

SDR and HDR video coding technology proposal by Sony

Karl Sharman

Masaru Ikeda



Summary of Proposal

- 10 new tools in 6 categories to enhance coding efficiency on top of JEM.
 1. Sign Prediction
 2. Use of multiple reference samples in intra prediction
 3. Modified PDPC Planar
 4. Transform Matrix Replacement
 5. Adaptive Multiple core Transforms for Luma and Chroma
 6. Adaptive Scaling for Transform and Quantization
 7. Affine MC
 8. Large CTU
 9. Extended Deblocking filter
 10. Modified Adaptive Loop Filter (ALF) Classification
- Common design between SDR and HDR.
- SDR: -2.41%, -4.85% and -5.14% for Y, U and V, respectively, and 378% and 131% for encoding and decoding time.
- HDR: -1.99%, -6.11% and -8.46% for Y, U and V, respectively, and 281% and 120% for encoding and decoding time.

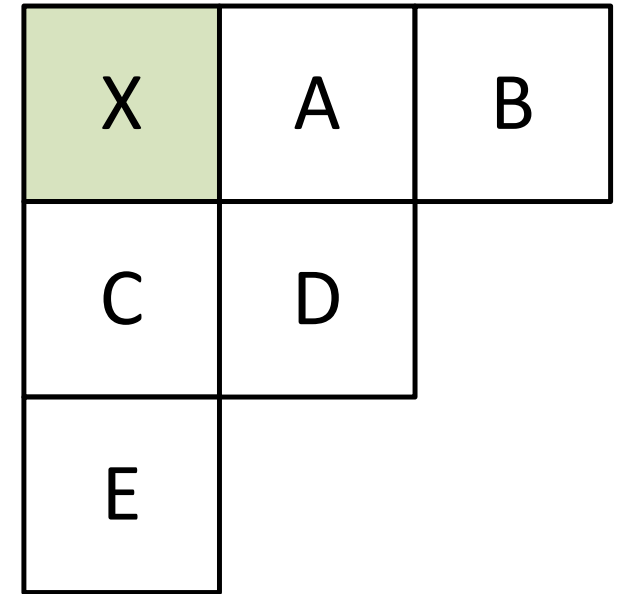
Coding tools

1. Entropy coding: sign prediction (1/2)

- In the spatial domain:
 - the values in the residual of inter blocks is, in general, distributed evenly in a block.
 - the values in the residual of intra blocks is, in general, lower nearer the predictors.
- This has implications in the frequency domain, and in particular, the sign bits can be predicted.
- Proposal:
 - The sign bits are coded using CABAC contexts when a prediction is made, with the value coded indicating whether the prediction was correct.
 - The prediction is made from the surrounding previously coded coefficients.
 - The prediction is dependent upon transform type (transform skip, EMT).

1. Entropy coding: sign prediction (2/2)

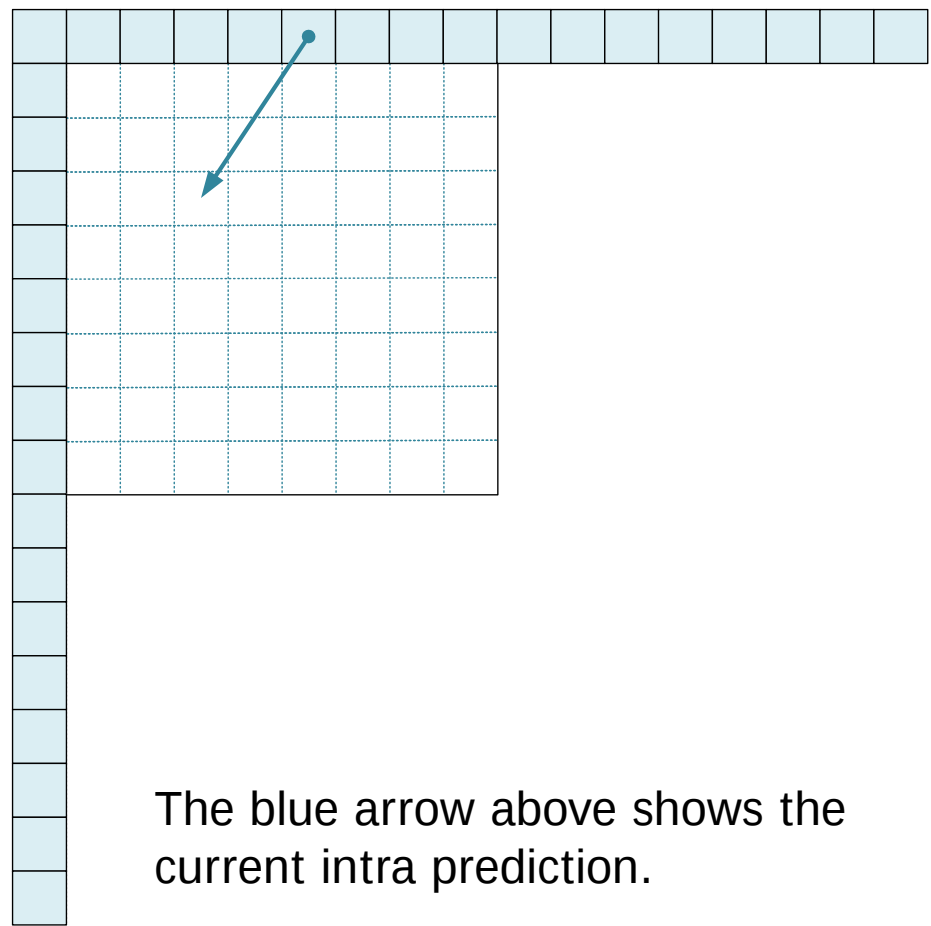
- In the CFP response, for all inter blocks,
 - Depending on the transform type (transform skip, EMT mode),
 - A prediction for coefficient 'X' is made from the previously coded coefficients A-E.
 - This is similar to significance map processing.
- In addition:
 - Sign bit hiding has been modified to hide the sign of the highest frequency coefficient.
 - The highest frequency coefficient is the least predictable; the lowest frequency coefficient is the most predictable.
- The signaling of EMT and NSST have been adjusted to use the lastSigPos, so that the EMT/NSST modes can be determined prior to decoding of coefficient signs.



