webm project

Google

Jim Bankoski,
Yue Chen,
Angie Chiang,
Alex Converse,
Adrian Grange,
Jingning Han,
Shunyao Li,
Zoe Liu,
***Debargha Mukherjee**,*
Hui Su,
Yunqing Wang,
Yaowu Xu

[Argon Designs,
Ronald Bultje]

An overview of new video coding tools in **VP10**

Debargha Mukherjee
Google, Inc.

Outline

- Introduction
- Coding Tools
- Current Results
- Conclusion

Outline

- Introduction
- Coding Tools
- Results
- Conclusion

Introduction: Towards VP10

- WebM project (2010): develop efficient, royalty-free video codecs
  - **VP8** (2010, Hangouts)
  - **VP9** (2013, YouTube)
- Typical recipe for generational progression of codecs:
  - 10 years interval - driven by hardware dev/adoption cycle
  - About 50% bandwidth reduction from previous generation
- Faster cycle needed to match the pace of innovation in online video
  - Explosive growth in online video (70% of Internet bandwidth)
- Royalty-free !

Enter VP10

Introduction: VP10 Technical Goals

- VP10 core objectives:
  - Advancement in compression efficiency
    - Generational improvement over VP9
  - Decoder complexity in check
    - Software decoding at 60fps for 1080p natural content videos.
- Support for various new applications beyond traditional
  - Improved coding performance for screen and gaming content.
  - Support emerging media applications: VR, lightfield.
- Timeline: tentatively ~3 years
  - Combination of a hw codec + a sw codec to be the new norm on devices.
  - Hardware cycle needs to shorten (**3 - 4** years?)

Introduction: Alliance for Open Media

- Alliance for Open Media - recently formed
  - Amazon
  - Cisco
  - Google
  - Intel
  - Microsoft
  - Mozilla
  - Netflix
  - ...

- Charter: to develop jointly a royalty-free codec.
  - VP10 + Daala + Thor + … = AoMedia codec

Outline

- Introduction
- Coding Tools
- Results
- Conclusion

Coding Tools

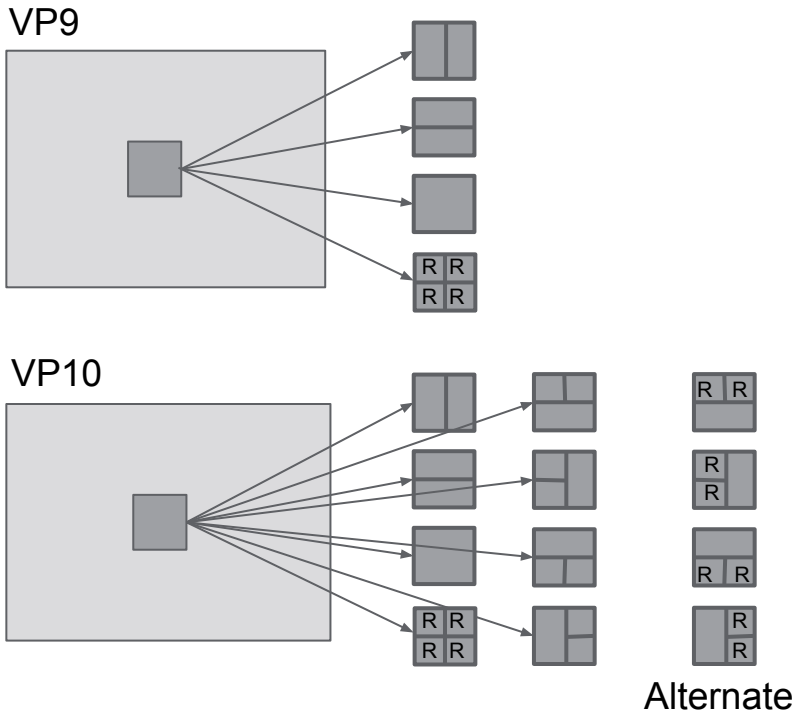
- Achieving a generational improvement in coding efficiency needs:
  - New coding tools, new ways of combining existing/new methods
- Several experiments and investigative threads underway
  - High bit-depth internal
  - Prediction tools
  - Transform coding tools
  - Screen content coding tools
  - Miscellaneous
- Disclaimer:
  - Nothing is set in stone, just a snapshot as of today! No IP review yet.
- Challenging task ahead !

