1      | 636         | 50         | 10        | 0              | 636          | M  | M  | M  |
| TVD-03-1      | 2334        | 50         | 10        | 0              | 500          | M  | M  | M  |
| TVD-03-2      | 2334        | 50         | 10        | 500            | 500          | M  | M  | M  |
| TVD-03-3      | 2334        | 50         | 10        | 1000           | 500          | M  | M  | M  |

## 03. Video Coding for Machines 표준화 동향

### VCM: 성능 평가 방법



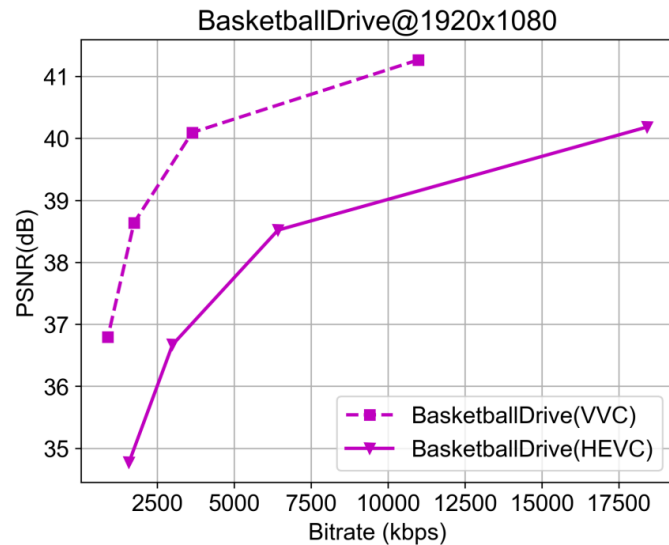
Pipeline of VCM evaluation framework for video datasets

- Machine tasks performance
  - ✓ Bitrate measurement: kbps
  - ✓ Object detection: mean Average Precision (mAP) by Faster R-CNN X101-FPN (part of Facebook AI Research's Detectron2)
  - ✓ Object tracking: Multiple Object Tracking Accuracy (MOTA) by JDE-1088x608

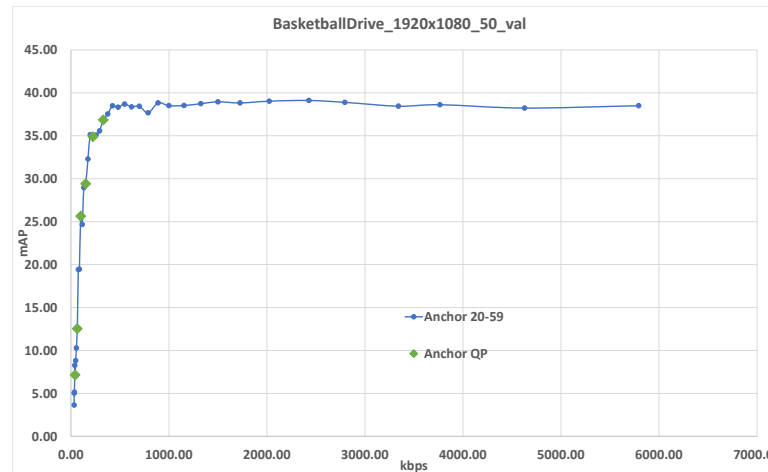
## 03. Video Coding for Machines 표준화 동향

### VCM: 성능 평가 방법

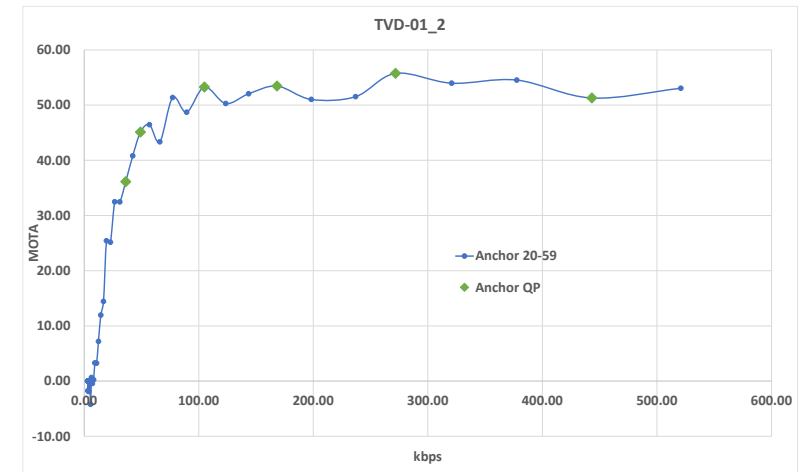
- 시퀀스별 6개의 성능 포인트를 이용하여 BD-rate mAP 또는 BD-rate MOTA를 계산하여 성능 비교
  - ✓ 어느 정도 비트율 이상에서 성능이 포화 상태에 이르는 경향이 있음
  - ✓ PSNR과는 달리 mAP, MOTA는 비트율의 증가에 따라 단조 증가하지 않음



