

# Multiple Hypothesis Based Motion Compensation for Three Dimensional Wavelet Coding of Video

Arvind Sundarajan, Joy Rajiv and Michael Braun  
May 29, 2002

## Motivation

- 3D Wavelet coding provides scalability in video quality and temporal resolution
- Motion compensation (MC) prior to temporal wavelet decomposition is necessary to ensure good energy compaction
  - Without MC, there is significant energy in the high-pass band and ghosting artifacts in the low-pass band
- Averaging multiple hypothesized MC images may provide robustness to interframe noise and lead to better compression

# Temporal Haar Transform

Without MC:

$$h_k = \frac{1}{2}[x_{2k+1} - x_{2k}]$$

$$l_k = x_{2k} + h_k$$

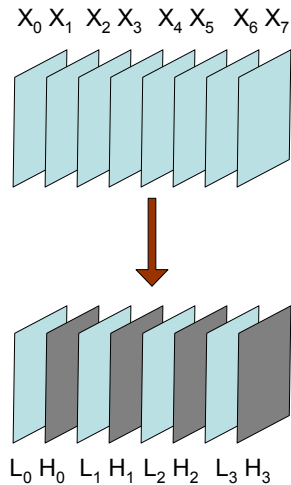
With MC:

$$h_k = \frac{1}{2}[x_{2k+1} - W_{12}(x_{2k})]$$

$$l_k = x_{2k} + W_{21}(h_k)$$

Lifting  $\rightarrow$  Warping function can be *anything* and we still achieve invertability.

We use simple block matching as warping function



## Warping Function for Motion Compensation



Image 1

## Warping Function for Motion Compensation



Image 4

## Warping Function for Motion Compensation



Image 1 warped to image 4

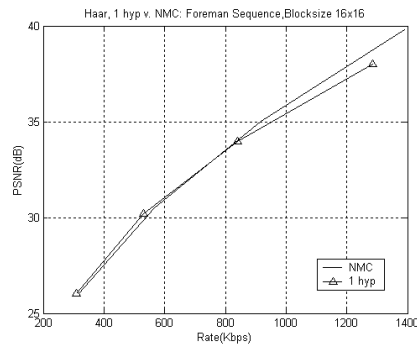
## Motion Vector Selection

- For the highpass temporal band:
  - $h_k = \frac{1}{2}[x_{2k+1} - W_{12}(x_{2k})]$
  - $W_{12} = \min_{W_{12}} \{ J = \lambda R_{hp|W_{12}} + D_{hp|W_{12}} \}$
  - $R_{hp|W_{12}} = R(\text{Motion vecs.} \mid W_{12}) + R(3D \text{ Wavelet coeffs} \mid W_{12})$
  - $D_{hp|W_{12}} = D(\text{reconstructed signal} \mid W_{12})$
- However, we do not know  $R_{hp}$  and  $D_{hp}$  until we spatially decompose and quantize.
  - We cannot spatially decompose until we have *all* motion vectors!
- Therefore our objective is to maximize the coding gain by minimizing  $E(h_k^2)$  for that block
  - $\rightarrow W_{12} = \min_{W_{12}} \{ J = \lambda R_{mv|W_{12}} + \text{MSE}[x_{2k+1}, W_{12}(x_{2k})] \}$

## Motion Vector Selection II

- Now for the lowpass temporal band:
  - $l_k = x_{2k} + W_{21}(h_k)$
  - $W_{21} = \min_{W_{21}} \{ J = \lambda R_{lp|W_{21}} + D_{lp|W_{21}} \}$
  - $R_{lp|W_{21}} = R(\text{Motion vecs.} \mid W_{21}) + R(3D \text{ Wavelet coeffs} \mid W_{21})$
  - $D_{lp|W_{21}} = D(\text{reconstructed signal} \mid W_{21})$
- Again, we will not know  $R_{lp}$  and  $D_{lp}$  until we spatially decompose and quantize.
- Therefore let's minimize  $\text{Var}(l_k \mid h_k)$  for that block
  - $\rightarrow W_{21} = \min_{W_{21}} \{ J = \lambda R_{mv|W_{21}} + \text{Var}[x_{2k} + W_{21}(h_k)] \}$
  - We hope this will act to minimize the overall variance of the lowpass temporal band.

# Initial Results



Performed *worse* than no motion compensation (NMC) for Foreman  
Perhaps there were ghosting artifacts in the Lowpass band...

## “Simplified” Temporal Wavelet

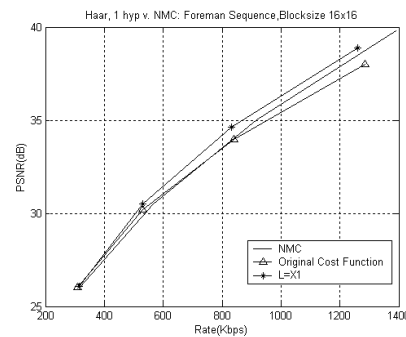
$$h_k = \frac{1}{2}[x_{2k+1} - W_{12}(x_{2k})]$$

$$l_k = x_{2k}$$

The lowpass  
essentially  
becomes an “I”  
frame

No more ghosting in

$$l_k \dots$$



## Best scheme for finding $W_{21}$

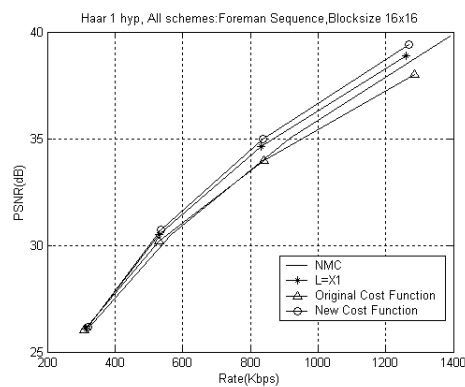
$$h_k = \frac{1}{2}[x_{2k+1} - W_{12}(x_{2k})]$$

$$l_k = x_{2k} + W_{21}(h_k)$$

Ignore  $h_k$  and simply find the “reverse trajectory”

$$W_{21} = \min_{W_{21}} \{ J = \lambda R_{mv|W_{21}} + \text{MSE}[x_{2k}, W_{21}(x_{2k+1})] \}$$

## Best scheme for finding $W_{21}$



→Use “reverse trajectory” strategy from now on, for best results.

