RATE-DISTORTION OPTIMIZED ILLUMINATION ESTIMATION FOR WAVELET-BASED VIDEO CODING

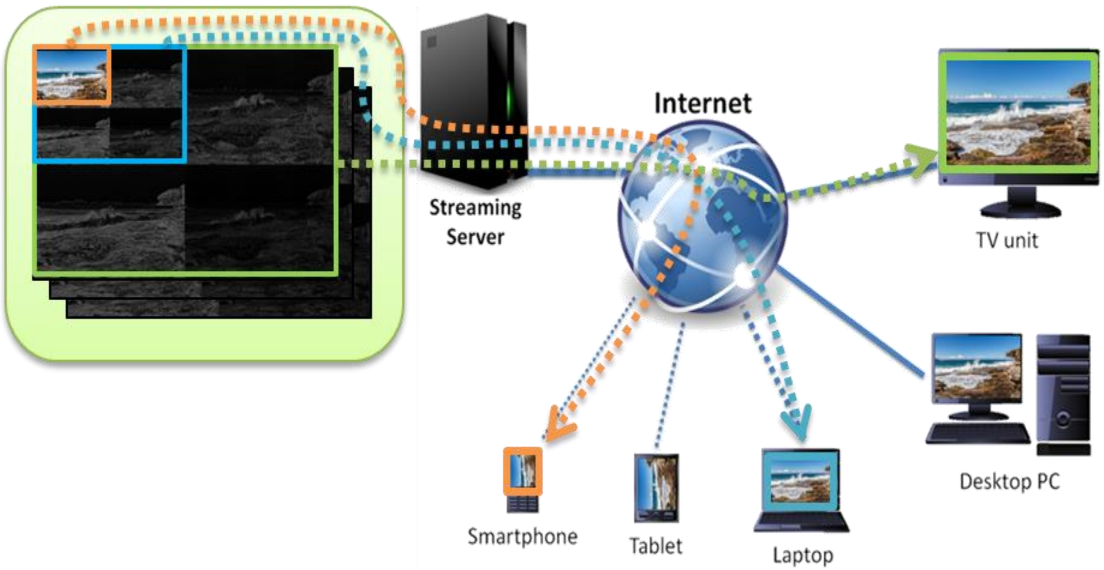
Maryam Haghighat, Reji Mathew, Aous Naman, Sean Young and David Taubman

University of New South Wales (UNSW), Sydney, Australia

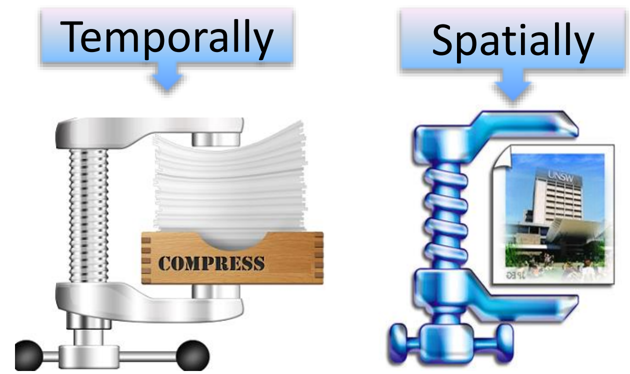
2018 IEEE International Conference on Acoustics, Speech and Signal Processing

