Adaptive Video Coding Using Bitplane Modeling And GFA Representation

Paul Bao, *Senior Member IEEE* and Xiaohu Ma
Nanyang Technological University

Abstract: In this paper, we present a novel video coding scheme based on the adaptive non-uniform bitplane modeling of video sequences in wavelet domain and the generalized finite automata (GFA) representation. Unlike the traditional block-based motion compensation coding, where a video sequence in GoPs is arranged as *I-, P-* and *B*-frames and a motion estimate is searched in one or a few reference frames at the fixed or predefined block sizes at a spatial domain metric, in the proposed scheme, a video sequence is represented in GoPs as an overall binary image by *bitplane modeling* the significant coefficients of the video sequence within subbands. The inter-frame, inter-level and inter-bitplane *similarities* inhabited in the binary image are then optimally explored, leading to a compact GFA representation of the bitplane modeled video sequence. Finally, all the transitions in the GFA representation are entropy encoded into a scalable bitstream. The proposed scheme significantly outperforms the H.26X series coding schemes in rate-distortion performance. It could achieve bitrate ranges at 4-5 Kbps and 15-18 Kbps for QCIF 10 Hz and QCIF 30 Hz sequences, respectively, a target unachievable by even the newly emerged H.264 standard.

1. Introduction

Motion estimate based video coding [1-4], whether in wavelet domain or spatial domain, fails to achieve a very low bitrate (in the range of 4-6 Kbps for QCIF 10 Hz format) with minimum motion compensation. Without motion compensation, however, the visual perception and the objective quality of the video tend to be very poor and unacceptable for most applications. This is mainly due to its incapability to explore *all* the potential similarity (redundancy) inhabited in a video sequence. The motion estimate could only explore the similarities between an artificially selected reference frame (I- or P-frame) and other frames at the fixed size (16×16) of macroblocks. Furthermore, the similarity is measured only inter-frame by a metric. Consequently the scheme based on this strategy would fail to explore and capture the *pairwise* inter-frame similarities in blocks of variable sizes. It also fails to capture the inter-bitplane transformation-based similarity, which usually forms the major contribution to the redundancy. In order to significantly improve the bitrates while attaining an acceptable visual quality or PSNR measurement, all the potential *similarities* in a video sequence should be defined and best possibly explored.

In this paper, we present a GFA modeling based video coding scheme aimed at optimally exploring all the similarities implicitly inhabited in video sequences, in particular, the transformation-based inter-bitplane similarity. The GFA modeling of the video sequences is capable of optimizing the rate-distortion performance of the proposed video coding.

2. Bitplane Modeling For Video

The GFA presentation of the video sequence in wavelet domain seemingly facilitates an optimal exploration of the *generalized* similarities. The GFA modeling takes full advantages of the binary fractal similarity of the video sequence in wavelet domain to form a Generalized Finite Automata (GFA)-based compact representation of video sequence. In this model, the interlevel (wavelet), inter bit-plane and interframe fractals in a GoP would be fully explored and the GFA will be entropy and streaming encoded at very low bitrates while retaining an acceptable perceptual quality of video.

2.1 Adaptive Non-uniform Quantization

Aiming at facilitating more spectrum redundancy to be captured in the bitplane model, we analyze the non-uniform quantization of the coefficients from the perspective of the Generalized Gaussian distribution. Consider a non-uniform quantization for coefficients as shown in figure 3. For simplicity, we only show the 4 most significant bits of the quantization. We may observe from figure 3 that the first two significant bits of the coefficients in the green interval complement to each other; furthermore the 2nd and 3rd significant bits of the coefficients in the red interval and equal to each other and finally the 2nd , 3rd and 4th significant bits of the coefficients in the blue interval equal to each other. Thus we can quantize the coefficients non-uniformly so that *most* coefficients have complement first 2 significant bits, equal 2nd, 3rd and/or 4th bits, facilitating subsequently a compact GFA representation of the video for an excellent rate-distortion performance. Obviously, the non-uniform quantizer so defined may lead to a larger quantization error e_Q. We would like to seek a quantizer Q and a similarity error model σ (to be explained in section 4.1) so that the overall distortion is minimized

$$\min_{R(Q,E)<R_T} D(Q,\sigma)$$

where $D(Q,\sigma)$ and $R(Q,\sigma)$ are the overall distortion and bitrate corresponding to quantizer Q and error model σ and R_T the target bitrate. This minimization problem can be formulated as the *Lagrangian* optimization on the cost function

$$J = D(Q,\sigma) + \lambda R(Q,\sigma)\,. \qquad (1)$$

However, it can be very time consuming to search for the quantization and error model parameters that minimize the cost function J. For each combination of the quantization and error model parameters, GFA modeling, coding of the basic blocks and entropy coding of GFA would have to be performed to obtain the corresponding rate and distortion. Due to the complexity for the calculation of the rate and distortion, the evaluation of cost function J for all the possible combinations is apparently impractical and an empirical approach in correlating the quantizer Q with the rate-distortion performance is essential for the bitplane modeling and the subsequent GFA representation.

