

Video Coding Focusing on Block Partitioning and Occlusion

Manoranjan Paul, *Member, IEEE*, and Manzur Murshed, *Member, IEEE*

Abstract—Among the existing block partitioning schemes, the pattern-based video coding (PVC) has already established its superiority at low bit-rate. Its innovative segmentation process with regular-shaped pattern templates is very fast as it avoids handling the exact shape of the moving objects. It also judiciously encodes the pattern-uncovered background segments capturing high level of interblock temporal redundancy without any motion compensation, which is favoured by the rate-distortion optimizer at low bit-rates. The existing PVC technique, however, uses a number of content-sensitive thresholds and thus setting them to any predefined values risks ignoring some of the macroblocks that would otherwise be encoded with patterns. Furthermore, occluded background can potentially degrade the performance of this technique. In this paper, a robust PVC scheme is proposed by removing all the content-sensitive thresholds, introducing a new similarity metric, considering multiple top-ranked patterns by the rate-distortion optimizer, and refining the Lagrangian multiplier of the H.264 standard for efficient embedding. A novel pattern-based residual encoding approach is also integrated to address the occlusion issue. Once embedded into the H.264 *Baseline* profile, the proposed PVC scheme improves the image quality perceptually significantly by at least 0.5 dB in low bit-rate video coding applications. A similar trend is observed for moderate to high bit-rate applications when the proposed scheme replaces the bi-directional predictive mode in the H.264 *High* profile.

Index Terms—Block partitioning, H.264, motion estimation, occlusion, pattern-based coding, rate-distortion optimisation, sub-blocking, uncovered background coding, video coding.

I. INTRODUCTION

VIDEO compression standards such as H.263 [1] and MPEG-2 [2] are inefficient while coding at low bit-rate due to their inability to exploit intrablock temporal redundancy (ITR). Fig. 1 shows that objects can partly cover a block, leaving highly redundant information in successive frames as background is almost static in co-located blocks. Inability to exploit ITR results in the entire 16×16 -pixel macroblock (MB) being coded with motion estimation (ME) and motion compensation (MC) regardless of whether there are moving objects in the MB.

The recent H.264/AVC [3] video coding standard has extended the block-based coding paradigm by introducing tree-

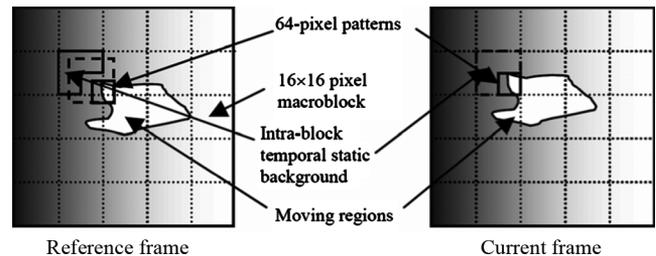


Fig. 1. Example on how pattern-based coding can exploit the intrablock temporal redundancy (ITR) to improve coding efficiency at low bit-rate.

structured variable block-size (TVBS) ME&MC to approximate the various motions within the MB more accurately by partitioning the 16×16 -pixel MB gradually to rectangular/square sub-blocks up to 4×4 pixels. We empirically observed in [4] that while coding head-and-shoulder type video sequences at low bit-rate, more than 70% MBs were never partitioned by the H.264 that would otherwise be at very high bit-rate. It can be easily observed that the possibility of choosing smaller block sizes diminishes as the target bit-rate is lowered. Consequently, coding efficiency improvement due to TVBS can no longer be realized for a low bit-rate target as larger blocks have to be chosen in most cases to keep the bit-rate in check but at the expense of inferior shape approximation.

Recently, many researchers [5]–[11] successfully introduced other forms of block partitioning to approximate the shape of a moving region (MR) even more closely to improve the compression efficiency (see Section II for details). But none of these techniques, including the H.264 standard, allows for encoding a block-partitioned segment by skipping ME&MC. Consequently they use unnecessary bits to encode almost zero-length motion vector with perceptually insignificant residual errors for the background segment. These bits are quite valuable at low bit-rate that could otherwise be spent wisely for encoding residual errors in perceptually significant segments.

These block partitioning techniques effectively divide a MB into two disjoint segments that are encoded with independent ME&MC. This is a significant improvement compared to the TVBS in the H.264 standard, which could use as many as 16 disjoint segments each with independent ME&MC. However, they are not suitable for low bit-rate video coding for the following reasons: (i) the penalty of extra bits to encode additional motion vectors and corresponding residual errors outweighs the marginal picture quality benefit at low bit-rate coding, especially when only one of the segment covers part of a moving object and the other segment covers almost static background with high ITR; and (ii) the computational complexity overhead for segmentation is also unjustified for low bit-rate video coding

Manuscript received August 10, 2008; revised August 28, 2009. First published September 29, 2009; current version published February 18, 2010. This work was supported in part by the Australian Research Council under Discovery Projects Grant DP0666456. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Antonio Ortega.

The authors are with the Gippsland School of Information Technology, Monash University, Churchill, Vic 3842, Australia (e-mail: manoranjan.paul@infotech.monash.edu.au; manzur@infotech.monash.edu.au).

Digital Object Identifier 10.1109/TIP.2009.2033406

applications, e.g., video conferencing with mobile phones. Note that the H.264 standard acknowledges the penalty of extra bits used by the motion vectors by imposing rate-distortion optimisation in motion search to keep the length of the motion vector smaller and disallowing B-frames, requiring two motion vectors for B macroblock, in the *Baseline* profile used widely in video conferencing and mobile applications.

The MPEG-4 [12] video standard exploits ITR at frame level by dividing the video frames into separate segments, comprising a background and one or more moving objects. Paradoxically, it also depends on computationally expensive segmentation and shape coding.

Wong *et al.* [13] first exploited ITR at block level and Paul *et al.* [14] then extended the idea of partitioning the MBs via a simplified segmentation process that again avoided handling the exact shape of the moving objects, so that popular MB-based motion estimation techniques could be applied. This pattern-based video coding (PVC) algorithm focuses on the MRs of the MBs, through the use of a set of regular 64-pixel pattern templates [see Fig. 2(a)]. The pattern templates were designed using 1s in 64 pixel positions and 0s in the remaining 192 pixel positions in a 16×16 -pixel MB. The MR of a MB is defined as a region comprising a collection of pixel positions where pixel intensity differs from its reference MB. Using some similarity measures, if the MR of a MB is found well covered by a particular pattern, then the MB can be classified as a region-active MB (RMB) and coded by considering only the 64 pixels of the pattern, with the remaining 192 pixels being skipped as static background. Embedding PVC in the H.264 standard as an extra mode provides higher compression for RMBs as the larger segment with high ITR is coded with no motion vector and residual error. The size of the patterns, 64 pixels, is carefully selected by empirical studies, which revealed that rate-distortion gain diminishes with larger patterns as the segment size with high ITR decreases. The MRs and their approximation using pattern templates are shown in Fig. 2(b) and (c), respectively, between frames one and two of the *Miss America* sequence.

