Assignment 2 Report

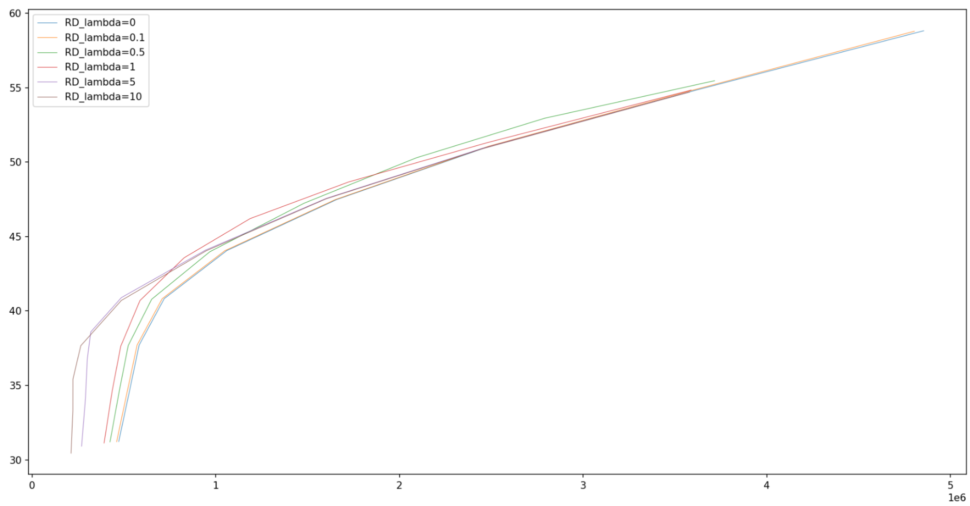
Feature 1: Multiple Reference Frame

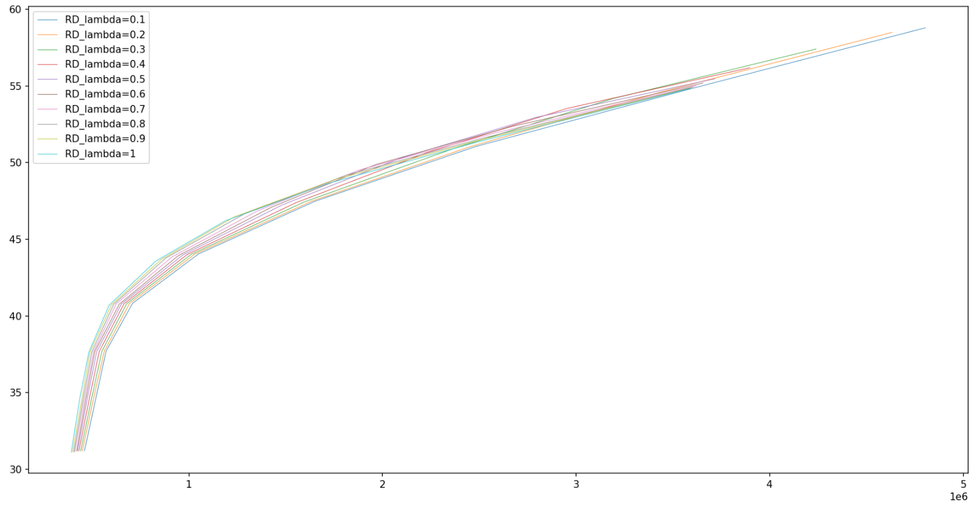
For feature1, we can use different color to indicate different ref frames to suggest that our feature work correctly:

In the image above, we configured the block size to 16, the search range to 4, qp to 4, and n frame to 3. Additionally, the representation uses a simple gray color for the previous first frame, light purple for the second previous frame, and dark purple for the third previous frame.

