Video Coding using Dual-Tree Wavelet Transform

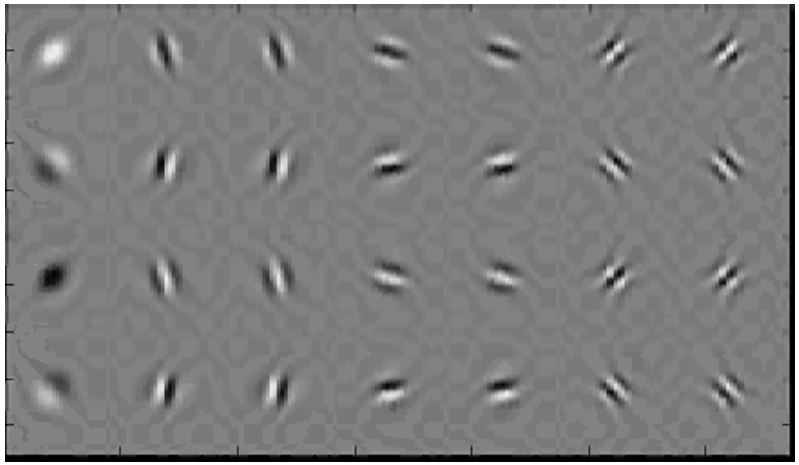
Beibei Wang[1], Yao Wang[1], Ivan Selesnick[1], Anthony Vetro[2]

1. Polytechnic University, Brooklyn, NY, 19020
2. MERL, Cambridge, MA 02139

Dual-tree DWT (DDWT)

- First proposed by Kingsbury, extended to 3-D by Selesnick
- 3-D DDWT is orientation and motion selective ☺
  - Each wavelet basis has a particular spatial orientation and motion direction.
- But it has more bases than the 3-D DWT (28 high subbands instead of 7, 4 low subbands instead of 1) ☹

Standard DWT



Dual-tree DWT

Why using 3D DDWT for video coding?

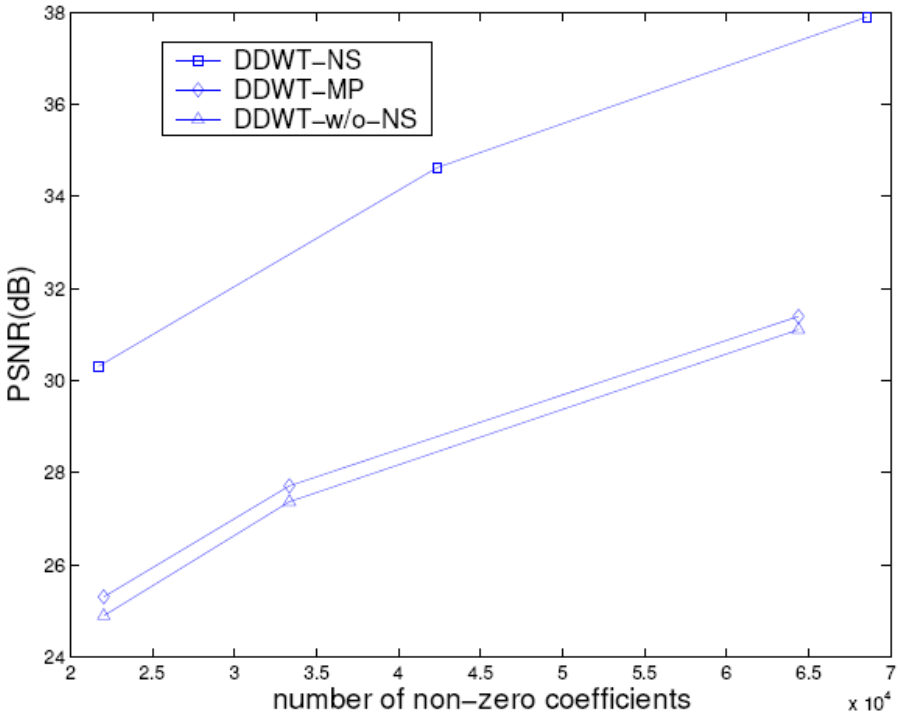
- Has the potential to represent video efficiently WITHOUT requiring motion estimation
- Has the computational efficiency of separable transforms
  - First apply separable DWT
  - Then linearly combines the resulting subbands
- Can offer full spatial, temporal, quality scalability
  - Such scalability is desirable considering the nature of the networks and users
  - More scalable than coders using motion estimation, as no motion vectors are coded

But…

- 3-D DDWT is an overcomplete transform
  - if using complex coefficients -> 8 : 1 redundancy
  - if only using real coefficients
    - 4 : 1 redundancy
    - Perfect Reconstruction
- Overcomplete transform doesn't necessarily mean inefficient coding
  - May require fewer significant coefficients to describe a signal

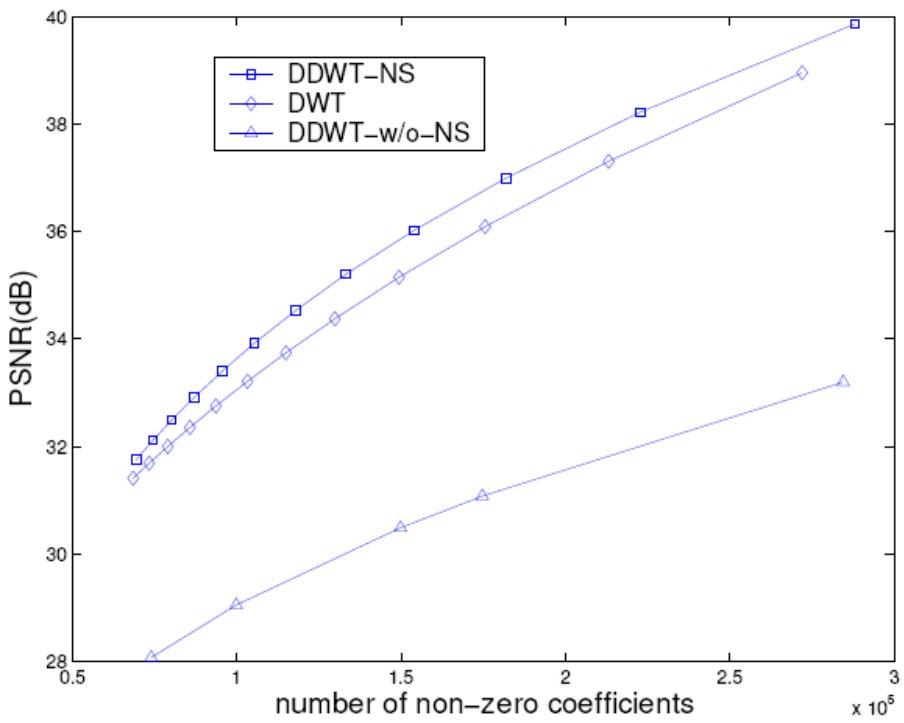
How to deduce the significant coefficients?

