IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 1, 2018, accepted March 2, 2018, date of publication March 16, 2018, date of current version May 24, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2816852

A Generalized Multi-Layer Framework for Video Coding to Select Prediction Parameters

MUHAMMAD ASIF[1], MAAZ BIN AHMAD[2], IMTIAZ A. TAJ[1], AND MUHAMMAD TAHIR[1]

[1]Department of Electrical Engineering, Capital University of Science and Technology, Islamabad 44000, Pakistan
[2]College of Computing and Information Sciences Department, PAF Karachi Institute of Economics and Technology, Karachi 75190, Pakistan

Corresponding author: Muhammad Asif (astz786@yahoo.com)

ABSTRACT The famous video coding standards of this era, such as high efficiency video coding (HEVC) and H.264/AVC, offer numerous coding parameters to enhance the compression ratio. These standards exploited a robust rate-distortion optimization (RDO) methodology to select the appropriate macroblock (MB) coding parameters, such as prediction type, modes, and block sizes. The exploitation of the RDO technique contributes significantly to increase the computational complexity of the coding process. In this paper, a generalized multi-layer framework is presented, which provides a hierarchical optimized way to select MB prediction parameters. Each layer of the proposed framework incorporates multiple innovative algorithms to shortlist the candidate prediction parameters prior to the RDO process. Moreover, in order to select the suitable prediction type and block size for intra-prediction, two techniques are proposed. The presented framework is flexible enough to accommodate various mode selection techniques that make it excellent choice to be used in the modern coding standards. The experimental results show that coding time is reduced up to 74% without significant loss in visual video.

INDEX TERMS RDO, intra-prediction, inter-prediction, H.264/AVC, HEVC, video coding.

I. INTRODUCTION

The revolutionary video coding standards High Efficiency Video Coding (HEVC) [1]–[3] and H.264/AVC [4] are created by the consolidated endeavors of ITU-T Video Coding Experts Group and ISO/IEC Moving Picture Experts Group. Better visual quality, compression ratio and enhanced peak signal-to-noise ratio (PSNR) are provided by these standards when contrasted with the past standards. The reason of these enhanced capabilities is the adaptation of numerous new schemes in these standards. These new schemes are: rate distortion optimization (RDO), directional intra prediction of blocks, de-blocks filtering, context-based adaptive binary arithmetic coding (CABAC) [5], multi-reference frame motion estimation (ME), variable block sizes ME and integer transform (DCT). The problem is that the consideration of these schemes enhances the computational complexity of encoder especially by RDO and variable block size motion estimation [6]–[8].

The fundamental coding unit is a macroblock (MB, i.e. 16×16 pixels) for a video frame processing in H.264/AVC standard. The following two are the prediction types for encoding the MBs:

- Intra-Predicted (I-MB): where reconstructed pixels of the neighboring MBs in the present frame are utilized for foreseeing an MB.
- Inter-Predicted (P-MB): where foreseeing of an MB is performed with the assistance of the reconstructed pixels from the previous frames

An H.264/AVC offers different coding modes with variable block sizes for intra and inter-prediction. These modes help in better portrayal of the temporal and spatial description of an MB. For luma intra-prediction, two block sizes i.e. 4×4 and 16×16 are used. Nine prediction modes are available for a luma 4×4 and for luma 16×16 and chroma 8×8 blocks, there are four modes. Eight of these nine prediction modes are directional and the staying one is DC i.e. mode 2 (for luma 4×4). Also for luma 16×16 and chroma 8×8, three out of four modes are directional and only one i.e. mode 2 is DC. A number of prediction block sizes i.e. 16×16, 16×8, 8×16, 8×8, 8×4, 4×8 and 4×4 are upheld for inter-prediction. Regardless of the block size, a technique called RDO is employed to find the best coding mode amongst the all conceivable mode combinations in H.264/AVC [9], [10]. The RDO computes a value called Rate Distortion cost (RDcost) for every conceivable prediction modes. The prediction mode having lowest value of RDcost is considered as the best mode. Along these lines, the number of conceivable intra-prediction modes for each MB are $N8 \times (16 \times N4 + N16) = 4 \times (16 \times 9 + 4) = 592$ and for inter-prediction, there are 20 conceivable modes. So, 592 and 20 RDO calculations are done by H.264/AVC in order to pick the most appropriate intra-prediction and inter-prediction mode. This exhaustive searching technique enhances the computational complexity of the encoder and becomes a hurdle for performing real-time encoding. Thus, there is a need to reduce this complexity in order to achieve efficient encoding for real-time applications.

This article proposes a general multi-layer framework for choosing suitable prediction parameters for an MB to diminish the computational complexity and overheads regarding RDO process. The displayed structure of the framework includes various techniques to pick suitable macroblock prediction type, appropriate block size selection for intra-prediction, SKIP mode early recognition and directional mode choice for both intra and inter prediction mode. Temporal and spatial statistics of video sequence, Quantization Parameter (QP), prediction modes of already coded neighboring MBs and thresholds based on motion-field statistics are exploited by these schemes before using the costly RDO process. The results obtained from experiments prove the fact that the presented framework significantly reduced the computational complexity and decreased the coding time while achieving negligible loss in video quality.

The remaining article is structured as follows: Literature review and statistical analysis are covered in section II and III, respectively. In section IV, the proposed multi-layer framework is illustrated followed by experimental results in section V. At the end, the conclusion is presented in section VI.

II. RELATED WORK

Several efforts have been performed by researchers to lessen the H.264/AVC encoder's computational complexity by proposing different schemes for selection of prediction parameters of an MB. The main theme of these schemes is to produce a smaller set of candidate modes for RDO calculation by eliminating the inappropriate modes and block sizes. Some of these schemes aim to early detect SKIP mode. Some states for SKIP mode early decision was discussed by Jeon's [11] and Choi's *et al.* [12]. SKIP mode was considered the best mode in it if all of these states were satisfied. So the remaining modes don't undergo the RDcost computation. To further improve the quality of this work, some other states obtained by exploiting adjacent MBs temporal and spatial coding details were proposed in [13]. Sum-of-Absolute-Transformed-Differences (SATD) concept was used by Saha *et al.* [14] for early detection of SKIP mode. In this technique all the modes were divided in to two groups i.e. group one comprises only SKIP mode and all other modes were in group two. First it is tested whether the SKIP mode is the optimal mode or not. If it is the optimal mode then all other modes don't undergo the RDcost computation. Otherwise, all the remaining modes go through the RDcost computation. This technique is not successful for the video sequences having detailed regions or fast motion.

Sum of Absolute Difference (SAD) concept with threshold value was deployed by some schemes [20], [21] to minimize the suitable inter-prediction modes list that have to undergo the RDO calculations. The approach of these schemes is simple, SAD value of an MB is matched against a threshold value to exclude the unnecessary modes. In [20] adaptive thresholds along with RDcost statistics were used to minimize that list while MB's activity related to its residual complexities (i.e. global and local) was exploited in [21] to shorten the inter-prediction modes list. Despite of the fact that schemes based on SAD are quite fast, yet to find the accurate threshold value is a difficult task. The quality of the video may be degraded significantly if an inappropriate value of the threshold is selected.

The concept of spatial homogeneity of an MB to find the appropriate modes of inter-prediction for RDcost calculation was used by Zhu *et al.* [22]. In this work, the Soble operator was applied on a down-sampled image to find the edge information (spatial homogeneity) of an MB. The scheme may give unsatisfactory results where the video sequence contains textured objects having smooth motion. Jing and Chau [23] tried to remove inappropriate modes of inter-prediction by taking in to account the temporal homogeneity concept of an MB. First, it is found whether the current MB belongs to the homogeneous portions or not. Mean Absolute Frame Difference (MAFD) and Mean Absolute Difference (MAD) values of the current MB are calculated to find it. The MAFD is the value obtained from the difference between this frame and previous frame while MAD is the value obtained from the difference between this MB and its same position's MB in the previous frame. Such category of techniques lacks in giving the guarantee for eliminating the unnecessary modes in the case of smooth motion background regions (e.g. video sequences captured by moving camera). Some schemes exploit both of the temporal and spatial homogeneity to find prediction modes. Wu *et al.* [24] used both of the spatial and temporal homogeneity concepts in the process of mode selection. They used MAD and Sobel operator to measure temporal and spatial homogeneity, respectively. Bharanitharan *et al.* [26] also developed a scheme targeting both of the spatial and temporal homogeneity by using block homogeneity concept to minimize inter-prediction candidate modes. 16×16 and 8×8 block patterns were used to calculate the MB homogeneity. Some techniques exploiting the concept of motion activity of an MB or adjacent MB's homogeneity are presented in [27] and [28]. In these techniques, if an MB is motion homogeneous or its motion activity is little then large block sizes are selected. Otherwise, the small block sizes are used in all other cases. An analysis about the status of motion activity in adjacent MBs (i.e. spatially and temporally adjacent) was performed by Zeng *et al.* [27] to find appropriate candidate modes. Liu *et al.* [28] applied 3 directional measures for motion homogeneity to find optimal inter-prediction mode. The use of normalized vector field to assess directional motion homogeneity was the prime feature of this scheme. At the 4×4 block level, motion estimation was performed to calculate motion vector field. More effective schemes may result if residual complexity concepts are also used inside these techniques beside motion homogeneity.

The major category of mode selection for intra-prediction comprises of techniques that make use of the local edge detection of blocks. All the techniques of this category differ in the way of finding local edge direction of blocks. Pan *et al.* [29] made use of Sobel operator to find local edge direction and its amplitude. As a result, a histogram was formed from which the most probable intra-prediction modes were selected for RDO process. Su *et al.* [30] calculated local edge direction with the help of integer transform. This transform was performed on actual frame. Adaptive threshold were also incorporated to have balance between complexity and compression.