It is observed that the distribution of wavelet coefficients of a large set of images and video sequences can be adequately described by *Generalized Gaussian distribution* (GGD) with the shape parameter β, ranged in (0.5,1). With the GGD which adequately describe the distribution of the wavelet coefficients, we can design the non-uniform quantizer aimed at optimizing the rate-distortion function (1) empirically. Different from the traditional goal of the quantizer for the data compression where the reconstruction values and quantization bins are selected so as to minimize the average distortion, the principle goal of the quantizer here instead is to find the reconstruction values $\hat{y} = \{\hat{y}_i^Q\}$ and quantization bins R_i $i = -m, \cdots -1, 0, 1, \cdots, m$ (m levels on each side plus a zero-zone bin) that minimize the final coding distortion comprised of the quantization error and the modeling error.

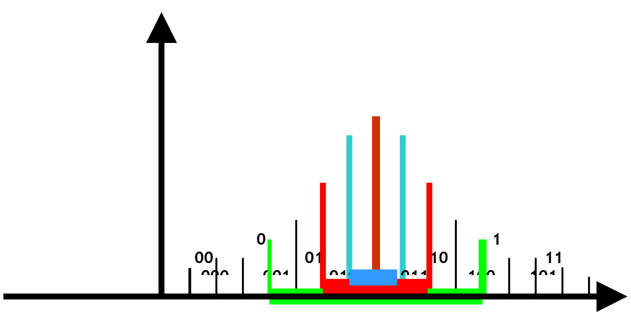

Figure 3. Non-uniform quantization for bitplane modeling

From the analysis on the quantization bitplane correlation (figure 3), we may design the quantizer *hierarchically* instead of sequentially as follows. We start at the initial range of all the positive coefficients (level-1) by partitioning the range into two (unequal) level-2 regions and proceed to the consequent levels iteratively. At level-i, we partition each of the level-i regions in into two level-($i+1$) regions until we reach level $r = \log m$. All the regions on level r will then form the quantization bins and their centroids will be the reconstruction values, respectively. Since we aim at promoting the similarities between the bitplanes in the bitplane model, we would like to partition the regions resulting in the bitplane similarities under a given error modeling. At level $i = 1$, we partition the region $R^{(1)}$ (the range of the positive coefficients) into two level-2 regions $R_1^{(2)}$ and $R_2^{(2)}$ so that the number of coefficients in $R_2^{(2)}$

$$\left|R_2^{(2)}\right| < \delta_1 .$$

Similarly, we partition region $R_1^{(2)}$ into two level-3 regions $R_1^{(3)}$ and $R_2^{(3)}$ so that

$$\left|R_2^{(3)}\right| < \delta_1 + \delta_2$$

and partition $R_1^{(3)}$ into two level-4 regions $R_1^{(4)}$ and $R_2^{(4)}$ so that

$$\left|R_2^{(4)}\right| < \delta_1 + \delta_2 + \delta_3 .$$

Since the quantization of the rest of regions is insignificant to the subsequent GFA modeling, we apply the uniform quantization to these regions. The centroid reconstruction values y_i for the quantization bin $R_i^{(r)}$ are calculated as

$$\hat{y}_i^Q = \frac{\int_{R_i^{(r)}} y GG_{\sigma_Y,\beta}(y)dy}{\int_{R_i^{(r)}} GG_{\sigma_Y,\beta}(y)dy} .$$

The negative coefficients are quantized in a symmetric ways and zero coefficients are kept as zeros. The resolution of the quantization is $r + 1 = \log N = \log(2m+1)$.

3. GFA Modeling of Binary Images

3.1 Genealogy graph representation of GFA

A genealogy graph is graph where all the nodes are classified by their characters into different generations. In a GFA genealogy graph, nodes are classified by the sizes of the subimages they represent: the original image sized at $n \times m$ is classified as the generation zero, represented by *root*; an image sized at $\frac{n \times m}{2^k \times 2^k}$ is classified as generation k; an image represented by a leaf node will have size 11×9.

3.2 Transition

An edge in the genealogy graph represents a GFA transition, labeled either by a triplet $\{j, i, t\}$, where i and j are quadrant indices and t the transformation, a quadruplet $\{s_i, q, s_t, t\}$, where s_i, q and s_t represent the in-state, quadrant and to-state, respectively or a septuplet $\{s_i, q, s_t, 9, x, y, t\}$,where (x, y) represents the motion vector and 9 is the transformation index for the motion estimate. The quadrants are indexed as 0,1,2, and 3. A triplet $\{j, i, t\}$, where $i < j$, represents a *self-transition* from quadrant j to quadrant i with transformation t, implying that quadrant j can be derived from quadrant i

(refer to the self transition on state 2 in figure 4). A quadruplet represents a transition from the current state (in-state) to one of its child states. A septuplet transition is an edge between the states on the same level (red edges in figure 4) and represents a motion-estimate transition. A transition from the current state to one if its grandchildren states on the genealogy graph (pink edges) indicates that the corresponding quadrant of the current will be recovered by 2^g replications of the to-state, where g is the number of generations between them. The multiple transitions from an in-state to the same to-state (subimage) will be represented by a single edge with the grouped transition parameters.

