A CONSTRAINED ADAPTIVE SCAN ORDER APPROACH TO TRANSFORM COEFFICIENT ENTROPY CODING

Ching-Han Chiang, Jingning Han, and Yaowu Xu

WebM Codec Team, Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043
Emails: {angiebird,jingning,yaowu}@google.com

ABSTRACT

Transform coefficient coding is a key module in modern video compression systems. Typically, a block of the quantized coefficients are processed in a pre-defined zig-zag order, starting from DC and sweeping through low frequency positions to high frequency ones. Correlation between magnitudes of adjacent coefficients is exploited via context based probability models to improve compression efficiency. Such scheme is premised on the assumption that spatial transforms compact energy towards lower frequency coefficients, and the scan pattern that follows a descending order of the likelihood of coefficients being non-zero provides more accurate probability modeling. However, a pre-defined zig-zag pattern that is agnostic to signal statistics may not be optimal. This work proposes an adaptive approach to generate scan pattern dynamically. Unlike prior attempts that directly sort a 2-D array of coefficient positions according to the appearance frequency of non-zero levels only, the proposed scheme employs a topological sort that also fully accounts for the spatial constraints due to the context dependency in entropy coding. A streamlined framework is designed for processing both intra and inter prediction residuals. This generic approach is experimentally shown to provide consistent coding performance gains across a wide range of test settings.

Index Terms— Adaptive coefficient scan, transform coding, topological sort

1. INTRODUCTION

Many modern video codecs employ a block-based coding scheme. In such a scheme, a frame is partitioned into a grid of blocks. Each individual block is coded using prediction generated from pixels in previously coded frames or previously coded pixels in the same frame. The prediction block is subtracted from the original pixel block, and the resulting residual block then undergoes spatial transform to further remove the remaining inter-pixel correlation, followed by coefficient quantization and entropy coding.

The sinusoidal transforms employed in the video/image codecs are low-complexity approximations to the Karhunen-Loeve transform (KLT). Hence, certain correlations exist between coefficients at different frequency positions. To code a 2-D array of quantized coefficients, ideally, one would use a joint probability of the coefficient vector for arithmetic coding, which requires an excessively giant codebook to maintain. In practice, the entropy coding engine decomposes the 2-D coefficient array into a sequence of individual coefficients prior to coding them. It exploits the fact that the magnitudes of adjacent coefficients are correlated by referencing the available spatial neighbor coefficients as context in probability models. The signs of transform coefficients are usually uncorrelated and are directly coded as binary symbols. The coding engine starts from lower frequency coefficients, where the quantized magnitudes are more likely to be non-zero, and proceeds through the 2D-array towards higher frequency coefficients. The process terminates at the last non-zero coefficient and discards the rest of the zeros, thereby minimizing the entropy cost on those tailing zero coefficients due to probability modeling noise.

Conventional designs use a pre-defined zig-zag scan order. Its deficiency in the context of intra prediction modes has been extensively studied in both discrete cosine transform [1]-[3] and wavelet transform[4] settings. A mode dependent scan order that skews towards vertical or horizontal directions for directional intra prediction modes is proposed in [1]. A joint optimization of the prediction, transform kernel, and coefficient scan order for entropy coding is introduced in [5]. Using localized adaptive scan order at block level has been proposed in [2]-[4] for image compression systems. In [6], the scan order for inter prediction residuals has been discussed. It draws the statistics from the raw (pre-quantized) transform coefficients and sorts the positions according to their likelihood of being non-zero values to produce a candidate scan order per frame. The resulting scan order is sent explicitly in the frame header to decoder. In addition, three other pre-defined scan patterns and the resulting patterns from previous frames will be evaluated to form a set of scan order candidates for encoder to select at coding block level. Recent work for higher dimension transform partitions the 2D-array into a set of 4x4 sub-blocks as coefficient groups (CGs). The coding engine processes each CG sequentially. All CGs share a single scan order selected from several pre-defined patterns [7].