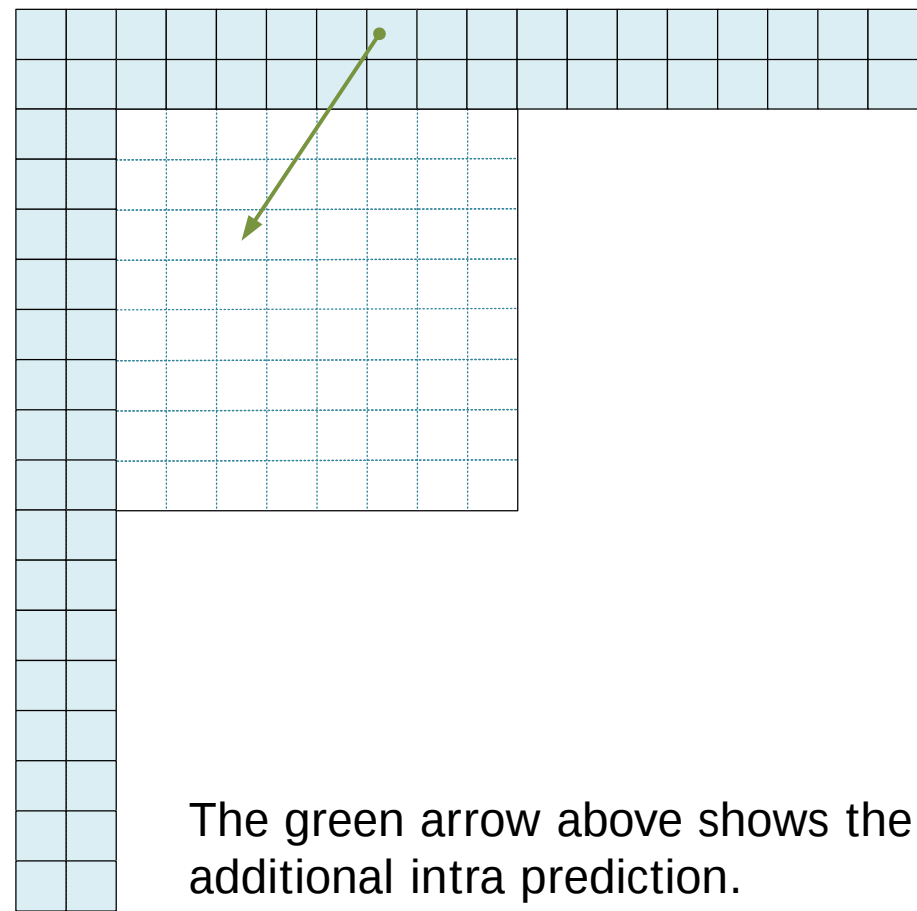
2. Use of multiple reference samples in intra prediction (1/2)

- Filtering the intra prediction reference data *along* the direction of prediction is potentially preferable to filtering *across* the feature.
 - Filtering is currently only done along the row/column, this is more likely to be *across* the direction of prediction.
- Proposal:
 - Use an additional row/column of samples to generate a second angular prediction and average the two predictions.
 - This second prediction is disabled for the top row of the CTU to limit the additional storage required – no extra line-stores needed.
 - As the second row/column is further away from the predicted sample position it may be less accurate so the second prediction is disabled for small blocks (4xN and Nx4).

