JCTVC-A121

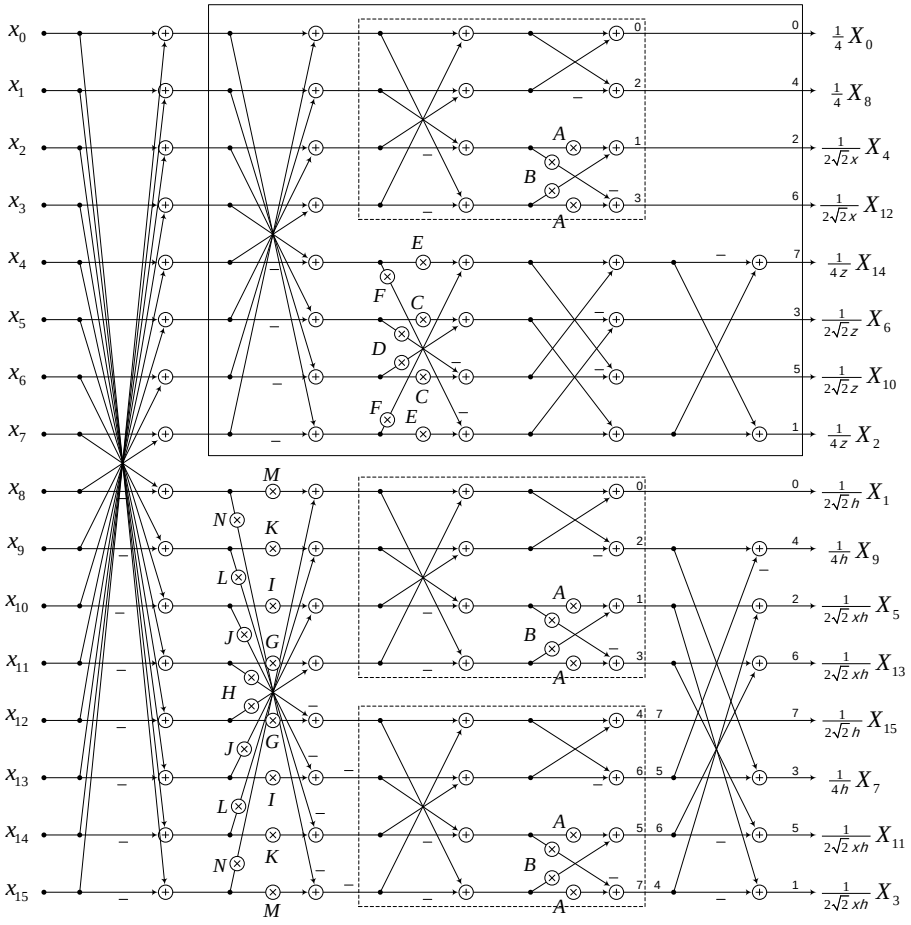
Marta Karczewicz, Peisong Chen,
Rahul Panchal, Wei-Jung-Chien, Shawn Wang,
Rajan Joshi

Overview

- JMKTA based, used tools:
  - Coding Using Extended Block Sizes (modified);
  - Motion Vector Competition (MVC);
  - Mode Dependent Directional Transform (MDDT);
  - Internal Bit Depth Increase (IBDI);
  - Switched Filters with Offset (modified);
  - In-Loop Filter (modified);
- Other additions, modifications:
  - Geometry partitioning;
  - Adaptive motion vector resolution;
  - Encoder simplifications.

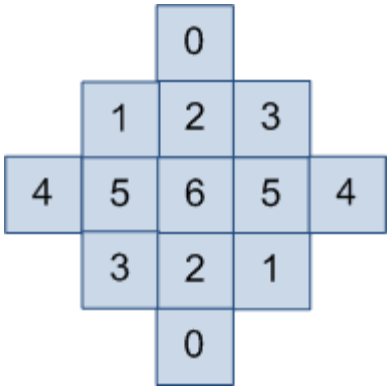
Extended Block Sizes & Large Transforms

- For partitions above 16x16 only 16x16 transform is used.
- Modified 16x8 and 16x16 transforms (VCEG-AL30).
  - Based on LLM factorization
  - Supporting a simple recursive factorization structure leading to faster implementation.
    - 16-point transform: 36 multiplications and 72 additions
  - Orthogonal after appropriate scaling has been applied.



In-Loop Filter

- Combination of two adaptive filter, post-filter and Qudtree-Based Adaptive Loop Filter (QALF) included in JMKTA (the combination described in VCEG-L27).
- QALF - quad-tree based block partitioning used to signal whether the luminance component of a block is filtered.
- Up to 16 filters can be used. For the filtered blocks, which filter is used for a given pixels depends on the value of Sum-Modified Laplacian.
- Symmetry of coefficients used to reduce overhead.