- Matching Pursuit [Mallat]:
  - Iteratively select the largest coefficient for the residual signal
- Noise Shaping [Kingsbury]:
  - Iteratively select coefficients larger than a threshold
  - modify selected coefficients to compensate for the loss of small coefs
  - gradually reduce the threshold.

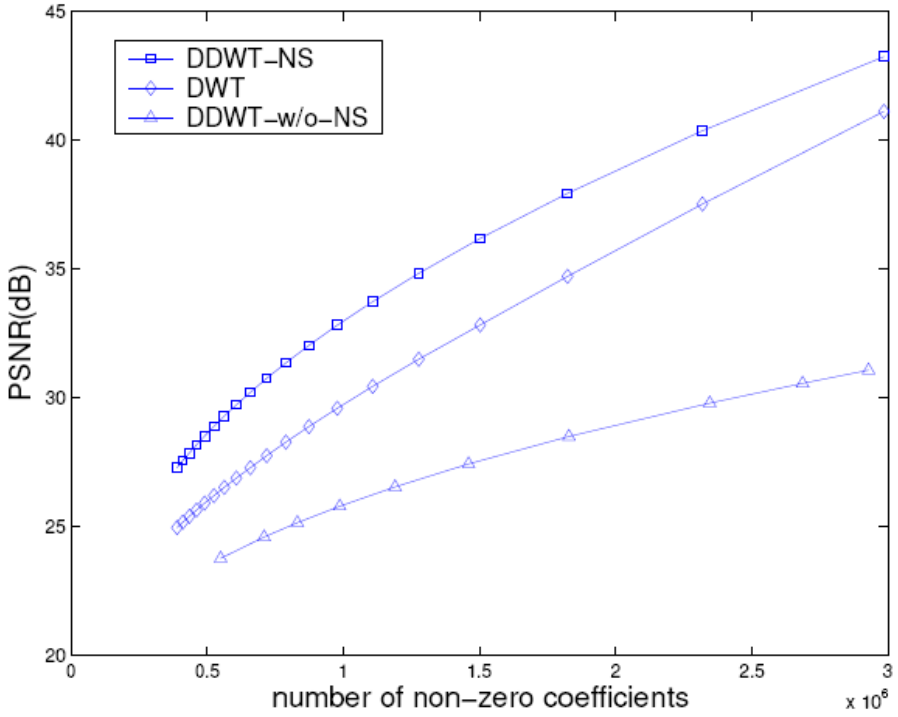


- MP requires extensive computation
- Compared to the results by simply choosing the largest *N* coefficients
  - MP provides only marginal gain
  - NS yielded much better image quality (5-6 dB higher)

Noise shaping applied to 3D DDWT



(a)Foreman (QCIF)



(b)Mobile_Calendar (CIF)

With the same number of retained coefficients, DDWT_NS yields higher PSNR than DWT!

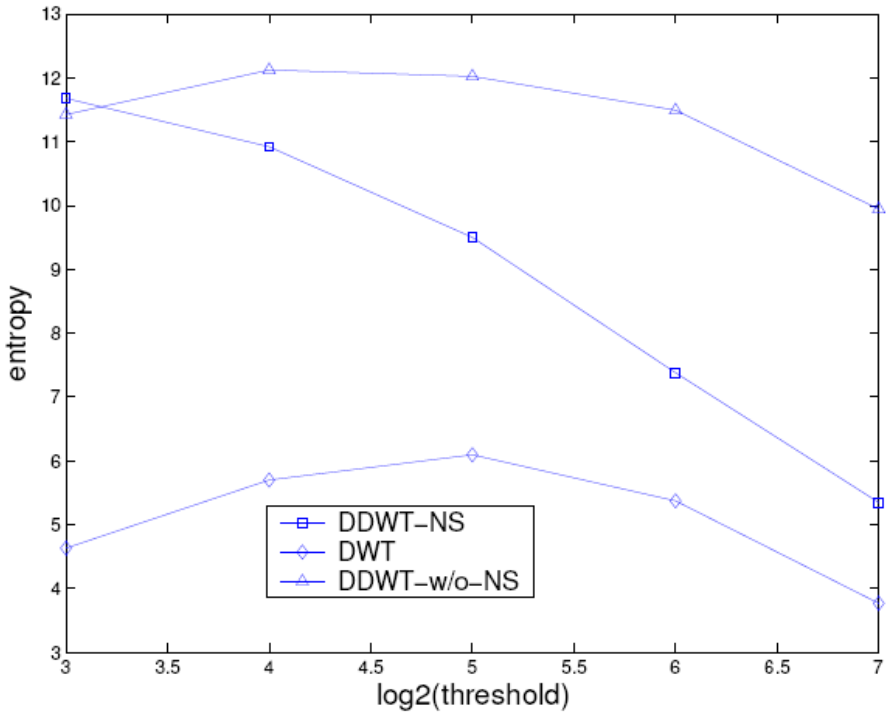
The Correlation Between Subbands

- The DDWT is a redundant transform
  - Subbands are expected to have non-negligible correlations.
- Wavelet coders code the location and magnitude information separately
  - Examine the correlation in the location and magnitude separately.

