Scalable video coding on DM36x

This document describes how to use temporal scalability in the beta H.264 ver 2.1 codec. It is assumed that the user has used prior version of H.264 encoder, viz ver 1.1 or ver 2.0

1.1 SVC-T feature in version 2.1 H.264 codec

New DM36x version 2.1 H.264 codec has SVC-T support. The salient features which allow temporal scalability are:

1.1.1 Encoder

- Hierarchical P Frame coding
  - Configurable number of Temporal layers
    - Up to 4 layers (e.g. -> 30fps, 15 fps, 7.5 and 3.75 fps)
    - Supports hierarchal encoding for both progressive and interlaced content
  - gaps_in_frame_num_value_allowed flag in H.264 SPS is set
    - Allows higher temporal layers to be removed
  - Rate control for hierarchical coding
    - QP allocation is similar to JSVM – higher quantization to upper layers.

- Optional SVC headers - Allows easy identification and extraction of temporal layers
  - Scalability info SEI message
    - Indicates the layers present and their dependencies.
  - Prefix NAL unit (Type 14, refer to section G.7.3.1.1 of H.264 standard)
• Indicates, to which scalable layer does this NAL belong.
  • In SVC-T context, this indicates the temporal layer that this picture belongs to.

• Flexible DPB management
  o User selection for DPB management –
    ▪ Sliding window with short term pictures
    ▪ Adaptive reference picture marking with long term frames

### 1.1.2 Decoder

• Can decode SVC-T stream produced by DM368 encoder
• It ignores the Prefix NAL unit and SEI Scalability information
• Expects the application layer or external module to give the desired temporal layer(s) for decode

### 1.2 Encoder API changes in version 2.1 codec for SVC-T

SVC-T codec interface with the application is controlled using the two create time parameters as below -

• IH264VENC_Params->numTemporalLayers
• IH264VENC_Params->svcSyntaxEnable

#### 1.2.1 numTemporalLayers:

This variable refers to the number of hierarchical layers. It can have any value from the range of 0 to 3

<table>
<thead>
<tr>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1 layer (Stream with framerate: F)</td>
</tr>
<tr>
<td>1</td>
<td>2 layer (Stream with framerate: F, F/2)</td>
</tr>
<tr>
<td>2</td>
<td>3 layer (Stream with framerate: F, F/2, F/4)</td>
</tr>
<tr>
<td>3</td>
<td>4 layer (Stream with framerate: F, F/2, F/4, F/8)</td>
</tr>
<tr>
<td>255</td>
<td>All P refer to the previous I or IDR frame (Stream with frame rate: F)</td>
</tr>
</tbody>
</table>

Where “F” is the targetFrameRate set though H264VENC_Dynamic_Params
EXAMPLES:

numTemporalLayers = 0 -> only one layer. Normal IPPP sequence. There is no temporal scalability as shown in figure below.

numTemporalLayers = 1 -> two layers of hierarchy, As shown in below figure.

numTemporalLayers = 2 -> three layers of hierarchy, as shown in the below figure.
numTemporalLayers = 3 -> four layers of hierarchy, as shown in the below figure

1.2.2 svcSyntaxEnable

This variable is used to turn SVC-T syntax ON or OFF and control DPB management. It can have any value from the range of 0 to 3.

Insertion of SVC syntax consists of

- Prefix NAL (refer to section G.7.3.1.1 of H.264 standard)
- scalability information SEI message (this is used for extracting different layers)

<table>
<thead>
<tr>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>SVC syntax is off. Short term reference frames are used. Default sliding window is used for DPB management.</td>
</tr>
<tr>
<td>1</td>
<td>SVC syntax is on. Short term reference frames are used. Default sliding window is used for DPB management.</td>
</tr>
<tr>
<td>2</td>
<td>SVC syntax is off. Long term reference frames are used. Adaptive reference marking is used for DPB management.</td>
</tr>
<tr>
<td>3</td>
<td>SVC syntax is on. Long term reference frames are used. Adaptive reference marking is used for DPB management.</td>
</tr>
</tbody>
</table>

When value of svcSyntaxEnable is 2 or 3, DPB management using adaptive reference marking is used. This consists of sending appropriate memory management control operation commands.
(MMCO commands) at various layers. In this mode of operation, long term frames are used and longTerm pictures have longTermIndex corresponding to their layer.

**Example:**

Consider an example with numTemporalLayers = 3 and svcSyntaxEnable = 3. The SVC structure is shown below (progressive content)

```
2       2       2     2
1                      1
0                                                            0
```

The pictures with the numbers above them are the longterm pictures used for reference and the number represents the longtermindex of the picture.

For interlaced content, LTI 0 and 1 are toggled for base layer and LTI 2 and 3 are used for upper layers respectively. This is to abide by H.264 adaptive buffer management constraints.

On the other hand, When value of svcSyntaxEnable is 0 or 1, DPB management is done using sliding window with only short term frames.

Below are few difference from application perspective when using the two different DPB schemes -

<table>
<thead>
<tr>
<th>Sliding window  (svcSyntaxEnable is 0 or 1)</th>
<th>Adaptive memory management  (svcSyntaxEnable is 2 or 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The overall DPB buffer requirement at the decoder end will be higher</td>
<td>More efficient in DPB buffer requirement</td>
</tr>
<tr>
<td>The temporal layers cannot be indentified unless informed through SVC syntax or though some external means.</td>
<td>LongTermIndex (LTI) can be used to identify the various temporal layers in absence of SVC syntax.</td>
</tr>
<tr>
<td>The decoding technique is relatively simpler.</td>
<td>The decoding technique is relatively complex; nevertheless, a universal decoder should decode it as the syntax is within H.264 specification.</td>
</tr>
</tbody>
</table>
Note:

IDR frame will reset the temporal GOP structure and will start a new GOP structure. Hence, it is advisable to have IDR Frame Interval (IH264VENC_DynamicParams->idrFrameInterval) in multiple of the GOP size. This is as per the H.264 specification, which resets the DPB at every IDR frame.

E.g. If numTemporalLayers = 2, IDRFrameInterval should be multiple of 4. if numTemporalLayers = 3, IDRFrameInterval should be multiple of 8.

IH264VENC_DynamicParams->intraFrameInterval does not have any effect on the temporal layer. It is treated as just another frame(similar to P frame) and can take any value.

1.3 FAQ

- What changes are needed in encoder configuration to switch between AVC and SVC-T

  Please refer to section 1.2 above.

- Do we have decoder supported for SVC-T as well?

  Our decoder can decode SVC-T stream. It ignores the additional headers and decodes the stream as if it were a hierarchical P stream.

- Does SVC-T work on both 300MHz DM365 and 400MHz DM368?

  Yes, it does on DM365 as well as DM368.

Useful links for DM36x codecs: