

Centralized Texture-Depth Packing SEI Message for H.264/AVC (Based on JCT3V-L0022)

Presenter : Jar-Ferr Yang

Center for Tomorrow Ubiquitous Cloud and Hypermedia Services
Institute of Computer and Communication Engineering
Department of Electrical Engineering
National Cheng Kung University, Taiwan

Contents

- **Centralized Texture-Depth Packing (CTDP) SEI Message Syntax**
- **CTDP Adoption Plans in Taiwan**
- **Conclusions**

Centralized Texture-Depth Packing (CTDP) SEI Message Syntax

CTDP SEI Syntax

centralized_texture-depth_packing(payloadSize) {	Descriptor
centralized_texture-depth_packing_cancel_flag	u(1)
if(! centralized_texture-depth_packing_cancel_flag) {	
centralized_texture-depth_packing_content_interpretation_type	u(3)
half_depth_line_number	u(10)
depth_spatial_flipping_flag	u(1)
depth_chroma_sampling_type	u(3)
depth_subpixel_arrangement_type	u(2)
depth_representation_type	ue(v)
baseline_dist_flag	u(1)
focal_length_flag	u(1)
z_near_flag	u(1)
z_far_flag	u(1)
d_min_flag	u(1)
d_max_flag	u(1)
if(baseline_dist_flag)	
view_syn_rep_info_element(BdistanceSign, BdistanceExp, BdistanceMantissa, BdistanceManLen)	
if(focal_length_flag)	
view_syn_rep_info_element(FlengthSign, FlengthExp, FlengthMantissa, FlengthManLen)	
if(z_near_flag)	
view_syn_rep_info_element(ZNearSign, ZNearExp, ZNearMantissa, ZNearManLen)	
if(z_far_flag)	
view_syn_rep_info_element (ZFarSign, ZFarExp, ZFarMantissa, ZFarManLen)	
if(d_min_flag)	
view_syn_rep_info_element (DMinSign, DMinExp, DMinMantissa, DMinManLen)	
if(d_max_flag)	
view_syn_rep_info_element (DMaxSign, DMaxExp, DMaxMantissa, DMaxManLen)	
if(depth_representation_type == 3) {	
depth_nonlinear_representation_num_minus1	ue(v)
for(i = 1; i <= depth_nonlinear_representation_num_minus1 + 1; i++)	
depth_nonlinear_representation_model[i]	
centralized_texture-depth_packing_persistence_flag	u(1)
}	
}	

centralized_texture- depth_packing_content_interpretation_type

Value	Interpretation
0	Indicates that the packed frame is arranged in order of reduced_depth_top (RD_T), resized_texture (RT), and reduced_depth_bottom (RD_B) from top to bottom called as CTDP TB type as illustrated in Figure D-X1.
1	Indicates that the packed frame is arranged in order of reduced_depth_left (RD_L), resized_texture (RT), and reduced_depth_right (RD_R) from left to right called as CTDP LR type as illustrated in Figure D-X2.
2	Indicates that the packed frame, which provides two texture-plus-depth views, is arranged in order of reduced_depth_leftview_top (RDL_T), reduced_depth_rightview_top (RDR_T), resized_texture_leftview (RT_L), resized_texture_rightview (RT_R), reduced_depth_leftview_bottom (RDL_B), and reduced_depth_rightview_bottom (RDR_B), from upper left, upper right to lower right called as CTDP 2TB type as illustrated in Figure D-X3.
3	Indicates that the packed frame, which provides two texture-plus-depth views, is arranged in order of reduced_depth_topview_left (RDT_L), reduced_depth_bottomview_left (RDB_L), resized_texture_topview (RT_T), resized_texture_bottomview (RT_B), reduced_depth_topview_right (RDT_R) and reduced_depth_bottomview_right (RDB_R) from upper left, upper center to lower right called as CTDP 2LR type as illustrated in Figure D-X4.



