NETVC Working Group
Internet-Draft
Intended status: Informational
Expires: May 19, 2018

N. Egge
L. Trudeau
Mozilla
D. Barr
Xiph.Org Foundation

November 15, 2017

Chroma From Luma Intra Prediction for NETVC
draft-egge-netvc-cfl-01

Abstract

Chroma from luma (CfL) prediction is a new and promising chroma-only intra predictor that models chroma pixels as a linear function of the coincident reconstructed luma pixels. In this document, we propose the CfL predictor adopted in Alliance Video 1 (AV1) to the NETVC working group. The proposed CfL distinguishes itself from prior art not only by reducing decoder complexity, but also by producing more accurate predictions. On average, CfL reduces the BD-rate, when measured with CIEDE2000, by 5% for still images and 2% for video sequences.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on May 19, 2018.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of

publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction 2
2. State of the Art in Chroma from Luma Prediction 4
3. Model Fitting the "AC" Contribution 5
4. Chroma "DC" Prediction for "DC" Contribution 6
5. Parameter Signaling 7
6. Experimental Results 8
7. Conclusion 10
8. Informative References 10
Authors' Addresses 12

1. Introduction

Still image and video compression is typically not performed using red, green, and blue (RGB) color primaries, but rather with a color space that separates luma from chroma. There are many reasons for this, notably that luma and chroma are less correlated than RGB, which favors compression; and also that the human visual system is less sensitive to chroma allowing one to reduce the resolution in the chromatic planes, a technique know as chroma subsampling [Wang01].

Another way to improve compression in still images and videos is to subtract a predictor from the pixels. When this predictor is derived from previously reconstructed information inside the current frame, it is referred to as an intra prediction tool. In contrast, an inter prediction tool uses information from previously reconstructed frames. For example, "DC" prediction is an intra prediction tool that predicts the pixels values in a block by averaging the values of neighboring pixels adjacent to the above and left borders of the block [Li14].

Chroma from luma (CfL) prediction is a new and promising chroma-only intra predictor that models chroma pixels as a linear function of the coincident reconstructed luma pixels [Kim10]. It was proposed for the HEVC video coding standard [Chen11b], but was ultimately rejected, as the decoder model fitting caused a considerable complexity increase.

More recently, CfL prediction was implemented in the Thor codec [Midtskogen16] as well as in the Daala codec [Egge15]. The

inherent conceptual differences in the Daala codec, when compared to HEVC, led to multiple innovative contributions by Egge and Valin [Egge15] to CfL prediction. Most notably a frequency domain implementation and the absence of decoder model fitting.

As both Thor and Daala are part of NETVC working group, a research initiative was established regarding CfL, the results of which are presented in this draft. The proposed CfL implementation not only builds on the innovations of [Egge15], but does so in a way that is compatible with the more conventional compression tools found in Alliance Video 1 (AV1). The following table details the key differences between LM Mode [Chen11b], Thor CfL [Midtskogen16], and Daala CfL [Egge15] (the previous version of this draft):

	LM Mode	Thor CfL	Daala CfL	AV1 CfL
Prediction Domain	Spatial	Spatial	Frequency	Spatial
Bitstream Signaling	No	No	Sign bit PVQ Gain	Signs + Index
Requires PVQ	No	No	Yes	No
Encoder Model Fitting	Yes	Yes	Via PVQ	Search
Decoder Model Fitting	Yes	Yes	No	No

This new implementation is considerably different from its predecessors. Its key contributions are:

- Parameter signaling, which avoids model fitting on the decoder and, as explained in Section 2, results in more precise predictions, as the chroma reference pixels are used for fitting (which is impossible when fitting on the decoder). The actual signaling is described in Section 5.

- Model fitting the "AC" contribution of the reconstructed luma pixels, as shown in Section 3, which simplifies the model and allows for a more precise fit.

- Chroma "DC" prediction for "DC" contribution, which requires no signaling and, as described in Section 4, is more precise.

