Fast ISP Mode Decision for the Versatile Video Coding Intra Prediction Using Machine Learning

Larissa Araújo
Video Technology Research Group (ViTech), Graduate Program in Computer Science (PPGC)
Federal University of Pelotas (UFPel)
Pelotas, Brazil
ldaaraujo@inf.ufpel.edu.br

Adson Duarte
Video Technology Research Group (ViTech), Graduate Program in Computer Science (PPGC)
Federal University of Pelotas (UFPel)
Pelotas, Brazil
airduarte@inf.ufpel.edu.br

Bruno Zatt
Video Technology Research Group (ViTech), Graduate Program in Computer Science (PPGC)
Federal University of Pelotas (UFPel)
Pelotas, Brazil
zatt@inf.ufpel.edu.br

Guilherme Correa
Video Technology Research Group (ViTech), Graduate Program in Computer Science (PPGC)
Federal University of Pelotas (UFPel)
Pelotas, Brazil
gcorrea@inf.ufpel.edu.br

Daniel Palomino
Video Technology Research Group (ViTech), Graduate Program in Computer Science (PPGC)
Federal University of Pelotas (UFPel)
Pelotas, Brazil
dpalomino@inf.ufpel.edu.br

ABSTRACT

The Versatile Video Coding (VVC) standard achieves high compression rates by introducing new encoding tools, such as the Intra Subpartition Prediction (ISP). However, the ISP increases the computational effort necessary to perform the mode decision of the intra prediction step. This paper proposes a fast intra-mode decision solution for the ISP using machine learning. A Decision Tree is employed to predict the most promising ISP modes to be optimal to avoid the costly RDO test of ISP modes that are less likely to be chosen. By reducing the number of modes fully evaluated by the RDO process, the proposed solution achieves an average time-saving of 3.15% with only 0.11% of coding efficiency loss when tested for the common test conditions of VVC. Unlike the related works, our solution avoids the time overhead of calculating image features by adopting features from the encoding process. Compared with related works, our solution presents competitive time-saving and coding efficiency results.

KEYWORDS

VVC, Intra Prediction, ISP, Machine Learning

1 INTRODUCTION

Digital videos have been fundamental in many areas, from entertainment and communication to surveillance applications and live broadcasts. A study reveals that during the third quarter of 2022, live streams featuring gaming-related content accumulated approximately 7.2 billion hours of content watched across leading streaming platforms [6]. In this context, video coding standards such as the Versatile Video Coding (VVC) [5] play a crucial role in enabling applications to manipulate high-definition videos for storage and transmission.

The VVC [5] is one of the latest and most advanced video coding standards. It offers superior bit-rate reduction without compromising visual quality when compared to its predecessor, the High Efficiency Video Coding (HEVC) standard [19]. This is possible due to several new encoding tools introduced in the standard, especially in the intra-prediction step of VVC. While VVC maintains the Planar, DC, and Angular directional modes from HEVC, it extends the number of Angular modes from 33 to 65. VVC also introduces the Matrix-weighted Intra Prediction (MIP) [18] and the Intra Subpartition Prediction (ISP) [8] to improve prediction accuracy. Combined with the Planar, DC, and Angular modes, the ISP tool enhances prediction granularity by processing a block through subpartitions in the horizontal or vertical directions. Despite the coding efficiency improvements introduced by the new encoding tools of VVC, there is a trade-off in encoding time, which makes it 34 times slower than HEVC [12]. Therefore, it is important to develop solutions to improve the encoding time by targeting the new intra-prediction tools in VVC, such as the ISP tool.

Some related works propose solutions to save time in the intra-mode decision in VVC, specifically focusing on the ISP tool. The main idea of these works is to avoid the costly evaluation performed by the Rate-Distortion Optimization (RDO) process for ISP modes that are less likely to be optimal. The works usually use heuristics or machine learning solutions to predict the most promising modes. For instance, in [14], a machine learning model is trained with a key feature computed over the image known as the Mean Absolute Sum of Transform coefficients. The model predicts whether the evaluation for each ISP mode is necessary or can be skipped. Another approach, presented in [17], involves training a Decision Tree model over another image feature, the block's variance. The Decision Tree predicts when the evaluation of all ISP candidates can be skipped. In [11], the texture complexity of the block is obtained through an image feature called Mean Absolute Deviation.

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2024). Juiz de Fora, Brazil. Porto Alegre: Brazilian Computer Society, 2024.

ISSN 2966-2753

Then, this feature is used to decide whether the evaluation of ISP candidates can be skipped. A heuristic algorithm is proposed in [13], where the list of ISP candidates is pre-pruned according to the shape of the block and the ISP subpartition direction.

Although all these works report time-saving results on the ISP mode decision, the solutions proposed by [14] and [11] rely on computing image features, inevitably introducing a processing time overhead to each one of these mode decision solutions — a fact that is not thoroughly discussed in these works. The solution proposed by [13] is the only one that does not rely on computing image features. Nevertheless, none of the related works employ solutions using features available at encoding time, such as the modes of neighbor blocks or rate-distortion costs calculated during the RDO process. In this work, we call these features "**encoding features**" since they are available during the encoding process.

This paper proposes a fast ISP mode decision solution using machine learning for the VVC intra prediction process. Our approach distinguishes from related works by using features available at encoding time, aka encoding features, as input for the machine learning model. This approach avoids the computations necessary to calculate image features as it is adopted by most of the related works. We classify the ISP candidates into two distinct classes based on their associated intra modes: ***(i) ISP Planar/DC***, comprising ISP subpartitions associated with Planar/DC modes, and ***(ii) ISP Angular***, comprising ISP subpartitions associated with Angular modes. Then, we train a Decision Tree model to predict between these two ISP classes using encoding features, such as rate-distortion costs for various intra-modes and neighbor decisions, which are accessible before the ISP mode decision. Leveraging these features allows our solution to avoid ISP modes that are less likely optimal, saving encoding time with negligible coding efficiency loss.