CTDP-TB



CTDP-LR

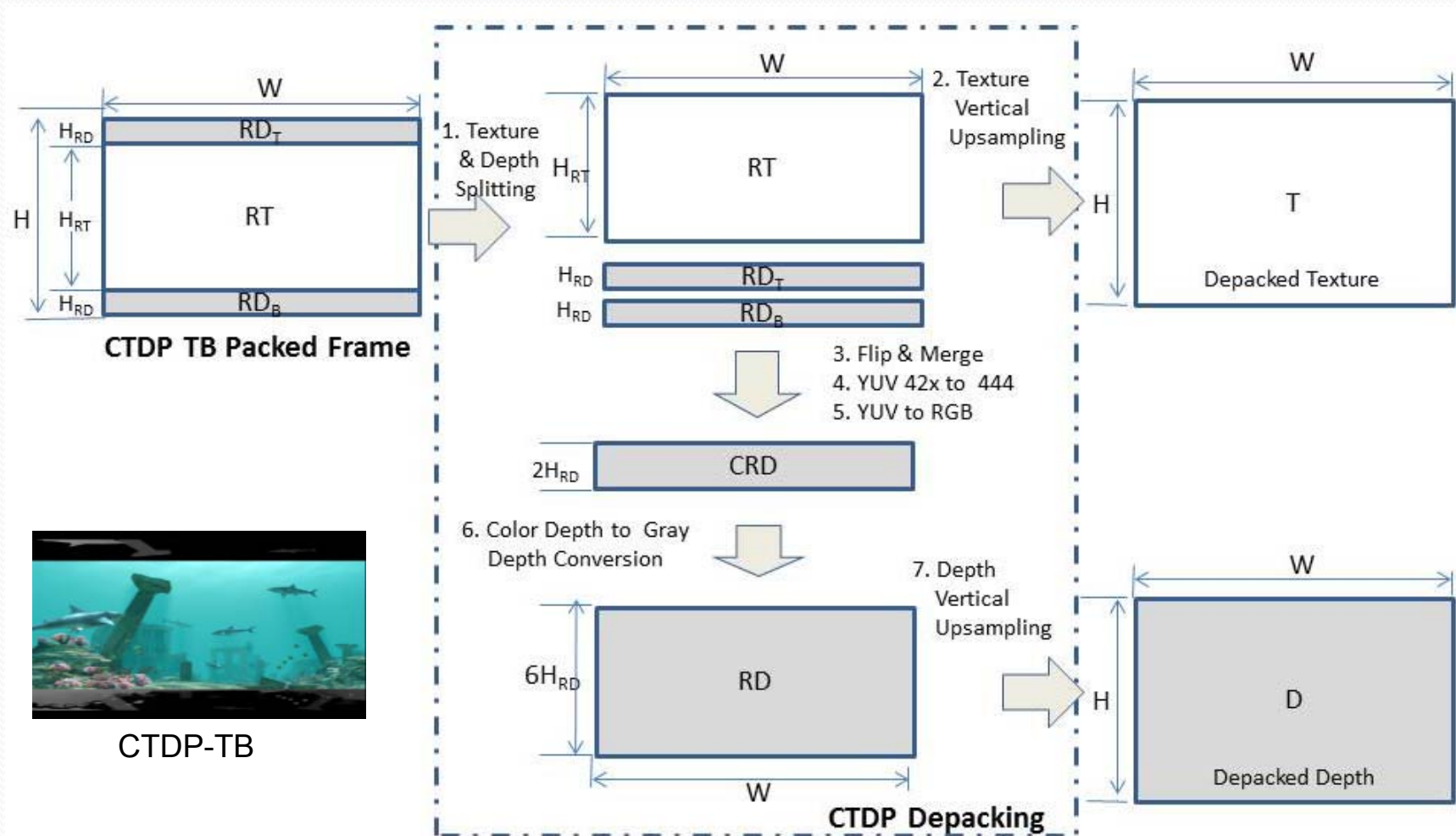


CTDP-2TB

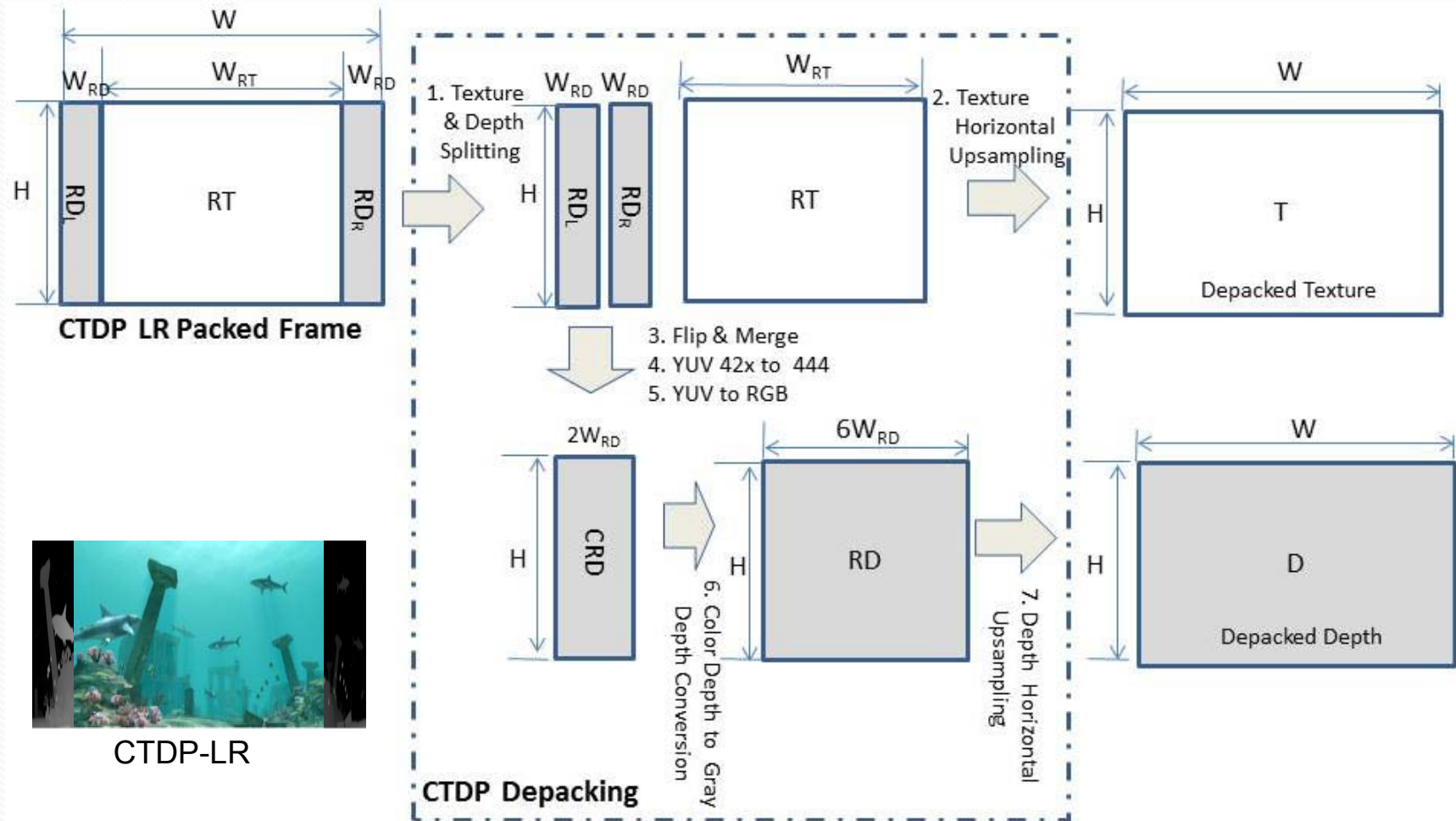


CTDP-2LR

