A NOVEL PATTERN IDENTIFICATION SCHEME USING DISTRIBUTED VIDEO CODING CONCEPTS

Manoranjan Paul and Manzur Murshed

Gippsland School of Information Technology, Monash University, Churchill, Vic-3842, Australia

E-mail: {Manoranjan.paul, Manzur.Murshed}@infotech.monash.edu.au

ABSTRACT

Pattern-based video coding focusing on moving region in a macroblock has already established its superiority over recent H.264 video coding standard at very low bit rate. Obviously, a large number of pattern templates approximate the moving regions better however, after a certain limit no coding gain is observed due to the increase number of pattern identification bits. Recently, distributed video coding schemes used syndrome coding to predict the original information in decoder using side information. In this paper a novel pattern identification scheme is proposed which predicts the pattern from the syndrome codes and side information in decoder so that actual pattern identification number is not needed in the bitstream. The experimental results confirm that this new scheme successfully improves the rate-distortion performance compared to the existing pattern-based video coding as well as H.264 standard. This new scheme will also open another window of syndrome coding application.

Index Terms—Video coding, pattern recognition, side information, distributed video coding, H.264, low bit rate.

1. INTRODUCTION

Applications such as video conference, video call, and Internet video streaming through limited bandwidth channel require very low bit rate (VLBR) video coding technique. The recent blockbased video coding standard H.264/AVC [1] has introduced variable block sizes (16×16, 16×8, 8×16, 8×8, 8×4, 4×8, and 4×4 pixels) motion estimation (ME) and motion compensation (MC) strategy to capture various motions within a 16×16 pixels macroblock (MB). However, this strategy failed to exploit intramacroblock temporal redundancy (ITR) (see Fig 1) which is static in successive frames. Indeed, few bits are used to signal zero residual error and zero motion for ITR to decoder, which is obviously significant when a sequence is encoded at VLBR. Pattern-based video coding (PVC) [2]-[5] exploited the ITR using binary patterns (see Fig 1) in partitioning the suitable MB via a simplified segmentation process that avoided handling the exact shape of the moving objects, so that popular MB-based ME&MC techniques could be applied. ME and MC are carried out using only pattern covered region. Thus, a significant compression is achieved compared to the H.264 as the pattern size is one-fourth of the original MB. A pattern codebook (PC) of 32 pattern templates is given in Fig 2.

The moving region (MR) of an MB has different shape as it depends on the part of object contained in the MB. Obviously, a large number of pattern templates approximate the MR more accurately but, after a certain limit no coding gain is observed due to the increase number of pattern identification codes (PIC). Variable length PIC using pattern frequency information and pattern co-occurrence matrix [4] used 4.62 and 4.15 bits per pattern represented MB (which is popularly know as regionactive MB (RMB)) instead of 5 bits for 32 patterns respectively. Hence, any strategy that reduces the PIC concomitantly improves the overall encoding performance by classifying more RMBs.





Very recently, computational complexity distribution of the existing video coding standards modeled with a heavy encoder and light decoder is challenged by a stark contrast distributed video coding (DVC) architecture [6][7][8][9] which is modeled with a light encoder and a heavy decoder. The DVC exploits the joint source-channel coding theorems developed by Slepian and Wolf [10] and Wyner and Ziv [11]. The main idea is to divide the source data into a finite number of sets, popularly known as *cosets*, and the original data can be retrieved in the decoder from the coset index and predicted data by applying nearest neighborhood technique. The predicted data is generated from the side information, for example, previously decoded data. If the maximum distance between actual and predicted data is E_{max} and the minimum distance of each coset at level *i* is $d_{min}=2^{i}$ then coset theorem states that the actual data will be retrieved without error if $E_{\text{max}} < d_{\min}/2$ is satisfied. In this case we need to divide data set into i+1 cosets. Compression of the scheme stems from transmitting only the coset index (syndrome) instead of the actual value.

