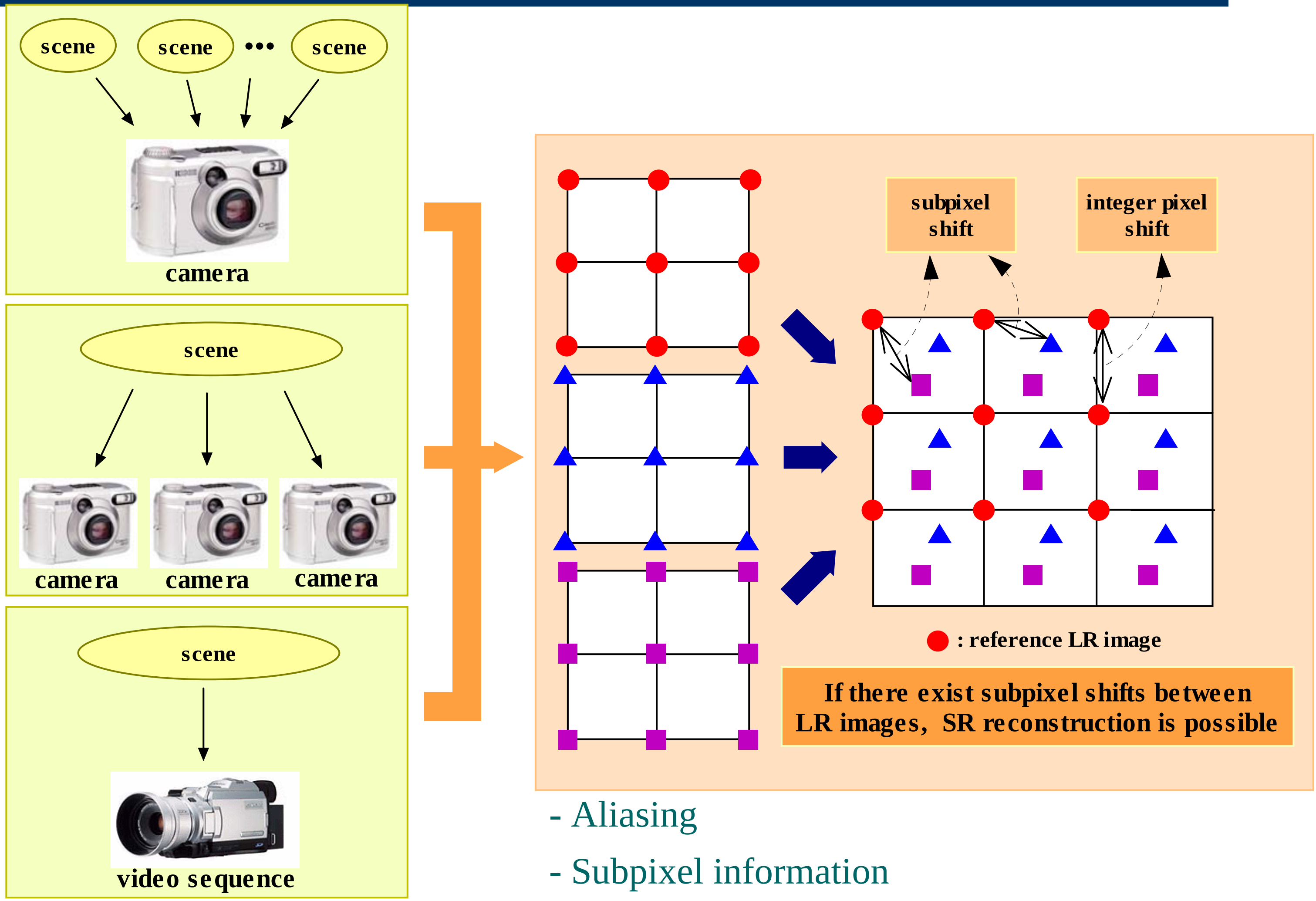
On Image and Video Super Resolution

Aggelos K. Katsaggelos
Professor
Northwestern University
Department of EECS
Evanston, IL 60208
http://www.ece.northwestern.edu/~aggk/
aggk@ece.northwestern.edu

