

# Google™ VP-Next

---

## Overview and Progress Update

# Overview

---



Work on Vp-Next started in earnest late in Q3 of 2011. It is an evolutionary development from VP8 not a completely new paradigm.

A focus on larger image formats poses new problems but also presents new opportunities.

- More macroblocks.... but the average complexity per MB is lower. In many cases, encoding information at the level of a 16x16 macroblocks is no longer appropriate.
- Larger scale transforms and prediction modes become increasingly important.

# Overview - Focus

---



- Scale better for larger images
- New prediction modes
- Entropy coding
- New Transforms
- Extended quality range
- Non transform and lossless coding methods

# Overview - Progress

---



# Segmentation

---



- Group together MBs that share common characteristics into segments.
- Optionally encode control flags and features at the segment level.
- Scales really well for large images.
- Most benefit when the segmentation is reasonably stable.
- Experiments with implicit segmentation strategies.

# Segment Codeable Features



## Currently Implemented / being tested.

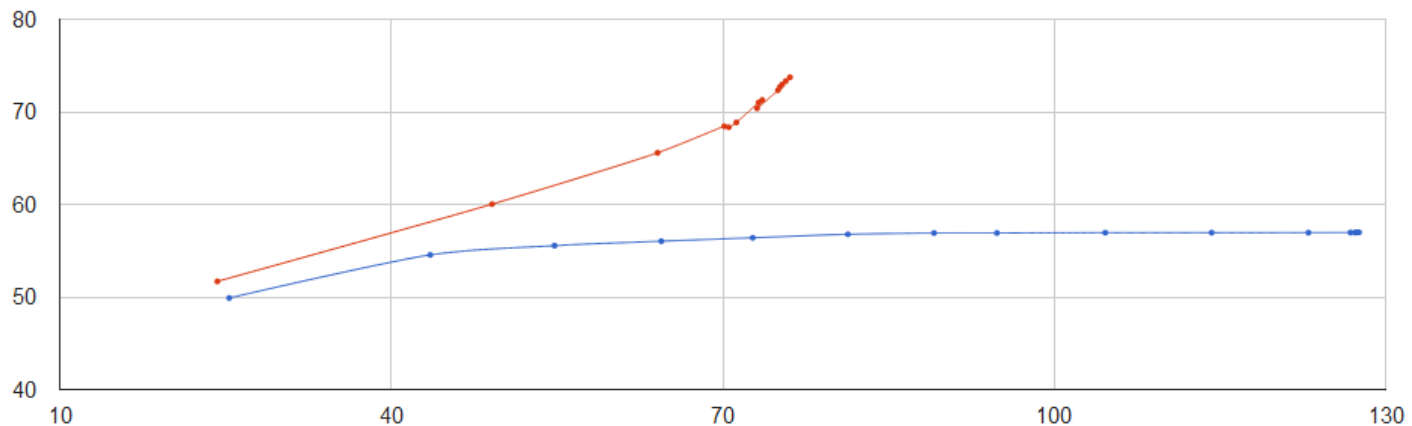
- Quantizer (also in VP8)
- Loop filter strength (also in VP8)
- Prediction Mode
- Reference frame
- Transform type
- End of block position

## Other possible additions

- SB pattern
- Motion
- Prediction Filters
- Custom entropy models

The biggest benefits so far have been at low data rates on low to moderate motion clips (>10% for many clips, some data points as high as 50%).

# Segmentation benefits



Segmentation combined with VP-Next's expanded quantizer range for a low motion clip

# Prediction



- Experiments with 1/8 pel and sub-pixel motion estimation
- Compound prediction using multiple predictors or reference frames.
- New prediction modes linked to new transforms
- Rotation filters instead of or alongside conventional fractional pel filtering
- Filtered prediction modes
- New Intra prediction functions and strategies.



We need to be much more adaptive to extremes and better behaved for large image formats. Coding of complex uncorrelated motion and very dense residual signals is an area where we need to improve.

Researching :-

- New predictive models and contexts.
- More efficient updates / adaptation.
- Segment level and SB level adjustments.

