Hierarchical Variable-sized Block Transform

JCTVC-B050

Bumshik Lee and Munchurl Kim
Korea Advanced Institute of Science and Technology (KAIST)

Hui Yong Kim, Jongho Kim and Jin Soo Choi
Electronic Telecommunications Research Institute (ETRI)

Spatial Analysis for Transform Type Selections

- Correlation coefficient between neighboring pixel values for predicted residuals

$$r(\tau) = E[x(n)x(n+\tau)] / \sigma^2$$

$r(1)$: pixel by pixel correlation coefficient

$r(2)$: every other pixels

Sequences	Resolution	QP	$r(1)$	$r(2)$	Proportions of transform types for non-SKIP blocks	
					8×8(%)	4×4(%)
City	QVGA (320×240)	23	0.195	-0.010	45.52	54.48
		27	0.249	-0.019	55.76	44.24
		33	0.368	0.008	61.71	38.29
	VGA (640×480)	23	0.194	-0.040	52.74	47.26
		27	0.247	-0.054	55.20	44.80
		33	0.341	-0.033	58.18	41.82
	720P (1280×720)	23	0.451	-0.067	78.79	21.21
		27	0.486	-0.048	71.10	28.90
		33	0.541	0.001	65.05	34.95
Bigships	QVGA (320×240)	23	0.308	0.018	50.80	49.20
		27	0.358	0.014	56.62	43.38
		33	0.452	0.048	60.00	40.00
	VGA (640×480)	23	0.388	0.016	53.80	46.20
		27	0.437	0.039	57.49	42.51
		33	0.489	0.063	61.09	38.91
	720 (1280×720)	23	0.542	0.143	58.04	45.43
		27	0.610	0.217	61.95	41.59
		33	0.658	0.269	65.11	36.55
ShuttleStart	QVGA (320×240)	23	54.57	0.059	54.57	45.43
		27	58.41	0.137	58.41	41.59
		33	63.45	0.178	63.45	36.55
	VGA (640×480)	23	0.437	0.064	53.30	46.70
		27	0.508	0.136	58.27	41.73
		33	0.554	0.176	64.96	35.04
	720 (1280×720)	23	0.627	0.184	66.90	33.10
		27	0.684	0.274	60.19	39.81
		33	0.726	0.336	64.84	35.16

Spatial Characteristics for Various Input Signals

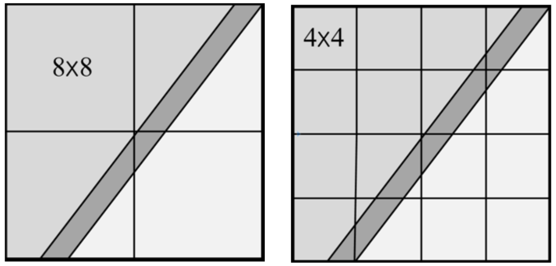
- According to the input texture
  - As the motion increases, *r* (pixel correlation) gets smaller
- According to the QP
  - For large QPs, *r* gets larger
- According to the spatial resolution
  - For higher resolutions, *r* gets larger

$\longrightarrow$ We can guess that large transform has advantage for the small motion, larger QP and higher resolution

$\longrightarrow$ Otherwise, large transform itself is not advantageous

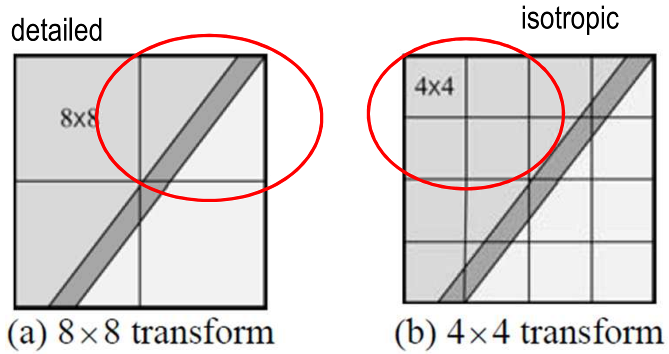
Limitation of single type transform

- Single Type Transform in H.264|MPEG-4 AVC



example

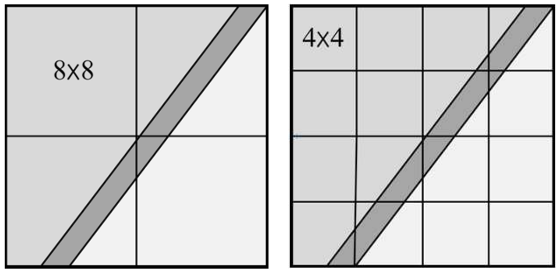
- Limitation of the structure



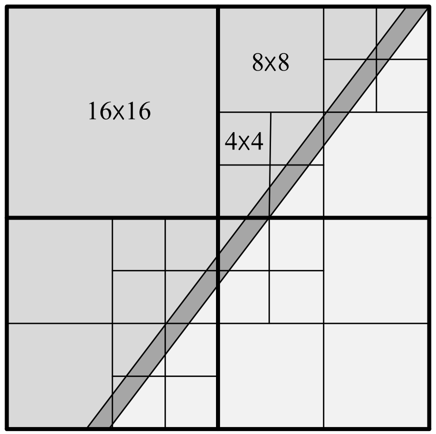
-Inefficient to adapt to the changing local characteristics
-Side information can be saved

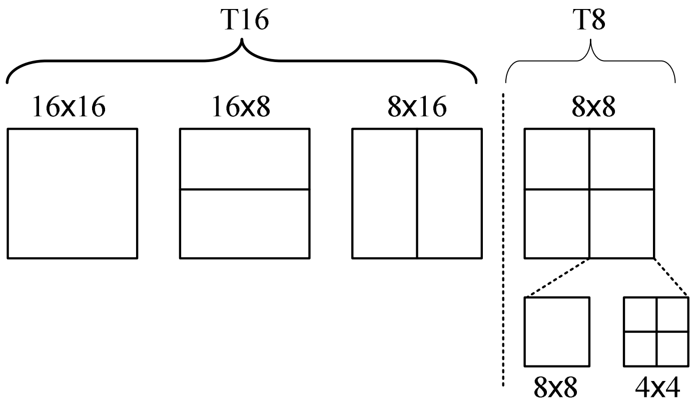
Hierarchical Variable-sized Block Transform (HVBT)

- Previous standard (H.264/AVC, WMV-9)