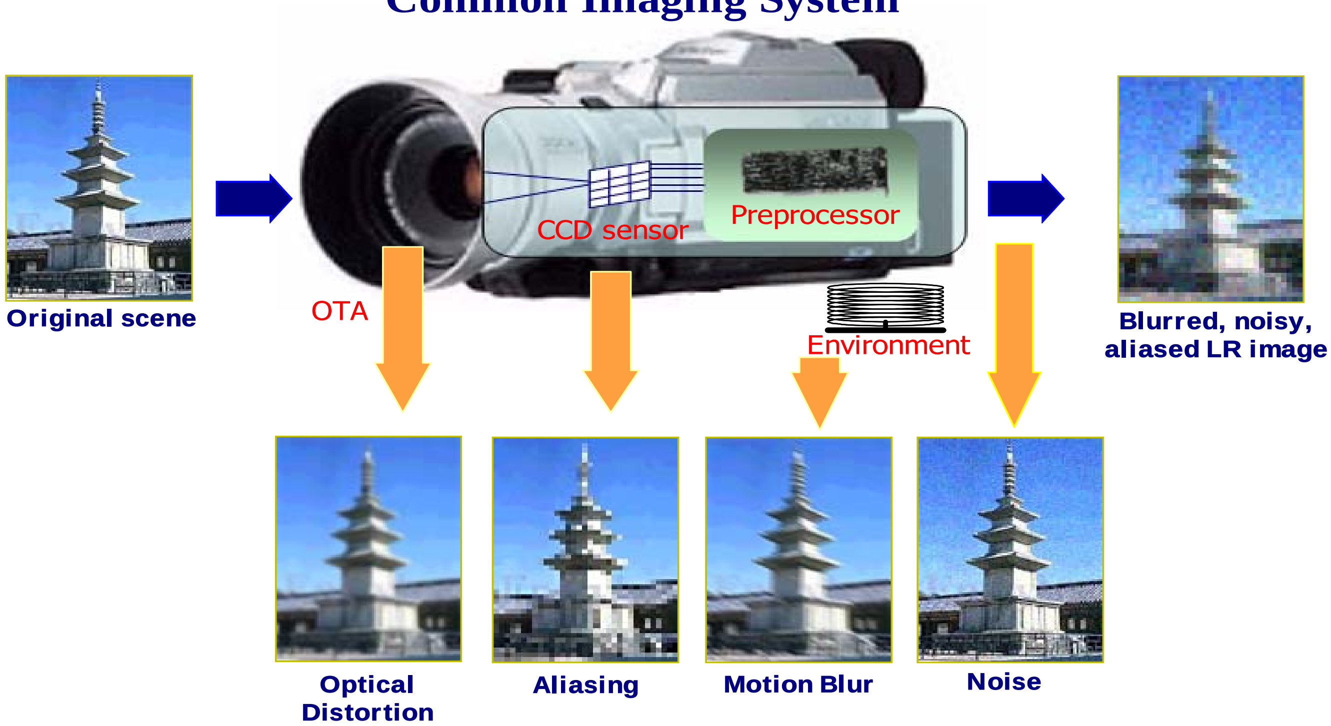
Basic Premise for Superresolution



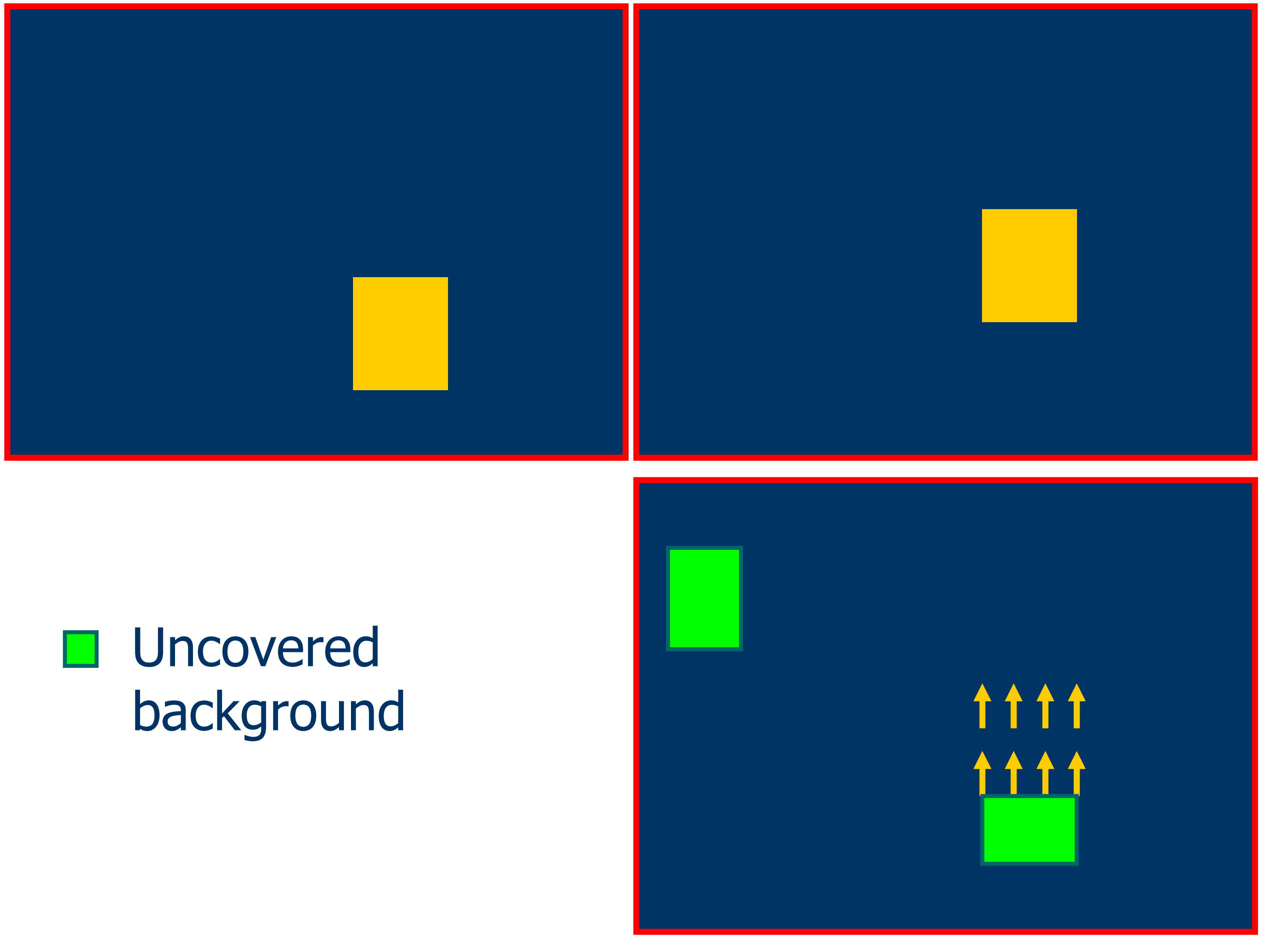
- Aliasing
- Subpixel information

Need for Resolution Enhancement

Common Imaging System



SR: How/Why does it work?



Super-Resolution: Why Does It Work?

- Super-Resolution
  - Recover Frequency Information that ***Exceeds*** the Frequency Range of the Channel (determined by bit rate)
  - How?
    - Aliasing embeds high-frequency information in the low-frequency coefficients
    - “Undoing” the aliasing recovers the high-frequencies
  - Multiple Observations are Assumed
    - Closely spaced cameras
    - Small perturbations in time
    - ***Inherent motion in the scene***

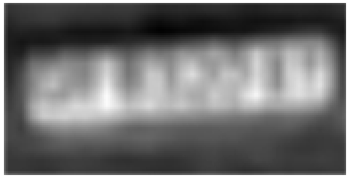
Super resolution from video

LR observations

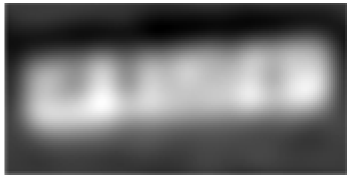
Interpolation

SR method

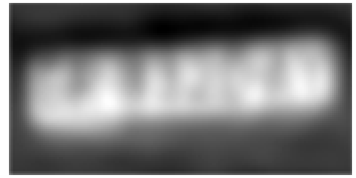
One More Example



bilinear



1 image



2 images

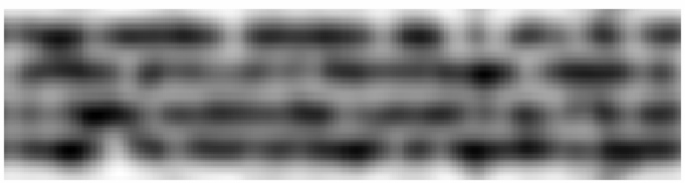
4 images

8 images

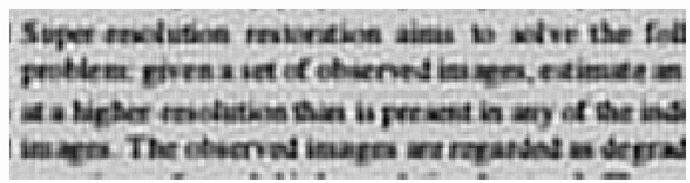
16 images

a

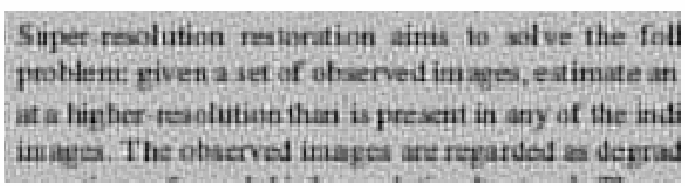
b



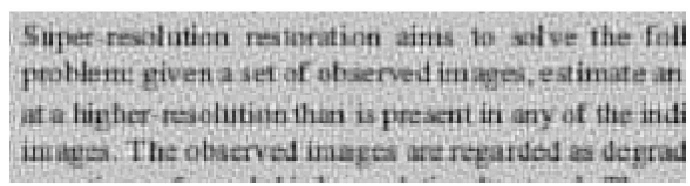
c



d



e



f

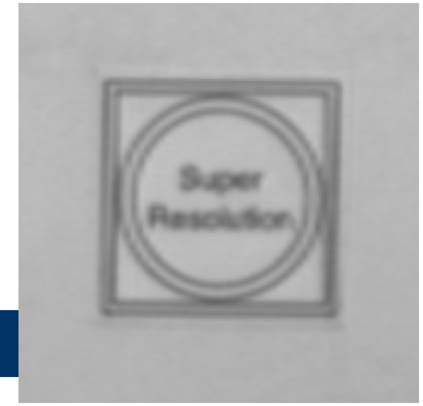
Cortijo, Villena, Molina, Katsaggelos, "Bayesian superresolution of text image sequences from low-resolution observations", Proc. IEEE 7th International Symposium on Signal Processing and Its Applications (ISSPA 2003), vol. I, pp. 421–424, 2003.

Super resolution from still low resolution images

(http://www.robots.ox.ac.uk/~improofs)

A sequence of 20 images of a laser-printed test pattern with random motion captured using a monochrome CCD video camera.

The best of the 20 images



Bicubic interpolation

When the images are registered into a common frame the pixel variations due to aliasing (under-sampling) are clearly visible.