Coding Tools: High Bit-Depth Internal

- A significant coding advantage in use of high bit-depth internal
- Most of the gain comes in the first 2 bits of precision
  - **Up to** 2-4% gain for coding 8-bit content with 10-/12-bit internal
- High bit-depth can become standard in a few years.
  - Studios producing 12-bit mezzanine formats, HDR coming.
- High bit-depth internal may be standard/default for VP10.
  - Can decode to 10-/12-bit or 8-bit depending on display capabilities from the same compressed bit-stream.

Coding Tools: Prediction Tools: Partition expansion

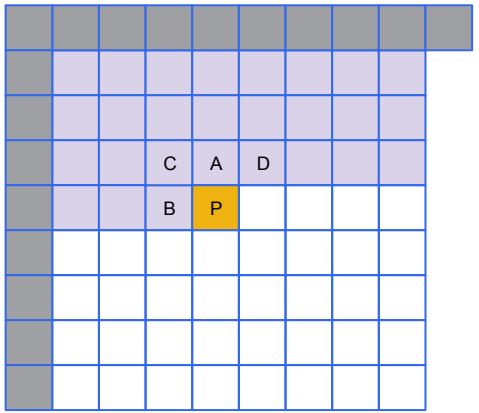
- Partition structure within frame
  - VP9
    - Uses 64x64 coding unit in frame
    - 4-way decomposition all the way until 4x4
    - End-blocks: superblock
  - VP10
    - Expands to a 8-way partition structure
    - *May go for larger coding unit 128x128* [Argon Designs]



Coding Tools: Prediction Tools: INTRA prediction

- INTRA prediction
  - FilterIntra [UCSB collaboration]
    - A new Intra prediction mode
    - Adds a set of intra-prediction modes using a recursive 3-/4-tap filtering framework from the left and top edges.
      - Filter weights are obtained by training
  - More directional prediction modes being added as well.

$$P = w_1 * A + w_2 * B + w_3 * C + w_4 * D$$

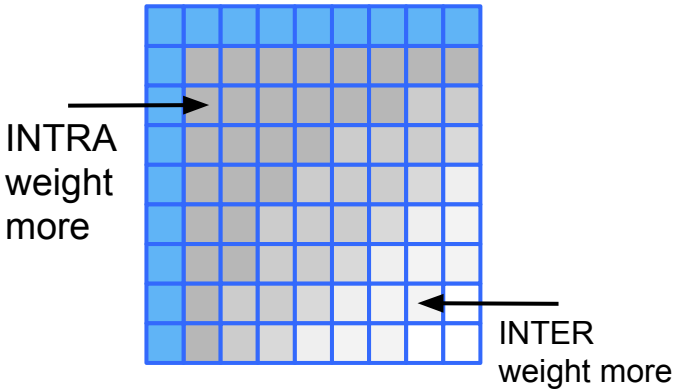


Coding Tools: Prediction Tools: INTER prediction

- VP9:
  - Select 3 reference frames for each coded frame from a pool of 8
  - Single ref modes: 1-ref specified with MV
  - Compound ref modes: 2 refs used with MV; Uniform (½, ½) weighting
- VP10 improvements:
  - Increases number of reference frames to 5 or 6
  - New single-ref modes
  - More extensive advancements in compound modes
    - Expands to allow combining INTER w/ INTRA
    - Expands to allow combining two INTERs in richer ways
      - Regarded as weighted combination schemes: *Smooth* or *Cliff*

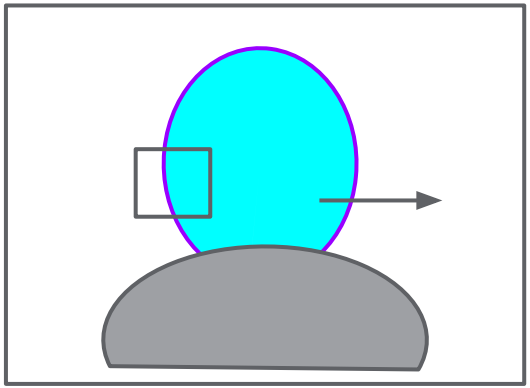
Coding Tools: Prediction Tools: InterIntra [Smooth]

- InterInter (Smooth):
  - Uniform ½-½ weighting of two inter predictors
  - Traditional compound Inter in VP9/VP10
- InterIntra (Smooth):
  - A compound prediction mode that combines an inter- and an intra-predictor.
  - The weights vary smoothly from intra to inter along the prediction direction



Coding Tools: Prediction Tools: Wedge Partition [Cliff]

