DIGITAL SUBBAND VIDEO

ENCODER INFORMATION
VERSION 2.0

October, 2024

2024 EMMIR
ENVEL GRAPHICS

Contact Information:

GitHub : https://github.com/LMP88959
YouTube: https://www.youtube.com/@EMMIR_KC/videos
Discord: https://discord.com/invite/hdYctSmyQJ

This document describes design decisions and features of the Envel Graphics DSV2 encoder and codec itself.

A. Motion Estimation

Full-Pixel Motion

The encoder uses hierarchical motion estimation (HME) to derive full-pixel motion vectors. HME is the “best of both worlds” in terms of performance and accuracy. It effectively captures high and low motion like a full search yet performs nearly as fast as a traditional speedy search like the three-step-search.

New in the DSV2 encoder:
At an effort level 1 or less, only left, right, top, and bottom blocks are checked. At effort levels greater than 1, the diagonal blocks are checked as well.

Half-Pixel Motion

Half-pixel motion vectors are derived using a fixed size half-pixel block search for speed. The search is a full search around the pixel’s eight neighbors but only checks a fixed size square in the center of the test block.

New in the DSV2 encoder:
At an effort level 3 or less, half-pixel motion estimation is not performed.
At an effort level less than 10, half-pixel and quarter-pixel motion estimation are skipped if the full-pixel SAD (sum of absolute differences) computation is sufficiently low, reducing encoding time significantly but providing worse video quality.

Quarter-Pixel Motion

Quarter-pixel motion vectors are derived using a fixed size quarter-pixel block search for speed. The search is a full search around the pixel’s eight neighbors but only checks a fixed size square in the center of the test block.

New in the DSV2 encoder:
At an effort level 7 or less, quarter-pixel motion estimation is not performed.
At an effort level less than 10, half-pixel and quarter-pixel motion estimation are skipped if the full-pixel SAD (sum of absolute differences) computation is sufficiently low, reducing encoding time significantly but providing worse video quality.
At effort level 8, quarter pixel motion estimation is only performed if the half-pixel motion vectors are sufficiently small (so as to capture very slow motion).

INTRA BLOCK DETERMINATION

Effective intra block determination is absolutely crucial to getting good looking video at low to medium bitrates. The encoder computes simple Human Visual System (HVS) based statistics on the block's luma and chroma to determine if it should be an intra block or an inter block.
The following statistics are computed:

- Variance
- Average
- Texture - average of horizontal and vertical differences in the block
- Span - difference between smallest and largest value in the block

The HVS tells us that error is mostly visible in flat/low texture regions, and so that is what intra block determination seeks to remove.

Intra Sub-block Determination

The DSV2 specification allows for an intra block to be split into 4 equal sub-blocks where each sub-block is either intra coded or inter coded with a zero motion vector. This is advantageous for scenes where there are 2D overlays on the video, like the score overlay in a football match. In cases where there is very high motion in the scene, these sub-blocks can be used to maintain a clear image of the score at the cost of accuracy around the edges of the overlay, which are generally not too visually important. The encoder computes an "intra metric" to help capture these static portions of the video. This metric is essentially a combination of texture and block similarity. A good candidate would be a sub-block that has high texture and a reference block that gives extremely accurate prediction for a majority of the sub-block.

New in the DSV2 encoder:
At an effort level 6 or greater, the sub-blocks are individually checked to determine if they contain enough of an error to mark the entire block as intra.

B. Stability Blocks

The encoder keeps track of the average motion of every block over a certain refresh period. Blocks who have an average motion of approximately zero over the period are marked as stable. Another factor that can mark a block as stable is if it is determined to be "high detail."

New in the DSV2 encoder:
Two new block flags were added: maintain and ringing. With psychovisual optimizations enabled, intra frames are analyzed to determine which blocks should get which flags.
Maintain blocks are generally low variance and low texture.
Stable blocks are generally blocks that are deemed to contain important high frequency information.
Ringing blocks are marked to use a ringing wavelet transform, this generally gives ugly artifacts at lower bit rates but works exceptionally well to synthesize apparent detail in high texture blocks. Generally used to make foliage seem more detailed.

C. Scene Change Detection

Compared to the DSV1 encoder, the DSV2 encoder implements a slightly more sophisticated method of scene change detection. Motion estimation takes place and computes how many sufficiently different blocks are in the frame and how many blocks would be marked as intra. It then takes into account the distance from the last intra frame and determines if the differences warrant the insertion of an intra frame instead of an inter frame.

D. Rate Control

For single pass average bitrate, the encoder knows how many bytes each frame needs on average to satisfy the given rate. It uses a simple control loop to ensure the video approximately hits its bitrate target without heavy oscillation or instability.
The loop utilizes a few pieces of information:

- ***avgbpf*** = average bytes per frame (BPF) so far
- ***isP*** = whether the current frame is intra or inter

The control loop finds how far (as a fraction called *delta*) the current ***avgbpf*** is from the target BPF and adjust the quality by *delta*. The *delta* value is limited harder if we are increasing the quality. If the frame ***isP*** then its lower quality bound is determined by the average quality of the inter frames minus four percent. This helps prevent the bit starvation that intra frames tend to impose on subsequent inter frames.

E. Motion Compensation

Motion compensation was a weak spot in DSV1, it could not represent large changes and the compensation itself was blocky. Both of these issues have been remedied in DSV2. Expanded Prediction Range Mode (EPRM) is a new compensation mode which doubles the range when storing block differences at the expense of fine precision. For example, instead of being able to represent {-3,-2,-1,0,1,2,3}, EPRM lets us represent {-6,-4,-2,0,2,4,6}.

To combat the blocky motion compensation, in-loop filtering was added. Simple smoothing and interpolation functions are performed on the block edges to blend it into its surroundings better. It also tries to avoid smoothing anything that should remain an edge.
DSV2 dynamically computes the filtering threshold (or strength) via a relatively simple algorithm which takes into account how disjoint the block is. If the block's motion vector is wildly different from its left and top neighbors, it is filtered more. In practice, this dynamic threshold computation has shown to be highly effective at determining how to filter blocks in a visually pleasing manner.

F. Metadata Statistics

All of these new block flags introduced in DSV2 come with some overhead, especially noticeable at lower bit rates. DSV2 allows the encoder to specify how the RLE encoded bits should be represented in each frame (if there are more zeros, RLE the zeros, and vice versa).