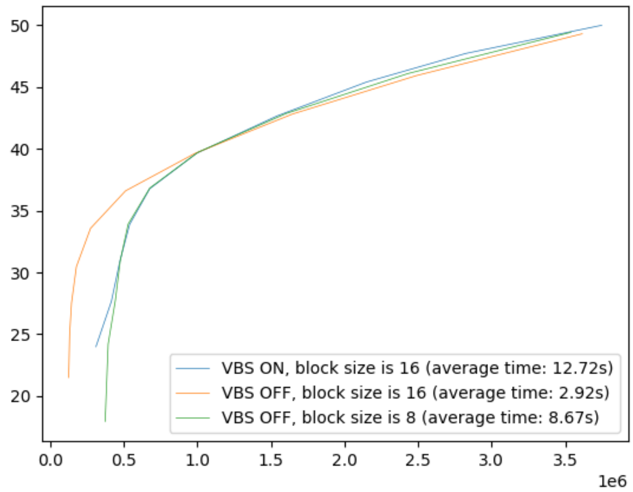
Furthermore, employing only QP=4 without any additional experiments, we graph per-frame distortion and encoded bitstream size for each varied setting of nRefFrames, ranging from 1 to 4, as outlined in the deliverable section:

Feature 2: Variable Block Size





We compared the effects of Variable Block Size (VBS) under different lambda settings and ultimately found that a lambda value of 0.3 yields the best overall results for a block size of 16 and a block search offset of 4.

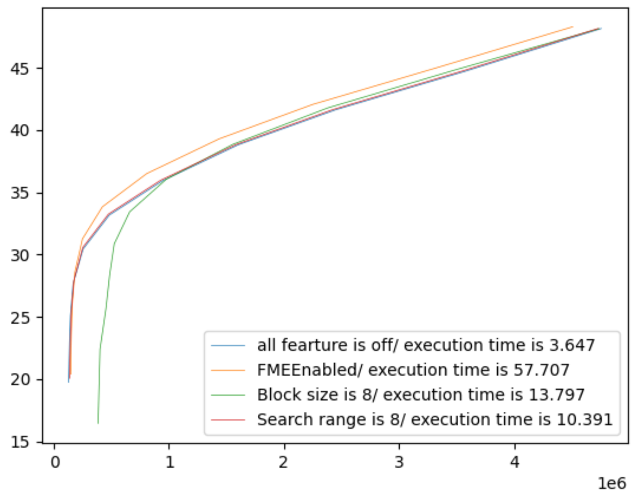


Using lambda = 0.3, we conducted comparative experiments, drew Rate-Distortion (RD) plots, and recorded the time consumption. The experiments included: 1. turning on VBS and setting the block size to 16, 2. turning off VBS and keeping the block size at 16, and 3. turning off VBS and setting the block size to 8. It was evident that the configuration with VBS turned on and a block size of 16 yielded the best results, outperforming both the all sub-blocks setting (VBS off, block size 8) and the all normal blocks setting (VBS off, block size 16). This demonstrates that our implementation of VBS is both correct and effective.

Feature 3: Fractional Motion Estimation

In the context of feature 3, we can create an R_D graph by contrasting FME-enabled scenarios with those where FME is disabled. To delve deeper into the assessment of

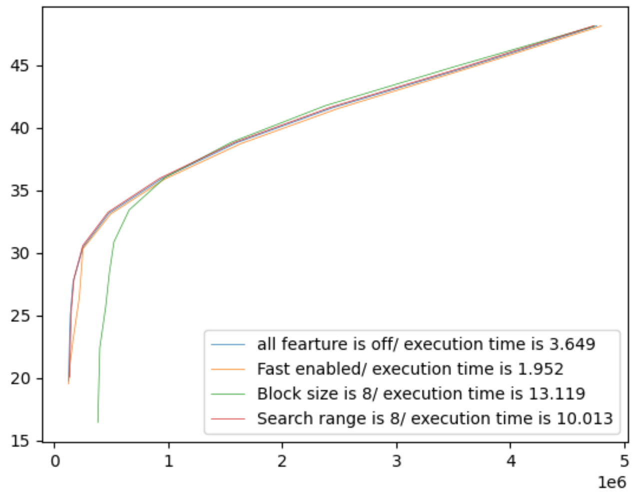
FME's impact, we additionally generate an R_D graph by varying the search range and block size to analyze performance differences.



Based on the above graph, it's evident that incorporating FME leads to improved performance, surpassing the outcomes achieved by merely employing smaller block sizes or larger search ranges. However, it comes at the expense of increased execution time.

Feature 4: Fast Motion Estimation

For feature 4, we can draw an R_D graph with Fast ME enabled compared to Fast ME disabled as follows and compare the execution time.

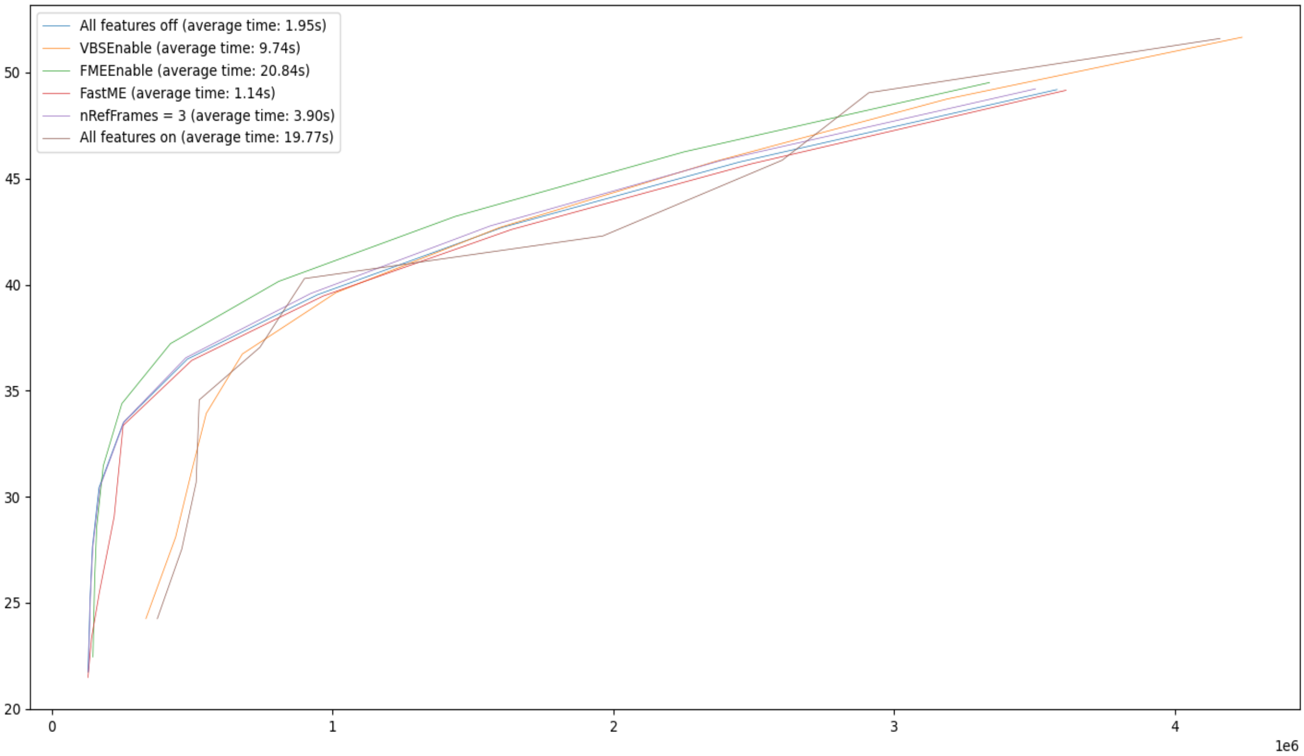


From the graph, we can see that implementing Fast Me will greatly decrease the execution time while maintaining the almost same performance.

