Layered Motion Compensation for Moving Image Compression

Gary Demos
Hollywood Post Alliance Rancho Mirage, California
21 Feb 2008
Part 1

High-Precision Floating-Point Hybrid-Transform Codec
DWT 9/7 Subband

Low Low

Optimal Filter (windowed sinc)

Low
DWT 9/7 Subbands (such as used in JPEG-2000 irreversible codec):

- Low Low band resembles a windowed sinc, but differs from it
- A typical subband layered hierarchy will alias in lower layers due to this difference, wasting bits, and making the image less useful (not really watchable at lower layers)
- Low and High phase centered on pixels, maximizing coded difference bits required for high subbands
- Band-split only defined at this pixel-centered (half-pixel offset) phase
- Only defined for factor-of-two resolution hierarchy
- High bands require 3/4 of the values of the reconstructed image pixel count (1/4 in the low-low band)
- Statistically divergent (i.e. some highly divergent outlier pixel values) when used in multi-layer configuration (such as in JPEG-2000). This is the biggest weakness of JPEG-2000 irreversible coding.
Symmetric half of variable-sharpness weighting function based upon windowed sinc filter sampling function of width ± 3.0 pixels.

Theoretical optimal filter is windowed sinc at full sharpness (shown dark-green).

Also, stars show the DWT 9/7 Low Band FIR Taps. Note that the DWT 9/7 low band 9 taps differ from the windowed sinc, and thus will alias.
Optimal Filter:

• Windowed sinc is optimal for low-pass and for upsize reconstruction, as well as pixel displacement

• Aliasing minimized at all levels

• Low-phase pixels can be centered between pixels, minimizing coded difference bits required for high subbands

• Can be applied at any phase, including very high sub-pixel accuracy (0.01 pixel accuracy is used here for resizing and displacement).

• Upsize requires full resolution, but coded bitsize is the main criterion (5/4 pixel count required at each step)

• Not restricted to factors of two (e.g. 1920x1080 & 1280x720 have a 3/2 & 2/3 relationship)

• Provides a feedback correction at every resolution level (unlike DWT 9/7). This is very helpful in bounding quantization error at every level (i.e. eliminating JPEG-2000 statistical divergence)
The Optimal Solution, A Combined System:

• Combination of both is best (each applied where best)
• An efficient and effective approach is to band-split using the optimal filter, then code the deltas using the DWT 9/7 transform at multiple levels with all bands present, followed by quantization (using finer quantization at the lower layers)
• The DWT 9/7 bi-orthogonal sub-band system provides orthogonal frequency bands for efficient layered quantization for compression coding, while retaining the benefits of the optimal filter when applied to the optimal-filter deltas.
• The use of the optimal (windowed-sinc) filter not only minimizes aliasing, but allows arbitrary phase (for translation, arbitrary resizing, etc), and supports arbitrary sub-pixel accuracy.
Example of combined Windowed-Sinc and DWT 9/7, using 3/2 resolution-enhancing layer:

Original 1920x1080

Upsized 1920x1080 (from decoded 1280x720)

Decoded 1280x720

Note: resolution-enhancing delta layer is phase-neutral, and is free of aliasing

Note: encoding shown, decoding is the implicit reverse process
Part 2

Flowfield-Driven Layered Motion Compensation System
Overall sharpness/softness for each channel for horizontal and vertical

RGB, YUV, or mv_x, mv_y, conf

Reduced Resolution

Variable Downsize

RGB, YUV, or mv_x, mv_y, conf

Variable Upsize

Overall sharpness/softness for each channel for horizontal and vertical

RGB, YUV, or mv_x, mv_y, conf

Increased Resolution

Quantized Transform
Encoder/Decoder

Coded Bits (in files or multiplex)

Decoded RGB, YUV deltas or mv_x mv_y conf

Quantized Transform
Stand-Alone Decoder

Coded Bits (in files or multiplex)

Decoded RGB, YUV deltas or mv_x mv_y conf
Toolkit (continued):