Super resolution image

Experimental Results

Bi-cubic interpolation
PNR=23.50 dB

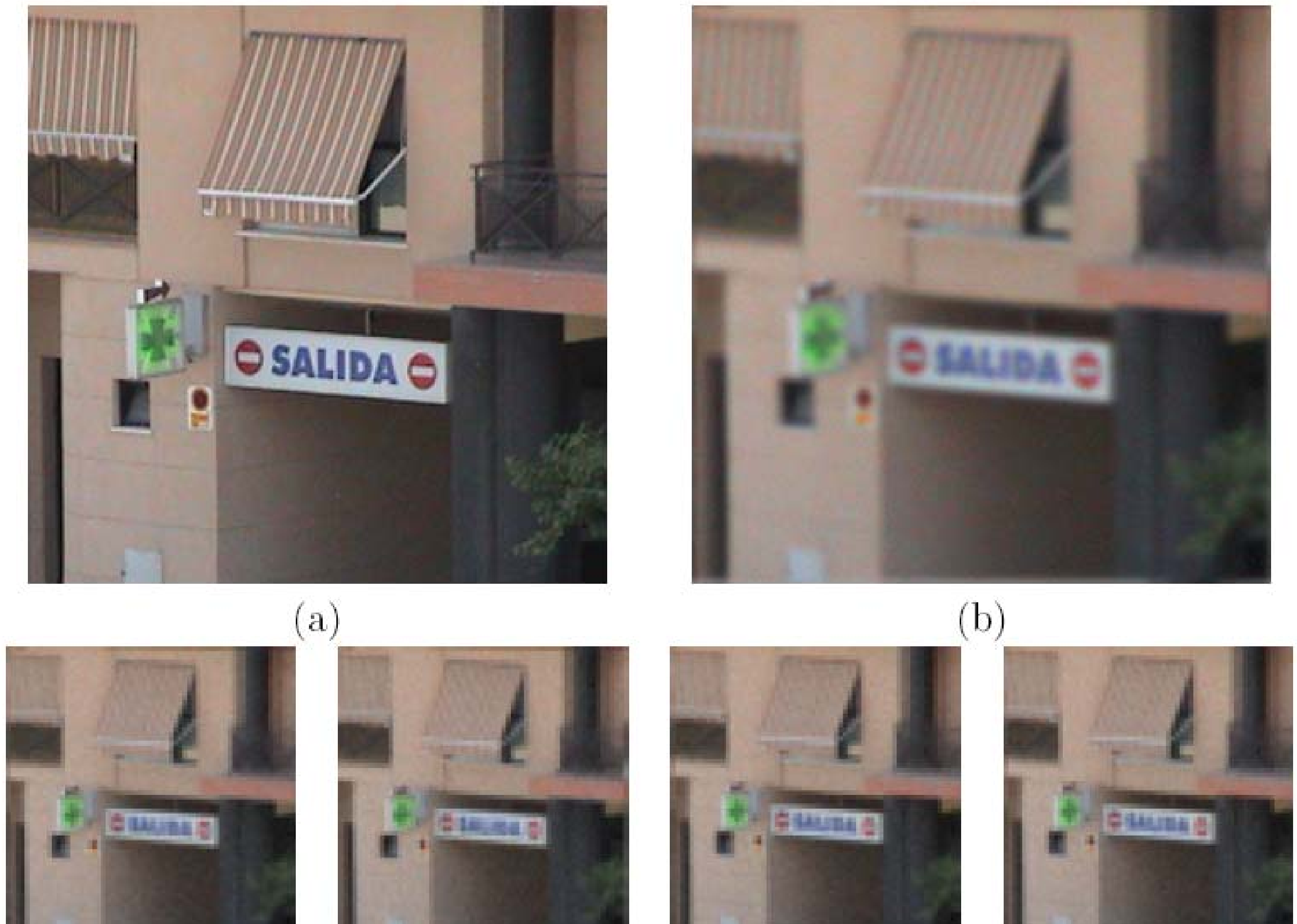
Proposed algorithm
PSNR=25.75 dB

Experimental Results

Bi-cubic interpolation
PSNR=30.05 dB

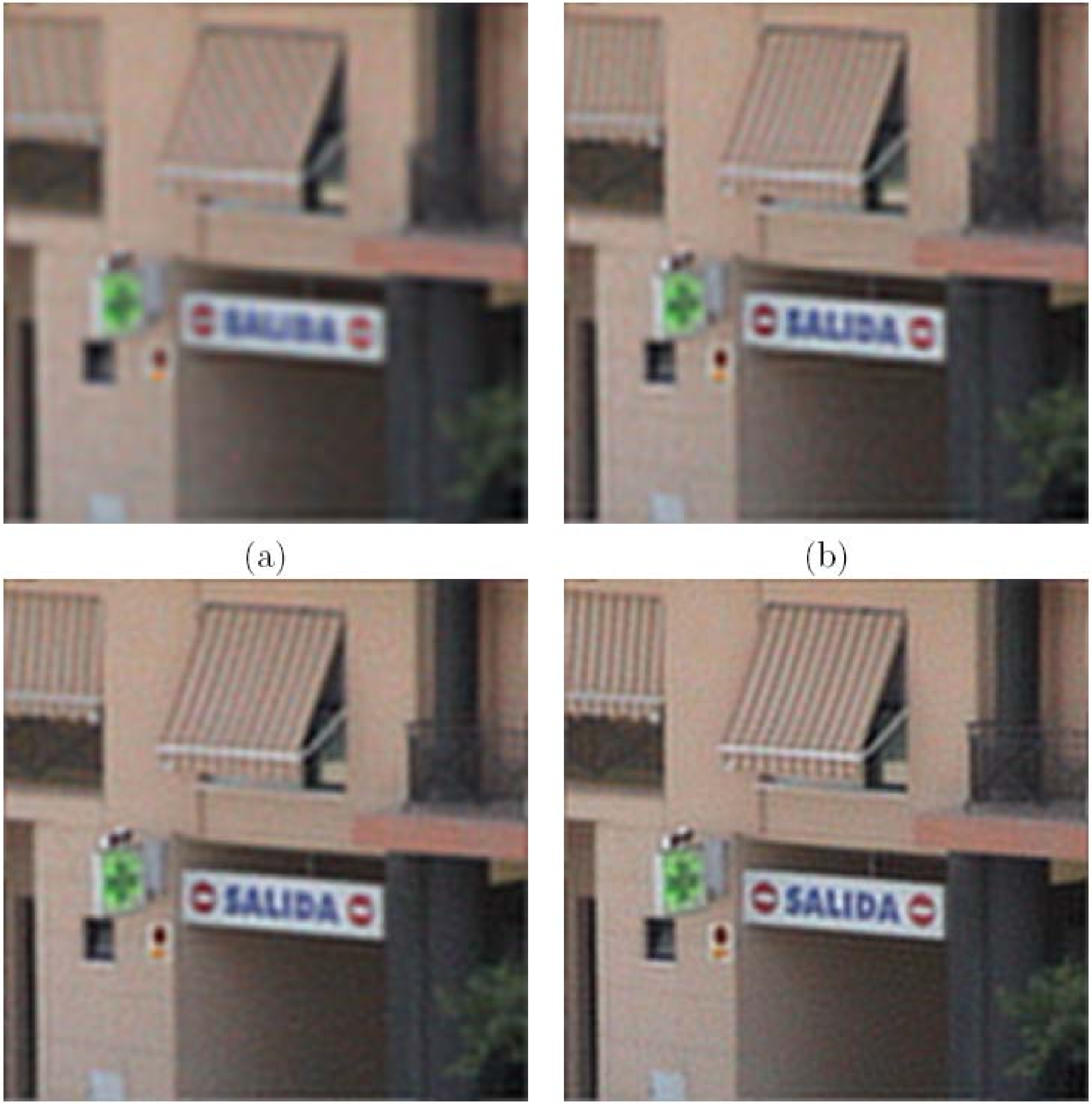
Proposed
PSNR=31.41dB

HR and blurred LR observations



(a) (b)

(c) (d) (e) (f)



(a) (b) (c) (d)

Using

(a) 1

(b) 2

(c) 4

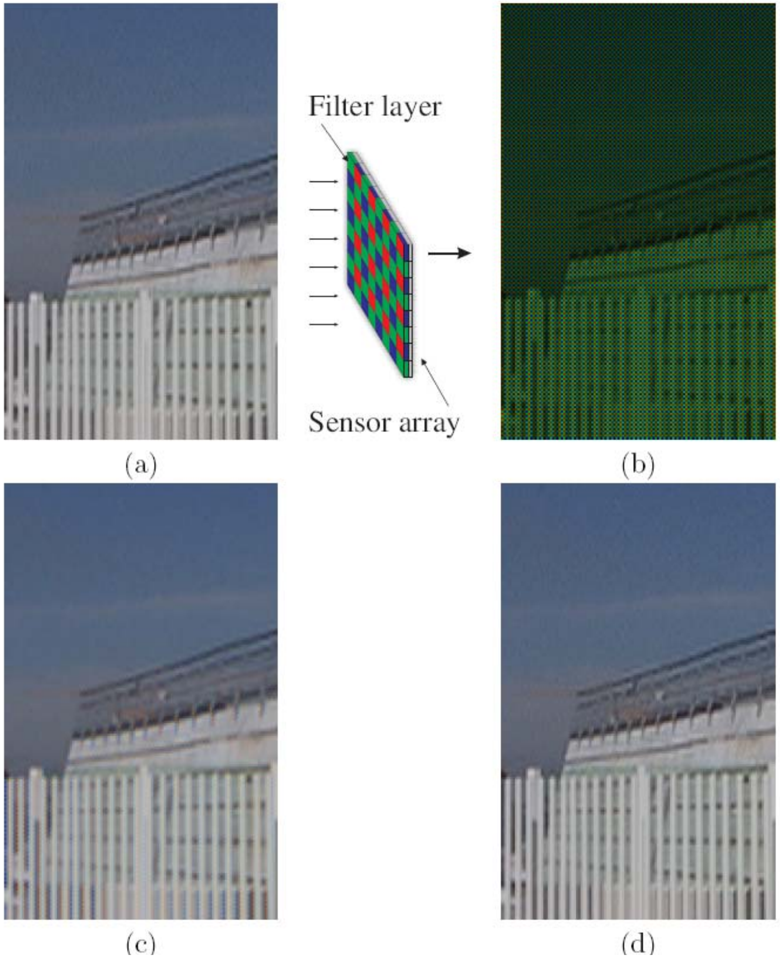
(d) 16

observations

Learning based SR

(a) (b) (c) (d)