- Proposed Structure with the Hierarchically Variable Transform Blocks





Order-16 ICT kernel in the HVBT

$$
\mathbf{T}_{\text{ICT, 16}} = \begin{bmatrix}
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
27 & 28 & 24 & 23 & 19 & 14 & 9 & 5 & -5 & -9 & -14 & -19 & -23 & -24 & -28 & -27 \\
24 & 20 & 12 & 8 & -8 & -12 & -20 & -24 & -24 & -20 & -12 & -8 & 8 & 12 & 20 & 24 \\
28 & 19 & 5 & -14 & -24 & -27 & -23 & -9 & 9 & 23 & 27 & 24 & 14 & -5 & -19 & -28 \\
16 & 8 & -8 & -16 & -16 & -8 & 8 & 16 & 16 & 8 & -8 & -16 & -16 & -8 & 8 & 16 \\
24 & 5 & -23 & -28 & -9 & 19 & 27 & 14 & -14 & -27 & -19 & 9 & 28 & 23 & -5 & -24 \\
20 & -8 & -24 & -12 & 12 & 24 & 8 & -20 & -20 & 8 & 24 & 12 & -12 & -24 & -8 & 20 \\
23 & -14 & -28 & 5 & 27 & 9 & -24 & -19 & 19 & 24 & -9 & -27 & -5 & 28 & 14 & -23 \\
16 & -16 & -16 & 16 & 16 & -16 & -16 & 16 & 16 & -16 & -16 & 16 & 16 & -16 & -16 & 16 \\
19 & -24 & -9 & 27 & -5 & -28 & 14 & 23 & -23 & -14 & 28 & 5 & -27 & 9 & 24 & -19 \\
12 & -24 & 8 & 20 & -20 & -8 & 24 & -12 & -12 & 24 & -8 & -20 & 20 & 8 & -24 & 12 \\
14 & -27 & 19 & 9 & -28 & 23 & 5 & -24 & 24 & -5 & -23 & 28 & -9 & -19 & 27 & -14 \\
8 & -16 & 16 & -8 & -8 & 16 & -16 & 8 & 8 & -16 & 16 & -8 & -8 & 16 & -16 & 8 \\
9 & -23 & 27 & -24 & 14 & 5 & -19 & 28 & -28 & 19 & -5 & -14 & 24 & -27 & 23 & -9 \\
8 & -12 & 20 & -24 & 24 & -20 & 12 & -8 & -8 & 12 & -20 & 24 & -24 & 20 & -12 & 8 \\
5 & -9 & 14 & -19 & 23 & -24 & 28 & -27 & 27 & -28 & 24 & -23 & 19 & -14 & 9 & -5
\end{bmatrix}
$$

The proposed HVBT

- Transform type is selected based on the RDO

$$\left\{\hat{\boldsymbol{\theta}}, T_{type}\right\} = \arg\min_{i} \sum_{j \in S_i} \left\{ \min\left[\left\| \mathbf{X} - \hat{\mathbf{X}} \right\|^2 + \lambda \cdot R\left(\hat{\boldsymbol{\theta}}_i^j, T_{type}\right) \right] \right\}$$

$\overline{\mathbf{x}}$: recon. pixel data in jth 16x16 block based on RDO decision of tx type

$\left\| \mathbf{x} - \overline{\mathbf{x}} \right\|$: reconstruction error for i[th] 8x8 block in a 16x16 block

λ : Lagrange multiplier

$\mathbf{X}$: pixel data

$\hat{\boldsymbol{\theta}}$: DCT-quantized coefficients

- Available transform type for the MB mode

MB modes	T16 Types	T8 Types
16×16	16×16	8×8, 4×4
16×8	16×8	8×8, 4×4
8×16	8×16	8×8, 4×4
8×8	N/A	8×8, 4×4
8×4, 4×8, 4×4	N/A	4×4

The Proposed Quadtree VBT

- Low Complexity Transform Type Decision Method

Obtain the J_{16} using the T16
for(8×8 blocks){
    Obtain the $J^{j}_{8\times 8}$ using 8×8 transform
    Obtain the $J^{j}_{4\times 4}$ using 4×4 transform
    Decide the transform type for j^{th} 8x8 block
    Obtain the J^{j}_{8} with the minimum RD cost
    if($J_{16} < \sum_j J^{j}_{8}$)
        Encode the transform type as 16×16
    Else
        Next 8×8 block
}
Decide the transform type for 16×16 block

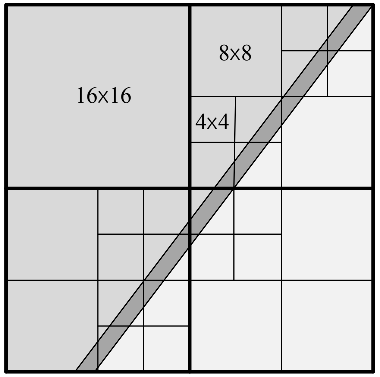
Top-down Approach

- Compared to H.264/AVC

Parameters	H.264/AVC	Proposed HVBT
Motion partition	16x16 – 4x4	16x16 – 4x4
Transform block	single size with 4x4 or 8x8	variable
cbp for luminance	4bits (1 bit per 8x8)	4bits (1 bit for 8x8)
Side information for transform types	1bit	1bit for T16 maximum 5 bits for T8

Side Information for HVBT

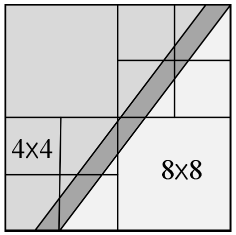
- The main problem for the HVBT is to send large amounts of side information



Side information : T16_flag (1 bit) + T8_flags(4 bits)

1 → **0** 1001 → **0** 1010 → **0** 0111

- Reduce of the amounts of side information for T8
  - The combination with luma *cbp* (coded block pattern)



example)
luma_cbp : 0 1 0 1 (from left-upper block)
T8_flag : 1 0 0 1 (from left-upper block)

Only 2 bits (01) are sent for non-zero cbp 8x8 blocks

  - For 8x4, 4x8, 4x4 sub-block modes, signaling bits are not sent

Experimental Setup

- JM11.0/KTA2.3
- Constraint Set 2 (Class B, Class C, Class D and E) with Beta Anchor
- GOP structure : IPPP
- Non-AVC tools
  - HPF (On or Off)
  - QALF (On or Off)
  - MVC (On or Off)
  - RDOQ(Off)
- QP (QP_P) range
  - Low QP range : 20, 24, 32 and 38
  - High QP range: 28, 31, 35 and 39

Experimental Setup

- Sets for experiments

QP_P	non-AVC tools On/Off: On	non-AVC tools On/Off: Off
Low QP (20, 24, 28, 32)	Set 0	Set 1
High QP (28, 31, 35, 39)	Set 2	Set 3

AHG-recommended condition

- HVBT is compared to the original H.264|MPEG-4 AVC and ST
  - Original H.264/MPEG-4 AVC
    - ✓ 4x4 and 8x8 transforms
  - ST (single-type transform)
    - ✓ Original H.264|MPEG-4 AVC + 16x16 transform kernel
    - ✓ 4x4, 8x8 and 16x16 transform kernels

Experimental Results

- Set 2 (High QP, non-AVC tools On)