15–20 April 2018 • Calgary, Alberta, Canada

Introduction | Scalability and Accessibility in Wavelet-based Video Coding



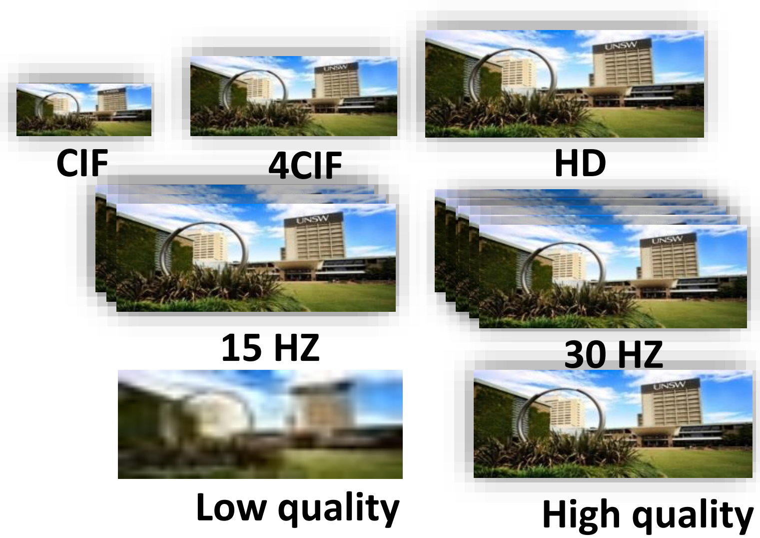
Video Compression



Spatial (resolution) Scalability

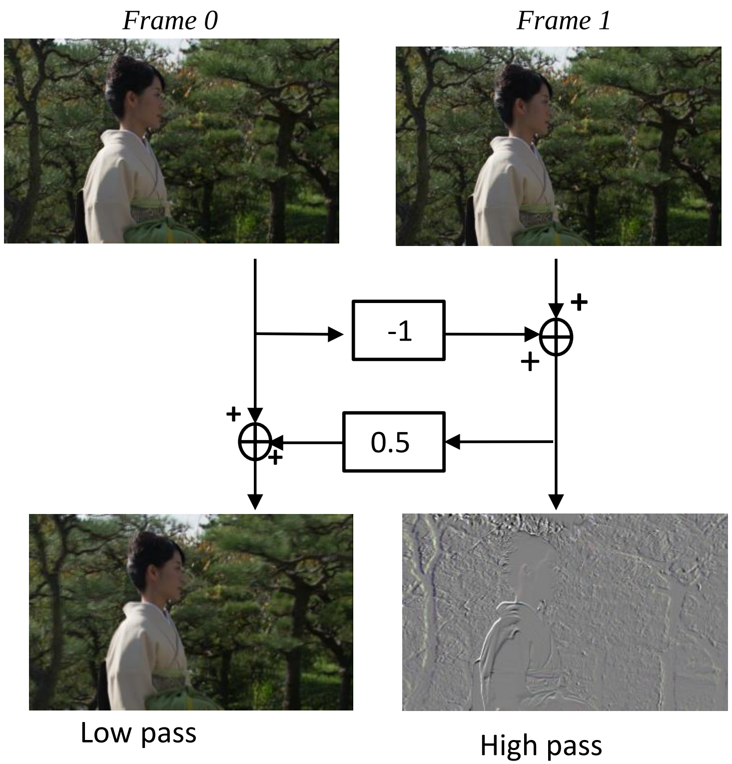
Temporal (frame rate) Scalability

Quality (SNR) Scalability



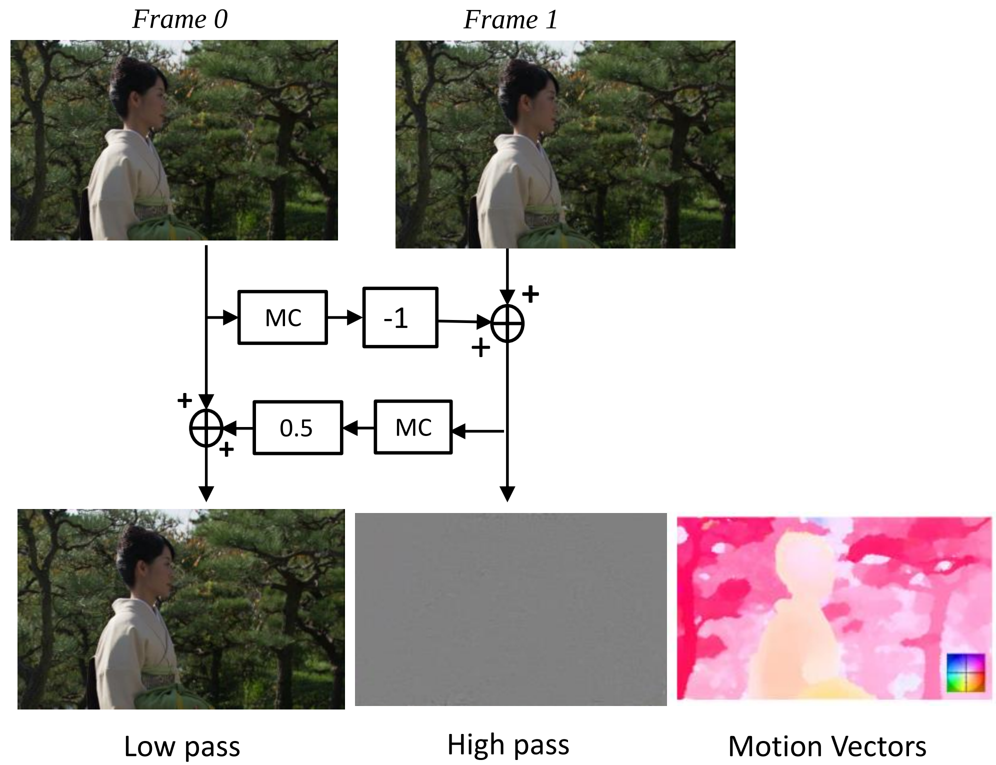
Decomposition of frames to sparser components

Temporal Haar transform



Not sparse enough

Motion Compensated Temporal Haar transform



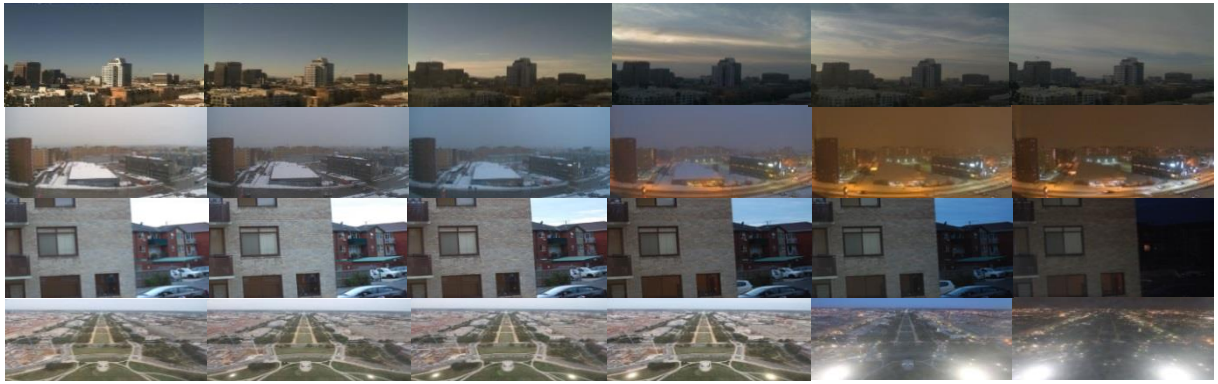
More sparsity than original frames

Image Sets with Illumination Variation

Application: Low frame rate surveillance video.