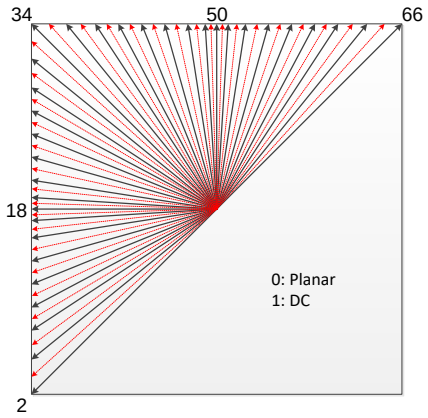


Figure 1: Intra modes in VVC.

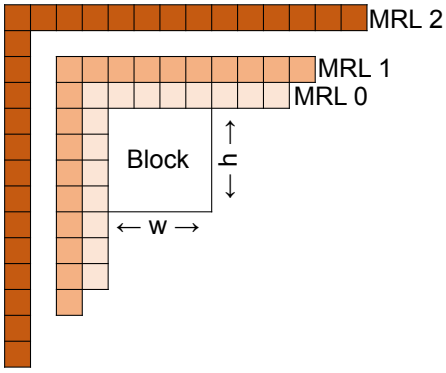


Figure 2: Multiple Reference Line tool.

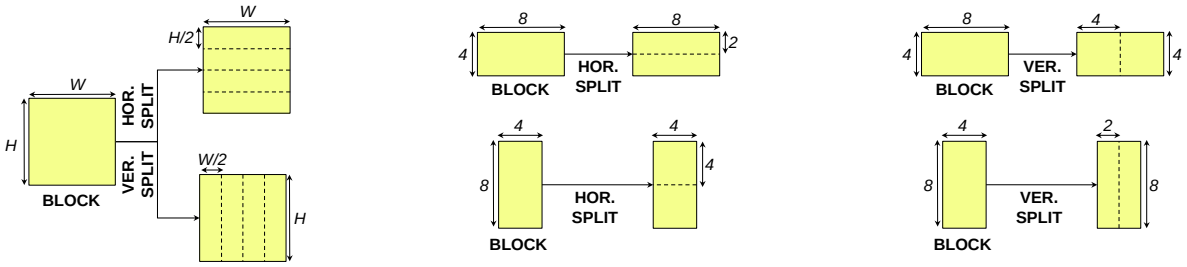
2 VVC INTRA SUBPARTITION PREDICTION

VVC introduced several innovations in the intra prediction. Firstly, it enabled the prediction of video frames through blocks of square and rectangular shapes, incorporating 17 block sizes. Considering the intra modes, the Planar and DC modes were preserved from the previous HEVC standard, while VVC expanded the Angular modes from 33 to 65, which are indicated by the red arrows in Figure 1. VVC also introduced a novel family of intra modes called MIP [18]. Alongside these, new tools that can be combined with the Planar, DC, and Angular modes were incorporated, such as the Multiple Reference Line (MRL) [7], which extends the number of available reference samples for intra prediction and is showcased in Figure 2, and the ISP [8], shown in Figure 3, which is the primary focus of this work.

The ISP tool enables more granular block prediction. For this purpose, the ISP tool horizontally or vertically divides the original image block into two or four subpartitions, depending on the block size, as illustrated in Figure 3. For blocks sized 8x4 and 4x8, the ISP tool generates only two subpartitions, either in the horizontal or vertical direction, as shown in Figure 3(b) and Figure 3(c), respectively. This restriction ensures that each subpartition contains at least 16 samples. For other block shapes, the ISP tool generates four subpartitions in either the horizontal or vertical direction, as shown in Figure 3(a). The prediction process for each subpartition occurs sequentially, and samples generated from the prediction of one subpartition are used as reference samples for the prediction of the next subpartitions.

While predictions are conducted individually for each subpartition, they must converge to the same intra mode, such as Planar, DC, or one of the Angular modes. This is done to improve the prediction accuracy of each subpartition and also to save the bits necessary to signal the intra modes in the bitstream since only one mode will be signaled. The addition of the ISP can improve coding efficiency by approximately 0.57% with a 12% increase in encoding time [8]. This increase in encoding time occurs because the ISP tool introduces an additional step in the intra-mode decision process of VVC. In this process, the encoder evaluates the Planar, DC, and Angular modes twice. First, a list of these modes is evaluated for the entire block by the RDO, and then, they are evaluated again for each possible ISP subpartition.

The intra-mode decision process determines the best intra mode for each block by evaluating several possible combinations through the Rate-Distortion Optimization (RDO) process [20]. However, this process is computationally intensive since the encoder must evaluate many intra-mode candidates. For each one of these modes,



(a) Four subpartitions. **(b) Two horizontal subpartitions.** **(c) Two vertical subpartitions.**

Figure 3: Intra Subpartitions Prediction Tool.

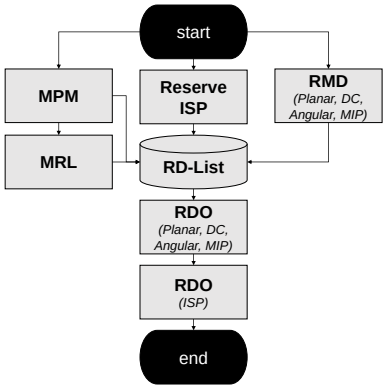


Figure 4: VTM standard intra mode decision.

the rate-distortion cost must be computed by the intra-mode decision. This cost is available only after the encoding steps, including prediction, direct and inverse transformation and quantization, and entropy coding. To handle this complexity, the reference software that implements VVC, called VVC Test Model (VTM) [4], incorporates the Rough Mode Decision (RMD) [21] and the Most Probable Modes (MPM) list [16]. The main idea behind the RMD and MPM steps is to generate the RD-List, a subset of the most promising modes. Only this subset is evaluated by the RDO process, avoiding the evaluation of modes that are less likely to be optimal.