Sequences		H.264/AVC vs. ST				H.264/AVC vs. HVBT			
		BDBR Avg(%)	BDPSNR Avg(dB)	BDBR High(%)	BDBR Low(%)	BDBR Avg(%)	BDPSNR Avg(dB)	BDBR High(%)	BDBR Low(%)
Class B	*Kimono1*	-1.06	0.04	-2.48	0.04	6.38	-0.22	0.58	11.91
	ParkScene	0.38	-0.01	-0.12	0.86	5.65	-0.18	1.05	11.03
	Cactus	-0.71	0.02	-1.62	-0.08	3.37	-0.09	-0.71	7.03
	BQTerrace	0.17	0.00	-0.39	0.48	6.16	-0.12	-0.09	12.41
	BasketballDrive	-1.51	0.04	-2.91	-0.59	1.09	-0.03	-2.92	4.55
	Average	-0.55	0.02	-1.50	0.14	4.53	-0.13	-0.42	9.39
Class C	*BQMall*	-0.89	0.04	-1.39	-0.31	0.70	-0.03	-1.51	3.40
	PartyScene	-0.35	0.01	-0.51	-0.13	-0.78	0.03	-1.51	0.54
	RaceHorses	-0.28	0.01	-0.43	-0.11	0.14	0.00	-1.24	2.06
	BasketballDrill	-0.92	0.04	-1.69	-0.29	1.20	-0.04	-1.03	4.28
	Average	-0.61	0.03	-1.01	-0.21	0.32	-0.01	-1.32	2.57
Class D	*BQSquare*	0.31	-0.01	0.43	-0.12	2.46	-0.09	-0.02	5.36
	RaceHorses	-0.16	0.01	-0.23	-0.06	-0.20	0.01	-0.88	0.84
	BasketballPass	-0.59	0.03	-0.82	-0.17	0.58	-0.02	-1.40	3.23
	BlowingBubbles	0.16	-0.01	0.01	0.16	0.84	-0.03	-0.13	2.55
	Average	-0.07	0.01	-0.15	-0.05	0.92	-0.03	-0.61	3.00
Class E	*Vidyo 1*	0.08	0.00	-0.40	0.34	17.84	-0.66	10.11	24.92
	Vidyo 3	0.62	-0.02	0.30	0.61	11.95	-0.46	5.81	16.83
	Vidyo 4	-0.30	0.01	-0.85	0.15	14.73	-0.49	8.41	20.52
	Average	0.13	0.00	-0.32	0.37	14.84	-0.54	8.11	20.76

Experimental Results

- Set 0 (Low QP, non-AVC tools On)

Sequences		H.264/AVC vs. ST				H.264/AVC vs. HVBT			
		BDBR Avg(%)	BDPSNR Avg(dB)	BDBR High(%)	BDBR Low(%)	BDBR Avg(%)	BDPSNR Avg(dB)	BDBR High(%)	BDBR Low(%)
Class B	*Kimono1*	-2.03	0.05	-1.31	-2.17	-1.52	0.04	-3.47	0.95
	ParkScene	-0.49	0.02	-0.68	-0.15	-1.18	0.04	-2.77	1.08
	Cactus	-1.14	0.03	-0.62	-1.86	-1.42	0.04	-1.58	-0.52
	BQTerrace	-0.45	0.01	-0.37	-0.59	-1.07	0.03	-1.51	0.19
	BasketballDrive	-2.74	0.06	-1.70	-3.09	-3.77	0.08	-3.15	-2.62
	Average	-1.37	0.03	-0.93	-1.57	-1.79	0.05	-2.50	-0.18
Class C	*BQMall*	-1.32	0.05	-0.97	-1.45	-2.37	0.09	-2.87	-1.49
	PartyScene	-0.45	0.02	-0.33	-0.48	-1.73	0.10	-1.78	-1.54
	RaceHorses	-0.32	0.02	0.05	-0.52	-1.55	0.08	-1.33	-1.43
	BasketballDrill	-2.64	0.11	-3.03	-1.88	-2.59	0.11	-3.96	-0.91
	Average	-1.18	0.05	-1.07	-1.08	-2.06	0.10	-2.48	-1.34
Class D	*BQSquare*	0.08	-0.01	-0.04	0.16	-0.43	0.03	-0.72	-0.01
	RaceHorses	-0.26	0.01	-0.19	-0.22	-1.42	0.08	-1.67	-0.89
	BasketballPass	-0.61	0.03	-0.65	-0.61	-1.45	0.08	-1.50	-1.33
	BlowingBubbles	-0.10	0.00	-0.13	-0.20	-1.05	0.05	-1.46	-0.61
	Average	-0.23	0.01	-0.25	-0.22	-1.09	0.06	-1.34	-0.71
Class E	*Vidyo 1*	-1.03	0.02	-1.68	-0.48	5.01	-0.11	-0.74	10.52
	Vidyo 3	-0.63	0.02	-1.81	0.44	0.70	-0.01	-3.47	5.98
	Vidyo 4	-1.67	0.04	-1.53	-1.24	2.36	-0.05	-1.53	8.36
	Average	-1.11	0.03	-1.67	-0.43	2.69	-0.06	-1.91	8.29

Experimental Results

- Set 1 (Low QP, non-AVC tools Off)

Sequences		H.264/AVC vs. ST				H.264/AVC vs. HVBT			
		BDBR Avg(%)	BDPSNR Avg(dB)	BDBR High(%)	BDBR Low(%)	BDBR Avg(%)	BDPSNR Avg(dB)	BDBR High(%)	BDBR Low(%)
Class B	*Kimono1*	-3.18	0.08	-1.09	-4.61	-4.78	0.12	-4.52	-4.72
	ParkScene	-1.42	0.05	-0.71	-2.23	-3.24	0.11	-3.45	-3.00
	Cactus	-0.94	0.03	0.13	-3.02	-2.47	0.06	-1.46	-3.22
	BQTerrace	-2.09	0.06	-0.80	-4.34	-3.61	0.11	-2.25	-5.65
	BasketballDrive	-3.53	0.08	-1.70	-5.11	-5.79	0.14	-3.47	-7.26
	Average	-2.23	0.06	-0.83	-3.86	-3.98	0.11	-3.03	-4.77
Class C	*BQMall*	-2.07	0.08	-1.00	-2.86	-3.93	0.16	-3.52	-4.05
	PartyScene	-0.58	0.03	-0.26	-0.90	-2.04	0.12	-1.81	-2.31
	RaceHorses	-0.53	0.03	0.05	-1.04	-1.96	0.10	-1.47	-2.24
	BasketballDrill	-4.10	0.18	-3.42	-4.63	-4.69	0.21	-4.66	-4.86
	Average	-1.82	0.08	-1.16	-2.36	-3.16	0.15	-2.86	-3.36
Class D	*BQSquare*	-0.25	0.01	-0.15	-0.42	-1.48	0.09	-1.49	-1.55
	RaceHorses	-0.48	0.03	-0.26	-0.54	-1.70	0.10	-1.90	-1.38
	BasketballPass	-1.60	0.09	-1.23	-1.76	-3.40	0.19	-3.24	-3.38
	BlowingBubbles	-0.54	0.02	-0.26	-0.97	-1.97	0.09	-1.87	-2.17
	Average	-0.72	0.04	-0.48	-0.92	-2.14	0.12	-2.12	-2.12
Class E	*Vidyo 1*	-9.85	0.26	-8.46	-8.71	-8.70	0.23	-10.50	-5.35
	Vidyo 3	-4.59	0.13	-4.23	-3.97	-6.22	0.17	-7.00	-3.58
	Vidyo 4	-4.92	0.12	-3.75	-4.96	-4.57	0.11	-5.52	-2.17
	Average	-6.45	0.17	-5.48	-5.88	-6.50	0.17	-7.68	-3.70