# Superblocks



- HD material often shows correlation over larger areas and is not well handled by VP8 (prediction & transform)
- Introduce the concept of Superblocks: 32x32 & ?64x64
- Mode, ref frame, MV, transform etc may be specified at SB level.
- Use Quadtree to signal coding level
- Relationship to segment-based coding:
  - Segment: good where the segmentation is stable over multiple frames
  - Superblocks: work well short-term but spatially correlated changes

# Super Block Status



- A framework is in place but is not used much yet.
- Coding follows SB rather than MB raster order
- Within a SB, MBs are encoded in raster order
- Cost of coding a SB with common mode, MV, ref-frame is compared to coding separately as 4xMBs (using RD)
- 2 Intra modes have been temporarily disabled because they require data from the above right MB that is not always available when coding using SBs

# Enhanced Interpolation



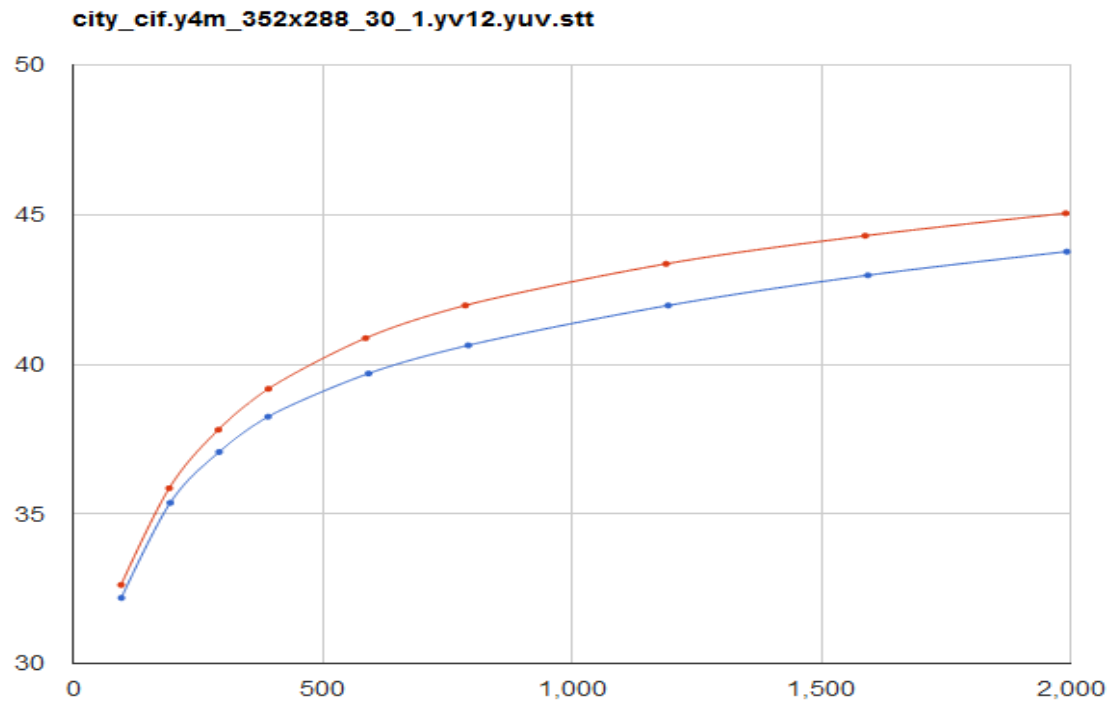
- Sub-pel precision and filters can have a big impact on compression efficiency.... BUT
  - Longer filters -> Better interpolation (but more cost)
  - Higher precision sup pel -> Better interpolation (but more cost)
- Codec Generational Progression
  - full pel ->
  - bilinear 1/2 pel ->
  - 4-6-tap, 1/4 pel (current generation) ->
  - 8-tap, 1/8 pel (next gen)

# VP-Next Enhanced Interpolation



- VP-Next:
  - Added a set of 8-tap filters (still experimenting with different variants)
  - Support for 1/8th pel precision
  - Filters selectable on a per-frame basis
- Experimental:
  - Implicit adaptive interpolation / filter selection, directional interpolation
  - Pre-calculated rotation filters
- Current Results:
  - ~2-3% average improvement but on a small subset the improvement is huge ( $\geq 20\%$ ).

# High Precision MVs



On this clip the improvement was over 20%

# New Entropy Coding



There have already been a number of changes to improve entropy coding efficiency (especially at low data rates)

- Contextual MB skip coding\
- Reference frame contextual coding
- Expanding the previous coef-contexts
- Modifications to coding of explicit segment map (differential coding option and contexts)
- Separate coding context for different frame types

## Next Steps:

- Major overhaul of how motion is coded

# Probability Updates



Explicit probability updates are expensive:-

- Probabilities can also be updated implicitly based on actual counts seen in the encoder and decoder

Other experiments to improve coding efficiency:- These will be particularly important if we want to update at a sub frame level such as segments or super blocks.