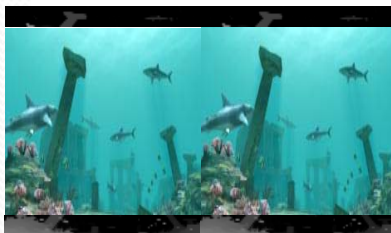
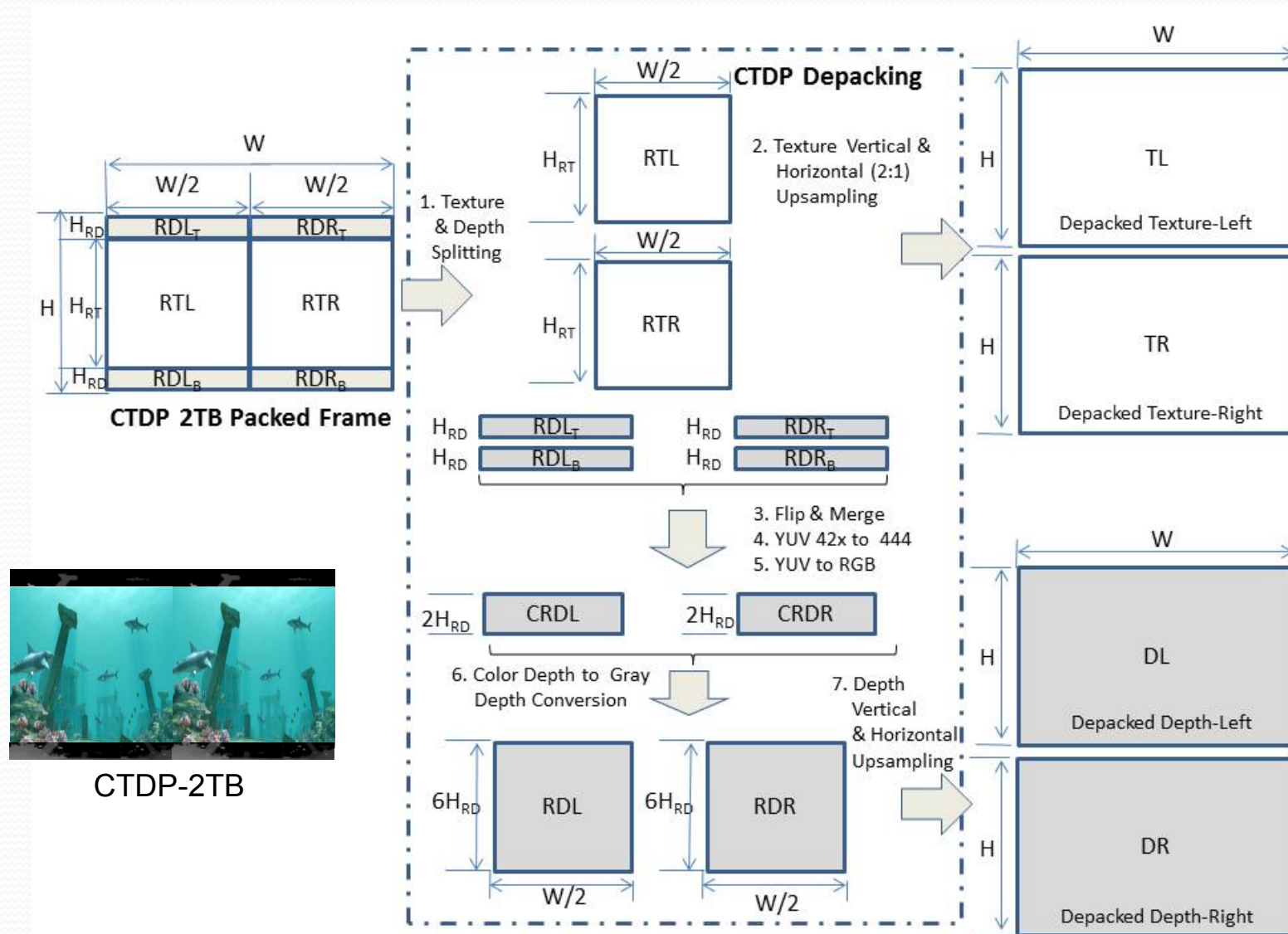
centralized_texture-depth_packing_content_interpretation_type = 0



