|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  14th Meeting: Vienna, AT, 25 July – 2 Aug. 2013 | Document: JCTVC-N1007 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Scalable High Efficiency Video Coding Test Model 3 (SHM 3)** | | |
| *Status:* | Output Document of JCT-VC | | |
| *Purpose:* | SHVC Test Model Description | | |
| *Author(s) or Contact(s):* | Jianle Chen, Qualcomm  Jill Boyce, Vidyo  Yan Ye, InterDigital  Miska M. Hannuksela, Nokia | Email: | [cjianle@qti.qualcomm.com](mailto:cjianle@qti.qualcomm.com)  [jill@vidyo.com](mailto:jill@vidyo.com)  [Yan.Ye@interdigital.com](mailto:Yan.Ye@interdigital.com)  [miska.hannuksela@nokia.com](mailto:miska.hannuksela@nokia.com) |
| *Source:* | Editors | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This document serves as a source of general tutorial information on SHVC Test Model 3 (SHM3).

**CONTENTS**

*Page*

[Abstract i](#_Toc366938551)

[1 Introduction 1](#_Toc366938552)

[2 Description of SHVC Test Model 1](#_Toc366938553)

[2.1 General overview of SHVC 1](#_Toc366938554)

[2.2 Coding of the enhancement layer in SHVC 2](#_Toc366938555)

[2.2.1 Inter layer texture prediction in SHVC 2](#_Toc366938556)

[2.2.2 Inter layer motion prediction in SHVC 3](#_Toc366938557)

[2.3 Resampling process of picture samples 3](#_Toc366938561)

[2.3.1 Downsampling process 3](#_Toc366938562)

[2.3.2 Upsampling process 5](#_Toc366938563)

[2.3.3 Extended Spatial Scalability 5](#_Toc366938564)

[3 Software 6](#_Toc366938565)

[3.1 Software repository 6](#_Toc366938566)

[3.2 Build System 6](#_Toc366938567)

[3.3 Software Structure 6](#_Toc366938568)

[4 Reference 6](#_Toc366938569)

# Introduction

At the 12th JCTVC meeting in Geneva, it was decided to include both of two inter-layer texture prediction signalling mechanisms, namely the “reference index based” method and the “textureRL based” method in SHVC test model 1 (SHM1) [1].

At the 13th Meeting in Incheon, it was decided to include “reference index based” method in SHVC Working Draft and retain the specification text of “textureRL based” method in SHVC test model 2 (SHM2) [2].

At the 14th Meeting in Vienna, it was decided to remove “textureRL based” method from SHVC test model 3 (SHM3). The draft text corresponding to SHM3 is provided in JCTVC-N1008 [3].

# Description of SHVC Test Model

## General overview of SHVC

SHVC is the scalable extensions of HEVC which currently provide features of temporal, SNR, spatial scalability and a combination of these scalabilities. The design of HEVC enables temporal scalability when a hierarchical temporal prediction structure is used. Hence the JCTVC currently concentrates on developing tools to support spatial and coarse grain SNR scalabilities in SHVC. Spatial and SNR scalabilities are enabled using a layered approach in SHVC. A general block diagram of a three spatial layer SHVC encoder is depicted in Figure 2‑1. SHVC adopts the multi-loop decoding framework. Pictures in different layers in the same access unit are coded into the bitstream in an ascending order of layer indices. The original input pictures are downsampled and coded into the base layer (BL) bitstream by using a conformant HEVC or AVC codec. To code the enhancement layer (EL) pictures, the already coded data in the lower reference layer(s) are used for inter-layer prediction to improve EL coding efficiency. In SHM3, both reconstructed picture samples and motion parameters from the reference layer(s) can be used for prediction. The inter-layer prediction tools currently supported by SHVC are described in the following sub-sections.



Figure 2‑1 High-level block diagram of an SHVC encoder.

## Coding of the enhancement layer in SHVC

The design philosophy of the SHVC standard is to achieve high scalable coding efficiency using relative simple system architecture. Another consideration is to keep the architecture design maximally aligned with the Multi-View extensions of HEVC (MV-HEVC) [4]. Therefore, the EL codec in SHVC does not allow low-level (block-level) changes to the single layer HEVC design. Instead, SHVC high-level syntax design has been modified in such a way that the reconstructed pictures (upsampled if necessary) from the reference layers having the same POC value as that of the current picture can be used as the inter-layer reference pictures for coding the current enhancement layer picture. This allows inter-layer prediction to be carried out without any low level coding process changes. The highlighted blocks shown in Figure 2‑1 are used to generate the resampled reference layer pictures; these blocks may be the only major additions necessary to support the SHVC codec.