Indoor images

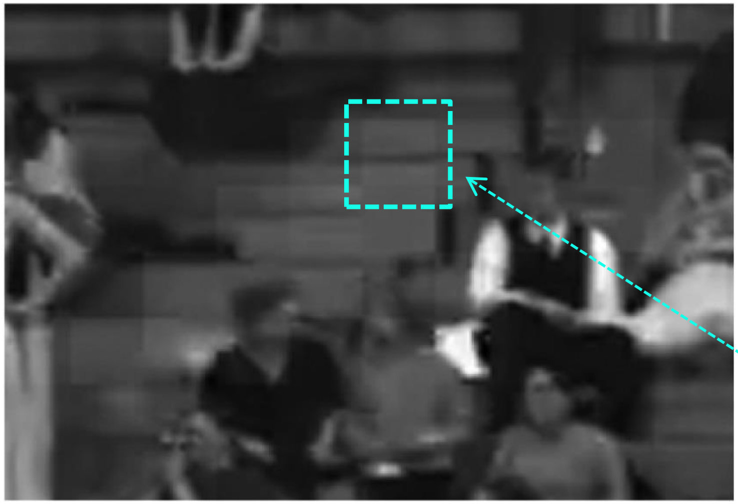
Outdoor images



$f_0[x]$

$f_1[x] = \alpha f_0[x] + \beta$

- Block-based method



Block-based illumination compensated frame

- Illumination compensation is considered in H.264 and HEVC standards in the form of weighted prediction to improve coding efficiency.
- Illumination change is usually modelled by a **scale** and an **offset** and is assumed to be constant within a block which can produce block boundary artefacts!

Contributions

Problems we need to handle:

1. Incorporating the illumination compensation into wavelet-based temporal transformations.

2. Estimation of the illumination field; developing a framework to decompose a sequence of frames into illumination variation fields and texture such that is efficient for compression.

3. Applying a highly scalable, embedded compression framework with R-D optimal termination points for frames and illumination information.

Lifting steps :

$$h[x] = f_1[x] - P(f_0[x])$$

$$l[x] = f_0[x] + U(h[x])$$

Prediction operator:

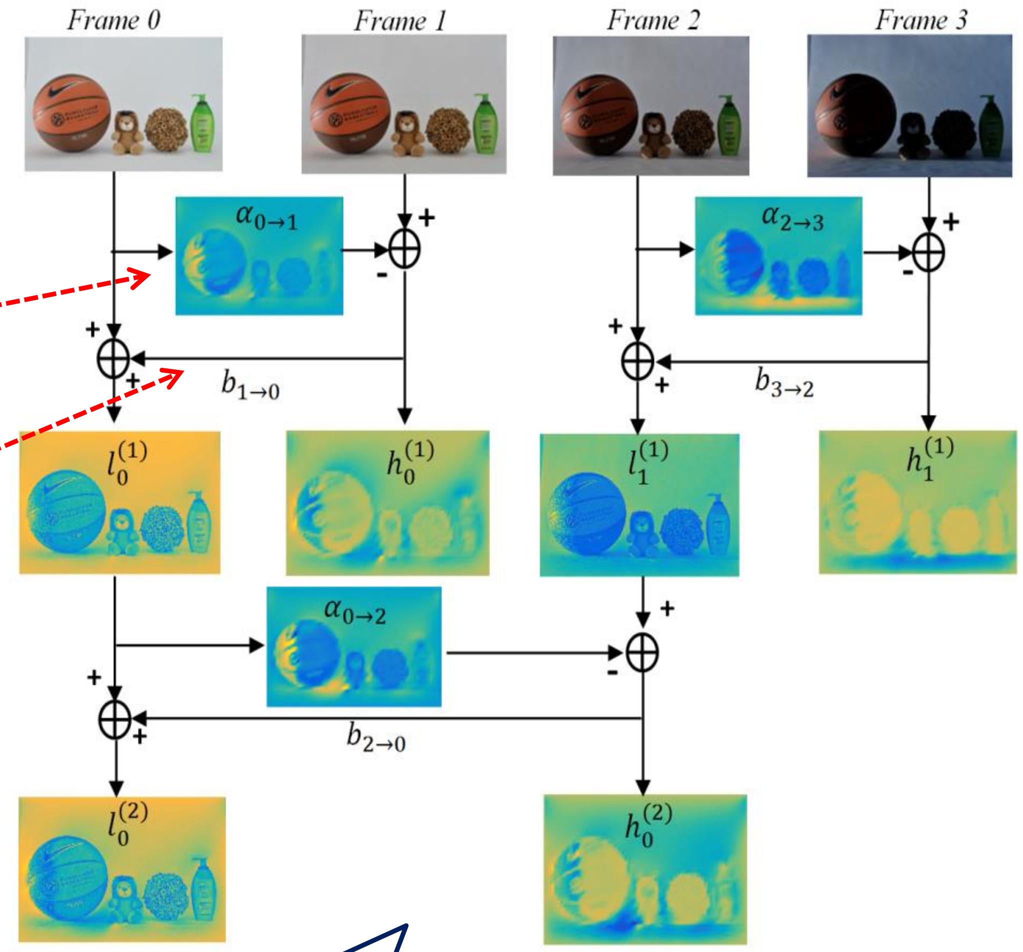
$$P(f_0[x]) = \alpha[x].f_0[x]$$

Update operator:

$$U(f_0[x]) = b[x].h[x]$$

$$b[x] = \frac{\alpha[x]}{1+\alpha^2[x]}$$

b minimizes temporal synthesis gain



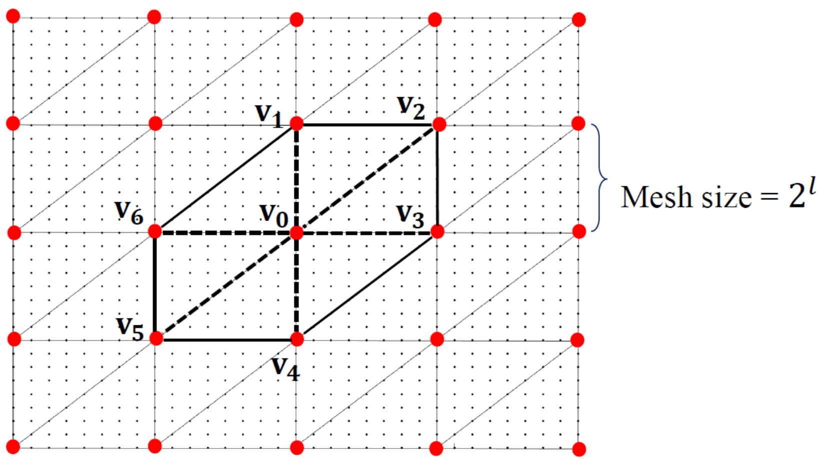
α and b are fixed values in Haar Transform

Problems we need to handle:

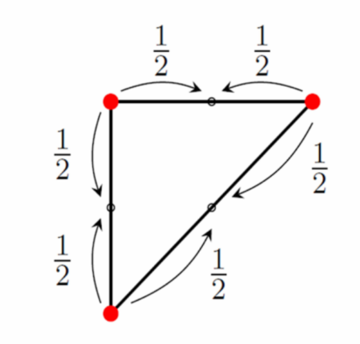
1. Incorporating the illumination information to wavelet-based temporal transformations.

2. Estimation of the illumination field; developing a framework to decompose a sequence of frames into illumination variation fields and texture such that is efficient for compression.

3. Applying a highly scalable, embedded compression framework with R-D optimal termination points for frames and illumination information.



Affine mesh



Affine interpolation

minimize

$$C = \| f_1 - \alpha f_0 - \beta \|_2^2 + \gamma \sum_{i=1}^{6} (\beta[\mathbf{v}_i] - \beta[\mathbf{v}_i])^2 + \gamma \sum_{i=1}^{6} (\alpha[\mathbf{v}_i] - \alpha[\mathbf{v}_i])^2$$

Distortion

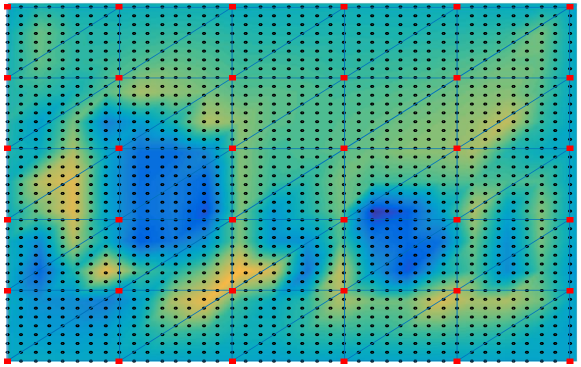
Regularization

Drawbacks of using a **fixed size** mesh and proposed solutions

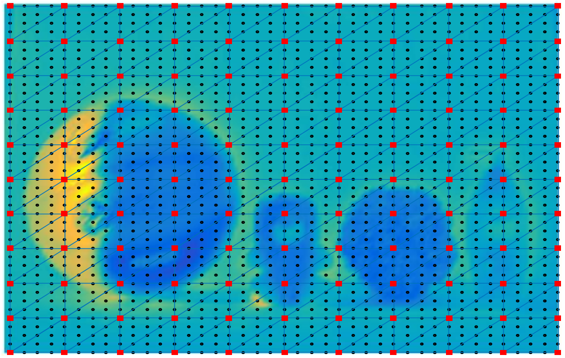
1. The illumination representation is limited to the same grid size over the whole frame.

2. The optimal mesh size is discovered by noting the corresponding R-D performance.

3. Coding efficiency is achieved using a coarse mesh or high regularization parameter.



Mesh size=8



Mesh size=4

In next step:

1. Illumination is estimated based upon the total rate-distortion cost of LIAT subband frames.
2. Smoothness of illumination field is discovered automatically such that rate-distortion is minimized.

- LIAT subband frames are all a function of α and subject to coding

Low-pass frame ($l(\alpha)$)	High-pass frame ($h(\alpha)$)	Illumination field (α)

Problem: Find α such that total coding cost is minimized

$$\underset{\alpha}{\operatorname{argmin}}\, J = J_\alpha + J_{h(\alpha)} + J_{l(\alpha)}$$

Coding cost of α — Coding cost of h — Coding cost of l

- How to model coding cost (J)?