비트율에 따른 PSNR 성능



비트율에 따른 object detection 성능



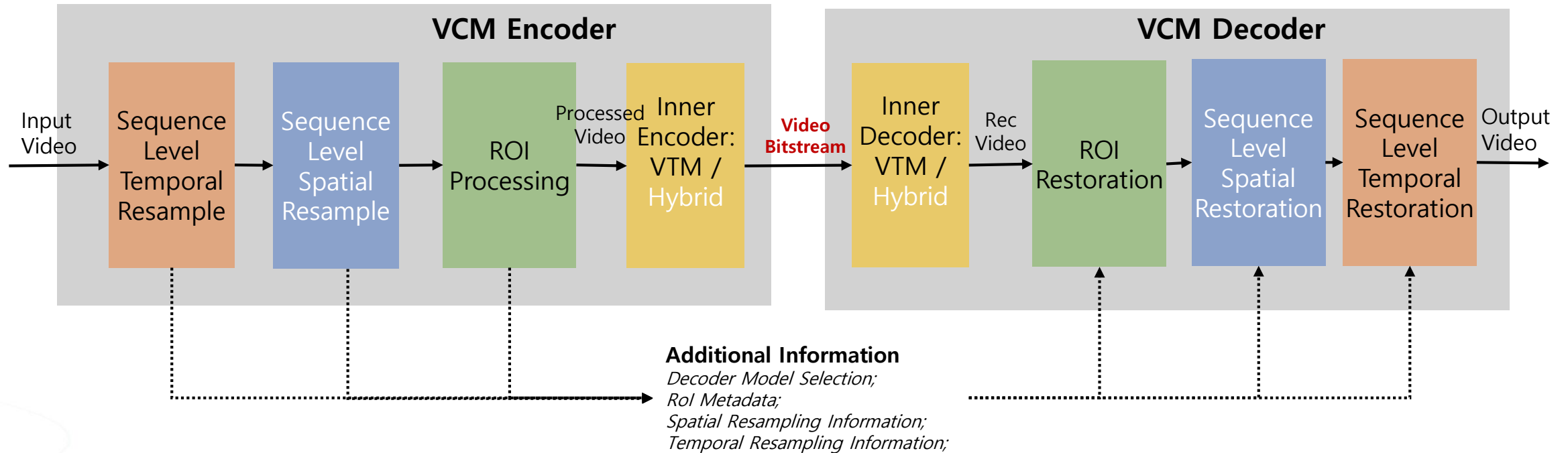
비트율에 따른 object tracking 성능

[출처] S. Bouaafia, R. Khemiri, F. E. Sayadi, "Rate-Distortion Performance Comparison: VVC vs. HEVC", SSD'21