centralized_texture-depth_packing_content_interpretation_type = 1

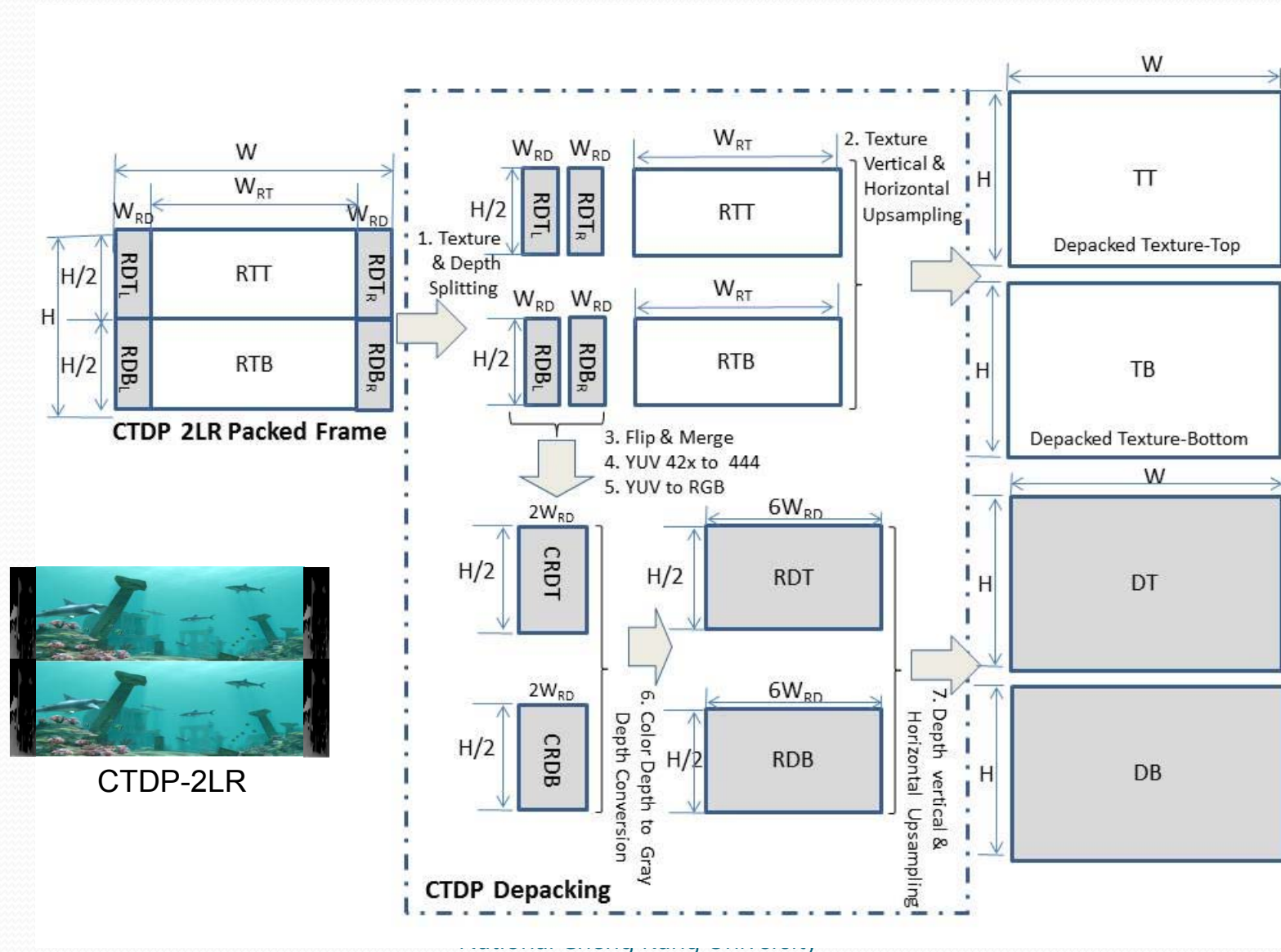


centralized_texture-depth_packing_content_interpretation_type = 2



CTDP-2TB

centralized_texture-depth_packing_content_interpretation_type = 3



half_depth_line_number

half_depth_line_number indicates the number of horizontal lines of the reduced depth top and bottom regions or the number of vertical lines of the reduced depth left and right regions. The numbers of horizontal and vertical lines of reduced texture regions can be derived accordingly. The detailed descriptions of sizes of reduced depth and reduced texture regions according to centralized_texture-depth_packing_content_interpretation_type.

Let W and H be the width and height as `pic_width_in_luma_samples` and `pic_height_in_luma_samples`, respectively, of the decoded frame output from the decoder in units of luma samples.

For CTDP-TB and CTDP-2TB packing types, the height of the reduced top and bottom depth regions is given as $H_{RD} = 2 \times \text{half_depth_line_number}$, for RD_T , RD_B , RDL_T , RDR_T , RDL_B , and RDR_B .

For CTDP-LR and CTDP-2LR packing types, the width of the reduced depth regions is defined by, $W_{RD} = 2 \times \text{half_depth_line_number}$, for RD_L , RD_R , RDT_L , RDT_R , RDB_L , or RDB_R .