For fastest runtime and decent results, use  $l_k = x_{2k}$

# Two Hypotheses

One hypothesis:

$$h_k = \frac{1}{2}[x_{2k+1} - W_{12}(x_{2k})], \quad l_k = x_{2k} + W_{21}(h_k)$$

Two hypotheses:

$$h_k = \frac{1}{2}[x_{2k+1} - \frac{1}{2}(W_{12,1}(x_{2k}) + W_{12,2}(x_{2k}))]$$

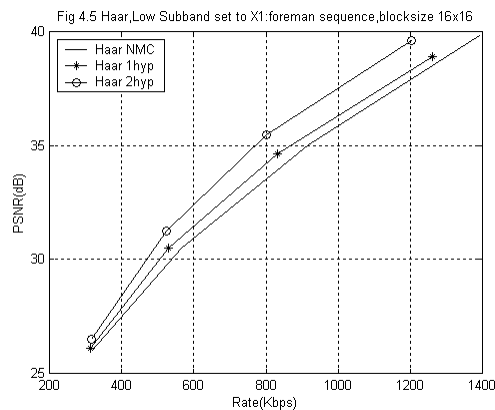
$$l_k = x_{2k} + \frac{1}{2}(W_{21,1}(h_k) + W_{21,2}(h_k))$$

Hold one hypothesis constant while minimizing cost function w.r.t. the other.

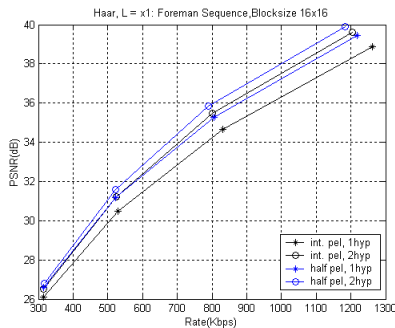
Repeat this until convergence upon two motion vectors.

Takes only 1.7 iterations, on average!

## Results for Two Hypotheses



# Half Pixel Accuracy



- Marginal gain due to ½ pel accuracy reduced when moving to 2 hypotheses
- Marginal gain due to moving to 2 hypotheses reduced when we use ½ pel accuracy
- 2 hypotheses, int pel still better than 1 hypothesis, half pel!

## Extension to 5/3 Temporal Wavelet

$$h_k = x_{2k+1} - \frac{1}{2}[x_{2k} + x_{2k+2}]$$

$$l_k = x_{2k} + \frac{1}{4}[h_{k-1} + h_k]$$

1 hypothesis:

$$h_k = x_{2k+1} - W_2(x_{2k}, x_{2k+2})$$

$$l_k = x_{2k} + \frac{1}{2}W_1(h_{k-1}, h_k)$$

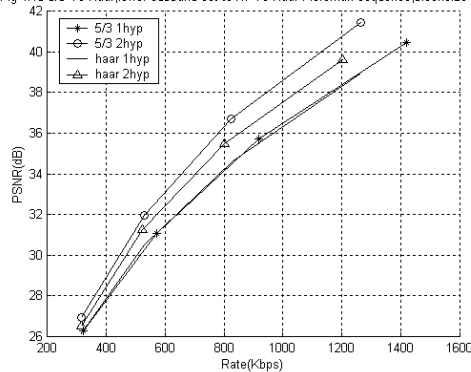
2 hypotheses:

$$h_k = x_{2k+1} - \frac{1}{2}[W_{1,1}(x_{2k}, x_{2k+2}) + W_{1,2}(x_{2k}, x_{2k+2})]$$

$$l_k = x_{2k} + \frac{1}{4}[W_{2,1}(h_{k-1}, h_k) + W_{2,2}(h_{k-1}, h_k)]$$

## Extension to 5/3 Temporal Wavelet

Fig 4.10 5/3 Vs Haar, lower subband set to X1 Vs Haar : foreman sequence, blocksize 16x16



- For 2 hypotheses, 5/3 much better than Haar

## Extension to 5/3 Temporal Wavelet

Alternative approach for 2 hypotheses:

$$h_k = x_{2k+1} - \frac{1}{2}[W_{2k,2k+1}(x_{2k}) + W_{2k+2,2k+1}(x_{2k+2})]$$

$$l_k = x_{2k} + \frac{1}{4}[W_{2k-1,2k}(h_{k-1}) + W_{2k+1,2k}(h_k)]$$

That is, always take 1 block from previous frame, 1 block from next frame

Advantage: Search over only 1 frame → reduce time by a factor of 2

Disadvantage: May not work when objects move into or out of scene

- Conclusion:
  - 3DWT provides scalability, MC provides better performance, lifting provides invertability
  - Two Hypotheses increases performance, requires ~1.7 additional searches for convergence
  - Reduced benefit at half-pel res., but 2 hyp. at integer res. still better than 1hyp. at half-pel
  - 5/3 better than Haar, since it is “bi-directional”
- Future Work:
  - More hypotheses
  - For 5/3, restrict origins of hypotheses
  - Try larger time kernels
  - Better multihypothesis search schemes
  - Other motion models (e.g. Mesh)