2. Use of multiple reference samples in intra prediction (2/2)



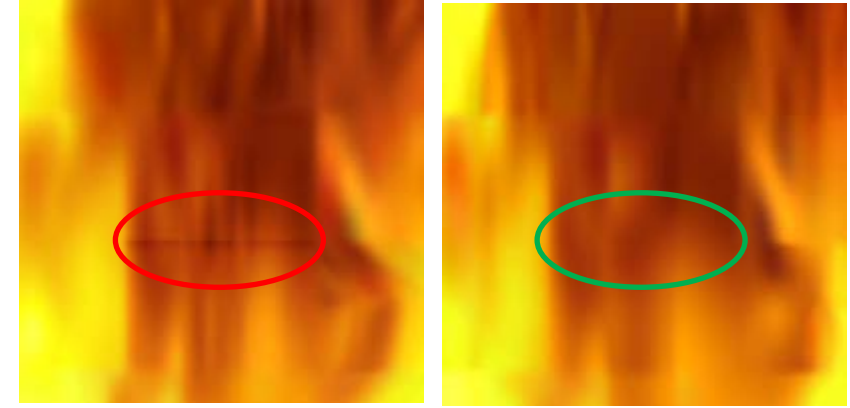
The blue arrow above shows the current intra prediction.



The green arrow above shows the additional intra prediction.

The two predictions are combined by averaging.

3. Modified PDPC Planar



- The PDPC planar tool applies a blend of the unfiltered reference samples with the sample planar-predicted from the filtered reference samples, with the amount of blend determined by the distance from the reference sample.
- In JEM, lines have been observed in intra predicted blocks in inter slices.
 - This occurs in blocks with a dimension ≥ 64
 - (Intra slices do not allow PDPC to be used in blocks that have a dimension > 32).
- Proposal (bug fix):
 - This is due to a logical shift of greater than 32 being applied to a 32-bit integer in the software.
 - To resolve this issue the derivations of Coeff_Top and Coeff_Left have been modified so that a 0 weighting is used when shiftRow or shiftCol respectively exceed 31.

4. Transform Matrix Replacement (1/2)

- Motivation
 - JEM has five types of primary transform: DCT2, DCT5, DCT8, DST1, DST7.
 - But, five unique transforms increases implementation cost.
- Proposal
 - Replace DCT5, DCT8 and DST1 with alternate transforms, as a result the number of transform matrix is two, DCT2 and DST7.

Replaced Transform	Alternate transform
DCT5	DCT2
DCT8	Flipped DST7 (called FlipDST7, identical to DCT8 except odd-position basis vectors' sign)
DST1	Transposed DST7 (called TrDST7, identical to DST6)

4. Transform Matrix Replacement (2/2)

- Transform matrix for each DCT2/DST7/FlipDST7/TrDST7 are defined as below:

$$\text{DCT2: } C_N^2[k, l] = \text{round} \left(2^m \frac{2}{\sqrt{2N}} \delta_k \cos \left\{ \frac{\left(l + \frac{1}{2}\right)k\pi}{N} \right\} \right),$$

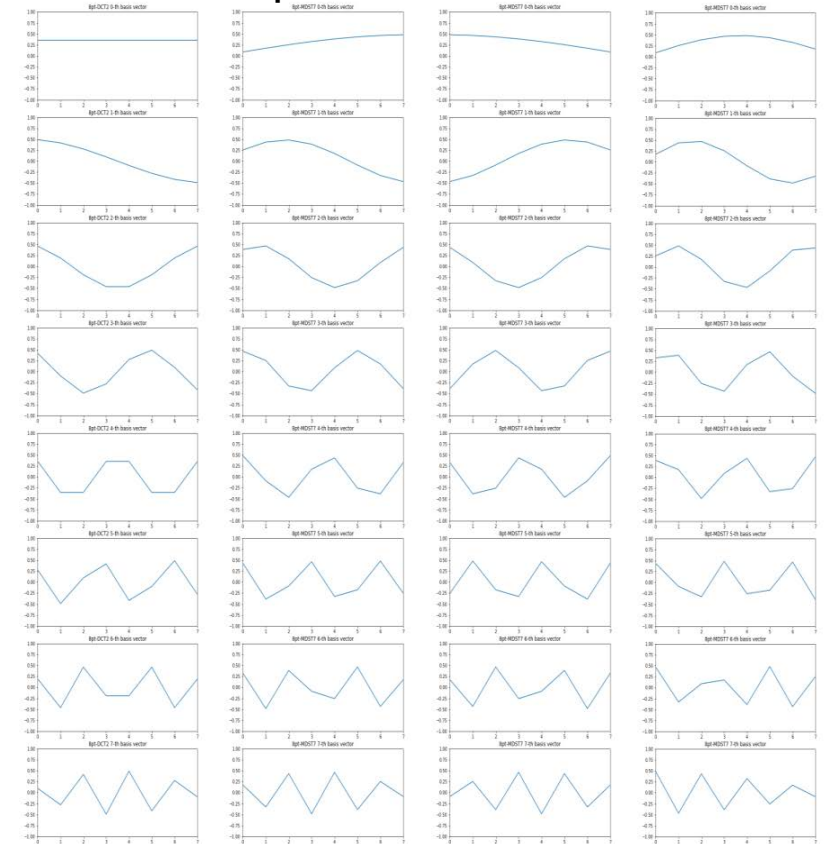
$$\text{DST7: } S_N^7[k, l] = \text{round} \left(2^m \frac{2}{\sqrt{2N+1}} \sin \left\{ \frac{2(l+1)\left(k + \frac{1}{2}\right)\pi}{2N+1} \right\} \right),$$