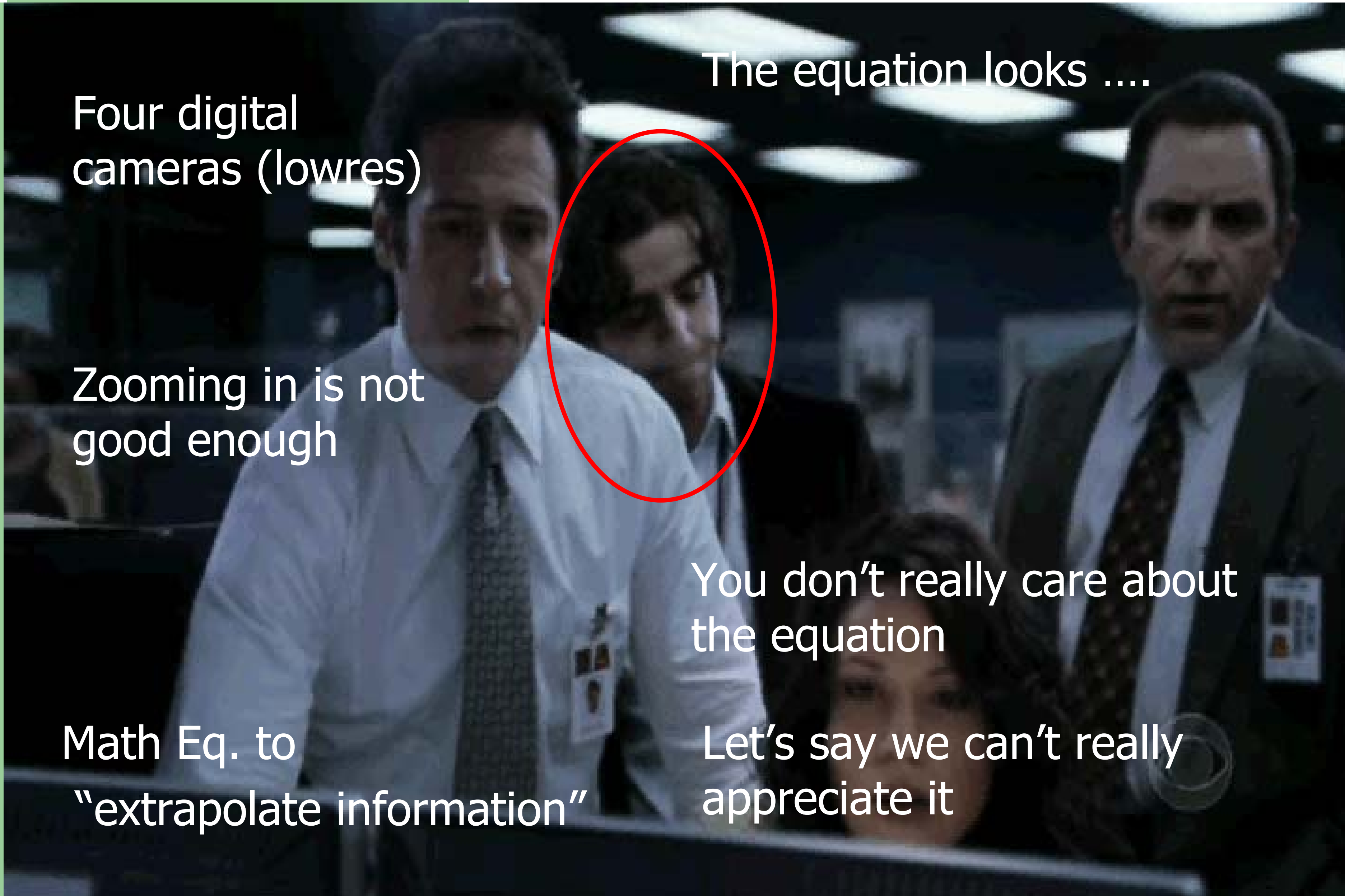
Demosaicking



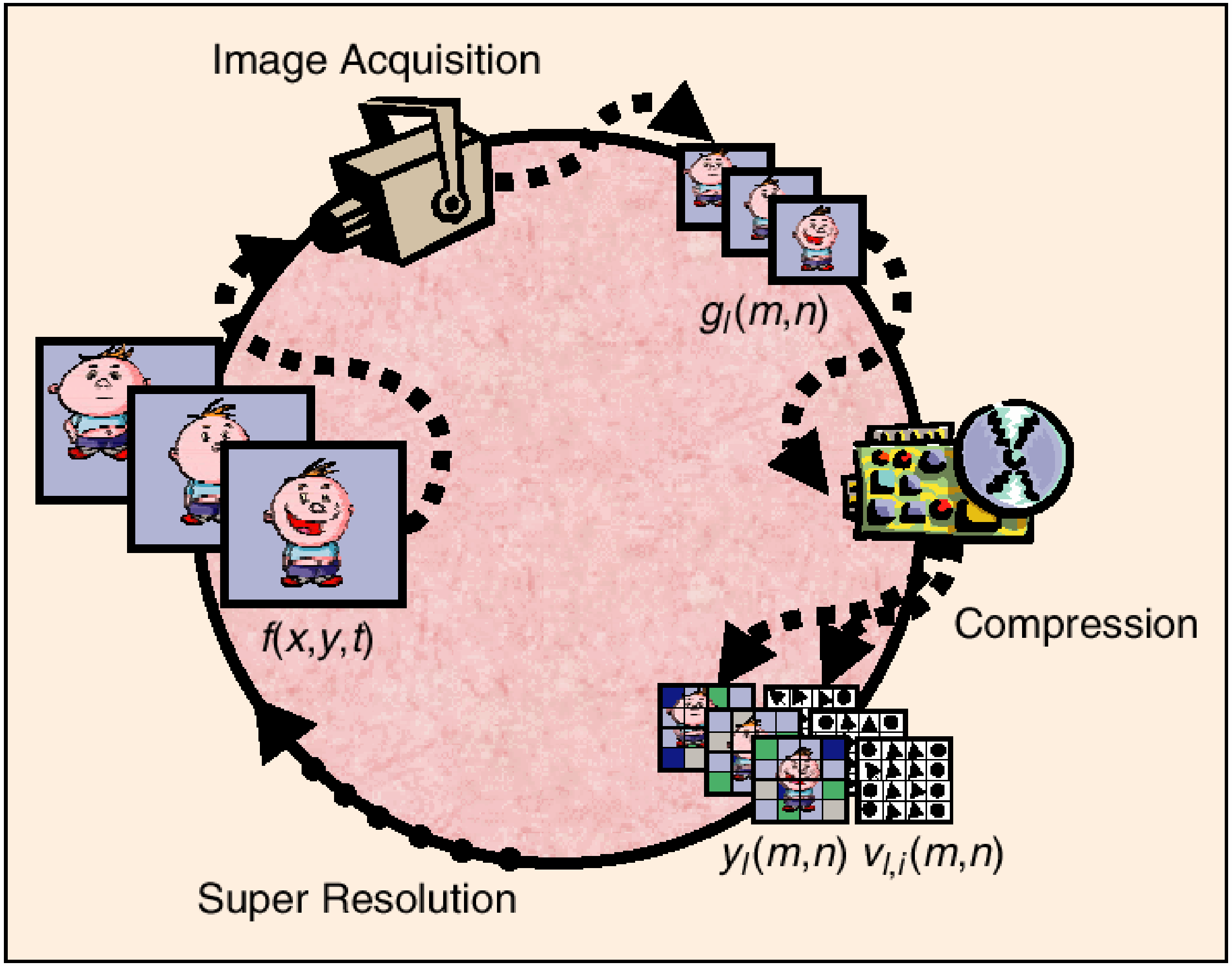
(a) (b) (c) (d)

Landsat ETM+

(a) (b) (c)


The equation looks ....
Four digital cameras (lowres)
Zooming in is not good enough
You don't really care about the equation
Math Eq. to "extrapolate information"
Let's say we can't really appreciate it

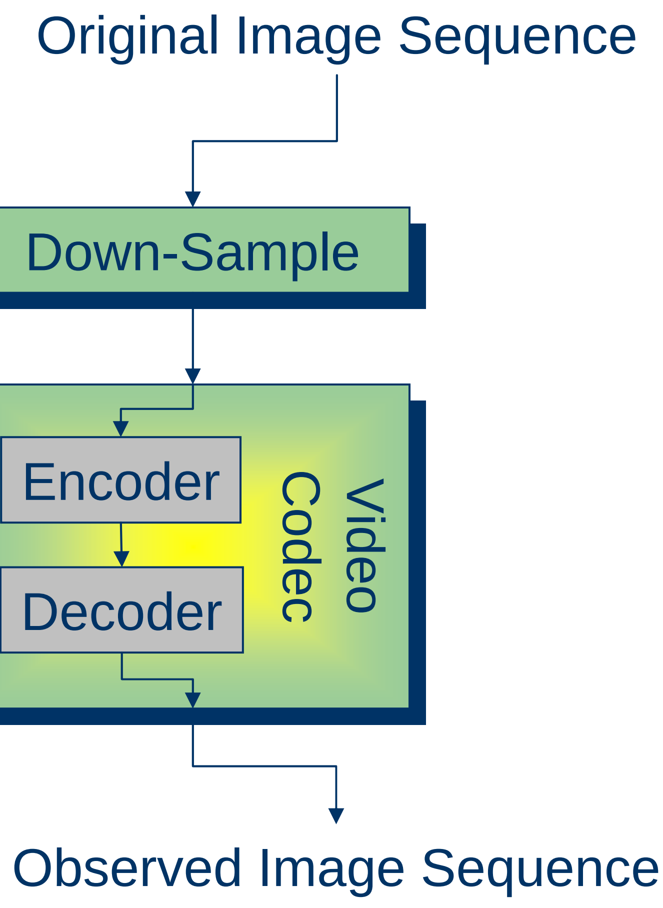
SR of Compressed Video System



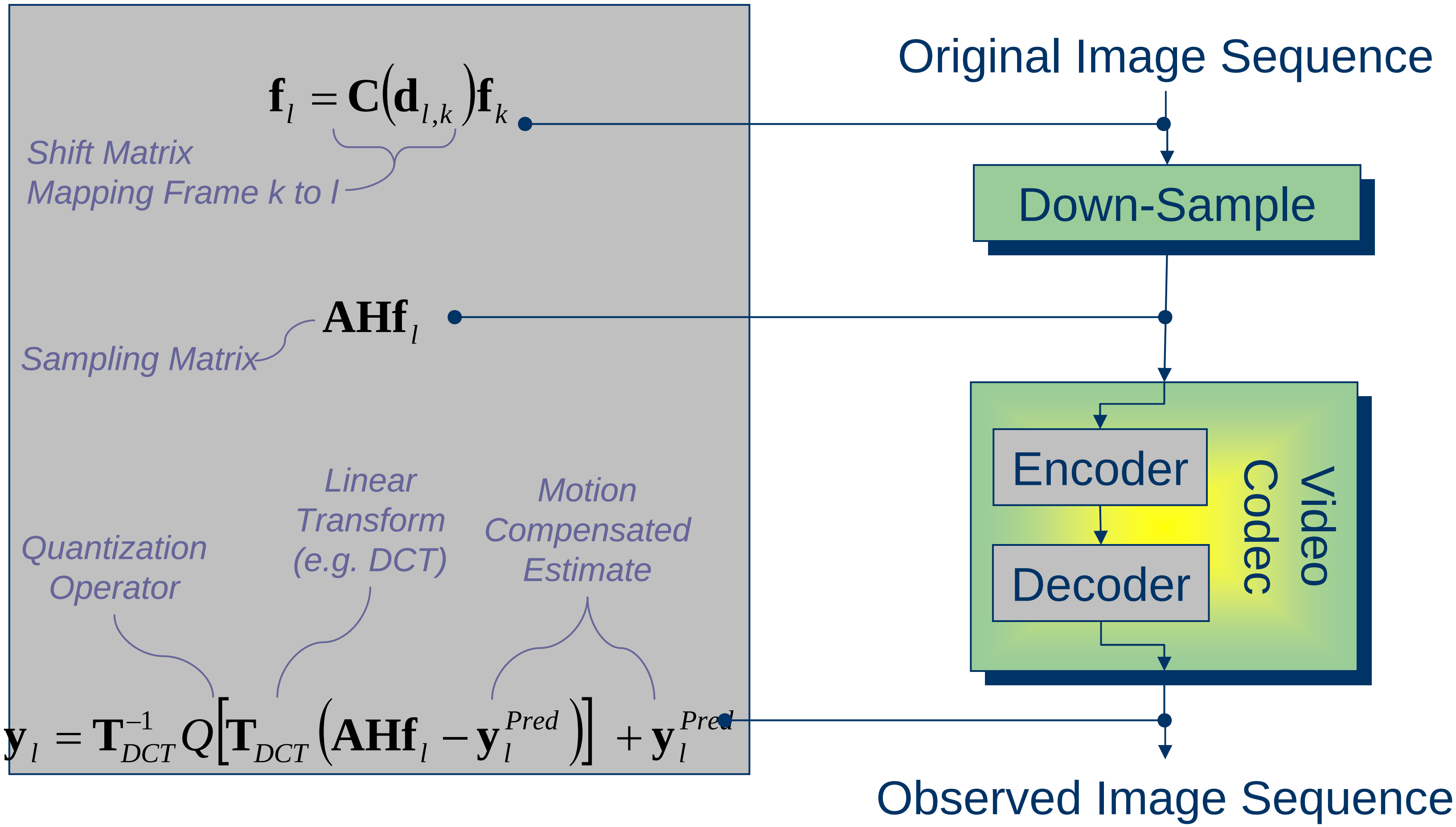
1. C.A. Segall, A.K. Katsaggelos, R. Molina, and J. Mateos, “Super-Resolution from Compressed Video,” in *Super-Resolution Imaging*, S. Chaudhuri, editor, Kluwer Academic Publishers: Boston, MA, p. 211-242, 2001.
2. C. A. Segall, R. Molina, and A. K. Katsaggelos, " High Resolution Images from a Low-Resolution Compressed Video," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp.37-48, May 2003.
3. C.A. Segall, A.K. Katsaggelos, R. Molina, and J. Mateos, “Bayesian Resolution Enhancement of Compressed Video,” *IEEE Trans. on Image Processing*, vol. 13, no. 7, pp. 898-911, July 2004.