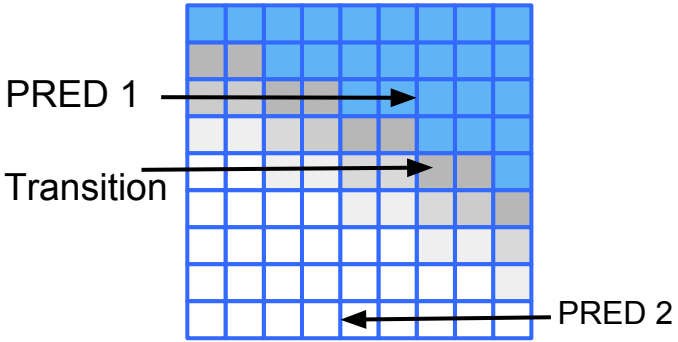
- In video, the boundary of a moving object can go through a block.
  - Boundary not necessarily horz or vert.
  - Going down to too small blocks inefficient.
- Solution:
  - Allow blocks to have oblique partitions
  - Predict each side with a different predictor
- Close connection to Wedgelets
  - Piecewise constant representation of an image
  - Extension to quad-tree partition



Coding Tools: Prediction Tools: Wedge Partition [Cliff]

- Wedge codebook
  - Set of oblique lines through a block.
- Convey:
  - wedge index + 2 predictors
- The predictors are combined with a cliff weighting scheme, where the *cliff* is oriented along the specified wedge.
  - *Overlapped wedge motion compensation*

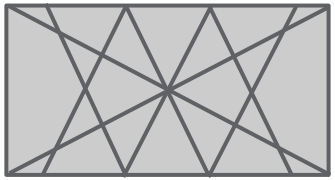
wedge codebook



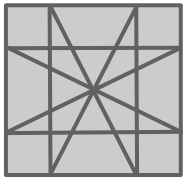
Coding Tools: Prediction Tools: Wedge Partition [Cliff]

- Wedge codebooks - size increases with increase in block size

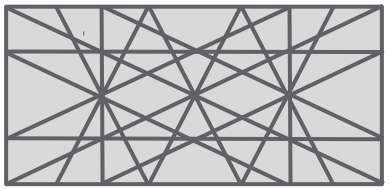
wedge codebook for 8x8



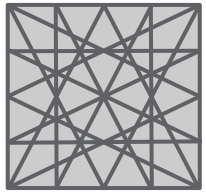
wedge codebook for 16x8, 8x16, 32x16, 16x32



