Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 4, 2017
S. Midtskogen
Cisco
October 31, 2016

Improved chroma prediction
draft-midtskogen-netvc-chromapred-02

Abstract

This document describes the technique used to improve the chroma prediction in the Thor video codec.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Modern video coding standards such as Thor [I-D.fuldseth-netvc-thor] form predictions for the luma channel (Y) and chroma channels (U and V) which are encoded separately (in that order). The prediction for each channel has spatial or temporal dependencies only in its own channel. Most of the perceived information of a video is to be found in the luma channel, but there still remain correlations between the luma and chroma channels. For instance, the same shape of an object can often be seen in all three channels, and if this correlation is not exploited, some structural information will be transmitted three times. Thor will attempt to improve the chroma prediction by finding linear relationships between the each of the initial chroma predictions and the luma prediction, and if certain criteria are satisfied, use that relationship to form a new prediction based on the reconstructed luma samples.

2. Definitions

2.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

3. Background

The improved predictions are derived from the reconstructed luma samples using a mapping. The underlying assumption is that the colours can be identified by their luminosities. Informally we can say that a new chroma prediction is formed from the reconstructed luma block painted with the colours of the initial chroma prediction.

There is often a linear correlation between the luma and chroma channel, so that a chroma sample c can be expressed by the linear function

$$c = a*y + b$$

Figure 1: Linear relationship

where y is the corresponding luma sample. This observation has been previously been used in techniques to convert YUV 4:2:0 and YUV 4:2:2 images to YUV 4:4:4, and in a (rejected) proposal for HEVC as a special intra mode. Thor, however, generalises the prediction, so it does not depend on the coding mode (i.e. whether inter or intra, or the kind of inter/intra mode).

Since it would be too costly to transmit the values a and b in the linear mapping, and since both the encoder and decoder must be able to compute identical predictions, a and b are derived from data available to both using linear regression.

4. Computing the improved prediction

Since the assumption that the correlation is the same in the predicted block and in the reconstructed block is not always true, the new prediction from luma might not be better even when there is a very good correlation in the predicted block. Therefore, we can only expected an improvement if the initial prediction is bad, and the luma residual is used as an estimate for this. The initial chroma prediction is kept unless the average squared difference between the reconstructed luma samples yr and the predicted y samples for an N*N prediction block is above 64:

```
_N_ _N_
\   \
/__ /__ (yr(i, j) - y(i, j)) ^ 2
i=1 j=1
------------------------------- > 64
               N*N
```

Figure 2: Requirement for improvement 1

The encoder and decoder must compute a and b using the same least square fit for an N*N prediction block, where y and c denote the luma and chroma samples in the initial prediction:

```
              _N_ _N_                                  _N_ _N_
              \   \                                    \   \
       Ysum = /__ /__ y(i, j)                   Csum = /__ /__ c(i, j)
              i=1 j=1                                  i=1 j=1

              _N_ _N_                                  _N_ _N_
              \   \                                    \   \
      YYsum = /__ /__ y(i, j) ^ 2             CCsum = /__ /__ c(i, j) ^ 2
              i=1 j=1                                  i=1 j=1

              _N_ _N_
              \   \
      YCsum = /__ /__ y(i, j) * c(i, j)
              i=1 j=1
```

Figure 3: Equations for linear regression 1

These sums will all be contained within a 32 bit signed integer when the internal bitdepth is 8. Otherwise 64 bit integers must be used. Then the following must be computed using 64 bit arithmetic regardless of bitdepth:

```
SSyy = YYsum - ((Ysum * Ysum) >> 2*log2(N))
SScc = CCsum - ((Csum * Csum) >> 2*log2(N))
SSyc = YCsum - ((Ysum * Csum) >> 2*log2(N))
```

Figure 4: Equations for linear regression 2

Still using 64 bit arithmetic, if

```
SSyy > 0 /\ 2 * SSyy * SSyy > SSyy * SScc
```

Figure 5: Requirement for improvement 2

then it is assumed that the correlation is reasonably good and a new prediction will be computed and used. Otherwise, the initial prediction will be kept. First, a and b must be computed. 2^15 is added to b to ensure correct rounding later on.

```
a = (SSyc << 16) / SSyy
b = (((Csum << 16) - a * Ysum) >> 2*log2(N)) + (1 << 15)
```

Figure 6: Equation for linear regression 3

The final operations are performed with 32 bit arithmetic, so a must be clipped to [-2^(31-B), 2^(31-B)], where B is the bitdepth, and b must be clipped to [-2^31, 2^31-1]. The a new chroma prediction c' is computed using the reconstructed luma samples yr, a and b, and a clipping function saturating the results to an 8 bit value:

```
c'(i, j) = clip((a * yr(i, j) + b) >> 16)
```

Figure 7: Improved chroma prediction

The above assumes 4:4:4 format. For the 4:2:0 format the predicted luma block must be subsampled first:

```
y'(i,j) = (y(2*i, 2*j)   + y(2*i+i, 2j) +
           y(2*i, 2*j+1) + y(2*i+1, 2*j+1) + 2) >> 2
```

Figure 8: Subsampling of predicted luma block

The resulting new chroma prediction must also be subsampled. The clipping is performed before the subsampling.

```
c'(i, j) = (clip((a*yr(2*i, 2*j) + b) >> 16) +
            clip((a*yr(2*i+1, 2*j) + b) >> 16) +
            clip((a*yr(2*i, 2*j+1) + b) >> 16) +
            clip((a*yr(2*i+1, 2*j+1) + b) >> 16) + 2) >> 2
```

Figure 9: Subsampling of improved chroma prediction

In intra mode the chroma prediction improvement must be performed right after each transform, since the new chroma reconstruction will be used to predict the next block.