Alternatively, a constrained adaptive scan order approach is proposed in this work to capture the statistics of prediction residual signal, while fully accounting for context dependency. Because the spatial transforms applied to the residual signal are approximations to the KLT, correlations between coefficients can not be completely eliminated. Hence, a coefficient may still need to refer to its neighbor coefficients as context information to improve probability modeling accuracy. For instance, the coding engine in VP8 [8] and its successor VP9 [9] use coefficients at the immediate above and left neighboring positions in the 2-D array as the probability model context for the current coefficient coding. The coefficient coding system used in HEVC [10, 11] employs a backward coding approach where the context depends on the right and bottom neighbors. Such context dependency implies certain constraints on the spatial scan order in the 2-D array.

This work devises an approach to obtain scan orders by combining frame level coefficient statistics and context dependencies. For each video frame, it first generates a draft scan order by ranking all coefficient positions in descending order of non-zero probabilities. It then translates the context dependency into a directed acyclic graph and integrates the graph into a topological sort to optimize the initial draft order into the final scan order for the frame. Clearly this approach is generally applicable to both intra and inter prediction residuals, and to various transform sizes and kernels. Experiments have shown that the proposed scheme provides consistent coding performance gains (on average 1%) across a wide range of test settings for both above mentioned mainstream coding engines.

2. MOTIVATION

To demonstrate the variations in signal statistics, we collect the quantized coefficients of 8x8 2D-DCT from different video clips. The non-zero probability at each frequency position is visualized in Fig.1. Clearly, the distributions are distinctive across various video clips, which substantiates the need for an adaptive scan order scheme. As we can see from the figure, the non-zero probability is not strictly monotonically decreasing along the vertical/horizontal axises. Hence, a direct ranking of the non-zero probability may lead to a non-continuous scan, which could potentially degrade the entropy coding efficiency. Such observations motivate our design of a constrained adaptive scan order.

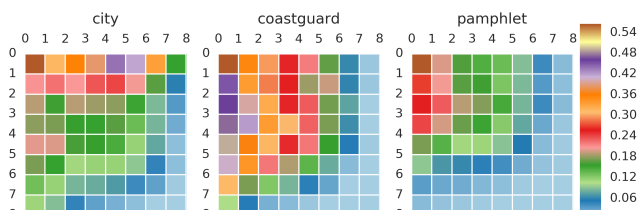

Fig. 1. Non-zero probability distribution across an 8X8 2D-DCT transform block.

3. CONSTRAINED ADAPTIVE SCAN ORDER

Two types of constraints will be employed to produce scan orders, namely, (a) the pre-defined context dependencies and (b) the estimated non-zero probabilities. In this section, we will first discusss these two constraints, then, we will describe our proposed scheme that uses a topological sort to incorporate these two factors together to generating adaptive scan orders. The discussion assumes the use of VP9 coefficient coding engine, but can be readily applied to other coding engines.

3.1. Context Dependency

The coefficient context dependencies are applied in entropy coding to exploit the remaining inter-coefficient correlations. To encode each coefficient, one would refer to its above and left neighbors, if they are available as context information, to retrieve the corresponding probability model. Such context dependencies can be represented by a directed acyclic graph (DAG) in a transform block as depicted in Fig. 2 (a).

The context dependency can be considered as a pre-defined hard constraint. Evidently, there are multiple solutions that satisfy such constraint. For example, two possible solutions that satisfy the above-left context dependency DAG for a 3×3 transform block are shown in Fig. 2 (b) and (c). This provides some flexibility to further optimize scan orders that would better capture the signal statistics.

3.2. Non-Zero Probability Estimation

Provided the non-zero probabilities of the frequency positions in a given frame, the coding engine would process individual coefficients in a descending order of the likelihood of non-zero value, to maximize the number of tailing zeros. The coding process terminates after processing the last non-zero coefficient and discards the rest tailing zero coefficients, thereby maximally reducing the entropy cost due to the probability modeling noise on less likely processed frequency positions.