wedge codebook for 16x16, 32x32



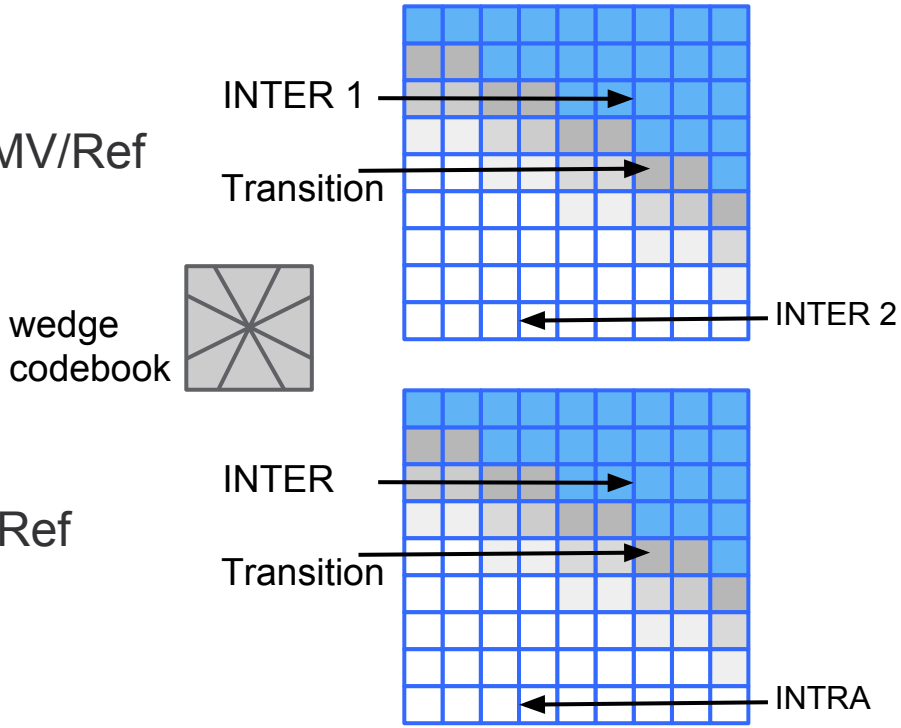
wedge codebook for 64x32, 32x64



wedge codebook for 64x64

Coding Tools: Prediction Tools: Wedge Partition Modes

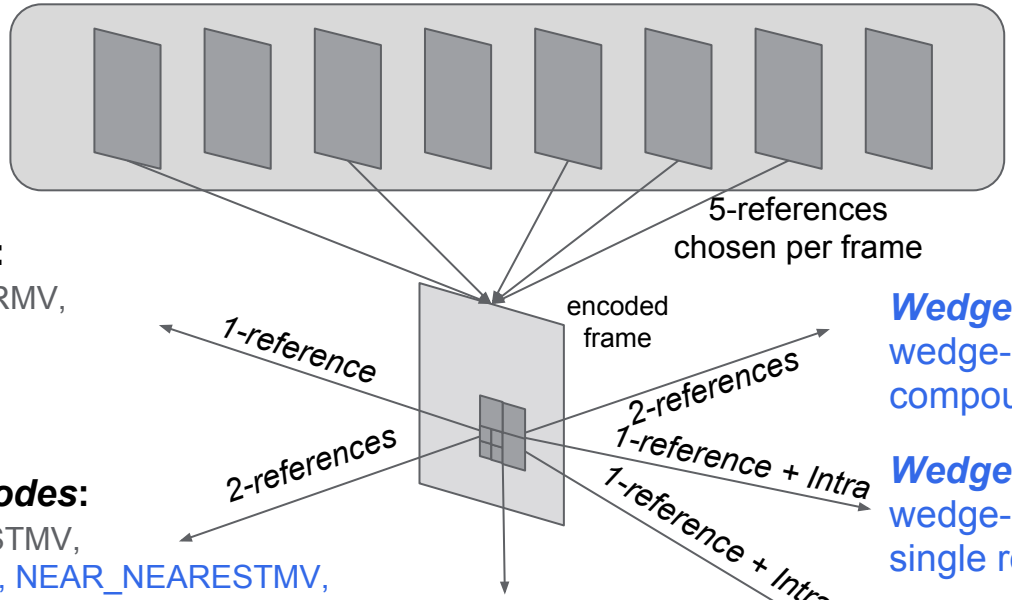
- Wedge-partition INTER-INTER
  - Convey wedge-index and two MV/Ref combinations
- Wedge-partition INTER-INTRA
  - Convey wedge-index, one MV/Ref and one INTRA mode



Coding Tools: Prediction Tools: Global Motion

- Global (camera) motion is common: pan/tilt/zoom/shake, gaming
  - Current generation codecs still use translational motion model
- Possible global camera models
  - Translation (2 params)
  - Translation + Rot + Zoom (4 params)
  - Affine (6 params)
  - Perspective (8 params)
- Global params in frame header; block-level warped pred modes
  - VP10: Extends VP9's ZEROMV definition.
  - Ongoing work: Ability to compute motion reliably is crucial

Coding Tools: VP10 Prediction Tools Summary



Single ref modes:
{NEARESTMV, NEARMV,
NEWMV, NEW2MV,
ZEROMV (GLOBAL)}

Compound ref modes:
{NEAREST_NEARESTMV,
NEAREST_NEARMV, NEAR_NEARESTMV,
NEW_NEARESTMV, NEAREST_NEWMV,
NEW_NEARMV, NEAR_NEWMV,
NEW_NEWMV, ZERO_ZEROMV (GLOBAL)}

INTRA modes:
{10 INTRA modes
8 filtered INTRA modes}

Wedge INTER-INTER modes:
wedge-index +
compound ref INTER mode

Wedge INTER-INTRA modes:
wedge-index + INTRA mode +
single ref INTER mode

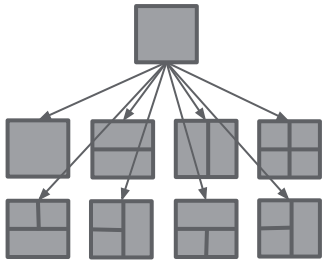
INTER-INTRA modes:
INTRA mode +
Single-ref INTER mode

Coding Tools: Transform Tools

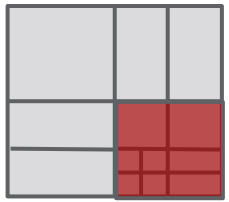
- Super transform
- Extended transform set
- Recursive transform units

Coding Tools: Transform Tools: Super Transform

- VP9: Transform size <= prediction size.
- VP10 Super transform: transforms cover multiple prediction blocks.
  - A combination of prediction and transform
- Partition syntax:
  - At any level of the partition tree, a bit indicates if a super transform is to be used at that level.
- Filter prediction block boundaries before applying super-transform - a new predictor
  - Recursive overlapped block motion compensation
  - Critical for efficiency



