Scalable Coding of Motion and Depth Fields with Shared Breakpoints

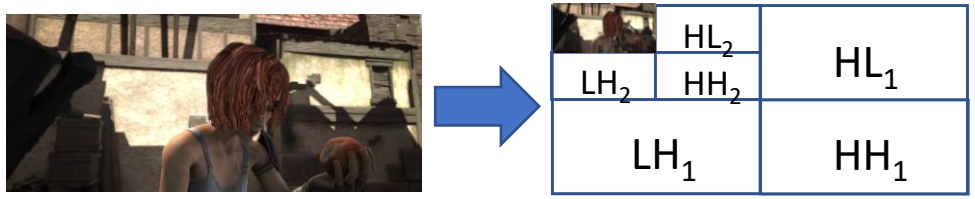
R. Mathew, Y. Li, D. Taubman

School of Electrical Engineering and Telecommunications
University of New South Wales, Sydney

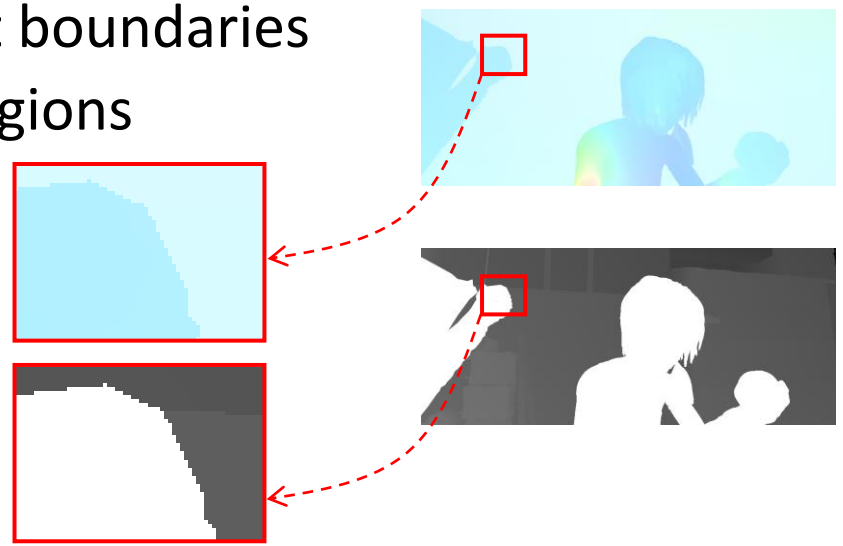
UNSW SYDNEY

Scalable Coding of Imagery

- JPEG 2000 enables scalable coding of images
  - Subband decomposition
  - Employing Discrete Wavelet Transform (DWT)

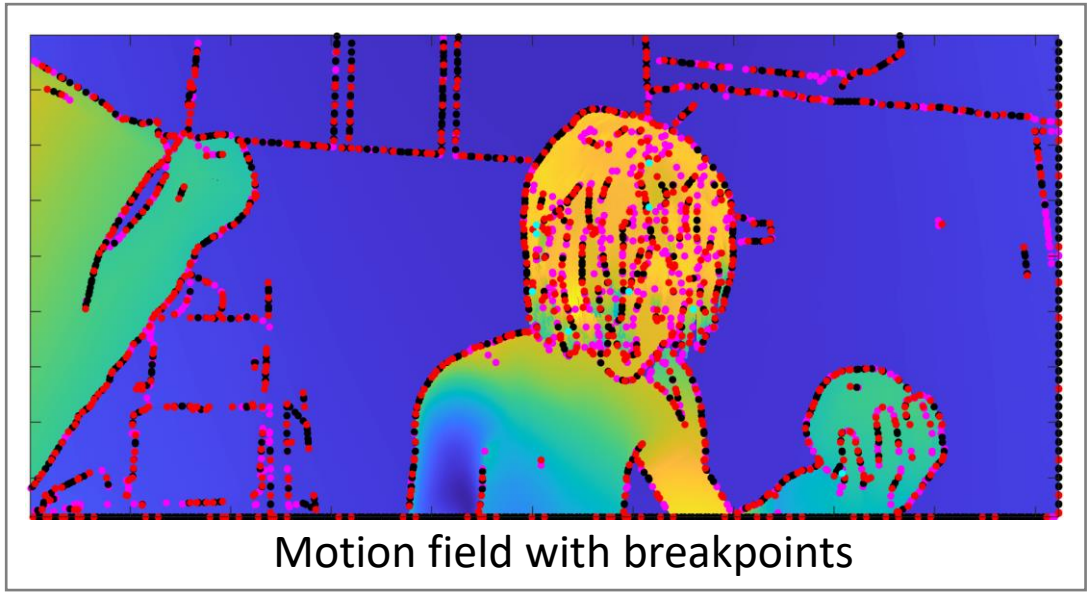


- Depth maps and motion flows: different characteristics to texture images
  - Sharp discontinuities at object boundaries
  - Smooth flows within object regions



Scalable Coding of Imagery

- New extension to JPEG 2000 is currently proposed
  - JPEG 2000 Part 17
  - For coding of discontinuous media
  - Completed Committee Draft of the standard