In [31] and [32], Non-normalized Haar transform (NHT) concept was deployed to discover the sub-blocks edges. This helped to reduce the candidate modes for RDO calculation. The requirement of pre-computations enhanced its overall complexity. Wang *et al.* [33] deployed the concept of descriptors of edge histogram. So Dominant Edge Strength (DES) was detected from it and only a few modes were chosen for RDO. A technique presented by Elyousfi *et al.* [34] used the directional information. This technique exploited the similarity in the dominating direction of a smaller and bigger block. For luma components, RDO calculation (RDcost) was performed for 4 modes. The resemblance from adjacent block mode along with this RDcost helped to find the most appropriate 4×4 modes. The appropriate candidate modes for 8×8 and 16×16 luma components were selected by finding the same dominant direction of smaller and bigger blocks. For chroma components, the DC mode is used. Byeongdu *et al.* [35] made use of Dominant edge direction (DED) for selecting appropriate intra mode. Pixel value addition and subtraction in both horizontal and vertical directions were done to calculate DED. So only for three most likely modes, RDO computation was performed. Another exciting technique to select intra-mode was mentioned by Elyousfi [36]. In this technique the gravity center vectors of blocks concept was utilized. The direction of these vectors helped to reduce the number of candidate modes for RDO calculation. Bharanitharan *et al.* [37] utilized the texture direction concept using directional change between two adjacent pixels. The inappropriate modes were eliminated with the help of four texture directions i.e. vertical, diagonal-down-left, horizontal and diagonal-downright.

It is evident from the discussion on existing schemes that the way of finding local edge direction affects the overall performance of these schemes. The video quality is also damaged by adapting these schemes though they help in decreasing the computational complexity. The theme of most of the schemes which reduce the RDO process complexity is almost same i.e. to shortlist the candidate modes for RDO for some prediction type (I-MB or P-MB). So, worst case scenarios of all the coding modes need to be performed by these schemes which became them less effective. In the average case scenarios, around half of the modes need to be evaluated. Even in the best case scenarios, at least one mode has to be computed separately for I-MB and P-MB. In short, these schemes still has a significant computational complexity and cannot be applied as such in multimedia real-time applications.

TABLE 1. Encoder configurations.

Configuration Parameter	Option
Encoder Profile	Main
RDO	Enable
Motion Estimation	Enable
Motion Estimation Algorithm	Fast search
Motion Vector (MV)	32 pels
MV Resolution	1/4-pel
No. of Reference Frames	1
Entropy Encoding	CABAC
Encoded Frames per Sequences	300
Encoding Structure	IPPPPP
Quantization Parameters	24, 28, 32, 36 and 40

So, there comes a requirement to create a generalized framework which should optimize the selection process of prediction parameters by removing most of the unnecessary block sizes and prediction modes before applying RDO processing. Also, there must not be any significant degradation in video quality by deploying this framework. In order to achieve these requirements, a framework is proposed in this paper which is a comprehensive one in its nature. It not only selects the appropriate prediction modes and prediction type but also finds the most suitable partitions of macroblock. A comprehensive mode elimination methodology is provided to significantly reduce the encoder's computational complexity.

III. OBSERVATION AND STATISTICAL ANALYSIS

The classification of the regions of video sequences is normally done by seeing their spatial and temporal characteristics. These regions may be categorized as slow motion, high motion, motionless, darker, brighter, textured, uniform motion, complex motion etc. A detailed experimental series is performed on different video sequences to gather the information required for doing statistical examination of prediction parameters selection. To accomplish this task, H.264/AVC reference software was used. The Table 1 shows the encoder configurations for acquiring data to perform statistical analysis.

Table 2 and Table 3 list the averaged probability of selecting each prediction parameter.

Table 2 portrays the probability of selecting partition of macroblock (i.e. 4×4 or 16×16) for intra-prediction. Majority of the MBs (i.e. 65.56%) are encoded as intra 4×4 and 34.44% are encoded as intra 16×16. The probability of selection of intra 4×4 is greater in *Foreman*, *Mobile*, *Parkrun* and

TABLE 2. Intra-prediction modes selection (%) probability.

Test Streams	Resolution	I16×16	I4×4
Stockholm		38.45	61.55
Shield	720p	31.74	68.26
Parkrun	(1280×720)	80.3	19.7
Vtc1nw		68.89	31.11
Football	(720×480)	39.84	60.16
Flower	NTSC	25.61	74.39
Mobile		83.89	16.11
Intros		65.14	34.86
Tempete		15.51	84.49
Mobile	(352×288)	6.05	93.95
Akiyo	CIF	54.13	45.87
Silent		26.58	73.42
MaD		54.36	45.64
Foreman		16.57	83.43
Coastguard		23.17	76.83
Container	(176×144)	39.21	60.79
Claire	QCIF	53.68	46.32
Highway		33.96	66.04
Mean		34.44	65.56

Tempete video sequences having scenes of high texture while for video sequences like *Akiyo*, *Claire*, *Intros* and *Vtc1nw* having homogeneous regions, Intra 16×16 selection probability is higher.

It is evident from Table 3 that for most of the MBs (i.e. 98.83%), inter-prediction type is selected while intra-prediction is selected for rest of the MBs. This table also concludes that for *Akiyo*, *Claire* and *Vtc1nw* video sequences (i.e. no motion or low motion sequences), SKIP mode is the most selected one and its selection percentage is 57.98%. It is clear from the table that smaller block sizes are chosen for object boundary or irregular motion cases. The large block sizes are chosen for high or low textured objects. The selection of larger block sizes increases as the QP is increased for both inter and intra-prediction.

In the motion estimation process, inter-prediction block sizes may help to reduce the residual or prediction error. The greater block sizes (e.g. 16×8, 8×16, 16×16) are more appropriate for prediction where an MB is part of still region or of homogeneous motion. Such MBs are common in the situations of static background, uniform motion of rigid object and smooth motion of moving background. The effect of selecting larger block sizes in these situations is the reduction of prediction error. On the other hand, if larger block sizes are used in case of high texture or complex motion, it results in greater residual error. So, larger block sizes should be avoided for such situations and small block sizes should be used as these have the capability to properly capture the complex motion. Some of the small block sizes are 8×4, 4×8, 8×8, and 4×4.

Fig.1 highlights an example of 2 frames of CIF resolution in which optimal inter-prediction modes are shown by using multiple block sizes for a corresponding MB. These modes are chosen with the help of reference software of the encoder. In the twenty ninth (29th) frame of *coastguard* stream the camera panning activity gives impression of background motion. As a result, most of the MBs are coded using 16×16 block sizes due to the reflection of homogeneous translation motion for shore and water in the background. As the ripple and boundary regions carry non-uniform motion, these are encoded using small block sizes as shown in Fig.1 (a). Similarly, the 40th frame of *Paris* stream carries static background. Large block sizes are used for encoding it due to absence of motion. The boundary region's MBs carrying different motion (e.g. clothes, face, head boundary regions) are encoded using small block sizes as shown in Fig. 1(b). The small block sizes are suitable in these regions due to their resistance against the residual error.

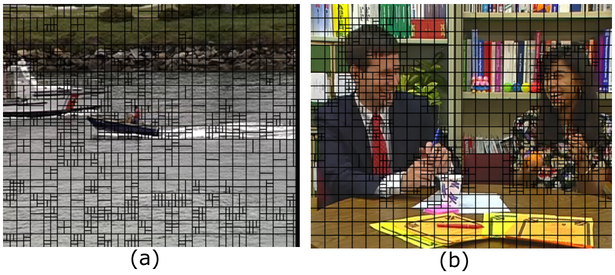


FIGURE 1. The optimal inter-prediction modes(a) Coastguard 29th frame (b) Paris 40th frame.

So it is evident from the above statistical analysis that the contents of the video (spatial and temporal statistics) play a vital role in the appropriate selection of prediction parameters. The SKIP mode is the prime candidate for encoding static or slow motion video sequences. The chances for selecting intra-prediction become higher in the case where MBs represent low motion region. Objects having homogeneous motion are normally encoded by larger block sizes in inter-prediction. Objects carrying irregular motion or for their boundary regions, the smaller block sizes become the better choice of selection. Intra 4×4 becomes the most suitable mode for MBs containing detailed regions while intra 16×16 becomes the ideal mode for MBs carrying smooth regions. There also exists a strong correlation among the coding mode of an MB and spatial and temporal statistics of its neighbors MBs. So these statistics of the adjacent MBs can be utilized efficiently in order to foresee most of the prediction parameters for a video frame.

IV. PROPOSED MULTI-LAYER FRAMEWORK

The presented framework consist of following five layers

- Layer 1: Features Extraction
- Layer 2: Prediction Type Decision
- Layer 3: Prior Mode Elimination
- Layer 4: Quick Mode Selection
- Layer 5: RDO Mode Exclusion

These layers of the presented muti-layer architecture contains various tools and techniques to extract spatial and temporal features and to select most probable prediction parameters. Fig. 2 shows the block diagram of the presented framework. The following sections describe in detail the functionally of each layer of the proposed framework.

TABLE 3. Prediction modes selection (%) probability.

Test Streams	Resolution	I16×16	I4×4	SKIP	16×16	16×8	8×16	8×8p
Shield		0.65	0.34	60.85	19.14	4.98	4.94	9.1
Parkrun	720p	0.14	0.11	34	23.06	7.04	7.59	28.06
Stockholm	(1280×720)	0.51	0.22	62.43	16.94	5.64	5.12	9.14
Intros		1.26	1.78	75.25	11.27	4.03	3.85	2.56
Football	(720×480)	3.54	6.39	39.75	20.34	8.53	9.91	11.54
Flower	NTSC	0.26	0.38	35.94	23.84	11.32	6.94	21.32
Mobile		0.27	0.15	27.28	23.67	9.98	10.01	28.64
Vtc1nw		0.01	0	94.54	3.05	0.91	0.79	0.7
MaD		0.09	0.14	77	12.39	3.6	3.92	2.86
Mobile	(352×288)	0.06	0.04	28.75	25.34	7.08	7.15	31.58
Akiyo	CIF	0	0	88.46	5.12	1.8	2.09	2.53
Silent		0.3	0.91	75.56	9.78	3.37	4.28	5.8
Tempete		0.62	1.37	32.44	25.4	8.79	7.64	23.74
Foreman		0.4	0.31	39.62	25.23	8.73	9.75	15.96
Coastguard	(176×144)	0.06	0.19	36.52	29.48	7.24	7.25	19.26
Claire	QCIF	0.01	0.07	84.18	6.98	2.36	2.17	4.23
Highway		0.03	0.1	67.77	15.84	5.78	3.61	6.87
Container		0.04	0.13	83.34	7.32	2.75	1.82	4.6
Mean		0.69	0.48	57.98	16.9	5.77	5.49	12.69

A. FEATURES EXTRACTION

According to statistical analysis performed in section III, macroblock prediction parameters are highly correlated with the spatial and temporal characteristics of the current MB and its neighboring MBs. In order to foretell about suitable prediction parameters, the following features are exploited in this work.

