https://doi.org/10.2352/ISSN.2470-1173.2017.2.VIPC-409

Adaptive Multi-Reference Prediction Using A Symmetric Framework

Zoe Liu*, Debargha Mukherjee*, Wei-Ting Lin+, Paul Wilkins*, Jingning Han*, Yaowu Xu*;

*Google, Inc., 1600 Amphitheatre Parkway, Mountain View, CA, USA 94043.

+University of California, Santa Barbara, CA 93106.

ABSTRACT

Google started the WebM Project in 2010 to develop open source, royalty--free video codecs designed specifically for media on the Web. Subsequently, Google jointly founded a consortium of major tech companies called the Alliance for Open Media (AOM) to develop a new codec AV1, aiming at a next edition codec that achieves at least a generational improvement in coding efficiency over VP9. This paper proposes a new coding tool as one of the many efforts devoted to AOM/AV1. In particular, we propose a second `ALTREF_FRAME` in the AV1 syntax, which brings the total reference frames to seven on top of the work presented in [11]. `ALTREF_FRAME` is a constructed, no-show reference obtained through temporal filtering of a look-ahead frame. The use of two `ALTREF_FRAME`s adds further flexibility to the multi-layer, multi-reference symmetric framework, and provides a great potential for the overall Rate-Distortion (RD) performance enhancement. The experimental results have been collected over several video test sets of various resolutions and characteristics both texture- and motion-wise, which demonstrate that the proposed approach achieves a consistent coding gain, compared against the AV1 baseline as well as against the results in [11]. For instance, using overall-PSNR as the distortion metric, an average bitrate saving of 5.880% in BDRate is obtained for the CIF-level resolution set, and 4.595% on average for the VGA-level resolution set.

Keywords: video coding, VP9, VP10, WebM, AV1, AOM, H.264, HEVC, multi-reference prediction, adaptive prediciton, `ALTREF_FAME`.

1. INTRODUCTION

Google started the WebM Project [1] in 2010 to develop open source, royalty--free video codecs designed specifically for media on the Web. The first codec released as part of the project was called VP8 [2] and is still used extensively in Google Hangouts. The second generation codec released by the WebM project, VP9 [3][4], -is currently served by YouTube, and enjoys billions of views per day. It achieves a coding efficiency similar to the latest video codec from MPEG entitled HEVC [5]. Realizing the need for even greater compression efficiency to cope with the growing demand for video on the web, the WebM team embarked on an ambitious project to develop a next