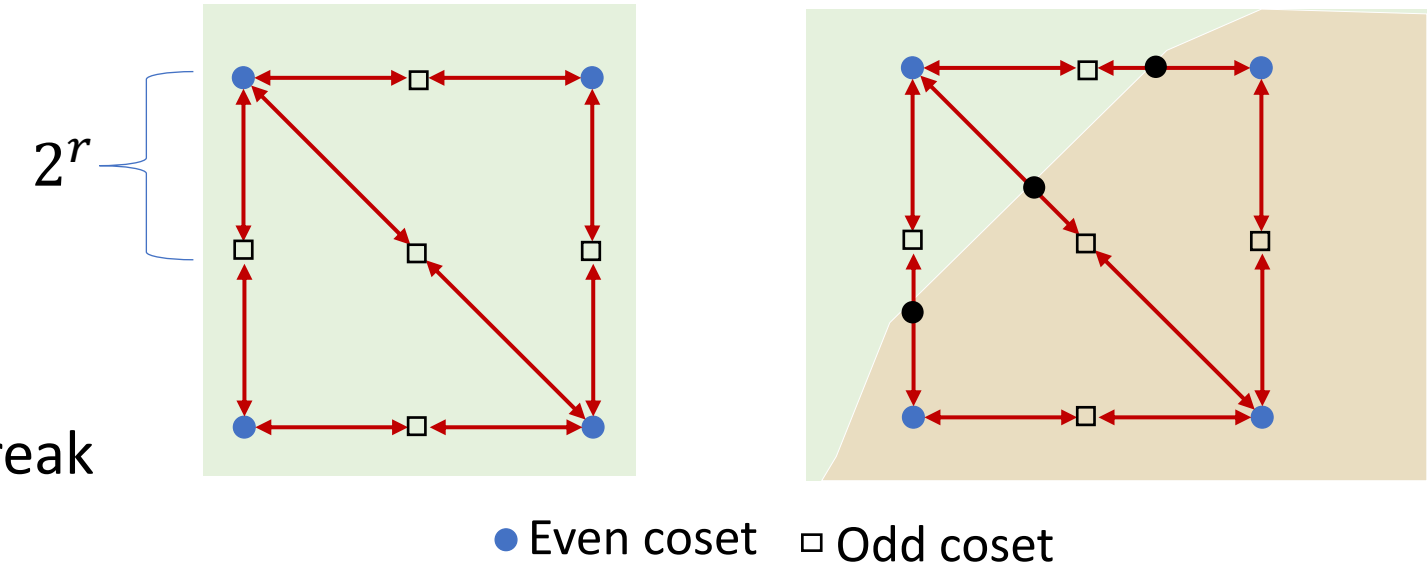


Motion field with breakpoints

- Key contribution
  - Breakpoint adaptive discrete wavelet transform (BPA-DWT)
  - Breakpoints: model object boundary geometry and
    adapt the DWT at the vicinity of object boundaries
    ensuring wavelet basis functions do not cross points of discontinuity

BPA-DWT

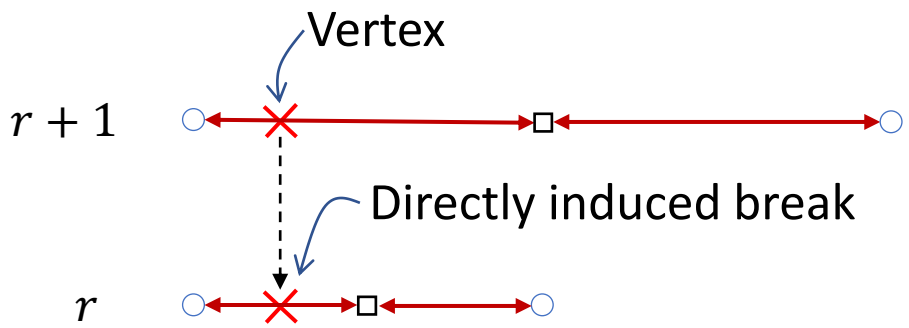
- JPEG 2000 Part 17
  - Two BPA-DWT defined
    - Quad-break
    - Tri-break
  - Focus of this paper is on Tri-break

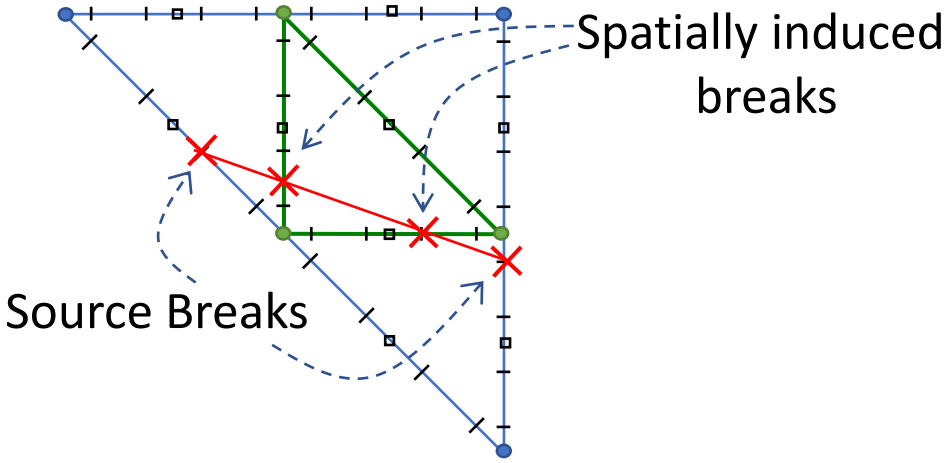


- Tri-break transform
  - Defined on a hierarchical triangular grid,
    - Interval spacing 2^r, at resolutions $r \in \{0, 1, 2, \ldots R\}$
  - Samples located at nodes: categorized into *even* and *odd* cosets
  - Breakpoints placed on arcs: modelling discontinuity boundary of the underlying flow