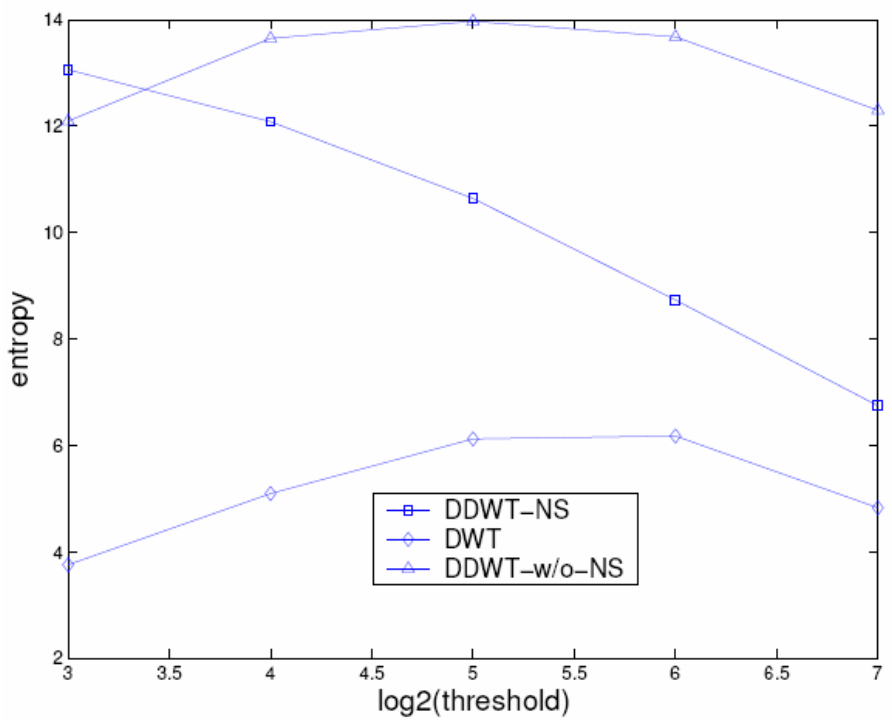
Correlation in Significance Maps

- Motivation:
  - Only a few subbands have significant energy for an object feature at a particular location
- How to verify this hypothesis?
  - The significance vector
    - For a given threshold *T*, set the significance bit to “1” if the corresponding wavelet coefficient is above *T*
    - For a given spatial location, the significance bits of all 28 subbands form a binary vector
  - The possible patterns of the significance vector are not random!
    - Evaluate the entropy of the significance vector
    - The vector entropy should be much lower than 28

Entropy of the Significance Vector



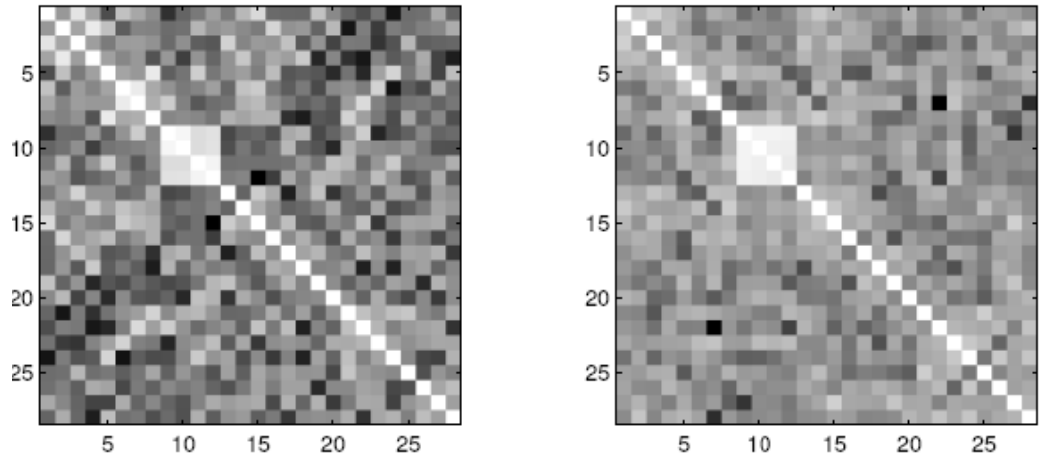
(a)Foreman (QCIF)



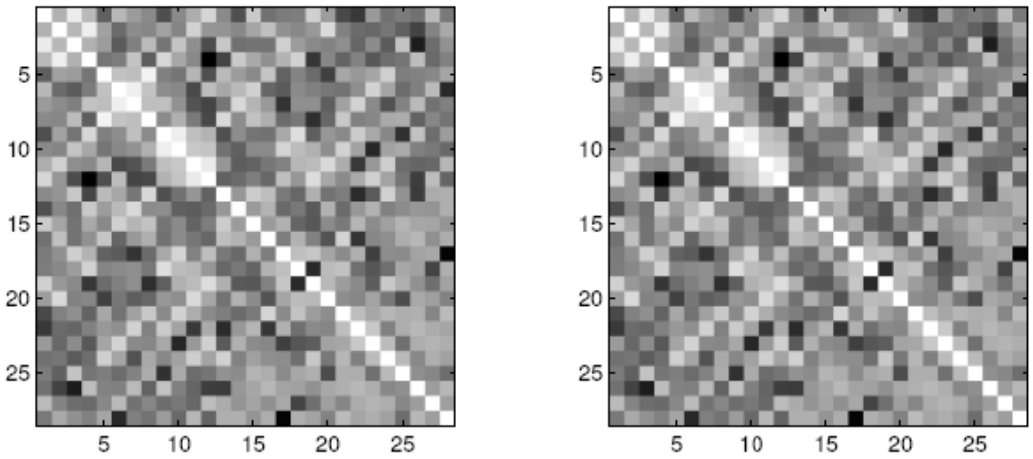
(b)Mobile_Calendar (CIF)

- DWT has 7 high subbands, the entropy is ~4-6
- DDWT has 28 high subbands, the entropy before noise shaping is ~10-12
- After noise shaping, the entropy is ~6 for T large
- The location information can be coded efficiently by vector coding across subbands!

Correlation in coefficient values



(a)Foreman (QCIF)



(b)Mobile_Calendar (CIF)

- Only a few subbands have strong correlation
- Other subbands are almost independent.
- After noise shaping, the correlation is reduced further

The correlation matrices of the 28 subbands
Left: w/o_NS; Right: with NS

➢The grayscale is logrithmically related to the absolute value of the correlation.

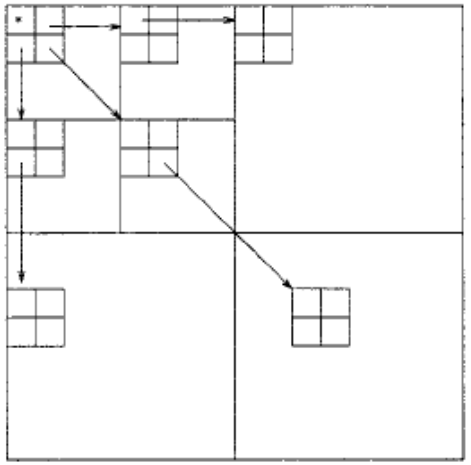
➢The brighter colors represent higher correlation.