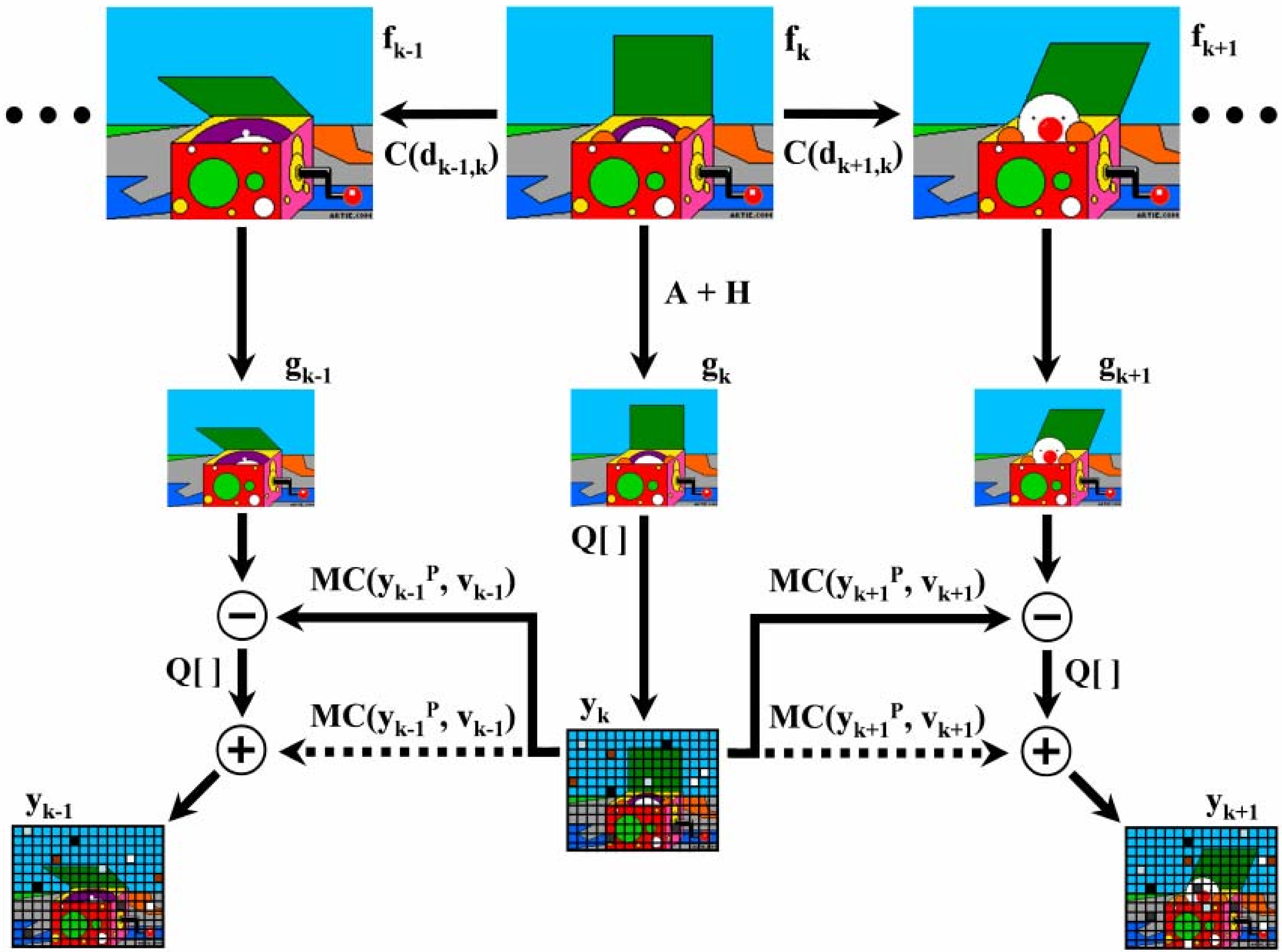
Problem Goal

- Recover High-Resolution Image from Sequence of Low-Resolution and Compressed Observations
- Exploit Information in the Compressed Bit-Stream
  - Transform Coefficients
  - Motion Vectors
  - Compression Modes
- Attenuate Compression Artifacts



System Model





f_{k-1}
f_k
f_{k+1}
C(d_{k-1,k})
C(d_{k+1,k})
A + H
g_{k-1}
g_k
g_{k+1}
Q[ ]
MC(y_{k-1}^P, v_{k-1})
MC(y_{k+1}^P, v_{k+1})
Q[ ]
Q[ ]
MC(y_{k-1}^P, v_{k-1})
y_k
MC(y_{k+1}^P, v_{k+1})
y_{k-1}
y_{k+1}

Input-Output Relations

$HR_l \Leftrightarrow HR_k$ $\quad f_l = C(d_{l,k}) f_k + n^r{}_{l,k}$ $\quad g_l = A \cdot H \cdot f_l + n_l$ $\quad LR_l \Leftrightarrow HR_l$

$$g_l = A \cdot H \cdot C\left(d_{l,k}\right) f_k + e_{l,k}$$

$LR_l \Leftrightarrow HR_k$

COMPRESSION

$CLR_l \Leftrightarrow LR_l$

$$y_l = T^{-1}\left[Q\left(T\left[g_l - MC_l\left(y_l^p, v_l\right)\right]\right)\right] + MC_l\left(y_l^p, v_l\right)$$

$CLR_l \Leftrightarrow HR_k$

$$y_l = T^{-1}\left[Q\left(T\left[A \cdot H \cdot C\left(d_{l,k}\right) f_k + e_{l,k} - MC_l\left(y_l^p, v_l\right)\right]\right)\right] + MC_l\left(y_l^p, v_l\right)$$

Solution Approach

- Formulation
  - Joint Estimate of Sub-Pixel Shifts and High Resolution Image
  - Given
    - Knowledge of Encoder Structure (Block-based)
    - Decoded Image
    - Motion Vectors
    - Mode (e.g., inter, intra, skip)
    - Quantization Intervals

High-Resolution Estimate

High-Resolution Image

Decoded Low-Resolution Observations

$$\hat{\mathbf{f}}_k, \hat{\mathbf{D}}_{TB,TF} = \arg\max_{\mathbf{f}_k,\mathbf{D}} p\left(\mathbf{f}_k, \mathbf{D}_{TB,TF} \mid \mathbf{Y}_{TB,TF}, \mathbf{V}_{TB,TF}\right)$$

Estimate of Sub-Pixel Shifts

Sub-Pixel Shifts

Transmitted Motion Vectors

Modeling the Observations

Bayes Rule

$$p(\mathbf{f}_k, \mathbf{D} \mid \mathbf{Y}, \mathbf{V}) = \frac{p(\mathbf{Y}, \mathbf{V} \mid \mathbf{f}_k, \mathbf{D})\, p(\mathbf{f}_k, \mathbf{D})}{p(\mathbf{Y}, \mathbf{V})}$$

Independence between decoded intensities and motion vectors

$$p(\mathbf{Y}, \mathbf{V} \mid \mathbf{f}_k, \mathbf{D}) = p(\mathbf{Y} \mid \mathbf{f}_k, \mathbf{D})\, p(\mathbf{V} \mid \mathbf{Y}, \mathbf{f}_k, \mathbf{D})$$

$$= \prod_l p(\mathbf{y}_l \mid \mathbf{f}_k, \mathbf{D})\, p(\mathbf{v}_l \mid \mathbf{Y}, \mathbf{f}_k, \mathbf{D})$$

Quantization noise

Provided by bitstream

Independence between original intensities and motion vectors

$$p(\mathbf{f}_k, \mathbf{D}) = p(\mathbf{f}_k)\, p(\mathbf{D})$$

Quantization Noise Models

Normal

$$p(\mathrm{y}_\mathrm{l} \mid \mathrm{f}_\mathrm{k}, \mathrm{d}_{\mathrm{l,k}}) \propto \exp\left[\frac{1}{2\sigma_l^2}\left\| \mathrm{y}_\mathrm{l} - \mathrm{AHC}(\mathrm{d}_{\mathrm{l,k}})\mathrm{f}_\mathrm{k} \right\|^2\right]$$