5. Performance

The improved chroma prediction may significantly improve the compression efficiency for images or video containing high correlations between the channels. It is particularly useful for encoding screen content, 4:4:4 content, high frequency content and "difficult" content where traditional prediction techniques perform poorly. Little quality change is seen for content not in these categories, but there is a general small increase in chroma PSNR.

An encoded configured for low delay and high complexity was used for the following results. The numbers have been computed using the Bjontegaard Delta Rate (BDR [BDR]). The rates for Y, U and V have been shown separately.

	4:4:4			4:2:0		
Sequence	Y	U	V	Y	U	V
cad_waveform	-21.3%	-27.0%	-24.0%	0.5%	-1.3%	-1.1%
pcb_layout	-9.2%	-13.3%	-10.6%	-1.6%	-3.1%	-3.5%
ppt_doc_xls	-6.3%	-14.1%	-12.7%	-0.1%	-0.8%	-0.8%
vc_doc_sharing	-2.9%	-6.4%	-6.9%	0.3%	-1.2%	-0.6%
web_browsing	-0.5%	-1.1%	-1.5%	0.3%	-0.5%	-1.0%
wordEditing	-1.8%	-5.9%	-4.8%	1.5%	1.2%	1.1%
park_joy	-0.5%	-2.6%	-0.9%	-0.0%	-0.8%	0.4%
old_town_cross	-0.1%	-2.2%	-1.2%	0.0%	-0.6%	-0.2%
Average	-5.3%	-9.1%	-7.8%	0.1%	-0.9%	-0.7%

Figure 10: Compression Performance, improved prediction for intra blocks only

	4:4:4			4:2:0		
Sequence	Y	U	V	Y	U	V
cad_waveform	-23.1%	-28.9%	-26.1%	-2.6%	-3.6%	-3.5%
pcb_layout	-21.0%	-29.0%	-21.0%	-5.4%	-7.9%	-5.4%
ppt_doc_xls	-9.0%	-19.0%	-17.5%	-0.2%	-0.2%	-1.2%
vc_doc_sharing	-4.7%	-9.6%	-9.6%	-0.1%	-1.0%	-0.4%
web_browsing	-0.6%	-1.5%	-1.5%	-0.5%	-1.2%	-1.2%
wordEditing	-11.3%	-13.7%	-11.7%	-3.0%	-4.2%	-3.2%
park_joy	-5.5%	-7.4%	-7.1%	-0.9%	-1.9%	-1.6%
old_town_cross	-1.7%	-3.6%	-2.2%	-0.3%	-4.1%	-1.6%
Average	-9.6%	-14.1%	-12.1%	-1.6%	-3.0%	-2.3%

Figure 11: Compression Performance, improved prediction using intra only coding

	4:4:4			4:2:0		
Sequence	Y	U	V	Y	U	V
cad_waveform	-11.5%	-14.4%	-12.7%	0.0%	-1.8%	-1.7%
pcb_layout	-3.2%	-5.5%	-4.8%	-0.9%	-2.4%	-3.4%
ppt_doc_xls	-0.1%	-0.7%	-0.3%	0.0%	-0.2%	-0.6%
vc_doc_sharing	-0.4%	-0.6%	-1.6%	-0.0%	-0.4%	-0.6%
web_browsing	0.1%	0.2%	0.1%	0.5%	-0.0%	-0.9%
wordEditing	-3.7%	-5.8%	-6.2%	0.4%	-0.9%	-1.4%
park_joy	-1.6%	-8.6%	-1.5%	0.0%	-3.5%	-0.2%
old_town_cross	-0.0%	-0.4%	-0.1%	0.0%	0.1%	-0.2%
Average	-2.5%	-4.5%	-3.4%	0.0%	-1.1%	-1.1%

Figure 12: Compression Performance, improved prediction for inter blocks only

	4:4:4			4:2:0		
Sequence	Y	U	V	Y	U	V
cad_waveform	-25.8%	-31.7%	-28.2%	-2.4%	-5.5%	-5.4%
pcb_layout	-11.5%	-16.1%	-13.5%	-2.4%	-4.1%	-5.6%
ppt_doc_xls	-6.3%	-14.3%	-13.2%	-0.2%	-0.8%	-0.8%
vc_doc_sharing	-3.0%	-6.7%	-8.2%	0.1%	-0.9%	-1.1%
web_browsing	-0.5%	-1.2%	-1.5%	0.2%	-0.3%	-2.0%
wordEditing	-3.4%	-6.8%	-6.6%	0.6%	-0.5%	-1.4%
park_joy	-1.7%	-9.2%	-1.7%	-0.0%	-4.0%	0.0%
old_town_cross	-0.1%	-2.2%	-1.0%	0.1%	-0.5%	-0.1%
Average	-6.5%	-11.0%	-9.2%	-0.5%	-2.1%	-2.0%

Figure 13: Compression Performance, improved prediction for intra and inter blocks

6. IANA Considerations

This document has no IANA considerations yet. TBD

7. Security Considerations

This document has no security considerations yet. TBD

8. Acknowledgments

The author would like to thank Arild Fuldseth and Mo Zanaty for reviewing this document, and Timothy Terriberry for pointing a couple of errors in the first draft.

9. References

9.1. Normative References

[I-D.fuldseth-netvc-thor]
Fuldseth, A., Bjontegaard, G., Midtskogen, S., Davies, T., and M. Zanaty, "Thor Video Codec", draft-fuldseth-netvc-thor-02 (work in progress), March 2016.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

9.2. Informative References

[BDR]   Bjontegaard, G., "Calculation of average PSNR differences between RD-curves", ITU-T SG16 Q6 VCEG-M33 , April 2001.

Author's Address

Steinar Midtskogen
Cisco
Lysaker
Norway

Email: stemidts@cisco.com