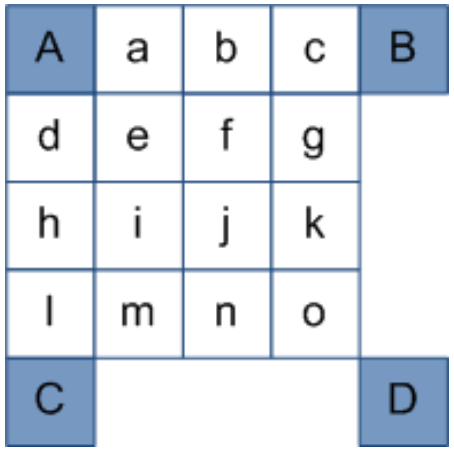


- Filter coefficients predicted from the coefficients used for the previous frames

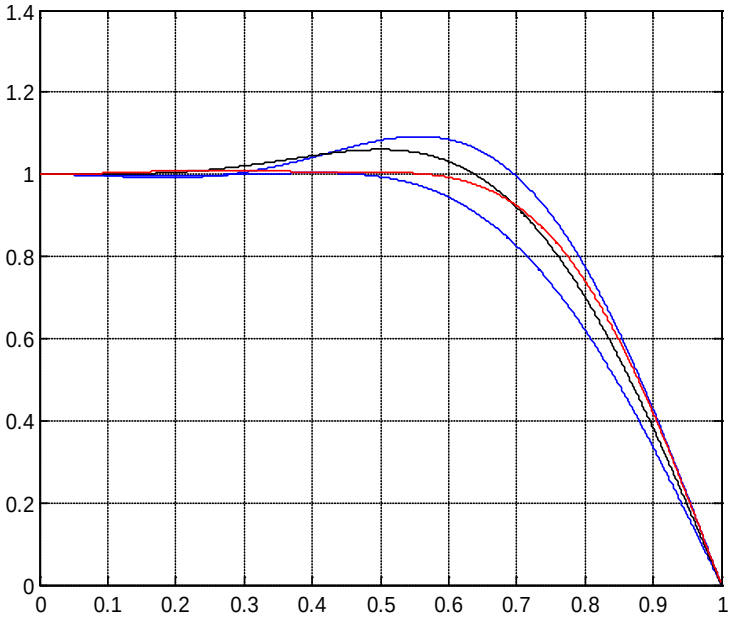
$$SML(i,j) = |2R(i,j) - R(i-1,j) - R(i+1,j)| + |2R(i,j) - R(i,j-1) - R(i,j+1)|$$

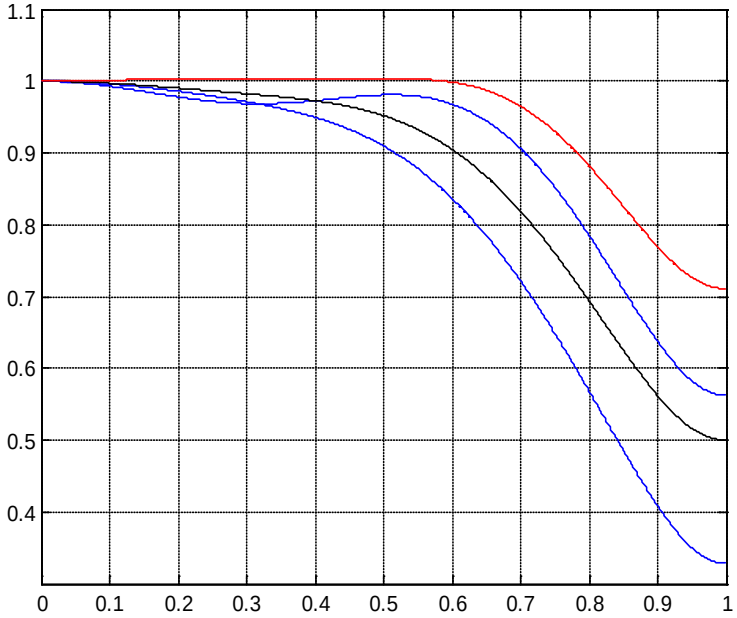
Switched Interpolation Filters With Offset

- Set of 4 fixed filters used:
  - 6-tap, 8-tap, and 2 sets each consisting of 4x4 non-separable filter for sub-pels e, f, g, I, j, k, m, n, o and 6-tap filter for sub-pels a, b, c, d, h, l.
- Each sup-pel position can use different filter (signaled per slice).
  - The filter selection is based on statistics gather for previously encoded frames.
- Each sup-pel position can have different value of DC offset assigned (signaled per slice).
  - Differences between DC values of the current and the reference frame for the entire frame (*DCframe*) and for each MB (*DCmb*) used.
  - If |*DCframe*|>*0* and |*DCframe*|<*1.5* number of sub-pel positions having offset of magnitude 1 is equal to *round(10·DCframe)*.
  - If |DCframe|>1.5 the sup-pel offsets have values from *DCmin* to *DCmax*. *DCmin* (*DCmax*) is the smallest (largest) value of *DCmb* assigned to least *n* MBs.