VTM performs the intra-mode decision following the steps of Figure 4. The RMD, MPM, and MRL steps jointly select up to eight intra modes to compose the RD-List. Both the RMD and MRL steps compute fast rate-distortion costs for the intra modes, selecting the six best ones. However, while the RMD evaluates the Planar, DC, Angular, and MIP modes with MRL set to zero, as shown in Figure 2, the MRL step evaluates the MPM modes twice. One with the MRL set to one and then with MRL set to two. The MPM step generates a list of six intra modes likely to be optimal for the current block based on the best intra modes in neighboring blocks [16]. The first MPM mode is always the Planar mode, while the remaining ones can be the DC or one of the Angular modes. If the first two MPM modes are not already in the RD-List, the MPM step adds them. At the end of the RD-List, the encoder reserves 16 positions evenly distributed between the horizontal and vertical subpartitions for ISP. VTM includes the same intra modes obtained from the RMD and MPM steps in the horizontal and vertical reserved positions, excluding the MIP modes. In addition, three Angular modes with the lowest costs during the RMD step, excluding the ones already in the RD-List, are also selected for the ISP evaluation. Then, the RDO evaluates all non-ISP modes (Planar, DC, Angular, and MIP modes) and only then starts the evaluation of the ISP modes.

While the RD-List is a subset of the most promising intra modes, only one will yield the best result in terms of coding efficiency. Even when we consider only the subset of ISP candidates present in the RD-List, the RDO must evaluate up to 16 modes for a single block to decide the best ISP mode. In this context, there is a need for solutions that target reducing the number of modes to be evaluated by the RDO process. The solutions must accurately predict the most promising ISP modes to reduce the computational effort required for the ISP mode decision with minimal loss of compression efficiency.

2.1 ISP Occurrence Rate Analysis

We performed two analyses to evaluate the occurrence rate of the Planar, DC, and Angular modes during the ISP step. The idea is to build our fast ISP mode decision solution based on the occurrence rate of each type of intra mode during the ISP process. We organized the ISP intra modes in two different classes: ***(i) ISP Planar/DC*** and ***(ii) ISP Angular***. The reason for this classification relies on the nature of each intra mode. While the Planar and DC modes are better for predicting homogeneous textures, the Angular modes are better for predicting directional textures.

In the first analysis, we computed the occurrence rate of each class organized by block size. The occurrence rate of a given class measures how often an intra mode within that class is the best when the final decision is an ISP. In the second analysis, we computed the frequency of Angular ISP candidates in the RD-List to understand the potential to avoid evaluating Angular ISP modes. We selected the same 15 videos used in the work of [9] for both analyses in this study. These videos were specifically chosen by the authors for their variety in motion and texture characteristics, as indicated by the Spatial Information (SI) and Temporal Information (TI) metrics [10]. Each video was encoded using VTM 18.0 [4] with the *All Intra* configuration and four Quantization Parameters (QP): 22, 27, 32, and 37. Subsequently, for each block, we extracted the final mode decision (ISP Planar/DC or ISP Angular) and the number of ISP candidates in the RD-List during encoding. To achieve this, we

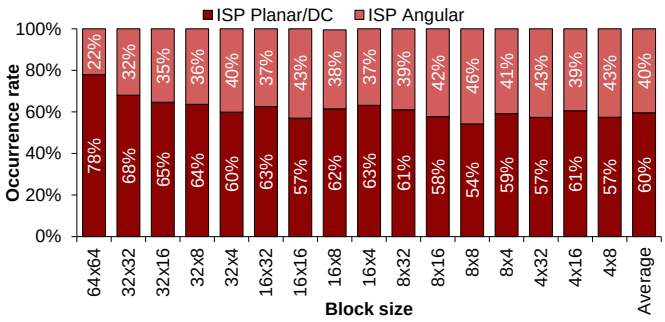


Figure 5: Occurrence rate of ISP Planar/DC and ISP Angular classes by block size.

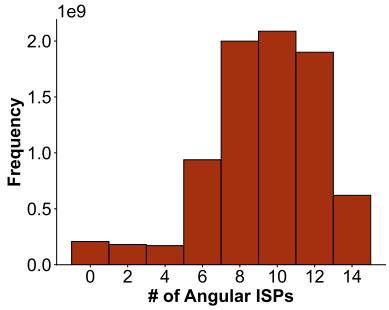


Figure 6: Frequency of angular ISPs in the RD-List.

modified the VTM code and inserted routines at specific points to collect the necessary data.

Figure 5 presents the occurrence rate in which the ***(i) ISP Planar/DC*** and ***(ii) ISP Angular*** classes yield the best rate-distortion result by block size. We can observe that the ***(i) ISP Planar/DC*** class has a higher occurrence rate when compared to the ***(ii) ISP Angular*** class for larger block sizes, particularly for 64x64 and 32x32. Furthermore, the ***(i) ISP Planar/DC*** class maintains a higher occurrence rate across all block sizes, obtaining 60% of the occurrence rate on average. In other words, when the ISP is chosen for prediction, the Planar and DC modes are more likely to be chosen.

In the second analysis in Figure 6, we can see the frequency of ISP candidates associated with Angular modes. We can observe that the number of ISP candidates associated with angular modes is always even. That happens because the VTM software evaluates the same intra modes for both horizontal and vertical subpartitions in the ISP. Besides, it is possible to notice that in most cases, there are 8, 10, and 12 ISP candidates associated with angular modes. This means that, in most cases, VTM will evaluate 8 to 12 ISP candidates associated with angular modes for a single block. This finding highlights the potential for reducing the computational effort of the intra mode decision in VVC by early predicting when the encoder can avoid evaluating the ISP candidates associated with the Angular modes.