The existing PVC schemes use three thresholds to (i) generate binary MRs; (ii) identify suitable MBs for pattern-mode coding based on the size of their MRs; and (iii) determine whether the best-matched pattern template sufficiently covers the MR of a suitable MB. The pixel intensity difference threshold T_I is used to generate the binary MR of a MB from the current and reference frame. A pixel position in the current frame is declared a moving point, if the intensity difference from its co-located pixel in the reference frame is more than T_I . All the moving points in a MB is collectively defined as its MR. The problem with this definition is that it cannot differentiate low to high intensity changes. When the best-matched similar pattern does not cover the entire MR of a MB, there is a possibility to include low intensity-changing moving points for motion compensation while leaving behind high intensity-changing points, resulting in higher perceptual errors in the background. Thus, treating all moving points equally degrades the performance. The moving region size threshold T_S is used to eliminate the MBs unsuitable for pattern representation at an early stage. The pattern similarity threshold T_P is then used to further eliminate some of the

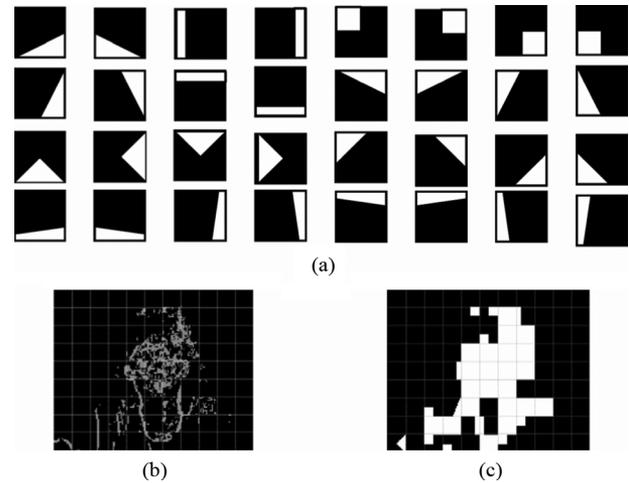


Fig. 2. Pattern templates and their effectiveness in approximating the moving regions. (a) The pattern codebook of 32 regular shaped patterns where the white and black segments are expected to cover moving regions and almost static background, respectively. (b) Moving regions between frames one and two of the *Miss America* sequence. (c) Corresponding pattern approximations.

suitable MBs whose dissimilarity with the best pattern is above this threshold. This was done under the assumption that the Lagrangian optimization function would have always disfavoured such a high distortion.

Although these thresholds were set after performing sensitivity analysis on a large number of standard and nonstandard video sequences, they impinge on extracting the best achievable rate-distortion performance. As such, we have recently observed that the suitability/unsuitability measures of the MBs have no direct bearing on the ultimate rate-distortion performance due to the diversified content of the video sequences. Obviously, the primary purpose of the above mentioned three thresholds was to reduce the number of MBs coded using pattern to control the computational complexity. Testing each MB for coding with the pattern mode requires extra motion estimation, which is the most computationally expensive part [15] in a coding scheme. Moreover, using fixed generalised thresholds for diversified video sequences will always have the possibility to overlook some potential RMBs. Clearly, eliminating the thresholds altogether by allowing all the MBs to be tested for the pattern mode and relying solely on the mode selection decision of the Lagrangian optimization function will provide a significantly improved rate-distortion performance over a wide range of motion activities. If the pattern motion is first predicted from the block motion and then refined within a much smaller search window, the computational complexity overhead of testing all MBs for the pattern mode can be easily kept in check.

Reusing the block motion vector for the pattern mode can potentially overcome another problem with the existing PVC technique related to occlusion (in fact this problem exists with any generic block matching coding algorithm) where the motion search for the MR covered by the best matched pattern template cannot find the true motion vector when the MR erroneously represents the background occupied by the moving object in the reference frame. Moreover, the 64-pixel patterns

cannot always cover arbitrary-shaped MRs completely, and, thus, there is always a possibility that the motion vector of the pattern may not be identical to that of the MR. To overcome this problem, instead of partitioning the block before the ME stage, we can perform the partitioning after the ME stage and immediately before the MC stage as follows. We need to find the pattern, which covers the maximum cumulative residual errors and then encode only this subset of the residual errors, leaving the remaining errors un-coded. Obviously, we degrade the image quality by ignoring some residual errors, but it is also the case when a conventional coding approach uses quantization. By considering the most important 64 residual errors in the block, the proposed approach would be able to use lesser quantization on those important residual errors for the same compression and, thus, achieve better rate-distortion performance, especially at low bit-rate video coding. By allowing two separate pattern modes, one for pre-ME partitioning and the other for post-ME partitioning, we can fairly assume that for nonocclusion situations, the MR would be encoded using former mode and for occlusion situations, the residual errors would be encoded using the latter mode.

In this paper, a novel threshold-free PVC scheme is proposed, which selects the final mode using solely the Lagrangian optimization function without considering the content of the video sequences. We have eliminated the pixel intensity difference threshold using a new similarity measure. It is important to note that, as the best pattern selection process relies on solely the similarity measures, it is not guaranteed that the best pattern will always result in the maximum compression, which also depends on the number, magnitude, and the orientation of the residual errors and the length of the pattern identification numbers. This paper thus introduces additional pattern modes, each selecting the pattern in order of the similarity ranking. We have also successfully handled occlusion specific problems by introducing pattern-based residual error coding. In addition, a new Lagrangian Multiplier (LM) is also developed as quite a number of additional modes are considered during encoding along with the H.264 *Baseline* profile. The experimental results confirm that this new scheme improves the rate-distortion performance significantly compared to the existing PVC technique and the H.264 *Baseline* profile, recommended specifically for some low-delay, low-complexity coding applications.

This paper is organized as follows. Section II provides the recent advancements on block partitioning. Section III illustrates the existing PVC technique. Section IV describes the proposed method, including occlusion problem handling, new similarity metric, multiple pattern modes, and the novel pattern-based residual error coding. Section V analyzes the detailed computational complexity. Section VI presents the experimental set up along with the comparative performance results. Section VII concludes the paper.

II. RELEVANT WORKS

Chen *et al.* [10] extended TVBS to include additional four partitioning with one L-shaped and one square segment to achieve some improvement in the picture quality. One of the limitations of segmenting MBs with 4×4 -pixel building blocks

as done in TVBS and [10] is that the partitioning boundaries cannot always approximate arbitrary shapes of moving objects efficiently.

Hung *et al.* [5] and Divorra *et al.* [6], [7] independently addressed this limitation of TVBS by introducing additional wedge-like partitioning where a MB is segmented using a straight line modelled by two parameters: orientation angle θ and distance ρ from the centre of the MB. The former used fixed-length codes to identify wedge-like partitions and also considered a limited number of partitions to improve computational complexity without any significant impact to picture quality. The latter improved compression efficiency further by encoding each of the parameters independently using variable-length predictive codes and setting the resolution of sampling for each parameter depending on the general quantization parameter (QP). A very limiting case with only four partitions ($\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$, and $\rho = 0$) was reported by Fukuhara *et al.* [11] even before the introduction of TVBS for very low bit-rate video coding.

Chen *et al.* [8] and Kim *et al.* [9] improved compression efficiency further with implicit block segmentation (IBS) and thus avoiding explicit encoding of the segmentation information. In both cases the segmentation of the current MB can be generated by the encoder and decoder using previously coded frames only. The former segmented the current MB based on the pixel illumination values and then encoded an extra motion vector, besides the two motion vectors for the segments, representing the closest matching segmentation in the reference frame. The latter used two motion vectors to identify two predicted blocks in the reference frame, segmented their difference based on the pixel difference values into multiple regions, and then interpolated each region in the current block from the two predicted blocks. The predicted blocks are searched in the reference frame with joint ME to find the optimised rate-distortion.

All these techniques effectively partition a MB into two disjoint segments (could be more than two in [9]) that are encoded with independent ME&MC. However, we argue that they are not suitable for low bit-rate video coding as the precious extra bits used for encoding the segment covering almost static background and the excessively high computational complexity of the segmentation process are not justified for low bit-rate applications.

To verify this argument, we have extended TVBS of the H.264 standard with additional block partitioning techniques to find their respective mode selection rates on a number of standard video sequences. Fig. 3 presents results on the *Foreman* QCIF 15 Hz sequence for three different extended schemes. In scheme i), PVC (threshold free) and TVBS modes were available for rate-distortion optimisation. At very low bit-rate, the PVC mode was selected predominantly. As the bit-rate was increased the TVBS mode achieved higher dominance. In the scheme (ii), IBS[9]¹ was introduced as an additional mode to scheme i). While the IBS mode was increasingly preferred to the TVBS mode as the bit-rate is increased, the PVC mode selection rate was unchanged from the first scheme. This means

¹This block partitioning scheme is selected for the comparison as it has reported the best compression efficiency among all the recent works on block partitioning.