Frame A           Frame B (reduced resolution)
Search Range, Horizontal And Vertical
Coarse Best Match mv_x, mv_y, confidence

Frame A           Frame B (full resolution)
Scaled coarse mv_x, mv_y, conf
Refinement small search range (centered at scaled coarse best match)
Fine Best Match mv_x, mv_y, confidence (e.g. for each 4x4 pixels)

Frame to be Displaced
Fine mv_x, mv_y, conf for each pixel
Per-Pixel Arbitrary Displacement and Filtering (Sharp/Soft)
Modulation for each pixel (for reso-enhancing layers) optional output
Predicted Displaced And Filtered Frame or
In-Place Sharp/Soft Filtered Frame
Basis of resizing and displacement filters is variable sharpness/softness windowed sinc

Symmetric half of variable-sharpness/softness weighting function based upon windowed sinc filter sampling function of width ± 3.0 pixels

For upsize and downsize, horizontal axis corresponds to lowest of the two resolutions. For displacement, resolution remains constant.
Start with hierarchical motion compensation

Searching for best match:
- Multiple match criteria
- Large search range, but slightly favor
  mv field smoothness and/or zero mv’s
- Include overlap (e.g. 8x8 for 4x4)
- Multiple shape distortions
  (per-pixel small deltas on block)
- Resulting in motion vector
  and confidence for each best-
  match block
- No need to retain shape distortion
  information

Block match, but with a “twist”

Also, Pincushion, Waist Pinch, and Other Useful Distortions Commonly Found In Motion
The result is a low-resolution motion vector field, with corresponding confidence values.
Unlike Block Displacement:

- Can track complex motion such as:
  - Zoom and travel
  - Atmospheric distortion
  - Camera rotation
  - Distorting objects
  - Parallax motion
  - Motion with lens distortion (e.g. wide-angle lens)
  - All of the above simultaneously!
  - This greatly expands the motion-tracking capability vs block matching

But…

- Cannot handle one object passing in front of another (e.g. distant person walking left in front of stationary background)
- One object seen through a reflective transparent window
- Chaotic motion such as blowing leaves or ocean waves
- Other non-trackable regional motion
Perform high resolution fine search with small search range, about the coarse low-resolution motion vector field

The result is a higher resolution motion vector flowfield, with corresponding confidence values. The resolution of this flowfield is typically a little below the resolution of the base layer.
Confidence:

- Quality of fine-search match
- AC and DC match criteria
- SAD (overall match), SSD (minimal outliers),
  Median of Absolute Differences (at least half have good match),
  weighted blend of the these (with appropriate scalings)
- Local divergence being low
If low confidence (a poor match):

- The residual after motion compensation will correct remaining difference. The quality of matching is therefore not directly visible.

- The goal of motion compensation is overall reduction of coded bits. Thus, perfect matching is not needed everywhere. Some matching will usually yield an overall benefit. The amount of benefit will be scene-dependent.

- For noise reduction preprocessing, low confidence can be used to defer back to the original image without any processing. The goal of noise reduction is also overall reduction of coded bits. As with motion compensation, some matching will usually yield an overall benefit. Since noise-reduction processing is potentially visible, caution is required such that pixels are only altered when matching confidence is high. Further, the amount of alteration is carefully bounded to a level corresponding approximately to the noise floor.
Sharp/Soft filtering:

- A function of “shutter-open” duration (e.g. 180 degree shutter)
- Long motion vectors have motion blur along direction of motion
- Softer filters can be used along direction of motion
- If a long motion vector is only sure to have blur only if there is high confidence associated with this motion vector
- Resolution-enhancing layers contain no detail in high-motion regions, only noise. Thus, resolution-enhancing delta layers can be modulated to reduce detail amplitude in regions of high motion when there is high match confidence.
- Sharp/soft filtering can be applied for both intra-coded frames as well as motion-compensated frames
- The amount of motion, in terms of number of pixels, is a function of resolution in a resolution layered system, and thus, motion vectors and sharp/soft filtering is applied with appropriate scale in each resolution layer.
Upsize mv’s and confidences, and scale motion vector field, for use at any resolution layer. Use variable upsize.