BPA-DWT

- Since Breakpoints adapt the transform
  - Need to be communicated along with subband coefficients
  - Only a small set of breakpoints are explicitly communicated – Vertices
  - Remaining are derived from Vertices
    - Directly induced breaks – from parent to child arcs in the hierarchical framework
    - Spatially induced breaks – across arcs at the same resolution in the hierarchy

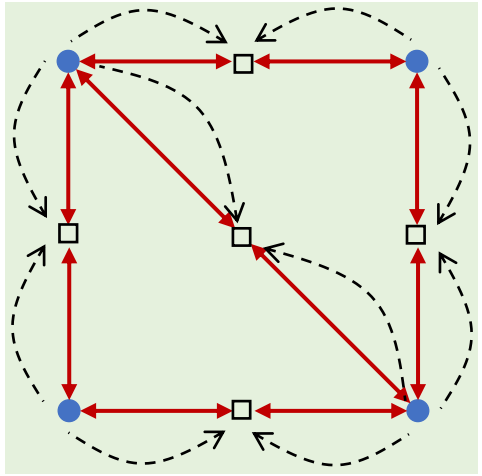


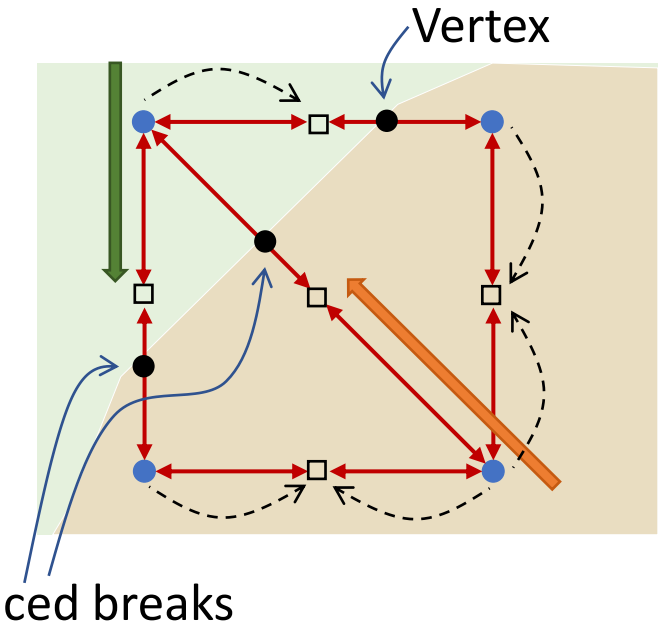


BPA-DWT

- Lifting steps

- Prediction Step
  - Even coset predicting odd coset
  - Arcs without break (smooth arcs)
    - Linear interpolation
  - Arcs with vertex
    - Constant extrapolation
  - Arcs with induced breaks
    - Linear extrapolation
    - Using gradients of nearby arcs in the same region as determined by breakpoint geometry