Experimental Results

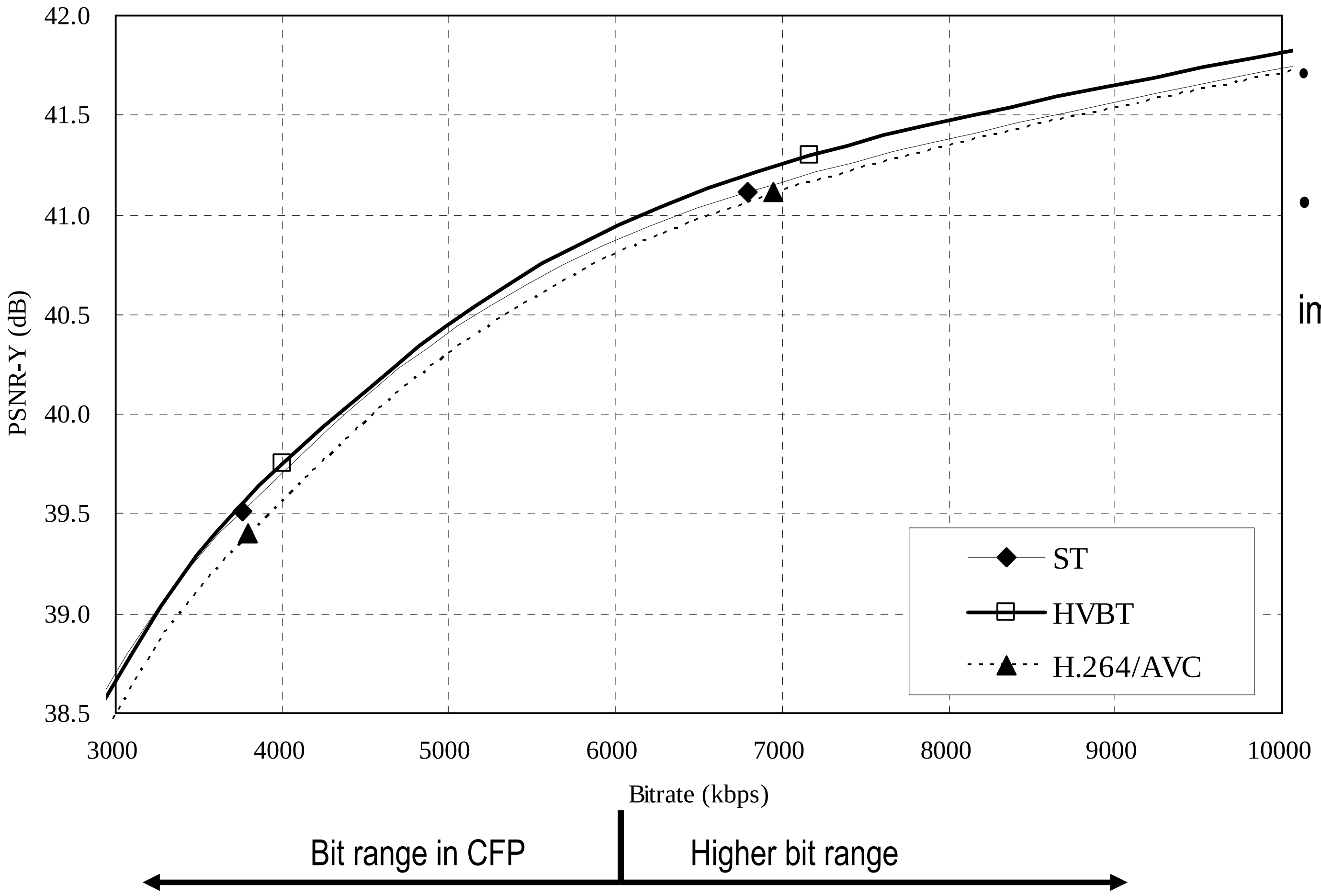
- Set 3 (High QP, non-AVC tools off)

Sequences		H.264/AVC vs. ST				H.264/AVC vs. HVBT			
		BDBR Avg(%)	BDPSNR Avg(dB)	BDBR High(%)	BDBR Low(%)	BDBR Avg(%)	BDPSNR Avg(dB)	BDBR High(%)	BDBR Low(%)
Class B	*Kimono1*	-3.18	0.08	-1.09	-4.61	-4.78	0.12	-4.52	-4.72
	ParkScene	-1.42	0.05	-0.71	-2.23	-3.24	0.11	-3.45	-3.00
	Cactus	-0.94	0.03	0.13	-3.02	-2.47	0.06	-1.46	-3.22
	BQTerrace	-2.09	0.06	-0.80	-4.34	-3.61	0.11	-2.25	-5.65
	BasketballDrive	-3.53	0.08	-1.70	-5.11	-5.79	0.14	-3.47	-7.26
	Average	-2.23	0.06	-0.83	-3.86	-3.98	0.11	-3.03	-4.77
Class C	*BQMall*	-2.07	0.08	-1.00	-2.86	-3.93	0.16	-3.52	-4.05
	PartyScene	-0.58	0.03	-0.26	-0.90	-2.04	0.12	-1.81	-2.31
	RaceHorses	-0.53	0.03	0.05	-1.04	-1.96	0.10	-1.47	-2.24
	BasketballDrill	-4.10	0.18	-3.42	-4.63	-4.69	0.21	-4.66	-4.86
	Average	-1.82	0.08	-1.16	-2.36	-3.16	0.15	-2.86	-3.36
Class D	*BQSquare*	-0.25	0.01	-0.15	-0.42	-1.48	0.09	-1.49	-1.55
	RaceHorses	-0.48	0.03	-0.26	-0.54	-1.70	0.10	-1.90	-1.38
	BasketballPass	-1.60	0.09	-1.23	-1.76	-3.40	0.19	-3.24	-3.38
	BlowingBubbles	-0.54	0.02	-0.26	-0.97	-1.97	0.09	-1.87	-2.17
	Average	-0.72	0.04	-0.48	-0.92	-2.14	0.12	-2.12	-2.12
Class E	*Vidyo 1*	-9.85	0.26	-8.46	-8.71	-8.70	0.23	-10.50	-5.35
	Vidyo 3	-4.59	0.13	-4.23	-3.97	-6.22	0.17	-7.00	-3.58
	Vidyo 4	-4.92	0.12	-3.75	-4.96	-4.57	0.11	-5.52	-2.17
	Average	-6.45	0.17	-5.48	-5.88	-6.50	0.17	-7.68	-3.70