## 03. Video Coding for Machines 표준화 동향

### VCMRS: VCM Reference Software

- 2022.11. VCMRS v0.1 배포
- 2023.11. VCMRS v0.7 배포 예정



VCMRS 구조

## 03. Video Coding for Machines 표준화 동향

### FCM: 공통 실험 조건

The configurations of the anchor for each dataset

| Machine Task                        | Configuration | QP                     |                        |
|-------------------------------------|---------------|------------------------|------------------------|
| Object detection on OpenImages      | AI            | 35, 37, 39, 41, 43, 45 |                        |
| Instance segmentation on OpenImages | AI            | 35, 37, 39, 41, 43, 45 |                        |
| Object tracking on TVD              | RA            | TVD-01                 | 22, 24, 26, 29, 31, 33 |
|                                     |               | TVD-02                 | 23, 25, 27, 28, 30, 31 |
|                                     |               | TVD-03                 | 25, 26, 27, 29, 30, 31 |
| Object Detection on SFU             | RA            | Traffic                | 37, 39, 41, 43, 45, 47 |
|                                     |               | Kimono                 | 41, 42, 43, 44, 45, 46 |
|                                     |               | ParkScene              | 32, 35, 38, 41, 44, 47 |
|                                     |               | Cactus                 | 40, 42, 44, 48, 49, 51 |
|                                     |               | BasketballDrive        | 32, 35, 38, 41, 44, 47 |
|                                     |               | BQTerrace              | 32, 35, 38, 41, 44, 47 |
|                                     |               | BasketballDrill        | 32, 35, 38, 41, 44, 47 |
|                                     |               | BQMall                 | 37, 39, 41, 43, 45, 46 |
|                                     |               | PartyScene             | 37, 40, 43, 46, 49, 51 |
|                                     |               | RaceHorsesC            | 39, 41, 43, 46, 49, 52 |
|                                     |               | BasketballPass         | 37, 39, 41, 43, 45, 46 |
|                                     |               | BQSquare               | 32, 35, 38, 41, 44, 47 |
|                                     |               | BlowingBubbles         | 40, 42, 44, 46, 49, 52 |
|                                     |               | RaceHorses             | 39, 41, 43, 46, 49, 52 |

|                          |    |    |                        |
|--------------------------|----|----|------------------------|
| Object tracking on HiEve | RA | 2  | 22, 25, 27, 29, 31, 34 |
|                          |    | 13 | 20, 22, 24, 26, 28, 29 |
|                          |    | 16 | 22, 24, 26, 28, 30, 31 |
|                          |    | 17 | 22, 23, 24, 26, 27, 28 |
|                          |    | 18 | 22, 25, 27, 29, 31, 34 |

