Daala: Building a Next-Generation Video Codec from Unconventional Technology

Jean-Marc Valin, Timothy B. Terriberry, Nathan E. Egge, Thomas J. Daede,
Yushin Cho, Christopher Montgomery, Michael Bebenita

mozilla research

xiph.org

The Daala Project

Project goals:

- Royalty-free video codec
- Replacing traditional tools with new/uncommon ones
- Exploring new ideas without constraints

Effort is now part of the Alliance for Open Media's (AOM) AV1 codec

Techniques

Main Daala techiques:

- Lapped transforms
- Overlapped-block motion compensation (OBMC)
- Perceptual vector quantization (PVQ)
- Chroma from luma (CfL) prediction
- Haar DC
- Multi-symbol entropy coding
- Deringing filter

In AV1 | Considered for AV1 | Not considered for AV1

Lapped Transforms

4x4 to 64x64 DCTs, with 4-point lapping

Advantages:

- No blocking artefacts
- Better energy compaction
- Perfect reconstruction

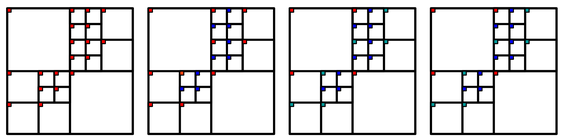
Disadvantages:

- Cannot use traditional intra prediction
- Have to use fixed size lapping (search)

Haar DC

Hierarchically code DC coefficients using Haar transform

Compensate for lack of intra predictor

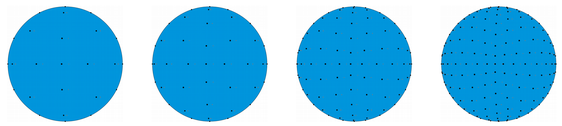


Perceptual Vector Quantization

Gain-shape vector quantization based on a spherical projection of the pyramid vector quantizer in N dimensions

- Gain = contrast
- Shape = details
- Number of pulses K based on gain

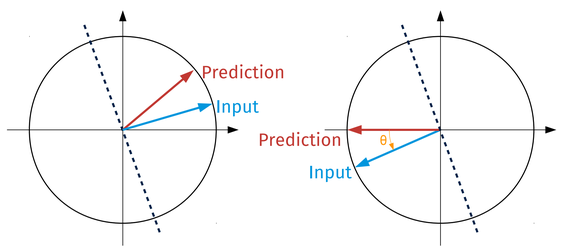
$$\mathbf{y} \in \mathbb{Z}^N : \sum_{i=0}^{N-1} |y_i| = K$$



K=3 (5.25 bits) K=4 (6.04 bits) K=6 (7.19 bits) K=8 (8.01 bits)

Using prediction:

- ~~Subtracting prediction from input~~
- Transform input using prediction
- Use Householder reflection to align prediction with one axis
- Introduce angle θ between input and prediction
- Code sphere in N-1 dimensions
- Optional *no reference* coding

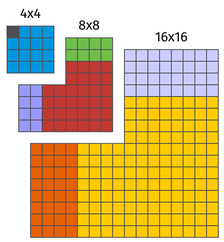


PVQ can take advantage of contrast masking

- Better resolution for small gain
- Quantize companded gain
- Can be done with no signaling

Split coeffs in *bands*

- DC coded separately
- Recursive subdivision
- Split per octave
- Split per direction

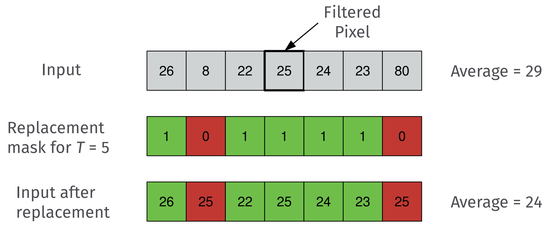


Deringing Filter

Overview:

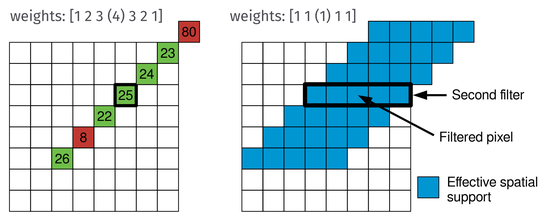
- Conditional replacement filter
- Directional 35-tap (5x7) separable filter
- Decoder-side direction estimation
- Computed on 8x8 blocks
- Strength signaled on superblocks (64x64)
- Completely vectorizable (SIMD)

Conditional replacement filter



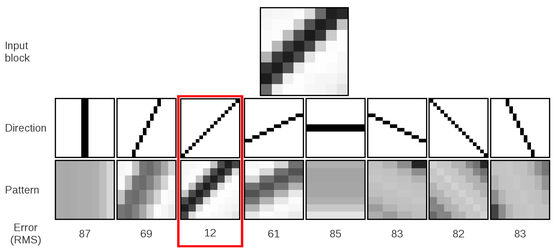
Directional Filtering:

- 7-tap filter along direction
- 5-tap filter across lines (lower threshold)



Direction estimation:

- Minimize error compared to directional line averages
- Fast, vectorizable algebraic simplifications

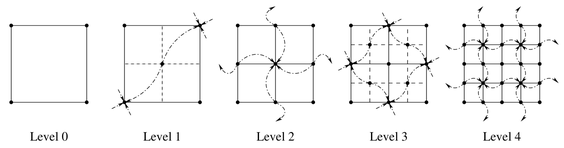


Threshold based on signaled global threshold, signaled superblock adjustment, non-signaled block variance measurement

OBMC

Blending based on 4-8 mesh

- Only double MVs for each level



Chroma from Luma (CfL)

Luma and chroma are highly correlated, so we can predict chroma from luma

- Chroma shape is predicted from luma
- Code gain and sign
- Luma optionally down-sampled (4:2:0)

Multi-Symbol Entropy Coder

Entropy decoding is a (serial) bottleneck in video decoding. We can reduce the cost by increasing the alphabet size and coding fewer symbols. Daala uses alphabet sizes up to 16.

Alliance for Open Media

AOM's new AV1 codec based on

- All of VP9 (Google)
- Parts of Daala (Xiph.Org/Mozilla)
- Parts of Thor (CISCO)
- New contributions

Results

Progress over the past 3 years