$$\text{FlipDST7: } \text{Flip}S_N^7[l, k] = S_N^7[l, N - k - 1]$$

$$\text{TrDST7: } \text{Tr}S_N^7[l, k] = S_N^7[k, l], \quad \text{for } k, l = 0, \dots, N-1$$

where δ_k is 0 for $k=0$ and 1 for otherwise, and m is the same as for JEM.

Example basis vectors N=8



DCT2

DST7

FlipDST7

TrDST7

5. Adaptive Multiple core Transforms for Luma and Chroma (1/2)

- Motivation:
 - In JEM, Adaptive Multiple core Transforms is used: DCT2, DCT5, DCT8, DST1, DST7 for luma, and DCT2 for chroma.
 - In general, multiple transforms can cover different residual characteristics and are expected to improve coding efficiency for not only luma but also chroma.
- Proposal
 - Apply AMT for both luma and chroma
 - Transform candidates are dependent on component and prediction modes.
 - An optimal transform type is chosen from the transform candidates by RDO search.

5. Adaptive Multiple core Transforms for Luma and Chroma (2/2)

- In JEM, transform index is signaled for H and V, separately.
 - In the response, the index is common between H and V.
 - $\text{trTypeH} = \text{TransformCandidates} [\text{prediction mode}][\text{index}][\text{H}]$
 - $\text{trTypeV} = \text{TransformCandidates} [\text{prediction mode}][\text{index}][\text{V}]$
- , where index is signaled except for inter chroma case. For inter chroma case, a new 1bit flag indicates that the index is the same as in collocated luma block or 0 (DCT2).

Mapping table for prediction mode to transform set

Intra Mode	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Transform set	0	1	5	3	5	3	5	3	5	3	5	3	5	3	7	7	7	7
Intra Mode	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	
Transform set	7	7	7	7	7	3	5	3	5	3	5	3	5	3	5	3	5	
Intra Mode	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
Transform set	2	4	2	4	2	4	2	4	2	4	2	6	6	6	6	6	6	6
Intra Mode	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67...	72	73	
Transform set	6	6	2	4	2	4	2	4	2	4	2	4	2	4	0	8		

For intra

For inter

Transform candidate per transform set

index	Transform Set (TrHxTrV)								
	0	1	2	3	4	5	6	7	8
0	DCT2xDCT2								
1	DST7xDST7	DST7xDST7	DST7xDST7	DST7xDST7	DST7xDST7	DST7xDST7	DST7xDST7	DST7xDST7	DST7xDST7
2	DST7xDCT2	DST7xDCT2	DCT2xDST7	DST7xDCT2	DCT2xDST7	DST7xDCT2	DCT2xDST7	DST7xDCT2	FlipDST7xDST7
3	DCT2xDST7	DCT2xDST7	FlipDST7xDST7	DST7xFlipDST7	FlipDST7xDST7	DST7xFlipDST7	FlipDST7xDST7	DST7xFlipDST7	DST7xFlipDST7
4	FlipDST7xDST7	FlipDST7xDST7	DST7xDCT2	DCT2xDST7	DST7xDCT2	DCT2xDST7	DST7xDCT2	DCT2xDST7	FlipDST7xFlipDST7
5	DST7xFlipDST7	DST7xFlipDST7	DST7xFlipDST7	FlipDST7xDST7	DST7xFlipDST7	FlipDST7xDST7	FlipDST7xDCT2	DST7xTrDST7	TrDST7xFlipDST7
6	DST7xTrDST7	DST7xTrDST7	DCT2xTrDST7	DST7xTrDST7	DCT2xTrDST7	DST7xTrDST7	DCT2xTrDST7	DCT2xTrDST7	DST7xDCT2
7	DCT2xTrDST7	DCT2xTrDST7	DST7xTrDST7	DCT2xTrDST7	DST7xTrDST7	DCT2xTrDST7	DST7xFlipDST7	FlipDST7xDST7	FlipDST7xDCT2
8	TrDST7xDST7	TrDST7xDST7	FlipDST7xDCT2	TrDST7xFlipDST7	FlipDST7xDCT2	TrDST7xFlipDST7	TrDST7xDST7	DCT2xFlipDST7	DCT2xDST7

