# New Prediction Techniques for Inter- and Intra-Frames of Advanced Video Coding

Chung-Cheng Lou*, Szu-Wei Lee†, Seung-Hwan Kim* and C.-C. Jay Kuo*
* University of Southern California, Los Angeles CA, USA
E-mail: {clou, Kimseung}@usc.edu, cckuo@sipi.usc.edu
† Intel Corp., Santa Clara CA, USA
E-mail: szu-wei.lee@intel.com

***Abstract*—A novel multi-order-residual-prediction (MORP) coding approach is proposed to improve spatial prediction efficiency in video coding. We observe that the compression ratio of a video coding algorithm depends on the nature of sequences as indicated by the ratio between inter and intra blocks in the bit-stream. When the percentage of intra blocks increases, the prediction efficiency decreases, thus leading to a poorer coding gain. In other words, one bottleneck of video coding lies in poor intra prediction efficiency. To address this issue, we propose an MORP coding scheme that adopts a second-order prediction scheme after the traditional first-order prediction. Different prediction techniques are adopted in different stages to tailor to the nature of the corresponding residual signals. The proposed MORP scheme outperforms H.264/AVC by a significant margin for the intra block coding and, thus, improves the overall coding efficiency especially for HD coding.**

## I. Introduction

The H.264/AVC standard contains many state-of-the-art techniques to achieve superior coding performance in low resolution and low bit-rate mobile applications. Recently, the market has witnessed a surging demand for high fidelity and high definition (HD) video contents. The increased data rate will impose higher pressure on networks, and more efficient coding techniques are needed. To address this demand, the Fidelity Range Extensions (FRExt) of H.264/AVC provide a number of enhanced capabilities to its base specification. A new successor, called High Efficiency Video Coding [1], is also in the standardization process to achieve a better coding gain. Besides the standardization work, there are other developments such as [2] and [3] that aim to achieve a better coding gain by cascading coding units of H.264/AVC to exploit the correlation of residual signals.

The effectiveness of rate reduction in video coding mainly comes from prediction. There are two prediction methods; namely, the intra-prediction and the inter-prediction. In this work, we first demonstrate that, when the video signal has weak inter-frame dependency and the bit-stream is dominated by intra modes, existing coding standards reach a performance bottleneck. Furthermore, we observe that the number of the intra blocks increases dramatically as the coding requirement moves toward high resolution (or finer QPs). To improve the efficiency of spatial prediction, a novel Multi-Order-Residual Prediction (MORP) scheme is proposed in this work, which improves the overall efficiency for inter- and intra-frames.

## II. Performance Analysis of H.264/AVC Prediction

The inter prediction is based on the assumption of object translation between adjacent frames while the intra prediction assumes smooth and homogeneous regions. Generally speaking, inter prediction performs better than intra prediction as shown in Table I, where the Blue Sky HD test sequence is encoded with only inter or intra modes at $QP = 20$. There exists a huge efficiency difference when the sequence is coded solely by the inter prediction (13.6 Mbps) or by the intra prediction (51.7 Mbps). For inter prediction, 4.5% and 8.8% of the bits are spent in the partitioning schemes and MVs, respectively. Yet, these prediction overheads effectively reduce the coding bits for prediction residuals. In contrast, the intra prediction has lower overhead (1.0%) since the predictive signal can be specified by one of several possible modes. Yet, it fails to provide rate saving in complex texture.

TABLE I
The bit allocation result when the Blue Sky HD (1920x1080) test sequence is coded by inter or intra prediction only at $QP = 20$.

| Bit Allocation | | Inter Prediction | Intra Prediction |
|---|---|---|---|
| Prediction Overhead | Mode | 4.473% | 0.966% |
| | Motion Info | 8.796% | NA |
| Residual | | 86.731% | 99.034% |
| Average Bit-rate | | 13.605 Mbps | 51.746 Mbps |

When a large portion of a video sequence lacks temporal dependency and the corresponding frames have to be coded as intra modes, the overall bit-rate tends to increase. There are two major factors causing this phenomenon: 1) frame size and 2) the QP value.

When a sequence has a small frame size such as CIF, the scene is represented by fewer pixels, with more image detail and less homogeneous area in a fixed region, *i*.e., a 16x16 MB. Therefore, intra prediction is less effective since there is no spatial similarity among blocks. On the other hand, the object movement can be tracked by motion compensated prediction (MCP) to reduce temporal redundancy. As shown in Table II, the inter mode percentage of sequences of the CIF resolution is a dominant majority as compared with sequences of higher resolution.