The $16 \times \text{half_depth_line_number}$ should not be larger than the number of the original horizontal or vertical lines of the luma frame to achieve reasonable depth and texture packing.

depth_spatial_flipping_flag

- depth_spatial_flipping_flag indicates that the reduced depth region are spatially flipped relative to its intended orientation for display or other such purposes

Value	Interpretation
0	No flipping
1	Flipping

No flipping



Flipping



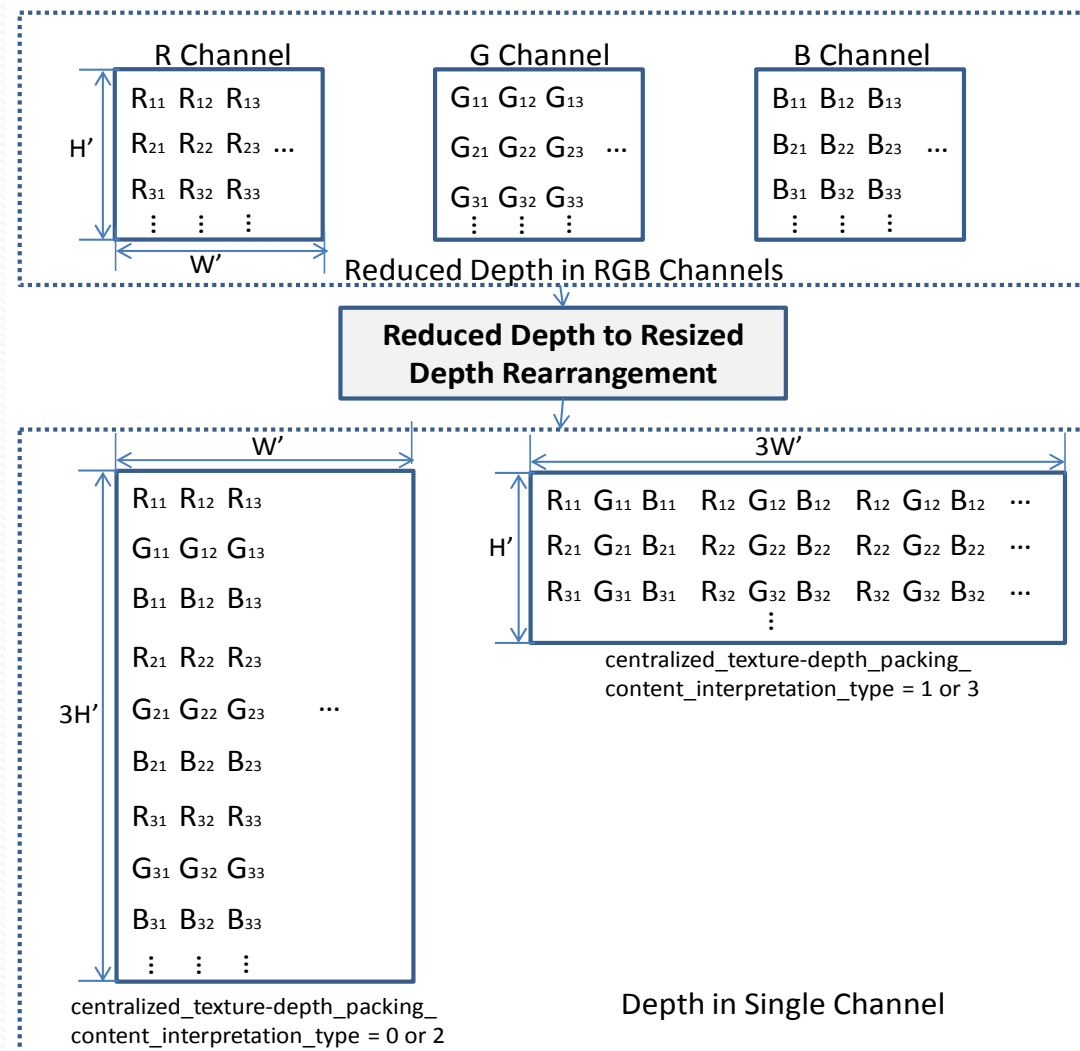
depth_chroma_sampling_type

- depth_chroma_sampling_type indicates that indicates that the depth chroma sampling method from 444 to 422 or 420 format is used by the CTDP packer as specified in Table D-X2. According to depth_chroma_sampling_type, YCbCr colour 422 or 420 format is converted to YCbCr colour 444 format.