3-D DUAL-TREE WAVELET VIDEO CODECS

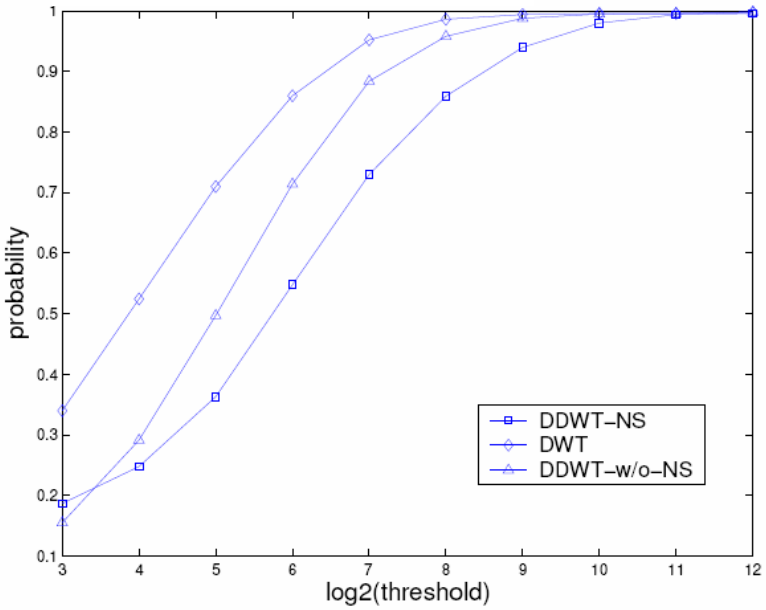
- Fewer coefficients do not necessarily mean fewer bits
  - whether the coefficient location/magnitude can be coded efficiently
  - More subbands in the 3-D DDWT
- Two video codecs using the 3-D DDWT
  - DDWT-SPIHT
    - applies the well-known 3D SPIHT on each of the four DDWT trees
  - DDWTVC
    - exploits the inter-subband correlation in the significance maps
    - code the sign and magnitude information within each subband separately.

DDWT-SPIHT

- 3-D SPIHT parent-children probability:
  - an insignificant parent does not have significant descendants
- Compared to DWT, DDWT has similar
  - Tree structure
  - parent-children probability
- Coding scheme
  - applied the 3-D SPIHT on each DDWT tree after noise shaping.



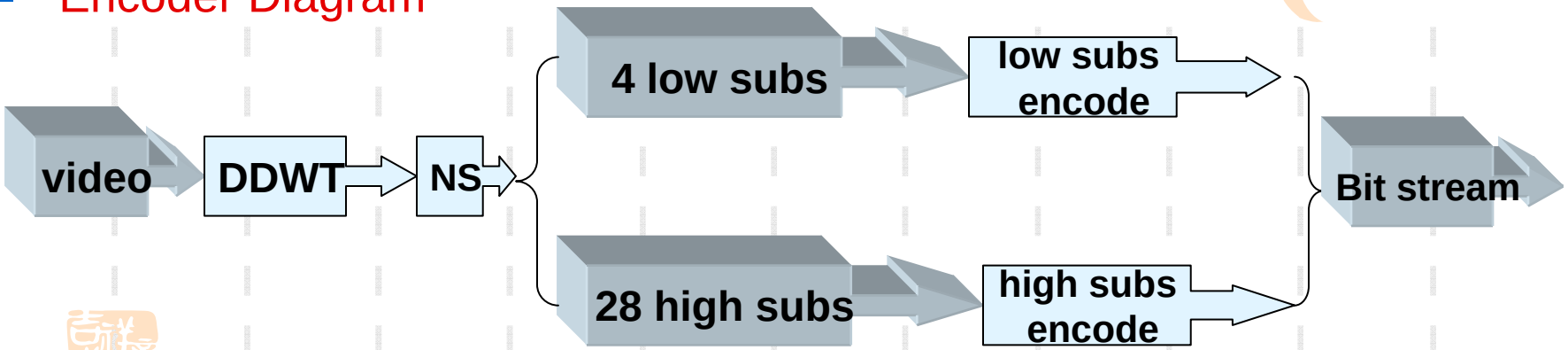
Parent-Children relationship (2-D)



Parent-Children Probability For “Forman"