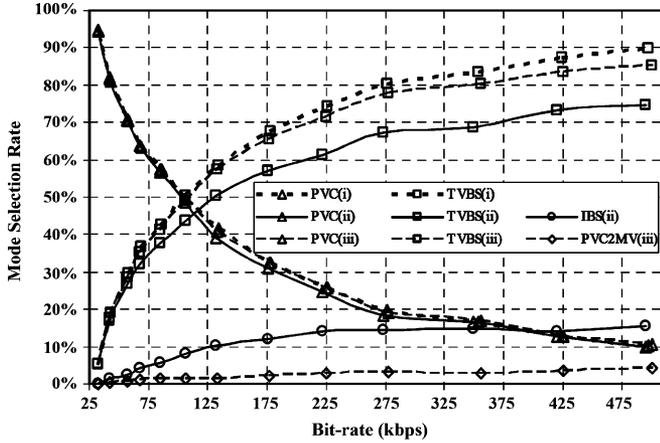


Fig. 3. Mode selection trends of (i) PVC and TVBS; (ii) PVC, TVBS, and IBS [9]; and (iii) PVC, TVBS, and modified PVC with two MVs block partitioning techniques on the *Foreman* QCIF 15 Hz video sequence.

that the IBS technique is unable to exploit the ITR of the static background at low bit-rate coding.

In scheme (iii), IBS was replaced from scheme (ii) with a modified PVC technique where the background segment was also coded with ME&MC. The PVC with two MVs (PVC2MV) mode shows a trend similar to IBS in scheme (ii) but at a significantly reduced level. In addition, we have observed that the average length of the pattern identification number is at least one bit less than the average bits required to encode the second motion vector representing the motion of the background segment for the low bit-rate range, up to 128 and 512 kbps for the QCIF and CIF formats, respectively. We can thus conclude that an implicit pattern selection technique based on a similar approach used in [9] to avoid encoding the pattern identification number will not achieve any coding efficiency. In fact, it can be concluded in general that the efficiency of coding the high-ITR background segments does not improve significantly with motion compensation.

III. THE EXISTING PATTERN-BASED VIDEO CODING

For the sake of completeness, this section briefly describes the existing pattern-based video coding (ePVC). The predefined 32 patterns in the pattern codebook (PC), shown in Fig. 2(a), approximate the MR in a MB with the white region representing the MR and the black regions representing the static background. Let $C_k(x, y)$ and $R_k(x, y)$ denote the k th MB of the current and reference frames of a video sequence, respectively, with frame size W pixels \times H lines where $0 \leq x, y \leq 15$ and $0 \leq k < W/16 \times H/16$. The MR $M_k(x, y)$ of the k th MB in the current frame is obtained [13] as follows [see

(1), shown at the bottom of the page] where B is a 3×3 unit matrix for the noise-reducing morphological closing operation \bullet and the threshold T_I is 2. Here, the value 1 or 0 indicates the corresponding pixel position belongs to the MR or the static region of the MB, respectively.

The dissimilarity of a pattern $P_n \in PC$ with the MR in the k th MB is defined as

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

where $|M_k|_1$ denotes the total number of 1s in M_k . Clearly, higher the similarity lower will be the value of $D_{k,n}$.

To avoid the pattern-based ME for all MBs, the set of eligible MBs, called *candidate* RMBs (CRMBs), are separated such that $8 \leq |M_k|_1 < T_S$. A CRMB is classified as an RMB and its moving region is represented by the i th pattern P_i such that

$$P_i = \arg \min_{P_n \in PC} ((D_{k,n}) | D_{k,n} < T_P). \quad (3)$$

Otherwise, the CRMB is classified as a normal MB, which is not coded with the pattern mode. To avoid more than four 4×4 blocks of discrete cosine transformation (DCT) calculations for the 64 residual error values per RMB, these values are rearranged into an 8×8 block. It avoids unnecessary DCT block transmission. A similar inverse procedure is performed during the decoding. After finding the bit rates and the corresponding distortion for all possible modes, the final decision is made based on the minimum Lagrangian cost function [16]. The existing PVC classifies 10% to 30% of the MBs as RMBs and improves the image quality by at least 0.2 dB compared to the H.264 *Baseline* profile at low bit-rate.

IV. THE PROPOSED NEW PATTERN-BASED VIDEO CODING

As mentioned earlier, the purpose of the three thresholds used in the ePVC was to control the computational complexity by reducing the number of MBs tested for possible coding with the pattern mode. Eliminating these thresholds and thus testing all the MBs for possible coding with the proposed new pattern-based coding (nPVC) can potentially improve the performance; but it demands resolving the following issues: (i) developing a new similarity metric, which does not require any binary MR; (ii) investigating the validity of the existing Lagrangian multiplier in this case; and (iii) controlling the computational complexity in real time. As the selection of the best pattern based on the similarity does not necessarily guarantee the maximum compression for all the cases, we also propose multiple pattern modes that select the pattern in order of the similarity ranking. Before discussing detailed solutions to these issues, we revisit the occlusion related problems with the ePVC, explained briefly in the introduction, in the next section.

$$M_k(x, y) = \begin{cases} 1, & \text{if } |(C_k(x, y) \bullet B) - (R_k(x, y) \bullet B)| > T_I \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

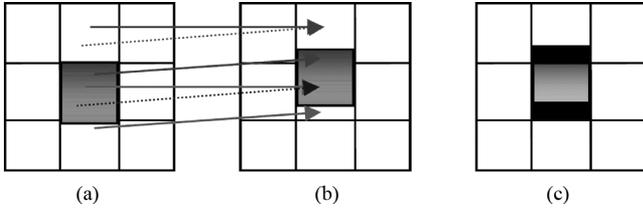


Fig. 4. Pattern matching by detecting the moving region. (a) An object in the reference frame; (b) the object in the current frame; and (c) moving region detection based on the intensity difference.

A. Occlusion Problems in the ePVC

The stationary background in the occluded region can be erroneously classified as a moving object by any block-based motion estimation when a part of the background is only visible in the current frame but occluded in the reference frame. Such faulty detections of the MRs result in incorrect motion vectors and eventually degrade the rate-distortion performance. Consider Fig. 4 where an object was in the middle of the reference frame (a) and then moved slightly upwards in the current frame (b). Fig. 4(c) shows the MR based on the intensity difference between the co-located pixels using the grayscale where pure white represents no difference, pure black represents the maximum possible difference, and gray shades represent differences in between. Note that the gray area in the MR is due to assuming the object not perfectly homogeneous. In the ePVC using the pixel intensity difference threshold T_I , only the black regions are considered as the MRs for the corresponding two MBs in Fig. 4(c) that will be referred to as the top and middle MB. It is perfectly alright for the top MB as this region would be covered by the best pattern as a portion of an object and ME&MC would be carried out based on this area (indicated by the solid upward arrow) and the rest of the MB would be copied from the co-located reference frame (indicated by the solid horizontal arrow).

However, for the middle MB, a part of the background in the current frame would be covered by the best pattern as a moving object (ignoring the lesser differences due to the non-homogeneity of the object) and ME&MC would be carried out based on this portion. The actual matching would eventually fail as the reference part for this MR is occluded in the reference frame. Due to the background homogeneity, this portion could be motion estimated from a different part of the background (indicated by the solid upward arrow), and, hence, the rest of the MB, which is a portion of the moving object erroneously assumed static, would now be copied from a different portion of the object in the co-located pixels of the reference frame (indicated by the solid horizontal arrow). This erroneous ME&MC process would definitely degrade the video quality as well as increase the bits for residual errors.

If we use the whole MB instead of pattern-covered moving region in the ME, a suitable motion vector could be estimated where the residual errors would be observed in the boundary-adjacent part (here the bottom part of the MB) of the background portion. If we then use pattern-matching on the residual errors after whole-block motion estimation to cover the maximum cumulative errors, encoding only the residual errors covered by the best pattern would save the bits significantly. This technique

effectively copies the pattern-uncovered region from the translated area, rightly assuming they are also moving according to the pattern-covered regions (indicated by the dashed upward arrows). Thus, pattern-based residual coding (PRC), detailed in Section D, would solve the occlusion problem observed in the middle MB of the example in Fig. 4. Moreover, after motion estimation (for nonocclusion cases), the residual errors may be found in a small region and for those cases the PRC would also provide better rate-distortion performance.