1) BRIGHTNESS (B)

The brightness of an MB is the average over all pixel intensities values of its luminance component Y(m,n). This statistical measure is used to classify an MB being a part of the dark or bright region of video frame. The following expression is used to compute the brightness of an MB.

$$B_{MB} = \frac{1}{WH}\sum_{m=1}^{W}\sum_{n=1}^{H} Y(m,n) \tag{1}$$

2) ZERO SAD (SAD_Z)

Zero SAD of an MB provides the information about its degree of motion (movement or stillness) with reference to its collocated MB in the previous frame of video sequence in displaying order. The following expression is used to calculate this parameter.

$$SAD_z = \sum_{u=1}^{W}\sum_{v=1}^{H} |Y(u,v) - Z(u,v)| \tag{2}$$

Where Y(u,v) and Z(u,v) indicate the luminance components of current and it's collocated MB in the foregoing frame, respectively.

3) VARIANCE (σ)

This statistical parameter is helpful to determine how each pixel with in an MB varies from the mean or neighboring pixel. Moreover, it can also be used to classify whether MB belongs to smooth or textured area. The variance of an MB is approximately assessed using following expression

$$\sigma_{MB} = \sum_{m=1}^{W}\sum_{n=1}^{H} |Y(m,n) - B_{MB}| \tag{3}$$

4) BLOCK EDGE DESCRIPTORS

The block edge descriptors including edge strength and direction are utilized to foretell intra-prediction mode. The block edge information at each pixel is determined by convolving block with vertical and horizontal Sobel filters. The output of Sobel filters is named as an edge map that consists of edge vectors and each vector is belonging to 4×4 block of a video frame. For each 4×4 block of a frame located at *p*th row and *q*th column the edge vector can be described as.

$$\begin{aligned} Ux_{p,q} &= Y_{p-1,q+1} + 2\times Y_{p,q+1} + Y_{p+1,q+1} \\ &\quad - Y_{p-1,q-1} - 2\times Y_{p,q-1} - Y_{p+1,q-1} \qquad (4) \\ Uy_{p,q} &= Y_{p+1,q-1} + 2\times Y_{p+1,q} + Y_{p+1,q+1} \\ &\quad - Y_{p-1,q-1} - 2\times Y_{p-1,q} - Y_{p-1,q+1} \qquad (5) \end{aligned}$$

Where $Ux_{p,q}$ and $Uy_{p,q}$ represent the intensity change in horizontal and vertical directions, respectively. The edge strength or magnitude $Z_{p,q}$ using $Ux_{p,q}$ and $Uy_{p,q}$ can be roughly computed as follows:

$$Z_{p,q} = |Ux_{p,q}| + |Uy_{p,q}| \tag{6}$$

The block edge direction $\theta_{p,q}$ can be determined by using following equation

$$\theta_{p,q} = \frac{180}{\pi} \times \arctan(Uy_{p,q}/Ux_{p,q}), \theta_{p,q} \in [0, 2\pi] \tag{7}$$

5) MOTION ACTIVITY (MA) AND RESIDUAL COMPLEXITY (RC)

Motion activity or homogeneity of an MB provides the information whether motion activity with in an MB is low (motion-homogeneous) or high (non-homogeneous). The residual complexity or prediction error of an MB is defined by sum

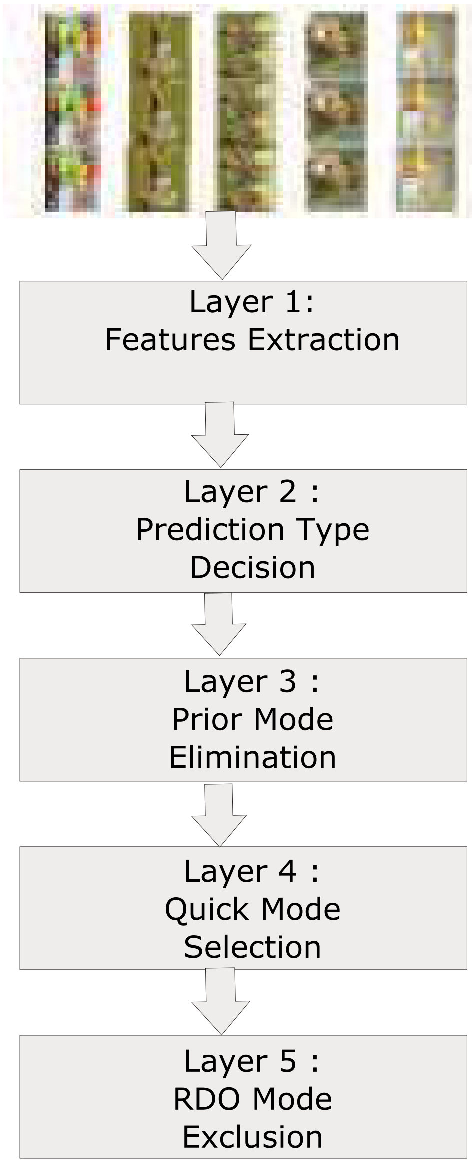


FIGURE 2. Block diagram of proposed multi-layer framework.

of absolute difference (SAD) of its constituent 8×8 blocks. These features are used to foretell inter-prediction modes and are computed using light weight 3-D Recursive Search (3-D RS) motion estimator [38] that converges toward the true motion at the 8×8 block level. If an $MB_{m,n}$ is situated at mth row and nth column of a video frame the motion vectors and SAD of its corresponding blocks are represented as $V_{p,q} = \{Vx_{p,q}, Vy_{p,q}\}, SAD_{p,q}, p \in [8m, 8m+1], q \in [8n, 8n+1]$. The following expressions are used to compute the MA and MH are of an MB.

The mean deviation of motion vectors can be computed as

$$D(P_t) = \frac{1}{2}\sum_{P_t}\{|Vx_{p,q} - \frac{1}{2}\sum_{P_t} Vx_{p,q}| + |Vy_{p,q} - \frac{1}{2}\sum_{P_t} Vy_{p,q}|\} \tag{8}$$

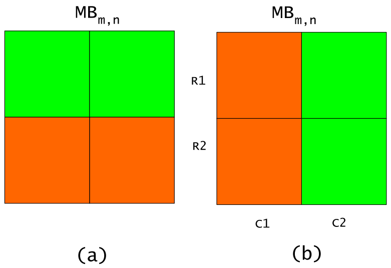


FIGURE 3. MB partitions (a) row wise (b) column wise.

The partition P_t can be either row R_t in Fig.3 (a) or column C_t in Fig.3(b). The vertical and horizontal MA and RC of an $MB_{p,q}$ are computed with the help of Eq.9, Eq.10 and Eq.11, respectively.

$$H\_MA_{p,q} = \frac{1}{2}\sum_{t=1}^{2} D(R_t) \tag{9}$$

$$V\_MA_{p,q} = \frac{1}{2}\sum_{t=1}^{2} D(C_t) \tag{10}$$

$$RC_{p,q} = \frac{1}{4}\sum SAD_{r,s} \tag{11}$$

The average MB residual complexity (AMB_RC) and frame residual complexity (FRC) are calculated with the help of Eq.12 and Eq.13, respectively.

$$AMB\_RC = \frac{1}{N}\sum_{k=1}^{N} SAD_k \tag{12}$$

$$FRC = \frac{1}{WH}\sum_{k=1}^{N} SAD_k \tag{13}$$

Where N represents the number of MBs in the frame and W, H are its width and height.

6) PREDICTION TYPE STATISTICS (PTS)

These statistics provide the information on how many spatial and temporal neighboring MBs of current MB are predicted as intra. The following expressions are used for collecting prediction type statistics

$$\begin{aligned} PTS_{spatial} &= isIntra(MB_L) + isIntra(MB_T) \\ &\quad + isIntra(MB_{TL}) + isIntra(MB_{TR}) \quad (14) \\ PTS_{temporal} &= isIntra(MB_T) + isIntra(MB_{TL}) \\ &\quad + isIntra(MB_{TR}) + isIntra(MB_R) \\ &\quad + isIntra(MB_L) + isIntra(MB_{DL}) \\ &\quad + isIntra(MB_D) + isIntra(MB_{DR}) \\ &\quad + isIntra(MB_{Collocated}) \quad (15) \end{aligned}$$

7) MOTION-FIELD STATISTICS (MFS)

The sum of absolute difference (SAD) of the spatial and temporal neighboring MBs of current MB are used to obtain

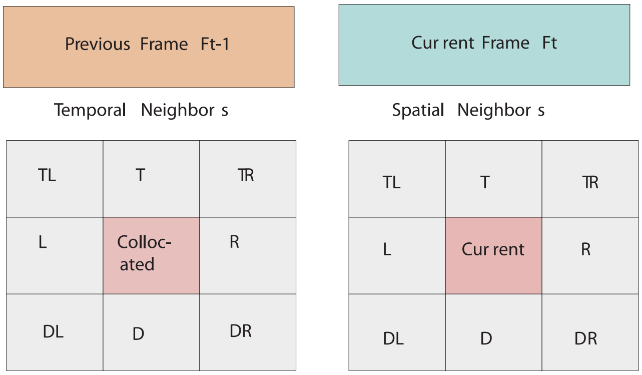


FIGURE 4. Neighboring macroblocks of current MB.

the motion-fields statistics. The following equations are used to compute the motion-fields statistics

$$MFS_{spatial} = \frac{1}{4}(SAD_T + SAD_L + SAD_{TL} + SAD_{TR}) \quad (16)$$

$$\begin{aligned} MFS_{temporal} = \frac{1}{9}(&SAD_T + SAD_L + SAD_{TL} \\ &+ SAD_{TR} + SAD_{DL} + SAD_R \\ &+ SAD_{Collocated} + SAD_{DR} + SAD_D) \quad (17) \end{aligned}$$

Fig.4 displays the current MB and its spatial and temporal neighboring MBs. In proposed methodology, a scaled down frame is used to compute the brightness, block edge statistics, variance and Zero SAD of an MB. The motive behind the exploitation of down sized image is to minimize the computations involve in feature extraction process. The frame down sizing factor is set to four. In down sized frame, one pixel indicates the 4×4 block of original frame and 4×4 block is equivalent to an MB i.e 16×16 block.