- We model the R-D cost (J) using the high rate model:

$$J = D + \lambda L$$

$$D = \sum_{s,n} D_{s,n} = \sum_{s,n} |y_{s,n}|^2 g_{s,n} e^{-aL_{s,n}}$$

$$L = \sum_{s,n} L_{s,n} = \sum_{s,n} \left(L_{s,n} + L^{\sigma}_{s,n}\right)$$

$L^{\sigma}_{s,n}$: required bits to communicate the significance state (zero or nonzero)

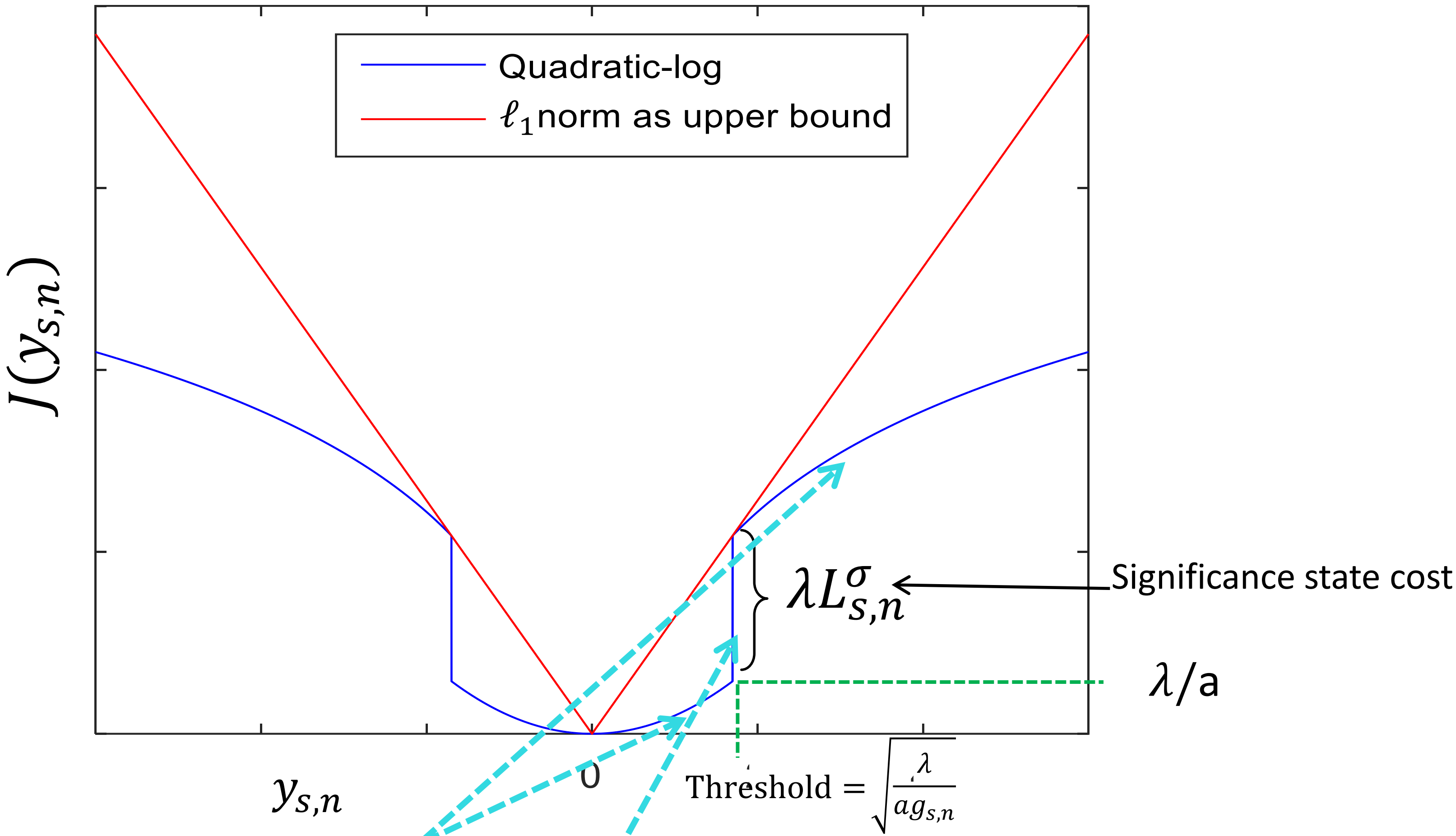
- $y_{s,n}$ transformed coefficient of the LIAT subband s at position n.
- $g_{s,n}$ product of the spatial and temporal synthesis gain.

- Cost function $J(y_{s,n})$ at R-D optimal operation point is:

$$J(y_{s,n}) = \begin{cases} g_{s,n}|y_{s,n}|^2 & \text{if } |y_{s,n}|^2 \leq \frac{\lambda}{a g_{s,n}} \\ \frac{\lambda}{a} + \frac{\lambda}{a} \ln\left(\frac{g_{s,n}|y_{s,n}|^2}{\lambda/a}\right) + \lambda L^{\sigma}_{s,n} & \text{otherwise} \end{cases}$$

Quadratic-log function

R-D Optimized Illumination Estimation in LIAT Framework



$$J(y_{s,n}) = \begin{cases} g_{s,n}|y_{s,n}|^2 & \text{if } |y_{s,n}|^2 \le \frac{\lambda}{a g_{s,n}} \\ \frac{\lambda}{a} + \frac{\lambda}{a}\ln\left(\frac{g_{s,n}|y_{s,n}|^2}{\lambda/a}\right) + \lambda L^{\sigma}_{s,n} & \text{otherwise} \end{cases}$$

Optimization Problem

$$\underset{\alpha}{\operatorname{argmin}}\ J = J_\alpha + J_{h(\alpha)} + J_{l(\alpha)}$$

$$\underset{\alpha}{\operatorname{argmin}}\ C(\alpha) = \|m_\alpha y_\alpha\|_1 + \|m_h y_{h(\alpha)}\|_1 + \|m_l y_{l(\alpha)}\|_1$$

$$m_{s,n} = \frac{\lambda/a + \lambda L^{\sigma}_{s,n}}{\sqrt{\frac{\lambda}{a g_{s,n}}}}$$

Slope of ℓ_1 function for each coefficient

- ✓ We solve the optimization problem using ADMM (alternating direction method of multipliers)

- ❖ We can interpret our compression inspired convex formulation as a way to effectively distribute the information in a sequence between multiplicative illumination terms and non-multiplicative residual terms.