For a sequence of higher resolution, the enlarged smooth area facilitates intra prediction as a block can be inferred by its neighbor pixels. Moreover, more film grain noise induced in the video capturing process appears in the video content [4]. This noise-like signal increases the residual cost of both prediction modes. Therefore, the intra prediction is able to achieve similar residual cost with less overhead. For 1920x1080 HD resolution, only less than 33% of a frame is coded by inter modes as shown in Table II.

TABLE II
MODE DISTRIBUTION AND BIT-RATE COMPARISON FOR VARIOUS SEQUENCES CODED AT $QP = 12$.

| Resolution | Sequence | Mode distribution | | Bit-rate |
|---|---|---|---|---|
| | | Inter mode | Intra mode | |
| CIF | Foreman | 83.51% | 16.49% | 5.459 Mbps |
| | Soccer | 86.84% | 13.16% | 6.556 Mbps |
| 1280x720 | Shields | 70.75% | 29.25% | 87.680 Mbps |
| | Stockholm | 56.65% | 43.35% | 89.520 Mbps |
| 1920x1080 | Pedestrian | 32.15% | 67.85% | 95.126 Mbps |
| | Rush Hour | 10.70% | 89.30% | 98.917 Mbps |
| | Riverbed | 00.75% | 99.25% | 157.781 Mbps |

The QP value is tightly coupled with picture quality. It also has an impact on the coding mode distribution. We perform a detailed analysis on the mode distribution of a wide range of QP values for the Rush Hour HD sequence in Fig. 1. For the low bit-rate coding ($QP > 24$), there are more than 70% inter and skip modes as the final modes. When the picture fidelity requirement is low, most MBs are coded as skip modes, which result in a significant gain in rate reduction. As the QP gets finer ($< 16$), the percentage of inter modes drops quickly to 30% or lower, while a large number of frames are coded as intra modes. It can also be seen from the bit-rate curve that the compression ratio is higher with lower quality requirement, yet the bit-rate goes up rapidly when the QP is set to be smaller than 16 due to the poor intra prediction performance.

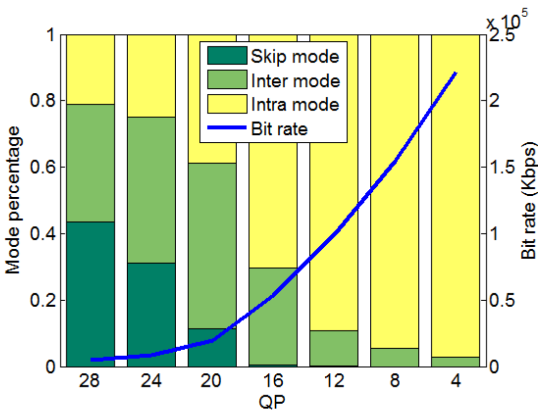



Fig. 1. Mode distribution and the bit-rate of the Rush Hour HD sequence with different QPs.

To conclude, the coding results of H.264/AVC for lower resolution sequences such as CIF at a coarse QP indicates that the performance gain is primarily achieved through the extensive use of efficient MCP. However, this coding performance does not hold for sequences of larger resolution and quantized by finer QP values. The efficiency of MCP degrades and the line-based intra prediction (LIP) cannot provide good rate savings. The two factors result in the degraded performance of HD video coding.

## III. Multi-Order-Residual Prediction Coding

In this section, we propose a novel Multi-Order-Residual-Prediction (MORP) scheme to address the inefficiency of the intra prediction for HD video.

### A. MORP Coding Architecture

The inefficiency of the existing H.264/AVC codec can also be observed from the prediction residual. A sample frame is taken from the Riverbed sequence of resolution 1920x1080 coded with $QP = 12$ as shown in Fig. 2(a). The prediction residual is plotted in Figs. 2(b) and 2(c). We see that the residual image still contains untreated structural information. This motivates us to apply a second order prediction to reduce residual bits furthermore.