B. PREDICTION TYPE DECISION

It is evident from statistical analysis presented in section III, prediction type decision is critical and contributed significantly to computational complexity of the RDO process. In order to minimize its contribution, this decision should be made at the start of prediction parameters selection process. This layer comprises of the proposed scheme to made decision regarding prediction type (I-MB or P-MB) of an MB. In which a machine learning-based solution is presented that classify each MB in one of the three pre-defined classes using Adaboost classifier. The training data that is used to model Adaboost classifier comprises of various spatial and temporal features including brightness, Zero SAD, variance, residual complexity, prediction type statistics and motion-field statistics. The training data is acquired from test streams listed in Table 4. Then the aforementioned spatial and temporal features of an MB and prediction type selected by the RDO process are considered as the learning set for the classifier.

For each test macroblock $MB_{p,q}$ with feature vector $V_{p,q} = \{B, SAD_z, \sigma, RC, PTS_{spatial}, PTS_{temporal}, MFS_{spatial}, MFS_{temporal}\}$, the decision about class is made based on class probability given by classifier. Suppose LH_C1 and LH_C2 represent the likelihood of occurrence of class 1 and Class 2, respectively. The class selection criteria is given in Table 5.

TABLE 4. Training sequences.

Test Streams	Resolution
Park_joy FourPeople In_to_tree Johnny KristenAndSara	720p (1280×720)
BasketballDrill PartyScene BQSquare BlowingBubbles RaceHorses	(720×480) NTSC
Waterfall Students Soccer Sign_irene Pamphlet	(352×288) CIF
Suzie Salesman Miss_am Ice Husky	(176×144) QCIF

TABLE 5. Prediction type decision.

Class	Condition	Prediction type
1	$LH\_C1 > LH\_C2$ & $LH\_C1 > T$	Intra
2	$LH\_C2 > LH\_C1$ & $LH\_C2 > T$	Inter
3	Otherwise	Consider Intra and Inter

TABLE 6. Intra-prediction block size selection.

Class	Condition	Candidate block size
1	$LH\_C1 > LH\_C2$ & $LH\_C1 > T$	4×4
2	$LH\_C2 > LH\_C1$ & $LH\_C2 > T$	16×16
3	Otherwise	16×16 and 4×4

The threshold T is adjusted to 0.6 after exhaustive simulations on various video streams. As a result of such categorization, the decision about prediction type for most of the MBs is made that significantly minimized the computational complexity of the RDO process.

C. PRIOR MODE ELIMINATION

Two main decisions are made at this layer of the framework. First one is related to inter-predicted MBs i.e SKIP mode early detection and second belongs to intra-predicted MBs i.e appropriate prediction block size (16×16 or 4×4) selection. These decisions eliminate most of the unlikely modes for both prediction types.

1) EARLY DETECTION OF SKIP MODE

An H.264 JM reference encoder [41] computes RDcost for all possible coding modes for each MB and encode it using mode that minimizes the RDcost. The statistical analysis shows that in case of inter-predicted MBs, SKIP mode overlooks among all the candidate modes particularly for video streams comprising homogeneous regions (slow-motion or still contents).

TABLE 7. Classes, conditions and candidate inter-prediction modes.

Class	Condition	Candidate Modes
1	$RC_{p,q} < T1 \,\&\, H\_MA_{p,q} < V1 \,\&\, V\_MA_{p,q} < V1$	16×16
2	$RC_{p,q} > T2 \,\&\, H\_MA_{p,q} > V2 \,\&\, V\_MA_{p,q} > V2$	16×16, 16×8, 8×16, 8×8p
3	$H\_MA_{p,q} > V\_MA_{p,q}$	16×16, 8×16
4	$H\_MA_{p,q} < V\_MA_{p,q}$	16×16, 16×8
5	$H\_MA_{p,q} = V\_MA_{p,q}$	16×16, 16×8, 8×16

TABLE 8. Primary intra-prediction modes for luma 4×4 block.

Mode	Condition
0	$\beta \in$] 258.75, 281.25] U] 78.75, 101.25]
1	$\beta \in$]348.75, 360.0] U [0.0, 11.25] U]168.75, 191.25]
3	$\beta \in$]213.75, 236.25] U] 33.75, 56.25]
4	$\beta \in$] 303.75, 326.25] U] 123.75, 146.25]
5	$\beta \in$] 281.25, 303.75] U] 101.25, 123.75]
6	$\beta \in$] 326.25, 348.75] U] 146.25, 168.75]
7	$\beta \in$] 236.25, 258.75] U] 56.25, 78.75]
8	$\beta \in$] 191.25, 213.75] U] 11.25, 33.75]

TABLE 9. Primary intra-prediction modes for luma 16×16 block.

Mode	Condition	Edge Strength
0	$\beta \in$]247.5, 292.5] U]67.5, 112.5]	EM0+= $Z_{i,j}$
1	$\beta \in$]337.5, 360] U [0, 22.5] U]157.5, 202.5]	EM1+= $Z_{i,j}$
3	else	EM3+= $Z_{i,j}$

TABLE 10. Primary intra-prediction modes for chroma 8×8 block.

Mode	Condition	Edge Strength
1	$\beta \in$]337.5, 360] U [0, 22.5] U]157.5, 202.5]	EM1+=$Z_{i,j}$
2	$\beta \in$]247.5, 292.5] U]67.5, 112.5]	EM2+=$Z_{i,j}$
3	else	EM3+=$Z_{i,j}$

This analysis demonstrates that procedure regarding detection of SKIP mode should be executed before beginning the inter-prediction mode selection process to avoid the RDcost computations for the remaining candidate modes that results in remarkably reduction of computational complexity. In this work, Jeon's technique [11] is exploited for SKIP mode early detection. The following are the steps of this algorithm

- Accomplish ME for an MB by exploiting 16×16 division and one reference frame. It gives motion vector and predicted MB as output.
- Calculate the residual/prediction error by subtracting predicted MB from actual MB data.
- Subtract the predicted motion vector (obtain from neighboring MBs motion vectors) from the actual motion vector (obtain through ME) in order to compute the motion vector difference (MVD).
- Transformed and quantized the prediction error to calculate the quantized coefficients.
- If MVD and quantized coefficients are zero than SKIP mode is selected. Otherwise, rest of the modes are candidate modes for inter prediction.

2) BLOCK SIZE SELECTION FOR INTRA-PREDICTION

In video coding for intra prediction, the selection of proper block size plays vital role to significantly decrease the prediction error and to enhance the coding efficiency. Generally, 4×4 is appropriate block size for MBs belonging to high textured or detailed regions and 16×16 is suitable for MBs being a part of smooth regions or low textured areas of a video sequence. In H.264/AVC, RDO is performed both for block sizes 16×16 and 4×4 without considering which block size is suitable. This approach increases overall computational complexity of the encoder. If it is possible to detect MBs belonging to smooth or low textured regions of video sequence then RDcost calculation for 4×4 prediction modes can be avoided because 16×16 block size is suitable for intra-prediction. Similarly, recognition of detailed or high textured MBs can be fruitful to skip RDcost computation for 16×16 prediction modes. Based on these observations, a technique is proposed to select an appropriate block for intra-prediction. In this technique, the task of suitable block size selection (4×4 or 16×16) is modeled as classification dilemma that utilizes AdaBoost classifier to categorize each MB into one of the three pre-defined classes. The training data that used to model Adaboost classifier comprises of two spatial features of an MB including brightness and variance. The training data is acquired from video streams listed in Table 4. Then the aforementioned features of an MB and block size for intra-prediction determined by the RDO are considered as the learning set for the classifier.

For each test macroblock $MB_{p,q}$ with feature vector $V_{p,q} = \{B, \sigma\}$, the decision about class is made based on class probability given by the classifier. Suppose LH_C1 and LH_C2 represent the likelihood of occurrence of class 1 and Class 2, respectively. The class selection criteria is mentioned in Table 6.

The threshold T is adjusted to 0.6 after exhaustive simulations on various video streams. As a result of such categorization, the decision about intra-prediction block size for most of the MBs is made that significantly minimizes the computational complexity of the RDO process and speed-up the coding process.

D. QUICK MODE SELECTION

The basic purpose of this layer is to reduce the candidate intra and inter-prediction modes for RDO process. It comprises two techniques one for inter-prediction mode selection [39] and other for intra-prediction mode selection [40]. The following sections briefly describe these techniques.

1) MODE SELECTION FOR INTER-PREDICTION

Inter-prediction mode selection algorithm is based on the observations that optimum inter-prediction mode for an

TABLE 11. Results for prediction type decision technique.

Test Streams	Resolution	*BDPSNR*	*BDBR*	*TS*	Class 1	Class 2	Class 3
Football		-0.045	1.072	-25.76	1.35	78.71	19.94
Flower		-0.002	0.023	-22.19	0.05	65.58	34.37
Mobile	NTSC	0.003	-0.062	-21.87	0.04	69.28	30.68
Vtc1nw	(720×480)	-0.011	0.34	-21.48	0	67.24	32.76
Galleon		-0.022	0.699	-22.93	0.64	71.44	27.92
Washdc		-0.018	0.419	-32.51	0.01	93.93	6.06
Mobile		-0.003	0.069	-29.02	0.2	89.7	10.1
Akiyo		0	0.001	-19.77	0.06	68	31.94
Paris	CIF	-0.004	0.072	-30.78	0.01	92.21	7.78
MaD	(352×288)	-0.025	0.548	-19.93	0.18	66.04	33.78
Tempet		-0.018	0.429	-30.15	0.27	89.57	10.16
Silent		-0.027	0.638	-27.11	0.46	87.09	12.45
Claire		0.037	-0.576	-20.56	0.01	66.83	33.16
Foreman		-0.009	0.167	-31.53	0.02	92.59	7.39
Container	QCIF	0	-0.021	-22.36	0	72.21	27.79
Coastguard	(176×144)	0.008	-0.249	-30.9	0.03	90.56	9.41
Highway		0.019	-0.508	-28.34	0.03	81.04	18.93
Hall		0	-0.011	-23.73	0	79.89	20.11
Mean		-0.006	0.169	-25.61	0.19	78.99	20.82

TABLE 12. Results of intra-prediction block size detection technique.