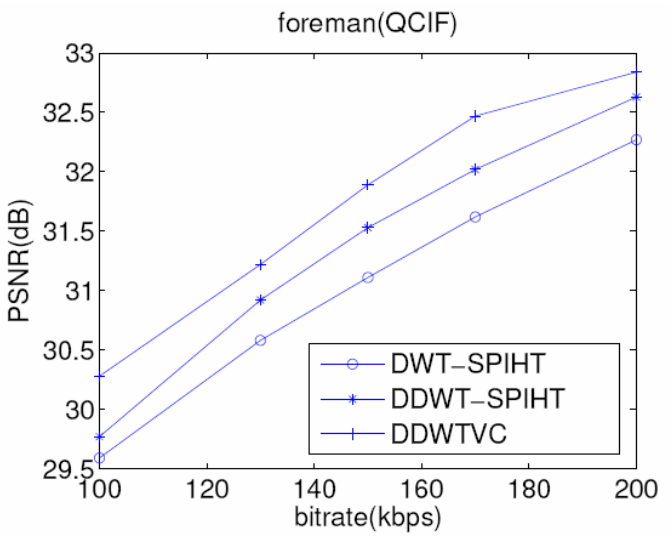
DDWTVC

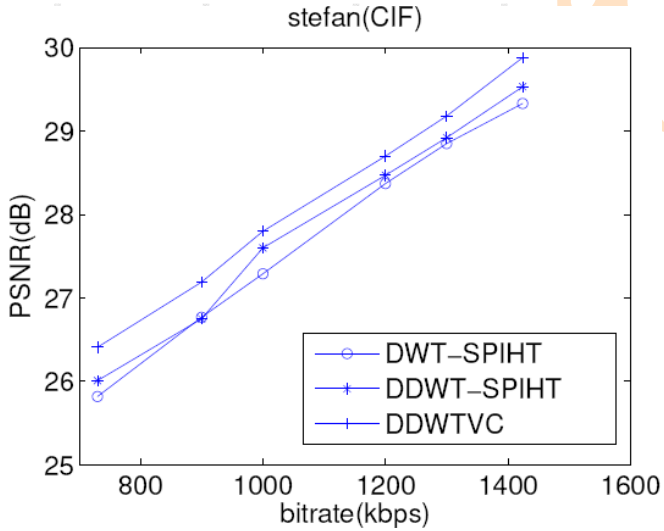
- Encoder Diagram

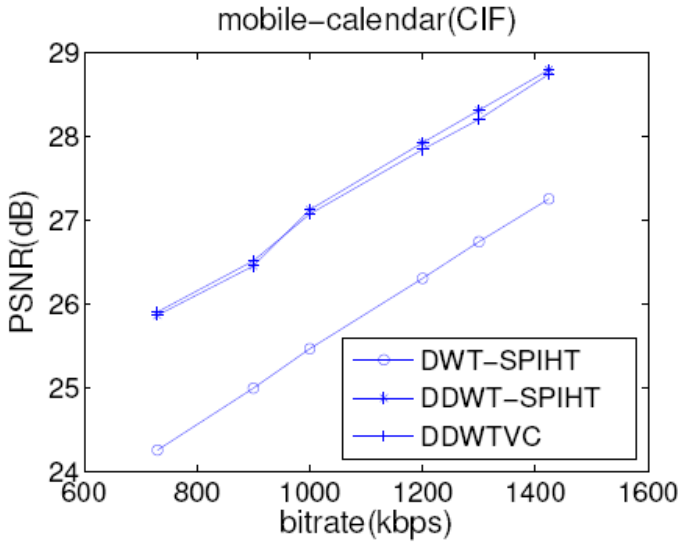


- Coding Algorithms
  - Bit plane coding as other wavelet-based coders
  - Significance Map
    - Arithmetic vector coding across subbands
  - Sign Information
    - Predict the sign based on the correlation between subbands
  - Magnitude Refinement
    - Using context modeling to exploit the spatial correlation among neighboring coefficients within the same subband.

Experimental results



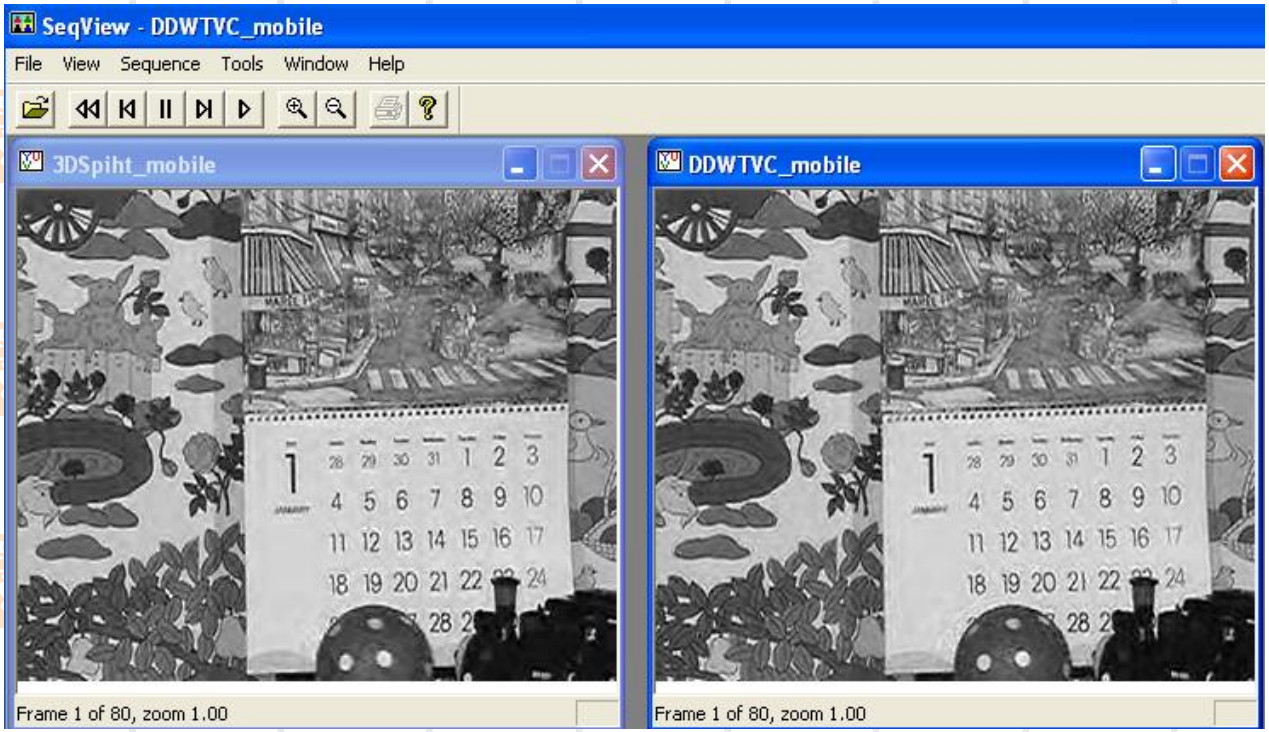




- Both DDWT-SPIHT and DDWTVC have better performance than DWT-SPIHT
- DDWTVC has comparable or better performance than DDWT-SPIHT