Experimental Results

- H.264/AVC vs. ST vs. HVBT

Kimono1 (1080P, Class B1) with non-AVC tools Off

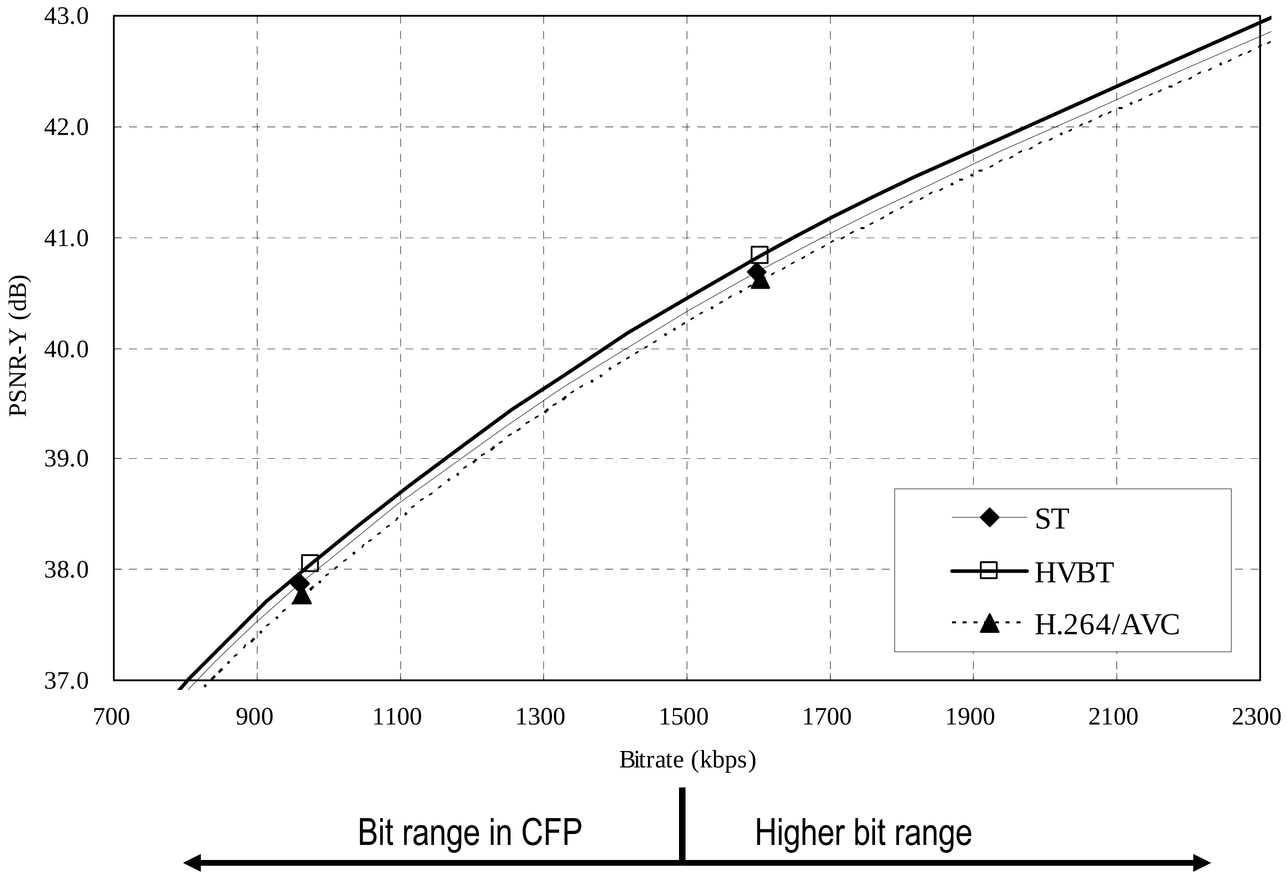


- For higher bit rates
  - HVBT is helpful
- For low bit rates
  - HVBT shows no improvement against ST

Experimental Results

- H.264/AVC vs. ST vs. HVBT

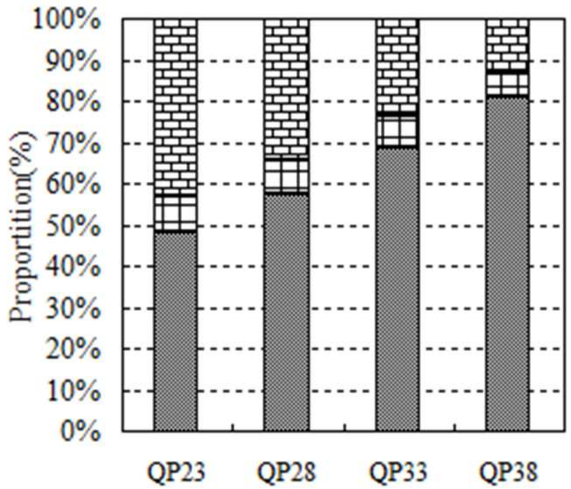
BasketballPass (416x240, Class D) with non-AVC tools Off



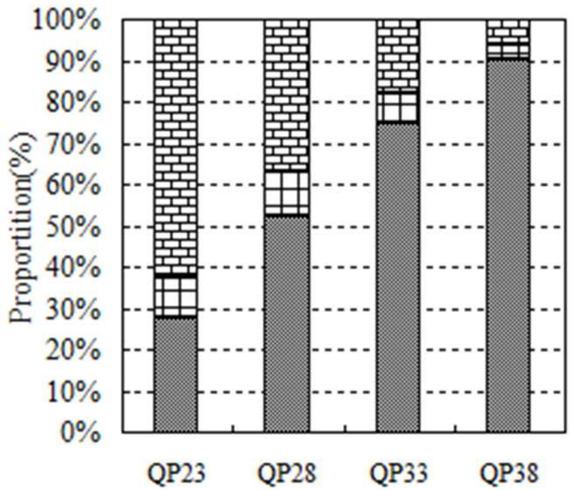
- For higher bit rates
  - HVBT is helpful
- For low bit rates
  - HVBT shows no improvement against ST