Considering the high occurrence rate of the ***(i) ISP Planar/DC*** class and the high frequency of the Angular ISP candidates in the RD-List, it is possible to reduce the computational effort of the ISP mode decision if an accurate machine learning model is employed to predict when the best ISP mode is Planar or DC. When this

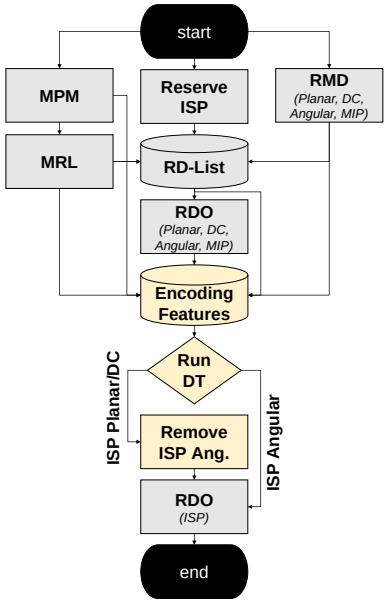


Figure 7: Proposed fast ISP mode decision solution.

happens, the RDO evaluation of many Angular ISP candidates can be skipped to save encoding time while minimizing the final coding efficiency loss.

3 FAST ISP MODE DECISION

This paper proposes a fast ISP mode decision using machine learning for the VVC standard. The main idea is to use the encoding features available at encoding time, thereby avoiding the additional overhead of computing image features. Based on the two analyses previously presented, our solution groups the ISP candidates according to their associated intra modes in two classes: ***(i) ISP Planar/DC*** class, containing the ISP candidates associated with the Planar and DC modes, and ***(ii) ISP Angular*** class, containing the ISP candidates associated with Angular modes. The goal was to ensure that the ISP Planar/DC class contains ISP candidates associated with Planar and DC modes, which are known to be effective at predicting homogeneous blocks, while the ISP Angular class includes ISP candidates associated with Angular modes, which are effective at predicting blocks with directional textures. After defining the classes, we train a Decision Tree offline using encoding features to predict between these two classes. Once training is complete, we integrate the final Decision Tree into the VTM. Whenever the Decision Tree predicts the ***(i) ISP Planar/DC*** class, the ISP candidates belonging to the ***(ii) ISP Angular*** class are not evaluated, saving encoding time. The choice for a Decision Tree model is justified for two main reasons. First, Decision Trees usually present good prediction accuracy for tabular datasets. Second, Decision Trees have a fast inference time. As our solution aims to save encoding time, we can not employ a complex model.

Figure 7 presents our solution, where the yellow rectangles represent the new steps introduced by the solution. Since the default intra-mode decision of VTM evaluates all non-ISP modes (Planar, DC, Angular, and MIP modes) and only then starts evaluating the

ISP modes, we added our solution between the RDO for the non-ISP modes and the RDO for the ISP modes. This way, the default decision is the same until the RDO evaluation of the non-ISP modes.

Then, the Decision Tree (DT) acquires a series of encoding features from the RMD, the MPM, the RD-List, and the RDO for non-ISP modes to predict one of the previously mentioned ISP classes. When the Decision Tree outcome is the ***(i) ISP Planar/DC*** class, our solution removes the ISP candidates associated with Angular modes from the RD-List, and the RDO is performed only for ISP candidates associated with Planar/DC modes. On the other hand, if the Decision Tree outcome is the ***(ii) ISP Angular*** class, all ISP candidates are evaluated. We decided never to remove the ISP candidates associated with Planar/DC modes to bring a balance between the time reduction and the coding efficiency loss, given the high occurrence rate of the ISP Planar/DC class presented in Figure 5.

3.1 Feature Extraction and Dataset Generation

Table 1 presents the encoding features extracted from VTM that were used in our Decision Tree model. The features are presented by name, description, the step where the feature is extracted from (RMD, MPM, RD-List or RDO), the type (Boolean, Decimal or Integer), the *min* and *max* values for the feature, and the number of values that the feature provides. RMD computes fast rate-distortion costs for the Planar, DC, Angular, and MIP modes. These fast costs hint at whether the ISP Planar/DC or ISP Angular classes will produce the best cost. This way, considering the best costs in RMD for the Planar, DC, Angular, and MIP modes, we extract the Sum of Absolute Differences (SAD), the Sum of Absolute Transformed Differences (SATD), the estimated number of bits, and the fast rate-distortion costs. Besides that, we also extract the x and y positions of the block, the best angular and MIP modes, and the best angular and DC MRL [7] numbers.

The MPM provides a list containing six intra modes. The first one is always the Planar; the remaining modes will be the DC or one of the Angular modes. These six intra modes are likely the best ones since they were the best for the left and upper neighboring blocks. Therefore, from the MPM, we extract the number of the five selected intra modes, excluding the Planar as it is constant, the number of the best intra modes in the left and upper neighboring blocks, eight boolean values indicating whether the best intra mode for the left and upper neighboring blocks is Planar, DC, Angular, or MIP, and a boolean value indicating if the DC is an MPM.

The VTM software distributes the non-ISP modes in the RD-List in ascending order according to their fast rate-distortion costs obtained from the RMD step. This way, the modes distribution order in the RD-List can also hint if the ISP Planar/DC or ISP Angular classes are likely to provide the best rate-distortion cost. Therefore, from the RD-List, we extract the position of the first occurrence of Planar, DC, Angular, and MIP modes and also the mode number of the first Angular and MIP modes.

Finally, since the RDO evaluation for non-ISP modes occurs before the evaluation of the ISP modes, all the complete rate-distortion costs computed by the RDO for non-ISP modes are available. This way, we extract the best rate-distortion costs obtained by the RDO step for the Planar, DC, Angular, and MIP modes. It is important to highlight that despite the high number of features used by our model, there is no need for additional computations since they are all available during the encoding process.

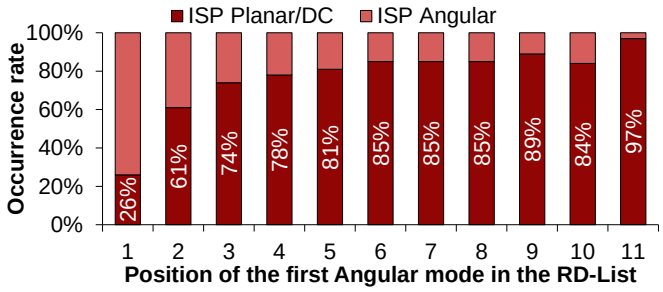


Figure 8: Occurrence rate of ISP classes according to the position of the first Angular mode in the RD-List.

