IEEE International Conference on Acoustics, Speech, and Signal Processing 2007

A New Bidirectionally Motion-Compensated Orthogonal Transform for Video Coding

Markus Flierl and Bernd Girod

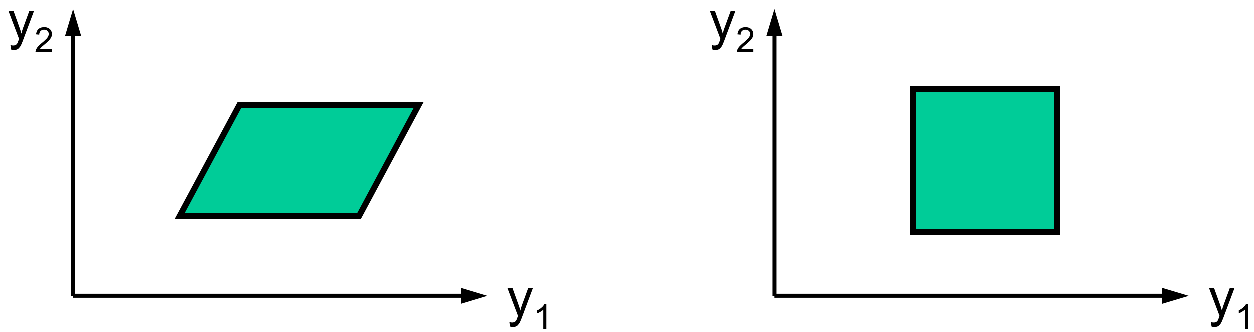
Max Planck Center for Visual Computing and Communication

Stanford University

Motivation

- Motion-compensated lifted Haar wavelet deviates substantially from orthonormality due to motion compensation

- Orthogonality offers good partition cell shapes



- **Goal:** Motion-adaptive transform that strictly maintains orthonormality while permitting flexible
  - unidirectional motion compensation **and**
  - bidirectional motion compensation

Outline

- Motion-Compensated Orthogonal Transform (MCOT)
- Bidirectionally MC incremental transform
- Special case: Unidirectionally MC incremental transform
- Bi-MCOT: Energy concentration constraint
- Dyadic transform for groups of pictures
- Experimental results

Orthogonal Video Transform

- Orthogonal transform for **triplets of input images**:

low band image →
high band image →
low band image →

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{pmatrix} = T \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{pmatrix}$$

- Factor *T* into a sequence of *k* **incremental transforms**:

$$T = T_k T_{k-1} \cdots T_\kappa \cdots T_2 T_1$$

- Each incremental transform is orthogonal: $T_\kappa T_\kappa^T = I$

- Incremental transforms generate a sequence of transformed image pairs:

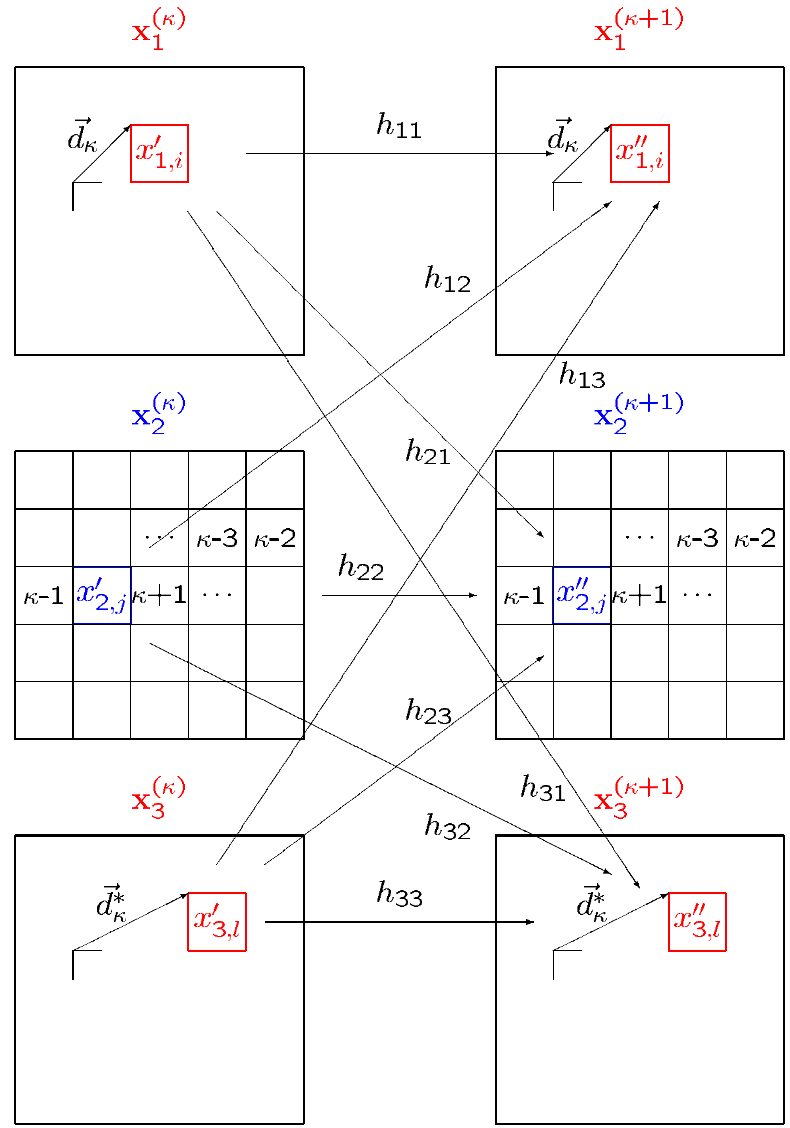
$$\begin{pmatrix} \mathbf{x}_1^{(\kappa+1)} \\ \mathbf{x}_2^{(\kappa+1)} \\ \mathbf{x}_3^{(\kappa+1)} \end{pmatrix} = T_\kappa \begin{pmatrix} \mathbf{x}_1^{(\kappa)} \\ \mathbf{x}_2^{(\kappa)} \\ \mathbf{x}_3^{(\kappa)} \end{pmatrix}$$

Bidirectionally MC Incremental Transform

low band image to-be

high band image to-be

low band image to-be



Bidirectionally MC Incremental Transform

$$T_\kappa = \begin{pmatrix}
\ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\
\cdots & 0 & h_{11} & 0 & \cdots & 0 & h_{12} & 0 & \cdots & 0 & h_{13} & 0 & \cdots \\
\cdots & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\
\cdots & 0 & h_{21} & 0 & \cdots & 0 & h_{22} & 0 & \cdots & 0 & h_{23} & 0 & \cdots \\
\cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
\cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots \\
\cdots & 0 & h_{31} & 0 & \cdots & 0 & h_{32} & 0 & \cdots & 0 & h_{33} & 0 & \cdots \\
\cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}$$

← *i*-th pixel in $\mathbf{x}_1$

← *j*-th pixel in $\mathbf{x}_2$

← *l*-th pixel in $\mathbf{x}_3$

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \quad \text{with} \quad HH^T = I$$

Unidirectionally MC Incremental Transform

$$
T_\kappa = \begin{pmatrix}
\ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\
\cdots & 0 & \color{red}{h_{11}} & 0 & \cdots & 0 & h_{12} & 0 & \cdots & 0 & 0 & 0 & \cdots \\
\cdots & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\
\cdots & 0 & h_{21} & 0 & \cdots & 0 & \color{blue}{h_{22}} & 0 & \cdots & 0 & 0 & 0 & \cdots \\
\cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
\cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots \\
\cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \color{red}{1} & 0 & \cdots \\
\cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
$$