Experimental Results

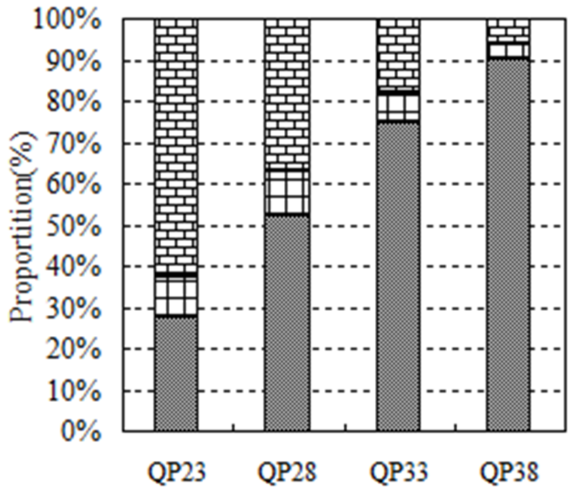
- Selected proportions of transform types



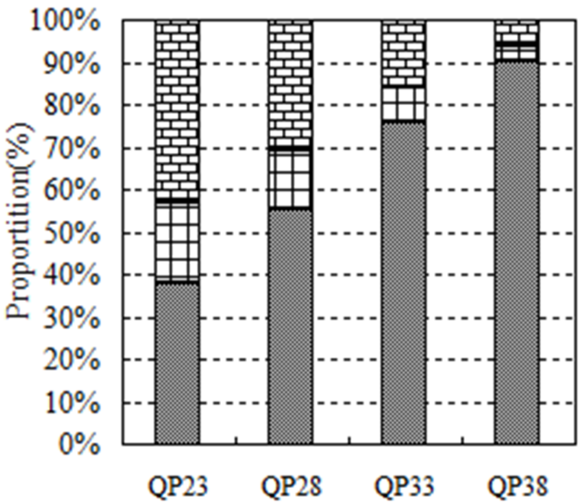
Cactus(1080P)



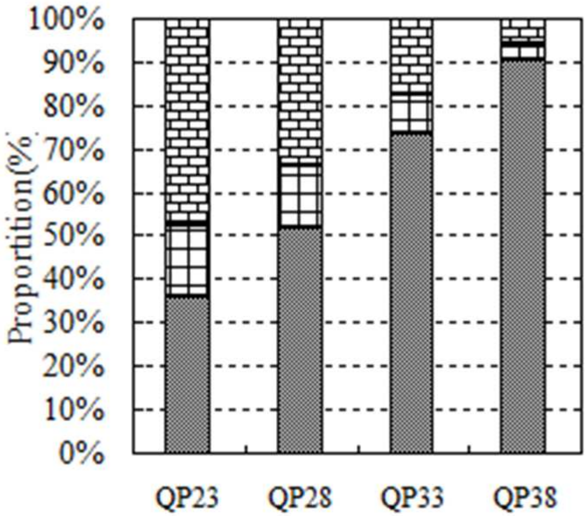
BQTerrace(1080P)



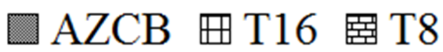
BasketballDrive(1080P)



BasketballDrill(832x480)



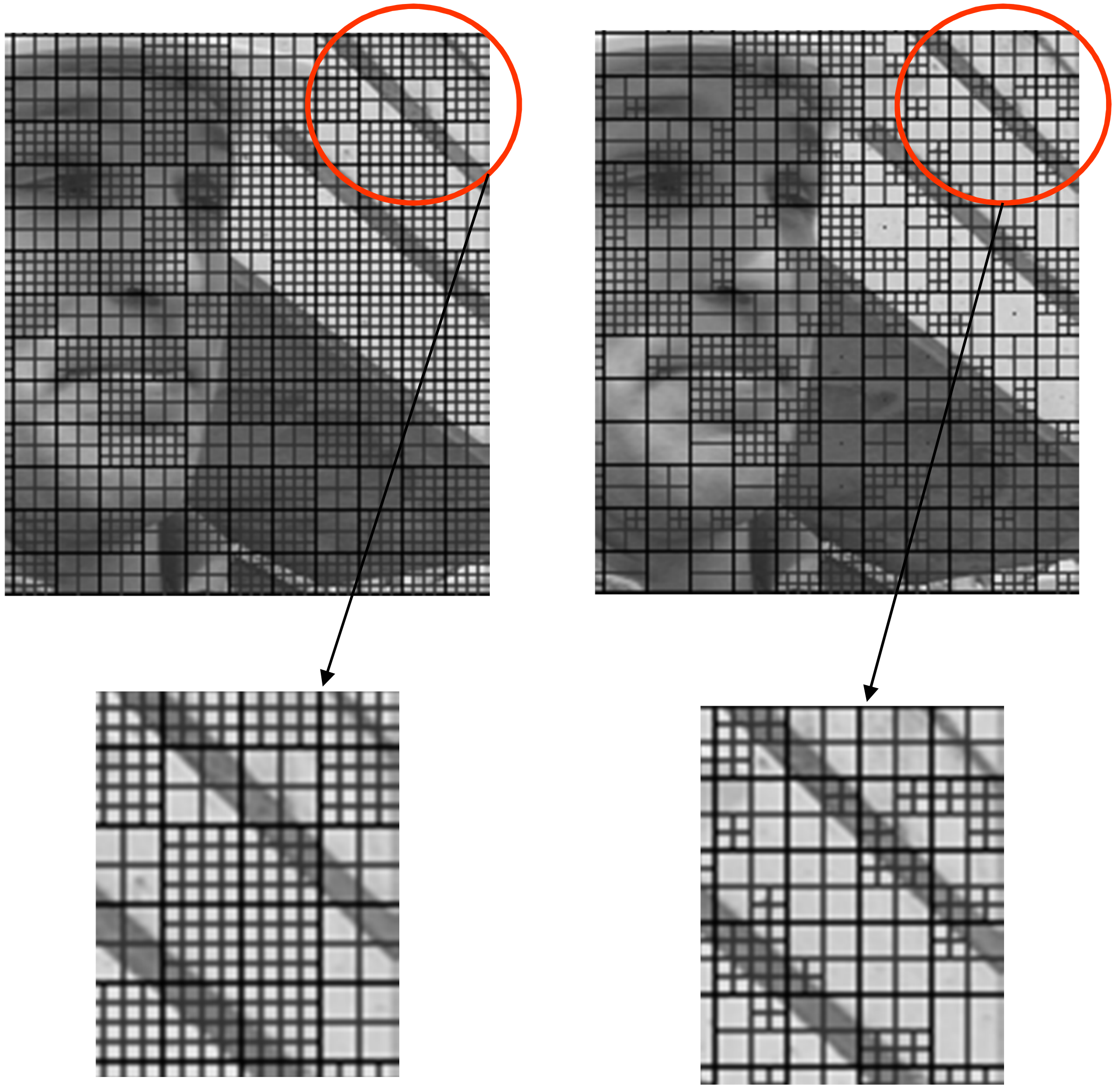
RaceHorses(832x480)



AZCB : SKIP + All Zero Coefficient Blocks for non-SKIP mode

Experimental Results

- *Foreman* CIF(H.264|MPEG-4 AVC vs. HVBT, QP20)



