Daala: A Perceptually-Driven Still Picture Codec

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

ICIP 2016

https://jmvalin.ca/slides/icip2016_slides.pdf

mozilla

xiph.org

Introduction

- Daala is a Royalty-free video codec

- Replaces traditional coding tools with new/uncommon ones

- Intra frames used as still picture codec
  - Evaluated at PCS 2015
  - Significantly improved since then

Daala Techniques

- Lapped transforms
  - 4x4 to 64x64 DCTs, 4-point lapping
- Multi-symbol arithmetic coding
  - Alphabet size up to 16 (fewer symbols to code)
- Perceptual vector quantization (PVQ)
  - Gain-shape quantization with spherical quantizer
  - Signal-free activity masking
- Chroma from luma (CfL) prediction
- Deringing filter

PCS 2015

- Did well one some images

- Not so well on others

0.25 bpp

bike
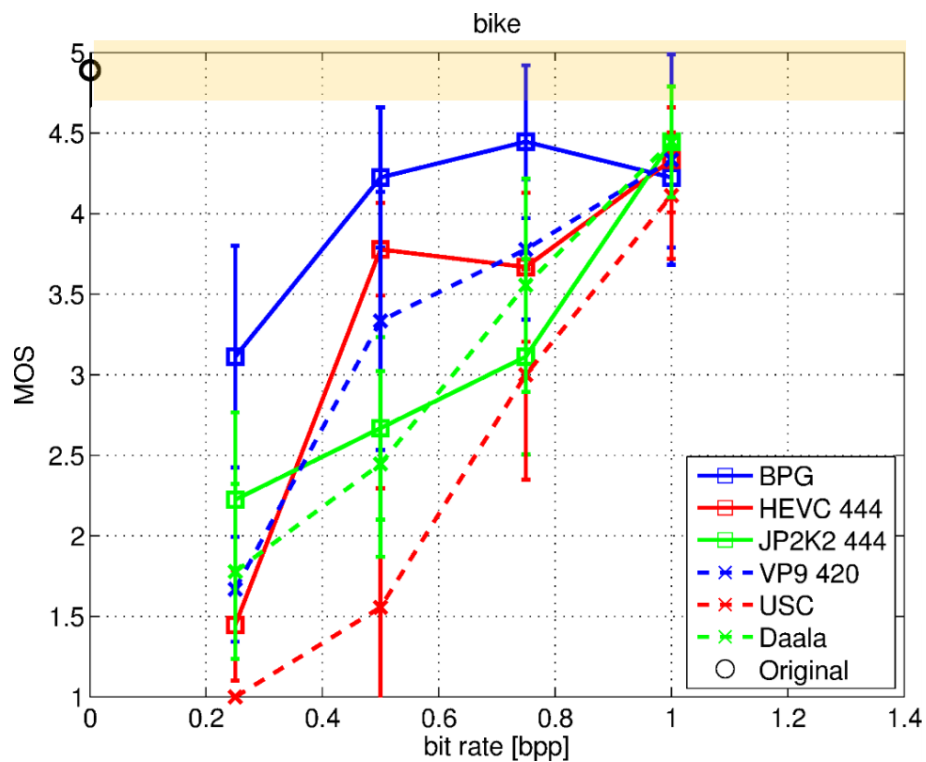
Improvements Since PCS 2015

- New deringing filter
- Finer chroma quantization