Finally, Section 6 presents detailed results of the compression gains of the proposed CfL prediction implementation in AV1.

2. State of the Art in Chroma from Luma Prediction

As described in [Kim10], CfL prediction models chroma pixels as a linear function of the coincident reconstructed luma pixels. More precisely, Let L be an M X N matrix of pixels in the luma plane; we define C to be the chroma pixels spatially coincident to L. Since L is not available to the decoder, the reconstructed luma pixels, L^r, corresponding to L are used instead. The chroma pixel prediction, C^p, produced by CfL uses the following linear equation:

```
C^p = alpha * L^r + beta
```

Some implementations of CfL [Kim10], [Chen11b] and [Midtskogen16] determine the linear model parameters alpha and beta using linear least-squares regression

```
                ___ ___                    ___ ___           ___ ___
                \   \                      \   \             \   \
alpha = (M*N) /__ /__ L^r(i,j)*C(i,j) - /__ /__ L^r(i,j) /__ /__ C(i,j)
                i=0 j=0                    i=0 j=0           i=0 j=0
         ---------------------------------------------------------------

                      ___ ___                  ___ ___
                      \   \                    \   \
                (M*N) /__ /__ (L^r(i,j))^2 - (/__ /__ L^r(i,j))^2
                      i=0 j=0                  i=0 j=0

                      ___ ___                 ___ ___
                      \   \                   \   \
                beta = /__ /__ C(i,j) - alpha /__ /__ L^r(i,j)
                      i=0 j=0                 i=0 j=0
                      ---------------------------------------
                                   (M*N)
```

We classify [Kim10], [Chen11b], and [Midtskogen16] as implicit implementations of CfL, since alpha and beta are not signaled in the bitstream, but are implied from the bitstream. The main advantage of the implicit implementation is the absence of signaling.

However, implicit implementations have numerous disadvantages. As mentioned before, computing least squares considerably increases decoder complexity. Another important disadvantage is that the chroma pixels, C, are not available when computing least squares on the decoder. As such, prediction error increases since neighboring reconstructed chroma pixels must be used instead.

In [Egge15], the authors argue that the advantages of explicit signaling considerably outweigh the signaling cost. Based on these

findings, we propose a hybrid approach that signals alpha and implies beta.

3. Model Fitting the "AC" Contribution

In [Egge15], Egge and Valin demonstrate the merits of separating the "DC" and "AC" contributions of the frequency domain CfL prediction. In the pixel domain, the "AC" contribution of a block can be obtained by subtracting it by its average.

An important advantage of the "AC" contribution is that it is zero mean, which results in significant simplifications to the least squares model parameter equations. More precisely, let L^AC$ be the zero-meaned reconstructed luma pixels. Because

```
___ ___
\   \
/__ /__ L_AC(i,j) = 0
i=0 j=0
```

substituting L^r by L_AC yields the following simplified model parameters equations:

```
                ___ ___
                \   \
alpha_AC = /__ /__ L_AC(i,j)*C(i,j)
           i=0 j=0
           ------------------------
           ___ ___
           \   \
           /__ /__ (L^r(i,j))^2
           i=0 j=0

               ___ ___
               \   \
     beta_AC = /__ /__ C(i,j)
               i=0 j=0
               --------------
                    (M*N)
```

We define the zero-mean chroma prediction, C_AC, like so

```
C_AC = alpha_AC * L_AC + beta_AC
```

When computing the zero-mean reconstructed pixels, the resulting values are stored using 1/8th precision fixed-point values. This ensures that even with 12-bit integer pixels, the average can be stored in a 16-bit signed integer.

By combining the luma subsampling step with the average subtraction step not only do the equations simplify, but the subsampling divisions and the corresponding rounding error are removed. The equation corresponding to the combination of both steps simplifies to:

```
                          __sy-1__sx-1
                          \     \
L_AC(i,j) =  8   (/__   /__L^r(sy*i+y,sx*j+x))
            -----  y=0   x=0
            sy*sx
       ___ ___        __sy-1__sx-1
       \   \          \     \
     - /__ /__   8   (/__   /__L^r(sy*i+y,sx*j+x))
       i=0 j=0 -----   y=0   x=0
               sy*sx
       ------------------------------------------
                         (M*N)
```

Note that this equation uses an integer division.