		Performance Analysis			Class Frequency (%)		
Test Streams	Resolution	*BDPSNR*	*BDBR*	*TS*	Class 1	Class 2	Class 3
Mobile		-0.023	0.268	-24.95	84.88	9.41	5.7
Flower	NTSC	-0.024	0.304	-18.33	63.59	22.8	13.61
Galleon		-0.025	0.626	-14.78	61.96	17.28	20.76
Washdc		-0.094	0.637	-18.71	86.51	2.93	10.56
Paris	CIF	-0.044	0.563	-15.13	76.24	4.4	19.36
Mobile		-0.009	0.099	-26.18	91.45	3.83	4.71
Tempet		-0.029	0.408	-24.27	87.43	5.58	6.98
Silent		-0.069	0.504	-15.6	78.48	3.48	18.04
Coastguard		-0.068	0.698	-13.53	73.54	0.42	26.04
Foreman	QCIF	-0.065	0.913	-22.17	88.04	3.56	8.44
Hall		-0.065	0.739	-14.4	69.48	6.86	23.65
Container		-0.027	0.82	-12.08	55.66	12.64	31.7
Mean		-0.045	0.548	-18.34	76.44	7.76	15.80

MB is highly correlated with its motion activity and residual complexity. In this approach, each $MB_{p,q}$ is classified into one of the five predefined classes when the specified conditions hold true. Based on its class, candidate inter-prediction modes are selected for RDO process. Table 7 enlists the conditions, corresponding classes and candidate inter- prediction modes.

The thresholds T1 and T2 are functions of average MB residual complexity (AMB_RC) and frame residual complexity (FRC). These thresholds are adjusted with the help of Eq.18 and Eq.19, respectively.

$$T1 = \frac{1}{FRC} \times W1 \times AMB\_RC \tag{18}$$

$$T2 = \frac{1}{FRC} \times W2 \times AMB\_RC \tag{19}$$

Where weights W1 and W2 are a function of quantization parameter (QP) and are adjusted according to Eq.20 and Eq.21, respectively.

$$W1 = e^{0.0085 \times QP} \tag{20}$$

$$W2 = e^{0.02 \times QP} \tag{21}$$

The value of thresholds V1 and V2 is adjusted to 0.5 and 1.0, respectively.

In case of class 2, further analysis is performed to take a decision whether 8×4, 4×8 and 4×4 (smaller block sizes) belong to set of candidate modes or not. For each 8×8 block B_k in $MB_{m,n}$ with SAD_k the decision about small prediction modes (8×4, 4×8 and 4×4) is made according to following criteria:

- If SAD_k < T3, Ignore all small prediction modes.
- If SAD_k > T4, Consider all small prediction modes.
- Otherwise, Consider 8×4 and 4×8

Where

$$T3 = \frac{1}{4} \times W1 \times AMB\_RC \tag{22}$$

$$T4 = \frac{1}{4} \times W2 \times AMB\_RC \tag{23}$$

2) MODE SELECTION FOR INTRA PREDICTION

The algorithm for intra prediction mode selection exploits the block edge information including edge strength and direction to short list most suitable modes for RDO process. For each 4×4 luma block, let θ and $\beta = \theta + \pi/2$ indicate its edge and prediction direction, respectively. Table 8 illustrates the primary intra-prediction mode corresponding to prediction direction.

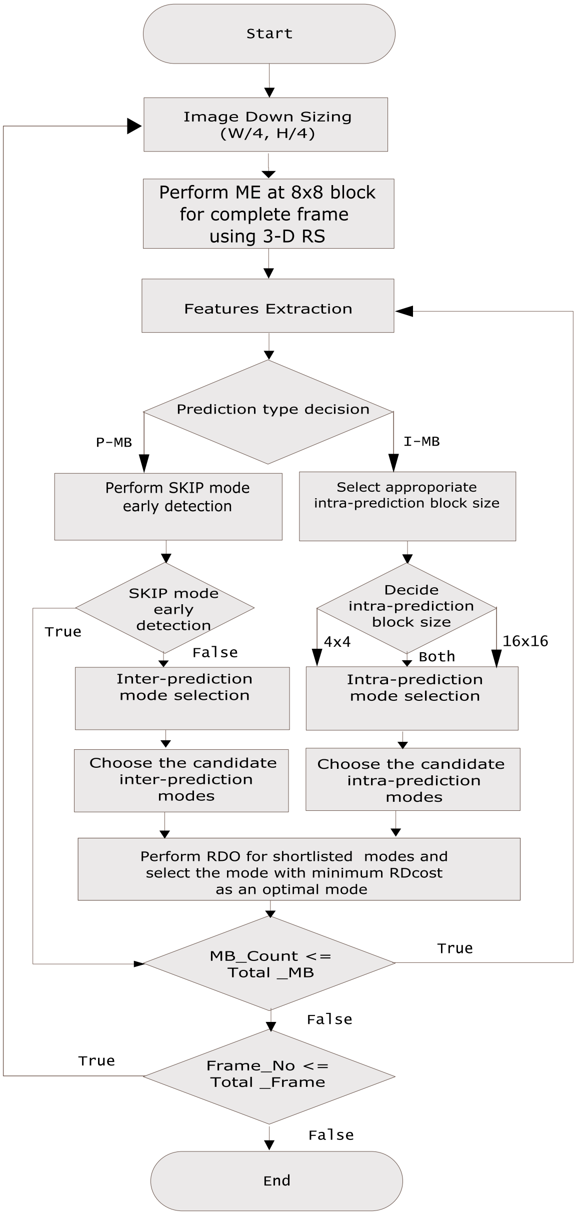

Start
Image Down Sizing (W/4, H/4)
Perform ME at 8x8 block for complete frame using 3-D RS
Features Extraction
Prediction type decision
P-MB
I-MB
Perform SKIP mode early detection
Select appropriate intra-prediction block size
SKIP mode early detection
True
False
Decide intra-prediction block size
4x4
Both
16x16
Inter-prediction mode selection
Intra-prediction mode selection
Choose the candidate inter-prediction modes
Choose the candidate intra-prediction modes
Perform RDO for shortlisted modes and select the mode with minimum RDcost as an optimal mode
MB_Count <= Total _MB
True
False
Frame_No <= Total _Frame
True
False
End

FIGURE 5. Flow diagram of the presented multi-layer framework.

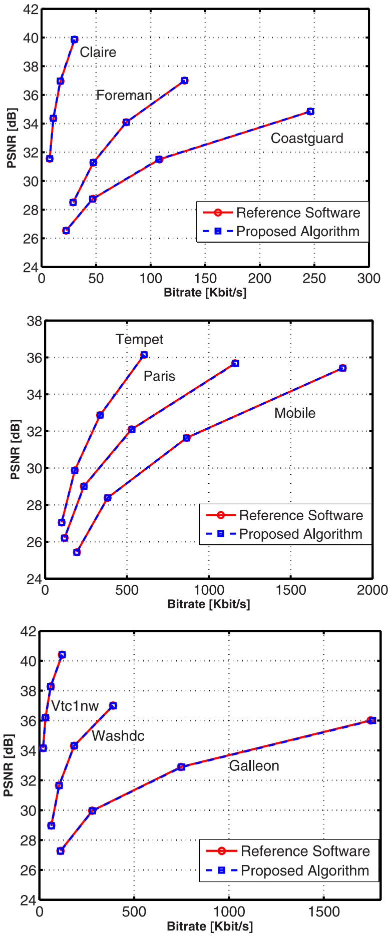


FIGURE 6. RD curves of JM reference algorithm and presented MB prediction type decision technique.

DC mode (2) is suitable for predicting smooth regions of video frame and has no dominant edge direction. Therefore, this mode is taken as a candidate prediction mode for all intra 4×4 blocks. Moreover, two neighboring modes of the primary prediction mode with respect to direction are also considered as candidate modes for RDO process. For example, if Mode 1 is the primary prediction mode for a block, then Mode 2, Mode 8 and Mode 6 will be three additional candidate prediction modes. In short, for each 4×4 block, 4 modes out of 9 go for RDcost calculation.

For each 16×16 luma and 8×8 chroma blocks, edge direction histogram is constructed with the help of their corresponding 4×4 blocks in order to determine the dominant edge direction. Table 9 enlists the mechanism to compute the histogram for luma 16×16.

Where $\beta = \theta + \pi/2$ is the prediction direction of each constituent 4×4 block of luma 16×16 component. Each bin of histogram adds up the edge strength of all 4×4 blocks having same prediction direction. The highest magnitude bin represents the most probable prediction direction for a block. The dominant prediction direction is used to select the primary prediction mode. In this algorithm, only those bins are considered for calculation of primary prediction mode that consists of at least five blocks.

For 8×8 chroma component, the same method is used as that of luma 16×16 except the mode order and there is no bound on similar direction blocks in histogram bin. Table 10 illustrates the histogram computation procedure for chroma 8×8 block. Based on the above described method, for both 16×16 luma block and 8×8 chroma block, one primary prediction mode is selected. The DC prediction mode is always candidate mode for them. Therefore, for luma 16×16 and chroma 8×8, two modes out of four are shortlisted for RDO process.