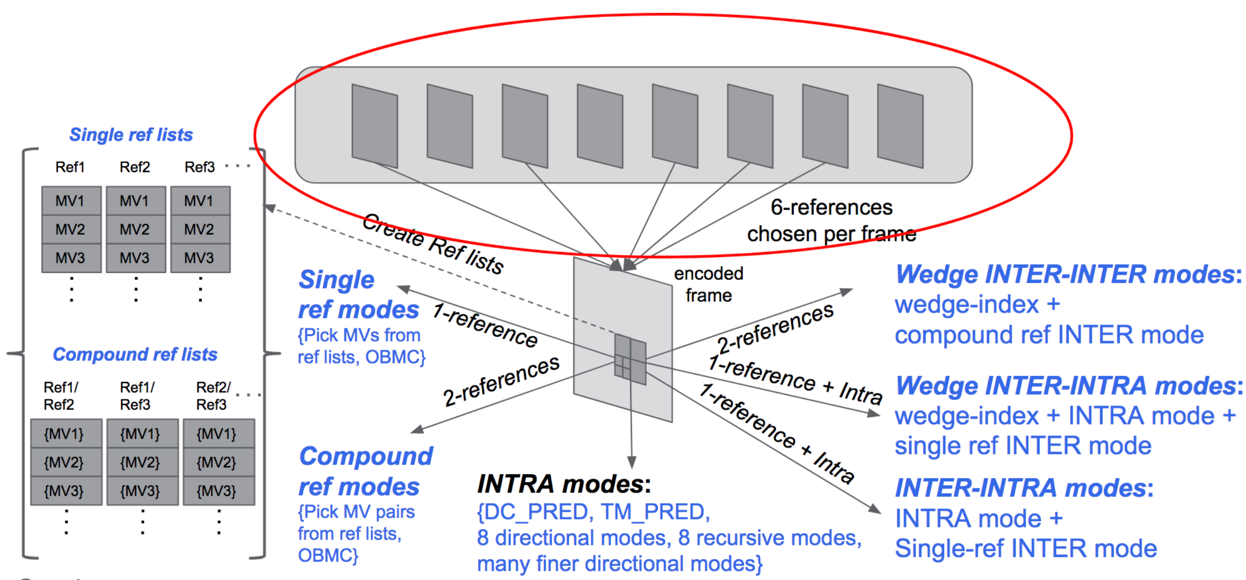

Figure 1. New coding tools explored by AOM/AV1

edition codec, VP10 [7], that achieves at least a generational improvement in coding efficiency over VP9. Starting from VP9, a set of new experimental coding tools have already been added to VP10 to achieve substantial coding gains. Subsequently, Google joined a consortium of major tech companies called the Alliance for Open Media (AOM) to jointly develop a new codec AV1 [8]. As a result, the VP10 effort has been largely merged with AV1.

Major tools that are being explored by AOM/AV1 are illustrated in Figure 1.

In this paper, we focus primarily on the use of the multiple reference prediction for the overall Rate-Distortion (RD) performance enhancement. Specifically, we describe tools that increase the flexibility and adaptability in selecting various combinations of reference frames, allowing the codec to handle a more diverse range of videos in terms of temporal correlations across successive frames.

2. SYMMETRIC FRAMEWORK USING MULTI-REFERENCE PREDICTION

Current VP9 codec uses three references for the encoding of each video frame, namely `LAST_FRAME`, `GOLDEN_FRAME`, and `ALTREF_FRAME`. In the work we presented in [11], a new coding tool is proposed and the number of references is extended from three to six, through the adding of three new references: `LAST2_FRAME`, `LAST3_FRAME`, and `BWDREF_FRAME`. In particular, `LAST2_FRAME` and `LAST3_FRAME` are two forward references, similar to `LAST_FRAME`, whereas `BWDREF_FRAME` is a backward reference, similar to `ALTREF_FRAME`. The use of `BWDREF_FRAME` exploits the existing `"show_existing_frame"` feature provided by VP9, to encode a look-ahead frame without applying temporal filtering, thus more applicable as a backward reference in a relatively shorter distance.

Specifically, in the single reference mode, where Inter-coded blocks use a single prediction obtained from one reference frame, the proposed coding tool allows each frame to choose from up to six reference frames. In the compound reference mode, where Inter-coded blocks use a combination of two predictions, obtained from a forward reference and a backward reference respectively, four choices are provided for the forward references and two for the backward references. Each video frame consequently is offered an extensively larger set of multi-reference prediction modes, thus leading to a greater potential for the RD performance improvement.

Secondly, through the use of `BWDREF_FRAME`, a symmetric framework of multi-reference prediction is established for the compound mode prediction: (1) A `BWDREF_FRAME` may be selected from a nearer future frame, as opposed to the `LAST_FRAME`; (2) A `BWDREF_FRAME` or `ALTREF_FRAME` may be selected from a farther future frame, as opposed to the `LAST2_FRAME`; and (3) An `ALTREF_FRAME` may be selected from a distant future frame, as opposed to the `GOLDEN_FRAME` in the distant past. Such framework provides an opportunity to encode a variety of videos with dynamic temporal correlation characteristics in a more adaptive and optimal way. An instantiation of the framework is illustrated in Figure 2.