In the proposed MORP scheme, picture redundancy is treated in two steps. The block diagram of the system is illustrated in Fig. 3. The first order prediction takes the captured video signal as the input, and the reference frame for the inter prediction comes from the summation of the two-order prediction frames and the residual frame, namely, the reconstructed frame. The second-order prediction takes the residual after the first-order prediction with small image features as the input, while the reference frame is taken from the coded second-order prediction frame and the residual frame. The pairing between the input and the reference frames in each order ensures the best match of signal characteristics.

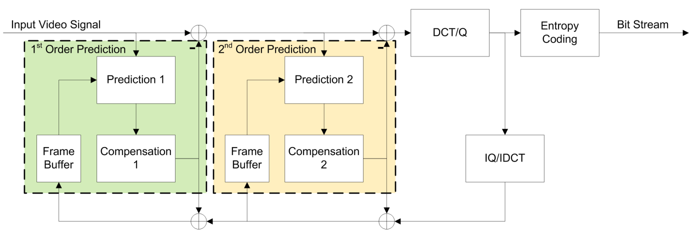


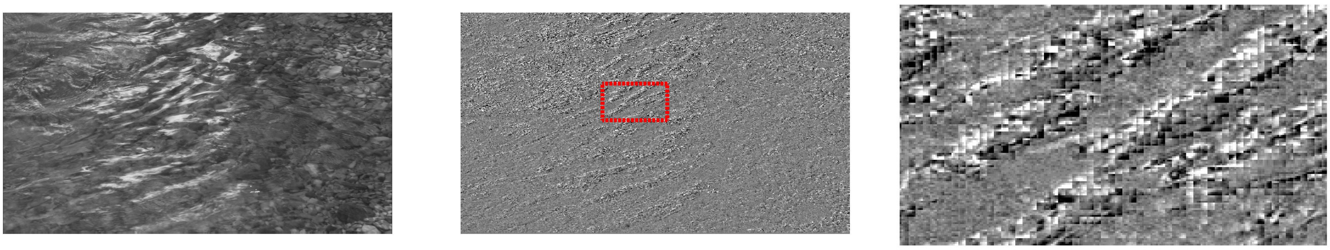

Fig. 3. The block diagram of the proposed MORP coding scheme.

### B. First-Order and Second-Order Prediction Techniques

To fit the characteristics of each residual images, the prediction techniques should be carefully chosen in the first-order and the second-order residual images. There are two major differences between the first-order and the second-order residual images.

First, the second-order residual contains less redundancy since most of them already exploited in the first-order prediction. The second-order residual contains the structural information plus the noise-like signal, which further lowers the similarities of the residuals and affects prediction efficiency.

Second, the dynamic range of the residual image is much narrower than that of the original input signal. In Fig. 4,



(a) A sample frame. (b) Prediction residual of the sample frame. (c) A close-up view of the rectangle area in (b).

Fig. 2. A sample frame of the prediction residual from the Riverbed HD sequence with $QP = 12$.

we compare the pixel histogram of the input signal and the prediction residual of the Riverbed sequence. The luminance of the input signal takes a value between 0 and 255 with a great majority falling in the range of 60 and 120 while the luminance of the residual is highly concentrated at 0. Note that one of the goals for prediction is to reduce the signal dynamic range. Given a narrower distribution of the residual, the second-order prediction technique should not have much overhead. Otherwise, the prediction gain will be significantly reduced.

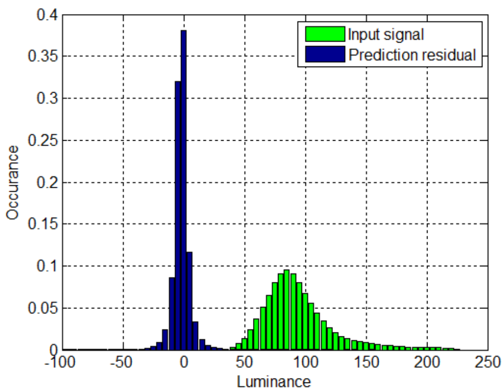



Fig. 4. The histogram of a sample frame of the Riverbed HD sequence with $QP = 12$.

Consider the two types of prediction methods: the translation-vector (TV) based and the line based. The TV based prediction provides more zero-skewed residuals as compared with the line based prediction, yet the overhead of TVs is large. When the image sequence contains great similarities between adjacent frames, the TVs based prediction can reduce image redundancy effectively. In contract, the line based prediction can predict blocks in a homogeneous region efficiently with little prediction overhead.