The PRC technique, however, is not suitable for cases like the example in the top MB in Fig. 4. The assumption that the pattern-uncovered region can be copied from the block-translated area in the reference frame is no longer valid when the pattern correctly approximates the MR and the background is more or less static between the frames. This problem can be easily resolved by allowing two separate pattern modes, one for pre-ME partitioning (similar to ePVC) and the other for post-ME partitioning (PRC). As the ultimate mode selection will be taken by the Lagrangian optimization function, we can fairly assume that for nonocclusion situations, the MR would be encoded using former mode and for occlusion situations, the residual errors would be encoded using the latter mode.

Using two separate pattern-mode codes in the encoded stream could be expensive in terms of extra bits needed for low bit-rate coding. If the pattern-uncovered regions for nonocclusion situations are copied from the translated area guided by the pattern motion, we can use only one pattern-mode code in the encoded stream as the decoding of both the pattern modes will be alike. Considering that the background is highly homogenous at low bit-rate, the subtle error in the misalignment of the background segment with high ITR will be offset by the bits saved for using only one extra mode code.

B. New Similarity Metric

The main purpose of the similarity metric is to find the best-match pattern for a given MR. The existing similarity metric selects the best pattern based on the minimum dissimilarity measured as the nonoverlapping region between the binary MR and the pattern using (2). When the best-matched similar pattern does not cover the whole MR of a MB, there is a possibility to exclude some high intensity difference moving points instead of low, which would be inefficient as alluded in the introduction. The proposed similarity metric does not require any pixel intensity difference threshold T_I , as it does not need any binary MR for selecting the best pattern. It determines the best pattern from the absolute error matrix between the current and co-located reference MBs calculated as

$$E_k(x, y) = |C_k(x, y) - R_k(x, y)|. \quad (4)$$

The similarity of a pattern $P_n \in PC$ with the MR in the k th MB is defined as

$$S_{k,n}(x, y) = \sum_{x=0}^{15} \sum_{y=0}^{15} |E_k(x, y) \times P_n(x, y)|. \quad (5)$$

The best pattern for an MR is then selected as

$$P_i = \arg \max_{P_i \in PC} (S_{k,n}). \quad (6)$$

0 3	0 1	0 0	0 0	0 1	1 0
2 8	0 1	1 0	0 1	0 0	0 0
(a)	(b)	(c)	(d)	(e)	(f)

Fig. 5. Comparative example of the new similarity metric against the existing one using 2×2 -pixel block. (a) The error matrix E_k . (b) The binary MR M_k . (c)–(f) Four 1-pixel patterns.

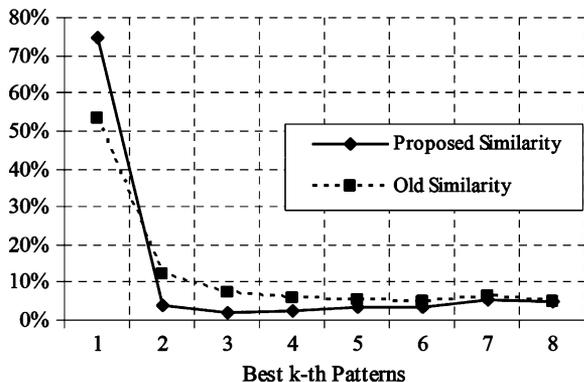


Fig. 6. Percentages of the RBMs finally selected by the k th best similar patterns using the proposed and old similarity metrics for the *Miss America* video sequence where the quantization parameter $QP = 30$ is used.

Another advantage of this new similarity metric over the existing one is that it considers the most dominating moving pixels where the pixel intensity differences are higher. Obviously, the same pattern will be selected in both cases where the MR of a MB is fully covered by the best pattern; otherwise different patterns may be selected by the two metrics. An example is given in Fig. 5 where an error matrix and a binary matrix are formed using (4) and (1) respectively assuming $T_I = 2$. The pattern in Fig. 5(d) or (e) will be selected using the old similarity metric; whereas only the pattern in Fig. 5(d) will be selected by the new similarity metric. Obviously, the pattern in Fig. 5(d) is the best choice for residual coding as it covers the most dominating pixel when the pattern motion is approximated.

C. Multiple Pattern Modes

As alluded in the introductory section, the best pattern selection process relying on the similarity measures does not guarantee that the best pattern will always result in the maximum compression. This is observed more frequently at higher bit-rates where the Lagrangian optimizer emphasises quality against bit-rates. This may also be observed at low bit-rate when motion vectors are nonzero. To eliminate this possibility, we propose multiple pattern modes that select the pattern in order of the similarity ranking. Since the similarity measure is still a good estimator, we only consider patterns ranked at the top by the similarity metric. Obviously, it will increase the computational complexity, which can be checked by using only a few top ranked patterns.

Fig. 6 shows the percentages of RBMs finally selected by the k th best similar patterns using the two different similarity matrices. From the figure it is clear that in both cases, the best similar pattern captures most of the RBMs. It is also clear that the best pattern captures 75% and 54% of RBMs with the proposed and the old similarity metric, respectively. As the best pattern using the new similarity metric still does not cover 25% of the

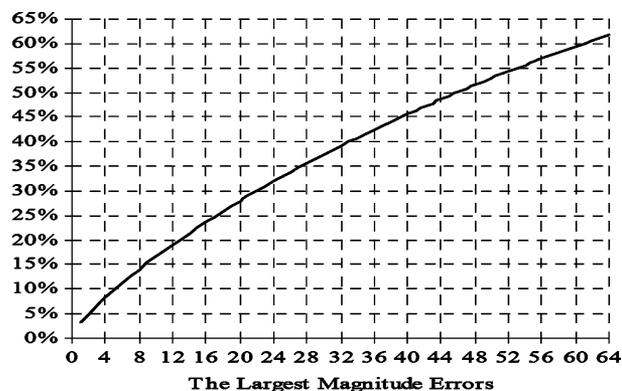


Fig. 7. Average percentage of total errors covered by the first k largest errors after motion estimation using the 16×16 block mode.

RMBs as the best alternative, using multiple pattern modes is justified. Note that no additional pattern-mode code is needed for the encoded stream as the decoding is indifferent to what pattern mode was eventually selected by the Lagrangian optimization function at the encoder.

D. Pattern-Based Residual Coding (PRC)

We have observed empirically that the residual errors are mostly concentrated in the boundary-adjointed regions after full block (16×16 pixels) motion estimation for MBs having both moving regions and static regions. Fig. 7 shows that the largest 64 errors cover on average 62% of total errors after full block motion estimation using all types of MBs. In this experiment, we have used the average values for all MBs using a number of standard QCIF video sequences. For obvious reason, the coverage percentage is less for smooth video sequences such as *Claire* and *Miss America*, and more for high motion video sequences such as *Foreman* and *Garden*. The coverage percentage is significantly higher for those MBs that are classified as RMBs by the PVC approach. This observation has motivated us to define a new similarity metric based on the pattern covered errors and to use PRC after block motion estimation.

Motion vector is calculated by the ePVC algorithm using pattern covered moving region (i.e., 64 pixels) instead of the whole 16×16 -pixel MB to reduce the computational time. But from the experimental results we have observed that the latter provides better results where the mismatch area between the pattern and the MR of a MB is relatively large. It is also true when occlusion has caused erroneous detection of MRs. With PRC we can easily use the motion vectors, generated by the 16×16 -pixel blocks. Thus, no extra computational time is needed for motion estimation. In the PRC algorithm we use the residual error between the current block and the motion estimated reference block, instead of the co-located reference block used by the ePVC algorithm. Then we find the best-matched pattern from the PC using the new similarity metric.