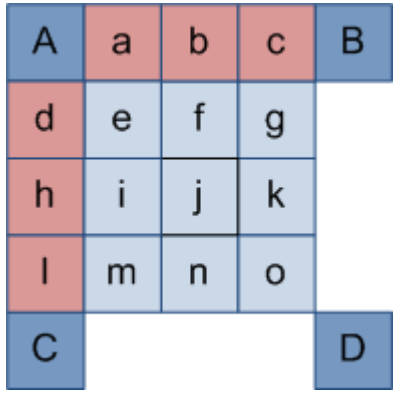
Switched Interpolation Filters With Offset

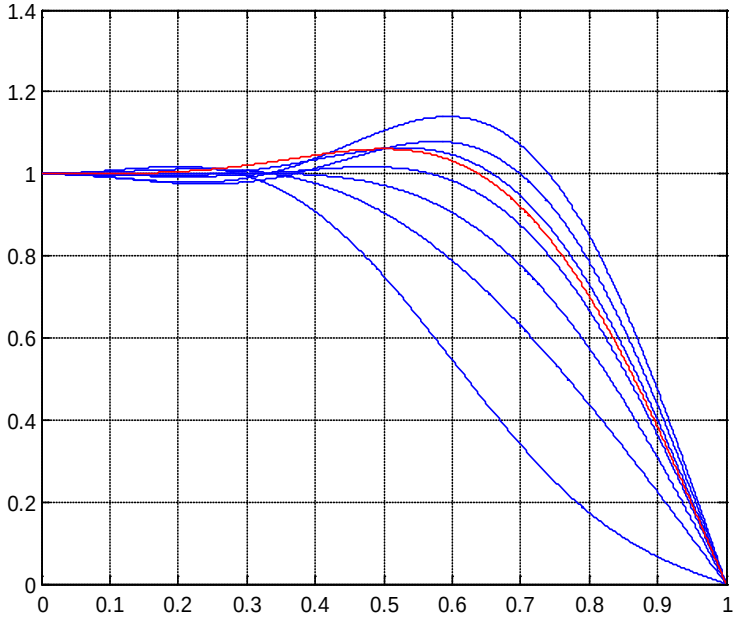


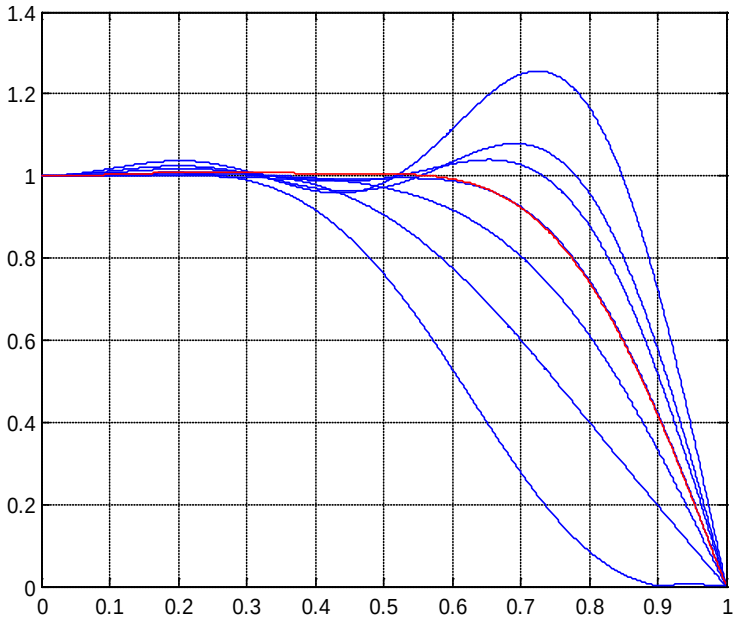


Simplified Switched Interpolation Filters With Offset

- Only separable filters used:
  - 8 6-tap filters,
  - 8 8-tap filters.
- 8-tap filters used only for sub-pel positions requiring only 1-D filtering.
- Loss: 0.7% on average.