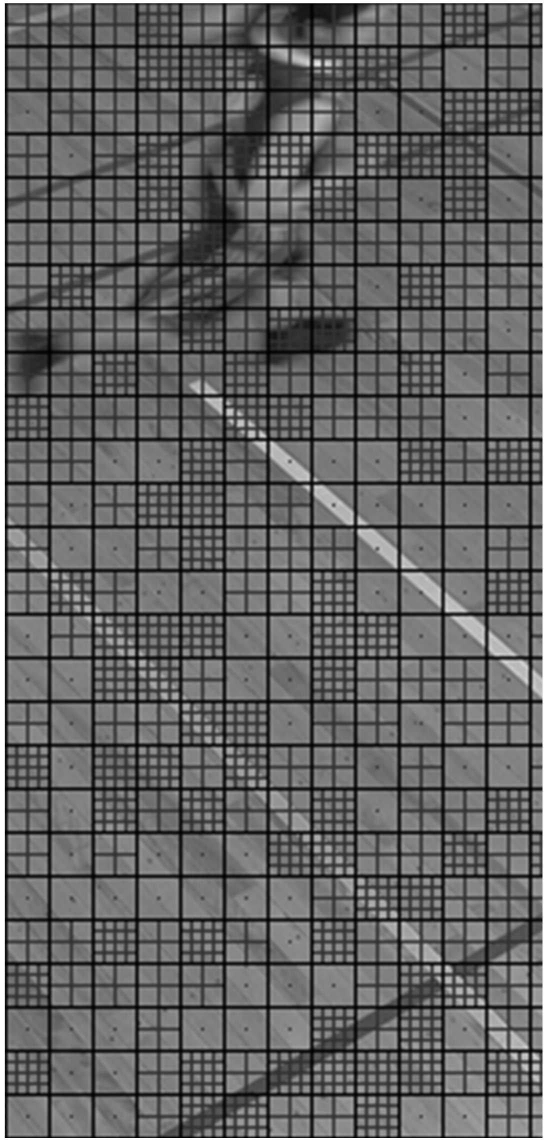
SKIP mode

16x16 Transform

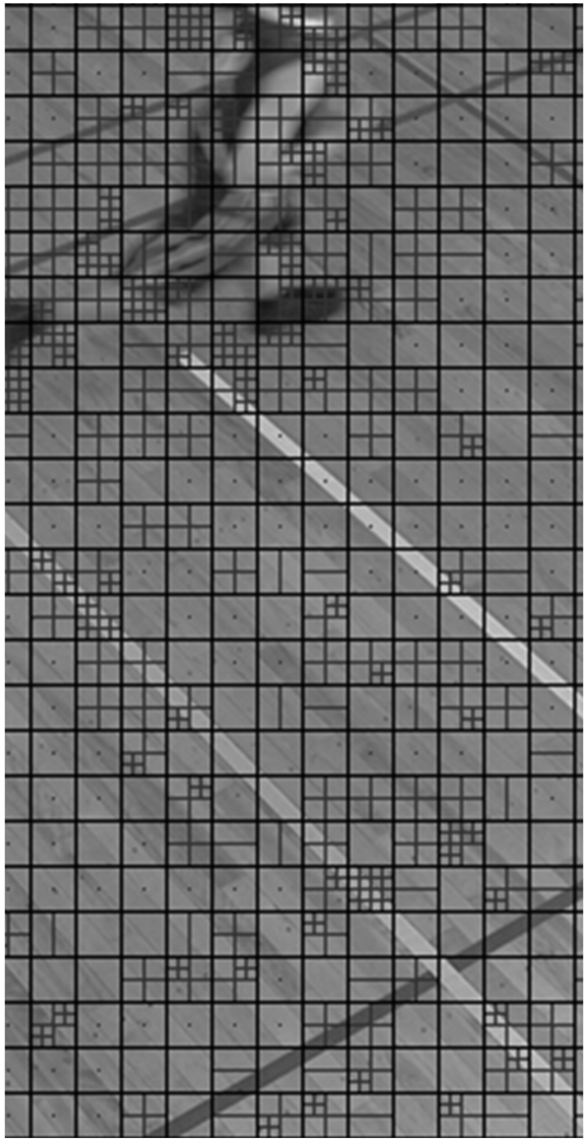
BasketballDrill (QP 24)

BasketballDrill (QP 34)

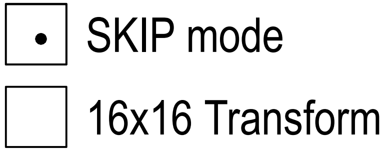
BasketballDrill (QP 24)