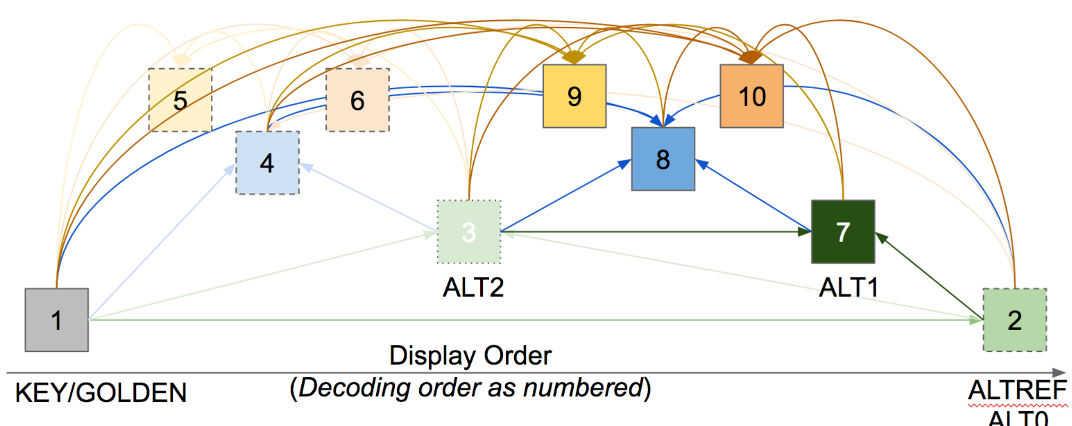

Figure 2(a). An example of video coding using a symmetric multi-reference prediction framework: Symmetric multi-reference prediction of a video clip in display order

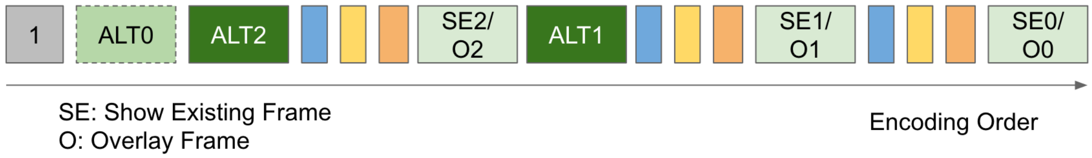

Figure 2(b). An example of video coding using a symmetric multi-reference prediction framework: Symmetric multi-reference prediction of a video clip in encoding order

3. ADAPTIVE PREDICTION USING SECOND ALTREF_FRAME

ALTREF_FRAME is a no-show frame usually constructed from a distant future frame through temporal filtering. An AV1 encoder may apply different temporal filtering strength to construct an ALTREF_FRAME, adapting to various motion smoothness levels across frames. A so-called Golden Frame (GF) group can be established, and all the frames within one GF group may share the same GOLDEN_FRAME and the same ALTREF_FRAME. LAST_FRAME may be updated constantly. When the distant future frame that provides ALTREF_FRAME is actually being coded, it is referred to as an OVERLAY frame but treated as a regular inter frame. OVERLAY frames usually cost fairly small amounts of bits as ALTREF_FRAME may serve as an ideal prediction.

We propose the use of a second ALTREF_FRAME in the new coding tool, namely ALTREF_FRAME2. The use of two ALTREF_FRAMEs allows the total number of reference frames further increase from six to seven. Moreover, the number of backward reference frames are more balanced with that of the forward reference frames, and a more ideal symmetric reference framework may hence be established. The introduction of ALTREF_FRAME2 has a great potential to better adapt to the various motion- / texture-wise characteristics in a video and provide more RD-optimized inter-predictors for the coding of each video frame.