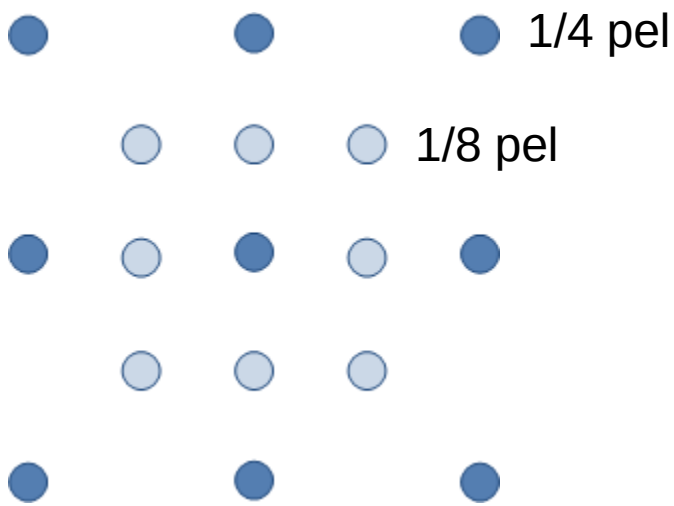




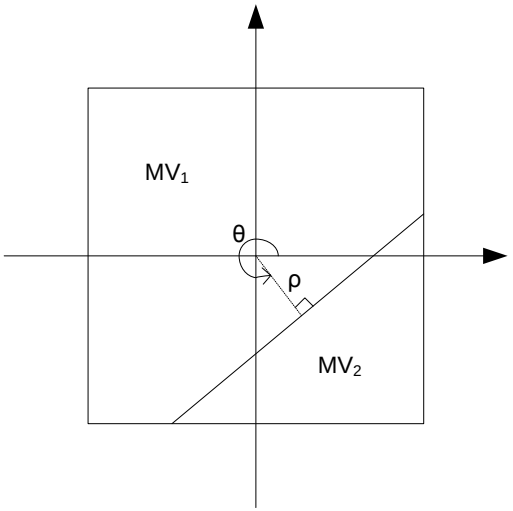


Motion Compensation

- Both 1/4 and 1/8 pel motion precision used, motion precision signaled per motion vector.
- MV prediction for the current block is formed with 1/8th pixel accuracy. If the current block has only 1/4th pixel motion accuracy, the MV prediction is converted to 1/4th pixel accuracy by right-shifting.
- 1/8 pel position obtained using bilinear interpolation of ¼ and ½ sub-pel positions (implemented through direct interpolation).



Geometry Partitioning



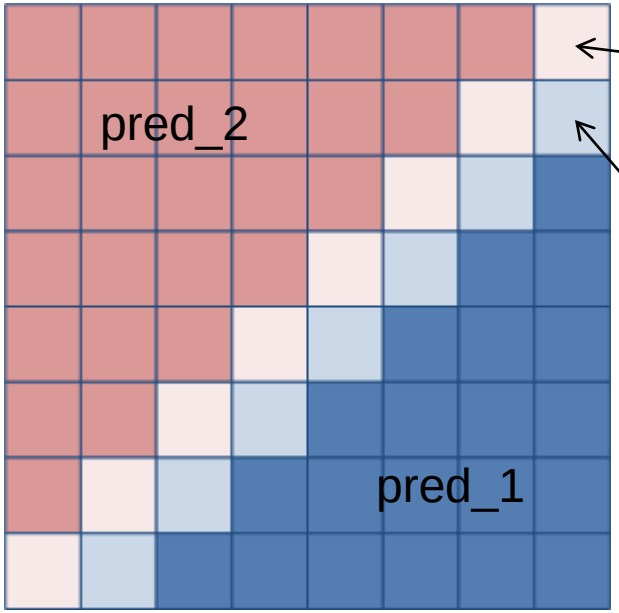
- Geometry used for:
  - 16x16, 32x32 and 64x64 blocks.
- Signaling:
  - Additional bit for 8x16,16x32, 32x64 modes.
- Division of the block:

$$y = ax + b = \frac{-1}{\tan(J)}x + \frac{r}{\sin(J)}, \text{where } \frac{-1}{\tan(J)} \text{ and } \frac{r}{\sin(J)} \text{ tabularized.}$$

- ρ values:
  - 16x16: 0 to 7,
  - 32x32: 0 to 15,
  - 64x64: 0 to 31.
- θ - 0 to 360 degree with step size of 11.25 degree (32 values).
- Signaling: fixed number of bits used for ρ and θ values, e.g., for 16x16 block 3 bits for distance and 5 bits for angle.

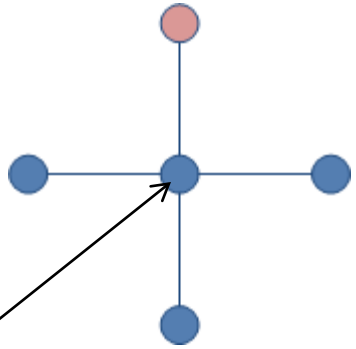
Geometry Partitioning

- Transform size:
  - 64x64 and 32x32 partition – 16x16.
  - 16x16 partition – 4x4, 8x8 and 16x16 (signaled).
- OBMC used during MC prediction.