Luma AMT for intra

Luma AMT for inter

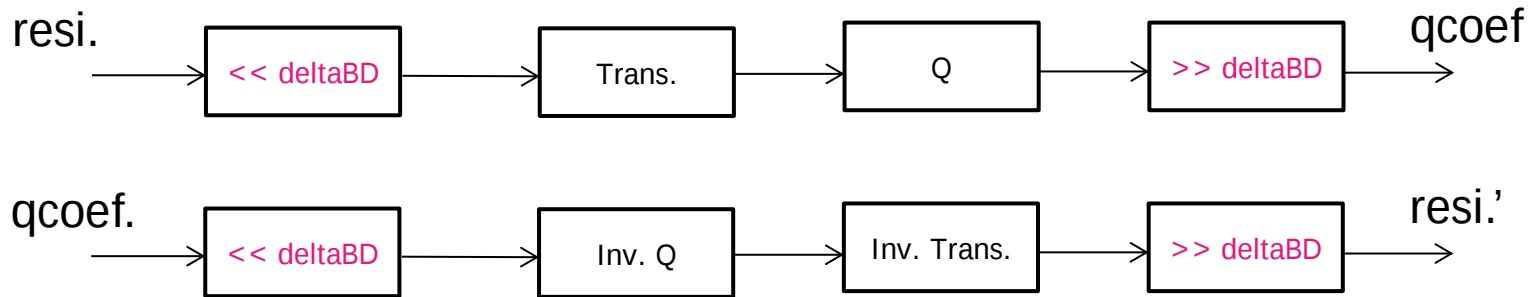
Chroma AMT for inter: collocated Luma index or DCT2

6. Adaptive Scaling for Transform and Quantization (1/2)

- Motivation
 - Actual dynamic range of pixel value is sometimes narrower than expected range, 1024/10bit.
 - Would like to extend calculation precision in transform and quantization using Adaptive Clipping information.
- Proposal
 - Rescale residues before transform and after quantization by an extra scaling factor δ_{BD} , which is derived using sequence-level bitdepth and slice-level min/max sample values signaled in slice header (Adaptive Clipping).

6. Adaptive Scaling for Transform and Quantization (2/2)

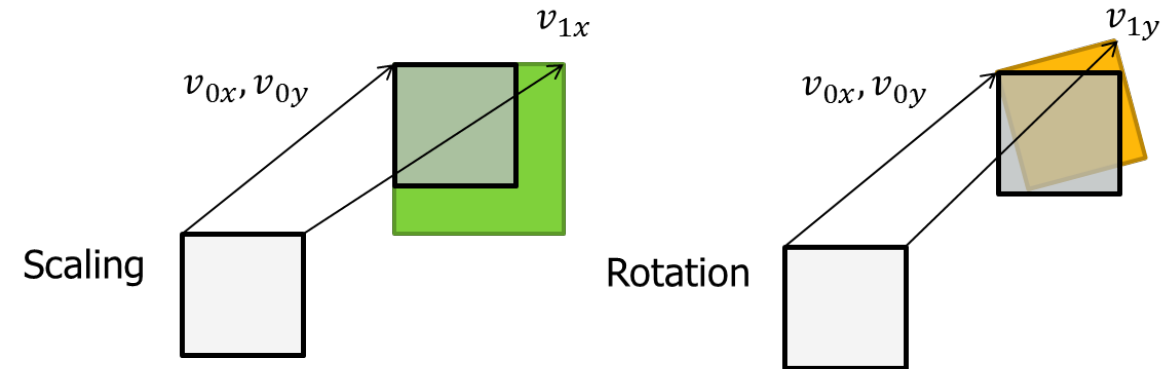
- The concept diagram of adaptive scaling for transform and quantization is as below:



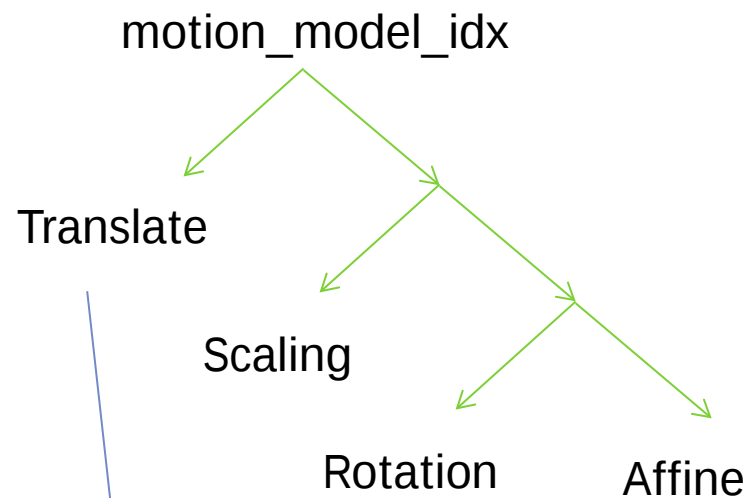
- The rescaling is given by shift operation.
- A scaling factor deltaBD is derived as below:
 - $\text{deltaBD} = \text{channelBitDepth} - \log_2\text{ResiRange}$
 - $\log_2\text{ResiRange} = \text{ceil}(\log_2(\text{maxSampleSignaled} - \text{minSampleSignaled}))$ in Adaptive Clipping