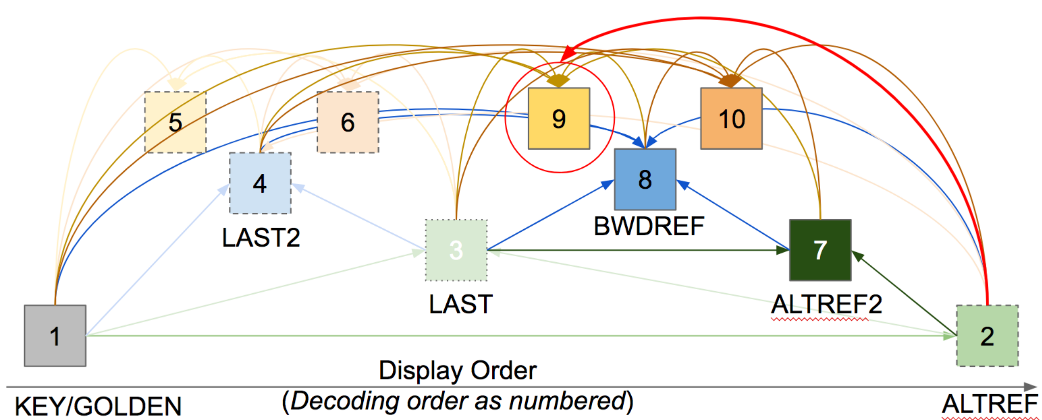

Figure 3. The use of a second ALTREF_FRAME (ALTREF2) in the hierarchical multi-layer framework

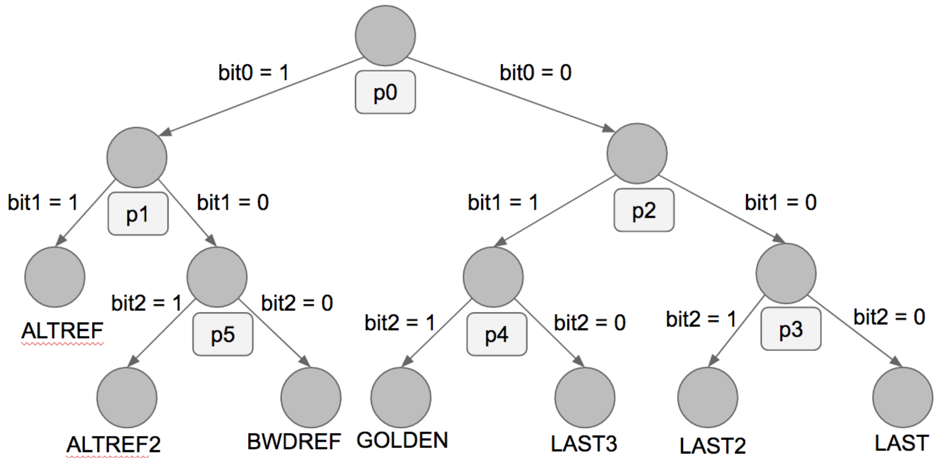

Figure 4(a). Binary tree structure design for context-based, bit-level entropy coding of the extended reference frames: Single reference prediction

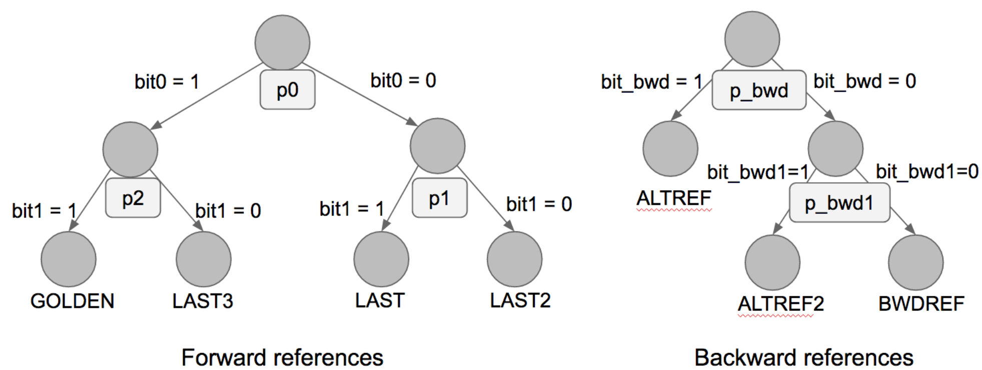

Figure 4(b). Binary tree structure design for context-based, bit-level entropy coding of the extended reference frames: Compound reference prediction

4. EXPERIMENTAL RESULTS

Results are collected over two standard video test sets with various resolutions and spatial / temporal characteristics, as presented in Table 1. Specifically, the set of *lowres* includes 40 videos of CIF resolution, and the set of *midres* includes 30 videos of 480p and 360p resolution. Each video contains 150 frames.