To obtain the 48 features presented in Table 1 for training the Decision Tree, we encoded the same 15 videos using the setup described in section 2.1. The dataset contains approximately 800,000 samples balanced by block size, QP, video, and class. Because there is a single dataset for all block sizes, we normalized the features related to rate-distortion costs according to Equation (1), where X is the set containing the rate-distortion cost-related features, namely the SAD, SATD, FracBits, RMD Cost, and RDO Cost in Table 1, w is the width of the block and h is the height of the block.

$$x\_norm = \frac{x}{w \cdot h}, \quad x \in X, \quad w, h \in \{4, 8, 16, 32, 64\} \tag{1}$$

To analyze the behavior of some of the features, we computed the information gain for each one of the encoding features considered in this work. Subsequently, the two features with the highest information gains were selected to be analyzed, which are (1) the position of the first Angular mode in the RD-List and (2) the intra mode that appears in the second position of the MPM list. The analysis of these two encoding features was conducted by computing the occurrence rate of the ISP Planar/DC and ISP Angular classes according to the feature's respective values.

Figure 8 illustrates the occurrence rate of the ISP Planar/DC and ISP Angular classes according to the position of the first Angular mode in the RD-List. One can notice that the first Angular mode occurs from the first to the eleventh position in the RD-List. When the first Angular mode occurs in the first position of the RD-List, the ISP Planar/DC exhibits an occurrence rate of 26%, indicating that in 74% of the cases, the ISP Angular class results in the best rate-distortion cost. However, as the position of the first Angular mode in the RD-List increases, the ISP Planar/DC class occurrence rate also increases. For instance, when the first Angular mode in the RD-List occurs from the second to the eleventh position, the ISP Planar/DC class always achieves a higher occurrence rate, peaking at 97% when the first Angular mode occurs in the eleventh position of the RD-List. In other words, as the position of the first Angular mode in the RD-List increases, it also increases the cases where VTM can avoid the RDO evaluation of the ISP Angular modes.

In Figure 9, the occurrence rate of the ISP Planar/DC and ISP Angular classes is shown according to the intra mode present in the second position of the MPM List. Since the second position of the MPM list can contain any intra mode among DC and Angular,

Table 1: Encoding features extracted from VTM.

Name	Description	Step	Type	Min	Max	# of values
BlockPosition	The x and y block positions.	RMD	Integer	0	4096	2
BestAngular	Best angular mode.	RMD	Integer	2	66	1
BestMIP	Best MIP mode.	RMD	Integer	0	7	1
MRLAngular	MRL reference line from the best angular mode .	RMD	Integer	0	2	1
MRLDC	MRL reference line from the best DC mode.	RMD	Integer	0	2	1
ModesMPM	MPM modes excluding Planar.	MPM	Integer	1	66	5
ModesPosition	Position of the first occurrence of each type of intra mode.	RD-List	Integer	1	11	4
FirstAngular	First angular mode in the RD-List.	RD-List	Integer	2	66	1
FirstMIP	First MIP mode in the RD-List.	RD-List	Integer	0	7	1
NeighborMode	Best intra mode number in the neighboring blocks.	MPM	Integer	0	66	2
NeighborType	Best intra mode type in the neighboring blocks.	MPM	Boolean	0	1	8
DCMPM	DC mode is an MPM.	MPM	Boolean	0	1	1
SAD	Sum of Absolute Differences.	RMD	Decimal	0.63	1390.38	4
SATD	Sum of Absolute Transformed Differences.	RMD	Decimal	0.52	497.81	4
FracBits	Estimated number of bits.	RMD	Decimal	1.19	15669.28	4
RMD Cost	Fast rate-distortion cost.	RMD	Decimal	0.65	502.01	4
RDO Cost	Complete rate-distortion cost.	RDO	Decimal	182.14	1881715.88	4
					Total	**48**

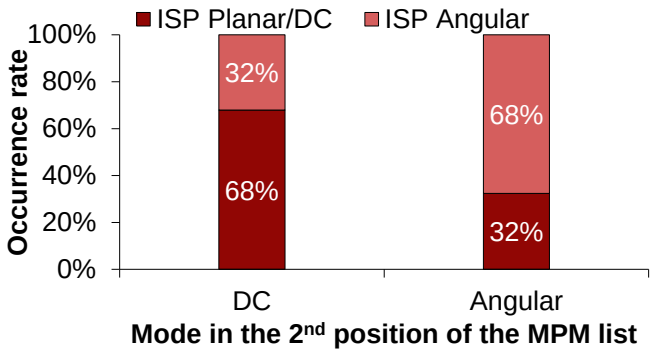


Figure 9: Occurrence rate of ISP classes according to the intra mode in second position on the MPM list.

we grouped the values of this feature into two categories: DC, with cases where the second position of the MPM list has the DC mode, and Angular, with cases where the second position of the MPM list has one of the Angular modes (modes 2 to 66 in Figure 1).

From Figure 9, one can see that when the second position of the MPM list has the DC mode, the ISP Planar/DC class achieves a 68% occurrence rate. This means that in only 32% of the cases, the ISP Angular class achieves the best rate-distortion cost. In contrast, when the second position of the MPM list has one of the Angular modes, the opposite occurs, and the ISP Angular class achieves a higher occurrence rate of 68%. Therefore, the analysis of this feature reveals that when the DC mode is the second in the MPM list, the ISP Planar/DC class has a higher chance of containing the best rate-distortion cost. As a result, when the DC mode is the second in the MPM list, VTM can avoid the RDO evaluation of the ISP Angular modes in 68% of the cases.