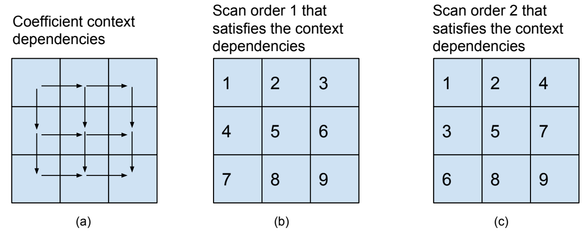

Fig. 2. (a) An example of the context dependency DAG of a 3×3 transform block. (b) and (c) Two example scan orders that comply with the context dependency DAG in (a).

However, explicitly sending such information incurs significant overhead in advanced video codecs, which typically involves multiple transform block sizes and kernels. We overcome this issue by using a moving window estimation scheme, where both encoder and decoder estimate current frame's probabilities using statistics from previously coded frames.

The non-zero probability estimation process is discussed here. Considering previously coded frame $(i-1)$, let its estimated non-zero probabilities be $P_e[i-1][r][c]$ and its observed non-zero probabilities be $P_c[i-1][r][c]$, where r and c denote the frequency indices in the vertical and horizontal directions respectively. The probabilities of frame i, $P_e[i][r][c]$, are estimated by

$$P_e[i][r][c] = (1-k) * P_e[i-1][r][c] + k * P_c[i-1][r][c], \quad (1)$$

where k is the update rate.

The observed non-zero probabilities of frame $(i-1)$, $P_c[i-1][r][c]$ in Eq. 1, can be obtained as follows. When coding frame $(i-1)$, we count the non-zero coefficient appearance $C[i-1][r][c]$ at frequency position (r, c) of the given transform size and type, as well as the transform block number M. The observed non-zero probabilities of the transform blocks in frame $(i-1)$ can be obtained by

$$P_c[i-1][r][c] = C[i-1][r][c]/M. \quad (2)$$

The estimated non-zero probabilities $P_e[i][r][c]$ can be considered as soft but dense constraints, since a scan order can be directly derived by sorting the probabilities $P_e[i][r][c]$. However, the scan order needs to be modified when it has conflicts with the context dependency DAG described in Sec. 3.1.

3.3. Topological Sort

Having built the context dependency DAG and the estimated non-zero probabilities $P_e[i][r][c]$, we can now design our adaptive scan order. Topological sorting based on depth-first search is a perfect algorithm for combining the context dependencies and the estimated non-zero probabilities to generate a desired adaptive scan order. We first sort the estimated non-zero probabilities $P_e[i][r][c]$ to produce a draft scan order. A topological sort is then applied to resolve the conflicts with the context dependency DAG.

The adaptive scan order design is implemented in two modules, TopologicalSort in Alg. 1 and ContextConflictSolver in Alg. 2. In both modules, each coefficient will have a coefficient index c_{idx} defined by $c_{idx} = r * len + c$ and a scan index s_{idx} corresponding to the adaptive scan order. In module TopologicalSort, each coefficient will also have a draft scan index ds_{idx} corresponding to the draft scan order.

In Alg. 1, procedure TopologicalSort has two inputs, draft scan order and ctx_dep, and one output $scan\_order$. The first input, draft scan order, maps from a draft scan index ds_{idx} to a coefficient index c_{idx} obtained by sorting the non-zero probabilities $P_e[i][r][c]$. The second input, ctx_dep, maps a coefficient index c_{idx} to its reference coefficient set. The output, $scan\_order$, which maps from a scan index s_{idx} to a coefficient index c_{idx}, is our desired adaptive scan order that complies with the provided context dependencies and largely follow the descending order of non-zero probabilities $P_e[i][r][c]$. The procedure TopologicalSort goes through each frequency position by draft scan order and applies ContextConflictSolver therein.

The procedure ContextConflictSolver in Alg. 2 recursively checks whether the context frequency coefficients that the current position depends on have been put into $scan\_order$. If not, it will then prioritize these parent indexes into $scan\_order$ table, followed by the current frequency position. Therefore, it ensures that all the preceding context information has been properly processed prior to the current frequency coefficient.