Uniform

$$p_{QC}(\mathrm{y}_\mathrm{l} \mid \mathrm{f}_\mathrm{k}, \mathrm{d}_{\mathrm{l,k}}) = \begin{cases} \mathrm{const} & \text{if } \left| \mathrm{T}\left[(\mathrm{AHC}(\mathrm{d}_{\mathrm{l,k}})\mathrm{f}_\mathrm{k} - \mathrm{MC}_\mathrm{l}(\mathrm{y}_\mathrm{l}^\mathrm{p}, \mathrm{v}_\mathrm{l})\right](i) \right| \le \dfrac{\mathrm{q(i)}}{2} \quad \forall \mathrm{i} \\ 0 & \text{otherwise} \end{cases}$$

Quantization Noise Models

Fidelity to Decoded Image

Decoded Image

$$p\left(\mathbf{Y}_{TB,TF} \mid \mathbf{f}_k, \mathbf{D}_{TB,TF}\right) \propto \prod_l \exp\left\{-\frac{1}{2}\left(\mathbf{y}_l - \mathbf{AHC}\left(\mathbf{d}_{l,k}\right)\mathbf{f}_k\right)^T \mathbf{K}_{Q,l}^{-1}\left(\mathbf{y}_l - \mathbf{AHC}\left(\mathbf{d}_{l,k}\right)\mathbf{f}_k\right)\right\}$$

Covariance Matrix for Quantization Noise in Frame l

Sub-sampling Matrix

Shift Between Frame k and l (Represented as a Matrix)

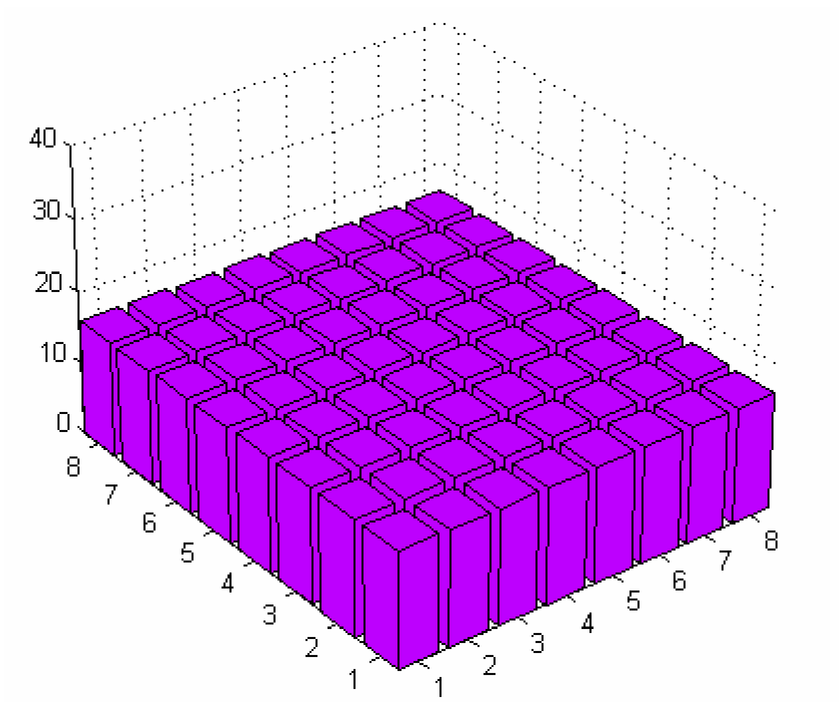
- Fidelity Constraint
  - Compression Assumed Primary Noise Process
  - Quantization Noise
    - Introduced in DCT Domain
    - Independent/Scalar Procedure
  - Spatial Representation

$$\mathbf{n}_{Spatial} = \mathbf{T}_{DCT}^{T}\mathbf{n}_{DCT}$$

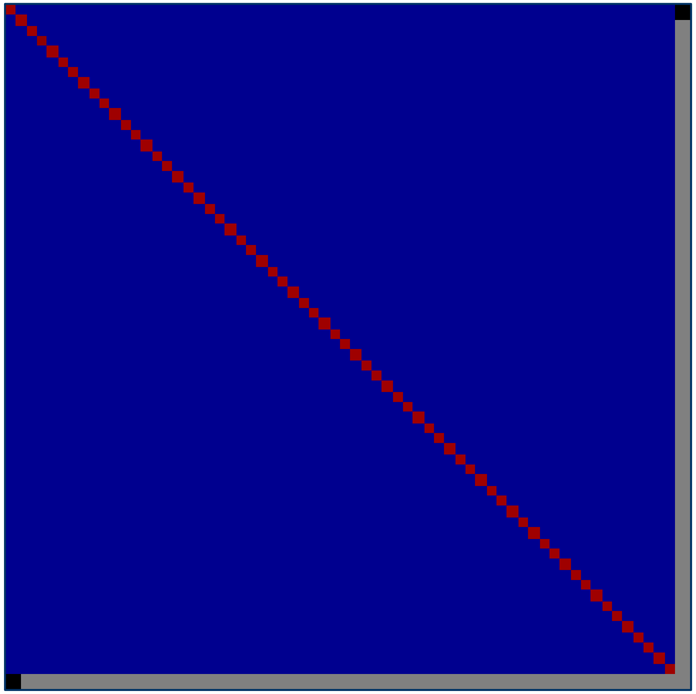
    - Inverse DCT: Linear
    - Noise: Independent
    - Tends to Gaussian distribution in spatial domain

Example Covariance

$$\mathbf{K}_Q = \mathbf{T}_{DCT}\mathbf{K}_{DCT}\mathbf{T}_{DCT}^T$$



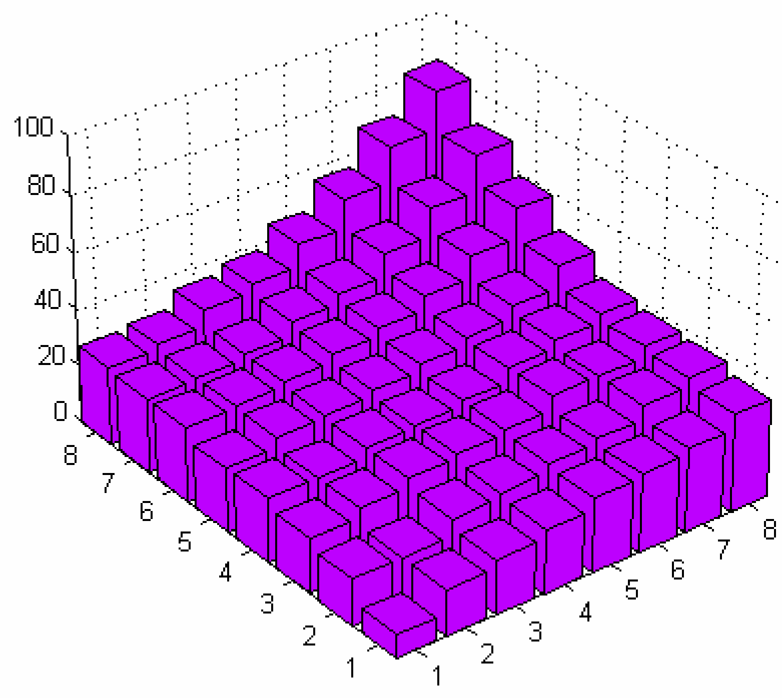
Quantizer in DCT Domain



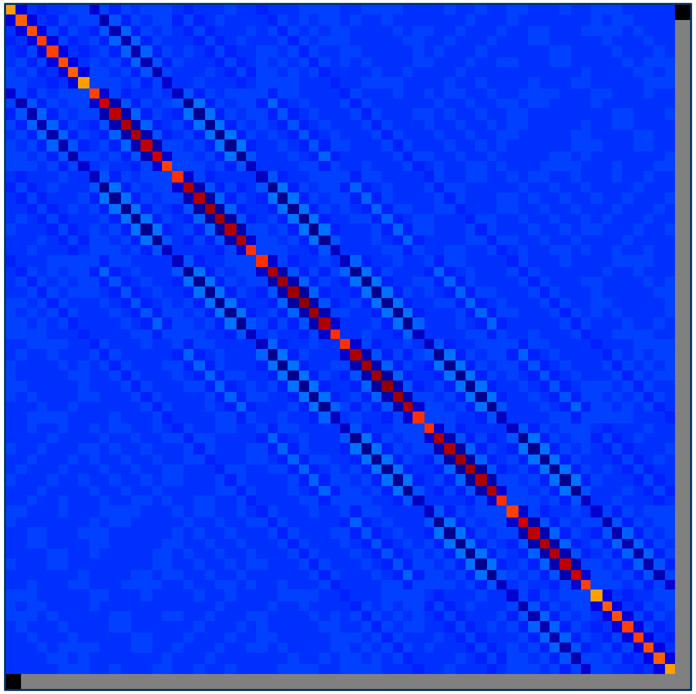
Covariance Matrix in Spatial Domain

Example Covariance

$$\mathbf{K}_Q = \mathbf{T}_{DCT}\mathbf{K}_{DCT}\mathbf{T}_{DCT}^T$$



Quantizer in DCT Domain



Covariance Matrix in Spatial Domain

Compression Motion Vector Models

LR motion vectors (used for compression) "similar" to HR motion!