E. New Lagrangian Multiplier With Pattern Modes in H.264

To select the best mode of variable block sizes motion estimation, the Lagrangian multiplier $\lambda = 0.85 \times 2^{(QP-12)/3}$ [16] is used in the H.264 standard where QP is the quantization parameter used. We are interested to find whether this multiplier needs

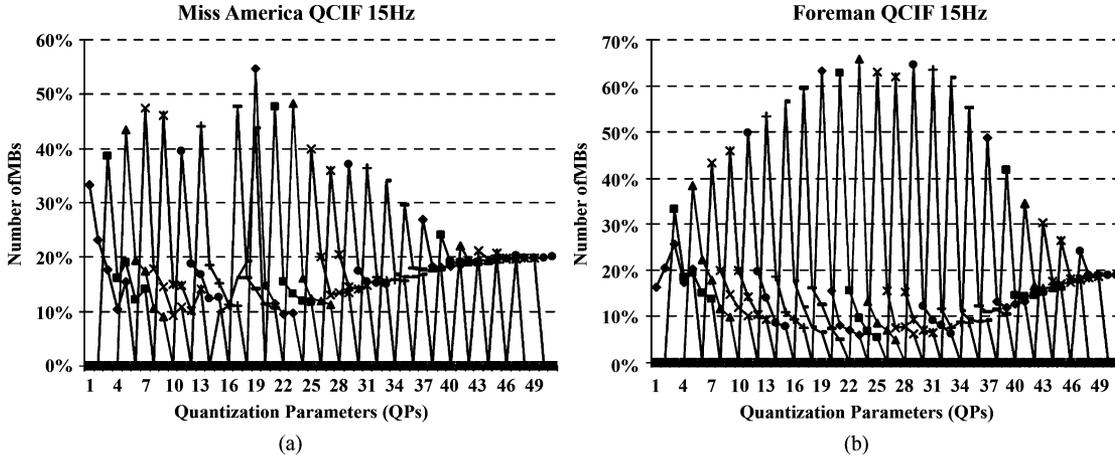


Fig. 8. Percentage of MBs selected by the Lagrangian optimization function against the five QP s used for a given set of λ while encoding the (a) *Miss America* and (b) *Foreman* sequences. Results for the same λ are presented in the same color.

any adjustment due to the inclusion of various pattern modes. To calculate the relationship of the Lagrangian multiplier with the QP exhaustively from the scratch, for a wide range of λ , we need to find the percentage of MBs classified using all possible quantization parameters QP s for each λ and then calculate the expected QP using weighted averaging on the percentage of MBs classified. In this way, for a particular λ we need to encode a video sequence 52 times as H.264 considers 52 QP values. To find the adjustment to the existing relationship in H.264, we can expect to speed up the process by considering only a subset of QP s around the one currently suggested by the existing inverse relationship. In fact a similar speed up technique was used to adjust the relationship from H.263 to H.26L [17]. We have used the subset of the five closest QP s, $QP - 2$, $QP - 1$, QP , $QP + 1$, and $QP + 2$ for each λ where QP is the quantization parameter obtained for that λ from the H.264 relationship. For example, if H.264 suggests $QP = 30$ for a given λ , and 3%, 5%, 15%, 42%, and 35% of MBs are selected by the Lagrangian optimization function against the five QP s 28, 29, 30, 31, and 32, respectively, then the expected QP is calculated as $3\% \times 28 + 5\% \times 29 + 15\% \times 30 + 42\% \times 31 + 35\% \times 32 = 31$.

Fig. 8 presents the percentage of MBs selected against the five QP s for a range of λ . For most of these λ s, the smallest QP is selected for the maximum number of MBs, which justifies adjusting the existing LM relationship of H.264 to $\lambda = 0.4 \times 2^{(QP-12)/3}$ due to the embedding of the pattern modes. As the pattern modes require fewer bits compared to the 16×16 mode, the reduced λ suggested by the proposed relationship signifies less importance in bits compared to the distortion in the minimization of Lagrangian cost function, which is consistent to maintain the same quality. We have also observed that for a given λ generated, QP is slightly large for relatively high motion video sequences compared to the smooth motion video sequences. However, this variation is not enough to introduce separate λ for various sequences.

V. COMPUTATIONAL COMPLEXITY

Let α , d , δ , and m be the number of MBs, motion search points used per MB, number of operations per search point (different operations are treated alike for the sake of simplicity),

and the average number of modes used by the H.264 *Baseline* profile for motion estimation, respectively. So, H.264 requires $\delta m d \alpha$ operations for motion estimation. It is often claimed by the researchers that motion estimation, irrespective of a scene's complexity, typically comprises around 60% of the processing overhead required to encode an inter picture with a software codec using the DCT [18], when full search is used. We may then argue that the H.264 *Baseline* profile ultimately requires

$$\frac{\delta m d \alpha}{60\%} = \frac{5}{3} \delta m d \alpha \text{ operations} \quad (7)$$

to complete the encoding.

Let $\tau \leq 1$ be the fraction of MBs classified as CRMBs that are further tested for pattern-based coding in the ePVC approach. Pattern motion search requires one-fourth time of block motion search as the pattern size is one-fourth of a MB. Thus, embedding the ePVC technique into H.264 requires extra $\delta d \tau \alpha / 4$ operations for motion estimation and, hence, altogether extra

$$\frac{5}{3} \frac{\delta d \tau \alpha}{4} = \frac{5}{12} \delta d \tau \alpha \text{ operations.} \quad (8)$$

So, the H.264 *Baseline* profile embedded with ePVC ultimately requires

$$\frac{5}{3} \delta m d \alpha + \frac{5}{12} \delta d \tau \alpha = \frac{5}{12} \delta d \alpha (4m + \tau) \text{ operations} \quad (9)$$

to complete the encoding.

To analyze the computational complexity of the proposed nPVC technique, we consider pre-ME (i.e., threshold-free ePVC) and post-ME (i.e., PRC) pattern matching separately. Let β_1 and β_2 be the number of pattern-modes used in pre-ME and post-ME pattern matching, respectively. By setting $\tau = 1$ in (8) as all the MBs are tested for pattern matching and multiplying with β_1 , we estimate that pre-ME pattern matching requires extra

$$\frac{5}{12} \delta d \alpha \beta_1 \text{ operations.} \quad (10)$$

Post-ME pattern matching does not require any motion estimation as it uses the already available 16×16 block mode motion

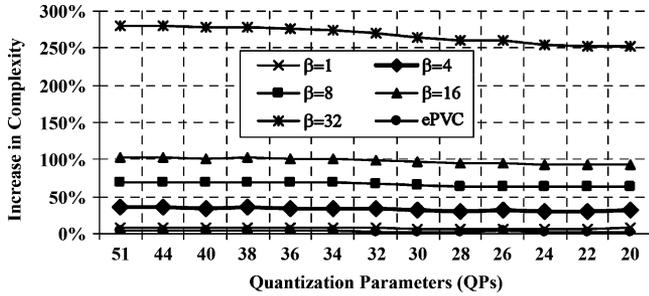


Fig. 9. Increase in computational complexity to embed the existing PVC (ePVC) and the proposed new PVC (nPVC) techniques using different number of pattern modes $\beta_1 = \beta_2 = \beta \in \{1, 4, 8, 16, 32\}$ into the H.264 *Baseline* profile for the *Miss America* video sequence.

vectors. We thus estimate that pre-ME pattern matching requires extra

$$\frac{5}{12} \delta d \alpha \beta_2 \times 40\% = \frac{1}{6} \delta d \alpha \beta_2 \text{ operations.} \quad (11)$$

So, the H.264 *Baseline* profile embedded with nPVC ultimately requires

$$\begin{aligned} \frac{5}{3} \delta m d \alpha + \frac{5}{12} \delta d \alpha \beta_1 + \frac{1}{6} \delta d \alpha \beta_2 \\ = \frac{5}{12} \delta d \alpha \left(4m + \beta_1 + \frac{2}{5} \beta_2 \right) \text{ operations} \quad (12) \end{aligned}$$

to complete the encoding.

Alternatively, we can conclude theoretically the ePVC and nPVC techniques require $0.25\tau/m$ and $(0.25\beta_1 + 0.1\beta_2)/m$ ($0.25\beta_1 + 0.1\beta_2$)/ m times more operations compared to the H.264 *Baseline* profile, respectively. Note that although we assume that the average number of *Baseline* profile modes m remains the same for the original H.264 and the two embedded schemes, in reality m is reduced significantly in the latter. It is due to the fact that while MBs are encoded using 8×8 or any smaller block mode in the H.264, the number of modes requiring ME&MC for each MB is seven. On the other hand, when a MB is encoded using a pattern mode, the number of modes requiring ME&MC would be four plus one-fourth of the number of pattern modes, e.g., if we use only four pattern modes, the average number of modes would be around five instead of seven. The experimental results on the *Miss America* sequence, presented in Fig. 9, confirm the validity of the theoretical analysis. For example, while $\tau = 0.4$, $\beta_1 = \beta_2 = \beta = 4$, and $m = 4.25$, theoretically ePVC and nPVC requires 2.4% and 35% more operations compared to H.264, respectively.