Algorithm 1 Modify Scan Order by Topological Sorting

Parameter:
len ▷ side length of the transform block
N ▷ len×len
c_{idx} ▷ coefficient index defined by $c_{idx} = r*\text{len}+c$
ds_{idx} ▷ draft scan order index
s_{idx} ▷ scan order index
visit[c_{idx}] ▷ table of coefficient scanned indicators
Input:
draft_scan_order[ds_{idx}] ▷ obtained by sorting $P_e[i][r][c]$.
ctx_dep[c_{idx}] ▷ coefficient context dependencies
Output:
scan_order[s_{idx}] ▷ adaptive scan order
Procedure: TopologicalSort
for c_{idx} = 0 to N-1 **do**
  visit[c_{idx}] = False
end for
s_{idx} = 0
for ds_{idx} = 0 to N-1 **do**
  c_{idx} = draft_scan_order[ds_{idx}]
  ContextConflictSolver(c_{idx}, ctx_dep, scan_order, visit, s_{idx})
end for

Algorithm 2 Recursive Context Conflict Solver

Input:
s_{idx} ▷ to-be-assigned scan order index
c_{idx} ▷ coefficient index
ctx_dep[c_{idx}] ▷ coefficient context dependencies
visit[c_{idx}] ▷ table of coefficient scanned indicators
Output:
s_{idx} ▷ increment it by one after it is assigned
scan_order[s_{idx}] ▷ adaptive scan order
Procedure: ContextConflictSolver
for each ctx_c_{idx} in ctx_dep[c_{idx}] **do**
  if visit[ctx_c_{idx}] is False **then**
    ContextConflictSolver(nb_c_{idx}, ctx_dep, scan_order, s_{idx})
  end if
end for
scan_order[s_{idx}] = c_{idx}
visit[c_{idx}] = True
s_{idx} = s_{idx} + 1

An example of the proposed adaptive scan order design for a 3×3 transform block is depicted in Fig. 3. Given that such design does not require an overhead rate cost, our implementation makes each transform size/kernel track its own statistics and maintain the scan order table accordingly. To minimize the increase in decode complexity, the scan order tables are updated on a frame by frame basis.

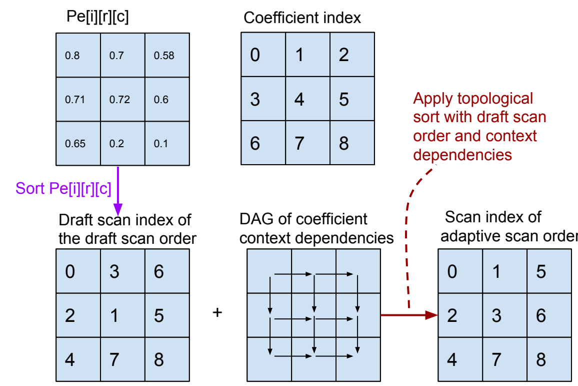

Fig. 3. An example of how to generate scan order from non-zero probabilities $Pe[i][r][c]$ and context dependencies.

4. EXPERIMENTAL RESULTS

We implemented our adaptive scan order scheme within the framework of VP9 coefficient coding engine (CCE). The original CCE scheme processes 2D-DCT coefficients in a forward zig-zag scan order and uses above and left neighbor coefficients whenever available as context information [9]. It employs either a row-biased or column-biased scan order for each directional intra prediction modes [5]. We replaced the scheme that uses a pre-defined zig-zag scan order and its variants for inter/intra modes with a streamlined scheme described in Sec. 3 that generates adaptive scan orders with an update rate of $k = \frac{1}{8}$. The scan order for each transform size and kernel is updated per frame. We also implemented and tested the proposed scheme within the framework of a context-based adaptive binary arithmetic coding (CABAC) system[10, 11] with a backward coefficient scan pattern, where transform coefficients are processed from higher frequency positions to lower frequency ones.