1/3*pred_1+2/3*pred_2

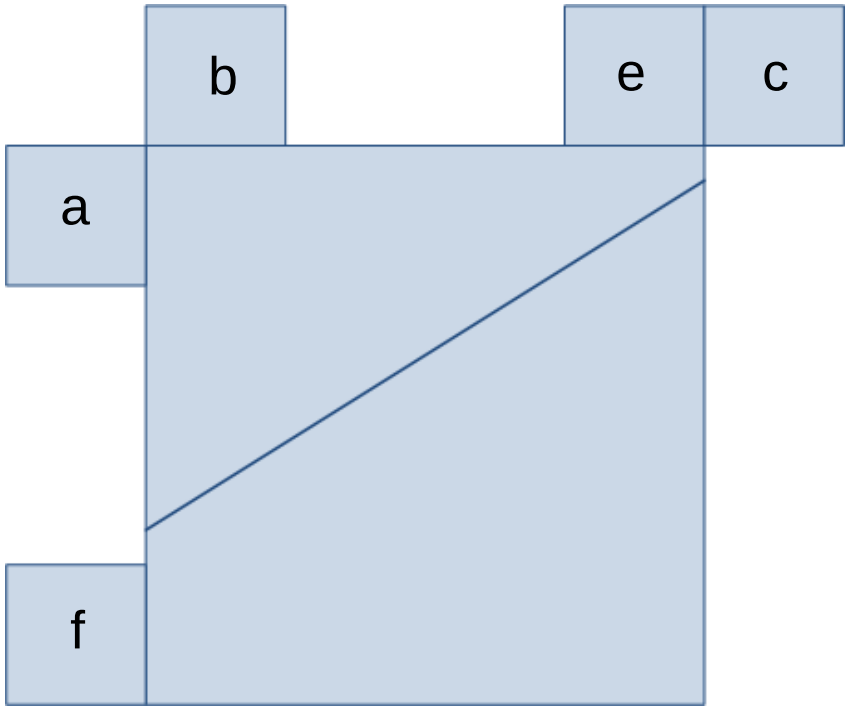
2/3*pred_1+1/3*pred_2



4-point connectivity used to determine edge pixel

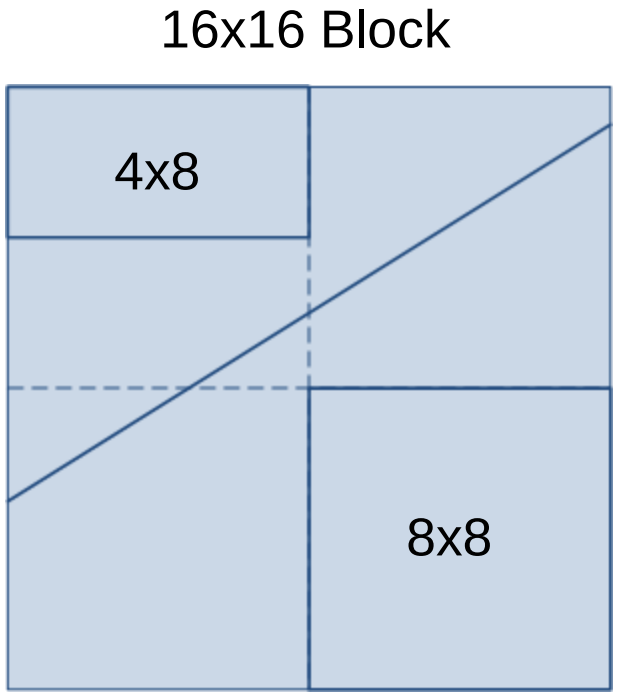
Geometry Partitioning – MV Prediction

```
if(a && b && c)
    return median(a, b, c);
else if(a)
    return mv_a;
else if(b)
    return mv_b;
else if(e)
    return mv_e;
else if(f)
    return mv_f;
else if(c)
    return mv_c;
else
    return median(a, b, c);
```