- Lapping reduced to 4 points
- 64x64 DCT
- Reduced-overhead entropy coder
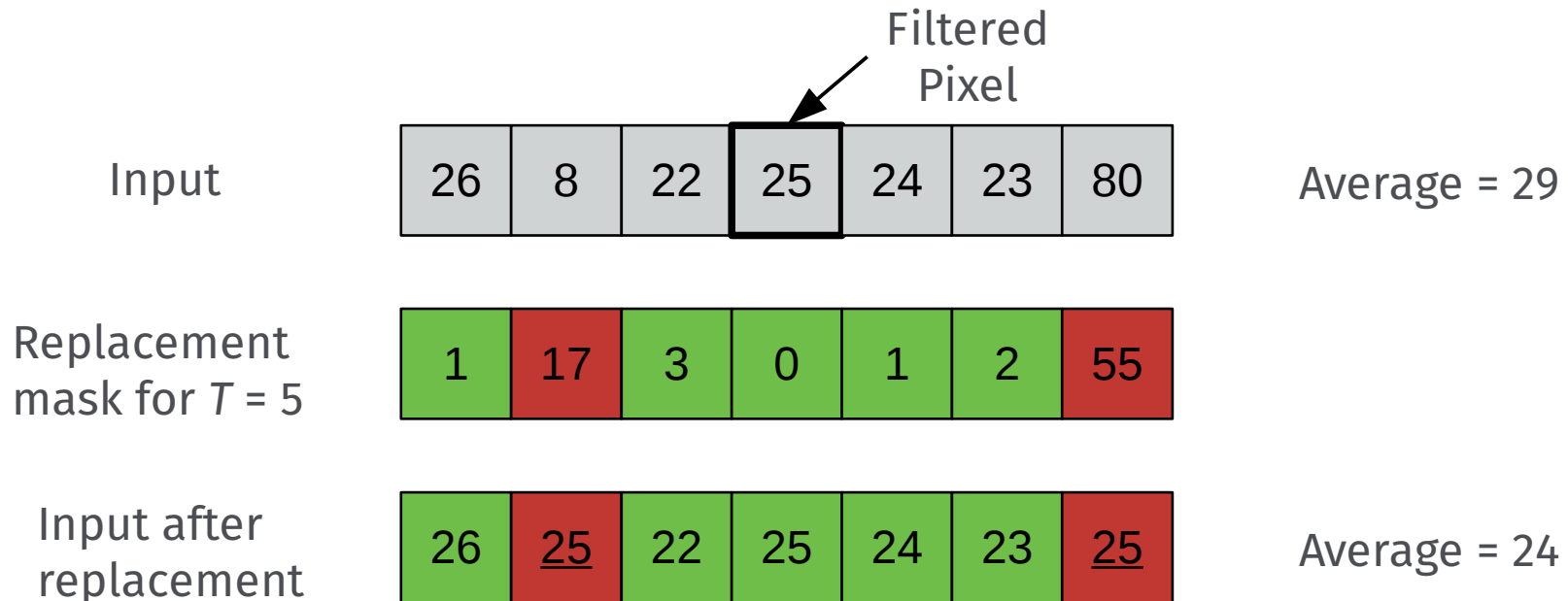
Deringing Filter Design Goals

- Smooth out ringing artifacts

- Preserve edges and texture

- Have reasonable complexity

  – Be easy to vectorize (SIMD)

- Originally designed for video

Deringing Filter Overview

- Computed on coded 8x8 blocks

- Conditional replacement filter

- Directional 35-tap separable filter

- Decoder-side direction estimation (no signaling)

- Filter strength signaled on coded superblocks (64x64)

Conditional Replacement Filter

- For each filtered pixel
  - Replace tap values too dissimilar to center pixel
  - $mask_{i,j} := |tap_{i,j} - center_i| < T$
  - IF $mask_{i,j} = 0$ THEN $tap_{i,j} := center_i$

Filtered Pixel

26	8	22	25	24	23	80

Input — Average = 29

1	17	3	0	1	2	55

Replacement mask for $T = 5$

26	25	22	25	24	23	25

Input after replacement — Average = 24

Conditional Replacement Filter (cont)