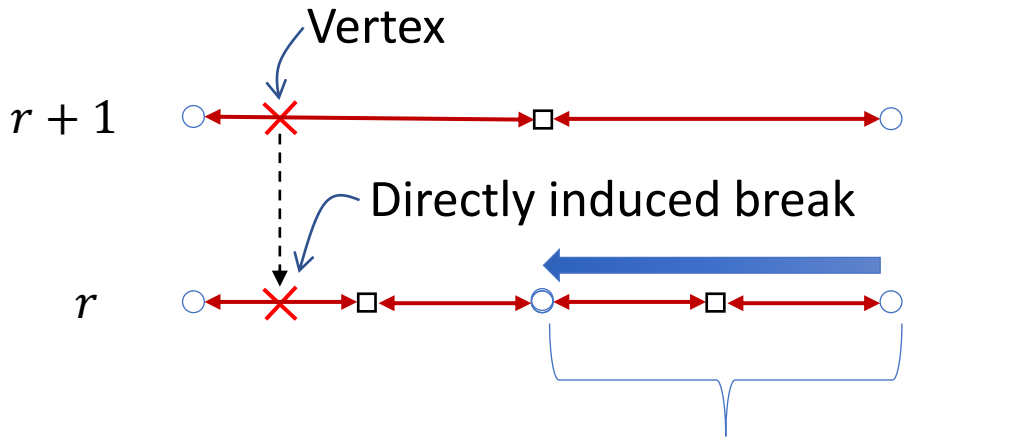




BPA-DWT

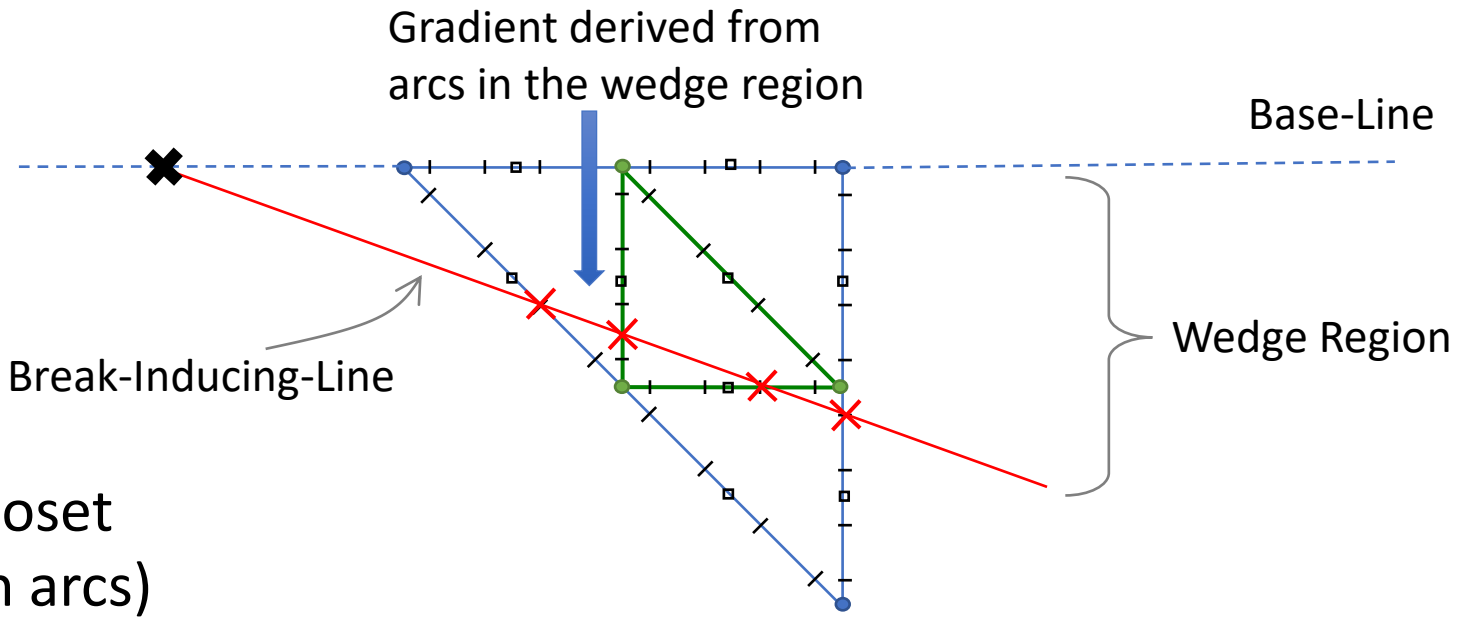
- Lifting steps

- Prediction Step
  - Even coset predicting odd coset
  - Arcs without break (smooth arcs)
    - Linear interpolation
  - Arcs with vertex
    - Constant extrapolation
  - Arcs with induced breaks
    - Linear extrapolation
    - Using gradients of nearby arcs in the same region as determined by breakpoint geometry
    - **Example: extrapolation for directly induced break**



BPA-DWT

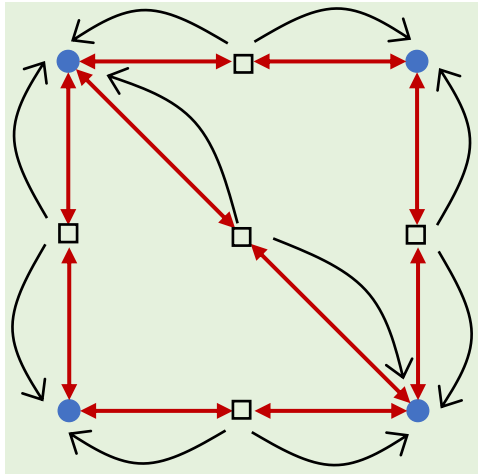
- Lifting steps

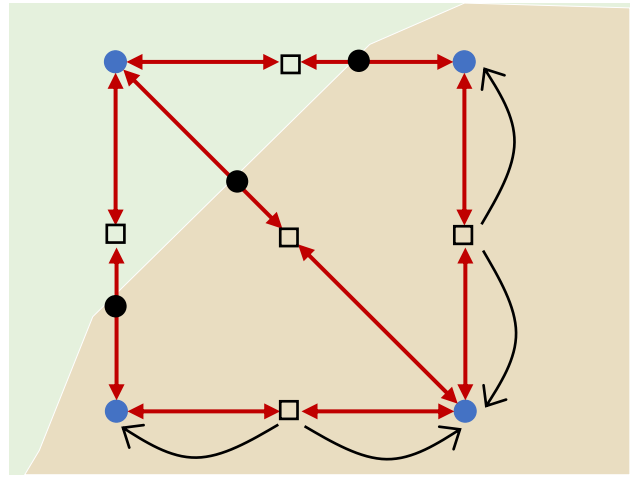


- Prediction Step
  - Even coset predicting odd coset
  - Arcs without break (smooth arcs)
    - Linear interpolation
  - Arcs with vertex
    - Constant extrapolation
  - Arcs with induced breaks
    - Linear extrapolation
    - Using gradients of nearby arcs in the same region as determined by breakpoint geometry
    - **Example: extrapolation for spatially induced break**

