

**ISO/IEC JTC 1/SC 29/WG 11**

**Coding of moving pictures and audio**

**Convenorship: UNI (Italy)**

**ISO/IEC JTC 1/SC 29/WG 11 N 18987**

**Document type: Approved WG 11 document**

**Title: Test Model 4 of Low Complexity Enhancement Video Coding**

**Status: Approved**

**Date of document: 2020-03-21**

**Source: Video**

**Expected action: N/A**

**No. of pages: 7**

**Email of convenor: leonardo@chiariglione.org**

**Committee URL: mpeg.chiariglione.org**

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC 1/SC 29/WG 11 N18987**

**Brussels, BE - January 2020**

|  |  |
| --- | --- |
| **Source:** | **Video** |
| **Title:** | **Test Model 4 of Low Complexity Enhancement Video Coding** |
| **Authors:** | **Simone Ferrara (V-Nova), Stefano Battista (Universita’ Politecnica delle Marche), Florian Maurer (RWTH Aachen University)** |

**Test Model 4 of Low Complexity Enhancement Video Coding**

# Introduction

This document provides description of the Test Model of Low Complexity Enhancement Video Coding (LTM 4.x). Brief description of design principles; utilized algorithm and software organization are provided.

# General description of the algorithm

The general structure of the encoding scheme uses a downsampled source signal encoded with a base codec, adds a first level of correction data to the decoded output of the base codec to generate a corrected picture, and then adds a further level of enhancement data to an upsampled version of the corrected picture.

The codec format uses a minimum number of simple coding tools. When combined synergistically, they can provide visual quality improvements when compared with a full resolution picture encoded with the base codec whilst at the same time generating flexibility in the way they can be used.

This document describes the encoding and decoding processes by assuming a 2-dimensional (2D) downscaling/upscaling between enhancement sub-layers 1 and 2 and no scaling between the reconstructed base codec video and enhancement sub-layer 1. Any other combination of the scaling options (1-dimensional, 2-dimensional or no scaling) can be used for the two available up/downscaler as well.

## Encoding process

The encoding process is depicted in the block diagram in Fig. 1. An input full resolution video is processed to generate various encoded streams. The following streams are produced by the encoder:

* a first encoded stream *(Encoded Base)* is produced by feeding a base codec (e.g., AVC, HEVC, or any other codec) with a downsampled version of the input video;
* a second encoded stream *(L-1 Coefficient Layers)* is produced by processing the residuals obtained by taking the difference between the reconstructed base codec video and the downsampled version of the input video (L-1 residuals);
* a third encoded stream *(L-2 Coefficient Layers)* is produced by processing the residuals obtained by taking the difference between an upsampled version of a corrected version of the reconstructed base coded video and the input video (L-2 residuals);
* a fourth encoded stream *(Temporal Layer)* is produced from the temporal processing on enhancement sub-layer 2 to instruct the decoder and
* a fifth stream *(Headers)* are produced for configuring the decoder.



Fig. 1 – Block diagram of encoding process

## Decoding process

The decoding process is depicted in the block diagram in Fig. 2. The decoder receives the five streams generated by the encoder. These streams are processed as described in the following:

* The headers are used to derive the decoder configuration. This includes settings for the decoding tools as well as the format of the output video.
* The encoded base stream (Encoded Base) is decoded by a base decoder corresponding to the base codec used in the encoder.
* This output is combined with the decoded residuals obtained from the L-1 Coefficient Layers.
* The combined video is upsampled and further combined with the decoded residuals obtained from the L-2 Coefficient Layers.
* Finally, residuals from the previous frame will be added according to a temporal mask as signaled in the Temporal Layer.



Fig. 2 – Block diagram of decoding process

# Description of encoder tools

## Downsampling

The downsampling process is applied to the input video to produce a downsampled video to be encoded by a base codec. The downsampling can be done either in both vertical and horizontal directions, only in the horizontal direction or alternatively cannot be applied.

## Residual encoding

The following paragraphs describe the steps required for encoding residuals.

As discussed in Section 2.1, there are two types of residuals. The L-1 residuals are obtained by taking a difference between the reconstructed base codec video and the downsampled version of the input video. The L-2 residuals are obtained by difference between an upsampled version of a corrected version of the reconstructed base coded video and the input video.

The same steps apply to both L-1 and L-2 residuals.

### Transform

There are two types of transforms that can be used in the encoding process. Both of these transforms leverage small kernels which are applied directly to the residuals. Fig. 3 shows a representation of how the residuals are transformed.