In summary, the analysis of these two features reveals the importance of encoding features in predicting when the RDO evaluation of the ISP Angular modes can be avoided by VTM. For instance, Figure 8 shows that when the first Angular mode in the RD-List occurs from the second to the eleventh position, VTM can avoid the RDO evaluation of the ISP Angular modes most of the time, given the higher occurrence rate of the ISP Planar/DC class. Specifically, when the first Angular mode occurs in the ninth, tenth, or eleventh position, VTM can avoid the RDO evaluation of the ISP Angular modes in 89%, 84%, and 97% of the cases, respectively. Similarly, Figure 9 demonstrates that when the DC mode is in the second position of the MPM list, VTM can avoid the RDO evaluation of the ISP Angular modes in 62% of the cases.

Although only the two features with the highest information gains were analyzed, similar behavior is expected from the remaining features, such as the positions of the first occurrence of the Planar, DC, and MIP modes, the rate-distortion costs associated with the best Planar, DC, Angular, and MIP modes, and the intra mode that occurs at each position of the MPM list. Since all these encoding features are already computed by VTM and available at encoding time, there is no additional overhead from computing features from the image, as is common in most related works.

3.2 Decision Tree Training

Using the Scikit-learn library [15], the dataset was split into 80% for training and validation, reserving the remaining 20% for testing. This division ensured that our model never saw 20% of the data throughout the training and validation stages. We performed the training and validation with two steps: a Random Search and a Grid Search. The Random Search [1] step involved the evaluation of 1,000 random combinations across a wide search space over the hyperparameters *criterion*, *min samples split*, *min samples leaf*, *max depth*, *max leaf nodes*, and *max features*. Each combination was

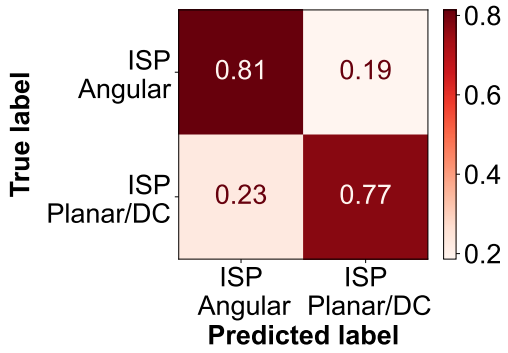


Figure 10: Confusion matrix for the final model in the test set.

evaluated using a 5-fold cross-validation approach over 80% of the data separated for the training and validation stages. Subsequently, we employed the Random Search results to compute the Pearson Correlation between each hyperparameter and the F1-score.

The *max leaf nodes* and the *max features* were the two hyperparameters with the highest correlation with an increased F1-score. Then, these two hyperparameters went for a refining phase in the Grid Search step, while the remaining stayed at their default values or the best values found in the Random Search. The Grid Search also considered the evaluation of combinations through a 5-fold cross-validation over the same 80% of the data separated for training and validation stages. The final model is obtained from the hyperparameter combination that generated the best F1-score in the Grid Search.

Figure 10 presents the confusion matrix obtained by the final Decision Tree model when evaluated under the 20% of the data not used during the Random Search and Grid Search steps, reserved for testing purposes. In the main diagonal, we have the accurate predictions made by the model, while down and above the main diagonal, we have the wrong predictions made by the model. The final model obtained accuracies of 77% and 81% for the ISP Planar/DC and ISP Angular classes, respectively.

Since we apply the Decision Tree model in the context of a solution to reduce the encoding time in video coding, we can classify the wrong predictions made by the model into two categories: **time errors** and **coding efficiency errors**. A **time error** happens when the model misclassifies an example of the ISP Planar/DC class in the ISP Angular class. There is no coding efficiency loss when that happens because our solution will not remove the ISP candidates associated with the Planar/DC modes from the RD-List. However, a **time error** occurs since the RDO evaluation for the ISP modes could be performed exclusively for the ISP candidates associated with the Planar/DC modes. On the other hand, a **coding efficiency error** happens when the model misclassifies an example belonging to the ISP Angular class in the ISP Planar/DC class. Then, our solution removes the ISP candidates associated with the Angular modes, reducing the encoding time but providing a loss of coding efficiency. By looking at Figure 10, **time errors** occur 23% of the time, while **coding efficiency errors** occur 19% of the time. Given the negligible difference between the two types of errors, these results highlight the model's effectiveness in balancing the trade-off between the time reduction and the loss of coding efficiency.

4 EXPERIMENTAL RESULTS

To evaluate the performance of our solution, we followed the Common Test Conditions (CTC) [3] of VVC, where we encoded 22 video sequences with the *All Intra* configuration and the QP values 22, 27, 32, and 37, both in the anchor VTM 18.0 and in the modified VTM 18.0 with our solution. The modified VTM 18.0 includes the final Decision Tree model, which is integrated to generate predictions for all processed blocks during encoding. We encoded the videos sequentially on a dedicated server with an Intel® Core™ i7-8700K processor and 16GB of RAM. **None** of the videos from the CTC of VVC were used to train the Decision Tree. Therefore, we evaluated our solution using videos that our model never saw. To obtain the performance of our solution, we calculate two metrics: the time-saving (TS), obtained by comparing the encoding time of the anchor with our solution, and the coding efficiency, measured in terms of BDBR [2], which calculates the bit-rate variation between two encoders considering the same visual quality.

Table 2 presents the results regarding Class, Video, TS, and BDBR. The classes are defined according to the Common Test Conditions (CTC) of VVC [3] and indicate the resolution of the videos. Videos in classes A1 and A2 have 4K resolution, class B videos have 1080p resolution, class C videos have 480p resolution, class D videos have 240p resolution, and class E videos have 720p resolution.

Our solution obtained a time saving of 3.15% with only 0.11% of coding efficiency loss on average. The best result is observed for the *Campfire* video sequence, with a time saving of 5.18% and a coding efficiency loss of 0.01%. This video achieved the best result because it has a simple texture containing many homogeneous areas, which is appropriate for the ISP Planar/DC class. As a result, our model predicts the ISP Planar/DC class more often, avoiding evaluating the ISP Angular class. Conversely, the *FourPeople* video sequence presented the worst result, with a time saving of 1.79% and a coding efficiency loss of 0.11%. This video has a more complex texture, presenting many edges that are more suitable for the ISP Angular class. Consequently, our model predicts this class more frequently, reducing the time-saving potential.