E. RDO MODE EXCLUSION

At this layer, conventional RDO technique is used to compute the RDcost for shortlisted candidate prediction parameters

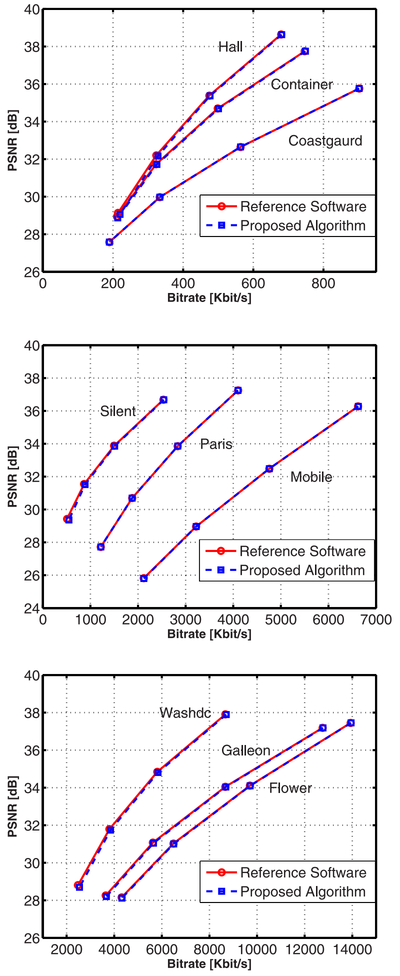


FIGURE 7. RD curves of JM reference algorithm and presented intra-prediction block size detection technique.

instead of all. The prediction parameters with least RDcost are selected for MB coding.

F. OVERALL ENCODING FLOW

The training process of Adaboost classifier is performed offline and trained models are loaded at the start of the encoding process. Fig.5 shows the encoding flow of the proposed multi-layer framework.

V. EXPERIMENTAL ANALYSIS

In order to assess the performance, JVT Reference Software [41] for H.264 is used to integrate the proposed framework. A series of experiments is done on intel core- i3 machine having 2 GB RAM. Multiple frame resolution video sequences are used carrying varying nature of motion contents. For example QCIF (144×176), CIF (352×288), NTSC (720×480), 720p (1280×720) and 1080p (1920×1080). There exist 4 different QP i.e. 40, 36, 32 and 28 for which every test sequence is encoded. −32 to +32 pels range is set for searching motion vector having resolution as 1/4 pels. In JM reference encoder, RDO is enabled and reference frame number is set as 5. For each of the QP value, the individual test sequence is run thrice and average result values are utilized.

TABLE 13. Results against IPPPPPP encoding structure.

Test Streams	Resolution	*BDBR*	*BDPSNR*	*TS*
Rush_hour	1080p (1920×1080)	0.913	-0.023	-69
Sunflower		0.467	-0.029	-73
Pedestrain_area		0.528	-0.008	-84
Riverbed		0.629	-0.035	-71
Parkrun	720p (1280×720)	0.856	-0.052	-65
Shields		0.363	-0.044	-70
Mobcol		0.128	-0.01	-66
Stockholm		0.153	-0.025	-69
Washdc	NTSC (720×480)	0.394	-0.015	-74
Vtc1nw		0.351	-0.039	-81
Galleon		0.126	-0.02	-70
Flower		0.949	-0.054	-65
Intros	CIF (352×288)	0.06	-0.016	-78
Tempet		0.852	-0.043	-66
Silent		0.123	-0.013	-80
Mobile		0.391	-0.029	-69
Akiyo		0.106	-0.012	-87
Football		0.789	-0.064	-67
Paris		0.956	-0.067	-65
Foreman		0.906	-0.065	-66
Container		0.518	-0.036	-75
News	QCIF (176×144)	0.413	-0.032	-76
Mother and daughter		0.11	-0.016	-82
Highway		0.251	-0.028	-79
Coastguard		0.567	-0.024	-67
Claire		0.458	-0.035	-85
Carphone		0.145	-0.012	-71
Miss-America		0.399	-0.039	-81
Mean		0.461	-0.032	-73.25

TABLE 14. Performance comparison against IPPPPP encoding structure (*TS*).

Test Streams	Resolution	Proposed	[21]	[20]	[11]
Mobile	CIF (352×288)	-69	-56	-29	-6
Akiyo		-87	-76	-62	-22
Paris		-65	-54	-40	-15
Foreman		-66	-62	-36	-10
Container		-75	-69	-55	-23
Football		-67	-48	-23	-10
Coastguard	QCIF (176×144)	-67	-47	-28	-7
Carphone		-71	-64	-39	-9
Claire		-85	-74	-65	-20
Miss-America		-81	-75	-60	-20
Highway		-79	-73	-53	-12
News		-76	-63	-57	-19
Mean		-74	-63	-46	-14

To compare the performance of the proposed framework with existing state of the art techniques, 3 metrics (i.e. speed-up in time (*TS*), Bjontegaard delta peak signal-to-noise ratio (*BDPSNR*) [42] and Bjontegaard delta bit-rate (*BDBR*)) are used. The equation below describes the calculation for *TS* value:

$$TS = \frac{T_p - T_r}{T_r} \times 100\% \tag{24}$$

Where

- T_p is the coding time of the proposed algorithm
- T_r is the coding time of the reference software
- The positive values of the performance measure metrics (*TS*, *BDPSNR* and *BDBR*) reflect increase

TABLE 15. Performance comparison against IPPPPP encoding structure (*BDPSNR*).

Test Streams	Resolution	Proposed	[21]	[20]	[11]
Mobile		-0.02	-0.04	-0.02	-0.01
Akiyo		-0.01	-0.03	0	0
Paris	CIF	-0.06	-0.02	-0.05	0
Foreman	(352×288)	-0.06	-0.04	-0.03	0
Container		-0.04	-0.02	-0.01	0.01
Football		-0.07	-0.04	0	0
Coastguard		-0.02	0.00	-0.04	-0.01
Carphone		-0.02	-0.09	-0.06	0.01
Claire	QCIF	-0.04	-0.08	0.01	0.01
Miss-America	(176×144)	-0.04	-0.06	0	0
Highway		-0.03	0.01	-0.08	-0.02
News		-0.03	-0.05	-0.03	-0.02
Mean		-0.037	-0.038	-0.026	-0.002

TABLE 16. Performance comparison against IPPPPP encoding structure (*BDBR*).

Test Streams	Resolution	Proposed	[21]	[20]	[11]
Mobile		0.39	0.87	0.34	0.17
Akiyo		0.11	-0.33	-0.14	-0.29
Paris	CIF	0.95	0.53	0.89	0.03
Foreman	(352×288)	0.91	0.34	0.49	-0.05
Container		0.52	-0.67	0.18	-0.34
Football		0.78	0.67	0.03	0.01
Coastguard		0.56	0.06	1.74	0.12
Carphone		0.14	0.10	0.84	-0.27
Claire	QCIF	0.45	-0.02	-0.17	0.1
Miss-America	(176×144)	0.39	-0.36	0.01	-0.14
Highway		0.25	1.50	2.55	0.67
News		0.36	-0.14	0.65	0.24
Mean		0.484	0.212	0.618	0.021

- The negative values of performance measure metrics reflect a decrease

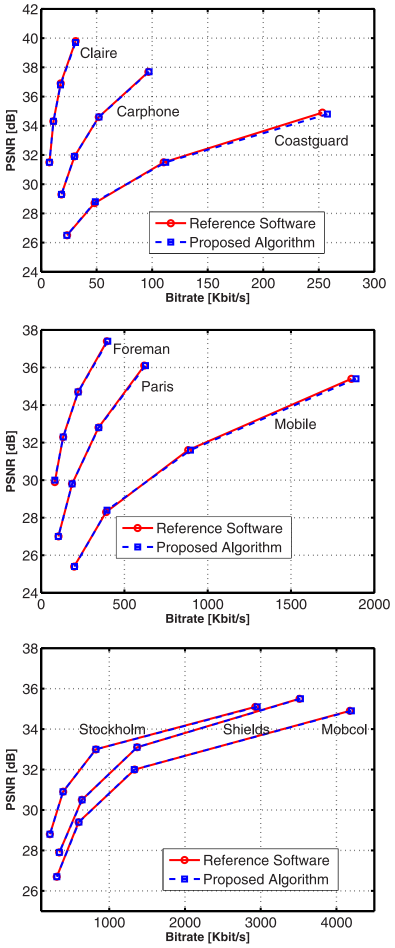


FIGURE 8. RD curves for IPPPPP encoding structure.

A. PREDICTION TYPE DECISION

The test streams of resolutions QCIF, CIF and NTSC each having 100 frames length are taken to judge the speed-up in RDO performance due to the selection of prediction type. The first most frame of each sequence is encoded as I-frame and all other frames are encoded as P-frames.

Table 11 describes the detailed results of the proposed scheme to make decision about prediction types. These results specify that the proposed scheme is faster than the searching method used in reference software by an amount of 25.61%. Also, it is evident from the table that there is only a slight increase of 0.169% in *BDBR* and a small reduction of 0.0006 *dB* of *BDPSNR* in the proposed scheme which is negligible. *Washdc* sequence gains the extreme speed-up of 32.51% while *Akiyo* sequence receives the lowest speed-up of 19.77% which shows the schemes works well for all of the test sequences. *Football* sequence contains high motion activity because of that it received the extreme gain in *BDBR* i.e. 1.072% and extreme *BDPSNR* loss i.e. 0.045. The MBs percentages falling in one of the 3 classes are also shown in Table 11. It is clear that the overall percentage of MBs which fall in Class 1 is quite small. The test sequences for which majority of the MBs fall in Class 2 received highest speed-up gains because RDcost is calculated only for inter-prediction. For example the speed-up of test sequences *Washdc* and *Foreman* are 32.51% and 31.53 %, respectively. If we look in to the number of MBs of these sequences which fall in Class 2, these are 93.93% for *Washdc* and 92.59% for Foreman. For sequences *MaD* and *Akiyo*, 33.78% and 31.94% of MBs fall in class 3, so the speed-up for these sequences is lower i.e. 19.93% and 19.77%, respectively. The reason for lower gain in speed-up is that RD cost is computed for inter as well as for intra prediction.

As shown from Fig 6, the RD performance of the proposed as well as reference software are very close to each other for multiple sequences.

B. BLOCK SIZE DETECTION FOR INTRA-PREDICTION

The test streams with three different resolution i.e. CIF, NTSC and QCIF are used to assess the gain in speed-up due to the presented Intra-Prediction Block Size Detection algorithm. All of these streams are of 100 frames length and all of the frames are encoded as I-frame.