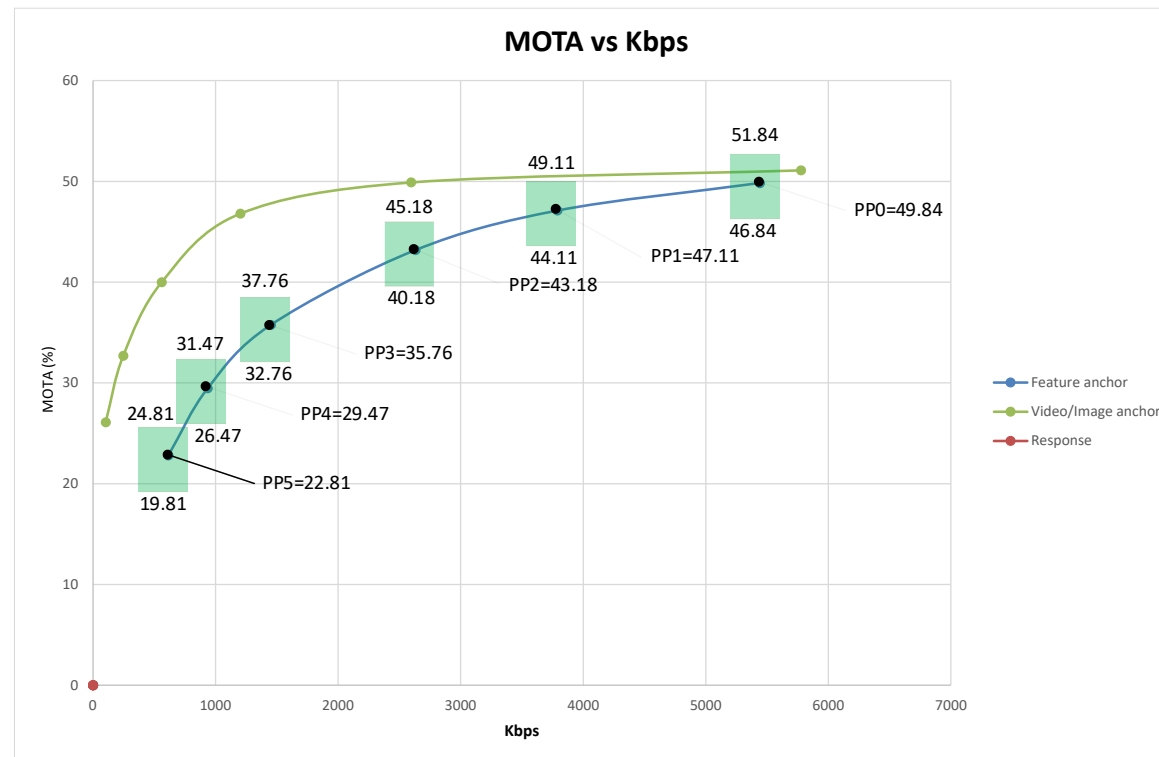
#### • 입력으로 사용할 feature

- ✓ Instance Segmentation: Mask R-CNN [2] X101-FPN at the P-layer split point for OpenImages
- ✓ Object tracking: JDE-1088x608 at layer indices 36, 61, and 74 for TVD videos and alternative split point at layers indices 75, 90, and 105 for HiEve videos
- ✓ Object detection: Faster-RCNN X101-FPN at the P-layer split point for SFU

## 03. Video Coding for Machines 표준화 동향

### FCM: 성능 평가 방법

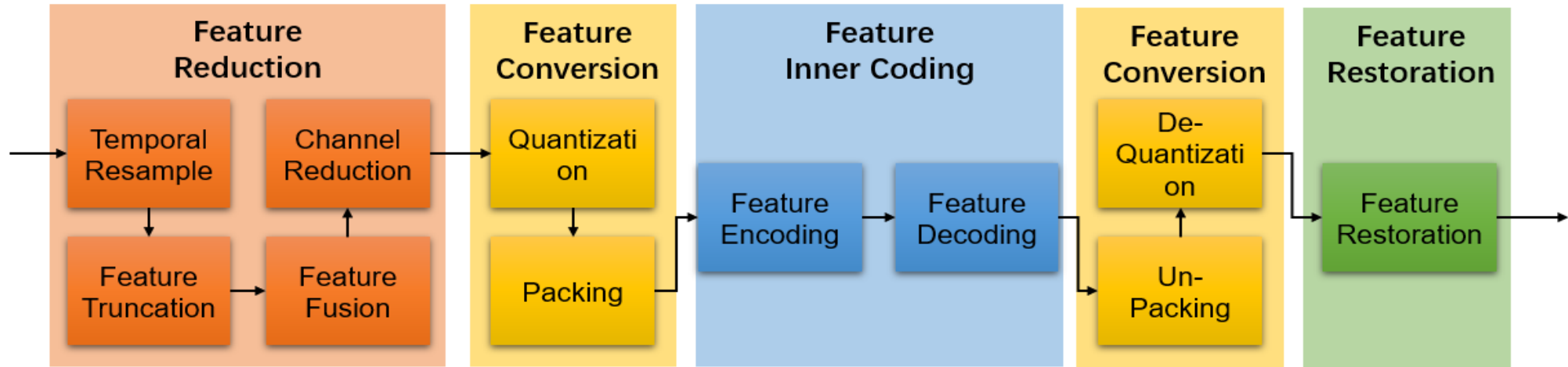
- 시퀀스별 6개의 성능 포인트를 이용하여 BD-rate mAP 또는 BD-rate MOTA를 계산하여 성능 비교
  - ✓ 제안 기술의 성능을 대응하는 앵커 결과의 특정 범위내에서 제시하도록 함



## 03. Video Coding for Machines 표준화 동향

### FCTM: Feature Compression Test Model

- 2023.11. FCTM 첫번째 버전 배포 예정



FCTM 구조

## 03. Video Coding for Machines 표준화 동향

### VCM, FCM 성능 수준(CfP 대응 기술 기준)

VCM: Task performance results (highest proposal)

| Task                  | BD-rate change |
|-----------------------|----------------|
| Object tracking       | -57%           |
| Instance segmentation | -45%           |
| Object detection      | -39%           |

FCM: Results of responses that passed crosscheck

| Proposal ID | Overall BD-Rate gain |                            |
|-------------|----------------------|----------------------------|
|             | Feature anchor       | Video anchor (informative) |
| P02         | -90.01%              | -42.70%                    |
| P03         | -80.59%              | 22.45%                     |
| P10         | -92.66%              | -57.83%                    |

## 03. Video Coding for Machines 표준화 동향

---

- 사람 중심에서 기계 중심 미디어로 기술 패러다임 변화를 주도하는 표준으로 지속적으로 발전할 것으로 예상



감사합니다.

정순흥 책임연구원, ETRI  
zeroone@etri.re.kr