Deliverables

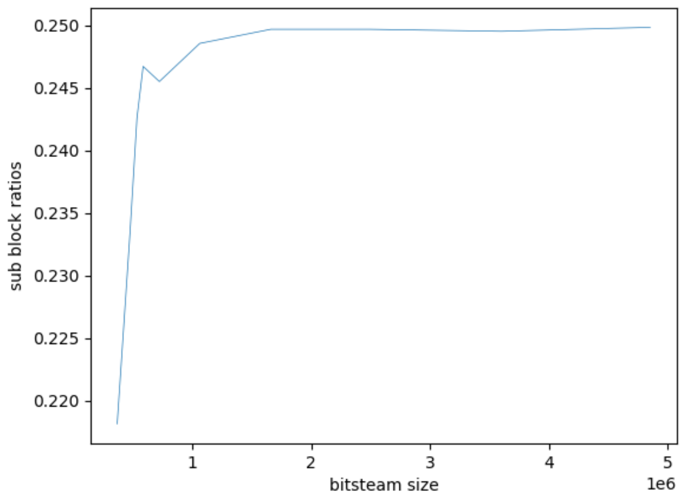
Part1

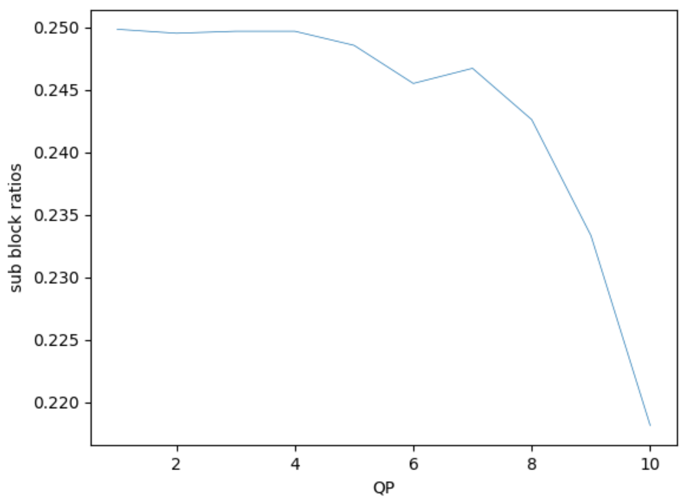
To delve deeper into the performance implications of the four features, we conducted encoding experiments using a constant setting (i_period = 8, block size = 16, search_range = 4) while enabling each feature individually. Subsequently, we generated R-D graphs to systematically compare the performances of the encoder under these different feature configurations. We also include the average execution time of the first 10 frames in the graph.



Part2

In our experimental analysis, we present a graphical representation illustrating the relationship between the ratio of sub-blocks and the tested QP (Quantization Parameter) values. Additionally, we explore the correlation between the sub-blocks ratio and the bitstream size. For this experiment, the following configurations were set: block size was fixed at 16, the block search offset was established at 4, and the I_Period was configured to 8. It's important to note that all features from assignment 2, except for the Variable Block Size (VBS), were disabled.