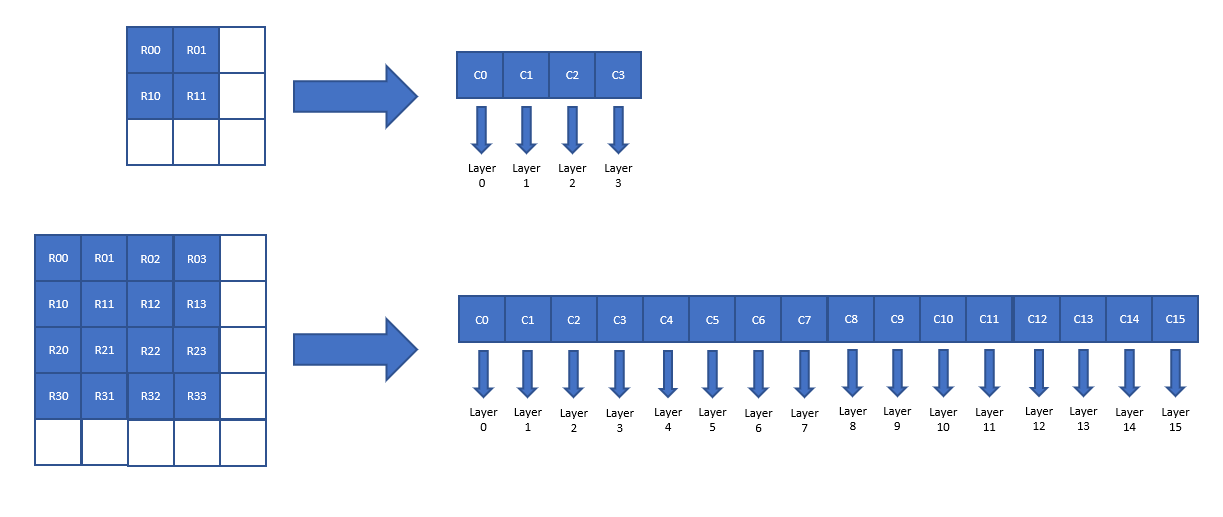


Fig. 3 – Schematic representation of transformation of residuals into coefficients and subsequent parsing into layers (called “coefficient groups” in [1]) (following quantization and entropy encoding). Top: 2x2 case, bottom: 4x4 case.

A first transform has a 2x2 kernel which is applied to each 2x2 block of residuals. The resulting coefficients are as follows:

=

A second transform has a 4x4 kernel which is applied to a 4x4 block of residuals. The resulting coefficients are as follows:

=

In the event the downsampling/upsampling is done in one direction only, the transforms are modified as follows:

=

=

### Quantization

The coefficients are then quantized using a quantizer. The quantizer may use a dead zone whose size changes relative to the quantization step.

### Entropy coding

The quantized coefficients are encoded using an entropy coder. In a scheme of entropy coding, the quantized coefficients are first encoded using run length encoding (RLE), then the encoded output is processed using a Prefix Code Encoder. There is also an option to encode using only RLE.

## Residual decoding

The input to this tool comprises the L-1 encoded residuals, which are passed through an entropy decoder, a de-quantizer and an inverse transform module. The operations performed by the entropy decoder are the inverse operations performed by the modules described in 3.2.1 - 3.2.3.

## L-1 filter

If a 4x4 transform is used, the decoded residuals may be fed to a filter module. In an implementation, the filter operates on each block of transformed residuals by applying a mask whose weights can be specified.

## Upsampling

The combination of the reconstructed L-1 residuals and base decoded video is upsampled in order to generate an upsampled reconstructed video. The upsampling method is selectable and signaled in the bytestream. Currently, there are four different upsampler kernels that could be selected.

## Predicted Residuals

Based on the Predicted Residual mode that was selected, the encoder will further modify the upsampled reconstructed video by applying Predicted Residuals (PRs). Predicted Residuals are based on the upsampled reconstructed video and on the pre-upsampling reconstructed lower resolution video.

## Temporal prediction

The temporal prediction uses a zero-motion vector prediction with a temporal buffer which stores residuals from the previous frame only.

The temporal prediction determines whether a block should be intra or inter coded. In particular, this is derived using an initial estimate provided via a temporal mask of whether (a) a tile (32x32 residuals) should be considered in its entirety an intra tile and, if not, (b) a transform block (2x2 or 4x4) should be considered an inter or intra block.

Specifically, the prediction verifies whether the number of inter blocks within a tile is above a certain threshold, in which case it decides to encode the whole tile as inter and refreshes the buffer. If that is not the case, if the tile is considered to be inter, then if the buffer is empty it encodes the whole tile as inter, if the buffer is not empty it encodes the whole tile as intra. On the other hand, if the tile is considered to be intra, it then runs a cost function to determine whether it is intra or inter.

## Level-2 (L-2) decoding

The input to this tool comprises the encoded L-2 residuals. The decoding process of the L-2 residuals are passed through an entropy decoder, a de-quantizer and an inverse transform module. The operations performed by these modules are the inverse operations performed by the modules described in 3.2.1 - 3.2.3. If the temporal selection mode has been selected, the residuals may be in part predicted from co-located residuals from a temporal buffer.

# Source Code and Information

The source code of the Test Model can be found in the following gitlab repository:

<http://mpegx.int-evry.fr/software/MPEG/LCEVC/LTM>.

A ReadMe file is provided on the same repository, containing information relative to the software implementation (LTM 4.0.0) and guidance on how to use it.

Note that the current configurations for the CTC [2] have been derived for testing the previous version of the software (LTM 3.0.1). A new configuration file for LTM 4.0.0 will be provided, as the current Test Model includes several improvements and significant changes to certain tools (e.g., quantization/dequantization) which require a different configuration file.

# References

[1] “Draft Text of ISO/IEC DIS 23094-2, Low Complexity Enhancement Video Coding”, ISO/IEC JTC1/SC29 WG11 Doc. w18986, Brussels, BE, January 2020.

[2] “Common Test Condition of Low Complexity Enhancement Video Coding”, ISO/IEC JTC1/SC29 WG11 Doc. w18988, Brussels, BE, January 2020.