The compression performance is evaluated on low-resolution (CIF, SIF), mid-resolution (480p, 4CIF) and high-resolution (720p, 1080p, XGA) datasets. The operating points are selected to target quality range between 35dB - 45dB in PSNR terms. The coding gains, in terms of Bjntegaard delta rate (BD-rate) reduction, over both VP9 forward coding engine (CCE) and CABAC-based backward coding engine (CABAC), are provided in Table 1-3. The performance gains from the proposed constrained adaptive scan order scheme are consistent across CCE and CABAC coding engines. It is about 1.0% coding gain on low-resolution, 0.9% coding gain on mid-resolution, and 1.1% coding gain on high-resolution datasets.

5. CONCLUSIONS

A constrained adaptive scan order scheme is proposed to better capture the true signal statistics, while fully accounting for the context based probability modeling. It provides a streamlined scan order design for both inter and intra prediction modes, and for all the transform size and kernel combinations. Experimental results have demonstrated consistent coding performance gains over a wide range of test settings.

Table 1. Coding gains of the proposed scheme over the pre-defined scan order scheme in terms of BD-rate reduction on low-resolution dataset.

		CCE	CABAC
	res	BDRate(%)	BDRate(%)
akiyo	CIF	0.229	0.404
basketballpass	240p	0.555	0.42
blowingbubbles	240p	0.393	0.537
bowing	CIF	0.703	0.451
bqsquare	240p	0.35	0.45
bridge_close	CIF	1.837	2.166
bridge_far	CIF	7.419	7.162
bus	CIF	1.04	1.255
cheer	SIF	0.442	0.34
city	CIF	0.67	1.038
coastguard	CIF	2.875	2.931
container	CIF	0.81	0.881
crew	CIF	0.62	1.068
deadline	CIF	0.568	0.676
flower	CIF	1.303	1.359
flowervase	240p	1.351	1.466
football	CIF	0.809	0.765
foreman	CIF	0.342	0.299
garden	SIF	0.992	0.916
hallmonitor	CIF	1.792	1.88
harbour	CIF	2.488	2.555
highway	CIF	3.267	2.819
husky	CIF	1.873	1.947
ice	CIF	0.386	0.628
keiba	240p	0.585	0.858
mobile	CIF	0.649	0.59
mobisode2	240p	0.415	0.832
motherdaughter	CIF	0.575	0.495
news	CIF	0.363	0.36
pamphlet	CIF	0.233	-0.189
paris	CIF	0.256	0.303
racehorses	240p	0.769	0.617
signirene	CIF	0.332	0.325
silent	CIF	0.253	0.238
soccer	CIF	0.927	0.685
stefan	SIF	2.123	2.137
students	CIF	0.11	0.246
tempete	CIF	0.556	0.569
tennis	SIF	-0.036	0.027
waterfall	CIF	0.371	0.471
OVERALL		1.04	1.074

Table 2. Coding gains of the adaptive scan order scheme over the pre-defined scan order scheme on mid-resolution dataset.

		CCE	CABAC
	res	BDRate(%)	BDRate(%)
aspen	480p	0.295	0.274
BQMall	480p	1.001	1.041
city	4CIF	1.222	1.412
controlled_burn	480p	0.372	0.06
crew	4CIF	1.077	1.009
crowd_run	480p	0.23	0.215
ducks_take_off	480p	5.273	5.335
Flowervase	480p	3.093	3.098
ice	4CIF	0.566	0.528
into_tree	480p	0.282	0.586
Mobisode2	480p	1.371	1.175
old_town_cross	480p	-0.252	-0.117
park_joy	480p	0.03	0.173
PartyScene	480p	0.434	0.47
red_kayak	480p	0.877	1.034
sintel_trailer	480p	0.287	0.074
snow_mnt	480p	-0.106	0.176
soccer	4CIF	2.385	2.078
speed_bag	480p	0.577	0.881
station2	480p	0.183	0.17
tears_of_steel1	480p	0.74	0.615
tears_of_steel2	480p	1.127	0.729
touchdown_pass	480p	0.209	0.181
OVERALL		0.925	0.922

Table 3. Coding gains of the adaptive scan order scheme over the pre-defined scan order scheme on high-resolution dataset.