Table 12 depicts the Intra-prediction detection results. It is evident from these results that a total gain of 18.34% is achieved in speed-up due to the proposed technique as compared to the search method used in the reference software.

TABLE 17. Performance comparison against all intra encoding structure (*TS*).

Test Streams	Resolution	Proposed	[37]	[33]	[29]
Container	QCIF (176×144)	-79.53	-67.74	-57.32	-56.36
News		-73.06	-68.02	-58.03	-55.34
Coastguard		-76.51	-69.26	-57.78	-55.03
Silent		-71.77	-69.46	-60.66	-65.17
Paris	CIF (352×288)	-71.44	-67.83	-58.78	-57.78
Mobile		-75.64	-68.89	-58.07	-59.09
Tempete		-75.95	-69.53	-56.86	-57.70
Stefan		-70.12	-67.04	-58.56	-57.97
Mean		-74.25	-68.47	-58.26	-58.06

TABLE 18. Performance comparison against all intra encoding structure (*BDPSNR*).

Test Streams	Resolution	Proposed	[37]	[33]	[29]
Container	QCIF (176×144)	-0.256	-0.241	-0.293	-0.234
News		-0.298	-0.285	-0.348	-0.294
Coastguard		-0.149	-0.181	-0.225	-0.106
Silent		-0.287	-0.232	-0.255	-0.183
Paris	CIF (352×288)	-0.268	-0.288	-0.274	-0.23
Mobile		-0.286	-0.250	-0.237	-0.255
Tempete		-0.254	-0.252	-0.254	-0.229
Stefan		-0.312	-0.290	-0.301	-0.242
Mean		-0.264	-0.252	-0.273	-0.221

TABLE 19. Performance comparison against all intra encoding structure (*BDBR*).

Test Streams	Resolution	Proposed	[37]	[33]	[29]
Container	QCIF (176×144)	3.685	2.976	4.44	3.695
News		3.621	3.439	4.451	3.902
Coastguard		2.585	2.820	4.034	2.361
Silent		4.992	4.013	4.58	3.54
Paris	CIF (352×288)	3.487	2.891	3.678	3.21
Mobile		3.104	2.860	2.871	3.168
Tempete		3.482	3.601	3.735	3.514
Stefan		3.721	3.730	3.889	3.717
Mean		3.585	3.291	3.960	3.388

Also the average increase in *BDBR* and decrease in *BDPSNR* is negligible i.e. 0.548% and 0.045%, respectively. Moreover, Table 11 reflects the number of blocks assigned to the different classes. Speed-up gain for encoder (*TS*) is more for Class 1 or Class 2 MBs. It can be seen for the video sequence Mobile, as there are 95.29% of MBs which are categorized in Class 1, 2 in CIF resolution. So the result is the overall speed-up gain of 26.18% with only 0.099% increase in *BDBR* and 0.009*dB* decrease in *BDPSNR* (both are negligible).

In Fig.7, the RD curves are drawn both for JM reference software and intra-prediction block size detection approach. It is clear from these graphs that the results of RD performance are quite close to each other for both of these schemes.

C. OVERALL PERFORMANCE

The performance comparison of the proposed framework with the previous works is done with the help of 5 multiple frame resolutions (i.e. NTSC, CIF, 720p, QCIF and 1080p). Simulation is done for both IPPPPP and all intra configurations.

TABLE 20. Results against all intra encoding structure.

Test Streams	Resolution	*BDPSNR*	*BDBR*	*TS*
Rush_hour	1080p (1920×1080)	-0.18	2.967	-72.67
Pedestrain_area		-0.202	2.679	-76.36
Sunflower		-0.231	3.408	-77.3
Riverbed		-0.228	3.367	-73.48
Parkrun	720p (1280×720)	-0.294	4.237	-64.97
Mobcol		-0.171	3.336	-70.29
Stockholm		-0.179	2.615	-70.85
Shields		-0.268	3.743	-68.01
Vtc1nw	NTSC (720×480)	-0.159	2.825	-80.03
Flower		-0.337	4.814	-66.35
Intros		-0.139	3.242	-75.17
Football		-0.259	4.528	-67.47
Galleon		-0.241	2.985	-74.26
Washdc		-0.262	3.326	-70.53
Mobile	CIF (352×288)	-0.286	3.104	-75.64
Akiyo		-0.155	2.107	-84.07
Paris		-0.268	3.487	-71.44
MaD		-0.131	2.264	-76.59
Tempet		-0.254	3.482	-75.95
Stefan		-0.312	3.721	-70.12
Carphone	QCIF (176×144)	-0.269	3.946	-74.61
Foreman		-0.312	4.649	-69.16
Coastguard		-0.149	2.585	-76.51
Claire		-0.13	2.318	-82.63
Silent		-0.287	4.992	-71.77
News		-0.298	3.621	-73.06
Container		-0.256	3.685	-79.53
Mean		-0.232	3.409	-73.66

1) EVALUATION AGAINST IPPPPPP ENCODING STRUCTURE

In this section, 100 frames are encoded in IPPPPP configuration for all test streams. It means that for each video sequence, the first frame is encoded as I-frame and remaining ones as P frames. The results in terms of *TS*, *BDPSNR* and *BDBR* are presented in Table 13 for multiple video streams. It is clear from these results that the proposed framework worked excellent and an overall 73.25% gain in speed-up is achieved for encoding time. 0.461% is the increase in *BDBR* and 0.032 *dB* is the decrease in *BDPSNR* (both are negligible). The proposed framework exhibits persistent performance for multiple resolutions video streams having 87% as the maximum gain in encoding speed and 65% as the minimum. This gain is achieved at the cost of slight increase in BDBR and slight decrease in *BDPSNR* (i.e. 0.956% and 0.067*dB*, respectively in the worst case scenario)

Table 14, 15, 16 are the comparison placeholder of the previous work with the proposed framework. These reflect that the proposed framework surpassed the Enrquez *et al.*'s [20], Jeon's [11] and Lee *et al.*'s [21] by saving encoding time up to 28%, 60% and 11%, respectively.

Fig.8 contains the RD curves of both of the proposed and reference software. It is clear from this figure that the results of both of these techniques for multiple video sequences are quite similar.

2) EVALUATION AGAINST ALL INTRA ENCODING STRUCTURE

In this section, 300 frames for all test streams are encoded as I-Frame for simulating results. Table 17, Table 18 and

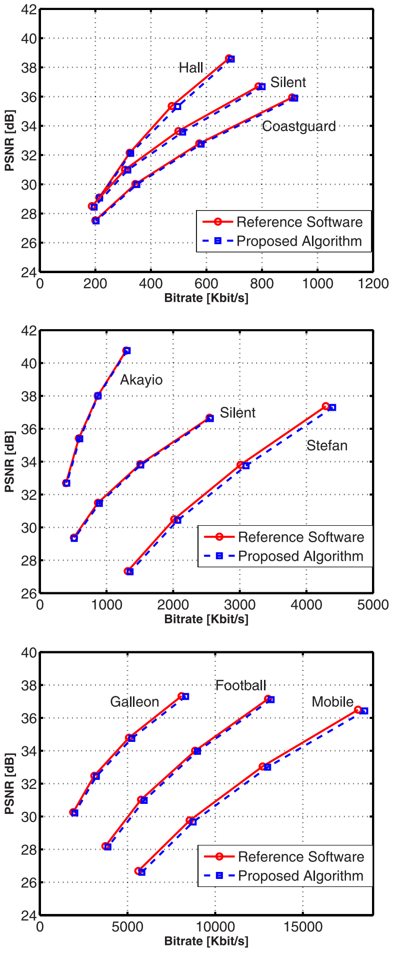


FIGURE 9. RD curves against all intra encoding structure.

Table 19 exhibits the performance comparison of the proposed framework in terms of *BDBR*, *BDPSNR* and *TS* with the previous classic schemes for intra configuration.

The results reflect the fact that the proposed framework saves TS of the encoder on an average by about 74.25%. The increase in *BDBR* is 3.585% and decrease in BDPSNR is 0.264*dB*. If the saved *TS* values for Wang *et al.*'s [33], Pan *et al.*'s [29] and Bharanitharan *et al.*'s [37] are considered than it shows that these schemes reduce *TS* by 58.26%, 58.06% and 68.47%, respectively. The *BDBR* values are increased by 3.960%, 3.338% and 3.291%, respectively. Similarly, the *BDPSNR* loss for these schemes is 0.273 *dB*, 0.221 *dB* and 0.252 *dB*, respectively. So the proposed framework surpasses these three classic techniques in terms of saving encoding time on the average by an amount 15.99%, 16.19% and 5.78%, respectively.

The proposed framework's comprehensive performance for multiple resolutions intra-frame sequences is depicted by Table 20. It shows that the proposed framework performs precisely for multiple resolutions test sequences and gains 73.66% average speed-up in encoding time. The average increase in *BDBR* is 3.409% and decrease in BDPSNR is 0.232*dB*.

The RD performance outcomes of the proposed framework are also just like the reference full search scheme for multiple video sequences, as shown by the RD curves of Fig.9.

VI. CONCLUSION

A computationally efficient multi-layer framework for macroblock prediction parameters selection is proposed in this paper. The presented approaches for selecting appropriate block size for intra-prediction and for prediction type decision are innovative and give acceptable results for variety of test sequences. The experimental results for H.264/AVC encoder show the effectiveness of the proposed framework as compared to the existing techniques. In terms of encoding efficiency, it outperformed the existing methodologies. The flexibility of the framework to adopt any combination of complexity reduction schemes makes it ideal to deploy in different situations. The proposed framework also works excellent in resource constraint environments like portable devices.

REFERENCES

[1] P. K. Podder, M. Paul, and M. Murshed, "Fast mode decision in the HEVC video coding standard by exploiting region with dominated motion and saliency features," *PLoS ONE*, vol. 11, no. 3, p. e0150673, 2016.

[2] S. Radicke, J.-U. Hahn, Q. Wang, and C. Grecos, "A parallel HEVC intra prediction algorithm for heterogeneous CPU+GPU platforms," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 103–119, Mar. 2016.