The rate-distortion performances by the nPVC algorithm using various number of pattern modes were studied for a number of standard video sequences. Fig. 10 presents the rate-distortion performances of the ePVC and nPVC embedding of H.264 for the *Miss America* video sequence. The image quality could be improved by 0.5 dB with 32 pattern-modes compared to the top-ranked four pattern-modes; but at the expense of three times more computational time (see Fig. 9) compared to H.264. As we observe the same for other video sequences, we recommend using $\beta = 4$ as a trade-off between the image quality and computational complexity.

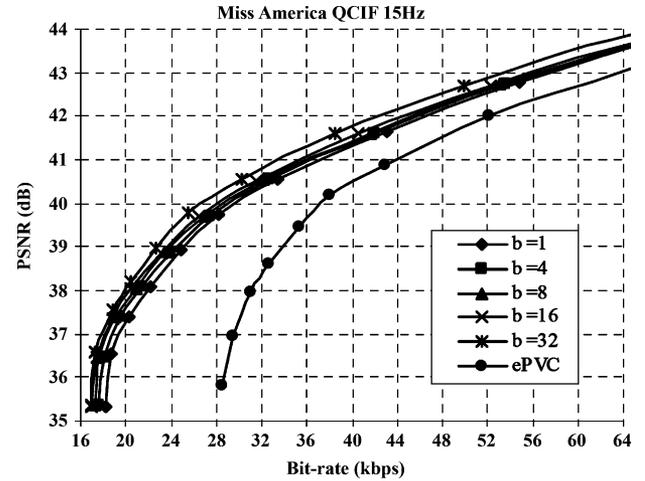


Fig. 10. Rate-distortion performance of embedding the existing PVC (ePVC) and the proposed new PVC (nPVC) techniques using different number of pattern modes $\beta_1 = \beta_2 = \beta \in \{1, 4, 8, 16, 32\}$ into the H.264 *Baseline* profile for the *Miss America* video sequence.

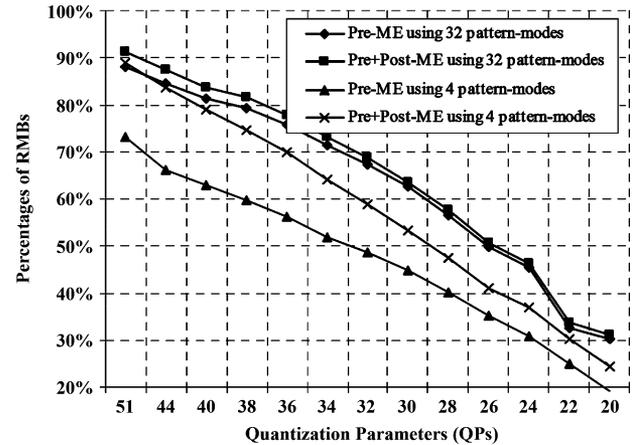


Fig. 11. Average percentage of MBs encoded as RBMs in the pre-ME stage and the combined pre-ME + post-ME stages with patterns by the nPVC algorithm using 32 and four pattern-modes.

Fig. 11 shows the average percentage of MBs encoded as RBMs in the pre-ME stage and the combined pre-ME + post-ME stages with patterns by the nPVC algorithm using 32 and four pattern-modes. Obviously, the number of RBMs decreases with QP in all cases as the smaller block-modes of H.264 are preferred by the Lagrangian optimization function to the pattern-modes for ensuring the quality of video at high bit-rate. The number of RBMs increases significantly for the combined stages compared to the pre-ME stage when four pattern-modes are used. This justifies the introduction of PRC in the proposed nPVC while using the recommended $\beta = 4$.

VI. EXPERIMENTAL RESULTS

We have implemented both the ePVC and nPVC algorithms embedded on the *Baseline* profile of the H.264/AVC standard (adapted from the JM 10.1 H.264/AVC reference software) with full-search motion estimation, group of picture (GOP) of 15 frames with the IPPP...P format, and up to quarter-pel accuracy for a number of standard and nonstandard QCIF and CIF

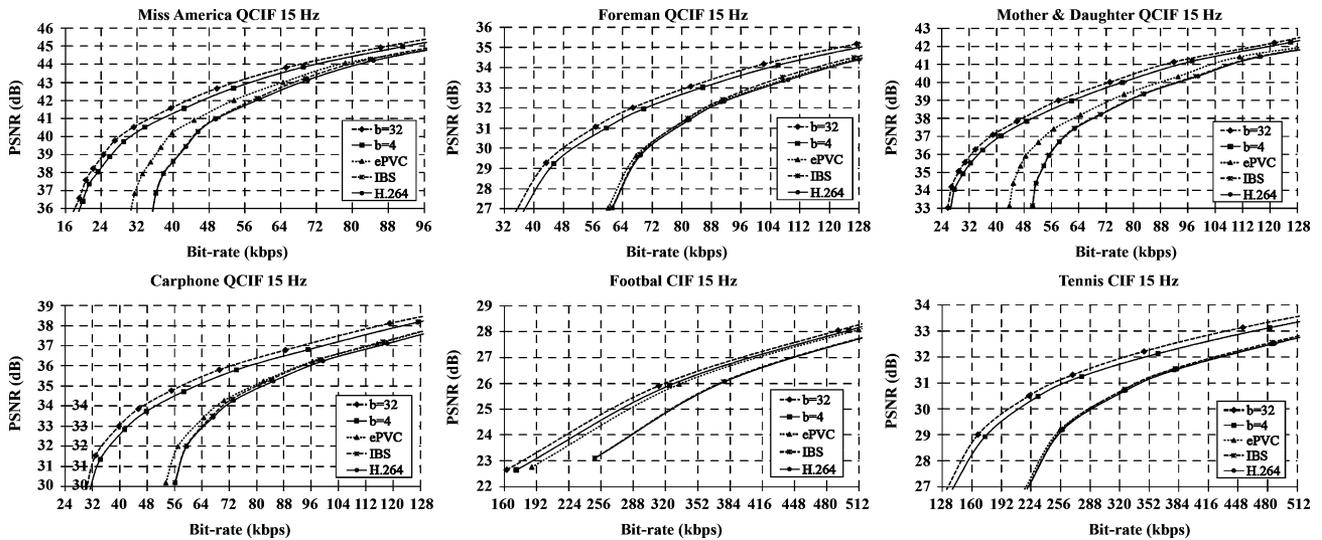


Fig. 12. Rate-distortion performance for six standard video sequences by the H.264 *Baseline* profile and its embedding with by the proposed new PVC (nPVC) technique using $\beta \in \{4, 32\}$ top-ranked pattern-modes in both pre-ME and post-ME pattern-matching, the existing PVC (ePVC) technique, and the implicit block segmentation (IBS) technique in [9].

[19] video sequences at 15 Hz, i.e., 15 frames per second (fps). The *Baseline* profile of H.264 is chosen as it is primarily designed for low-delay and low-complexity coding applications especially in videoconferencing and mobile applications. The proposed scheme aims to improve the rate-distortion performance for the same applications at low bit-rates, up to 128 and 512 kbps for the QCIF and CIF formats, respectively, with performance-constrained devices.

We have used the same PC in Fig. 2(a) for ePVC and nPVC implementations. In both cases, only one pattern-mode code is used in the encoded stream despite considering eight and 64 distinct pattern-modes by the Lagrangian optimization function in the nPVC with $\beta = 4$ and 32, respectively. This is due to the indifferent decoding of the pre-ME and post-ME pattern-modes, and the β top-ranked pattern-modes that have been explained in Sections IV-A and IV-C, respectively.