7. Affine MC (1/2)

- Motivation
 - Affine transform motion compensation can improve the quality of inter prediction
 - But there is a trade-off between the quality and an overhead
 - In the response
 - would like to reduce the overhead of motion information
 - And would like to add the option of sub-block size for any applications.
- Proposal
 - Define 3 parameter models, scaling mode and rotation mode, in addition to current affine mode
 - Adaptively use 3 or 4 parameter model.
 - Sub-block size for above modes can be changed in SPS



7. Affine MC (2/2)



8. Large CTU

- Motivation
 - For low bitrate applications, it's good to use larger CTU to reduce an overhead
 - An inter prediction was improved by using tools such as Affine MC and FRUC.
 - Large prediction unit will be efficient.
 - But, a large transform makes complexity higher.
- Proposal
 - Extend the CTU size up to 256x256.
 - To avoid large transform, the cbf is implicitly set to 0 when the size is greater than 128.

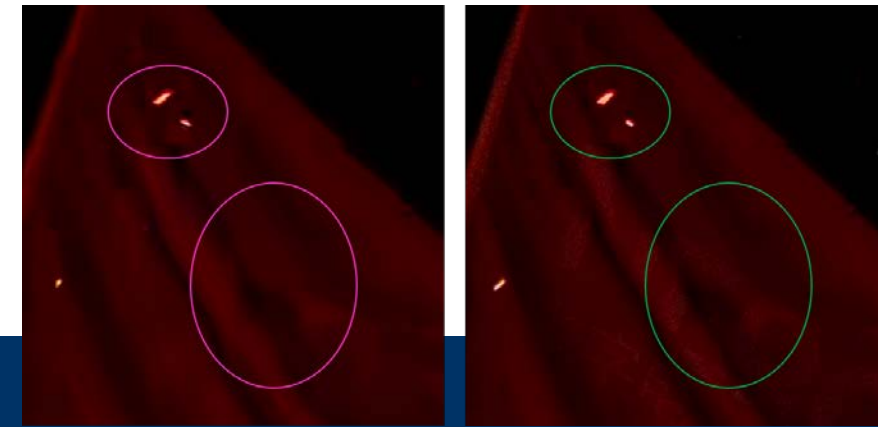
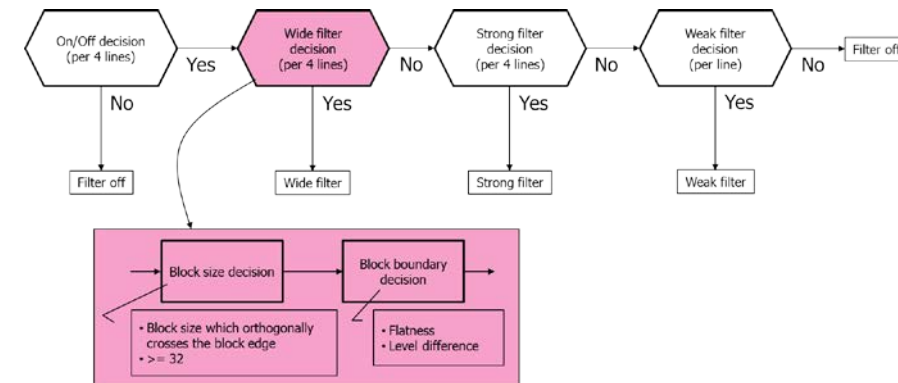
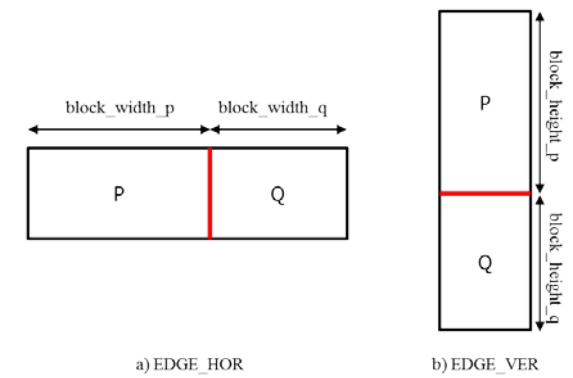
9. Extended Deblocking filter (1/2)

- Motivation
 - Blocking artifacts at large block boundaries can be more visible in both luma and chroma.
 - Those artifacts can be reduced by longer filter
- Proposal
 - In addition to current filter, a long filter is applied to both luma and chroma.
 - The extended filter can be applied to 7 samples in each side of the boundaries when the block width/height which orthogonally crosses the block edge is equal to or larger than 32.
 - The extended filter is also introduced to chroma which can be applied to 3 samples in each side.
 - The decisions are extended, accordingly.