BPA-DWT

- Lifting steps



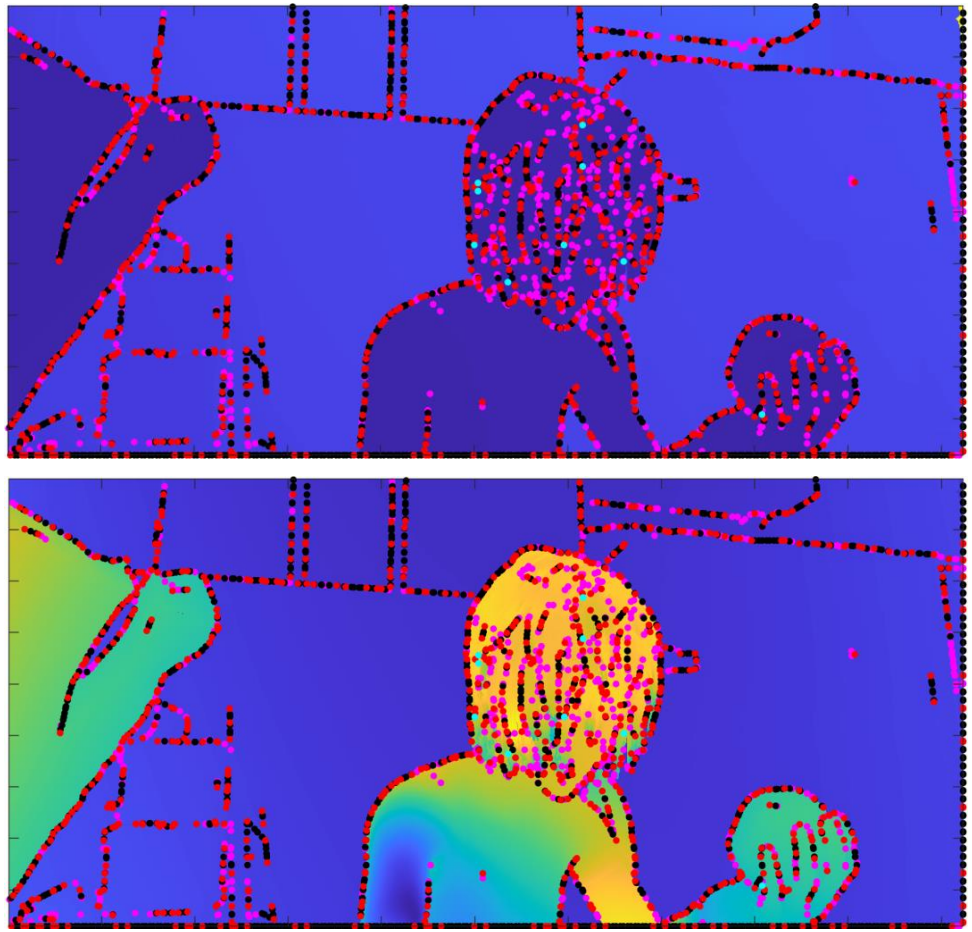


- Update Step
  - Feeds back to even coset partial prediction residuals determined during the predict step
  - Update step skipped for arcs with breaks

- After Completion of lifting steps at resolution r
  - Odd coset: high-pass subbands
  - Even coset: a low-pass or approximation subband,
    - input to the next coarser resolution r+1.

Coding Depth Maps and Motion Flows

- Multi-view video encoders require both
  - Motion – exploit temporal correlation between frames,
  - Depth – take advantage of inter-view correlation
- Application:
  - Coding video captured by a High Density Camera Array
- Observation:
  - Common boundaries between motion and depth
- Propose:
  - Common breakpoint representation for tri-break



Depth map (top) and motion field (bot) with shared breakpoints.

Coding Depth Maps and Motion Flows

- Estimation of Shared Breakpoints
  - Placement of vertices: goal to minimize total Lagrangian cost J_V
  - Calculated across all 3 input components
    - Depth
    - Motion – Horizontal and vertical components
- Start breakpoint estimation at the finest level
- Successively progress to coarser levels in the hierarchical framework
  - pruning away vertices of child arcs if new geometry at parent level reduces J_V