$$p\,(v_{l,k} \mid f_k, d_{l,k}, y_l) = \begin{cases} \text{const} & \text{if} \quad |\, v_{l,k}(j) - [A_D d_{l,k}](j)\,| \leq \Delta \\ 0 & \text{otherwise} \end{cases}$$

$$p\,(v_{l,k} \mid f_k, d_{l,k}, y_l) \propto \exp\left[- \frac{\gamma_l^2}{2} \left\| A_u v_{l,k} - d_{l,k} \right\|^2 \right]$$

Compression Motion Vector Models

Fidelity to Transmitted Motion Vectors

$$p\left(\mathbf{V}_{TB,TF} \mid \mathbf{Y}_{TB,TF}, \mathbf{f}_k, \mathbf{D}_{TB,TF}\right) \propto \prod_l \exp\left\{-\frac{1}{2}\left(\mathbf{y}_l^{\mathrm{Pred}} - \mathbf{AH}C\left(\mathbf{d}_{l,k}\right)\mathbf{f}_k\right)^T \mathbf{K}_{MV,l}^{-1}\left(\mathbf{y}_l^{\mathrm{Pred}} - \mathbf{AH}C\left(\mathbf{d}_{l,k}\right)\mathbf{f}_k\right)\right\}$$

Covariance Matrix for Displaced Frame Difference+ in Frame l

Sub-sampling Matrix

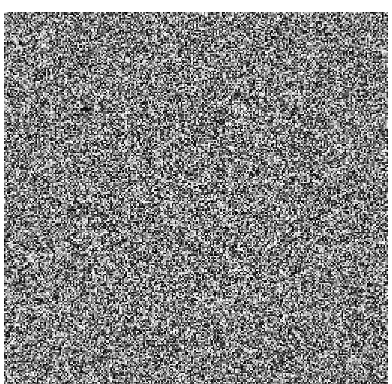
Shift Between Frame k and l (Represented as a Matrix)

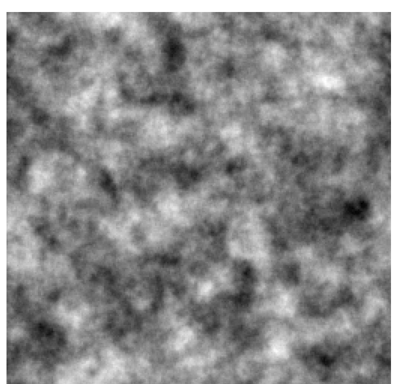
- Fidelity Constraint
  - Motion Vectors
    - Noisy observation of actual displacements
    - Compression Low $\Rightarrow$ Motion Vectors and Actual Shifts are Similar
    - Significant Image Features $\Rightarrow$ Motion Vectors and Actual Shifts are Similar

TF and TB: Previous and future frames in the reconstruction

Intensity Prior Models

- Images Generally Smooth
- Model Encoder Structure
  - Ignores Correlations Across Block Boundaries
  - Leads to Well Known “Blocking Artifact”
  - Blocking Errors Rarely Present in Original Data





A Priori Intensity Information

High-Resolution Image

$$p(\mathbf{f}_k) \propto \exp\left\{-\left(\lambda_1\left\|\mathbf{Q}_1\,\mathbf{f}_k\right\|^2 + \lambda_2\left\|\mathbf{Q}_2\,\mathbf{A}\mathbf{H}\,\mathbf{f}_k\right\|^2\right)\right\}$$

High Pass Operator (to Penalize High Frequency Information)

Difference Operator Across Block Boundaries (to Penalize Uncorrelated Boundary Pixels)

Motion Prior Models

- *A Priori* Information
  - Model of Displacements in Original Image Frame
    - Displacements are generally correlated
  - Other Examples
    - Displacements are usually continuous across block boundaries
    - Displacements are usually correlated between Frames

$$p(\mathrm{d}_{\mathrm{l,k}}) = \mathrm{const}$$

A Priori Motion Information

$$p\left(\mathbf{D}_{TB,TF}\right) \propto \exp\left\{\sum_{l=k-TB}^{k+TF} -\left(\lambda_3 \left\|\mathbf{Q}_3 \mathbf{d}_{l,k}\right\|^2\right)\right\}$$

High Pass Operator (to Penalize High Frequency Information)

Optimization Problem

Taking logarithms, the maximization becomes the minimization

$$\hat{\mathbf{f}}_k,\quad \hat{\mathbf{D}} =$$

$$\arg\min_{\mathbf{f}_k,\mathbf{D}} \left\{ \sum_{l=k-TB}^{k+TF} \left(\mathbf{y}_l - \mathbf{AH}C\left(\mathbf{d}_{l,k}\right)\mathbf{f}_k\right)^T \mathbf{K}_{Q,l}^{-1}\left(\mathbf{y}_l - \mathbf{AH}C\left(\mathbf{d}_{l,k}\right)\mathbf{f}_k\right) \right.$$

$$+ \sum_{l=k-TB}^{k+TF} \left(\mathbf{y}_l^{\text{Pred}} - \mathbf{AH}C\left(\mathbf{d}_{l,k}\right)\mathbf{f}_k\right)^T \mathbf{K}_{MV,l}^{-1}\left(\mathbf{y}_l^{\text{Pred}} - \mathbf{AH}C\left(\mathbf{d}_{l,k}\right)\mathbf{f}_k\right)$$

$$\left. + \lambda_1\left\|\mathbf{Q}_1\mathbf{f}_k\right\|^2 + \lambda_2\left\|\mathbf{Q}_2\mathbf{AHf}_k\right\|^2 + \lambda_3\left\|\mathbf{Q}_3\mathbf{d}_{l,k}\right\|^2 \right\}$$

TF and TB: Previous and future frames in the reconstruction

One Solution Approach

Cyclic-Coordinate Descent

Given $\hat{f}_k^q$,

$$\hat{D}^q = \arg\max_D p(\hat{f}_k^q, D) p(Y, V \mid \hat{f}_k^q, D)$$

Given D^q,

$$\hat{f}_k^{q+1} = \arg\max_{f_k} p(f_k, \hat{D}^q) p(Y, V \mid f_k, \hat{D}^q)$$

1. Estimate Sub-Pixel Shifts
   1. Assume Hi-Res Image Known
   2. Gradient Descent Algorithm
2. Estimate Hi-Res Image
   1. Utilize Computed Sub-Pixel Shifts
   2. Gradient Descent Algorithm
3. Refine Sub-Pixel Estimate
   1. Use Recent Hi-Res Estimate
4. Goto 2 until Termination

Solution Approach

- Sub-pixel shifts are found by

$$\hat{\mathbf{d}}_{l,k}^{m+1} = \hat{\mathbf{d}}_{l,k}^{m} - \alpha_d^{l,k} \left\{ \underbrace{\frac{\partial C\left(\hat{\mathbf{d}}_{l,k}^{m}\right)\mathbf{f}_k}{\partial \hat{\mathbf{d}}_{l,k}^{m}}}_{\text{Computed Numerically}} (\mathbf{AH})^T \left[\mathbf{K}_{Q,l}^{-1}\left(\mathbf{y}_l - \mathbf{AH}C\left(\hat{\mathbf{d}}_{l,k}^{m}\right)\mathbf{f}_k\right) + \mathbf{K}_{MV,l}^{-1}\left(\mathbf{y}_l^{\mathrm{Pred}} - \mathbf{AH}C\left(\hat{\mathbf{d}}_{l,k}^{m}\right)\mathbf{f}_k\right) \right] + \lambda_3 \mathbf{Q}_3^T \mathbf{Q}_3 \hat{\mathbf{d}}_{l,k}^{m} \right\}$$