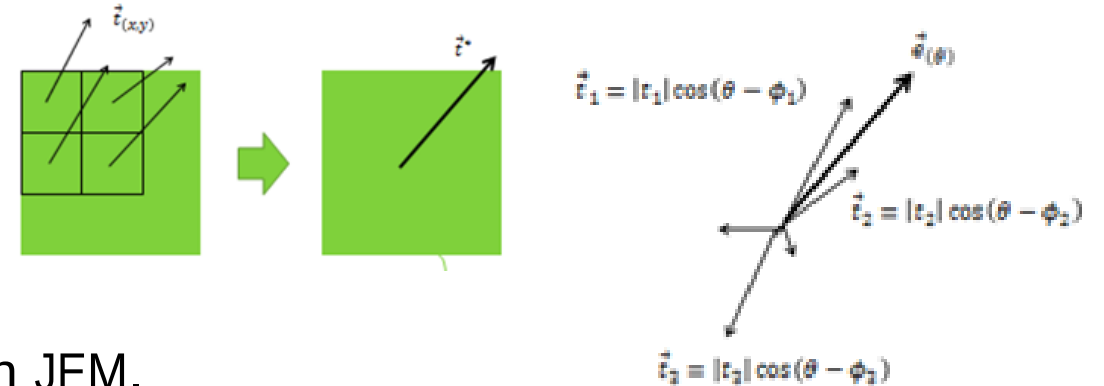
$$\begin{aligned}p6' &= (7*p7+2*p6+p5+p4+p3+p2+p1+p0+q0+8) \gg 4 \\p5' &= (6*p7+p6+2*p5+p4+p3+p2+p1+p0+q0+q1+8) \gg 4 \\p4' &= (5*p7+p6+p5+2*p4+p3+p2+p1+p0+q0+q1+q2+8) \gg 4 \\p3' &= (4*p7+p6+p5+p4+2*p3+p2+p1+p0+q0+q1+q2+q3+8) \gg 4 \\p2' &= (3*p7+p6+p5+p4+p3+2*p2+p1+p0+q0+q1+q2+q3+q4+8) \gg 4 \\p1' &= (2*p7+p6+p5+p4+p3+p2+2*p1+p0+q0+q1+q2+q3+q4+q5+8) \gg 4 \\p0' &= (p7+p6+p5+p4+p3+p2+p1+2*p0+q0+q1+q2+q3+q4+q5+q6+8) \gg 4\end{aligned}$$
$$\begin{aligned}p_2' &= (3*p_3+2*p_2+p_1+p_0+q_0+4) \gg 3 \\p_1' &= (2*p_3+p_2+2*p_1+p_0+q_0+q_1+4) \gg 3 \\p_0' &= (p_3+p_2+p_1+2*p_0+q_0+q_1+q_2+4) \gg 3\end{aligned}$$

9. Extended Deblocking filter (2/2)

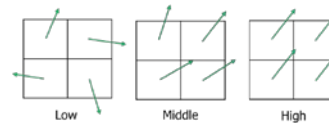
- Modified decision
 - Large block decision is introduced.
 - The thresholds are 32 and 16 for luma and chroma, respectively.
 - The strong filter decision is:
 - Extended to luma wide filter.
 - Introduced to chroma wide filter.
 - DF on/off and strong filter decision
- Chroma artifacts
 - JEM decoded image at R1.
 - Replaced luma domain with original in order to visualize chroma artifacts.



10. Modified ALF Classification



- Motivation
 - Classification is one of key areas in ALF.
 - Alternative feature is tested compared with ALF in JEM.
- Proposal
 - Horizontal and vertical gradient are calculated using Sobel filter.
 - Representative vector is derived so that the differences between representative and each vector are minimized in reference area (e.g. 6x6).
 - $\arg \max_{\theta} J(\theta) = \arg \max_{\theta} \sum_i \langle \vec{e}(\theta), \vec{t}_i \rangle = \arg \max_{\theta} \sum_i |t_i|^2 \cos(\theta - \phi_i)^2$
 - The similarity is calculated between the derived vector and each vector using the inner product and give three categories, low/middle/high, by threshold parameters, th1 and th2.



- Total number of class and the granularity of classification are the same as for JEM.