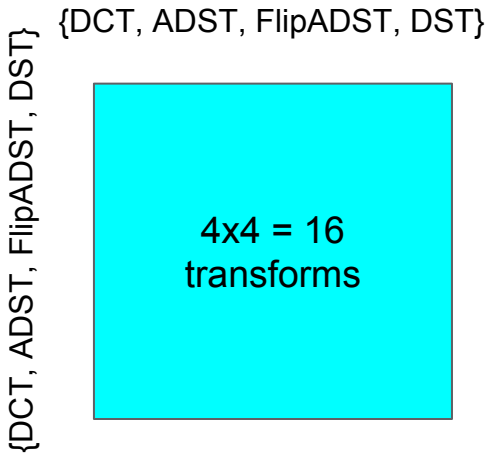
Partition Tree



Uses one large transform

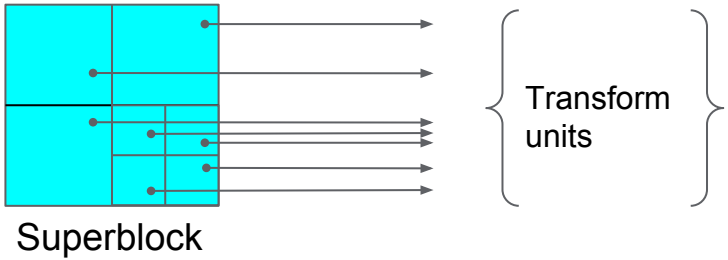
Coding Tools: Transform Tools: Extended Transforms

- VP9 transforms
  - DCT exclusively for INTER
  - Hybrid DCT/ADST (ADST = DST-IV) for INTRA blocks
- VP10 Extended transforms for INTER
  - Expands INTER transforms to all combinations of DCT/ADST/DST
    - 16 transforms for 4x4/8x8/16x16 sizes
    - Overheads: Indicate transform used
    - Alternate transforms being explored for 32x32 and 64x64 - Wavelet/DCT hybrids
  - Transform Skip mode - for screen content

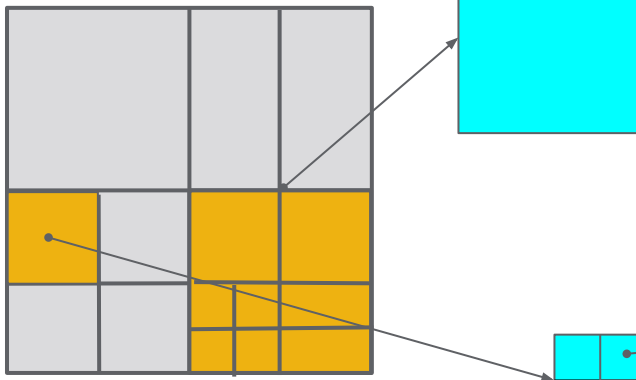


Coding Tools: Transform Tools: Recursive transform units

- VP9:
  - Transforms within a superblock are all of same size < superblock size
- VP10:
  - Allows transforms within a single superblock to have tree-structured partitions
  - More flexibility - can isolate part of a larger block with residue concentration

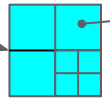


Coding Tools: Transform Tools: Summary



Prediction Partitioning

Super-transform [single transform for multiple prediction blocks]: Recursive OBMC for multiple prediction blocks, followed by single transform



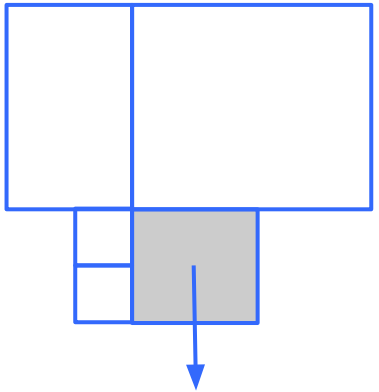
Recursive transform unit [multiple transforms for single prediction unit]: Tree-structured decomposition of transform units within end prediction block.

Choose one of 16 actual transforms

{DCT, ADST, FilpADST, DST}
x
{DCT, ADST, FilpADST, DST}

Coding Tools: Miscellaneous: Copy Mode

- Copy Mode
  - Copies entire motion/mode information from neighboring blocks
  - Creates a list of candidate info from neighbors.
  - Picks one that matches, or transmits a new one.
  - Some overlap/redundancy with NEARESTMV and NEARMV
    - To be harmonized.