Although the compression ratio of the DVC can theoretically be as efficient as that of the conventional heavy encoders, this upper bound can only be achieved with the accurate estimation of inter-frame correlation structures. In real videos however the correlation structures are highly spatialvarying, potentially unbounded in magnitude, and imprecisely known leading to significant imperfection. Moreover, being a block-based codec, this new technique is obviously unable to exploit the intra-block temporal redundancy as alluded for the existing standards. So far no distributed video coding schemes are comparable to the conventional coding in terms of quality.

In this paper, we use distributed coding concept to encode PIC of 32 patterns. If we applied this theorem straightforward to encode 32 patterns, we will not get any compression as the maximum prediction error is 16 for 32 patterns which require 32 cosets. Obviously smaller number of cosets will provide compression for those RMBs where accurate pattern retrieval is possible. Fig 2 shows all 32 patterns. It is interesting to observe that the gravitational centres of moving region of all patterns are almost uniformly distributed in 360⁰ with respect to centre. We will exploit this angular distance for coset formation as the dissimilarity between two patterns is directly proportional to the angular difference between them. We will use side information from the previously decoded frames as we experimentally observed that there is a strong correlation between two successive co-located MBs in terms of pattern similarity. As the same predicted system can be designed in the both encoder and decoder, we will not lose any information if we discard those RMBs where pattern could not be retrieved accurately.



Fig 2: The pattern codebook of 32, 64-pixel patterns (numbering from left to right and top to bottom), defined in 16×16 blocks, where the white (moving) region represents 1 and black (static) region represents 0 [4].

In this paper a novel PIC scheme is proposed which successfully exploited distributed video coding concepts for retrieving the original pattern in decoder without knowledge of its index. Using this approach a significant reduction of the PIC size is achieved (5 - log₂(number of cosets)) per RMB and as a result the rate-distortion performance is improved compared to the existing pattern-based coding algorithm as well as H.264 video coding standard. Note that, the proposed system is a conventional coding system and very similar to current PVC technique with only exception is it applied distributed coding concepts to retrieve the pattern index at the decoder.

This paper is organized as follows: Section 2 describes coset formation using the angular information. Section 3 describes the side information generation. Section 4 illustrates pattern retrieve, and Section 5 analyses simulations set up and results. Section 6 concludes the paper.

2. COSET FORMATION WITH 32 PATTERNS

The main target in coset formation is to maximize the distance among the patterns within a coset and minimize the distance among the cosets. A small number of cosets provide better compression as they require less number of bits (syndrome codes) to encode coset index, however the probability of getting exact pattern will be less because the distance among the patterns in a coset is less. On the other hand, if the number of cosets is large, the syndrome code size will be large but the probability of getting exact pattern will be high. Four ways namely 2×16, 4×8, 8×4, and 16×2 in coset×pattern format, 32 patterns can be divided. By considering all 64 '1's in a pattern will make the coset formation better. But this strategy is an intractable problem as we need to consider all $^{256}C_{64}$ combinations. Thus, coset formation technique needs to be converted into a simple problem where the gravitational center of a pattern will represent corresponding whole pattern. Although, by doing this we could not utilize all available information in coset formation, it simplifies the formation step. It is easily observed that the gravitational centers (GCs) of patterns (considering only white moving region) are almost symmetrical around 360⁰ (see Fig 3).



Fig 3: Positions of gravitational centers of patterns in 16×16 block and the numbers indicate the patterns.

Fig 3 shows the positions of GCs of 32 patterns. Patterns could be ranked based on the angular distance of their gravitational centers from the positive X-axis where middle point of MB is considered as origin. For example, patterns 4, 18, 28, 10, and 6 are ranked as 1, 2, 3, 4, and 5. In order to form $n = 2^k$ cosets of 32/n patterns, every *n*th ranked pattern will be grouped into one coset where *k* is an integer. The various sizes of cosets and their minimum distances (d_{min}) using angular distance are shown in Table 1. As we mention earlier, if the predicted pattern is within $d_{min}/2$ distance with the correct pattern then syndrome coding technique will find actual pattern.