- There is no need to always transmit the exact new probability - search for the least expensive updates with the most benefit.
- Model distribution of probabilities and encode an update to the model rather than individual probability values.
- Have a set of baseline models or distributions that can be used as baseline sets chosen using an index or a value such as  $Q$ .



# 8x8 transform:



## Work associated with 8x8 transform:

- Tuning: transform precisions, adjustments to trellis quant
- Add 8x8 intra prediction modes, new loop filters
- Currently we associate transform size decision with prediction modes on Inter coded frame: 16x16 prediction modes-> 8x8, all other prediction modes-> 4x4
- Rate distortion mode decision
- Context conversion between 8x8 and 4x4 coded MB

Sizeable improvement, especially on HD clips

# Other transforms related changes



- Support for a lossless mode using a reversible variant of the Walsh Hadamard transform.
- Significant improvements in the round trip error for the existing DCT and increased quantizer range / precision (even in its lossy mode Vp-Next can get up to ~70dB vs 50-52dB for VP8).
- Work has started on implementing and integrating a 16x16 transform option.
- Ongoing experiments into alternative transforms including adaptive / directional variants.

# new loop filter



- Designed to help the 8x8 transform coded MBs and reduce blocking in high resolution clips
  - Used only on 8x8 transformed block boundaries and replaces previous macroblock edge filter.
  - use existing filter on 4x4 transform block boundaries
- Test for "flat" areas
  - mask :- thresholded flat block detection
  - filtering :-  
filter taps: [1, 1, 1, 2, 1, 1, 1]/8 for flat content  
else standard loop filter
- Helps 8x8 transformed MBs

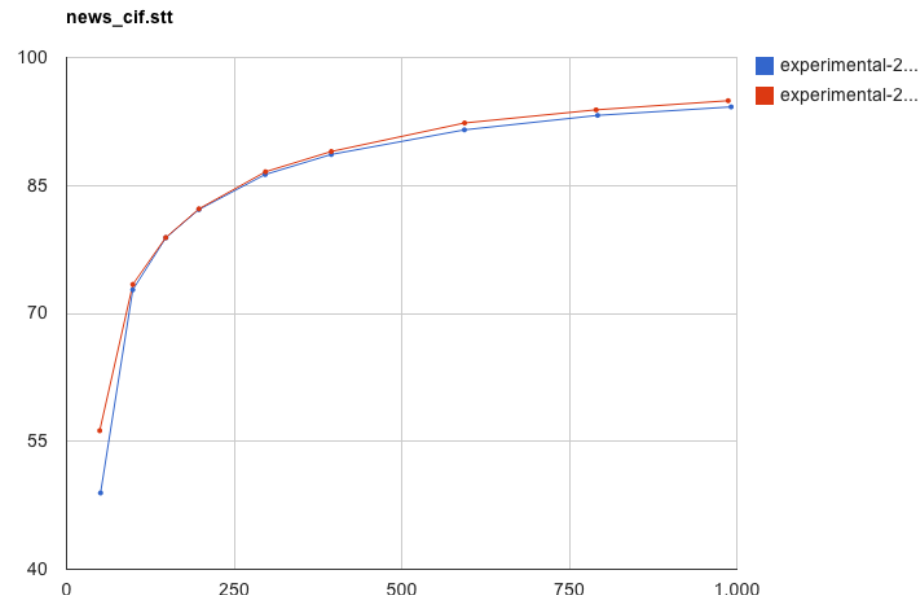


# Predictive Segmentation Map



Changes to the static segment:

- For the reference frame (ARF):
  - decrease Q
  - lower loopfilter strength
- For other frames:
  - force skip or increase Q
  - force ARF as ref frame
  - force mode (zeromv)



# Compound prediction



Combine two "single" (intra or inter) predictors to create a new predictor (inter/inter, intra/intra or inter/intra).