← *i*-th pixel in $\mathbf{x}_1$

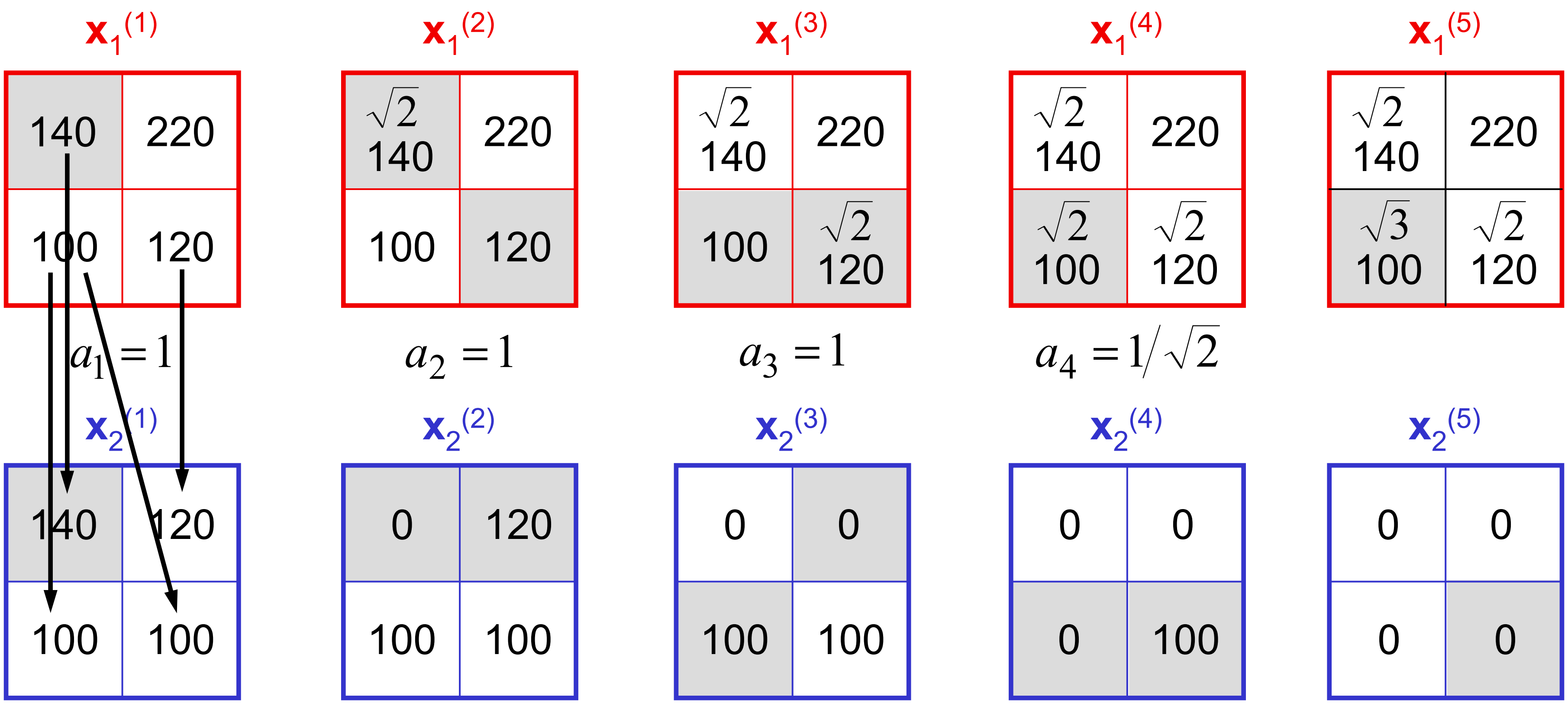
← *j*-th pixel in $\mathbf{x}_2$

← *l*-th pixel in $\mathbf{x}_3$

$$
H = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} = \frac{1}{\sqrt{1+a^2}} \begin{pmatrix} 1 & a \\ -a & 1 \end{pmatrix} \quad \text{as} \quad HH^T = I
$$

a: decorrelation factor

Example: Unidirectionally MCOT



$$\begin{bmatrix} x''_{1,i} \\ x''_{2,j} \end{bmatrix} = \frac{1}{\sqrt{1+a_\kappa^2}} \begin{bmatrix} 1 & a_\kappa \\ -a_\kappa & 1 \end{bmatrix} \begin{bmatrix} x'_{1,i} \\ x'_{2,j} \end{bmatrix}$$

Uni-MCOT: Energy Concentration Constraint

- Choose decorrelation factor for each incremental transform such that the energy in the high band to-be is removed

- Assume that pixel $x_{2,j}$ is connected to pixel $x_{1,i}$, i.e., $x_{2,j} = x_{1,I}$

- Resulting high band pixel to-be $x''_{2,j}$ shall be zero:

$$\begin{pmatrix} u_1 x_{1,i} \\ 0 \end{pmatrix} = H \begin{pmatrix} v_1 x_{1,i} \\ v_2 x_{1,i} \end{pmatrix}$$

- Energy conservation: $u_1^2 = v_1^2 + v_2^2$
- Decorrelation factor depends only on v_1 and v_2

Definition of Scale Counters

- Let n_1, n_2 be the scale counters for pixel $x_{1,i}$, $x_{2,j}$
- n_1, n_2 simply count how often the pixel $x_{1,i}$, $x_{2,j}$ are used as reference for motion compensation

