|  |  |
| --- | --- |
|  | Moving Picture, Audio and Data Coding by Artificial Intelligencewww.mpai.community |

**Public Document**

|  |  |
| --- | --- |
| **N376** | 2021/09/30 |
| **Source** | Leonardo Chiariglione, Roberto Iacoviello |
| **Title** | Extending the scope of MPAI AI-based Video Coding standards |
| **Target** | MPAI Members |

Video coding standards exist since 40 years. The first two standards – H.120 and H.261 – were taken over by MPEG-1 and then by MPEG-2. The subsequent 18.5 years-old MPEG-4 AVC is still the dominant video codec in the industry.

It is a matter of fact that, 8.5 years after its publication, MPEG-H HEVC decoders are installed in most TV sets and widely used, HEVC codecs are installed in mobile handsets and hardly used and 12% of internet video content – according to the last published statistics – uses HEVC.

The latest VVC codec is going to have significant competition from codecs based on different licensing models. Currently, some groups continue the practice of just adding technologies – this time also AI based – to existing standards. The practical use of future standards based on this approach is anybody’s guess.

A key element driving the establishment of MPAI has been licensability of data – in particular video – coding standards. Therefore, MPAI has focused its video coding activities on the *MPAI-EVC Evidence Project* seeking to replace or improve existing MPEG-5 EVC tools with AI tools. The latest results confirm the gains that can be obtained by using AI tools for video coding.

Based on public announcements, MPAI expects that the licensing landscape of the MPEG-5 EVC standard will be significantly simplified. The basic MPEG-5 EVC profile is expected to be royalty free and, in a matter of months, three major companies are due to publish their MPEG-5 EVC patent licences.

MPAI’s strategy is to start from a high-performance “clean-sheet” data-processing-based coding scheme and add AI-enabled improvements to it, instead of starting from a scheme where data processing technologies give insignificant improvements and are overloaded by IP problems.

Once the MPAI-EVC Evidence Project will demonstrate that AI tools can improve the MPEG-5 EVC efficiency by at least 25%, MPAI will be in a position to initiate work on its own MPAI-EVC standard. The functional requirements already developed need only to be revised while the frame­work licence needs to be developed before a Call for Technology can be issued.

Thus, MPAI-EVC can cover the short-to-medium term video coding needs.

There is consensus in the video coding research community – and some papers make claims grounded on results – that so-called End-to-End (E2E) video coding schemes can yield significant­ly higher performance. However, many issues need to be examined, e.g., how such schemes can be adapted to a standard-based codec (see Annex 1 for a first analysis). End-to-End E2E VC promises AI-based video coding standard with significantly higher performance in the longer term.

As a technical body unconstrained by IP legacy and whose mission is to provide efficient and usable data coding standards, MPAI should initiate the study of what we can call *End-to-End Video Coding* (MPAI-EEV). This decision would be an answer to the needs of the many who need not only environments where academic knowledge is promoted but also a body that develops common understanding, models and eventually standards-oriented End-to-End video coding.

The MPAI-EVC Evidence Project should continue and new resources should be found to support the new activity. MPAI-EEV should be considered at the Interest Collection stage.

**Annex 1 – About End-to-End (E2E) video coding**

Deep learning is a powerful tool to devise new architectures (alternative to the classic block-based hybrid coding framework) that offer higher compression, or more efficient extraction of relevant information from the data.

Motivated by the recent advances in deep learning, several E2E image/video coding schemes based have been developed. By replacing the traditional video coding chain with a fully AI-based architecture, it is expected that higher compression will be obtained and that compressed videos are more visually pleasing because they do not suffer from blocking artifacts and pixelation [1].

Generally speaking, the main features of E2E schemes are [1]:

* Generalisation of motion estimation to perform compensation beyond simple translation.
* Joint optimisation of all transmitted data (motion information and residuals).



*Figure 1 – An End-to-End deep video compression scheme*

*Figure 1* depicts a possible End-to-End deep video coding scheme. It predicts the current frame using a trained network (Deep Inter Coding) and uses two auto-encoder neural networks to compress motion information (Deep Inter Coding) and prediction residuals (Deep Residual Coding). The entire architecture is jointly optimised with a single loss function, i.e. the joint rate-distortion optimisation (Rate Distortion Optimisation aims to achieve higher reconstructed frame quality for a given number of bits for compression).

Research is moving towards experimenting more general and hopefully more effective architec­tures. *Figure 2* depicts a single neural network trained to compress any video source and capable to learn and manage the trade-off between bitrate and final quality [2].



*Figure 2 – A general E2E architecture*

As designing such a network might be a challenging goal, it may be necessary to face the End-to-End challenge step by step.

**Papers Results**

MPAI has carried out a literature survey on End-to-End video coding (M533). *Table 1* summarises the gain (expressed in terms of the Bjontegaard Delta on the rate, BD-Rate) with respect to HEVC of the two deep video architectures considered.

*Table 1* shows that the average voding efficiency improvement of two reported E2E deep video architecture is 31.59% compared to HEVC.

*Table 1 – Gain of some End-to-End deep video architectures vs. AVC*

|  |  |  |
| --- | --- | --- |
| **Paper** | **Test condition** | **BD- Rate** |
| An End-to-End Learning Framework for Video Compression | AVC | -25.06% |

*Table 2*shows the coparison with HEVC of two selected E2E architecture and it shows an average coding efficiency improvement of 32.06%.

Taking into account the results above, the MPAI-EVC Evidence Project supports the new activity based on E2E video coding.

*Table 2 – Gain of some End-to-End deep video architectures vs. HEVC*

|  |  |  |
| --- | --- | --- |
| **Paper** | **Test condition** | **BD- Rate** |
| ELF-VC: Efficient Learned Flexible-Rate Video Coding | HEVC Low Delay | -26% |
| Neural Video Compression Using Spatio-Temporal Priors | HEVC Main profile | -38.12% |
| **Average** | -32,06% |

Taking into account the results above, the MPAI-EVC Evidence Project supports the new activity based on E2E video coding.

**References**

[1]G. Lu et al., *An End-to-End Learning Framework for Video Compression*, in “IEEE Transac­tions on Pattern Analysis and Machine Intelligence“, **DOI**: 10.1109/TPAMI.2020.2988453

[2] Jacob et al., *Deep Learning Approach to Video Compression*, in “2019 IEEE Bombay Sec­tion Signature Conference (IBSSC)“, 2019, **DOI**: 10.1109/IBSSC47189.2019.8973035