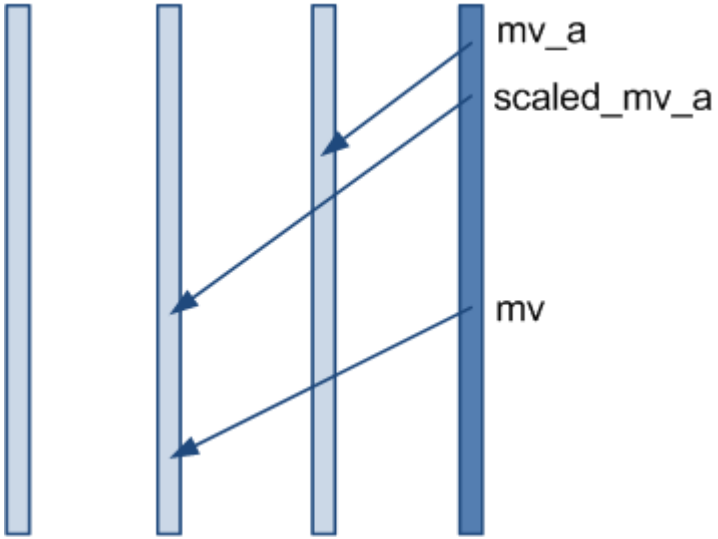
Geometry Partitioning - Encoder

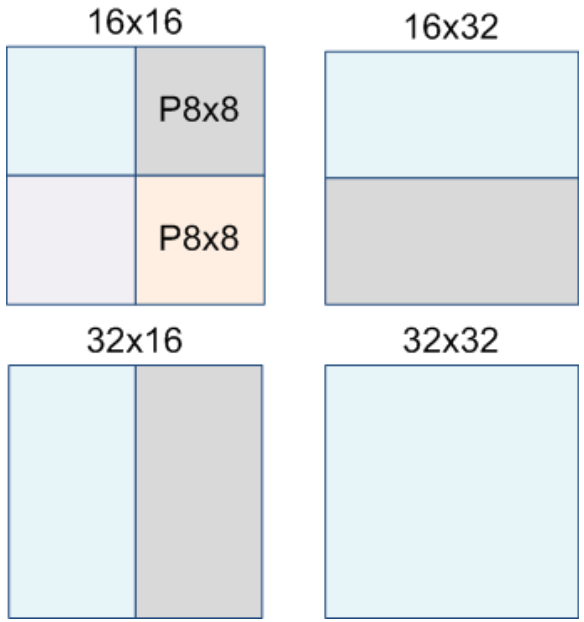
- Motion vectors from block matching are reused to find initial set of best partitions.
- For the selected partitions mv are estimated using EPZS.
- 16x16 blocks:
  - 16 best partitions selected.
- 32x32 and 64x64 blocks:
  - Subsample both angle θ and distance ρ by 2.
  - Select 2 best partitions using mv obtained by block matching.
  - Estimated motion for 18 partitions – 2 selected ones and their neighboring ones(+/- 11.25 degrees for θ and +/- 1 for ρ).
- Using refinement (from +/- 1 to +/- 1/8 pel) instead of EPZS introduces loss <0.5%.



Other Modifications

- Addition of P_Direct mode (zero motion, non-zero cbp).
- Signaling of MB modes for B frames.
  - Partition (16x16, 16x8, etc.) and list (L0, L1, Bi) information signaled separately.
- Scaled motion vector prediction.
- Encoder simplifications.
  - Direct and Skip modes calculated. If Skip chosen no further modes evaluated.
  - If 2 NxN partitions are further divided, larger partitions including these NxN partitions, e.g., 2Nx2N, not checked.





	JM16.2 HierP	Proposal Low Rates	Proposal High Rates	Proposal QP=22,27,32,37	Proposal Bug Fixes QP=22,27,32,37
1080p	Kimono	-40.42	-44.78	-41.12	-40.89
	ParkScene	-29.48	-26.28	-27.75	-27.98
	Cactus	-31.71	-31.57	-31.02	-31.77
	Basketball	-41.39	-40.69	-39.40	-39.86
	BQSquare	-45.58	-50.43	-45.26	-45.80
	Avg_1080p	**-37.72**	**-38.75**	**-36.91**	**-37.26**
WVGA	Basketball	-28.92	-28.35	-29.21	-30.28
	BQSquare	-31.92	-30.87	-31.75	-32.13
	PartyScene	-28.54	-24.10	-23.27	-23.66
	RaceHorses	-27.96	-26.76	-29.55	-29.74
	Avg_WVGA	**-29.34**	**-27.52**	**-28.45**	**-28.95**
WQVGA	Basketball	-23.29	-22.70	-24.57	-24.79
	BQSquare	-34.85	-32.96	-32.49	-32.76
	PartyScene	-15.42	-16.66	-18.95	-19.45
	RaceHorses	-20.44	-20.27	-24.71	-24.82
	Avg_WQVGA	**-23.50**	**-23.15**	**-25.18**	**-25.46**
720p	Vidyo1	-46.57	-46.24	-38.78	-39.76
	Vidyo3	-40.15	-42.55	-37.39	-38.98
	Vidyo4	-37.30	-45.31	-38.90	-39.87
	Avg_WVGA	**-41.34**	**-44.70**	**-38.36**	**-39.54**
	Overall Avg	**-32.75**	**-33.16**	**-32.13**	**-32.66**