[3] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm, and T. Wiegand, *High Efficiency Video Coding (HEVC) Text Specification Draft 9*, document JCTVCK1003, ITU-T/ISO/IEC, Joint Collaborative Team on Video Coding, 2012.

[4] *Information Technology-Coding of Audio-Visual Objects Part 10. Advanced Video Coding*, Standard ISO/IEC, ISO/IEC 1449610, 2003.

[5] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[6] J. Ostermann *et al.*, "Video coding with H.264/AVC: Tools, performance, and complexity," *IEEE Trans. Circuits Syst. Mag.*, vol. 4, no. 1, pp. 7–28, 1st Quart., 2004.

[7] S. Momcilovic, N. Roma, and L. Sousa, "Multi-level parallelization of advanced video coding on hybrid CPU+GPU platforms," in *Proc. 10th Int. Workshop Algorithms*, 2012, pp. 165–174.

[8] M. Asif, M. Farooq, and I. A. Taj, "Optimized implementation of motion compensation for H.264 decoder," in *Proc. 5th Int. Conf. Comput. Sci. Converg. Inf. Technol. (ICCIT)*, Nov./Dec. 2010, pp. 216–221.

[9] 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, no. 7, pp. 688–703, Jul. 2003.

[10] L. Zhao, L. Zhang, S. Ma, and D. Zhao, "Fast mode decision algorithm for intra prediction in HEVC," in *Proc. Vis. Commun. Image Process. (VCIP)*, Nov. 2011, pp. 1–4.

[11] B. Jeon, *Fast Mode Decision for H.264 (ISO/IEC, 2003) ISO/IEC JTC1/SC29/WG11 and ITU-T SG16*, document JVT-J033, 2003.

[12] I. Choi, J. Lee, and B. Jeon, "Fast coding mode selection with rate-distortion optimization for MPEG-4 part-10 AVC/H.264," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 12, pp. 1557–1561, Dec. 2006.

[13] C. Grecos and M. Y. Yang, "Fast inter mode prediction for P slices in the H264 video coding standard," *IEEE Trans. Broadcast.*, vol. 51, no. 2, pp. 256–263, Jun. 2005.

[14] A. Saha, K. Mallick, J. Mukherjee, and S. Sural, "SKIP prediction for fast rate distortion optimization in H.264," *IEEE Trans. Consum. Electron.*, vol. 53, no. 3, pp. 1153–1160, Aug. 2007.

[15] X. Jing and L.-P. Chau, "An efficient inter mode decision approach for H.264 video coding," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jun. 2004, pp. 1111–1114.

[16] Y. H. Kim, J. W. Yoo, S. W. Lee, J. Shin, J. Paik, and H. K. Jung, "Adaptive mode decision for H.264 encoder," *Electron. Lett.*, vol. 40, no. 19, pp. 1172–1173, Sep. 2004.

[17] J. Bu, S. Lou, C. Chen, and J. Zhu, "A predictive block-size mode selection for inter frame in H.264," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2006, pp. 917–920.

[18] C.-H. Kuo, M. Shen, and C.-C. J. Kuo, "Fast inter-prediction mode decision and motion search for H.264," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jun. 2004, pp. 663–666.

[19] B. Feng, G.-X. Zhu, and W.-Y. Liu, "Fast adaptive inter-prediction mode decision method for H.264 based on spatial correlation," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2006, pp. 1804–1807.

[20] E. M. Enriquez, A. J. Moreno, and F. D. D. Maria, " An adaptive algorithm for fast inter mode decision in the H.264/AVC video coding standard," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 826–834, May 2010.

[21] J. Lee, S. Kim, K. Lim, H. J. Kim, and S. Lee, "Fast intermode decision algorithm based on general and local residual complexity in H.264/AVC," *EURASIP J. Image Video Process.*, vol. 2013, Dec. 2013, Art. no. 30. [Online]. Available: https://doi.org/10.1186/1687-5281-2013-30

[22] D. Zhu, Q. Dai, and R. Ding, "Fast inter prediction mode decision for H.264," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun. 2004, pp. 1123–1126.

[23] X. Jing and L.-P. Chau, "Fast approach for H.264 inter mode decision," *Electron. Lett.*, vol. 40, no. 17, pp. 1050–1052, Aug. 2004.

[24] D. Wu *et al.*, "Fast intermode decision in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 7, pp. 953–958, Jul. 2005.

[25] D. Wu, S. Wu, K. P. Lim, F. Pan, Z. G. Li, and X. Lin, "Block INTER mode decision for fast encoding of H.264," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2004, pp. 181–184.

[26] K. Bharanitharan, B.-D. Liu, and J.-F. Yang, "Classified region algorithm for fast intermode decision in H.264/AVC encoder," *EURASIP J. Adv. Signal Process.*, vol. 2010, Dec. 2016, Art. no. 150809. [Online]. Available: https://link.springer.com/article/10.1155/2010/150809

[27] H. Zeng, C. Cai, and K.-K. Ma, "Fast mode decision for H.264/AVC based on macroblock motion activity," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 4, pp. 491–499, Apr. 2009.

[28] Z. Liu, L. Shen, and Z. Zhang, "An efficient intermode decision algorithm based on motion homogeneity for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 1, pp. 128–132, Jan. 2009.

[29] F. Pan *et al.*, "Fast mode decision algorithm for intraprediction in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 7, pp. 813–822, Jul. 2005.

[30] R. Su, G. Liu, and T. Zhang, "Fast mode decision algorithm for intra prediction in H.264/AVC with integer transform and adaptive threshold," *J. Signal Image Video Process.*, vol. 1, no. 1, pp. 11–27, 2007.

[31] Z. Wei, H. Li, and K. N. Ngan, "An efficient intra mode selection algorithm for H.264 based on fast edge classification," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2007, pp. 3630–3633.

[32] H. Li, K. N. Ngan, and Z. Wei, "Fast and efficient method for block edge classification and its application in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 6, pp. 756–768, Jun. 2008.

[33] J.-C. Wang, J.-F. Wang, J.-F. Yang, and J.-T. Chen, "A fast mode decision algorithm and its VLSI design for H.264/AVC intra-prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 10, pp. 1414–1422, Oct. 2007.

[34] A. Elyousfi, A. Tamtaoui, and E. Bouyakhf, "A new fast intra prediction mode decision algorithm for H.264/AVC encoders," *Int. J. Comput. Syst. Sci. Eng.*, vol. 4, no. 1, pp. 89–95, 2008.

[35] B. La, M. Eom, and Y. Choe, "Dominant edge direction based fast intra mode decision in the H.264/AVC encoder," *J. Zhejiang Univ. Sci. A*, vol. 10, no. 6, pp. 767–777, 2009.

[36] A. Elyousfi, "Fast gravity direction-based ultra-fast intra prediction algorithm for H.264/AVC video coding," in *Signal Image and Video Processing (SIViP)*, vol. 7. New York, NY, USA: Springer, 2011, pp. 53–65. [Online]. Available https://link.springer.com/content/pdf/10.1007/s11760-011-0232-x.pdf

[37] K. Bharanitharan, J.-R. Ding, B.-W. Chen, and J.-F. Wang, "Selective intra block size decision and fast intra mode decision algorithms for H.264/AVC encoder," *IEICE Trans. Inf. Syst.*, vol. 95, no. 11, pp. 2720–2723, 2012.

[38] G. de Haan, P. W. A. C. Biezen, H. Huijgen, and O. A. Ojo, "True-motion estimation with 3-D recursive search block matching," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 5, pp. 368–379, Oct. 1993.

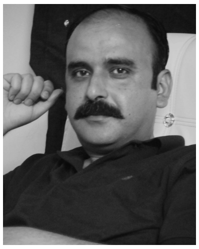
[39] M. Asif, A. I. Taj, S. M. Ziauddin, M. B. Ahmad, and M. Tahir, "An efficient scheme for intra prediction block size and mode selection in advanced video coding," in *Proc. Frontiers Inf. Technol. (FIT)*, Dec. 2015, pp. 211–215.

[40] M. Asif, A. I. Taj, S. M. Ziauddin, M. B. Ahmad, and A. Raza, "An efficient inter prediction mode selection scheme for advanced video coding based on motion homogeneity and residual complexity," *IEEJ Trans. Elect. Electron. Eng.*, vol. 11, no. 6, pp. 760–767, 2016.

[41] *H.264/AVC JM Reference Software Version 12.2*, Joint Video Term (JVT), Fraunhofer, Heinrich-Hertz-Institute, May 2015. [Online]. Available: http://iphome.hhi.de/suehring/tml/download/old_jm/

[42] G. Bjontegaard, *Calculation of Average PSNR Differences Between RD Curves*, document VCEG-M33, ITU-T SG16 Q6 Video Coding Experts Group (VCEG), Austin, TX, USA, Apr. 2001. [Online]. Available: http://wftp3.itu.int/av-arch/video-site/0104_Aus/VCEG-M33.doc

MUHAMMAD ASIF received the Ph.D. degree in electrical engineering from the Capital University of Science and Technology, Islamabad, Pakistan, in 2016. He has contributed over 15 research papers. His current research interests include image and video processing, parallel processing, embedded system optimization, and network security.



MAAZ BIN AHMAD received the Ph.D. degree in computer engineering from the Centre for Advanced Studies in Engineering, Islamabad, Pakistan, in 2014. He has authored over 15 research papers. His research interests include network security, video processing, and multimedia.

IMTIAZ A. TAJ received the M.Sc. and Ph.D. degrees in electronics and information engineering from Hokkaido University, Japan, in 2001. He is currently a Professor with the Department of Electrical of Engineering, Capital University of Science and Technology, Islamabad, Pakistan. He has authored over 40 research papers, including 20 in reputed international journals. His research interests include computer vision, image processing, video processing, pattern recognition, biometrics, and optics.

MUHAMMAD TAHIR received the B.S. degree in electrical engineering from UET Lahore, Pakistan, and the M.S. degree in computer engineering from LUMS, Lahore. He is currently pursuing the Ph.D. degree with the Capital University of Science and Technology, Islamabad, Pakistan. His areas of interest include video codec's, code optimization for embedded platforms, image processing, and digital design.

• • •