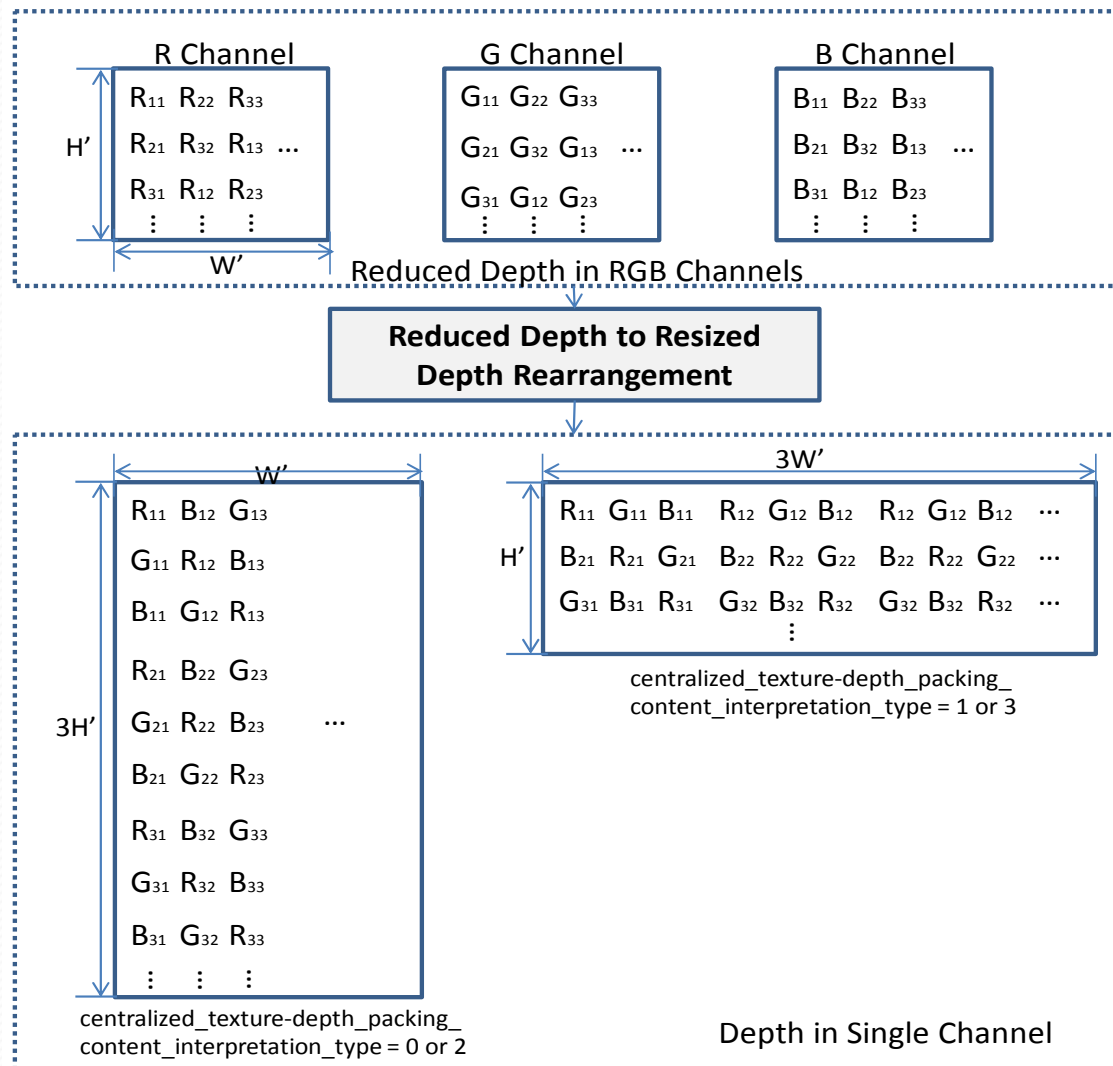
Value	Interpretation
0	CTDP depacker is informed that the direct sample is used to convert the chroma sample from 444 to 422 or 420 format according to marked chroma sample location..
1	CTDP depacker is informed that the mean of two (four) samples is used to convert the chroma sample from 444 to 422 (420) format.

Values of depth_chroma_sampling_type that do not appear in Table D X2 are reserved for future specification by ITU-T |ISO/IEC and shall not be present in bitstreams conforming to this version of this Specification. Decoders shall ignore centralized texture-depth packing SEI messages that contain reserved values of depth_chroma_sampling_type.

depth_subpixel_arrangement_type = 0



depth_subpixel_arrangement_type = 1



depth_representation_type

- depth_representation_type specifies the representation definition of decoded luma samples of auxiliary pictures. In the table, disparity specifies the horizontal displacement between two texture views and Z value specifies the distance from a camera.

Value	Interpretation
0	Each decoded luma sample value of an auxiliary picture represents an inverse of Z value that is uniformly quantized into the range of 0 to maxVal, inclusive. When z_far_flag is equal to 1, the luma sample value equal to 0 represents the inverse of ZFar (specified below). When z_near_flag is equal to 1, the luma sample value equal to maxVal represents the inverse of ZNear (specified below).
1	Each decoded luma sample value of an auxiliary picture represents disparity that is uniformly quantized into the range of 0 to maxVal, inclusive. When d_min_flag is equal to 1, the luma sample value equal to 0 represents DMin (specified below). When d_max_flag is equal to 1, the luma sample value equal to maxVal represents DMax (specified below).
2	Each decoded luma sample value of an auxiliary picture represents a Z value uniformly quantized into the range of 0 to maxVal, inclusive. When z_far_flag is equal to 1, the luma sample value equal to 0 corresponds to ZFar (specified below). When z_near_flag is equal to 1, the luma sample value equal to maxVal represents ZNear (specified below).
3	Each decoded luma sample value of an auxiliary picture represents a nonlinearly mapped disparity, normalized in range from 0 to maxVal, as specified by depth_nonlinear_representation_num_minus1 and depth_nonlinear_representation_model[i]. When d_min_flag is equal to 1, the luma sample value equal to 0 represents DMin (specified below). When d_max_flag is equal to 1, the luma sample value equal to maxVal represents DMax (specified below).
Other values	Reserved for future use

The variable maxVal is set equal to $(1 \ll (8 + \text{bit_depth_luma_minus8})) - 1$, where bit_depth_luma_minus8 is the value included in or inferred for the active SPS of the layer with nuh_layer_id equal to targetLayerId.

baseline_dist_flag equal to 0 specifies that the syntax elements specifying the baseline distance are not present in the syntax structure. **baseline_dist_flag** equal to 1 specifies that the syntax elements specifying the baseline distance value are present in the syntax structure.

focal_length_flag equal to 0 specifies that the syntax elements specifying the focal length are not present in the syntax structure. **focal_length_flag** equal to 1 specifies that the syntax elements specifying the focal length value are present in the syntax structure.