Coding Tools: Miscellaneous: Loop Postfilter

- VP9 has only a deblocking filter
- VP10: Allows applying a nonlinear filter in loop after deblocking
  - Currently a bilateral filter
    - Send parameters corresponding to spatial and range Gaussian variances.
    - Need a fast implementation of the filter.
  - Experiments with other filters ongoing
    - Adaptive Wiener + bilateral combinations

Coding Tools: Miscellaneous, Eclectic

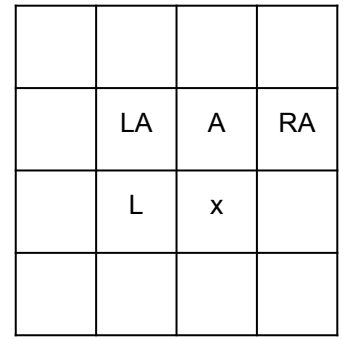
- 64x64 transforms
  - Small gains for high-def with 64x64 DCT, wavelets under investigation
- Quantization
  - Non-uniform for the first few quantization bins - small gains
- Super-resolution block coding mode - preliminary work
  - Reduced resolution coding for certain blocks
  - Interpolate + data-dependent filter applied for reconstruction.
- ANS (Asymmetric Numeral Systems) in lieu of arith coding for coefs
  - Decoding speed
- Tile based random access
  - Ability to random access [VR, lightfields]

Coding Tools: Screen Content Tools

- Screen content videos
  - Contains lots of text and graphics
- Conventional coding tools designed for natural videos are not appropriate
- Explore coding tools suitable for screen content video coding
  - Palette-based coding
  - Non-transform coding
  - Intra block motion

Coding Tools: Screen Content Tools: Palette Coding Mode

- Screen content often contains few colors but complicated shapes
- Palette representation: distinct colors + color index map
- Palette-based coding mode:
  - Obtain base colors and color indices
    - k-means clustering; choose k in [2, 8], so as to minimize RD cost.
  - Encode base colors
    - Rolling dictionary of colors at frame-level
  - Encode color indices
    - Use causal neighbors as context



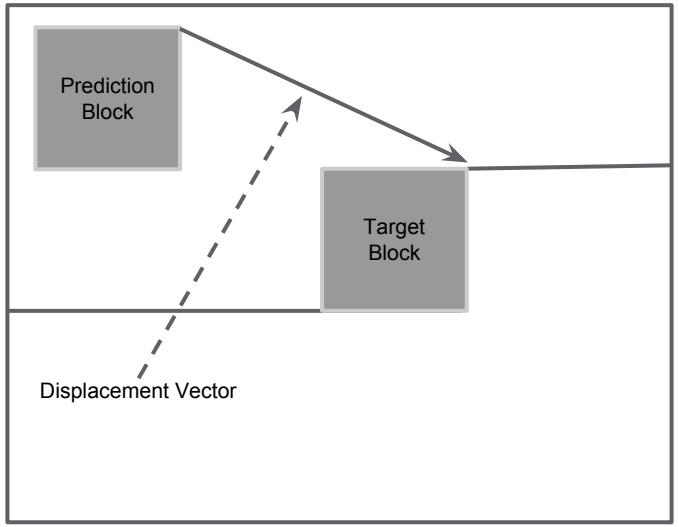
- Regarded as an extra intra prediction mode

Coding Tools: Screen Content Tools: Non-Transform Coding

- Transform coding is widely used in video codecs
  - Compact representation with most energy at low frequencies
- May not work well with screen content videos
- Non-transform coding:
  - Skip transform and code prediction residue signal in pixel domain
    - Can be used for all INTRA and INTER modes
  - Intra prediction special case [for Hor/Vert/TM]
    - Directional recursive DPCM

Coding Tools: Screen Content Tools: Intra Block Motion

- Motivation: repeated patterns and textures
- Use previously coded blocks of the current frame as a predictor
- Like motion prediction but within the current frame



Outline

- Introduction
- Coding Tools
- Results
- Developer Guide

Results - so far

- Common Test Conditions:
  - Bitrate control, 2-pass, 150 frame kf interval,
  - Speed settings: very-slow (X264/X265), cpu-used 0 (VP9/VP10)
- TestSets:
  - Natural content
    - CIF
      - https://x20web.corp.google.com/~debargha/vp10_results/x264_x265_vp9_derflr.html
      - https://x20web.corp.google.com/~debargha/vp10_results/vp10sep282015_derflr.html
    - 480P
      - https://x20web.corp.google.com/~debargha/vp10_results/x264_x265_vp9_hevcmr.html
      - https://x20web.corp.google.com/~debargha/vp10_results/vp10sep192015_hevcmr.html
  - Screen content
    - Mixed HD
      - https://x20web.corp.google.com/~debargha/vp10_results/20150610_scAllTools_screen.html