Fig. 12 presents the rate-distortion performance for six standard video sequences, namely *Miss America*, *Foreman*, *Mother & Daughter*, *Carphone*, *Football*, and *Tennis*, by the H.264 *Baseline* profile standard and its embedding with the proposed nPVC technique using $\beta \in \{4, 32\}$ top-ranked pattern-modes in both pre-ME and post-ME pattern-matching, the existing ePVC technique, and the implicit block segmentation (IBS) [9] embedding, which has reported the best compression efficiency among all the recent works on block partitioning. Of these six sequences, four were in the QCIF and the remaining two were in the CIF format. These frames sizes are mostly used in the low bit-rate applications such as YouTube and mobile video telephony. Clearly, the nPVC embedding achieved superior rate-distortion performance for all the test sequences.

While the performance curves in Fig. 12 shows the image quality achieved by different techniques for a wide range of bit-rates, we are particularly interested to compare the image quality of the nPVC embedding using $\beta = 4$ against the H.264 *Baseline* profile and the ePVC embedding. Let us consider the maximum and its half bit-rates for the low bit-rate range considered for the QCIF and CIF formats as the reference points without any loss

TABLE I
IMAGE QUALITY IMPROVEMENT BY THE PROPOSED NEW PVC (nPVC) EMBEDDING USING $\beta = 4$ AGAINST THE H.264 *BASELINE* PROFILE AND THE EXISTING PVC (ePVC) EMBEDDING AT THE MAXIMUM AND ITS HALF BIT-RATES CONSIDERED FOR THE QCIF AND CIF FORMATS

	H.264		ePVC	
	64 kbps	128 kbps	256 kbps	512 kbps
<i>Miss America</i>	0.8 dB	0.4 dB	0.5 dB	0.3 dB
<i>Foreman</i>	3.0 dB	2.2 dB	0.6 dB	0.5 dB
<i>Mother & Daughter</i>	1.7 dB	0.8 dB	0.6 dB	0.4 dB
<i>Carphone</i>	2.1 dB	1.5 dB	0.7 dB	0.5 dB
<i>Football</i>	1.3 dB	0.2 dB	0.5 dB	0.1 dB
<i>Tennis</i>	1.5 dB	0.3 dB	0.6 dB	0.1 dB

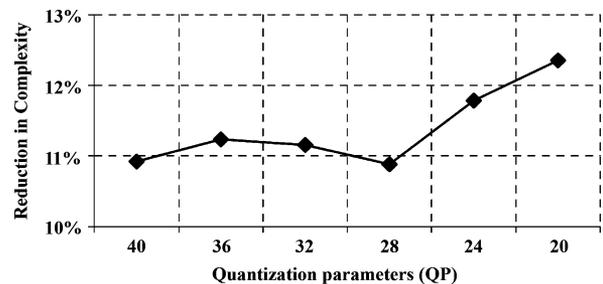


Fig. 13. Average computational complexity reduction when the proposed nPVC mode using $\beta = 4$ replaces the bi-directional prediction mode in the H.264 *High* profile for standard 4CIF video sequences.

of generality. Improvement in the image quality by the nPVC embedding at these reference points are presented in Table I. With the four QCIF sequences, the nPVC embedding improved image quality 0.8 to 3.0 dB at 64 kbps and 0.4 to 0.8 dB at 128 kbps compared to H.264; and 0.4 to 2.2 dB at 64 kbps and 0.3 to 0.5 dB at 128 kbps compared to the ePVC embedding. The improvement was relatively less for the two CIF sequences with high motion activities—1.3 to 1.5 dB at 256 kbps and 0.5 to 0.6 dB at 512 kbps compared to H.264; and 0.2 to 0.3 dB at 256 kbps and 0.1 dB at 512 kbps compared to the ePVC embedding.

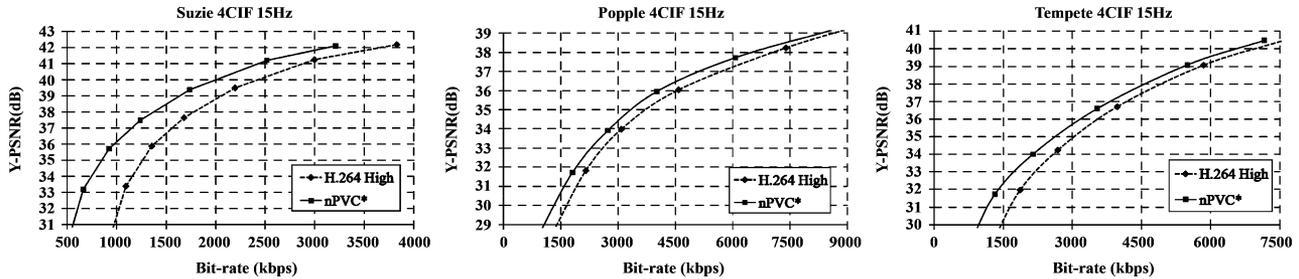


Fig. 14. Rate-distortion performance for three 4CIF standard video sequences by the H.264 *High* profile and its embedding with the proposed new PVC nPVC* technique using $\beta = 4$ minus the bi-directional predictive mode.

While the improvement diminished with higher bit-rates, in all the cases the proposed embedding improved H.264 image quality by at least 0.5 dB, which is perceptually significant. For the CIF sequences, the improvement relative to the ePVC embedding was not perceptually recognisable at relatively high bit-rates. This is primarily due to less than 10% MBs ultimately encoded with the pattern-modes at such high bit-rates (see Fig. 3), and, thus, it should not be implicated undermining the changes introduced in the proposed technique such as avoiding all the thresholds and pattern-based residual coding. Note that embedding of the IBS technique into the H.264 *Baseline* profile improved the compression rate by no more than 5% at the low bit-rates. This performance is almost half to what was reported in [9] for relatively high bit-rates. This result once again proves that encoding of the relatively static high-ITR background segments of the MBs with ME&MC is not favoured by the rate-distortion optimization function in low bit-rate coding.

Intuitively the pattern-based video coding technique is an alternative to a “partial H.264 skip mode.” The presented coding gains are achieved because the proposed method significantly reduces the number of DCT coefficients (using four 4×4 DCT blocks) and the static regions have their DCT coefficients skipped altogether. The rate-distortion optimization process eventually favours the proposed mode at low bit-rates as the pattern mode tends to use fewer bits compared to the other modes.

As the proposed nPVC embedding into the H.264 *Baseline* profile increases the computational complexity by as much as 35% (see Fig. 9), it is worthwhile to investigate whether the pattern based coding mode pays off against the bi-directional prediction mode when embedded into the H.264 *High* profile. To accomplish this, we embedded the proposed nPVC mode using $\beta = 4$ into the H.264 *High* profile standard by replacing the bi-directional prediction mode. Let this embedding be identified as nPVC* to differentiate it from the proposed nPVC embedding into the H.264 *Baseline* profile. While we have continued using GOP of 15 frames with the IPPP...P format for nPVC*, we have had to use GOP with the IBPBP...BP format for the H.264 *High* profile to accommodate bi-directional prediction.

Fig. 13 presents the average computational complexity reduction (conducted according to the guidelines proposed in [20] and [21]) using the proposed nPVC* embedding against the H.264 *High* profile for three 4CIF (720×576) standard video sequences, namely *Suzie*, *Popple*, and *Tempete*. While both the pattern mode using $\beta = 4$ and the bi-directional prediction

TABLE II
IMAGE QUALITY IMPROVEMENT BY THE PROPOSED NEW PVC nPVC* EMBEDDING USING $\beta = 4$ AGAINST THE H.264 *High* PROFILE AT TWO DIFFERENT HIGH BIT RATES WITH 4CIF STANDARD VIDEO SEQUENCES

	1500 kbps	3000 kbps
<i>Suzie</i>	1.8 dB	0.5 dB
<i>Popple</i>	1.1 dB	0.6 dB
<i>Tempete</i>	2.0 dB	0.7 dB

mode require almost the same ME & MC operations, a significant number of MBs are eventually encoded using the patterns instead of 8×8 mode. For those MBs, a significant amount of operations in ME&MC are saved in nPVC* compared to the bi-directional prediction for further smaller modes (such as 8×4 , 4×8 and 4×4). Consequently, on average 11% computational complexity is saved by nPVC*.