		CCE	CABAC
	res	BDRate(%)	BDRate(%)
basketballdrive	1080p	1.992	1.887
blue_sky	1080p	1.113	1.087
bqterrace	1080p	0.954	0.919
cactus	1080p	0.566	0.507
chinaspeed	XGA	0.267	0.458
city	720p	0.38	0.935
crew	720p	1.253	1.156
crowd_run	1080p	0.548	0.611
cyclists	720p	0.149	0.152
dinner	1080p	1.235	1.445
ducks_take_off	1080p	4.332	4.399
factory	1080p	0.424	0.414
fourpeople	720p	1.398	1.211
in_to_tree	1080p	0.893	0.656
jets	720p	1.191	1.212
johnny	720p	3.049	3.232
kimono1	1080p	1.259	1.421
kristenandsara	720p	2.585	2.71
life	1080p	0.033	-0.073
mobcal	720p	0.159	0.03
night	720p	0.552	0.632
old_town_cross	720p	-0.013	0.144
parkjoy	1080p	0.964	0.89
parkrun	720p	0.85	0.98
parkscene	1080p	0.569	0.393
ped	1080p	1.078	1.072
riverbed	1080p	2.335	2.343
rush_hour	1080p	1.456	1.406
sheriff	720p	2.138	2.127
shields	720p	0.43	0.802
station2	1080p	1.369	1.548
stockholm_ter	720p	0.744	0.893
sunflower	720p	0.394	0.312
tennis	1080p	0.997	0.854
tractor	1080p	1.563	1.541
vidyo1	720p	0.918	0.845
vidyo3	720p	2.158	2.266
vidyo4	720p	0.52	0.481
OVERALL		1.126	1.155

6. REFERENCES

[1] Xiaopeng Fan, Yan Lu, and Wen Gao, "A novel coefficient scanning scheme for directional spatial prediction-based image compression," *IEEE International Conference on Multimedia*, pp. 557–560, 2003.

[2] Hsin-Hui Chen, Ying-Wun Huang, and Jian-Jiun Ding, "Local prediction based adaptive scanning for JPEG and H.264/AVC Intra coding," *IEEE International Conference on Image Processing*, pp. 1636–1640, 2013.

[3] Vanessa Testoni, Max H. M. Costa, Darko Kirovski, and Henrique S. Malvar, "On the adaptive coefficient scanning of JPEG XR/HD photo," *Data Compression Conference*, pp. 69–78, 2010.

[4] James S. Walker and Truong Q. Nguyen, "Adaptive scanning methods for wavelet difference reduction in lossy image compression," *IEEE International Conference on Image Processing*, pp. 182–185, 2000.

[5] Jingning Han, Ankur Saxena, Vinay Melkote, and Kenneth Rose, "Jointly optimized spatial prediction and block transform for video and image coding," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1874–1884, 2012.

[6] Qichao Sun, Xiaoyang Wu, and Lu Yu, "An adaptive coefficient scanning scheme for inter-prediction coding," *Picture Coding Symposium*, 2007.

[7] Joel Sole, Rajan Joshi, Nguyen Nguyen, Tianying Ji, and Marta Karczewicz, "Transform coefficient coding in HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1765–1777, 2012.

[8] James Bankoski, Paul Wilkins, and Yaowu Xu, "Technical overview of VP8, an open source video codec for the web," *IEEE International Conference on Multimedia*, 2011.

[9] Debargha Mukherjee et. al., "A technical overview of VP9- the latest open-source video codec," *SMPTE Motion Imaging Journal*, vol. 124, no. 1, pp. 44–54, 2015.

[10] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the high efficient video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[11] Detlev Marpe, Heiko Schwarz, and Thomas Wiegand, "Context-based adaptive binary arithmetic coding in the H. 264/AVC video compression standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 620–636, 2003.