Table 1: Various size of cosets using angular distance and corresponding minimum distance among the cosets

Coset	Patterns	
[30,21,15,20,31,8,2,17,25,23,9,4,28,6,13,11]	15	
4	[19,5,3,24,12,7,18,22] [30,14,31,2, 25,9,28,13]	270
	[14,32,16,26,1,27,10,29][21,20,8,17,23,4,6,11]	37
8	[19,3,12,18] [30,31,25,28] [14,16,1,10] [21,8, 23, 6]	0.50
	[5,24,7,22] [15,2,9,13] [32,26,27, 29] [20, 17,4,11]	85
16	[19,12] [30,25] [14,1] [21,23] [5,7] [15,9] [32, 27]	
	[20,4] [3,18] [31,28] [16,10] [8,6] [24,22] [2,13]	175°
	[26,29] [17,11]	
		0

The Table 1 shows the minimum angular distances 13^0 , 37^0 , 85^0 , and 175^0 using 2, 4, 8, and 16 cosets respectively. Theoretically the distances would be 22.5^0 , 45^0 , 90^0 , and 180^0 , but already designed patterns are not fully symmetrical around the circle.

3. SIDE INFORMATION GENERATION

If $C_k(x,y)$ and $R_k(x,y)$ denote the k^{th} MB of the current and reference frames respectively of a video sequence, where $0 \le x, y \le 15$, the moving region $M_k(x, y)$ of the k^{th} MB in the current frame is obtained by:

$$M_k(x, y) = T(|C_k(x, y) \bullet B - R_k(x, y) \bullet B|)$$
(1)

where *B* is a 3×3 unit matrix for the morphological closing operation •, which is applied to reduce noise, and the thresholding function T(v) = 1 if v > 2, and is 0 otherwise. If we generate a binary matrix for whole current frame with respect to reference frame, we will get moving region for entire current frame. Fig 4 shows two moving region between frame 1 & 3 in (a) and frame 3 & 5 in (b). If we carefully observed these two

moving regions we will see the similarity between two in terms of shape and size in MB level. From this observation we are motivated that side information could be generated from previously decoder successive two frames. Note that we consider alternative frames as we use 15 frames per second at ratedistortion calculation. Obviously more similarity will be observed between adjacent two frames.



Fig 4: The moving region between frame 1 and 3 (a) and frame 3 and 5 (b) for Miss America video sequence using (1).

This co-relation can be quantified using pattern similarity. Let $|Q|_{\ell}$ be the total number of ℓ 's in the matrix Q. The dissimilarity of a pattern $P_n \in PC$ with the MR in the *k*th MB can be defined as:

$$D_{k,n} = |M_k|_1 - |M_k \wedge P_n|_1.$$
 (2)

Clearly, higher the similarity lower will be the value of $D_{k,n}$. The best similar pattern P_i will be selected for the MR M_k as follows:





Fig 5: Pattern prediction by k-th best pattern of current MB using colocated MB and motion-translated co-located MB respectively.

To find the relation between the best pattern of current MB and the pattern for co-located MB in the previous decoded frame, we investigate the rank of the current best pattern for the co-located MBs. Fig 5 shows the normalized ranking. It reveals that in 40% cases same pattern is selected as the best pattern for both the current and co-located MBs. Instead of co-located MB, we also investigate the MB which is actually referenced using motion vector of the current MB (termed as motion-translated co-located MB). The experimental results confirm that later approach provides slightly better results due to the motion corelation. Obviously for zero motion vector, the result is same in both cases. The figure also shows that more than 60% cases we can find the current best pattern within the first-fourth best for the motion-translated co-located MB. Thus, we can successfully predict the actual pattern for current MB more frequently by comparing this kind of side information and syndrome coding (i.e., coset index). For convention video coding, decoder is embedded within the encoder, thus, when we are coding an MB by pattern mode, we know that whether the prediction is successful. If it is not successful we may not consider the MB as a RMB and encode it using other H.264 modes. Final mode including pattern mode is selected by the Lagrangian optimization function.