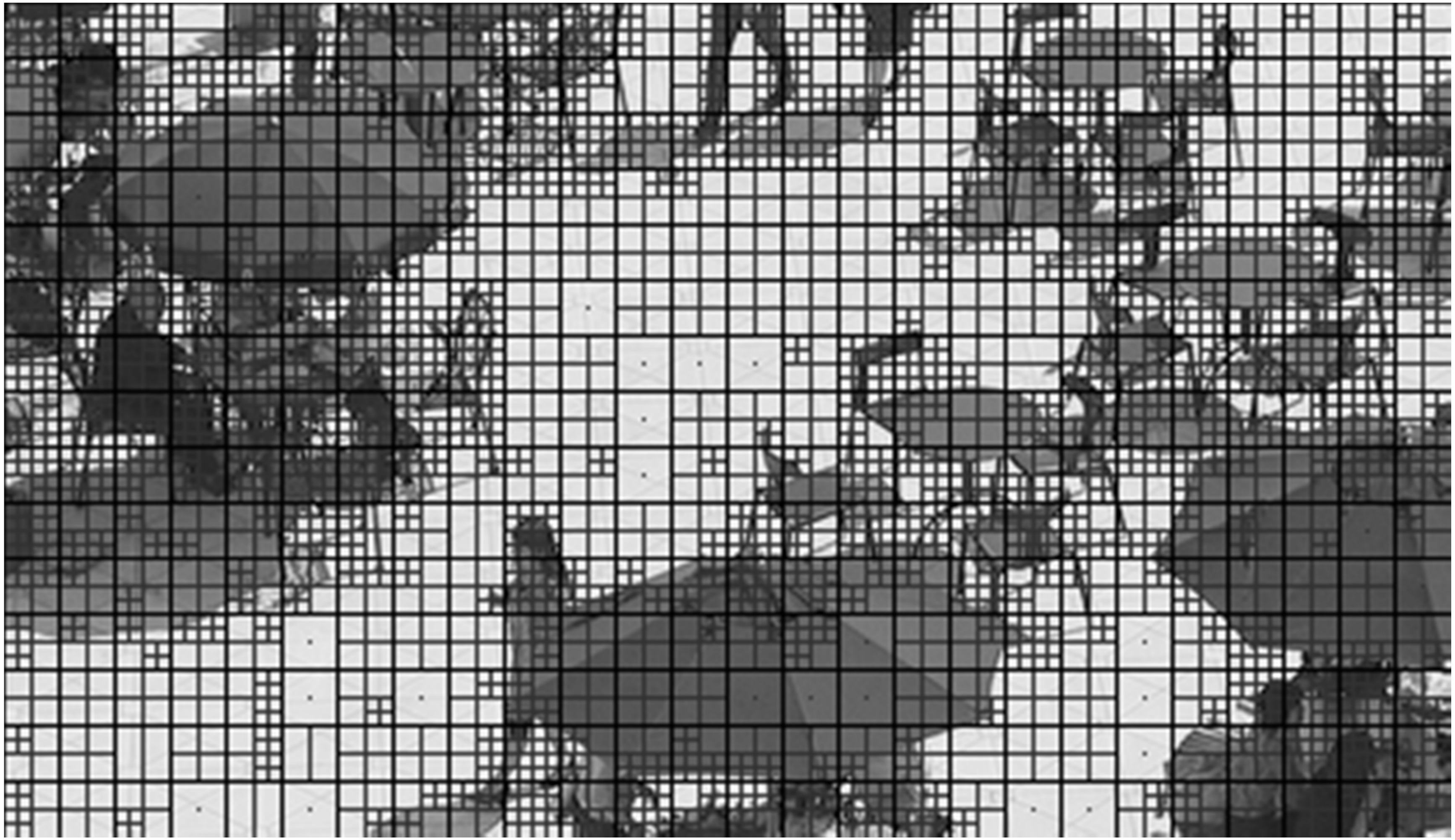
H.264/AVC



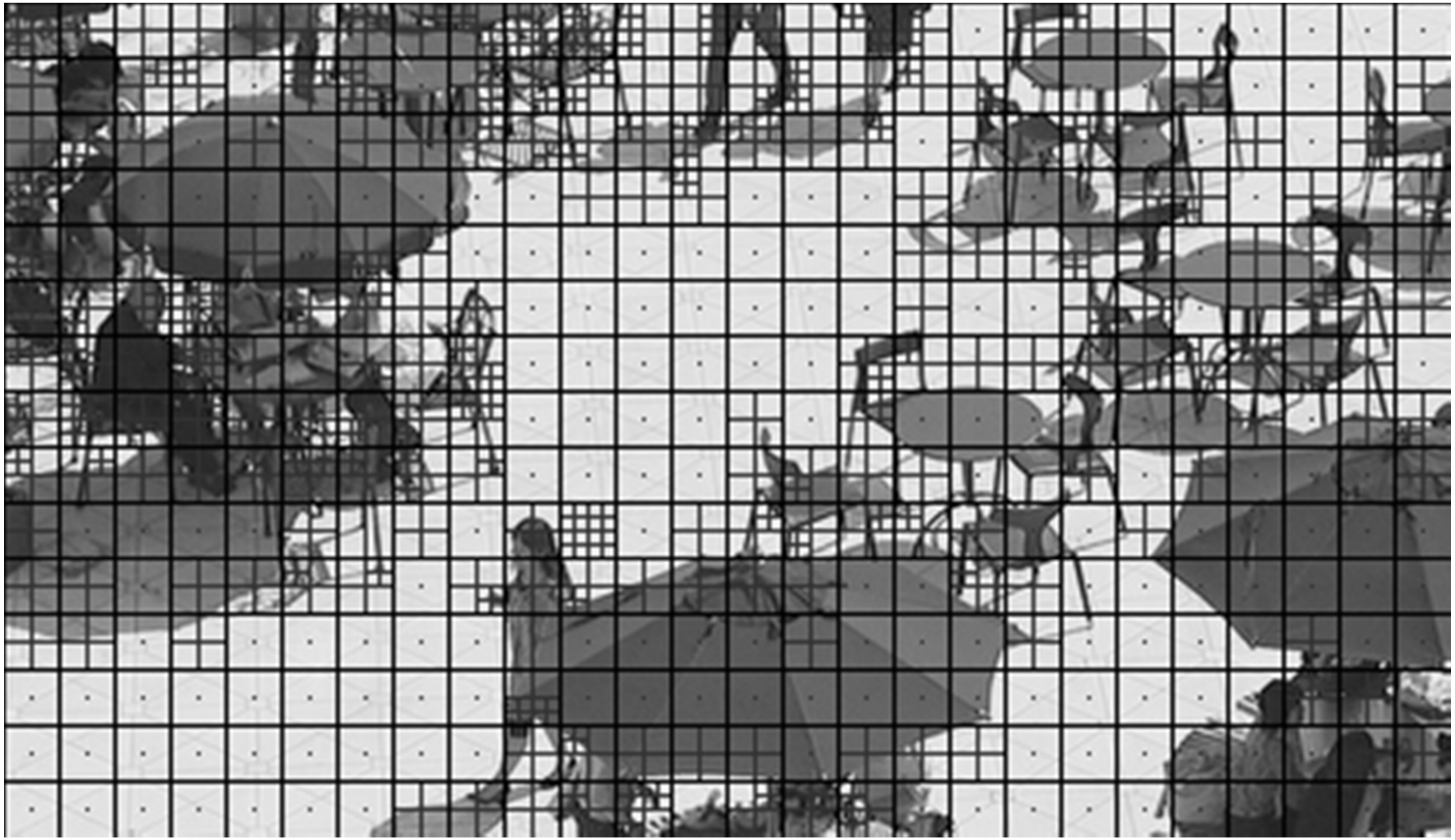
HVBT



BQSquare (QP 20)



BQSquare (QP 34)



Summary

- The proposed HVBT scheme shows the RD performance improvements consistently in a high bit range regardless of whether the non-AVC tools were ON or OFF.
- Its best performance is obtained in a low QP range with the non-AVC tools turned OFF.
- The weak points of the current HVBT scheme are:
  - its RD performance is degraded in low bit ranges. Especially, this is noticeably observed when the non-AVC tools are ON.
  - The reason for this performance degradation is because the high QP and non-AVC tools tend to lower the energy of ICT coefficients, which leads to the SKIP modes and large block modes to be more preferably selected than the modes by hierarchical transform partitions.

Future Plan

- We presented our preliminary results and analyzed the performance of our proposed HVBT scheme, which showed somewhat limited performance improvements under the current test conditions.
- Nevertheless, there are some possibilities of improving the proposed HVBT scheme:
  - (1) its signaling syntax of transforms types are not optimized, which can further be improved in the future TM architecture;
  - (2) The maximum size of the transform kernels of HVBT is limited to order-16, which can be combined with the transform kernels of larger sizes in conjunction with the scalable syntax in the future TM architecture.

Some Issues

- The HVBT study has been performed with the set of QP values (28, 31, 35, 39) which seems to be favorably shifted towards a lower bit range.
- It is worthwhile to consider an appropriate range of QP values;
- The test sequences to be used for the transform experiments seem to lean toward a set of complex scenes which may drive some tools to overfit a particular data set.
- Therefore, it is also worthwhile to consider more appropriate sets of test sequences.