### Inter layer texture prediction in SHVC

In SHVC, inter-layer texture prediction is invoked by including the inter-layer reference pictures from the reference layers (resampled if necessary), together with the temporal reference pictures, in the reference picture lists of the enhancement layer picture. At the Prediction Unit (PU) level, the signalled one or two reference picture indices are used to indicate whether the current PU is predicted from temporal reference pictures, from inter-layer reference pictures, or from a combination of both. When a PU is predicted from at least one inter-layer reference picture, there is a bitstream conformance constraint that requires the motion vectors associated with the inter-layer reference picture(s) to be zero.

The initial reference picture lists in SHM3 are constructed as follows. For reference picture list 1 (L1), the temporal references are first added into the reference list in the same manner as the reference picture list construction in HEVC. After that, the inter-layer reference picture(s) is added at the end of L1 as long term reference picture(s). For reference picture list 0 (L0), the inter layer reference picture(s) is inserted between the set of short-term positive temporal (forward) reference pictures and the set of short-term negative temporal (backward) reference pictures. The inter-layer reference picture(s) is added to the reference picture list L0 when the current enhancement-layer picture is coded as P-Slice, and is added to both reference picture lists L0 and L1 when the current enhancement-layer picture is coded as B-Slice.

### Inter layer motion prediction in SHVC

In SHVC, inter-layer motion parameter prediction can be invoked by setting an inter-layer reference picture as the collocated reference picture for TMVP derivation. When spatial scalability is used between the current enhancement layer and its reference layer, motion field mapping process is performed to derive the motion field for the inter layer reference pciture; no additional block level decoding process modification to TMVP derivation is needed at the enhancement layer.

In motion field mapping process, the motion field of the inter-layer reference picture is obtained based on the compressed motion field of the lower resolution reference layer picture. The motion parameters (including MVs and reference indices) and prediction mode for each 16×16 block of the inter-layer reference picture are derived from the corresponding motion parameters and prediction mode of the collocated block in the reference layer picture. The 16×16 block size is chosen to be compliant with the HEVC TMVP derivation process, where compressed motion field of 16×16 blocks of the temporal reference picture is used.

As shown in Figure 2‑2, where each grid in the enhancement layer (right) picture represents an 8×8 block and each grid in the reference (left) layer represent a 4×4 block, the collocated 16×16 block in the reference layer picture is derived as follows:

1. The collocated sample location of the center sample of the 16×16 block in the reference layer picture is denoted as (xRL, yRL).
2. The location (xRL, yRL) is then rounded to a 16×16 block by using an unbalanced offset of 4, as follows,

xRL = ( ( xRL + 4 ) >> 4 ) << 4 (2‑1)

yRL = ( ( yRL + 4 ) >> 4 ) << 4 (2‑2)

With the un-balanced offset in the rounding operation, the values of (xRL, yRL) are rounded to appropriate 16×16 block with the top-left location indicated by 1, 2, 3 or 4 in Figure 2‑2.

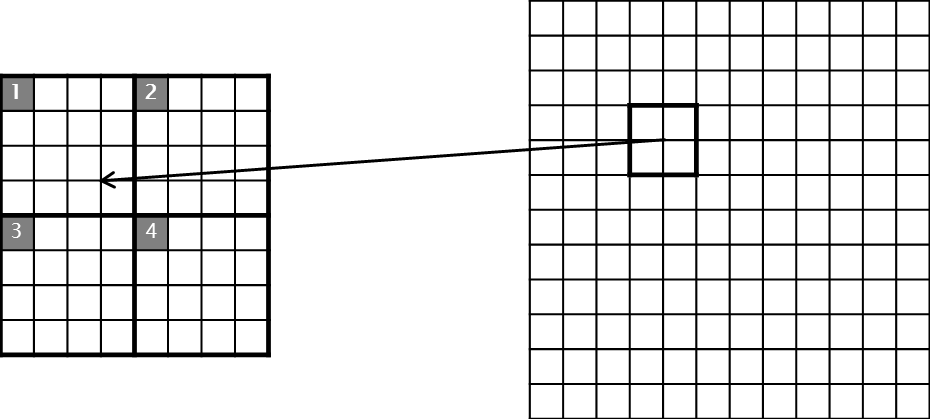


Figure 2‑2 Derivation of collocated 16×16 block in refrence layer

In SHVC, inter-layer motion parameter prediction is disabled when the reference layer is the base layer and is coded using an AVC codec.