In the previous equation, sx and sy are the subsampling steps for the x and y axes, respectively. The proposed CfL only supports 4:2:0, 4:2:2, 4:4:0 and 4:4:4 chroma subsamplings [Wang01], for which:

sy*sx in {1, 2, 4}.

Also, because both M and N are powers of two, M * N is also a power of two. It follows that the previous integer divisions can be replaced by bit shift operations.

4. Chroma "DC" Prediction for "DC" Contribution

Switching the linear model to use zero mean reconstructed luma pixels also changes beta_AC, to the extent that it now only depends on C. More precisely, beta_AC is the average of the chroma pixels.

The chroma pixel average for a given block is not available in the decoder. However, there already exists an intra prediction tool that predicts this average. When applied to the chroma plane, the "DC" prediction predicts the pixel values in a block by averaging the values of neighboring pixels adjacent to the above and left borders of the block [Li14].

Concretely, the output of the chroma "DC" predictor can be injected inside the proposed CfL implementation as an approximation for beta_AC.

The proposed CfL prediction is expressed as follows:

$$CfL(alpha) = alpha * L_AC + DC_PRED.$$

5. Parameter Signaling

Signaling the scaling parameters allows encoder-only fitting of the linear model. This reduces decoder complexity and results in a more precise prediction, as the best scaling parameter can be determined based on the reference chroma pixels which are only available to the encoder. The scaling parameters for both chromatic planes are jointly coded using the following scheme.

First, we signal the joint sign of both scaling parameters. A sign is either negative, zero, or positive. In the proposed scheme, signaling (zero, zero) is not permitted as it results in "DC" prediction. It follows that the joint sign requires an eight-value symbol.

As for each scaling parameter, a 16-value symbol is used to represent values ranging from 0 to 2 with a step of 1/8th. The entropy coding details are beyond the scope of this document; however, it is important to note that a 16-value symbol fully utilizes the capabilities of the multi-symbol entropy encoder [Valin16]. Finally, scaling parameters are signaled only if they are non-zero.

Signaling the scaling parameters fundamentally changes their selection. In this context, the least-squares regression used in [Kim10], [Chen11b], and [Midtskogen16] does not yield an RD-optimal solution as it ignores the trade-off between the rate and the distortion of the scaling parameters.

For the proposed CfL prediction, the scaling parameter is determined using the same rate-distortion optimization mechanics as other coding tools and parameters of AV1. Concretely, given a set of scaling parameters A, the selected scaling parameter is the one that minimizes the trade-off between the rate and the distortion

$$alpha = \underset{a \text{ in } A}{\operatorname{argmin}} \left(D(CfL(a)) + lambda\ R(a) \right).$$

In the previous equation, the distortion, D, is the sum of the squared error between the reconstructed chroma pixels and the reference chroma pixels. Whereas, the rate, R, is the number of bits required to encode the scaling parameter and the residual coefficients. Furthermore, lambda is the weighing coefficient between rate and distortion used by AV1.

6. Experimental Results

To ensure a valid evaluation of coding efficiency gains, our testing methodology conforms to that of [Daede17]. All simulation parameters and a detailed sequence-by-sequence breakdown for all the results presented in this paper are available online at [AWCY]. Furthermore, the bitstreams generated in these simulations can be retrieved and analyzed online at [Analyzer].

The following tables show the average percent rate difference measured using the Bjontegaard rate difference, also known as BD-rate [Bjontegaard01]. The BD-rate is measured using the following objective metrics: PSNR, PSNR-HVS [Egiazarian2006], SSIM [Wang04], CIEDE2000 [Yang12] and MSSIM [Wang03]. Of all the previous metrics, only the CIEDE2000 considers both luma and chroma planes. It is also important to note that the distance measured by this metric is perceptually uniform [Yang12].