Contributions

Problems we need to handle:

1. Incorporate the illumination information to wavelet-based temporal transformations.

2. Estimation of the illumination field; developing a framework to decompose a sequence of frames into illumination variation fields and texture such that is efficient for compression.

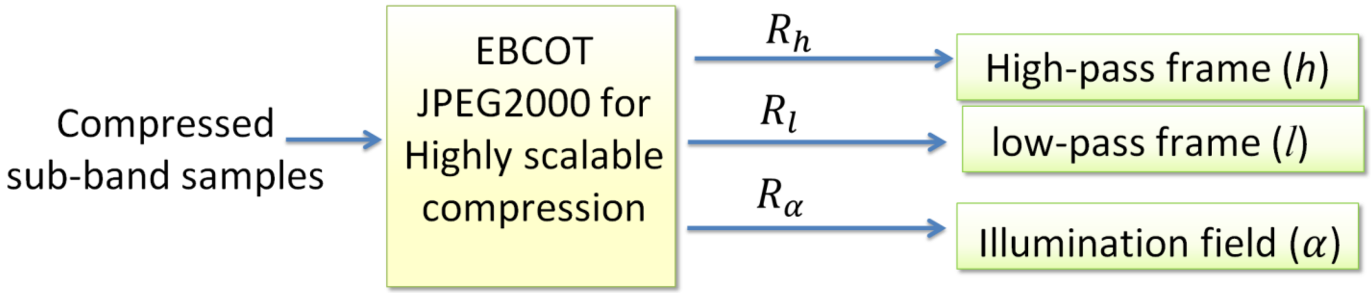
3. Applying a highly scalable, embedded compression framework with R-D optimal termination points for frames and illumination information.

✓ The rate-allocation problem is to find the optimal truncation points in Embedded Block Coding with Optimized Truncation (EBCOT) so as to minimize overall distortion s.t. an overall bit-rate constraint:

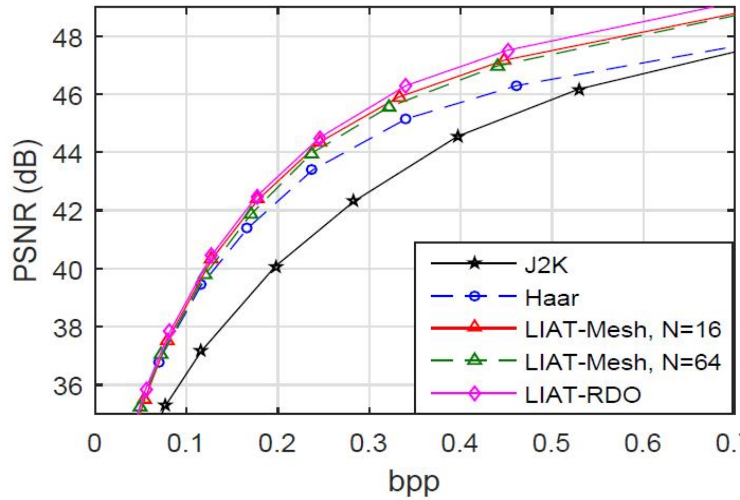
$$\min \; J(\lambda_j) = D(\lambda_j) + \lambda_j \sum_s R_s(\lambda_j)$$

λ_j= Lagrangian multiplier associated with the quality layer *j*.

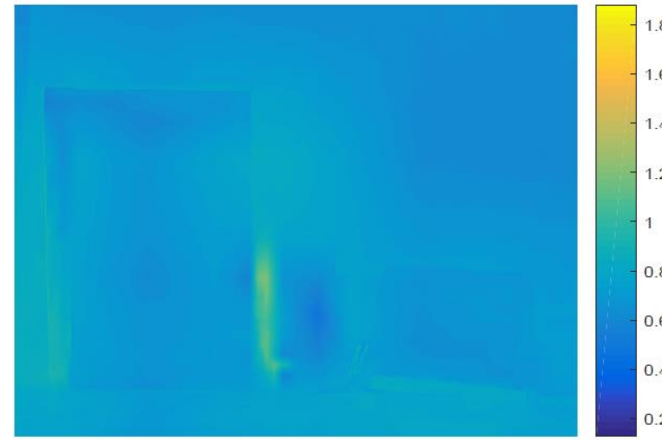
✓ Using a linearized distortion model, we achieve a R-D optimal bit allocation through some gain factors which reflect how the error in each LIAT subband spreads in the final reconstructed frames.