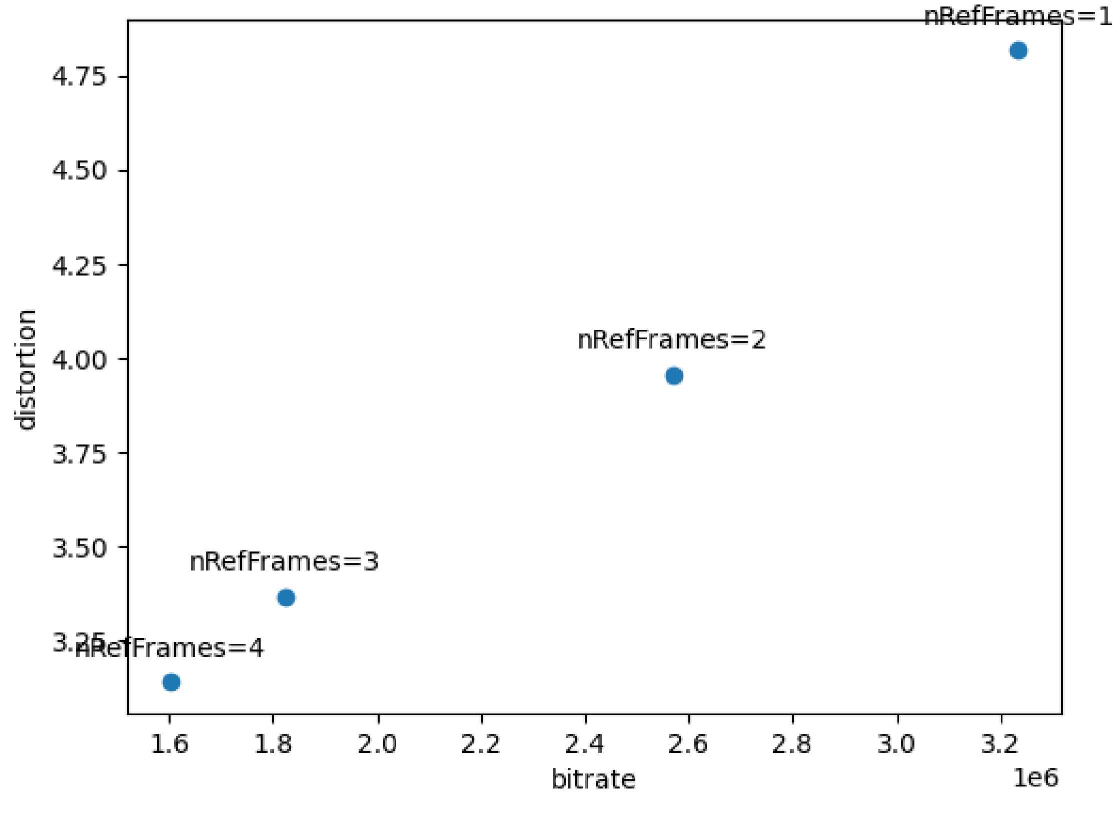




Part3

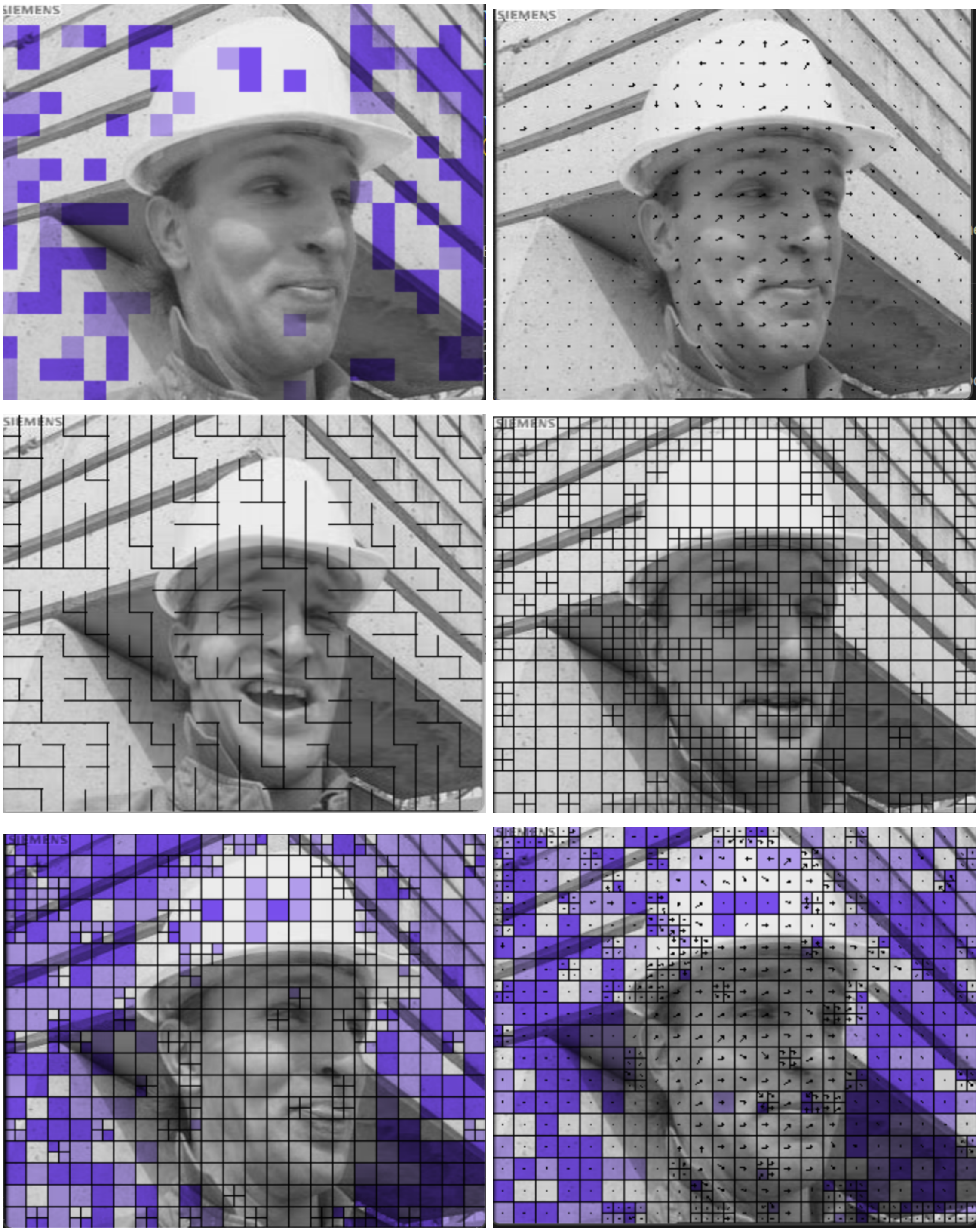
We also focused on the test file 'synthetic.yuv' as provided in Quercus. Our analysis was conducted using a fixed QP (Quantization Parameter) value of 4, without varying this parameter for other experiments. The task involved plotting graphs that depict the per-frame distortion and the encoded bitstream size for each distinct setting of nRefFrames, ranging from 1 to 4.