Results

- Natural content
  - ***derflr*** set
  - CIF resolution

Only 14% BDRATE reduction achieved by VP9/HEVC over H.264

VP10 Comparison Charts

Method For Combining Points ◉ BDRATE ○ BDSNR

Average size reduction to get the same quality ◉ Average PSNR ○ Global PSNR ○ SSIM

File	nextgen_base_derflr	x265_derflr
akiyo_cif	-15.3480%	-4.9823%
bowing_cif	-34.6290%	-21.0813%
bridge_close_cif	18.2070%	3.6871%
bridge_far_cif	-28.7400%	-6.8335%
bus_cif	-9.8902%	-17.9321%
cheer_sif	-8.1010%	-12.8426%
city_cif	-24.6617%	-24.0624%
coastguard_cif	-4.6579%	-16.1088%
container_cif	-17.8110%	-8.4100%
crew_cif	-13.2779%	-18.1356%
deadline_cif	-9.9536%	-7.9179%
flower_cif	-12.0326%	-18.3567%
football_cif	-16.1945%	-17.7656%
foreman_cif	-19.6390%	-21.8524%
hallmonitor_cif	-29.4419%	-10.1798%
harbour_cif	-4.3556%	-10.9181%
highway_cif	-25.3511%	-17.0959%
husky_cif	-2.8015%	-10.8634%
ice_cif	-8.3581%	-15.9563%
mobile_cif	-22.4415%	-25.3881%
motherdaughter_cif	-14.5129%	-9.9254%
news_cif	-7.6271%	-10.0628%
pamphlet_cif	-26.7722%	-13.9213%
paris_cif	-1.2702%	-4.5470%
signirene_cif	-12.4390%	-12.6387%
silent_cif	-5.7766%	-8.0149%
soccer_cif	-16.3208%	-19.1230%
stefan_sif	-11.4539%	-15.6092%
students_cif	-9.9624%	-11.2216%
tempete_cif	-12.7963%	-19.3881%
tennis_sif	-9.8148%	-11.7325%
waterfall_cif	-37.9443%	-31.6954%
OVERALL	-14.2553%	-14.0899%

Baseline X.264

Compare X.264, VP9, X.265

Google

Results

- Natural content
  - ***derflr*** set
  - CIF resolution

VP10 (12 bit) achieves over VP9 close to 85% of the gain HEVC/VP9 achieves over H.264.

VP10 Comparison Charts

Method For Combining Points ◉ BDRATE ○ BDSNR
Average size reduction to get the same quality ◉ Average PSNR ○ Global PSNR ○ SSIM

File	nextgen_sep28all_derflr	nextgen_sep28allhbd12_derflr
akiyo_cif	-5.7214%	-11.0286%
bowing_cif	-6.1115%	-11.8154%
bridge_close_cif	-10.6923%	-12.9670%
bridge_far_cif	-13.0134%	-21.6711%
bus_cif	-11.2815%	-12.8926%
cheer_sif	-10.5350%	-10.8140%
city_cif	-8.3586%	-14.1268%
coastguard_cif	-6.6235%	-9.3010%
container_cif	-6.1375%	-10.3084%
crew_cif	-7.6629%	-11.3940%
deadline_cif	-7.8187%	-12.0031%
flower_cif	-9.0569%	-10.0121%
football_cif	-7.3405%	-8.4849%
foreman_cif	-9.9084%	-14.4255%
hallmonitor_cif	-7.7728%	-15.1399%
harbour_cif	-8.0173%	-9.5331%
highway_cif	-7.2854%	-13.3626%
husky_cif	-8.7819%	-9.0297%
ice_cif	-7.7516%	-8.8581%
mobile_cif	-12.1276%	-13.5775%
motherdaughter_cif	-7.4689%	-17.2787%
news_cif	-7.7021%	-10.6919%
pamphlet_cif	-7.1292%	-12.4876%
paris_cif	-8.7221%	-10.8581%
signirene_cif	-9.8391%	-12.2144%
silent_cif	-7.6036%	-11.0562%
soccer_cif	-8.9300%	-13.2276%
stefan_sif	-10.8138%	-11.8555%
students_cif	-7.5903%	-14.8345%
tempete_cif	-9.9698%	-12.1170%
tennis_sif	-7.9806%	-11.6937%
waterfall_cif	-7.8285%	-17.4415%
OVERALL	-8.5493%	-12.3907%