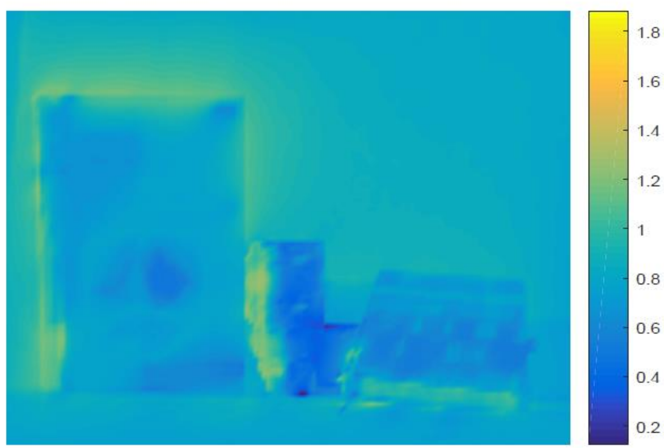
R-D optimized vs Mesh based illumination estimation



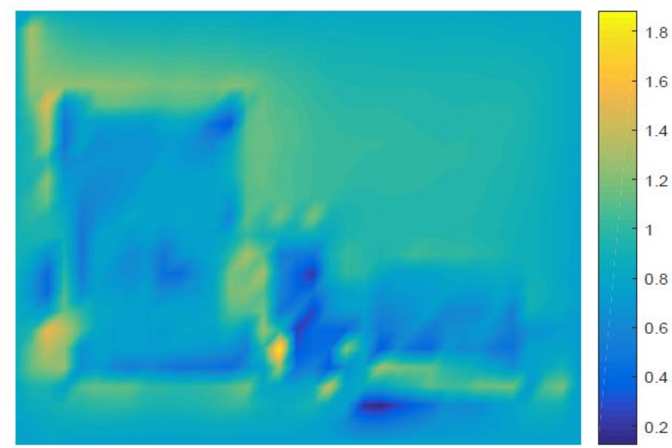
(a) R-D results



(b) α, R-D optimized

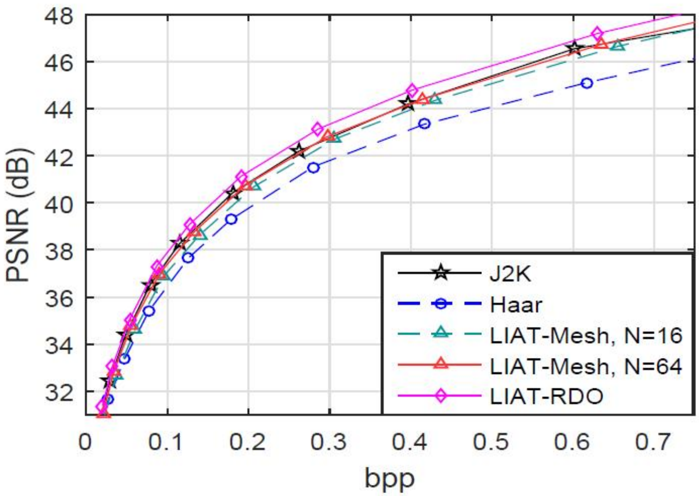


(c) α, mesh size= 16

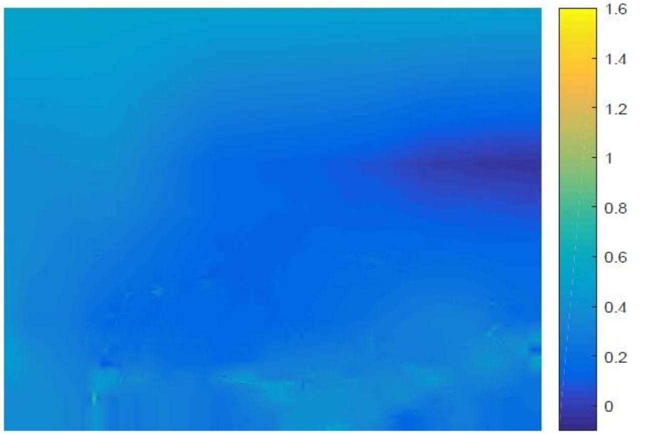


(d) α, mesh size= 64

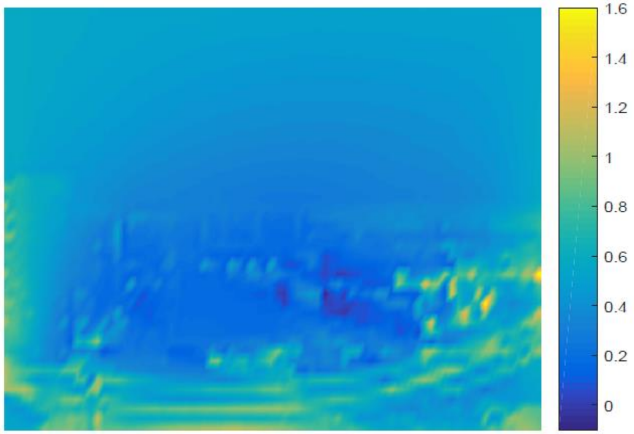
R-D optimized vs Mesh based illumination estimation