- In the beginning, the scale counter is $n = 0$ and the scale factor is $v = 1$

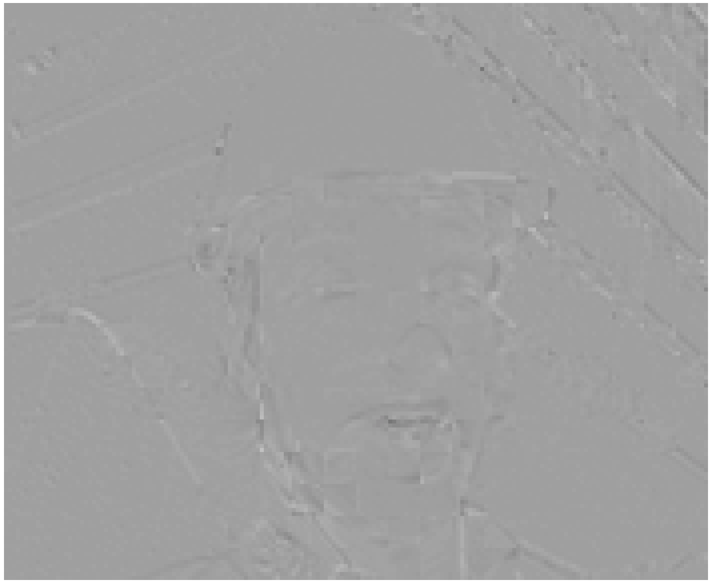
- For arbitrary scale counter m and n, the scale factors are

$$u = \sqrt{m+1} \quad \text{and} \quad v = \sqrt{n+1}$$

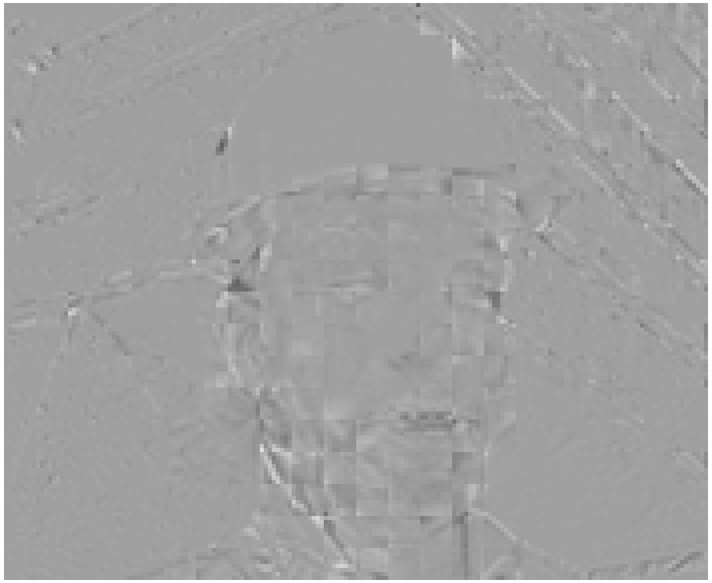
- Example: Scale counter update rule for Uni-MCOT:

$$m_1 = n_1 + n_2 + 1$$

Uni-MCOT: Experimental Results



temporal high band
first decomposition level



temporal high band
second decomposition level

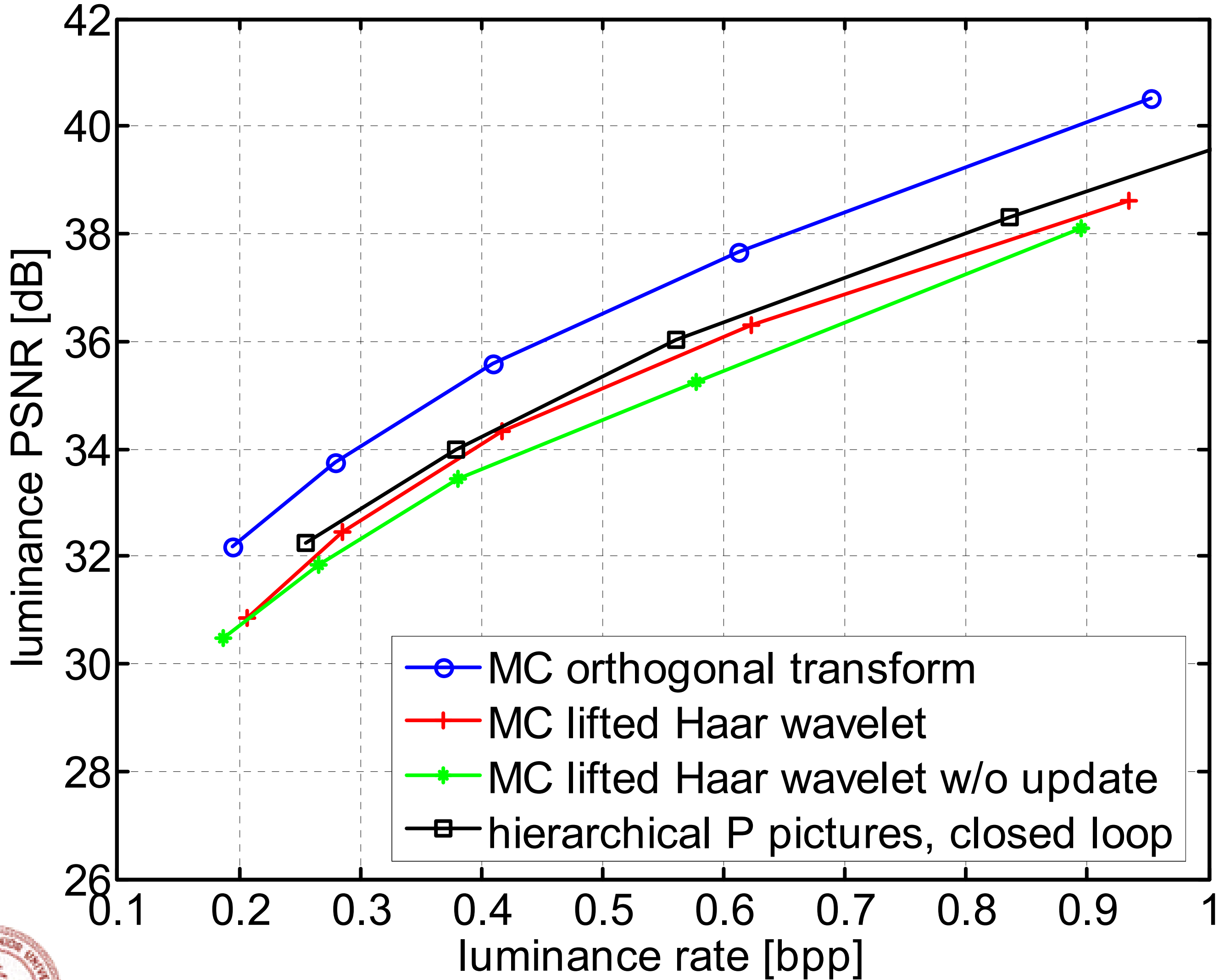
Uni-MCOT: Experimental Results

temporal low band
second decomposition level

rescaled temporal low band
second decomposition level

$$v = \sqrt{n+1}$$

Uni-MCOT: Experimental Results



Foreman

QCIF

30 fps

288 frames

GOP size K=16

8x8 block motion

Bi-MCOT: Euler's Rotation Theorem

- Any rotation in 3D can be given as a composition of rotations about three axes, i.e., $H = H_3 H_2 H_1$
- We choose the following composition:

$$
\begin{aligned}
H \;=\; & \begin{pmatrix} \cos(\psi) & 0 & \sin(\psi) \\ 0 & 1 & 0 \\ -\sin(\psi) & 0 & \cos(\psi) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix} \\
& \cdot \begin{pmatrix} \cos(\phi) & 0 & \sin(\phi) \\ 0 & 1 & 0 \\ -\sin(\phi) & 0 & \cos(\phi) \end{pmatrix}
\end{aligned}
$$

- Euler angles ψ, θ, ϕ are determined by the energy concentration constraint

Bi-MCOT: Energy Concentration Constraint

- Assume that pixel $x_{2,j}$ is connected to pixels $x_{1,i}$ and $x_{3,l}$, i.e., $x_{1,i} = x_{2,j} = x_{3,l}$
- Zero-energy constraint for the high band pixel:

$$\begin{pmatrix} u_1 x_{1,i} \\ 0 \\ u_3 x_{1,i} \end{pmatrix} = H_3 H_2 H_1 \begin{pmatrix} v_1 x_{1,i} \\ v_2 x_{1,i} \\ v_3 x_{1,i} \end{pmatrix}$$