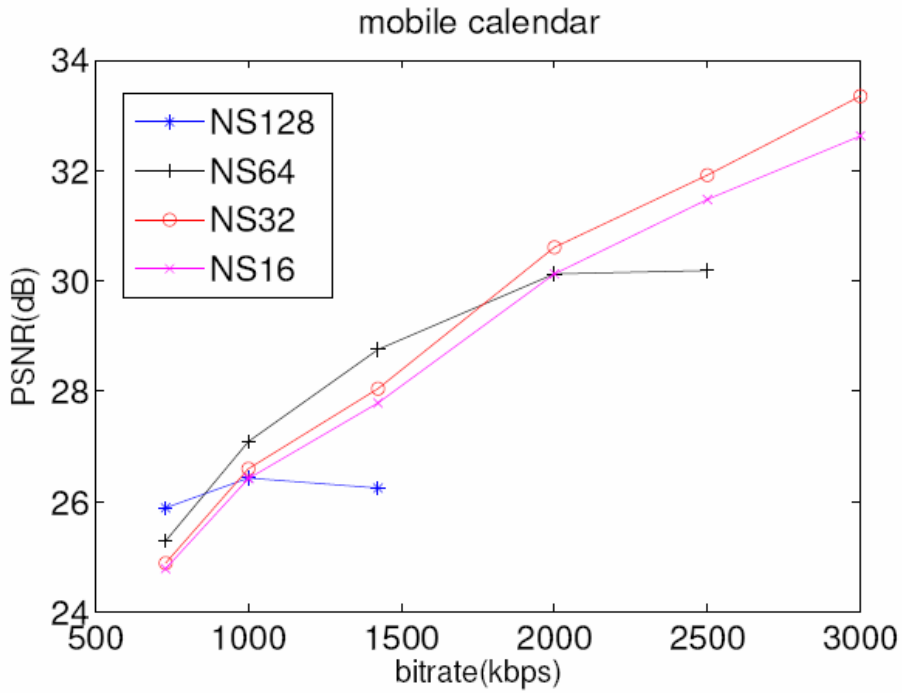
Sample video sequences

- Subjectively,
  - Both DDWT-SPIHT and DDWTVC preserve edge and motion information better than DWT- SPIHT
  - DWT-SPIHT exhibits blurs in some regions and when there are a lot of motions.



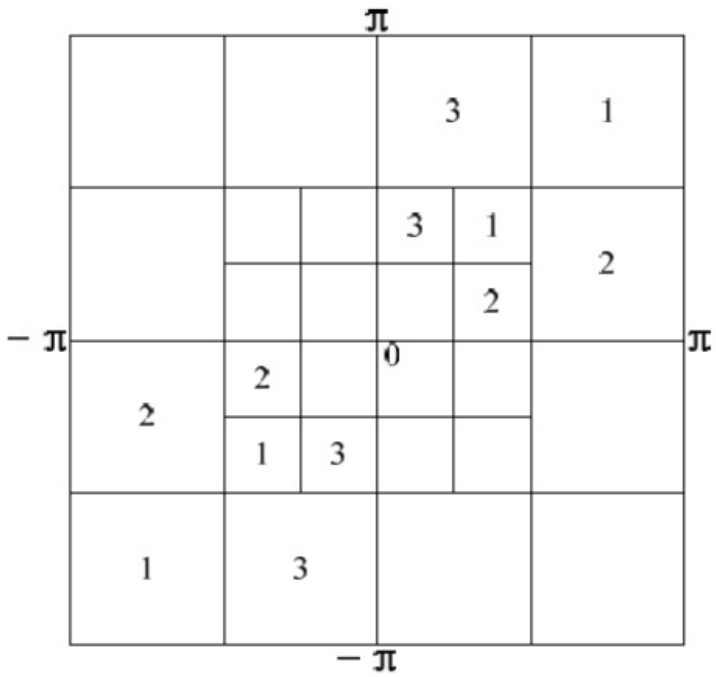
Scalability

- With the coefficients derived from a chosen threshold,
  - DDWTVC produces a fully scalable bit stream
- Noise shaping modifies previously chosen large coefficients
  - R-D Optimal only for the highest bit rate associated with this threshold.

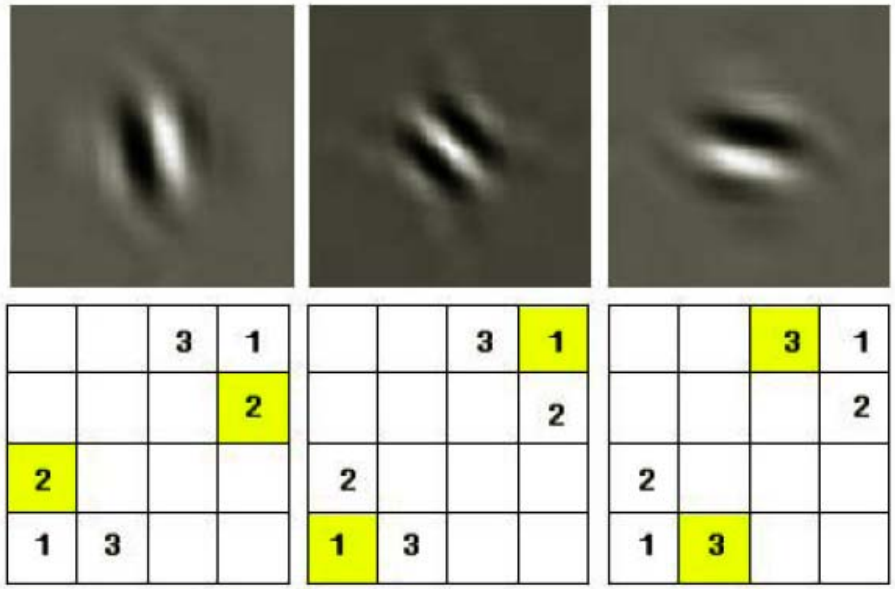


- 1 dB coding efficiency penalty for full scalability (for threshold 32).