The graph can be interpreted as follows: The "synthetic" video is an artificial video comprising entirely unrelated frames with no inherent connection. However, by configuring a larger nRefFrames setting, we can extend the search range back into previous multiple frames, identifying potential correlations and encoding the video from that point. If a favourable match is found, the distortion decreases. Consequently, a larger nRefFrames setting increases the likelihood of discovering matching previous frames, ultimately leading to improved performance.

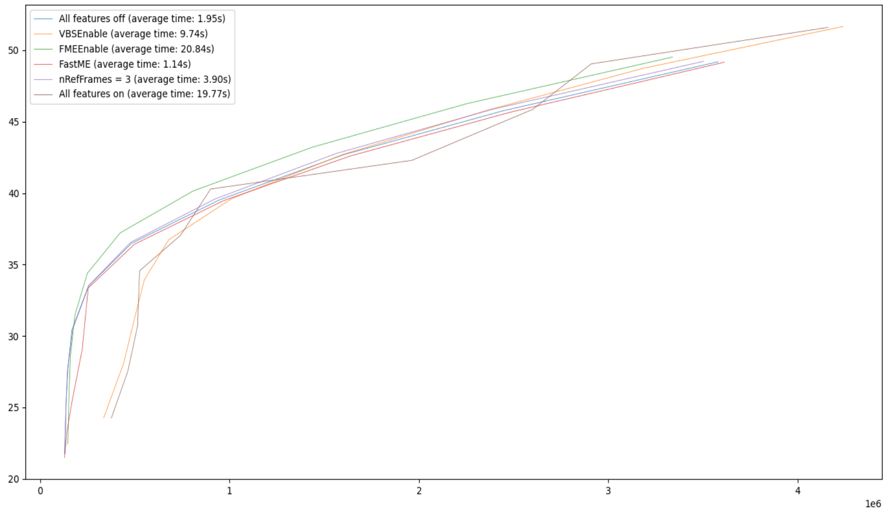


Part4

We have implemented visualisations for block boundaries, which allow for the visualisation of variable block sizes, moving vectors of P-frames, modes of I-frames and the reference frame per block. Each of these features can be independently toggled on and off. Below are the screenshots demonstrating our results.



What effect do Multiple Reference Frames, Variable Block Size, and Fractional Motion Estimation have on encoding speed and quality?



From the R-D graph above, we can conclude the effect of different features for first 10 frames of video foreman.yuv as follows:

	Quality	Speed
Multiple Reference Frame	Slightly improve the quality	Slow down the speed
Variable Block Size	Greatly improve quality	Slow down the speed
Fractional Motion Estimation	Greatly Improve quality	Largely slow down the speed
Fast Motion Estimation	Almost the same	Greatly speed up

We can infer from the graph that VBS and FME have the most significant influence on video quality, whereas Fractional Motion Estimation has the most pronounced impact on encoding speed.

How do they compare to using a smaller block size and/or a larger search range?

The graph in Feature 3 illustrates that employing a smaller block size or opting for a larger search range can enhance performance. Nevertheless, the performance improvement from these adjustments is not as substantial as that achieved through the use of FME or VBS features. Furthermore, the utilisation of FME or VBS does not preclude the possibility of selecting smaller block sizes or larger search ranges; the decision should be guided by the specific usage scenario and quality requirements. When choosing appropriate encoding settings, it's crucial to engage in trade-offs to arrive at an informed decision.

Could the effect of Multiple Reference Frames depend on content type? Can you think about what kind of non-artificial video could benefit the most from it?

The effect of Multiple Reference Frames can largely depend on content type. From my understanding, multiple reference frames will have a more significant effect in situations where scenes change rapidly so that the encoder can search for a better match in multiple references. However, if it is a static scene, using multiple references will not have much impact on the video quality and will waste resources finding matches in multiple identical ref_frames.