4. PATTERN RETRIEVE

Sometimes after motion estimation the residual error is almost zero or less significant due to the smooth translational motion and coarse quantization. Moreover, zero motion vector is observed for a significant number of cases in the typical video sequences. These two features we can exploit in our pattern prediction scheme to increase the pattern retrieval accuracy. In the existing PVC scheme, only the pattern covered MR is considered for the ME&MC. The rest of the region of an MB is copied from the co-located MB as a static region from the decoded frame. When a pattern could not accurately approximate the MR of an MB, the pattern mode does not provide the best rate-distortion performance due to the pattern mismatch error. Thus, our proposal is to consider the whole MB instead of pattern covered MR for ME&MC. By doing this, we not only eliminate pattern miss-approximation error but also reduce the extra ME computational cost as we already have motion vector for the entire MB from the 16×16 mode. Thus, when residual error is zero after quantization, retrieving a wrong pattern through our prediction scheme could not affect in rate-distortion performance.

In our proposed scheme, we determine the best pattern for the current MB by (2) and (3). Then we find the coset index of this pattern. We also find the best pattern (i.e., side information) from the motion-translated co-located MB in the decoded frame. From the coset index and side information, we apply nearest neighborhood technique to find the predicted pattern. If the predicted pattern and the actual pattern (i.e., the best pattern for the current MB) are same pattern, we will encode this MB by the best pattern. Otherwise we will try with second/third/fourth best pattern at the same way. If no one is successful, we will encode this MB by other H.264 modes instead of pattern mode.



Fig 6: Percentage of accurate pattern prediction by first four best patterns using different cosets (left) and pattern prediction by the first four best patterns (right).

5. SIMULATION RESULTS

The existing PVC, proposed scheme (termed as PVCcosetX {2, 4, 8, 16}), and the H.264 algorithms are implemented based on the Baseline profile of H.264/AVC [12] with full search fractional pel motion estimation for the number of standard QCIF [12] video sequences at 15 frames per second (fps). The experimental results reveal that 85~100% correct decision can be achieved up to 20 quantization parameter (QP) [12] (see Fig 6) using all four best patterns. For obvious reason, the prediction is improving with the number of cosets (see Fig 6 (left)). Whereas only the best pattern can predict more than 60% patterns accurately when QP = 20 (see Fig 6 (right)). It is clear that, using four patterns instead of one will increase the computation, however, it can be reduced using already available motion estimation information of 16×16. Thus, proposed scheme is comparable with the existing PVC in terms of computational complexity. The number of RMBs is decreasing with bit rates (see Fig 7). Thus, rate-distortion gain using the proposed and the

PVC schemes have diminishing return with the bit rates compared to the H.264.



Fig 7: Average number of RMBs with different QPs using eight standard video sequences.

The new PVCcosetX scheme reduces the bitstream by 29% (by 16 cosets) to 33% (by 2 cosets) and 8% (by 16 cosets) to 13% (by 2 cosets) on average (see Table 2) compared to the H.264 and the existing PVC which is significant especially at very low bit rate. Note that the PVC using more cosets performs better with bit rates as it accommodates more RMBs but the PVC with 4 cosets will provide reasonable performance for all bit rates. The PVC with 2 cosets will outperform in all cases if we can use better side information. A rate-distortion performance is also shown for Miss America video sequence using different cosets in Fig 8. The figure confirms that more than 0.25 dB PSNR is improved at 64 kbps using 2 cosets.