Table 1. BDRate reduction using the metric of overall-PSNR by `ALTREF_FRAME2`

Set of *lowres*		Set of *midres*	
Against AV1 baseline	Against [11]	Against AV1 baseline	Against [11]
-5.880%	-0.277%	-4.585%	-0.158%

Table 2. Bitrate reduction using `ALTREF_FRAME2` (*lowres* set)

	BWDREF + ALTREF (%)			**BWDREF + ALTREF + ALTREF2 (%)**		
File	**avg_psnr**	**ovr_psnr**	**ssim**	**avg_psnr**	**ovr_psnr**	**ssim**
akiyo_cif.y4m	-5.483	-6.286	-1.512	-5.639	-6.426	-2.103
basketballpass_240p.y4m	-5.172	-5.208	-2.589	-5.478	-5.502	-3.048
blowingbubbles_240p.y4m	-6.940	-6.662	-5.797	-7.028	-6.755	-5.836
bowing_cif.y4m	-2.730	-3.734	-3.313	-3.160	-4.187	-3.688
bqsquare_240p.y4m	-10.187	-10.376	-6.651	-10.203	-10.413	-6.670
bridge_close_cif.y4m	-6.730	-5.583	-7.844	-7.422	-6.311	-8.716
bridge_far_cif.y4m	-5.495	-5.894	-7.623	-6.092	-6.511	-8.138
bus_cif.y4m	-4.988	-4.807	-3.456	-5.147	-4.961	-3.738
cheer_sif.y4m	-3.158	-3.177	-0.302	-3.881	-3.917	-1.088
city_cif.y4m	-4.988	-5.126	-2.027	-4.963	-5.116	-1.970
coastguard_cif.y4m	-9.831	-9.767	-11.453	-9.970	-9.902	-11.582
container_cif.y4m	-11.715	-12.781	-13.278	-11.886	-12.974	-13.648
crew_cif.y4m	-3.481	-3.634	-3.144	-3.596	-3.751	-3.175
deadline_cif.y4m	-3.658	-4.275	-1.361	-3.963	-4.612	-1.375
flower_cif.y4m	-12.568	-13.093	-6.024	-12.640	-13.188	-6.104
flowervase_240p.y4m	-8.978	-9.158	-8.202	-8.906	-9.076	-8.026
football_cif.y4m	-1.312	-0.128	-0.970	-1.342	-0.179	-1.038
foreman_cif.y4m	-4.245	-4.347	-0.736	-4.759	-4.863	-1.095
garden_sif.y4m	-8.368	-8.729	-5.558	-8.572	-8.931	-5.684
hallmonitor_cif.y4m	-1.679	-0.288	-0.148	-2.590	-1.212	-1.397