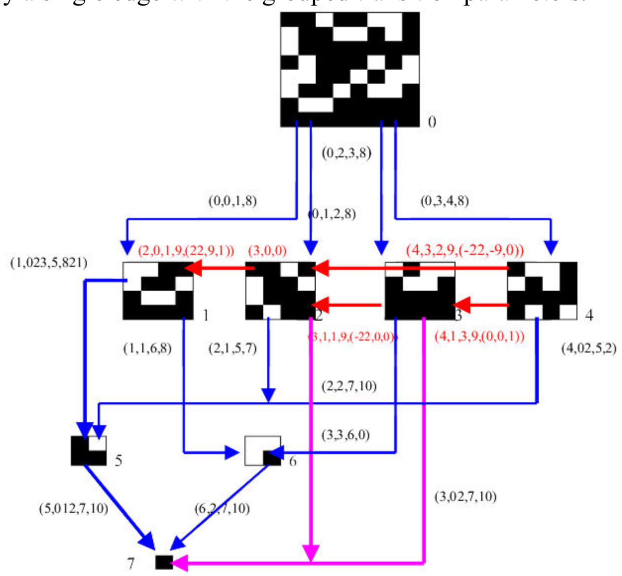

Figure 4. GFA modeling of the chessboard

4. Experiment

The preliminary results on the *carphone* sequence have achieved a rate at 5 Kbps for 10 Hz or 18 Kbps 30 Hz with PSNRs ranged between 29-33 dB, respectively, showing an excellent potential in achieving the targeted performance by the proposed video coding scheme.

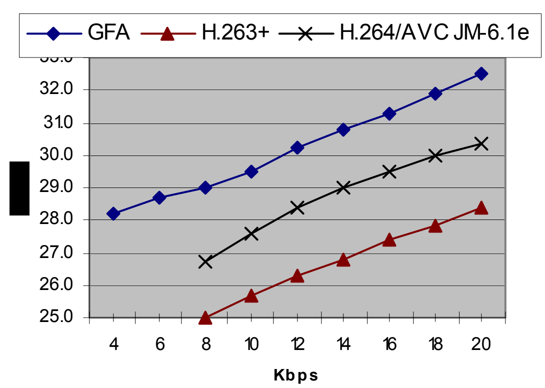

Figure 5. Comparisons between the GFA scheme, H.263+ and H.264 JM-6.1e for QCIF *Carphone* at 8Hz

Table I: Rate-distortion

	Carphone					
PSNR	32.16	31.94	29.93	29.48	28.99	28.48
CR	70.56	83.54	215.23	257.62	314.83	462.20
Bit-rate (8 Hz)	33.67	28.44	11.04	9.22	7.55	5.14
Bit-rate (30 Hz)	126.27	106.65	41.39	34.58	28.30	19.27

6. Conclusion

We propose a video coding scheme using the bitplane modeling and GFA representation aimed at optimally exploring interframe, interlevel and interbitplane similarities inhabited in video sequence. The proposed scheme significantly outperforms the H.26X series coding schemes in rate-distortion performance. It could achieve bitrate ranges at 4-5 Kbps and 15-18 Kbps for QCIF 10Hz and QCIF 30Hz sequences, respectively, a target unachievable by even the newly emerged H.264 standard. Numerous multimedia communication applications, previously unpractical, would be envisaged with this supremacy in bitrates.

References

[1] R. Schafer, T. Wiegand and H. Schwarz, The emerging H.264/AVC standard, http://www.ebu.ch/trev_293-schaefer.pdf

[2] S. Cho and W. A. Pearlman, " A full-featured, error-resilient, scalable wavelet video codec based on the set partitioning in hierarchical trees (SPIHT) algorithm", *IEEE Trans. CSVT*, Vol. 12 No. 3, pp. 157 –171, 2002

[3] S. H. Yoon, K. Chellapilla, and S.S. Rao, "Wavelet based scalable video system using bit-plane encoding" *Proceedings of 1997 IEEE International Symposium on Circuits and Systems,* Vol. 2, pp. 1133-1136, 1997

[4] U. Hafner, J. Albert, S. Frank, and M. Unger, "Weighted finite automata for video compression", *IEEE Journal on Selected Areas in Communications*, Vol. 16, No. 1, pp.108-119, Jan. 1998

Figure 6. Reconstructed *Carphone* (PSNR=32.9752 & 34.32 Kbps 8Hz)