Isotropic DDWT Decomposition

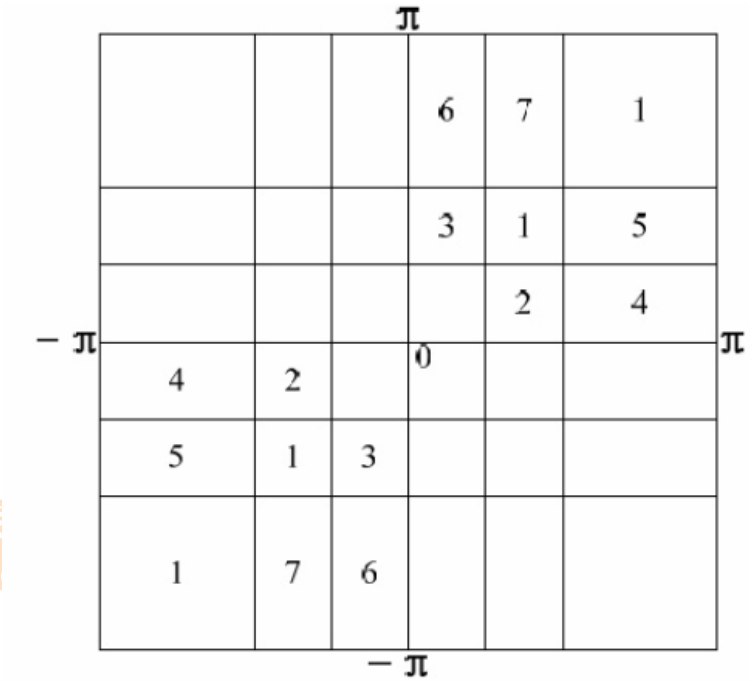


(a)Isotropic tiling

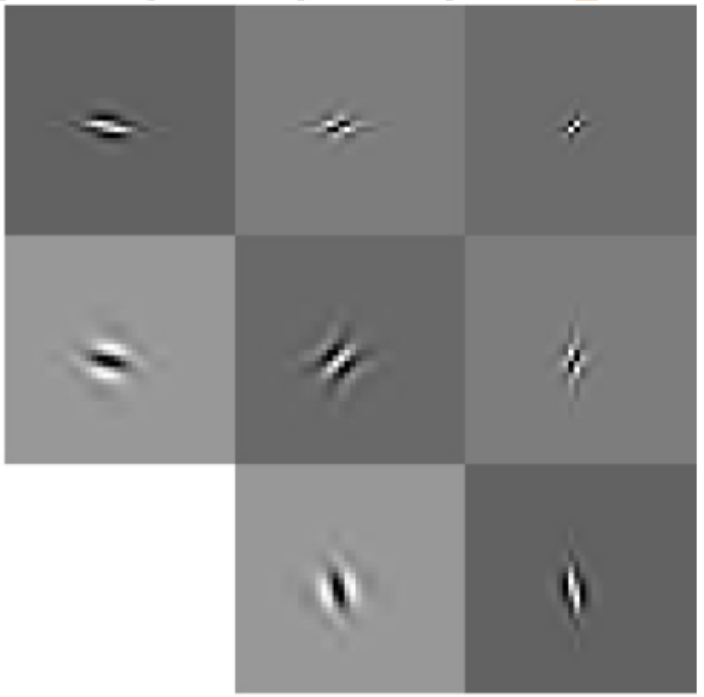


Typical wavelets associated with the isotropic 2-D DDWT.

Anisotropic DDWT Decomposition

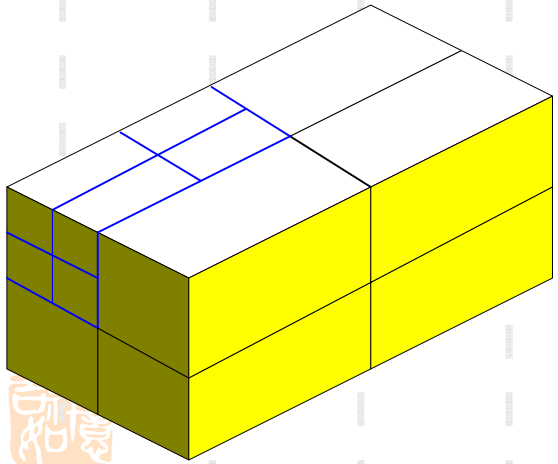


(b)Anisotropic tiling

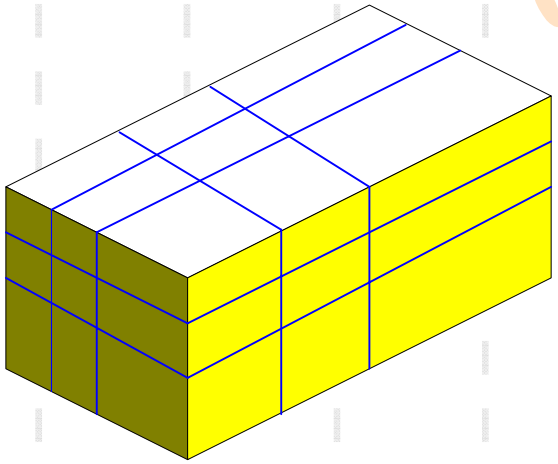


Typical wavelets associated with 2-D anisotropic DDWT

Anisotropic DDWT Decomposition (Con'd)