harbour_cif.y4m	-7.805	-7.899	-9.595	-7.868	-7.961	-9.664
highway_cif.y4m	-3.813	-2.432	-2.652	-4.567	-3.234	-3.476
husky_cif.y4m	-3.938	-4.175	-3.397	-4.272	-4.503	-3.848
ice_cif.y4m	-3.900	-4.894	-3.380	-3.839	-4.814	-3.296
keiba_240p.y4m	-1.371	-0.761	-0.146	-1.574	-0.967	-0.495
mobile_cif.y4m	-12.176	-12.207	-8.717	-12.633	-12.713	-8.729
mobisode2_240p.y4m	-10.475	-11.881	-11.053	-10.365	-11.805	-10.932
motherdaughter_cif.y4m	-4.625	-4.796	-3.069	-5.362	-5.548	-3.775
news_cif.y4m	-2.798	-3.086	1.298	-2.914	-3.195	0.801
pamphlet_cif.y4m	-0.793	-1.319	-0.073	-0.513	-0.990	0.323
paris_cif.y4m	-2.938	-3.434	1.508	-3.105	-3.593	1.292
racehorses_240p.y4m	-1.414	-1.496	0.076	-1.694	-1.766	-0.124
signirene_cif.y4m	-5.128	-5.537	-3.291	-5.330	-5.746	-3.517
silent_cif.y4m	-3.322	-3.466	-2.018	-3.688	-3.840	-2.418
soccer_cif.y4m	-1.081	-0.945	1.995	-1.590	-1.426	1.101
stefan_sif.y4m	-7.504	-7.198	-5.411	-7.853	-7.569	-5.623
students_cif.y4m	-6.025	-6.691	-4.647	-6.230	-6.921	-5.021
tempete_cif.y4m	-9.561	-9.410	-8.374	-9.705	-9.564	-8.458
tennis_sif.y4m	-2.367	-2.855	-1.220	-2.722	-3.181	-1.611
waterfall_cif.y4m	-7.222	-7.137	-3.542	-7.201	-7.091	-3.428
{OVERALL}	**-5.499**	**-5.617**	**-3.942**	**-5.757**	**-5.880**	**-4.225**

Table 3. Bitrate reduction using `ALTREF_FRAME2` (*midres* set)

File	`BWDREF + ALTREF (%)`			`BWDREF + ALTREF + ALTREF2 (%)`		
	avg_psnr	**ovr_psnr**	**ssim**	**avg_psnr**	**ovr_psnr**	**ssim**
BQMall_832x480_60.y4m	-5.800	-6.275	-4.836	-6.111	-6.609	-5.055
BasketballDrillText_832x480_50.y4m	-3.492	-3.971	-0.310	-3.781	-4.261	-0.553
BasketballDrill_832x480_50.y4m	-2.921	-3.307	0.254	-3.167	-3.544	-0.008
Flowervase_832x480_30.y4m	-4.591	-3.969	-3.336	-4.652	-4.058	-3.642
Keiba_832x480_30.y4m	-2.056	-1.440	-1.669	-2.547	-1.924	-2.463

Mobisode2_832x480_30.y4m	-3.031	-2.616	-1.126	-2.604	-2.132	-0.547
PartyScene_832x480_50.y4m	-5.156	-5.647	-3.363	-5.131	-5.629	-3.313
RaceHorses_832x480_30.y4m	-0.513	-0.415	-0.255	-0.717	-0.652	-0.499
aspen_480p.y4m	-1.126	-2.398	-0.160	-1.766	-2.946	-0.585
city_4cif_30fps.y4m	-4.513	-4.675	-1.709	-4.579	-4.744	-1.622
controlled_burn_480p.y4m	-0.238	-1.085	1.968	-0.062	-0.978	2.529
crew_4cif_30fps.y4m	-1.920	-2.718	-0.731	-2.060	-2.863	-0.925
crowd_run_480p.y4m	-10.183	-11.134	-7.026	-10.185	-11.128	-6.916
ducks_take_off_480p.y4m	-17.949	-18.939	-17.798	-17.914	-18.915	-17.727
harbour_4cif_30fps.y4m	-7.822	-8.328	-8.240	-7.913	-8.386	-8.298
ice_4cif_30fps.y4m	-2.410	-3.196	-0.014	-2.520	-3.318	-0.042
into_tree_480p.y4m	-2.137	-2.207	-1.656	-1.701	-1.784	-1.133
old_town_cross_480p.y4m	-4.227	-4.337	-2.139	-3.909	-4.032	-1.685
park_joy_480p.y4m	-7.665	-9.293	-5.023	-8.040	-9.675	-6.271
red_kayak_480p.y4m	0.394	2.592	0.906	0.371	2.548	0.828
rush_field_cuts_480p.y4m	-8.649	-9.423	-6.059	-9.001	-9.783	-6.665
sintel_trailer_2k_480p24.y4m	-15.340	-5.376	-3.921	-15.218	-5.165	-3.920
snow_mnt_480p.y4m	-0.678	-0.408	0.895	-0.792	-0.529	1.074
soccer_4cif_30fps.y4m	-0.766	-1.440	1.765	-1.419	-2.083	0.794
speed_bag_480p.y4m	-6.108	-7.737	-5.560	-7.137	-8.616	-6.624
station2_480p25.y4m	-1.796	-1.792	-1.123	-1.823	-1.814	-1.046
tears_of_steel1_480p.y4m	-3.775	-4.141	-0.996	-4.053	-4.422	-1.276
tears_of_steel2_480p.y4m	-5.400	-6.614	-2.024	-5.465	-6.703	-2.139
touchdown_pass_480p.y4m	-1.084	-1.942	3.368	-1.625	-2.527	2.496
west_wind_easy_480p.y4m	-1.236	-1.062	-2.632	-1.376	-1.181	-2.565
{OVERALL}	**-4.406**	**-4.443**	**-2.418**	**-4.563**	**-4.595**	**-2.593**