## Resampling process of picture samples

### Downsampling process

This section introduces the downsampling process used to generate the base layer input video sequences from the enhancement layer input video sequences used in the SHVC common test conditions for ratio 1.5 and ratio 2 spatial scalability.

Figure 2‑3 (a) and (b) show the relative sampling grids of the enhancement layer picture and the base layer picture for ratio 2 and ratio 1.5 spatial scalabilities, where, white squares, white triangles, blue circles and blue triangles represent EL luma samples, EL chroma samples, BL luma samples, and BL chroma samples, respectively. Both the enhancement layer and base layer video sequences are assumed to be in YUV4:2:0 format.

**(a) ratio 2 (b) ratio 1.5**

Figure 2‑3 Relative sampling grids between EL picture and BL picture for 2x and 1.5x spatial ratio

To generate base layer pictures, 12-tap, 2D separable downsampling filters are applied to the enhancement layer pictures. The downsampling filter coefficients for ratio 1.5 and for ratio 2 are shown in Table 2‑1 and Table 2‑2, respectively. The same downsampling filters are applied to luma and to chroma components. For each sample location (*x, y*) in the base layer picture, the following process is used to select the horizontal filter phase *ph* and vertical filter phase *pv*

 (2‑3)

, where  (2‑4)

The filters with phase *ph* and phase *pv* from Table 2‑1 or Table 2‑2 are selected and applied in cascade, first in the horizontal dimension and then in the vertical dimension. For ratio 1.5, 8-phase filter is used; 4 of the 8 phases are not used by the downsampling process, therefore their filter coefficients are not shown in Table 2‑1. For ratio 2, 4-phase filter is used; 2 of the 4 phases are not used by the downsampling process and their filter coefficients are not shown in Table 2‑2.

Additional detailed information on the downsampling process, including the source code, can be found in [4].

Table 2‑1 – Ratio 1.5 downsampling filter coefficients

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| phase i | Downsampling filter coefficients | | | | | | | | | | | |
| f15[i,0] | f15[i,1] | f15[i,2] | f15[i,3] | f15[i,4] | f15[i,5] | f15[i,6] | f15[i,7] | f15[i,8] | f15[i,9] | f15[i,10] | f15[i,11] |
| 0 | 0 | 5 | -6 | -10 | 37 | 76 | 37 | -10 | -6 | 5 | 0 | 0 |
| 1 | -1 | 5 | -3 | -12 | 29 | 75 | 45 | -7 | -8 | 5 | 0 | 0 |
| 2 | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| 3 | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| 4 | -1 | 3 | 2 | -13 | 8 | 65 | 65 | 8 | -13 | 2 | 3 | -1 |
| 5 | -1 | 2 | 3 | -12 | 2 | 59 | 70 | 14 | -13 | 1 | 4 | -1 |
| 6 | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| 7 | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a |

Table 2‑2 – Ratio 2 downsampling filter coefficients

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| phase i | Downsampling filter coefficients | | | | | | | | | | | |
| f20[i,0] | f20[i,1] | f20[i,2] | f20[i,3] | f20[i,4] | f20[i,5] | f20[i,6] | f20[i,7] | f20[i,8] | f20[i,9] | f20[i,10] | f20[i,11] |
| 0 | 2 | -3 | -9 | 6 | 39 | 58 | 39 | 6 | -9 | -3 | 2 | 0 |
| 1 | 1 | -1 | -8 | -1 | 31 | 57 | 47 | 13 | -7 | -5 | 1 | 0 |
| 2 | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| 3 | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a |

### Upsampling process

This section introduces the upsampling process applied to the reference layer reconstructed picture, in order to provide inter layer texture prediction for the enhancement layer, when the reference layer and the enhancement layer pictures have different spatial resolutions.