For the first-order prediction, the input signal contains a large amount of similarities which can be easily exploited. Thus, the block based MCP in H.264/AVC is adopted for the inter modes with its proven coding performance to reduce the inter-frame dependency with a flexible block size. For the intra modes, besides the LIP, the block based intra prediction (BIP) [5] is also adopted for structured image features such as image texture or edges. The rate-distortion analysis is used to choose either one of the two.

The second-order prediction deals with the residual that has less image structure but more noise-like contents. For the inter prediction, MCP is still adopted to exploit the small inter-frame dependency. For the intra prediction, we only adopt the LIP due to its low prediction overhead. However, there is a slight difference between the first-order LIP and the second-order LIP. That is, the default DC prediction value of the second-order LIP is set to 0 (instead of 127) to reflect the actual DC value of the residual. The BIP is not used mainly due to its high prediction overhead.

Besides the inter or intra prediction, a bypass mode is also implemented, where the second order prediction process is skipped. The bypass mode is likely to happen when the skip mode is chosen in the first order prediction or when the first order prediction already provides good coding performance.

### C. Rate-Distortion Optimization (RDO)

Mathematically, to obtain the best trade-off between the distortion and bit-rate saving, a Lagrangian cost function can be used to find the minimal cost $J$ among the available modes as

$$J(s, p_1, p_2, r) = D(s, p_1, p_2, r) + \lambda(R_{pred1} + R_{pred2} + R_{res}(r)), \quad (1)$$

where $s, p_1, p_2, r$ denote the original, the first-order prediction, the second-order prediction, and the quantized residual blocks, $D(s, p_1, p_2, r)$ is the distortion between the original and the reconstructed signal, $R_{pred1}$ and $R_{pred2}$ are the prediction overhead bits for the two predictions, $R_{res}(r)$ is the bits to encode the quantized residual signal, and $\lambda$ is the Lagrangian multiplier.

However, the Lagrangian function in Eq. (1) involves joint optimization between the two prediction orders with high computational complexity. Here, we approximate $J$ by two decoupled functions $J_1$ and $J_2$ so that each prediction can be processed individually. That is,

$$J_i(t_i, p_i, r_i) = D(t_i, p_i, r_i) + \lambda(R_{predi} + R_{res}(r_i)), \quad i = 1, 2, \quad (2)$$

where $t_1 = s$ and $t_2 = s - p_1$ representing the input signal of each order. With the simplification in Eq. (2), the RDO is accomplished in two stages sequentially. In the first stage, the original video signal $s$ is fed into the first-order prediction with $J_1$ as the RDO objective function assuming no further prediction. Note that the prediction residual $r_1$ in $J_1$ is the quantized residual after the first order prediction, and it is only for the RDO purpose and not included in the final bit stream for transmission. After the first-order prediction, the residual, $(s - p_1)$, is fed into the second-order prediction as input $t_2$, where $J_2$ is used as the objective function in the RDO analysis. The quantized residual $r_2$ is coded to form the final output bit-stream.

## IV. Experimental Results

The proposed MORP coding scheme was implemented in the reference software JM16.2 to demonstrate its effectiveness. The benchmark JM software encoded the sequences with the high profile and CABAC was used for the entropy coding.

The prediction techniques in the first-order and the second-order are different in the proposed MORP scheme. To understand the contribution of each individual prediction, the Riverbed sequence was first coded using intra modes only with four different settings: 1) H.264/AVC, 2) H.264/AVC with BIP support, 3) MORP with only LIP in both orders, and 4) the proposed MORP scheme. The results are shown in Fig. 5. The curve "H.264+BIP" outperforms "H.264" by 2-3 dB under lower bit-rate ($< 50$ Mbps), as the BIP is able to reduce the coefficient bits by TVs when the picture fidelity requirement is low. When the bit-rate becomes high ($> 50$ Mbps), the proposed MORP is able to provide significant rate reduction even without BIP. When the two factors join together, the curve "MORP" outperforms H.264/AVC in all bit-rate range.