- Express filter in terms of difference to center

$$y(n) = x(n) + \frac{1}{\sum_k w_k} \sum_{k, k \neq 0} w_k R\left(x(n+k) - x(n), T\right)$$
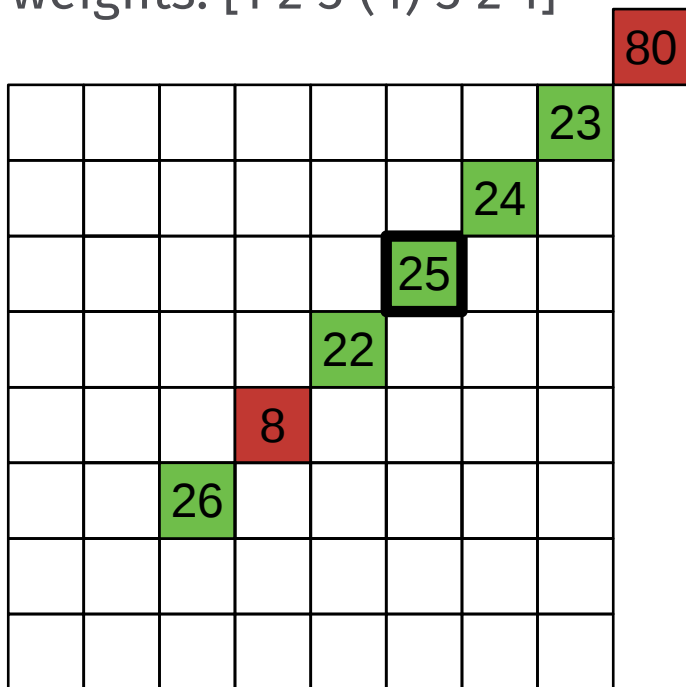
$$R(x, T) = \begin{cases} x & , \ |x| < T \\ 0 & , \ \text{otherwise} \end{cases}$$

- Weight normalization is constant
- $R(x,T)$ easy to compute
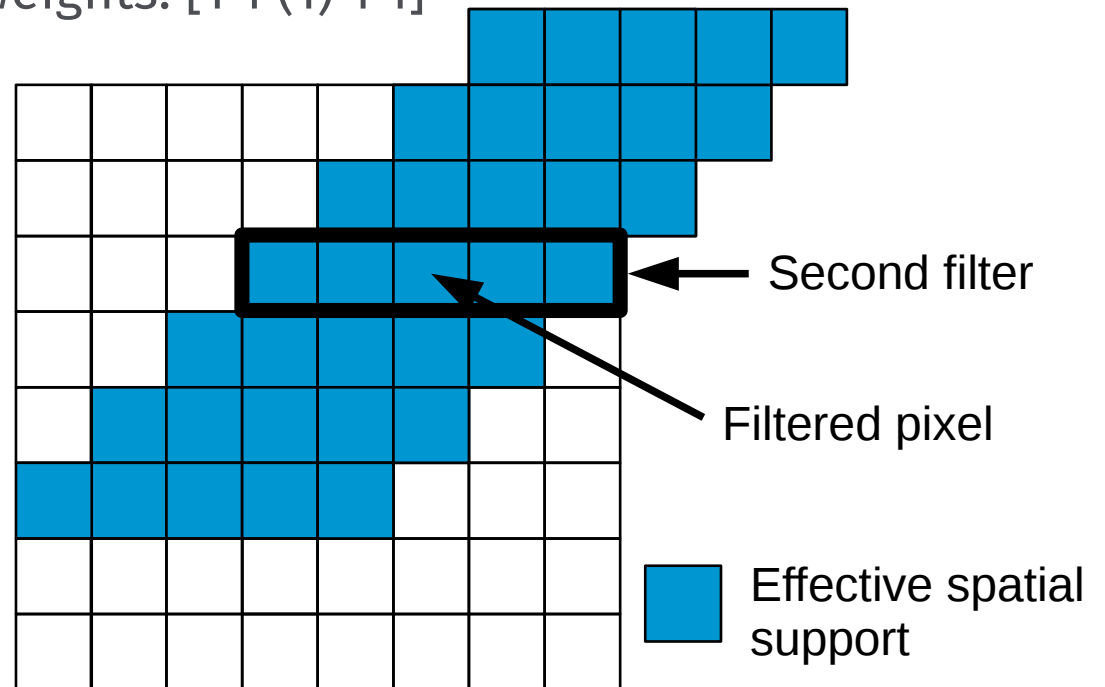- Vectorizes completely