- Currently, it combines two inter predictors using a different reference frame and the same mode
- Predictors are averaged to generate a new predictor



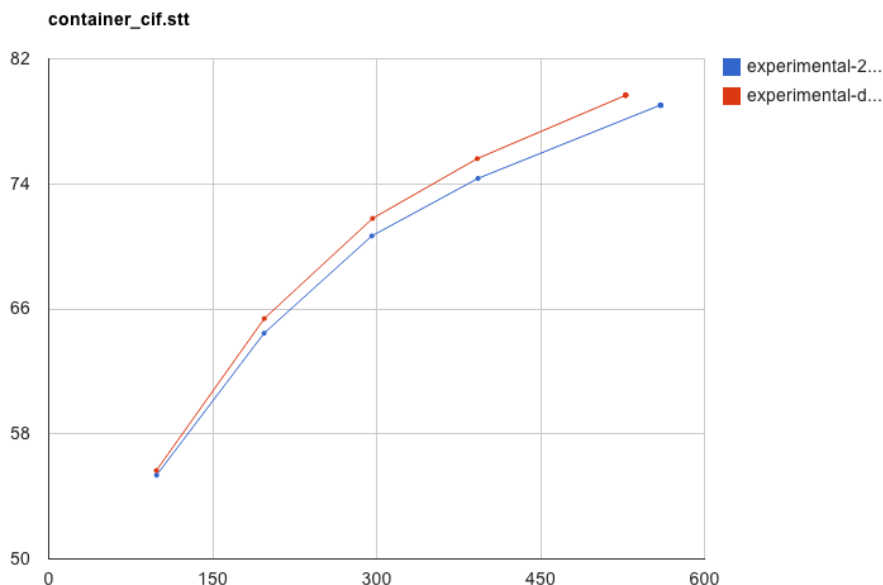
```
0 0 0 0 0 0 0 0 0 0 1 1 1 0 1 0 0 0 0 0 0 0
0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 1
0 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 0
0 1 1 0 1 1 0 0 0 0 0 0 0 1 1 1 1 0 1 1 1
0 1 1 1 1 1 1 0 0 1 0 0 0 1 1 1 1 1 1 1 1
0 1 1 1 0 0 0 0 1 0 0 1 1 1 1 1 0 1 1 1 0
1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1
0 1 1 0 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0
1 0 1 1 1 1 0 1 1 1 0 1 0 0 1 0 1 0 1 1 1
1 1 1 1 0 1 1 0 1 1 1 0 0 1 1 1 0 1 1 1 1
1 1 1 0 1 0 0 1 0 0 0 1 1 1 0 1 1 0 1 0 0
1 1 1 0 0 1 0 0 1 1 1 1 1 1 0 1 0 0 0 1 1
1 1 1 1 0 0 1 0 0 1 0 0 0 0 0 1 0 0 0 1 1
1 1 1 0 1 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1
1 1 1 0 0 1 0 1 1 0 0 1 0 0 1 1 0 1 0 0 0
1 1 0 1 1 1 1 1 1 0 0 0 1 1 0 0 0 1 1 1 0
1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 0 1 0 1
1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 0 0 0
```

# Compound prediction



Combine two "single" (intra or inter) predictors to create a new predictor (inter/inter, intra/intra or inter/intra).

- Currently, it combines two inter predictors using a different reference frame and the same mode.
- Predictors are averaged to generate a new predictor