As required in [Daede17], for individual feature changes in libaom, we use quantizers: 20, 32, 43, and 55. We present results for three test sets: Objective-1-fast [Daede17], Subset1 [Testset] and Twitch [Testset].

In the following table, we present the results for the Subset1 test set[AWCYSubset1]. This test set contains still images, which are ideal to evaluate the chroma intra prediction gains of CfL when compared to other intra prediction tools in AV1.

	PSNR	PSNR Cb	PSNR Cr	PSNR HVS	SSIM	MS SSIM	CIEDE 2000
Average	-0.53	-12.87	-10.75	-0.31	-0.34	-0.34	-4.87

For still images, when compared to all of the other intra prediction tools of AV1 combined, CfL prediction reduces the rate by an average of 5% for the same level of visual quality measured by CIEDE2000.

For video sequences, next table breaks down the results obtained over the objective-1-fast test set [AWCYObjective1].

	PSNR	PSNR Cb	PSNR Cr	PSNR HVS	SSIM	MS SSIM	CIEDE 2000
Average	-0.43	-5.85	-5.51	-0.42	-0.38	-0.40	-2.41
1080p	-0.32	-6.80	-5.31	-0.37	-0.28	-0.31	-2.52
1080psc	-1.82	-17.76	-12.00	-1.72	-1.71	-1.75	-8.22
360p	-0.15	-2.17	-6.45	-0.05	-0.10	-0.04	-0.80
720p	-0.12	-1.08	-1.23	-0.11	-0.07	-0.12	-0.52

Not only does CfL yield better intra frames, which produces a better reference for inter prediction tools, but it also improves chroma intra prediction in inter frames. We observed CfL predictions in inter frames when the predicted content was not available in the reference frames. As such, CfL prediction reduces the rate of video sequences by an average of 2% for the same level of visual quality when measured with CIEDE2000.

The average rate reductions for 1080psc are considerably higher than those of other types of content. This indicates that CfL prediction considerably outperforms other AV1 predictors for screen content coding. As shown in the following table, the results on the Twitch test set [AWCYTwitch], which contains only gaming-based screen content, corroborates this finding.

	PSNR	PSNR Cb	PSNR Cr	PSNR HVS	SSIM	MS SSIM	CIEDE 2000
Average	-1.01	-15.58	-9.96	-0.93	-0.90	-0.81	-5.74

Furthermore, individual sequences in the Twitch test set show considerable gains. We present the results for Minecraft_10_120f (Mine), GTAV_0_120F (GTAV), and Starcraft_10_120f (Star) in the following table. It would appear that CfL prediction is particularly efficient for sequences of the game Minecraft both sequences reduces the average rate by 20% for the same level of visual quality measured by CIEDE2000.

	PSNR	PSNR Cb	PSNR Cr	PSNR HVS	SSIM	MS SSIM	CIEDE 2000
Mine	-3.76	-31.44	-25.54	-3.13	-3.68	-3.28	-20.69
GTAV	-1.11	-15.39	-5.57	-1.11	-1.01	-1.04	-5.88
Star	-1.41	-6.18	-6.21	-1.43	-1.38	-1.43	-4.15

7. Conclusion

In this document, we presented the chroma from luma prediction tool adopted in AV1 that we proposed for NETVC. This new implementation is considerably different from its predecessors. Its key contributions are: parameter signaling, model fitting the "AC" contribution of the reconstructed luma pixels, and chroma "DC" prediction for "DC" contribution. Not only do these contributions reduce decoder complexity, but they also reduce prediction error; resulting in a 5% average reduction in BD-rate, when measured with CIEDE2000, for still images, and 2% for video sequences.

Possible improvements to CfL for AV2 include non-linear prediction models and motion-compensated CfL.

8. Informative References

[Analyzer]
Bebenita, M., "AV1 Bitstream Analyzer", Mozilla https://arewecompressedyet.com/analyzer/, n.d..