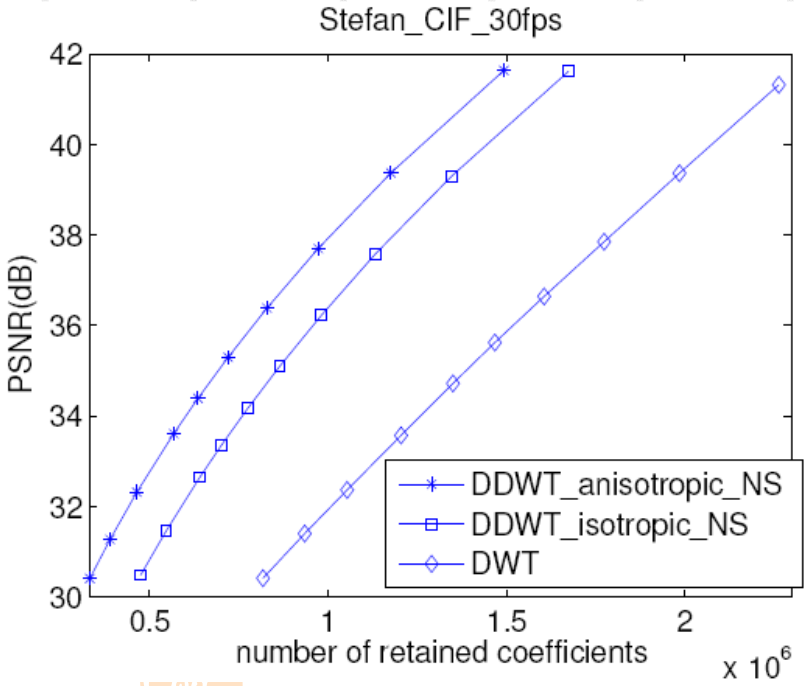
Isotropic decomposition

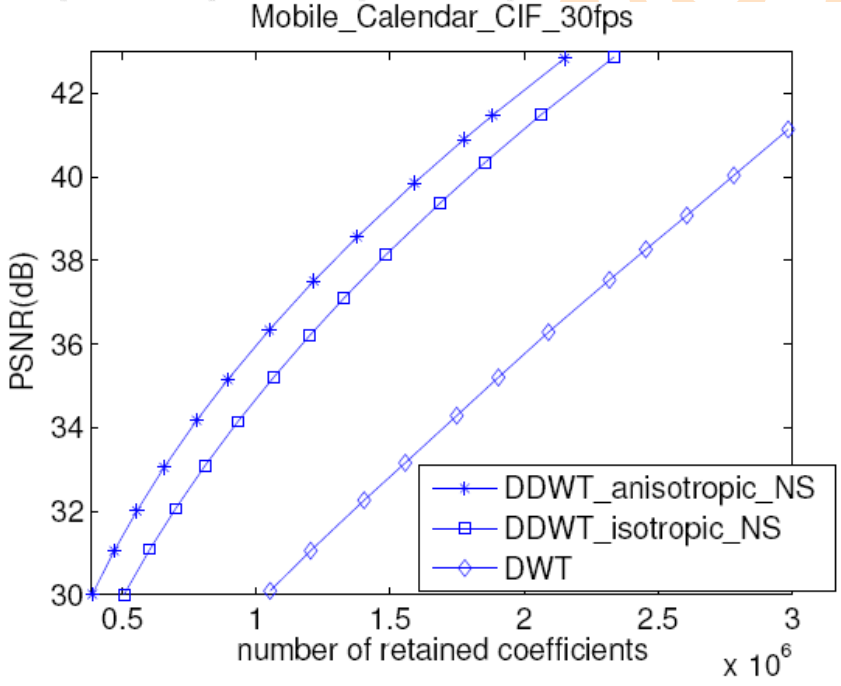


Anisotropic decomposition

- Anisotropic decomposition splits not only subband LLL, but subbands LLH, LHL, HLL, HLH, HHL, LHH
- Anisotropic decomposition allows different number of decompositions along temporal, horizontal and vertical directions

Anisotropic DDWT for Video Representation



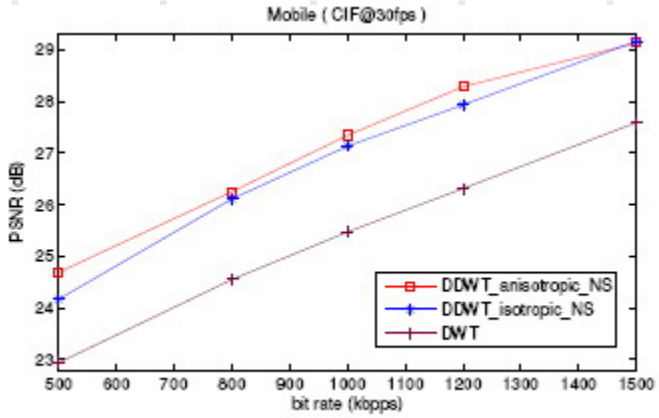


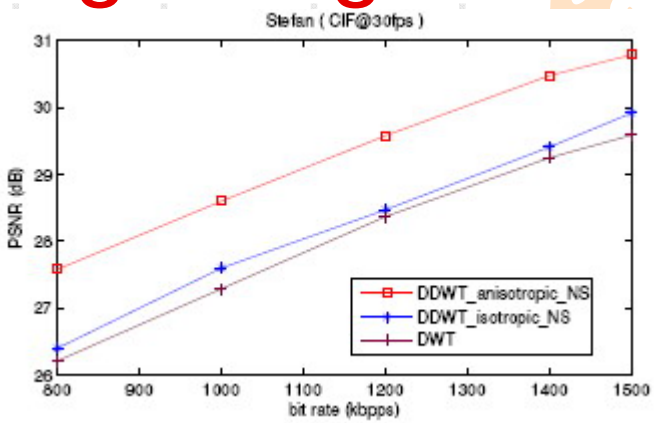
Mobile-Calendar (CIF)

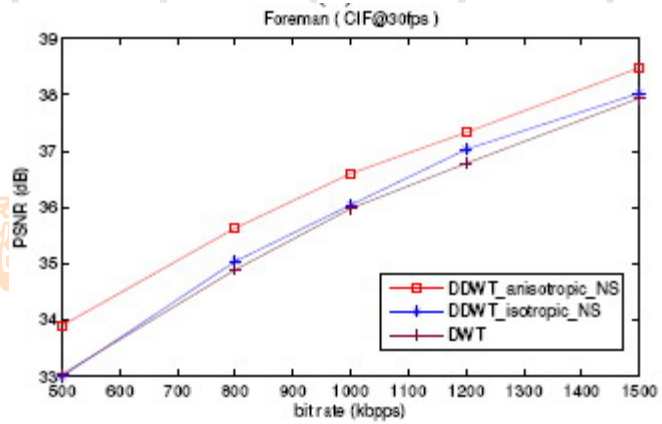
Stefan (CIF)

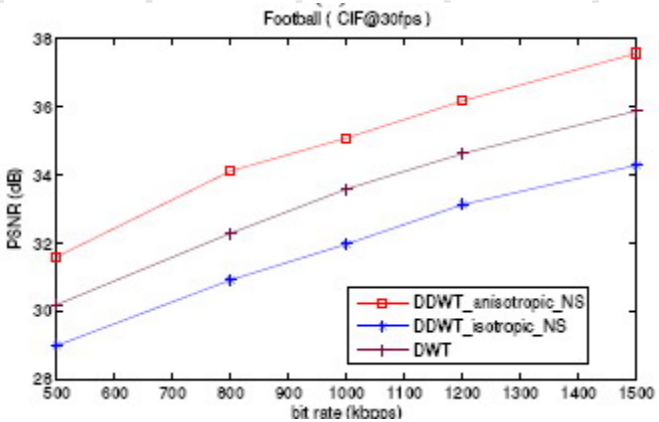
Anisotropic decomposition has better PSNR performance after Noise Shaping

ADDWT Video Coding using SPIHT









For smoother motion sequences

Both DDWT-SPIHT and ADDWT-SPIHT achieve higher PSNR (up to 2 dB) than the DWT-SPIHT

ADDWT outperforms the DDWT up to 1 dB.

For higher motion sequences

DDWT-SPIHT is worse than DWT-SPIHT

ADDWT-SPIHT provides significant gains (up to 3 dB) over the DDWT and 2 dB gain over DWT-SPIHT

Conclusion

- 3-D DDWT has the potential for efficient video coding WITHOUT motion estimation!
- Noise shaping can reduce the number of coefficients to below that required by DWT (for the same video quality).
- Strong correlation in the location of significant coefficients across subbands, but not in the values
- Both DDWT-SPIHT and DDWTVC are better than DWT-SPIHT, both objectively and subjectively.
- Anisotropic structure needs fewer coefficients to achieve the same PSNR than the isotropic structure