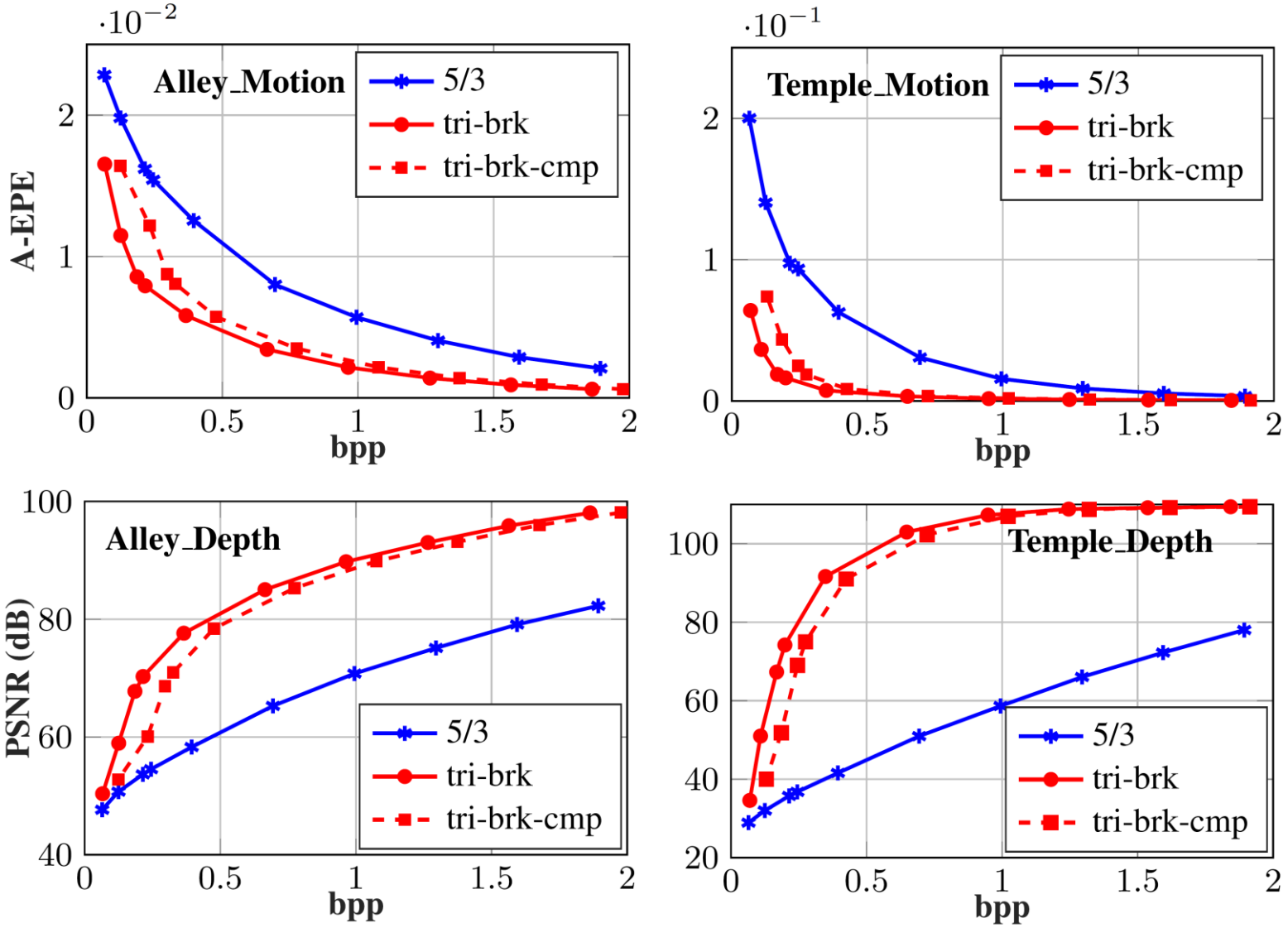
Coding Depth Maps and Motion Flows

- R-D Results
  - Coding Depth Maps and Motion Flows
  - Corresponding to a frame of Sintel* sequences: Alley, Ambush and Temple

  - Rate (bpp): includes rate for
    - depth and motion subband coefficients, and
    - shared breakpoints
  - PSNR: metric for decoded depth map quality
  - Average End Point Error (A-EPE): metric for decoded motion field distortion

* D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "*A naturalistic open source movie for optical flow evaluation*," in European Conf. on Comp. Vis., 2012.

Coding Depth Maps and Motion Flows



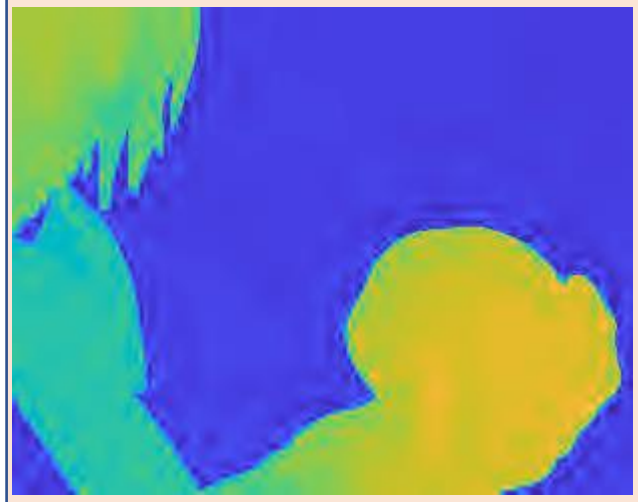
R-D comparisons

5/3: JPEG 2000 with 5/3 DWT

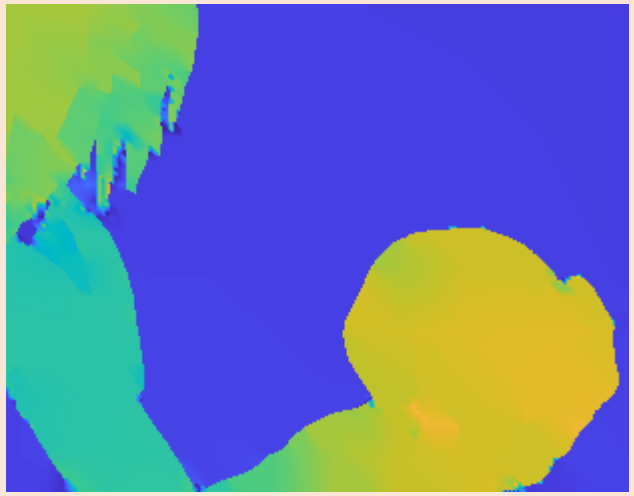
tri-brk: subband coefficients coded using JPEG 2000 and shared breakpoints coded with embedded bit-plane coding as defined in JPEG 2000 Part 17 extension.

tri-brk-cmp: breakpoints are estimated and communicated separately for each component.