[AWCY] "Are We Compressed Yet?", Xiph.Org Foundation https://arewecompressedyet.com, n.d..

[AWCYObjective1]
Trudeau, L., "Results of Chroma from Luma over the Objective-1-fast test set", Are We Compressed Yet? https://doi.org/10.6084/m9.figshare.5577778.v1, November 2017.

[AWCYSubset1]
Trudeau, L., "Results of Chroma from Luma over the Subset1 test set", Are We Compressed Yet? https://doi.org/10.6084/m9.figshare.5577661.v2, November 2017.

[AWCYTwitch]
Trudeau, L., "Results of Chroma from Luma over the twitch test set", Are We Compressed Yet? https://doi.org/10.6084/m9.figshare.5577946.v1, November 2017.

[Bjontegaard01]
Bjontegaard, G., "Calculation of average PSNR differences between RD-curves", Video Coding Experts Group (VCEG) of ITU-T VCEG-M33, 2001.

[Chen11b] Chen, J., Seregin, V., Han, W., Kim, J., and B. Jeon, "CE6.a.4: Chroma intra prediction by reconstructed luma samples", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-E266, March 2011.

[Daede17] Daede, T., Norkin, A., and I. Brailovsky, "Video Codec Testing and Quality Measurement", IETF NETVC Internet-Draft draft-ietf-netvc-testing-05, March 2017.

[Egge15] Egge, N. and J. Valin, "Predicting chroma from luma with frequency domain intra prediction", Proceedings of SPIE 9410, Visual Information Processing and Communication VI, March 2015.

[Egiazarian2006]
Egiazarian, K., Astola, J., Ponomarenko, N., Lukin, V., Battisti, F., and M. Carli, "Two new full-reference quality metrics based on HVS", Proceedings of the Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM, January 2006.

[Kim10] Kim, J., Park, S., Choi, Y., Jeon, Y., and B. Jeon, "New intra chroma prediction using inter-channel correlation", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-B021, January 2010.

[Li14] Ze-Nian, L., Drew, M., and J. Liu, "Fundamentals of Multimedia", ISBN 3319052896, org Springer Publishing Company, Incorporated, edition 2nd, 2014.

[Midtskogen16]
Midtskogen, S., "Improved chroma prediction", draft-midtskogen-netvc-chromapred-02 IETF NETVC Internet-Draft, October 2016.

[Testset] Daede, T., "Test Sets", Hosted by the Xiph.org Foundation https://people.xiph.org/~tdaede/sets/, n.d..

[Valin16] Valin, J., Terriberry, T., Egge, N., Daede, T., Cho, Y., Montgomery, C., and M. Bebenita, "Daala: Building A Next-Generation Video Codec From Unconventional Technology", Multimedia signal processing (MMSP) workshop arXiv:1608.01947, September 2016.

[Wang01] Wang, Y., Zhang, Y., and J. Ostermann, "Video Processing and Communications", ISBN 23132985, Prentice Hall PTR, Upper Saddle River, NJ, USA, edition 1st, 2001.

[Wang03] Wang, Z., Simoncelli, E., and A. Bovik, "Multiscale structural similarity for image quality assessment", The 37th Asilomar Conference on Signals, Systems Computers Volume 2, November 2003.

[Wang04] Wang, Z., Bovik, A., Sheikh, H., and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", issn 1057-7149, IEEE transactions on image processing Volume 13, number 4, April 2004.

[Yang12] Yang, Y., Ming, J., and N. Yu, "Color Image Quality Assessment Based on CIEDE2000", Advances in multimedia Article ID 273723, 2012.

Authors' Addresses

Nathan E. Egge
Mozilla
331 E Evelyn Ave
Mountain View 94041
USA

Email: negge@mozilla.com

Luc N. Trudeau
Mozilla
331 E Evelyn Ave
Mountain View 94041
USA

Email: luc@trud.ca

David M. Barr
Xiph.Org Foundation
21 College Hill Road
Somerville, MA 1124
USA

Email: b@rr-dav.id.au