z_near_flag equal to 0 specifies that the syntax elements specifying the nearest depth value are not present in the syntax structure. **z_near_flag** equal to 1 specifies that the syntax elements specifying the nearest depth value are present in the syntax structure.

z_far_flag equal to 0 specifies that the syntax elements specifying the farthest depth value are not present in the syntax structure. **z_far_flag** equal to 1 specifies that the syntax elements specifying the farthest depth value are present in the syntax structure.

d_min_flag equal to 0 specifies that the syntax elements specifying the minimum disparity value are not present in the syntax structure. **d_min_flag** equal to 1 specifies that the syntax elements specifying the minimum disparity value are present in the syntax structure.

d_max_flag equal to 0 specifies that the syntax elements specifying the maximum disparity value are not present in the syntax structure. **d_max_flag** equal to 1 specifies that the syntax elements specifying the maximum disparity value are present in the syntax structure.

View synthesis representation information element syntax

view_syn_rep_info_element(OutSign, OutExp, OutMantissa, OutManLen) {	Descriptor
da_sign_flag	u(1)
da_exponent	u(7)
da_mantissa_len_minus1	u(5)
da_mantissa	u(v)
}	

The syntax structure specifies the value of an element in the view synthesis representation information SEI message.

The syntax structure sets the values of the OutSign, OutExp, OutMantissa and OutManLen variables that represent a floating-point value. When the syntax structure is included in another syntax structure, the variable names OutSign, OutExp, OutMantissa and OutManLen are to be interpreted as being replaced by the variable names used when the syntax structure is included.

da_sign_flag equal to 0 indicates that the sign of the floating-point value is positive. da_sign_flag equal to 1 indicates that the sign is negative. The variable OutSign is set equal to da_sign_flag.

da_exponent specifies the exponent of the floating-point value. The value of da_exponent shall be in the range of 0 to $2^7 - 2$, inclusive. The value $2^7 - 1$ is reserved for future use by ITU-T | ISO/IEC. Decoders shall treat the value $2^7 - 1$ as indicating an unspecified value. The variable OutExp is set equal to da_exponent.

da_mantissa_len_minus1 plus 1 specifies the number of bits in the da_mantissa syntax element. The value of da_mantissa_len_minus1 shall be in the range of 0 to 31, inclusive. The variable OutManLen is set equal to da_mantissa_len_minus1 + 1.

da_mantissa specifies the mantissa of the floating-point value. The variable OutMantissa is set equal to da_mantissa.

Association between depth parameter variables and syntax elements

x	s	e	n	v
Bdistance	BdistanceSign	BdistanceExp	BdistanceMantissa	BdistanceManLen
Flength	FlengthSign	FlengthExp	FlengthMantissa	FlengthManLen
ZNear	ZNearSign	ZNearExp	ZNearMantissa	ZNearManLen
ZFar	ZFarSign	ZFarExp	ZFarMantissa	ZFarManLen
DMax	DMaxSign	DMaxExp	DMaxMantissa	DMaxManLen
DMin	DMinSign	DMinExp	DMinMantissa	DMinManLen

The variables in the x column of table are derived from the respective variables in the s, e, n and v columns as follows:

- If the value of e is in the range of 0 to 127, exclusive, x is set equal to $(-1)^s * 2^{e-31} * (1 + n \div 2^v)$.
- Otherwise (e is equal to 0), x is set equal to $(-1)^s * 2^{-(30+v)} * n$.

The DMin and DMax values, when present, are specified in units of a luma sample width of the coded picture.

The units for the ZNear and ZFar, when present, are identical but unspecified.

When depth_representation_type is equal to 0 or 2, the disparity D can further be obtained from the value of Z, Bdistance, and Flength, where $D = Bdistance \times Flength \div Z$

depth_nonlinear_representation_num_minus1 plus 2 specifies the number of piecewise linear segments for mapping of depth values to a scale that is uniformly quantized in terms of disparity.

depth_nonlinear_representation_model[i] for i ranging from 0 to $\text{depth_nonlinear_representation_num_minus1} + 2$, inclusive, specify the piecewise linear segments for mapping of decoded luma sample values of an auxiliary picture to a scale that is uniformly quantized in terms of disparity. The values of **depth_nonlinear_representation_model[0]** and **depth_nonlinear_representation_model[depth_nonlinear_representation_num_minus1 + 2]** are both inferred to be equal to 0.

centralized_texture_depth_packing_persistence_flag specifies the persistence of the centralized texture-depth packing arrangement SEI message.

centralized_texture-depth_packing_persistence_flag equal to 0 specifies that the centralized texture-depth packing arrangement SEI message applies to the current decoded frame only.

centralized_texture-depth_packing_persistence_flag equal to 1 specifies that the centralized texture-depth packing arrangement SEI message persists in output order until any of the following conditions are true:

- A new CVS begins.
- The bitstream ends.
- A frame in an access unit containing a centralized texture-depth packing SEI message with the same

Adoption Plans of CTDP Formats

ChungHwa Telecom's Adoption Plan

Chunghwa Telecom (CHT) Company with about 1.4 million MOD users has signed an agreement to adopt CTDP format as the following:

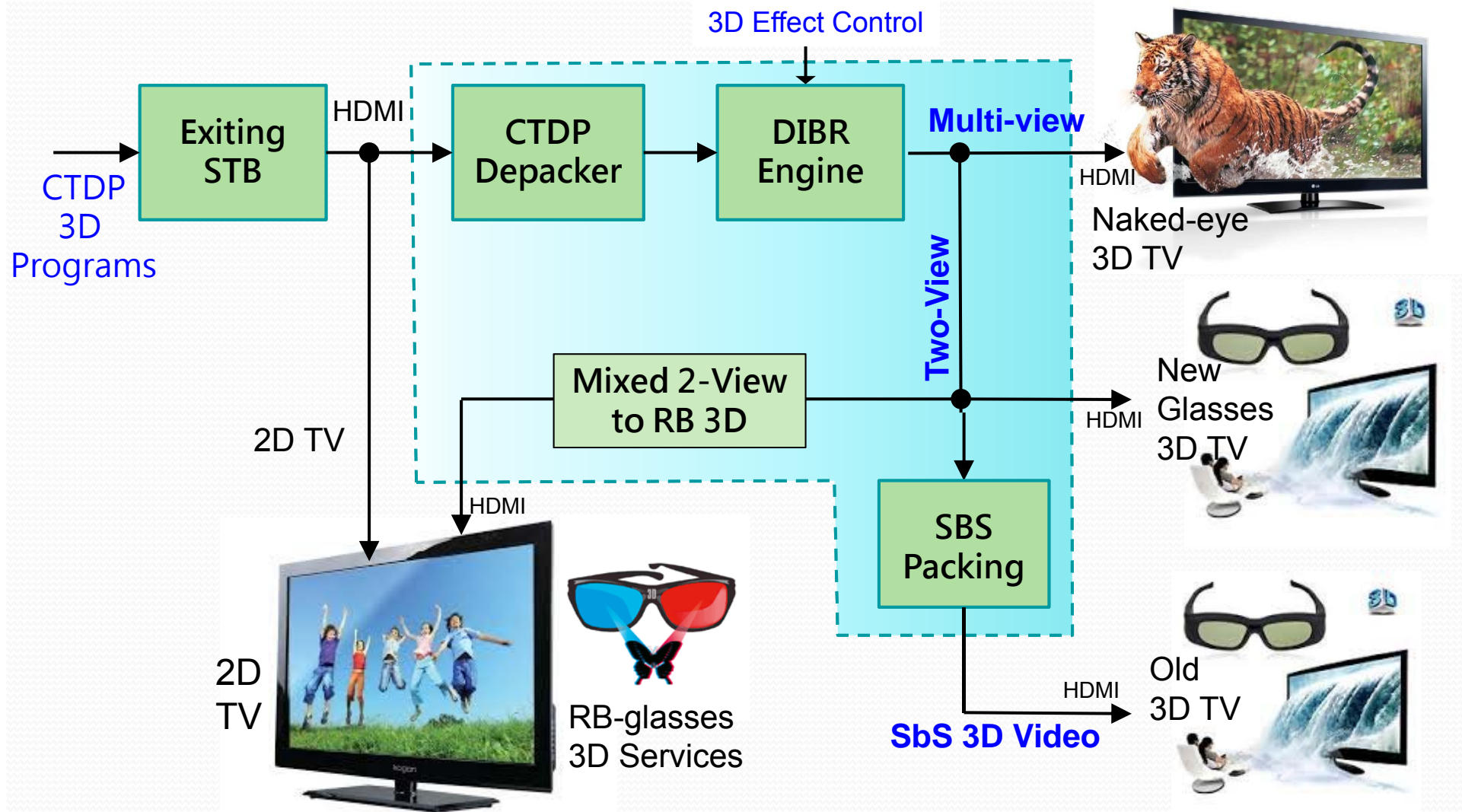
1. 09/01/2015: Propose and confirm the test and adoption detailed plan
2. 02/01/2016: Test of downloaded 3D CTDP services with computer clients
(Internal test in CHT and TOUCH Center only)
3. 05/01/2016: Services of 3D MOD services for computer clients
(Charge and open to the public and final CTDP format selection)
4. 11/01/2017: Test of 3D MOD services through Set-top-Boxes with separated CTDP depacking module (Open for selected users)
5. 11/01/2018 : Test of 3D MOD services through Set-top-Boxes with Integrated CTDP depacking module (charged 3D services)

Hsin Yeong An's Adoption Plan

Hsin Yeong An (HYA) Cable TV Company with about 145 thousand users is a local cable company near the National Cheng Kung University. The test and adoption plan with time lines are designed as the following:

1. 09/01/2015: Propose and confirm the test and adoption detailed plan
2. 02/01/2016: Test of downloaded 3D CTDP services with computer clients
(Internal test in HYA and TOUCH Center only)
3. 11/01/2016: Services of 3D CATV services for computer clients
(Charge and open to the public and final CTDP format selection)
4. 11/01/2017: Test of 3D CATV services through Set-top-Boxes with separated CTDP depacking module (Open for selected users)
5. 11/01/2018 : Test of 3D CATV services through Set-top-Boxes with Integrated CTDP depacking module (charged 3D services)

First Stage of 3D Service Delivery in Taiwan



Conclusions

- This contribution proposes modified CTDP SEI Message based on the prior contribution JCT3V-L0022.
- Before the available of 3D broadcasting systems, we believe that the proposed CTDP formats with SEI Message could help to deliver 3D videos in the current 2D broadcasting systems simply and efficiently.
- Once CTDP SEI Message is **accepted by MPEG**, ChungHwa Telecom MOD System and Hsin-Yeong An Cable TV Company will adopt CTDP formats for their 3D video services in Taiwan.

Thanks for your kind attention

Q&A