For high-definition videos, such as those in classes A1, A2, and B, our solution achieves the highest time saving results, as shown by the averages for these specific classes in Table 2. For instance, in the A1 class, which includes 4K resolution videos, our solution obtains the highest average time saving of 4.24% with only a 0.03% loss in coding efficiency. Similarly, in classes A2 and B, which contain 4K and 1080p video resolutions, our solution achieves the second and third highest average time saving of 3.48% and 3.55%, respectively, with only a 0.06% and 0.11% loss in coding efficiency. These results are extremely important because high-definition videos require the most significant encoding times. In other words, our solution effectively saves time in the encoding of videos where it is most critical while introducing only minor losses in coding efficiency.

To summarize, our solution demonstrates the capability to reduce encoding time with minimal loss in coding efficiency across all evaluated video sequences. This is possible through a Decision Tree model, which leverages features directly extracted from the

Table 2: Time-saving and coding efficiency results.

Class	Video	TS	BDBR
A1	Tango2	4.07%	0.05%
	FoodMarket4	3.48%	0.02%
	CampFire	5.18%	0.01%
A2	CatRobot	2.79%	0.09%
	DaylightRoad2	3.30%	0.06%
	ParkRunning3	4.36%	0.04%
B	MarketPlace	5.06%	0.05%
	RitualDance	3.80%	0.13%
	Cactus	4.25%	0.13%
	BasketballDrive	2.24%	0.14%
	BQTerrace	2.40%	0.09%
C	RaceHorsesC	4.69%	0.14%
	BQMall	1.92%	0.21%
	PartyScene	3.06%	0.14%
	BasketballDrill	2.08%	0.14%
D	RaceHorses	3.13%	0.05%
	BQSquare	2.21%	0.14%
	BlowingBubbles	3.82%	0.14%
	BasketballPass	1.79%	0.16%
E	FourPeople	1.49%	0.18%
	Johnny	2.10%	0.19%
	KristenAndSara	2.19%	0.13%
	Average (A1)	**4.24%**	**0.03%**
	Average (A2)	**3.48%**	**0.06%**
	Average (B)	**3.55%**	**0.11%**
	Average (Overall)	**3.15%**	**0.11%**

encoding process, thereby circumventing the need for additional computations to generate input features.

Table 3 compares our solution with related works, considering the software version, time saving (TS), BDBR, and the TS/BDBR ratio, representing the trade-off between time saving and coding efficiency loss. The solutions proposed by both [17] and [13] obtained better results in terms of time saving. However, they also obtained a higher coding efficiency loss than our solution. Specifically, [17] demonstrated a smaller TS/BDBR trade-off than our solution, while [13] showed a trade-off similar to ours. On the other hand, the solutions presented by [14] and [11] achieve a similar BDBR when compared to our solution while providing greater time saving and TS/BDBR trade-off.

It is essential to observe that the impact of the time spent computing the input features used by the solutions in [14]-[11] is not discussed. These include calculating image features such as block variance, mean absolute deviation (which entails summing the differences between the value of each luminance sample and the mean of all samples), and the mean absolute sum of transform coefficients. The latter involves the summation of all transformed coefficients within the block.

Table 3: Comparison with related works.

Solution	Software	TS	BDBR	TS/BDBR
Our	**VTM 18.0**	**3.15%**	**0.11%**	**28.64**
Park [14]	VTM 11.0	7.20%	0.08%	90.00
Saldanha [17]	VTM 10.0	8.32%	0.31%	26.84
Liu [11]	VTM 08.0	7.00%	0.09%	77.78
Park [13]	VTM 09.0	12.11%	0.43%	28.16

Considering the complexity of calculating such features, the time-saving results obtained by these works might vary significantly on different computing systems. For instance, many of these image features could be calculated in parallel with the encoding process on systems with an embedded GPU. However, if a GPU is unavailable, the image features should be calculated sequentially before the proposed ISP mode decisions, adding significant computations and reducing the potential for time-saving. Unlike these related works, our solution circumvents time overhead by exclusively utilizing features derived from the encoding process, achieving competitive results even without the power of the image features. To the best of our knowledge, our work is the first in its approach to incorporating encoding features to minimize the number of evaluated ISP candidates in the intra mode decision of VVC.

5 CONCLUSION

This paper presented a fast ISP mode decision solution using machine learning for the VVC standard. An analysis of ISP mode occurrence revealed the prevalence of Planar and DC modes when the ISP tool is used. In contrast, we found a high frequency of many Angular modes in the candidate modes list for ISP. From these findings, we decided to group the ISP modes into two classes according to their associated intra mode: ISP Planar/DC and ISP Angular. Then, we employed a Decision Tree trained with encoding features to predict between these classes. Whenever the Decision Tree predicts the ISP Planar/DC class, our solution avoids evaluating the modes in the ISP Angular class, reducing the encoding time. The experimental results demonstrate the effectiveness of the proposed solution in reducing encoding time while preserving coding efficiency. Compared to related works, our solution presents competitive results and avoids additional computations to generate the input features for the proposed fast mode decision solution.

ACKNOWLEDGMENTS

The authors thank FAPERGS, CNPq and CAPES (Finance Code 001) Brazilian research support agencies that financed this investigation.

REFERENCES

[1] James Bergstra and Yoshua Bengio. 2012. Random search for hyper-parameter optimization. *Journal of machine learning research* 13, 2 (2012), 281–305. https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf

[2] Gisle Bjontegaard. 2001. Calculation of average PSNR differences between RD-curves. https://www.itu.int/wftp3/av-arch/video-site/0104_Aus/VCEG-M33.doc VCEG-M33.