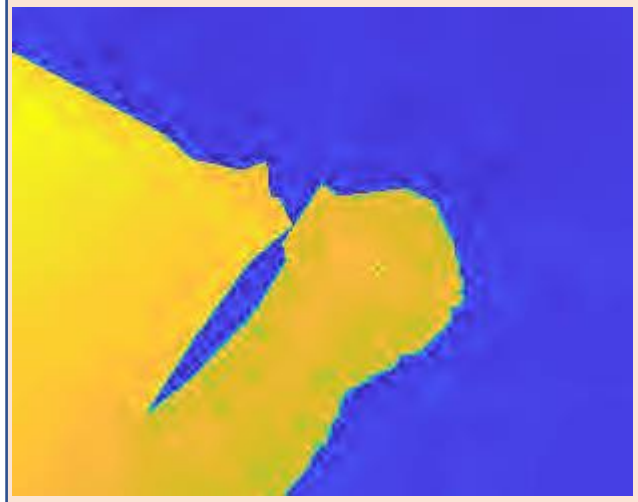
Alley - Decoded Examples at 0.045 bpp



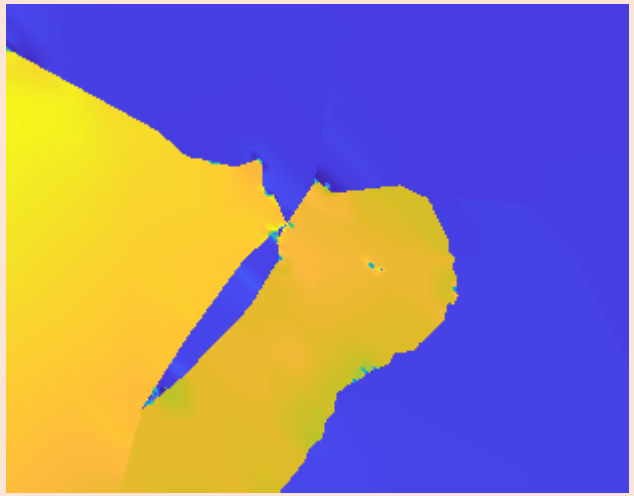
Decoded disparity (5_3_DWT)



Decoded disparity (tri-break)



Decoded disparity (5_3_DWT)



Decoded disparity (tri-break)

Temple - Decoded Examples at 0.044 bpp

Decoded disparity (5_3_DWT)

Decoded disparity (tri-break)

Triangular Mesh Representation

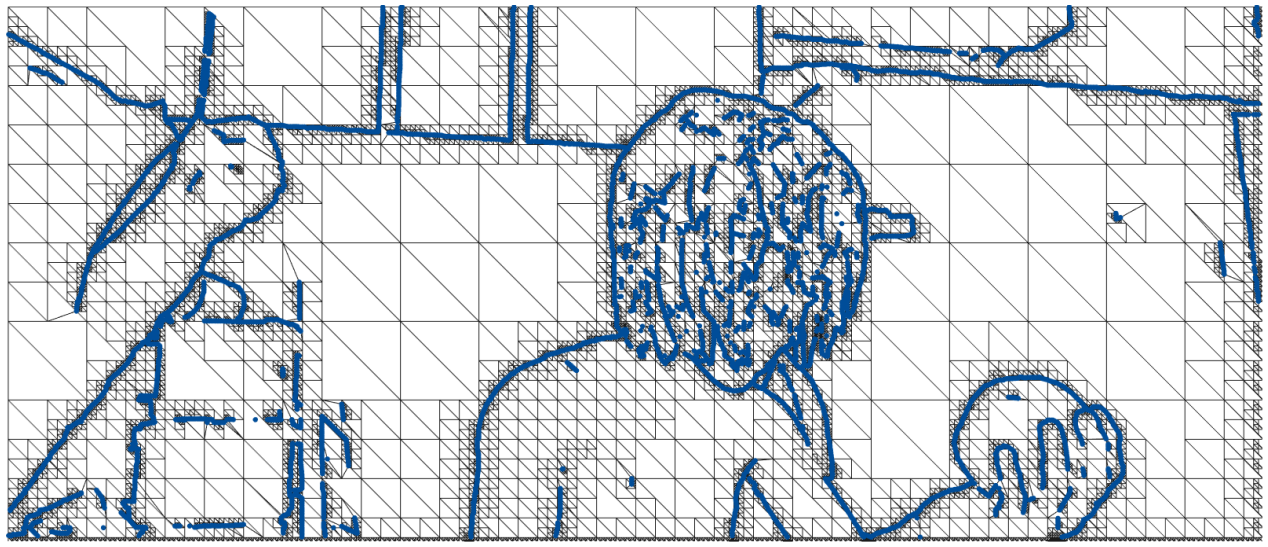
- Subband coefficients and breakpoints - decoded directly onto a triangular mesh.
  - Enabling triangular mesh based warping - most suited to GPU based architectures