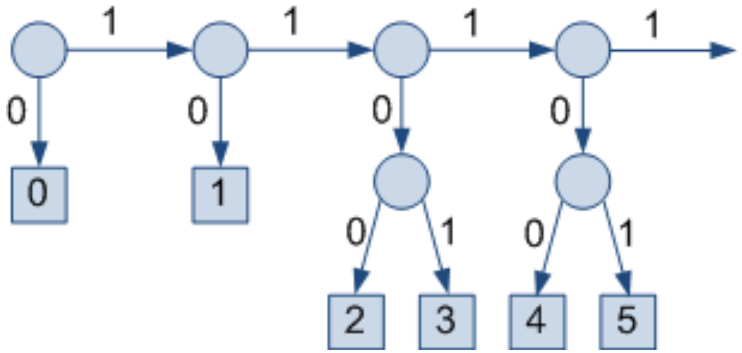
	JM16.2 HierB	Proposal Low Rates	Proposal High Rates	Proposal QP=22,27,32,37	Proposal Bug Fixes QP=22,27,32,37
4kx2k	Traffic	-33.23	-31.15	-33.23	
	PeopleOnStreet	-19.55	-20.33	-20.33	
	Avg_4kx2k	**-26.39**	**-25.74**	**-26.78**	
1080p	Kimono	-39.38	-39.79	-43.65	-43.68
	ParkScene	-29.00	-26.58	-32.15	-32.21
	Cactus	-31.24	-30.62	-32.99	-33.08
	Basketball	-35.67	-34.90	-38.86	-39.01
	BQSquare	-34.45	-45.36	-37.53	-37.63
	Avg_1080p	**-33.95**	**-35.45**	**-37.04**	**-37.12**
WVGA	Basketball	-30.60	-30.69	-29.95	-30.38
	BQSquare	-33.46	-31.84	-30.67	-30.81
	PartyScene	-33.16	-32.10	-32.74	-32.79
	RaceHorses	-29.69	-26.81	-24.93	-24.98
	Avg_WVGA	**-31.73**	**-30.36**	**-29.57**	**-29.74**
WQVGA	Basketball	-22.81	-21.73	-21.48	-21.61
	BQSquare	-43.55	-44.41	-41.28	-41.35
	PartyScene	-26.32	-27.08	-25.59	-25.72
	RaceHorses	-21.05	-19.83	-20.92	-20.96
	Avg_WQVGA	**-28.43**	**-28.26**	**-27.32**	**-27.41**
	Overall Avg	**-30.88**	**-30.88**	**-31.09**	**-31.86**

Simplifications – Decoder Complexity

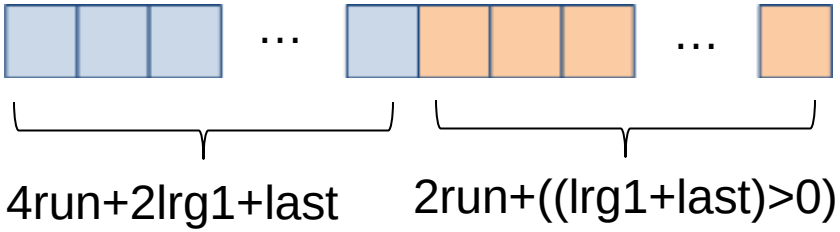
- Removed IBDI.
  - Requires 12 instead of 8 bit storage of the reference frame, increased memory bandwidth requirements.
- Set maximum size for in-loop filter to be 5x5, removed chroma filtering.
- Replaced CABAC with VLC, VLC modifications comparing to AVC:
  - Reference frame index (refIdx) combined with signaling of mv resolution. In case of VLC 1/8 resolution used only when refIdx=0.
  - CBP:
    - For 64x64 and 32x32 partitions, single bit sent signaling is the entire partition coded or not-coded.
    - For 16x16 partitions:
      - some_luma_blocks_coded + 2nc (nc defined as in AVC);
      - transform size;
      - 4 (2) bits signaling which 8x8 (16x8) blocks coded.

Simplifications – Decoder Complexity

- VLC modifications comparing to AVC (cont.):
  - Transform coefficients coding using:
    - 3D VLCs: 4run+2lrg1+last, if run<run_{max};
    - 2D VLCs: 2run+((lrg1+last)>0).
  - VLC for each context signaled per slice, based on previously gathered statistic.
    - Parameterized tree codes – extension of Golomb codes.
    - Contexts - scan position dependent.
  - 3D VLCs – index assigned to a symbols can be modified during encoding.



m=1, d=1, w=2



	JM16.2 HierB	Proposal QP=22,27,32,37	Proposal no IBDI	Proposal no IBDI 5x5 Loop Filt
1080p	Kimono	-42.69	-41.61	-41.55
	ParkScene	-31.17	-30.86	-30.41
	Cactus	-32.15	-31.25	-31.05
	Basketball	-38.06	-37.24	-36.91
	BQSquare	-36.42	-34.26	-33.10
	Avg_1080p	**-36.10**	**-35.04**	**-34.60**
WVGA	Basketball	-29.67	-28.48	-28.34
	BQSquare	-29.81	-29.50	-28.82
	PartyScene	-32.13	-31.28	-29.84
	RaceHorses	-24.72	-23.80	-23.14
	Avg_WVGA	**-29.08**	**-28.27**	**-27.54**
WQVGA	Basketball	-21.04	-20.86	-20.29
	BQSquare	-40.82	-39.40	-36.52
	PartyScene	-25.06	-24.93	-23.94
	RaceHorses	-20.49	-19.85	-18.96
	Avg_WQVGA	**-26.85**	**-26.26**	**-24.93**
	Overall Avg	**-31.09**	**-30.26**	**-29.45**