Table 2: Coding performance using PVCcoset, PVC, and H.264 algorithms

Video	PSNR	Bit Rate (kbps)					
	(dB)	PVCcoset			PVC	H.264	
		2	4	8	16		
Miss	36	14.5	15.5	16.5	17.0	19	33
America	40	26.5	27.5	27.5	28.5	31	43
	43	55.5	56.2	56.5	57.2	59	67
	44	69.8	70.2	70.6	71.2	73	79
Foreman	27	30.5	32.0	33.0	34.0	39	58
	30	44.0	45.0	46.0	47.0	51	67
	32	64.0	65.0	66.0	67.0	71	84
	34	80.5	81.0	81.2	82.0	86	98
Claire	35	15.2	16.2	17.2	18.2	21	35
	39	23.8	24.5	25.0	25.5	29	40
	41	37.0	37.5	37.8	38.0	42	51
	43	54.5	54.5	54.8	55.0	59	67
Salesman	31	22.0	23.0	24.0	26.0	29	47
	35	36.5	37.5	38.5	39.5	42	61
	37	53.4	53.5	54.5	55.0	58	75
	39	77.0	78.0	78.0	78.5	81	100

6. CONCLUSIONS

In this paper, angular distance and coset information are used to predict the pattern in decoder without sending the PIC. Using this concept, the PVC can accommodate more number of patterns for better approximation of moving region within an MB without increasing the PIC size. As a result, the ratedistortion performance is improved compared to the existing methods as well as the H.264 video coding standard. This new scheme will open another window of syndrome coding applications.



Fig 8: Rate-distortion performance using the H.264, existing PVC, and proposed PVCcosetX {2,4,8,16} for Miss America video sequence.

7. REFERENCES

[1] ITU-T Rec. H.264/ISO/IEC 14496-10 AVC. Joint Video Team (JVT) of ISO MPEG and ITU-T VCEG, JVT-G050, 2003.

[2] Fukuhara T., K. Asai, and T. Murakami, "Very low bit-rate video coding with block partitioning and adaptive selection of two timedifferential frame memories," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 7, 212–220, 1997.

[3] Wong K.-W., K. –M. Lam, and W. –C. Siu, "An efficient low bitrate video-coding algorithm focusing on moving regions," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11(10), 1128–1134, 2001.

[4] Paul M., M. Murshed, and L. Dooley, "A real-time pattern selection algorithm for very low bit-rate video coding using relevance and similarity metrics," *IEEE Trans. Circuits and System on Video Technology*, vol. 15(6), 753–761, 2005.

[5] Paul M. and M. Murshed, "An Optimal Content-based Pattern Generation algorithm," IEEE Signal Processing Letters, vol. 14(12), pp. 904-907, 2007.

[6] Aron A. and B. Girod, "Towards practical wyner ziv coding of video," Proc. of IEEE Int. Con. on Image Proc, vol. 3, 869–872, 2003.

[7] Chien W. J., L. J. Karam, and G. P. Abousleman, "Distributed video coding with lossy side information," Proc. of IEEE Int. Conference on Acoustics, Speech, and Signal Proc. vol. 2, 69–72, 2006.

[8] Girod B., A. Aaron, S. Rane, and D. Robello-Monedero, "Distributed video coding," Proc. IEEE, Special Issue on Advances in Video Coding and Delivery, 1–11, 2004.

[9] Puri R., and K. Ramchandran, "PRISM: a video coding architecture based on distributed compression principles," Proc. Allerton Conf. on Compression, Control and Computing, 2002.

[10] Slepian D., and J. k. Wolf, "Noiseless coding of correlated information sources," IEEE Trans. On Information Theory, vol. 19, 471–480, 1973.

[11] Wyner A., and J. Ziv, "The rate-distortion function for source coding with slide information at the decoder," IEEE Trans. On Information Theory, vol. 22, 1–10, 2003.Maragos, P., "Tutorial on advances in morphological image processing and analysis," *Opt. Eng.*, vol. 26(7), 623–632, 1987.

[12] Richardson, I. E. G., "H.264 and MPEG-4 video compression," *Wiley Press*, 2003.