- Coarse to fine strategy – follows decoding order
- Mesh elements at a parent level are sub-divided based on appearance of
  - non-zero coefficients at odd coset locations, or
  - novel breakpoints in the form vertices

- Mesh created for
  - scalar depth maps (prior work*) and
  - motion flows – extension of prior work to vector fields

* Y. Li, R. Mathew, and D. Taubman, “Scalable mesh representation for depth from breakpoint-adaptive wavelet coding,” in 2020 IEEE International Workshop on Multimedia Signal Processing (MMSP), 2020.

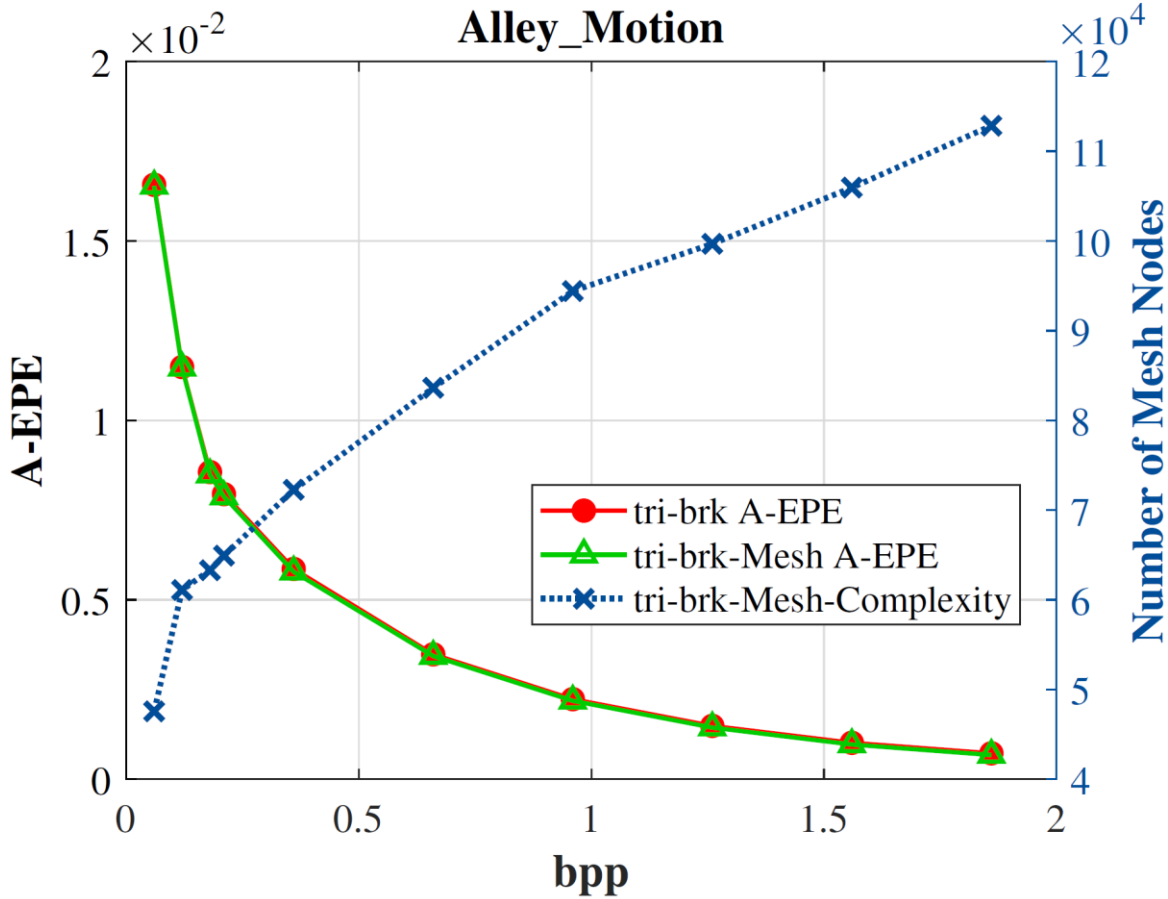
Triangular Mesh Representation



Large triangles: regions of smooth motion
Smaller triangles: object boundaries or complex motion

Triangular cell: affine interpolation of vectors at nodes

Compact representation: vectors needed only at nodes



Mesh model is **accurate**, compared to sample-based field.
Complexity: number of elements is scalable with bit-rate.

Conclusion

- Proposed JPEG 2000 Part 17 extensions
  - For coding of discontinuous media
  - Tri-break – breakpoint adaptive wavelet transform defined on a triangular grid
- Tri-break transform provides improved R-D performance
  - For coding piecewise smooth depth maps and motion flows
- Results show viability of sharing breakpoints
  - For depth maps and motion fields anchored at a common frame
- Triangular mesh representation
  - Able to decode breakpoints and subband coefficients directly onto a triangular mesh
  - Enables view warping based on triangular cells - suited to GPU based architectures