Given the upsampling ratio N in both directions, the upsampling process is conceptually approximated by interpolating the reference layer reconstructed picture to 16 times its size in both directions and decimating the 16x picture with the ratio M, where M=16/N. To perform 16x upsampling, 16-phase interpolation filters are used. Detailed filter coefficients are provided in Table H‑1 and Table H‑2 in [3], for luma and for chroma, respectively. The upsampling filter are designed to be backward compatible with the filters used for motion compensation interpolation process in HEVC, where 8-tap filters are applied for luma component and 4-tap filters are applied for chroma component. The filters used in HEVC motion compensation interpolation process are kept unchanged in the upsampling process at the corresponding ½- and ¼-pixel fractional positions, and new filter coefficients are added for the new phases that do not exist in the HEVC motion compensation interpolation process. The current SHM only supports ratio 1.5 and ratio 2 spatial scalability, which means only a subset of the 16-phase interpolation filters are needed. The phase filters not used in either ratio 2 or ratio 1.5 upsampling process are not defined; instead, they are marked as “n/a” in Table H‑1 and Table H‑2 in [3].

In actual implementation, the 16x interpolation is not performed for every sample; instead, only those samples in the 16x picture that will be kept after decimation are interpolated. To generate a sample located at (*x, y*) in the enhancement layer picture, its corresponding position (*x*16*, y*16) in the virtual 16x picture before decimation is first found by using the method specified in sub-clause H.6.2 in [3]. Then, the sample value at (*x*16*, y*16) is interpolated by applying the appropriate phase filter to the support region in the reference layer reconstructed picture. The process to determine the reference layer support region and the appropriate phase filters are specified in sub-clause H.8.4.1.3 and H.8.4.1.4 in [3], for luma and for chroma, respectively. Table 2‑3 summarizes the filter phases used for ratio 2 and ratio 1.5x upsampling, for luma and chroma, and for each of the horizontal and vertical directions.

Table 2‑3 – Phase filters used for ratio 1.5 and ratio 2 upsampling

|  |  |  |
| --- | --- | --- |
| Ratio | Color Component & Direction | Applicable Phase |
| 2 | Luma Horizontal & Vertical | 0, 8 |
| Chroma Horizontal | 0, 8 |
| Chroma Vertical | 6, 14 |
| 1.5 | Luma Horizontal & Vertical | 0, 5, 11 |
| Chroma Horizontal | 0, 5, 11 |
| Chroma Vertical | 4, 9, 15 |

### Extended Spatial Scalability

In the scalable extension of ITU-T H.264 | ISO/IEC 14496-10, Extended Spatial Scalability (ESS) enables a generalized relation between successive spatial layers. A picture of a reference spatial layer may represent a cropped area of the higher resolution picture and the ratio between successive spatial layers can be any value. In SHM3, ESS enables the cropping mode between the reference layer and the current enhancement layer, but the spatial ratio between the two layers is restricted to be 1, 1.5 or 2. The mechanism to support arbitrary spatial ratio in SHVC is under consideration by JCTVC. The geometrical parameters defining the cropping window are signaled at the sequence level.

Figure 2‑4 illustrates the sample location relations between the reference layer and enhancement layer. In SHM3, the picture samples upsampling process is first applied to the reference layer picture to form upsampled picture within the scaled windows. Then, the scaled reference layer picture is further padded in horizontal and/or vertical directions to create an inter-layer reference picture having the same resolution as the current enhancement layer picture. In motion mapping process, the motion information outside of scaled base layer window is marked as unavailable by setting the block prediction mode to intra prediction mode.