	JM16.2 HierP	Proposal QP=22,27,32,37	Proposal no IBDI	Proposal no IBDI 5x5 Loop Filt
1080p	Kimono	-40.89	-40.08	-39.20
	ParkScene	-27.98	-25.89	-25.80
	Cactus	-31.77	-29.94	-29.81
	Basketball	-39.86	-38.17	-37.13
	BQSquare	-45.80	-43.58	-41.42
	Avg_1080p	**-37.26**	**-35.53**	**-34.67**
WVGA	Basketball	-30.28	-28.95	-27.27
	BQSquare	-32.13	-30.61	-30.29
	PartyScene	-23.66	-22.39	-21.41
	RaceHorses	-29.74	-28.92	-28.04
	Avg_WVGA	**-28.95**	**-27.72**	**-26.75**
WQVGA	Basketball	-24.79	-23.49	-23.49
	BQSquare	-32.76	-32.06	-29.65
	PartyScene	-19.45	-18.10	-17.87
	RaceHorses	-24.82	-23.90	-23.52
	Avg_WQVGA	**-25.46**	**-24.39**	**-23.63**
720p	Vidyo1	-39.76	-36.08	-36.00
	Vidyo3	-38.98	-35.32	-34.95
	Vidyo4	-39.87	-36.00	-35.50
	Avg_WVGA	**-39.54**	**-35.80**	**-35.48**
	Overall Avg	**-32.66**	**-30.84**	**-30.08**

		JM16.2 IPPP High Profile	Proposal QP=22,27,32,37	Proposal no IBDI 5x5 Loop Filt	Proposal no IBDI 5x5 Loop Filt no MVC VLC	JM16.2 IPPP High Profile VLC
1080p	Kimono	-36.13	-34.17	-28.75	25.48
	ParkScene	-25.06	-22.61	-19.55	12.59
	Cactus	-27.41	-25.25	-20.61	14.20
	Basketball	-33.59	-30.38	-24.87	17.37
	BQSquare	-36.75	-33.18	-25.67	4.60
	Avg_1080p	**-31.79**	**-29.12**	**-23.89**	**14.85**
WVGA	Basketball	-28.36	-24.87	-21.38	11.53
	BQSquare	-26.47	-24.09	-21.07	10.70
	PartyScene	-25.47	-23.14	-19.13	8.76
	RaceHorses	-19.79	-17.73	-15.07	8.56
	Avg_WVGA	**-25.02**	**-22.46**	**-19.16**	**9.89**
WQVGA	Basketball	-17.99	-16.56	-13.20	10.55
	BQSquare	-38.55	-35.34	-27.65	11.40
	PartyScene	-19.72	-18.10	-15.78	8.92
	RaceHorses	-15.20	-13.59	-10.11	9.04
	Avg_WQVGA	**-22.87**	**-20.90**	**-16.69**	**9.98**
720p	Vidyo1	-37.72	-33.36	-28.74	12.03
	Vidyo3	-34.72	-30.27	-26.90	14.43
	Vidyo4	-37.26	-32.52	-27.79	12.08
	Avg_WVGA	**-36.57**	**-32.05**	**-27.81**	**12.85**
	Overall Avg	**-28.76**	**-25.95**	**-21.64**	**12.02**

Decoder Complexity

IPPP					
Vidyo3 720p 150 frames					
	QP22	QP27	QP32	QP37	Avg Time
JM16.2	10.45	8.03	8.22	8.03	**8.68**
JM KTA - AVC	23.59	22.52	22.28	21.67	**22.52**
JM KTA	40.70	38.17	36.14	32.27	**36.82**
Proposal	21.69	19.39	17.32	15.04	**18.36**
Proposal (Simplified)	16.05	14.23	13.08	12.88	**14.06**

HierB					
Vidyo3 720p 145 frames					
	QP22	QP27	QP32	QP37	Avg Time
JM16.2	10.50	9.57	9.48	9.30	**9.71**
JM KTA - AVC	29.06	29.66	28.22	30.81	**29.44**
JM KTA	45.53	43.06	40.92	35.67	**41.30**
Proposal	24.71	22.35	20.75	20.44	**22.06**
Proposal (Simplified)	19.78	17.78	15.90	18.88	**18.09**

Encoder Complexity

IPPP					
Vidyo3 720p 150 frames					
	QP22	QP27	QP32	QP37	Avg Time
JM16.2	1624	1613	1761	2000	**1749**
JM KTA - AVC	3714	4765	5939	5845	**5066**
JM KTA	28913	27096	26091	30050	**28038**
QTM w/ Geom	12304	10583	9169	8094	**10038**
QTM w/ Geom Simplified	8216	7304	6727	6244	**7123**

HierB					
Vidyo3 720p 145 frames					
	QP22	QP27	QP32	QP37	Avg Time
JM16.2	3981	2659	3842	3443	**3481**
JM KTA - AVC	13518	13367	12201	11915	**12750**
JM KTA	43797	43578	42969	42103	**43112**
QTM w/ Geom	16668	14301	12222	11407	**13650**
QTM w/ Geom Simplified	10269	9407	8648	8293	**9154**