8-bit internal

12-bit internal

Baseline VP9

Compare VP9, VP10 (8-bit), VP10 (12-bit)

Results

- Natural content
  - ***hevcmr*** set
  - 480P resolution

29% BDRATE reduction achieved by VP9/HEVC over H.264

VP10 Comparison Charts

Compare X.264, VP9, X.265

Method For Combining Points (•) BDRATE ( ) BDSNR

Average size reduction to get the same quality (•) Average PSNR ( ) Global PSNR ( ) SSIM

File	nextgen_base_hevcmr	x265_hevcmr
BQMall_832x480_60	-26.7649%	-27.8190%
BasketballDrillText_832x480_50	-33.9045%	-26.7766%
BasketballDrill_832x480_50	-36.8228%	-28.8834%
Flowervase_832x480_30	-31.1168%	-27.9726%
Keiba_832x480_30	-23.8019%	-33.5314%
Mobisode2_832x480_30	-35.1188%	-31.8524%
PartyScene_832x480_50	-31.4817%	-26.0603%
RaceHorses_832x480_30	-20.1191%	-28.9091%
OVERALL	-29.8913%	-28.9756%

Baseline X.264

Results

- Natural content
  - ***hevcmr*** set
  - 480P resolution

VP10 (12 bit) achieves over VP9, close to 50% of the gain HEVC/VP9 achieves over H. 264.

VP10 Comparison Charts

Compare VP9, VP10 (8-bit), VP10 (12-bit)

Method For Combining Points (•) BDRATE ( ) BDSNR

Average size reduction to get the same quality (•) Average PSNR ( ) Global PSNR ( ) SSIM

File	nextgen_allsep18_hevcmr	nextgen_allsep19hbd12_hevcmr
BQMall_832x480_60	-11.3543%	-12.5508%
BasketballDrillText_832x480_50	-12.9620%	-14.4182%
BasketballDrill_832x480_50	-12.0227%	-13.7783%
Flowervase_832x480_30	-11.7762%	-15.8829%
Keiba_832x480_30	-12.7879%	-14.2440%
Mobisode2_832x480_30	-14.6805%	-20.0822%
PartyScene_832x480_50	-10.3321%	-10.4347%
RaceHorses_832x480_30	-10.2327%	-11.6352%
OVERALL	-12.0186%	-14.1283%

8-bit internal

12-bit internal

Baseline VP9

Results

- Screen content
  - ***screen_content*** set
  - Mixed HD resolutions

VP10 (8 bit) achieves close to 20% BDRATE improvement over VP9.

8-bit internal

Codec Comparison Results

Method For Combining Points ○ Average of bitrates difference ○ BDSNR ◉ BDRATE
Average size reduction to get the same quality ◉ Average PSNR ○ Global PSNR ○ SSIM

File	scAllTools_screen
Basketball_Screen	-27.0123%
MissionControlClip2	-22.8170%
MissionControlClip3	-28.0156%
chromoting_numbers	-22.7495%
excel	-33.7551%
gimp	-24.2557%
liquify	-13.2265%
sc_SlideShow	-13.7199%
sc_console	-27.7857%
sc_desktop	-37.7534%
sc_flyingGraphics	-11.2332%
sc_map	-10.5910%
sc_programming	-17.4035%
sc_robot	-1.1726%
sc_socialNetworkMap	-1.4187%
sc_web_browsing	-26.6617%
OVERALL	-19.9732%

Results: Summary

- Summary:
  - **derflr**: low resolution CIF sequences
    - Both VP9/HEVC achieve only 14% BDRATE reduction over H.264
      - On this set, hard to get gains on.
    - VP10: 12% BDRATE reduction over VP9
  - **hevcmr**: 480p resolution HEVC test sequences
    - On this set, both VP9/HEVC achieve 29-30% BDRATE over H.264
    - VP10: 14% BDRATE reduction over VP9
  - **screen_content**: screen content material
    - VP10: 20% BDRATE reduction over VP9
- Overall, good progress at low resolution, less so at higher resolutions.

Outline

- Introduction
- Coding Tools
- Results
- Conclusion

Conclusions

- Some progress towards a next generation codec
  - Still long way to go still esp. for higher resolutions.
  - Need a few big ideas
- Coding efficiency improvement coming at cost of search complexity on encoder side
  - Okay for VOD, Not okay for video conferencing
  - Suggestion for the next codec - split tools based on encode search space
- Welcome to join the VP10 effort!
- All development in the open.