- Energy conservation: $u_1^2 + u_3^2 = v_1^2 + v_2^2 + v_3^2$
- Euler angles depend on the scale factors u and v
  - Averaging with equal weight
  - Burden low-band pixels equally

Bi-MCOT: Scale Counter Update Rule

- After each incremental transform, scale counters have to be updated for modified pixels
- Scale counter update rule for Bi-MCOT:

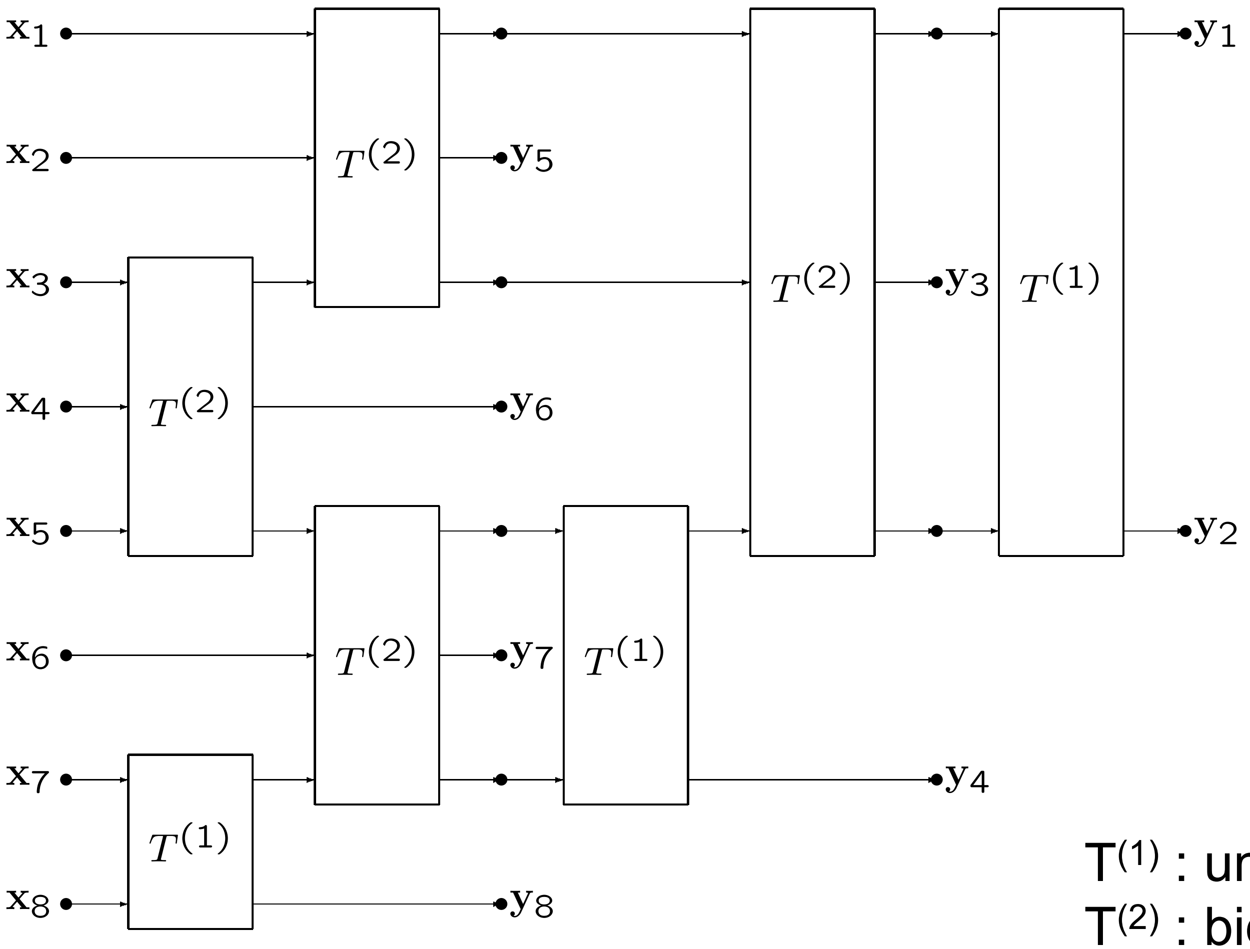
$$m_1 = n_1 + \frac{n_2 + 1}{2} \quad \text{and} \quad m_3 = n_3 + \frac{n_2 + 1}{2}$$