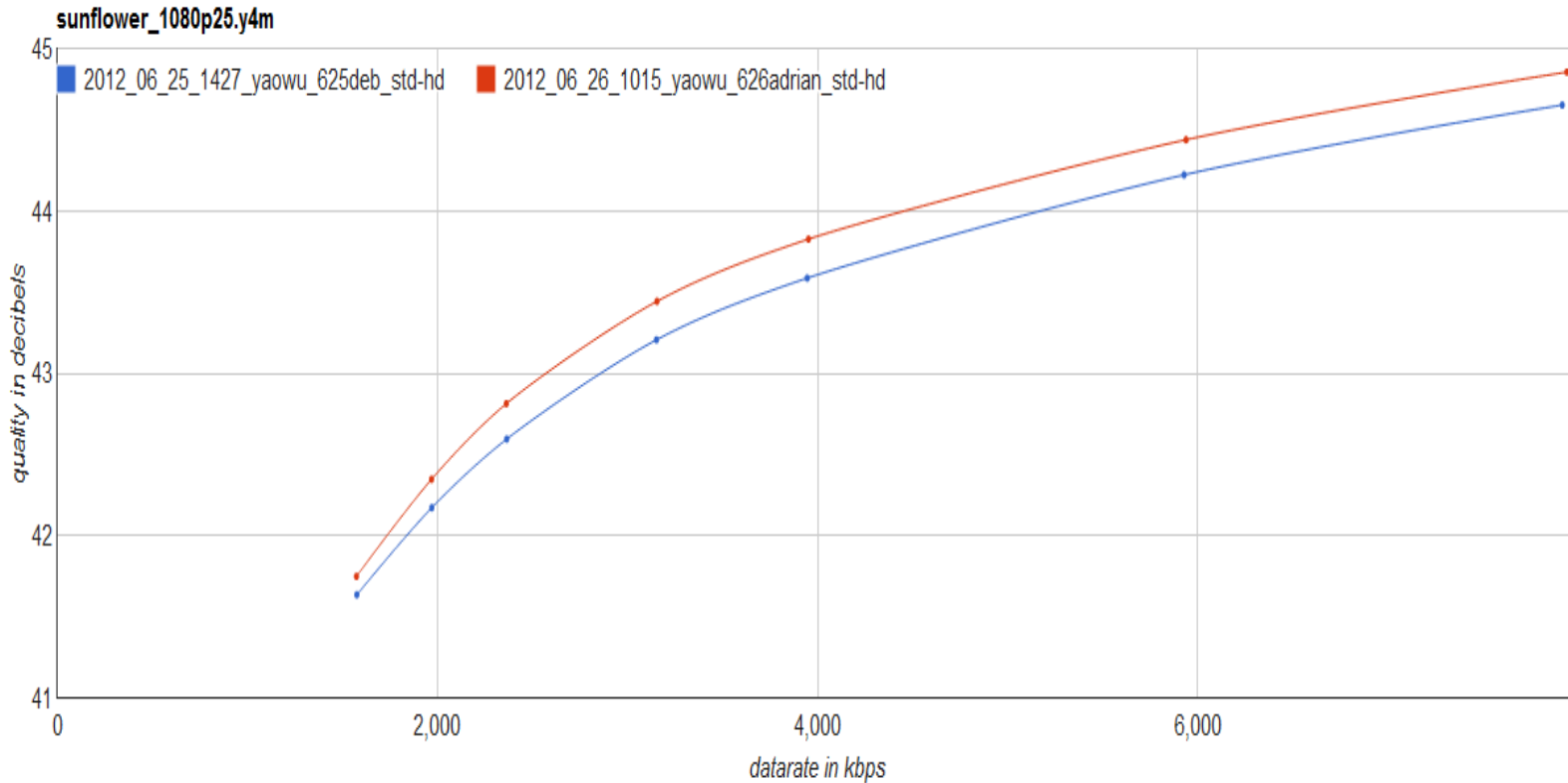
# Prediction Filtering



- Optionally apply a filter to the output of an existing prediction mode to create a modified predictor.
- As currently implemented there is some overlap with other methods and a high signalling overhead. However, there is still an average gain of ~2% and as much as 10% for some clips.
- Future experiments will include:
  - filter weight and shape determined by the direction and amplitude of motion.
  - selective use instead of fractional pel filters.
  - segment level, SB or implicit coding of filter flag.



# Prediction Filtering



# Enhanced Intra Prediction Experiment

---

- Goal is to increase the quality of prediction for intra coded blocks within a mainly inter coded frame.
- Encode the frame as two passes.
  - First encode all blocks that are clearly best coded using INTER modes.
  - Then fill in the INTRA macroblocks as a second pass where they now have access to more "neighbours" for prediction.
- Pros
  - We have already shown that there can be a good increase in the quality of the INTRA prediction Block and hence possible quality gain.
- Cons
  - Makes decode process more complex, especially in hardware.

Any Questions ?





# Background

---



- Video over IP has come of age - the novelty has worn off !
- HD expected even for mobile devices.
- Rising processing power but this is still an issue especially for mobile devices.
- In the future IP video is likely to take the lead on quality / resolution etc.

# But....



- Not all the world has Multi mb/s broadband and 3/4G
- We should not neglect small resolutions and low bandwidth use cases.
- Not everybody has a device that is less than 6 months old.

# VP-Next Objectives

---



- Build a compelling successor to VP8 as a next generation codec for the web.
- Smaller / Faster / Lighter ???  
(well ok... only a little bit slower / heavier)
- Vp-next IS still an evolutionary development not a completely new paradigm.
- HD is the new "sweet spot".