Compression Performance

SDR

	Constraint Set 1					
	Over HM16.16			Over JEM7.0		
	BD-rate(Y)	BD-rate(U)	BD-rate(V)	BD-rate(Y)	BD-rate(U)	BD-rate(V)
FoodMarket4	-34.68%	-45.90%	-49.51%	-2.13%	-4.09%	-5.16%
CatRobot1	-41.04%	-55.19%	-48.40%	-2.98%	-7.08%	-5.53%
DaylightRoad2	-42.29%	-58.79%	-44.14%	-3.37%	-4.60%	-5.51%
ParkRunning3	-32.08%	-27.25%	-30.22%	-1.43%	-2.89%	-4.00%
Campfire	-35.92%	-36.87%	-52.35%	-2.41%	-3.35%	-5.66%
BQTerrace	-32.69%	-47.71%	-56.38%	-3.38%	-4.89%	-4.68%
RitualDance	-28.65%	-39.56%	-42.90%	-1.86%	-4.26%	-4.09%
MarketPlace	-30.75%	-49.97%	-47.38%	-2.26%	-5.50%	-5.35%
BasketballDrive	-32.57%	-48.48%	-46.09%	-2.09%	-5.82%	-4.85%
Cactus	-36.34%	-48.34%	-44.57%	-2.17%	-5.99%	-6.57%
Average SDR-A	-37.20 %	-44.80 %	-44.93 %	-2.46 %	-4.40 %	-5.17 %
Average SDR-B	-32.20 %	-46.81 %	-47.46 %	-2.35 %	-5.29 %	-5.11 %
Average all	-34.70 %	-45.81 %	-46.19 %	-2.41 %	-4.85 %	-5.14 %
Enc Time[%]	3012%			378%		
Dec Time[%]	995%			131%		

	Constraint Set 2					
	Over HM16.16			Over JEM7.0		
	BD-rate(Y)	BD-rate(U)	BD-rate(V)	BD-rate(Y)	BD-rate(U)	BD-rate(V)
BQTerrace	-29.17%	-49.07%	-52.91%	-4.14%	-4.36%	-4.06%
RitualDance	-22.94%	-38.29%	-39.74%	-1.60%	-6.28%	-6.70%
MarketPlace	-21.78%	-49.29%	-47.78%	-1.40%	-11.40%	-11.33%
BasketballDrive	-26.89%	-48.56%	-46.60%	-1.74%	-5.36%	-4.96%
Cactus	-29.92%	-45.91%	-44.92%	-2.37%	-6.30%	-9.67%
Average SDR-B	-26.14%	-46.22%	-46.39%	-2.25%	-6.74%	-7.34%
Enc Time[%]	2286%			284%		
Dec Time[%]	851%			128%		

HDR (PSNR based BD-rate)

	Constraint Set 1					
	Over HM16.16			Over JEM7.0		
	BD-rate			BD-rate		
	psnrY	psnrU	psnrV	psnrY	psnrU	psnrV
DayStreet	-35.74%	-55.97%	-68.32%	-2.60%	-5.03%	-6.83%
PeopleInShoppingCenter	-29.50%	-57.69%	-55.65%	-1.84%	-4.53%	-4.51%
SunsetBeach	-24.04%	-77.84%	-82.33%	-2.61%	-5.88%	-11.87%
Market	-24.75%	-40.94%	-38.73%	-1.72%	-3.05%	-6.36%
ShowGirl2	-28.00%	-66.68%	-40.59%	-3.14%	-23.37%	-24.36%
Hurdles	-31.36%	-53.69%	-50.94%	-0.73%	1.46%	0.86%
Starting	-29.73%	-32.90%	-26.72%	-1.49%	-1.47%	-5.17%
Cosmos1	-24.11%	-68.19%	-80.92%	-1.82%	-7.00%	-9.44%
Average HDR-A	-29.76%	-63.83%	-68.77%	-2.35%	-5.14%	-7.73%
Average HDR-B	-27.59%	-52.48%	-47.58%	-1.78%	-6.69%	-8.89%
Average all	-28.40%	-56.74%	-55.53%	-1.99%	-6.11%	-8.46%
Enc Time[%]	2243%			281%		
Dec Time[%]	911%			120%		

Summary of Proposal

- 10 new tools in 6 categories to enhance coding efficiency on top of JEM.
 1. Sign Prediction
 2. Use of multiple reference samples in intra prediction
 3. Modified PDPC Planar
 4. Transform Matrix Replacement
 5. Adaptive Multiple core Transforms for Luma and Chroma
 6. Adaptive Scaling for Transform and Quantization
 7. Affine MC
 8. Large CTU
 9. Extended Deblocking filter
 10. Modified Adaptive Loop Filter (ALF) Classification
- Common design between SDR and HDR.
- SDR: -2.41%, -4.85% and -5.14% for Y, U and V, respectively, and 378% and 131% for encoding and decoding time.
- HDR: -1.99%, -6.11% and -8.46% for Y, U and V, respectively, and 281% and 120% for encoding and decoding time.