Directional Filtering

- 7-tap filter along direction
- 5-tap filter across lines (lower threshold)
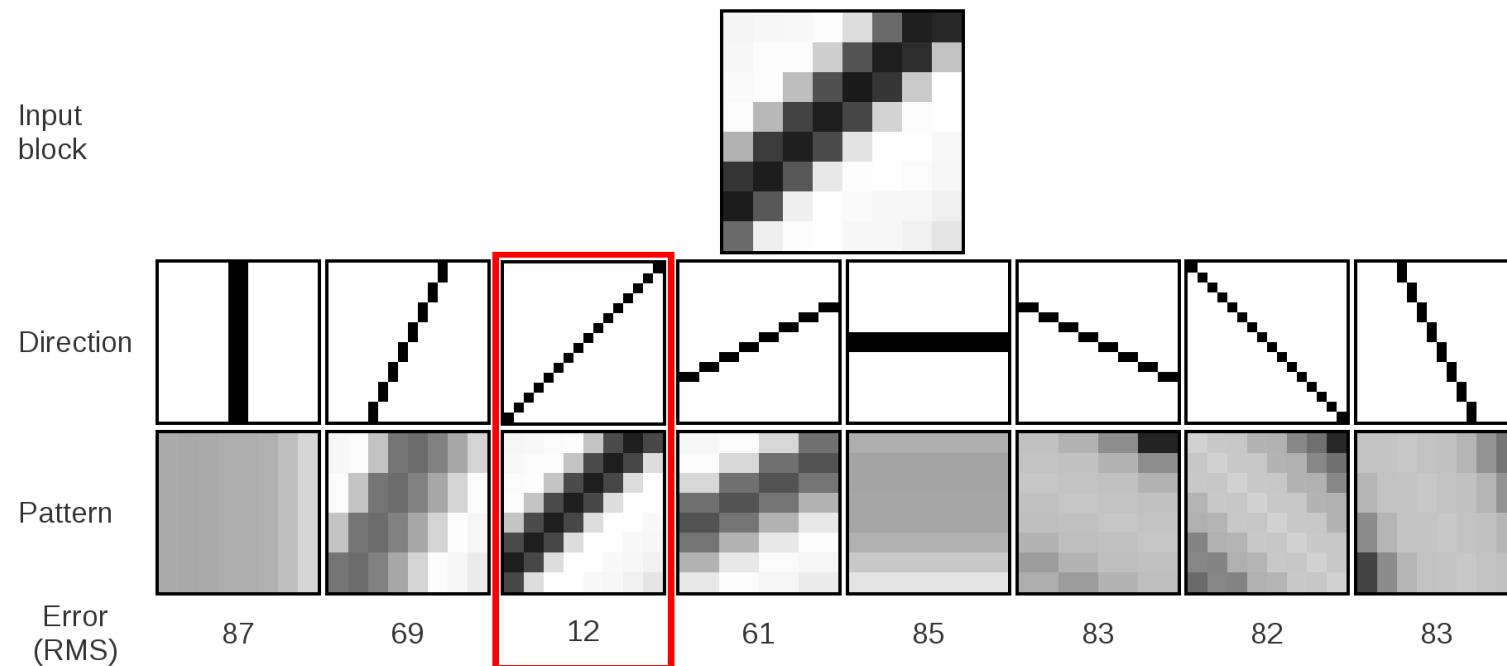
weights: [1 2 3 (4) 3 2 1]

weights: [1 1 (1) 1 1]

Second filter

Filtered pixel

Effective spatial support

Direction Estimation

- Minimize error between decoded image and directional line averages

- Fast, vectorizable algebraic simplifications

Input block

Direction

Pattern

Error (RMS)   87   69   12   61   85   83   82   83

Before Deringing

After Deringing

Results

- Compared to PCS 2015 (BD-rate)
  - New deringing filter
  - ~~Finer chroma quantization~~
  - Lapping reduced to 4 points
  - 64x64 DCT
  - Reduced-overhead entropy coder

Metric	Low (%)	Medium (%)	High (%)
PSNR	-6.3	-6.9	-7.9
PSNR-HVS	-6.7	-6.6	-6.3
SSIM	-4.8	-5.3	-6.5
FAST-SSIM	-2.4	-2.2	-0.6

Bike 0.25 bpp (PCS 2015)

Bike 0.25 bpp (ICIP 2016)

Future Work

- Experiment with intra prediction

- Improve non-photographic content coding

- Add alpha channel support

- Support for other color spaces (e.g. YCgCo)

- Define container format

- Address results from ICIP 2016 Grand Challenge

Conclusion

- Daala could be a good starting point for a royalty-free still image codec

Questions?