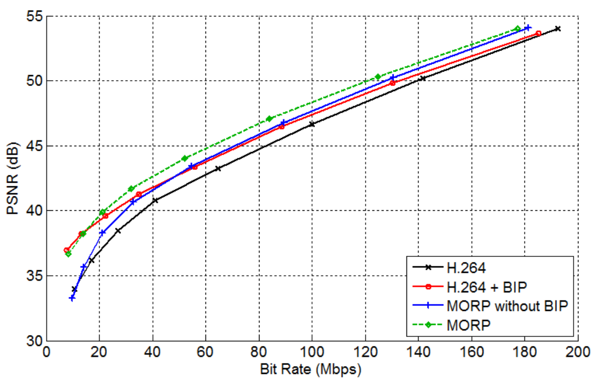



Fig. 5. The rate-distortion comparison between the proposed MORP scheme and the benchmark codec when the Riverbed sequence of resolution 1920x1080 is coded by intra modes only.

The average bit-rate and PSNR difference of the proposed MORP algorithm are summarized in Table III, where the test sequences of resolution 1920x1080 were encoded by two schemes: 1) the I frames only and 2) the normal I/P setting. When the test sequences are coded by intra modes, the average bit-rate reduction is 21.64%. Tractor and Riverbed sequences have more than 30% bit saving due to their irregular picture features. When both inter and intra modes are adopted, the overall improvement depends on the intra bit saving and the intra mode percentage. For test sequences with a lower percentage of intra modes and a larger percentage of inter and skip modes, like Station and Ducks Take Off, the saving is below 9%. For other sequences with higher intra mode percentage, like Riverbed, the saving can be as high as 19%.

TABLE III
The coding efficiency comparison of the proposed MORP scheme and H.264/AVC.

| Sequence | I Frame Only | I/P Frame | |
|---|---|---|---|
| | Δ Bit Rate | Δ Bit Rate | Intra Mode |
| Station | -19.71% | -08.88% | 30.67% |
| Ducks Take Off | -19.17% | -06.51% | 31.91% |
| Tractor | -36.50% | -13.71% | 34.26% |
| Old Town Cross | -10.12% | -05.69% | 41.78% |
| Blue Sky | -14.22% | -06.69% | 47.37% |
| Rush Hour | -20.34% | -10.24% | 50.22% |
| Pedestrian | -20.24% | -11.00% | 53.00% |
| Riverbed | -32.81% | -19.28% | 94.07% |
| **Average** | -21.64% | -10.25% | 47.91% |

The overall encoding complexity may increase a little due to the use of the second order prediction, which is performed to improve the coding efficiency of the intra MB. If there are 47.91% intra-coded MBs as shown in Table III, and the encoding complexity of intra-prediction is 0.54% of the whole encoding system [6], then the additional encoding complexity due to the second order prediction is about 0.26%. We plan to study the reduction of the encoding complexity for the second order prediction in the future.

## V. Conclusion

A novel Multi-Order-Residual-Prediction (MORP) scheme was proposed to improve video coding efficiency. The prediction residual was further exploited by the second-order prediction, and different prediction techniques were developed to reduce residual redundancy in different orders. It was shown by experimental results that the proposed MORP algorithm reduces the bit-rate by 10.25% as compared with H.264/AVC.

## References


[1] JCT-VC. Architectural outline of proposed high efficiency video coding design elements. JCTVC Meeting (JCTVC-A202), 2010.

[2] S. Li, S. Chen, J. Wang, and L. Yu. Second order prediction on h. 264/AVC. In Proceedings of the 27th conference on Picture Coding Symposium, pages 85–88. Institute of Electrical and Electronics Engineers Inc., The, 2009.

[3] J.W. Kang, C.-C. Lou, S.-H. Kim, and C.-C. Jay Kuo. Improved FOR/SOR-Based video coding and its performance analysis. In IEEE International Conference on Image Processing, to be published, 2011.

[4] A. Amer and E. Dubois. Fast and reliable structure-oriented video noise estimation. IEEE Transactions on Circuits and Systems for Video Technology, 15(1):113–118, 2005.

[5] Y. Dai, Q. Zhang, A. Tourapis, and C.-C. Jay Kuo. Efficient block-based intra prediction for image coding with 2D geometrical manipulations. In Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, pages 2916–2919, 2008.

[6] T. C Chen, H. C Fang, C. J Lian, C. H Tsai, Y. W Huang, T. W Chen, C. Y Chen, Y. H Chen, C. Y Tsai, and L. G Chen. Algorithm analysis and architecture design for HDTV applications-a look at the h. 264/AVC video compressor system. Circuits and Devices Magazine, IEEE, 22(3):22–31, 2006.