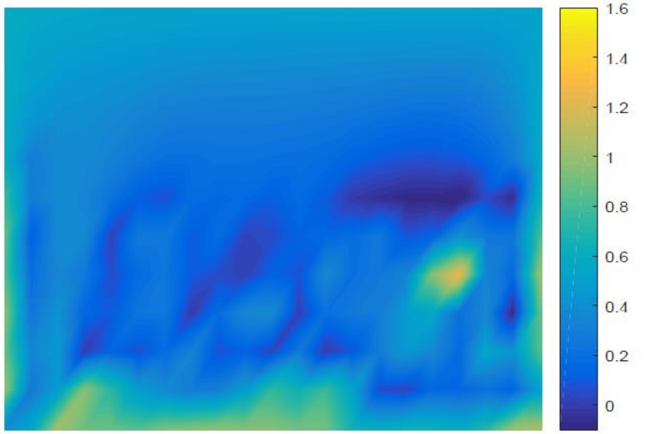
(a) R-D results



(b) α, R-D optimized

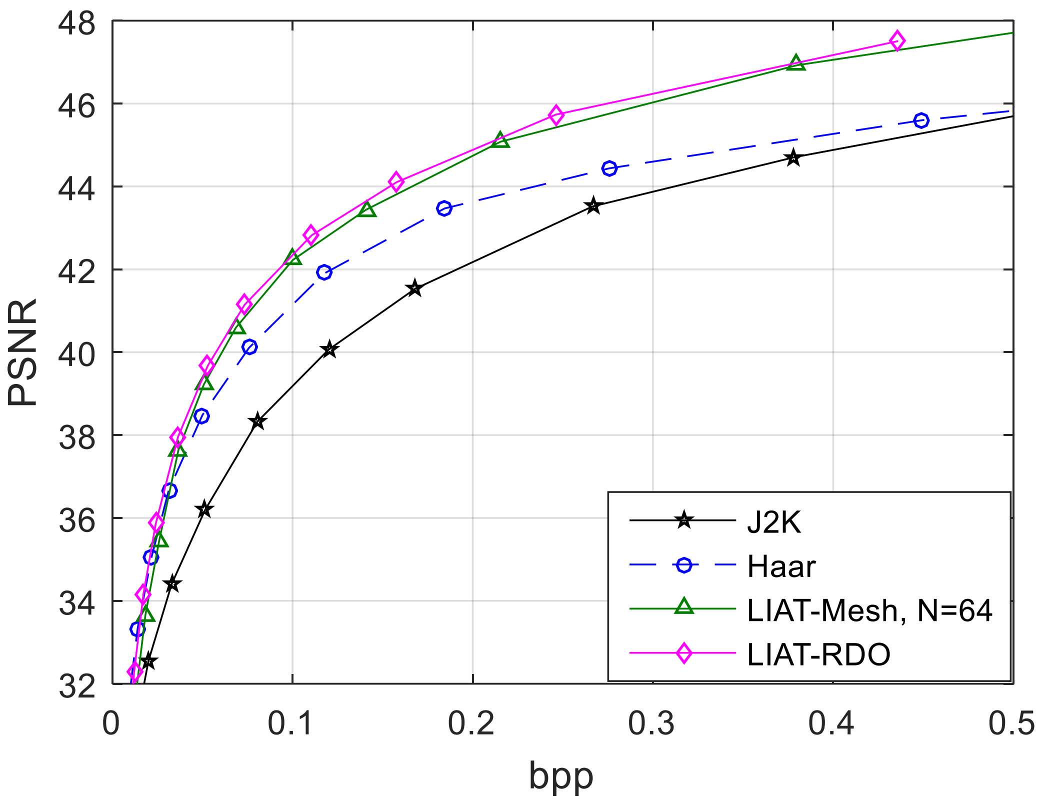


(c) α, mesh size= 16

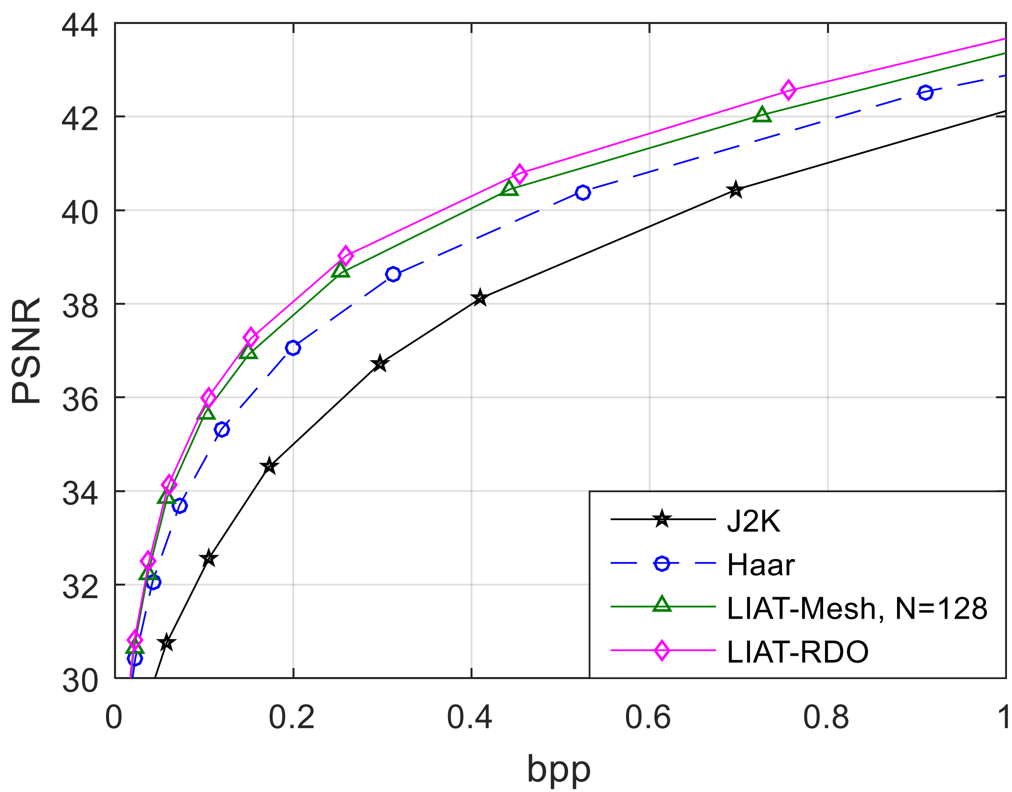


(d) α, mesh size= 64

Results for two temporal levels of LIAT



Results for two temporal levels of LIAT



Thanks for your attention!

Questions?

Contact: maryam.haghighat@unsw.edu.au