Fig. 14 presents the rate-distortion performance for the abovementioned three 4CIF standard video sequences by the H.264 *High* profile standard and its embedding with the proposed nPVC* technique using $\beta = 4$ top-ranked pattern-modes. While the improvement diminished with higher bit-rates, nPVC* convincingly outperformed the H.264 *High* profile for the entire range of moderate to high bit-rates. Improvement in the image quality by the nPVC* embedding at two high reference bit-rates of 1500 and 3000 kbps are presented in Table II. The nPVC* embedding improved image quality 1.1 to 2.0 dB at 1500 kbps and 0.5 to 0.9 dB at 3000 kbps compared to H.264 *High* profile. So, replacing the bi-directional predictive coding with the proposed pattern based coding improves computational complexity as well as rate-distortion performance at moderate to high bit-rate on high resolution video sequences.

VII. CONCLUSION

In this paper, a new pattern-based video coding technique has been proposed, which successfully avoids content-sensitive thresholds that have been used in the previous pattern-based attempts. By avoiding the thresholds, this technique fully exploits the Lagrangian optimization function in selecting the best mode. In addition, we have also introduced a new similarity metric, multiple pattern modes, and a modified Lagrangian multiplier. A novel pattern-based residual encoding approach has also been proposed to address the occlusion issue. All these changes and additions have resulted in a much superior pattern-based coding technique. Once embedded into the H.264 *Baseline* profile, this technique can improve the image quality perceptually significantly by at least 0.5 dB in applications at low bit-rates, up to 128 and 512 kbps for the QCIF and CIF formats, respectively, with 35% increase in the computational complexity. We can even reduce the computational complexity by at least 10%

while improving the image quality by at least 0.5 dB while encoding higher resolution video sequences (4CIF) at moderate to high bit-rates when the proposed pattern-based coding technique replaces the bi-directional prediction mode in the H.264 High profile.

Handling the occlusion problem is itself a significant research challenge having many open problems yet to be solved by its primary research field, computer vision. To our knowledge, this paper is the first attempt to address this problem at block-partitioning level video coding. While occlusion may also occur due to overlapping of two or more objects, we have focused primarily on the background occlusion problem as at low bit-rates, extra bits to support additional segments needed to identify and tackle occlusion problems in general are often deemed excessive by the rate-distortion optimizer. We have also shown that the effectiveness of any block partitioning technique diminishes at high bit-rates and, hence, addressing other occlusion problems is more important when coding video at significantly higher bit-rates than considered by this paper.

ACKNOWLEDGMENT

The authors would like to thank Associate Editor A. Ortega and the anonymous referees for their insightful comments, suggestions, and criticism that considerably improved the quality of this article.

REFERENCES

- [1] *ITU-T Recommendation H.263 Video coding for low bit-rate communication*, Version 2 Version 2, 1998.
- [2] *ISO/IEC 13818, MPEG-2 International Standard*, 1995.
- [3] *ITU-T Rec. H.264/ISO/IEC 14496-10 AVC*, Joint Video Team (JVT) of ISO MPEG and ITU-T VCEG, JVT-G050, 2003.
- [4] M. Paul and M. M. Murshed, "Superior VLBR video coding using pattern templates for moving objects instead of variable-bloc size in H.264," in *Proc. 7th IEEE Int. Conf. Signal Processing*, Beijing, China, 2004, pp. 717–720.
- [5] E. M. Hung, L. Ricardo, D. Queiroz, and D. Mukherjee, "On macroblock partition for motion compensation," in *Proc. IEEE Int. Conf. Image Processing*, 2006, pp. 1697–1700.
- [6] O. D. Escoda, P. Yin, C. Dai, and X. Li, "Geometry-adaptive block partitioning for video coding," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2007, pp. 1-657–660.
- [7] O. D. Escoda, P. Yin, and C. Gomila, "Hierarchical B-frame results on geometry-adaptive block partitioning," presented at the VCEG-AH16 Proposal, ITU/SG16/Q6/VCEG, Antalya, Turkey, Jan. 2008.
- [8] J. Chen, S. Lee, K.-H Lee, and W.-J Han, "Object boundary based motion partition for video coding," presented at the Picture Coding Symp., 2007.
- [9] J. H. Kim, A. Ortega, P. Yin, P. Pandit, and C. Gomila, "Motion compensation based on implicit block segmentation," presented at the IEEE Int. Conf. Image Processing, 2008.
- [10] S. Chen, Q. Sun, X. Wu, and L. Yu, "L-shaped segmentations in motion-compensated prediction of H.264," presented at the IEEE Int. Conf. Circuits and Systems, 2008.
- [11] T. Fukuhara, K. Asai, and T. Murakami, "Very low bit-rate video coding with block partitioning and adaptive selection of two time-differential frame memories," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 212–220, 1997.
- [12] *ISO/IEC N4030, MPEG-4 International Standard*, 2001.
- [13] K. -W. Wong, K. -M. Lam, and W. -C. Siu, "An efficient low bit-rate video-coding algorithm focusing on moving regions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 1128–1134, 2001.
- [14] M. Paul, 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 Syst. Video Technol.*, vol. 15, no. 6, pp. 753–761, Jun. 2005.
- [15] T. Shanableh and M. Ghanbari, "Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formats," *IEEE Trans. Multimedia*, vol. 2, pp. 101–110, 2000.
- [16] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 688–702, 2003.
- [17] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *Proc. IEEE Int. Conf. Image Processing*, 2001, pp. 542–545.
- [18] Y. Q. Shi and H. Sun, *Image and Video Compression for Multimedia Engineering Fundamentals, Algorithms, and Standards*. Boca Raton, FL: CRC, 1999.
- [19] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression*. New York: Wiley, 2003.
- [20] M. Horowitz, "A coding efficiency-computational complexity analysis of KTA 1.9r1 coding tools," presented at the Q.6/SG16 (VCEG), 35th Meet., Berlin, Germany, Jul. 16–18, 2008.
- [21] T. K. Tan, G. Sullivan, and T. Wedi, "Recommended simulation common conditions for coding efficiency experiments revision 4," presented at the Q.6/SG16 (VCEG), 36th Meeting, San Diego, CA, Jul. 8–10, 2008.



Manoranjan Paul (M'03) received the B.Sc.Eng. (hons.) degree in computer science and engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 1997, and the Ph.D. degree from the Monash University, Churchill, Australia, in 2005.

He joined the Computer Science and Engineering Department, Ahsanullah University of Science and Technology, Dhaka, Bangladesh, as a Lecturer in 1997 and was promoted to Assistant Professor in 2000. He worked as a Research Fellow in the University of New South Wales, ADFA, Canberra, Australia, in 2005–2006, and Monash University, Australia, in 2006–2009. He is currently working as a Research Fellow at the School of Computer Engineering, Nanyang Technological University, Singapore. His major research interests are in the fields of image/video coding, multimedia communications, computer vision, and vertical handoff. He has published more than 35 refereed international journals, book chapters, and conference publications.

Dr. Paul became a member of IEEE in 2003 and Australian Computer Society in 2005. He has served as a guest editor of the *Journal of Multimedia*.



Manzur Murshed (M'96) received the B.Sc.Eng. (hons.) degree in computer science and engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 1994, and the Ph.D. degree in computer science from the Australian National University, Canberra, Australia, in 1999.

He is currently an Associate Professor and the Head of Gippsland School of Information Technology, Monash University, Australia, where his major research interests are in the fields of video coding and indexing, image processing, wireless and multimedia communications, and video surveillance. He has published more than 120 journal and peer-reviewed research publications.

Dr. Murshed served the *Journal of Multimedia* as a guest editor and refereed articles for many reputed journals and conference proceedings. He delivered two invited keynote speeches at the 2004 IEEE International Conference on Computer and Information Technology, Dhaka, and the 2007 EII Workshop on Video Signal Processing and Communication, Gippsland, Australia. He has received several large research grants, including two Discovery Project grants from the Australian Research Council. He received the 2007 Vice-Chancellor's Knowledge Transfer Excellence award (commendation) from the University of Melbourne, a 2006 Faculty of IT Excellence in Research by ECR award from Monash University, and the 1994 University Gold Medal from BUET.