[3] Frank Bossen, Jill Boyce, Karsten Sühring, Xiang Li, and Vadim Seregin. 2020. VTM common test conditions and software reference configurations for SDR video. Retrieved Aug 22, 2024 from https://jvet-experts.org/doc_end_user/current_document.php?id=10545 JVET-T2010-v1.

[4] Frank Bossen, Karsten Suehring, and Xiang Li. 2018. VTM reference software for VVC. Retrieved Aug 22, 2024 from https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM

[5] Benjamin Bross, Ye-Kui Wang, Yan Ye, Shan Liu, Jianle Chen, Gary J. Sullivan, and Jens-Rainer Ohm. 2021. Overview of the Versatile Video Coding (VVC) Standard and its Applications. *IEEE Transactions on Circuits and Systems for Video Technology* 31, 10 (2021), 3736–3764. https://doi.org/10.1109/TCSVT.2021.3101953

[6] L. Ceci. 2023. Live streaming - Statistics & Facts. Retrieved Jun 20, 2023 from https://www.statista.com/topics/8906/live-streaming/#topicOverview

[7] Yao-Jen Chang, Hong-Jheng Jhu, Hui-Yu Jiang, Liang Zhao, Xin Zhao, Xiang Li, Shan Liu, Benjamin Bross, Paul Keydel, Heiko Schwarz, Detlev Marpe, and Thomas Wiegand. 2019. Multiple Reference Line Coding for Most Probable Modes in Intra Prediction. In *2019 Data Compression Conference (DCC)*. IEEE, Snowbird, UT, USA, 559–559. https://doi.org/10.1109/DCC.2019.00071

[8] Santiago De-Luxán-Hernández, Valeri George, Jackie Ma, Tung Nguyen, Heiko Schwarz, Detlev Marpe, and Thomas Wiegand. 2019. An Intra Subpartition Coding Mode for VVC. In *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, Taipei, Taiwan, 1203–1207. https://doi.org/10.1109/ICIP.2019.8803777

[9] Adson Duarte, Bruno Zatt, Guilherme Correa, and Daniel Palomino. 2023. Fast Intra Mode Decision Using Machine Learning for the Versatile Video Coding Standard. In *2023 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, Monterey, CA, USA, 1–5. https://doi.org/10.1109/ISCAS46773.2023.10181769

[10] International Telecommunication Union (ITU). 2023. Subjective video quality assessment methods for multimedia applications. Retrieved Aug. 22, 2024 from https://www.itu.int/rec/T-REC-P.910-202310-I/en

[11] Zhi Liu, Mengjun Dong, Xiao Guan, Mengmeng Zhang, and Ruoyu Wang. 2021. Fast ISP coding mode optimization algorithm based on CU texture complexity for VVC. *EURASIP Journal on Image and Video Processing* 2021 (07 2021). https://doi.org/10.1186/s13640-021-00564-4

[12] Alexandre Mercat, Arttu Mäkinen, Joose Sainio, Ari Lemmetti, Marko Viitanen, and Jarno Vanne. 2021. Comparative Rate-Distortion-Complexity Analysis of VVC and HEVC Video Codecs. *IEEE Access* 9 (2021), 67813–67828. https://doi.org/10.1109/ACCESS.2021.3077116

[13] Jeeyoon Park, Bumyoon Kim, and Byeungwoo Jeon. 2020. Fast VVC intra prediction mode decision based on block shapes. In *Applications of Digital Image Processing XLIII*, Andrew G. Tescher and Touradj Ebrahimi (Eds.), Vol. 11510. International Society for Optics and Photonics, SPIE, Basel, Switzerland, 115102H. https://doi.org/10.1117/12.2567919

[14] Jeeyoon Park, Bumyoon Kim, Jeehwan Lee, and Byeungwoo Jeon. 2022. Machine Learning-Based Early Skip Decision for Intra Subpartition Prediction in VVC. *IEEE Access* 10 (2022), 111052–111065. https://doi.org/10.1109/ACCESS.2022.3215163

[15] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 85 (2011), 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html

[16] Jonathan Pfaff, Alexey Filippov, Shan Liu, Xin Zhao, Jianle Chen, Santiago De-Luxán-Hernández, Thomas Wiegand, Vasily Rufitskiy, Adarsh Krishnan Ramasubramonian, and Geert Van der Auwera. 2021. Intra Prediction and Mode Coding in VVC. *IEEE Transactions on Circuits and Systems for Video Technology* 31, 10 (2021), 3834–3847. https://doi.org/10.1109/TCSVT.2021.3072430

[17] Mário Saldanha, Gustavo Sanchez, César Marcon, and Luciano Agostini. 2021. Learning-Based Complexity Reduction Scheme for VVC Intra-Frame Prediction. In *2021 International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, Munich, Germany, 1–5. https://doi.org/10.1109/VCIP53242.2021.9675394

[18] Michael Schäfer, Björn Stallenberger, Jonathan Pfaff, Philipp Helle, Heiko Schwarz, Detlev Marpe, and Thomas Wiegand. 2019. An Affine-Linear Intra Prediction With Complexity Constraints. In *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, Taipei, Taiwan, 1089–1093. https://doi.org/10.1109/ICIP.2019.8803724

[19] Ícaro Siqueira, Guilherme Correa, and Mateus Grellert. 2020. Rate-distortion and complexity comparison of HEVC and VVC video encoders. In *2020 IEEE 11th Latin American Symposium on Circuits & Systems (LASCAS)*. IEEE, San Jose, Costa Rica, 1–4.

[20] G.J. Sullivan and T. Wiegand. 1998. Rate-distortion optimization for video compression. *IEEE Signal Processing Magazine* 15, 6 (November 1998), 74–90. https://doi.org/10.1109/79.733497

[21] Liang Zhao, Li Zhang, Siwei Ma, and Debin Zhao. 2011. Fast mode decision algorithm for intra prediction in HEVC. In *2011 Visual Communications and Image Processing (VCIP)*. IEEE, Tainan, Taiwan, 1–4. https://doi.org/10.1109/VCIP.2011.6115979