5. EXPERIMENTAL RESULTS

In this paper, we propose a multi-layer, multi-reference symmetric framework for AOM/AV1, an effort for the next generational royal free, open source video codec. In particular, through the use of two `ALTREF_FRAMEs`, together with five other reference frames, the framework provides a great potential for the overall Rate-Distortion (RD) performance enhancement. The experimental results have been collected over several video test sets of various resolutions and characteristics both texture- and motion-wise, which demonstrate a consistent coding gain by the proposed approach, compared against the AV1 baseline as well as against the results in [11]. For instance, using overall-PSNR as the distortion metric, an average bitrate saving of 5.880% in BDRate is obtained for the CIF-level resolution set, and 4.595% on average for the VGA-level resolution set.

REFERENCES

[1] http://www.webmproject.org/

[2] J. Bankoski, J. Koleszar, L. Quillio, J. Salonen, P. Wilkins, Y. Xu, *VP8 Data Format and Decoding Guide*, RFC 6386, http://datatracker.ietf.org/doc/rfc6386/

[3] D. Mukherjee, J. Bankoski, R. S. Bultje, A. Grange, J. Han, J. Koleszar, P. Wilkins, Y. Xu, “The latest open-source video codec VP9 - an overview and preliminary results,” *Proc. IEEE Picture Coding Symp.*, pp. 390-93, San Jose, Dec. 2013.

[4] D. Mukherjee, J. Bankoski, R. S. Bultje, A. Grange, J. Han, J. Koleszar, P. Wilkins, Y. Xu, “A Technical Overview of VP9 - the latest open-source video codec,” *SMPTE Motion Imaging Journal*, Jan/Feb 2015.

[5] Gary J. Sullivan, JensRainer Ohm, WooJin Han, and Thomas Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard,” *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 22, No. 12, Dec 2012.

[6] Thomas Wiegand, Gary J. Sullivan, Gisle Bjøntegaard; Ajay Luthra. "Overview of the H.264/AVC Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13 No. 7, Jan 2011.

[7] D. Mukherjee, H. Su, J. Bankoski, A. Converse, J. Han, Z. Liu, Y. Xu, “An overview of video coding tools under consideration for VP10: the successor to VP9,” *Proc. SPIE, Applications of Digital Image Processing XXXVIII*, vol. 9599, Sep 2015.

[8] http://www.aomedia.org/

[9] M. Paul, W. Lin, C.-T. Lau, and B.S. Lee, “A long-term reference frame for hierarchical B-picture-based video coding,” *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 24, No. 10, Oct. 2014.

[10] G. Bjøntegaard, “Calculation of average psnr differences between rdcurves,” *VCEGM33*, 13th VCEG meeting, Austin, Texas, March 2001.

[11] Wei-Ting Lin, Zoe Liu, Debargha Mukherjee, Jingning Han, Paul Wilkins, Yaowu Xu, Jim Bankoski, and Ken Rose, “Efficient AV1 video coding using a multi-layer framework,” submitted to *IEEE International Conference in Image Processing*, Beijing, China, September, 2017.