Result: one motion vector and one confidence for every pixel, at each and every resolution layer, including when layers have complex numerical size relationships (e.g. 720x480, 1280x720, 1920x1080)
Multi-Resolution Layered Coding Using Optimal Resizing Filters and Quantized Codec (e.g. DWT 9/7 bi-orthogonal subband Wavelet Coding) with Flowfield Motion Compensation (note error-bounding feedback at each resolution-enhancing layer)
Flowfield-Predicted (forward) Layered Coding
Example Prediction Dependency

\[ D_1 = \text{Decode Time 1} \]
\[ D_2 = \text{Decode Time 2} \]
\[ D_3 = \text{Decode Time 3} \]
\[ D_4 = \text{Decode Time 4} \]
Flowfield-Driven Motion Compensation Feature Summary:

• Layered motion compensation using flowfields
• Built on basic toolset
• The key tool provides per-pixel displacement to 0.01 pixel
• Hierarchical motion search using block-matching
• Interpolation of mv’s using smooth variable upsize filter
• Sharp/soft filtering based on conf and mv length
• Modulation of resolution-enhancing layers based on conf and mv length
• Tracks complex motion
Results:

Intra coding significantly improved (> ~3x?, scene dependent) vs. the compression ratio of JPEG-2000
  - without the statistical divergence (i.e. strictly-bounded error)
  - benefiting substantially from flowfield intra pre-processing

Flowfield mc overall compression gain of between 1.2x and 2.5x vs. intra coding (typically 1.75x). For F-frames 3x is common and for M-frames 10x is common for high-detail scenes with slow smooth motion

Flowfield mc coding is thus achieving compression ratios approaching that of block-mc DCT-based 4:2:0 codecs (e.g. AVC/H.264, VC-1, MPEG-2) at Hi-Def and 4k while providing 4:4:4 and without block artifacts

Greatly expanded feature set compared to previous codecs (e.g. precision, dynamic range, error-bounding, layers, lossless residual)
Results (variable bitrate average, constant quantization):

<table>
<thead>
<tr>
<th>Scene</th>
<th>Bitrate</th>
<th>Gain vs. Intra</th>
<th>Compression Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pax 1920x1080/30P (layered as 1920x1080, 1280x720, 640x360)</td>
<td>9.54mbps</td>
<td>1.22 : 1</td>
<td>197.1 : 1 (10b), 315.4 : 1 (16b)</td>
</tr>
<tr>
<td>MCM 1920x1080/30P (layered as 1920x1080, 960x540)</td>
<td>33.63mbps</td>
<td>1.52 : 1</td>
<td>55.9 : 1 (10b), 89.4 : 1 (16b)</td>
</tr>
<tr>
<td>White man 1920x1080/30P (layered as 1920x1080, 960x540)</td>
<td>10.27mbps</td>
<td>1.38 : 1</td>
<td>183.0 : 1 (10b), 292.8 : 1 (16b)</td>
</tr>
<tr>
<td>Garden Music 1280x720/60P (layered as 1280x720, 640x360)</td>
<td>19.53mbps</td>
<td>2.19 : 1</td>
<td>68.0 : 1 (8b)</td>
</tr>
<tr>
<td>Mystic India Trailer 4096x2928/24P (layered as 4096x2928, 2048x1464, 1024x732)</td>
<td>66.05mbps</td>
<td>1.50 : 1</td>
<td>130.7 : 1 (10b)</td>
</tr>
<tr>
<td>STEM MMR4 4096x1714/24P (layered as 4096x1714, 2048x858, 1024x430)</td>
<td>54.88mbps</td>
<td>1.41 : 1</td>
<td>111.4 : 1 (12b)</td>
</tr>
</tbody>
</table>

**Average** over these test scenes (averaged equally) (note: amount of gain vs. intra is adjustable)  
1.54 : 1  
146.2 : 1
References:


garyd@alumni.caltech.edu