- Recall: scale counter update rule for Uni-MCOT:

$$m_1 = n_1 + n_2 + 1$$

- Example: first decomposition level, i.e., $n_2 = 0$
  - Uni-MCOT increases one scale counter by 1
  - Bi-MCOT increases two scale counters by 0.5 each

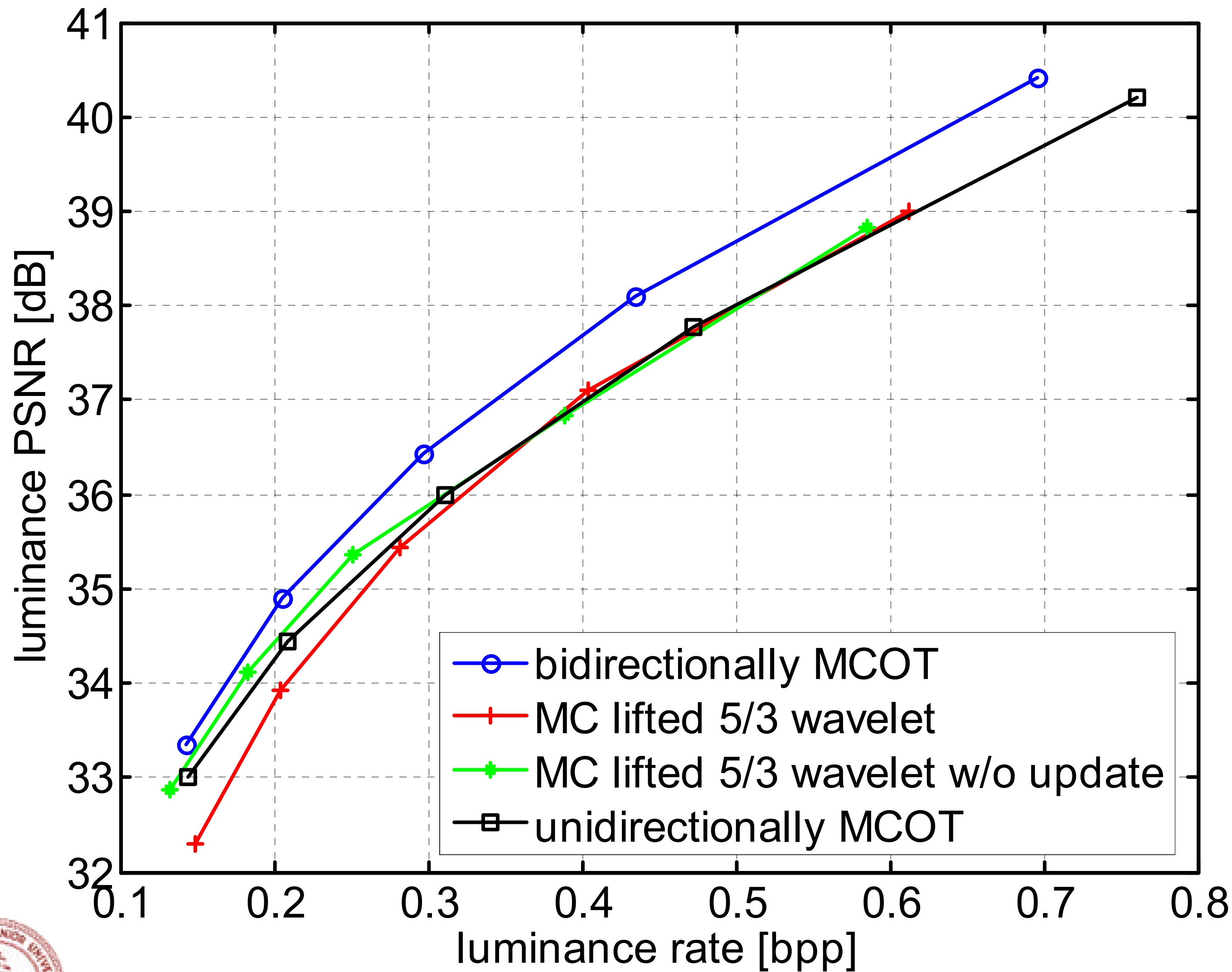
Dyadic Transform for Groups of Pictures



$T^{(1)}$: unidirectionally MCOT
$T^{(2)}$: bidirectionally MCOT

Example: GOP size K=8

Bi-MCOT: Experimental Results



Foreman

CIF

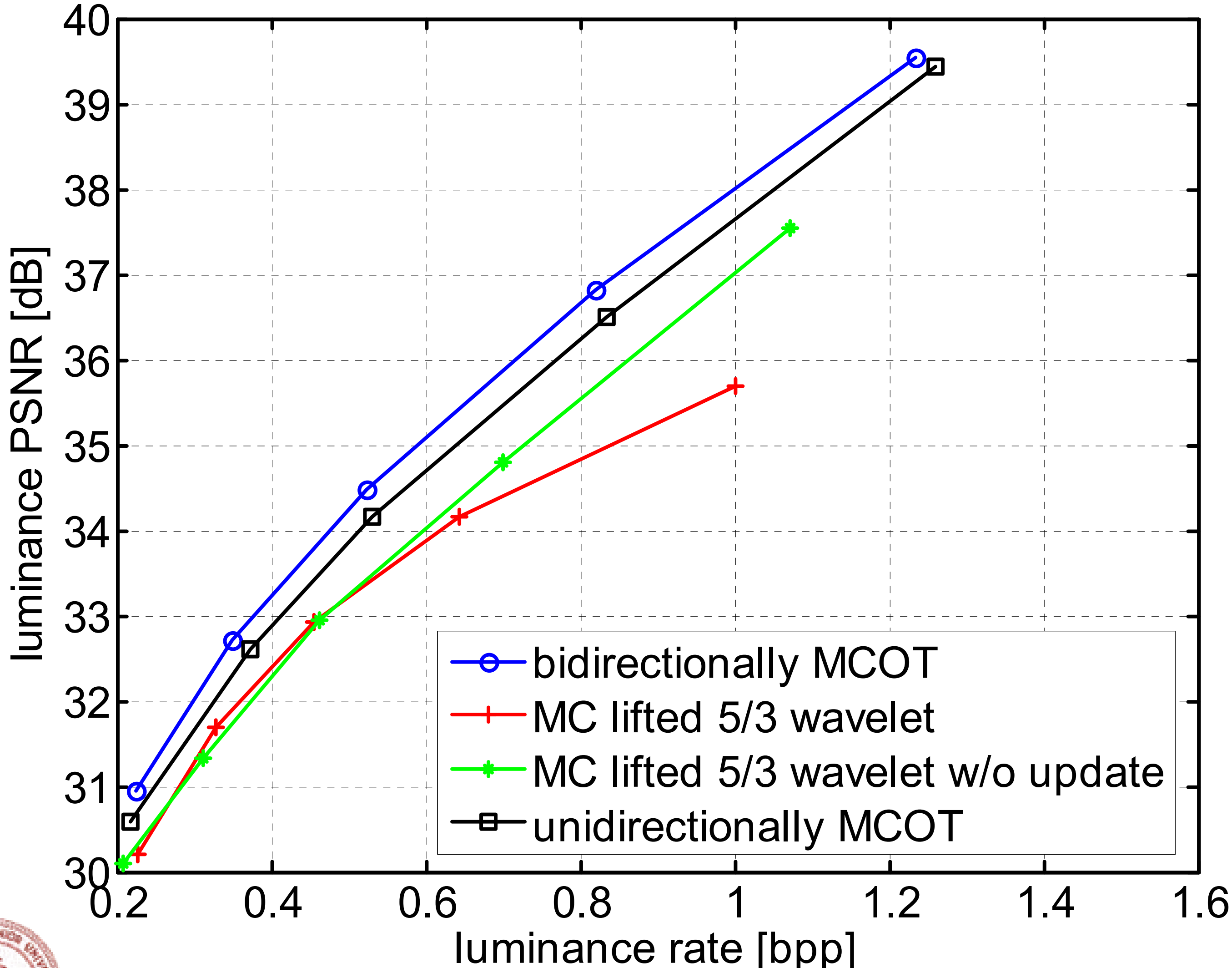
30 fps

288 frames

GOP size K=16

8x8 block motion

Bi-MCOT: Experimental Results



Soccer

CIF

30 fps

64 frames

GOP size K=16

8x8 block motion

Conclusions

- Motion-compensated orthogonal video transform which permits bidirectional motion compensation
- Highly flexible incremental transforms
- Systematic construction with Euler rotations
- Energy concentration constraint
- Results for integer-pel accurate bidirectionally MCOT
- Extensions to sub-pel accurate motion compensation are possible – see paper at DCC 2007

Further Reading

http://www.orthogonalvideo.org