Ideal use cases for multiple references will be rapidly changing scenes such as moving landscape videos where scenes change constantly. In such videos, using multiple reference frames can help us find better matches in ref_frames and can contribute to producing higher-quality videos. Moreover, multiple reference frames can be beneficial in dynamic videos or videos that have some fast-moving objects.

What trade-offs are involved in using a feature like Fractional Motion Estimation, other than performance?

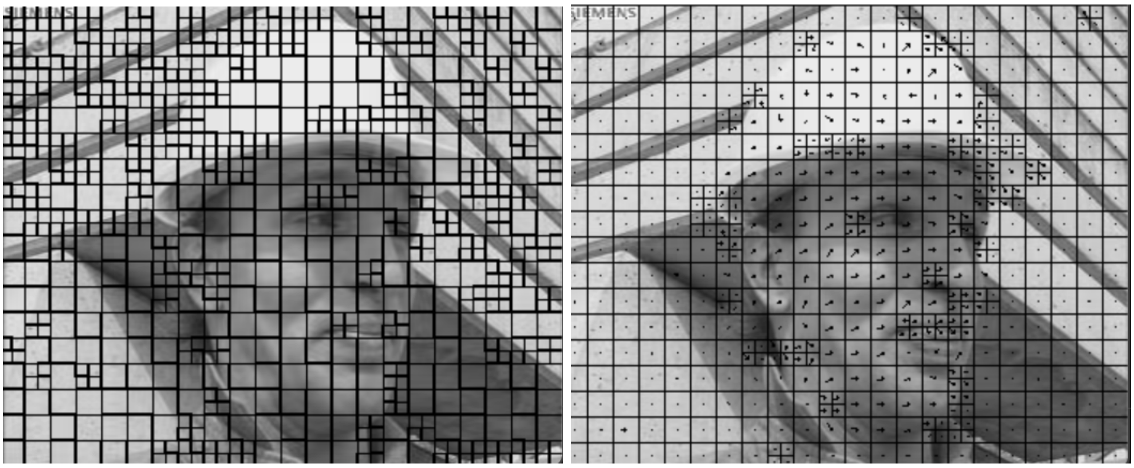
There are various trade-offs involved in using Fractional Motion Estimation besides pure performance:

computing resources limitations
The smaller the fractional factor is, the larger the computational burden it will bring, which will include reconstructing larger ref_frames and searching matches in a larger search range. Such an increase in computing requirements can be crucial in scenarios where computing resources are limited.
bitrate constraints
Using FME can result in larger bitrates, which can be a burden in scenarios where the bandwidth of the network is limited, such as in a live stream. In such cases, a large bitstream can lead to discontinuous video on the receiver's side.
Nonlinear benefits
The advantage gained from FME is nonlinear; an increased search does not guarantee better quality. There is a specific threshold where we can achieve a good balance between cost and results.

Some video standards support up to 1/8-pixel ME, would it be a good idea to use such a feature unconditionally? Why?

It is not a good idea to use features like 1/8-pixel ME unconditionally. As described in the previous question, there are several trade-offs in FME. In conditions like live video transformation or scenarios where computing resources are limited or bandwidth is restricted, it may not be a good idea to use the FME feature.

For Variable Block Size, what kind of areas get larger block sizes in intra-frames? Are they the same in Inter frames? Why?



Intra-Frames: In intra-frame encoding, larger block sizes are typically assigned to areas within objects, where there is less variation and more uniformity. This is because these areas don't require as much detail to be accurately represented. Conversely, smaller block sizes are favored at object boundaries or in areas with rich details. This is due to the need for capturing finer details and variations, such as textures or edges, which are more prevalent at these locations. The more complex or detailed a region is, the smaller the blocks needed to effectively capture the varying information within that single block.

Inter-Frames: For inter-frames, the logic is somewhat similar but also takes into account the motion dynamics. In areas within objects, which can be approximated as rigid bodies, larger blocks are sufficient since the movement can be well represented by a single motion vector. However, at the boundaries of objects or in areas with significant detail (like the edges between eyes, cheeks, and the background), the motion is often non-uniform. In these areas, different sub-blocks within a larger block might move in varying directions or speeds. Therefore, smaller block sizes are more advantageous in these regions to accurately capture the complex motion patterns.