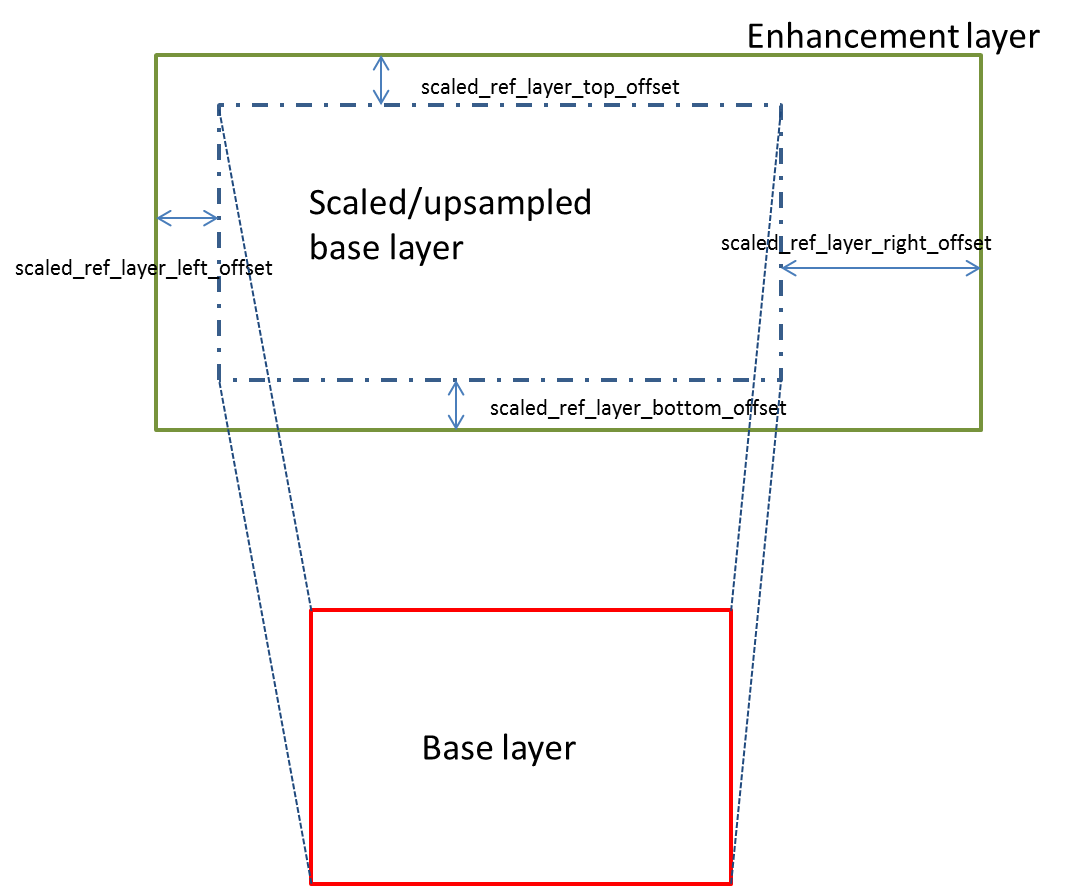


Figure 2‑4 – Relations between enhancement layer and reference layer with ESS

# Software

## Software repository

The source code for the software is available in the following SVN repository.

<https://hevc.hhi.fraunhofer.de/svn/svn_SHVCSoftware/>

For tool integration branch for a company can be obtained by contacting:

Seregin, Vadim ([vseregin@qti.qualcomm.com](mailto:vseregin@qti.qualcomm.com))

## Build System

The software can be built under linux using make. For Windows, solutions for different versions of Microsoft Visual Studio are provided.

## Software Structure

The SHVC Test Model Software has the same software structure to the current HEVC test model HM software, which includes the following applications and libraries for encoding, decoding:

* Applications:
  + TAppEncoder, executable for bit stream generation
  + TAppDecoder, executable for reconstruction.
  + TAppCommon, common functions for configuration file parsing .
* Libraries:
  + TLibEncoder, encoding functionalities
  + TLibDecoder, decoding functionalities
  + TLibCommon, common functionalities
  + TLibVideoIO, video input/output functionalities

# Reference

1. J. Chen, J. Boyce, Y. Yan and M. M. Hannuksela, “SHVC Test Model 1 (SHM 1)”, JCTVC-L1007, 12th JCTVC Meeting, Geneva, CH, Jan. 2013
2. J. Chen, J. Boyce, Y. Yan and M. M. Hannuksela, “SHVC Test Model 2 (SHM 2)”, JCTVC-M1007, 13th JCTVC Meeting, Incheon, KR, Apr. 2013
3. J. Chen, J. Boyce, Y. Yan and M. M. Hannuksela, “Scalable High Efficiency Video Coding Draft 3”, JCTVC-N1008, 14th JCTVC Meeting Vienna, AT, Jul. 2013
4. G. Tech, K. Wegner, Y. Chen, M. Hannuksela, J. Boyce, “MV-HEVC Draft Text 5”, JCT3V-E1004, 5th JCT3V Meeting Vienna, AT, Jul. 2013
5. J. Dong, Y. He, and Y. Ye, "Downsampling filter for anchor generation for scalable extensions of HEVC", m24499, 100th MPEG meeting, Geneva, CH, Apr. 2012.