α: relaxation parameter, chosen to ensure convergence

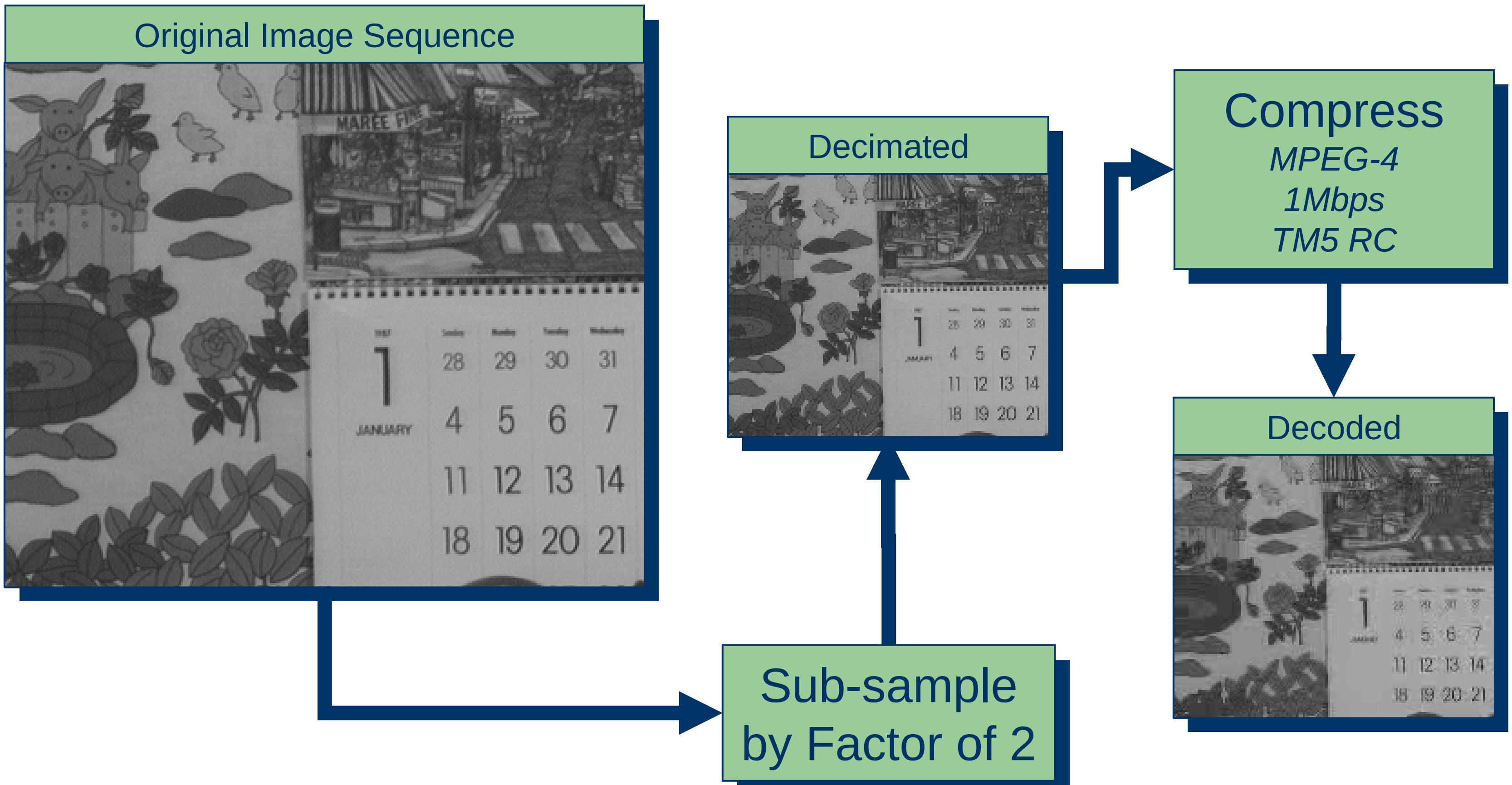
Solution Approach

- High-resolution estimates are found by

$$\hat{\mathbf{f}}_k^{n+1} = \hat{\mathbf{f}}_k^n - \alpha_f \left\{ \sum_{l=k-TB}^{k+TF} \mathbf{C}^T\left(\mathbf{d}_{l,k}\right)(\mathbf{AH})^T \left[\mathbf{K}_{Q,l}^{-1}\left(\mathbf{y}_l - \mathbf{AHC}\left(\mathbf{d}_{l,k}\right)\hat{\mathbf{f}}_k^n\right) \right. \right.$$

$$\left. \left. + \mathbf{K}_{MV,l}^{-1}\left(\mathbf{y}_l^{\text{Pred}} - \mathbf{AHC}\left(\mathbf{d}_{l,k}\right)\hat{\mathbf{f}}_k^n\right)\right] + \lambda_1 \mathbf{Q}_1^T \mathbf{Q}_1 \hat{\mathbf{f}}_k^n + \lambda_2 (\mathbf{Q}_2\mathbf{AH})^T(\mathbf{Q}_2\mathbf{AH})\hat{\mathbf{f}}_k^n \right\},$$

Simulations



Simulations

Compress
MPEG-4
1Mbps
TM5 RC

Decoded

Bi-Linear Interpolation

Simulations

Compress
MPEG-4
1Mbps
TM5 RC

Decoded

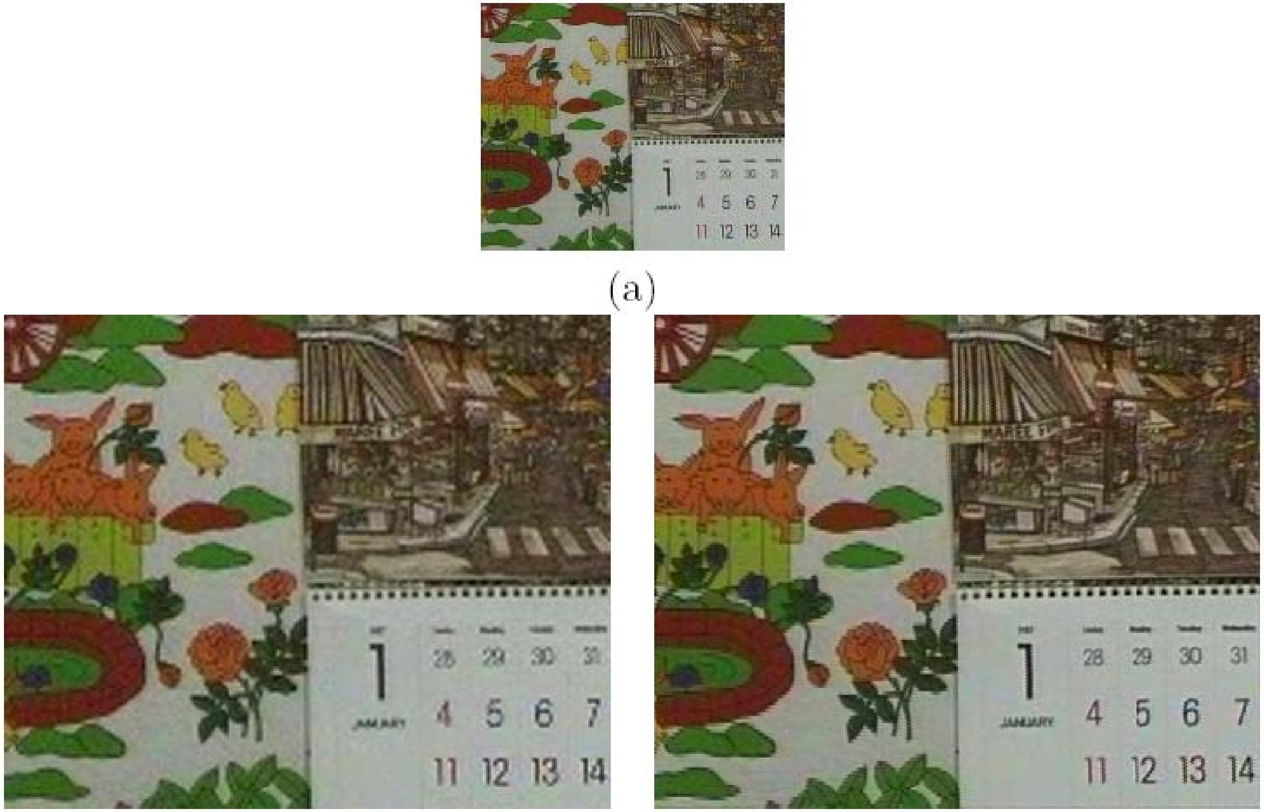
Proposed Algorithm

Simulations

Bi-Linear Interpolation

Proposed Algorithm

Experimantal Results

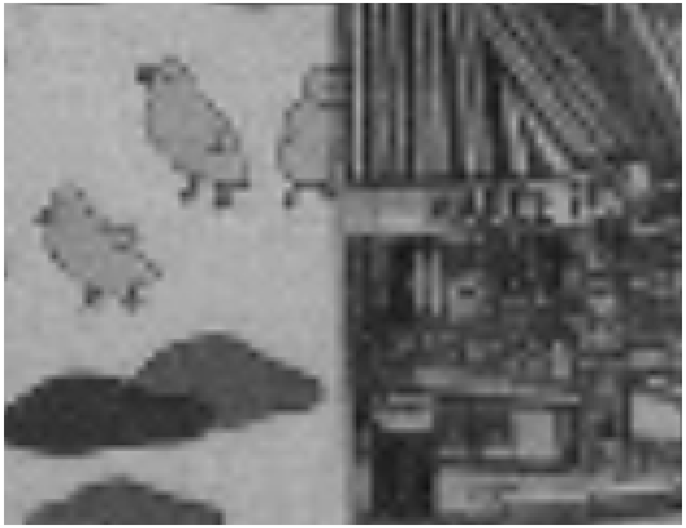


(a)

(b)

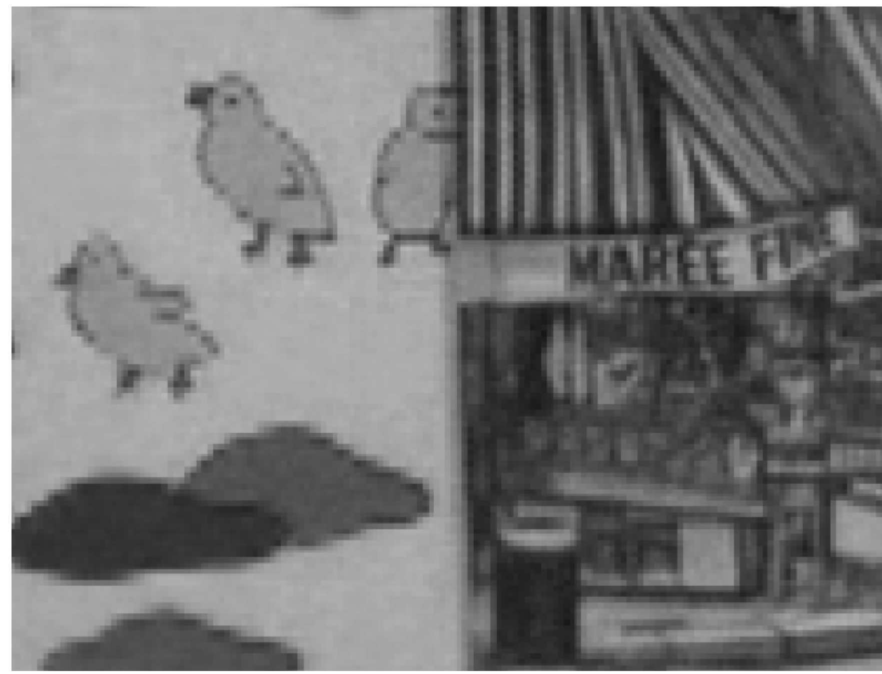
(c)

Examples



Simulations

Bilinear Interpolation



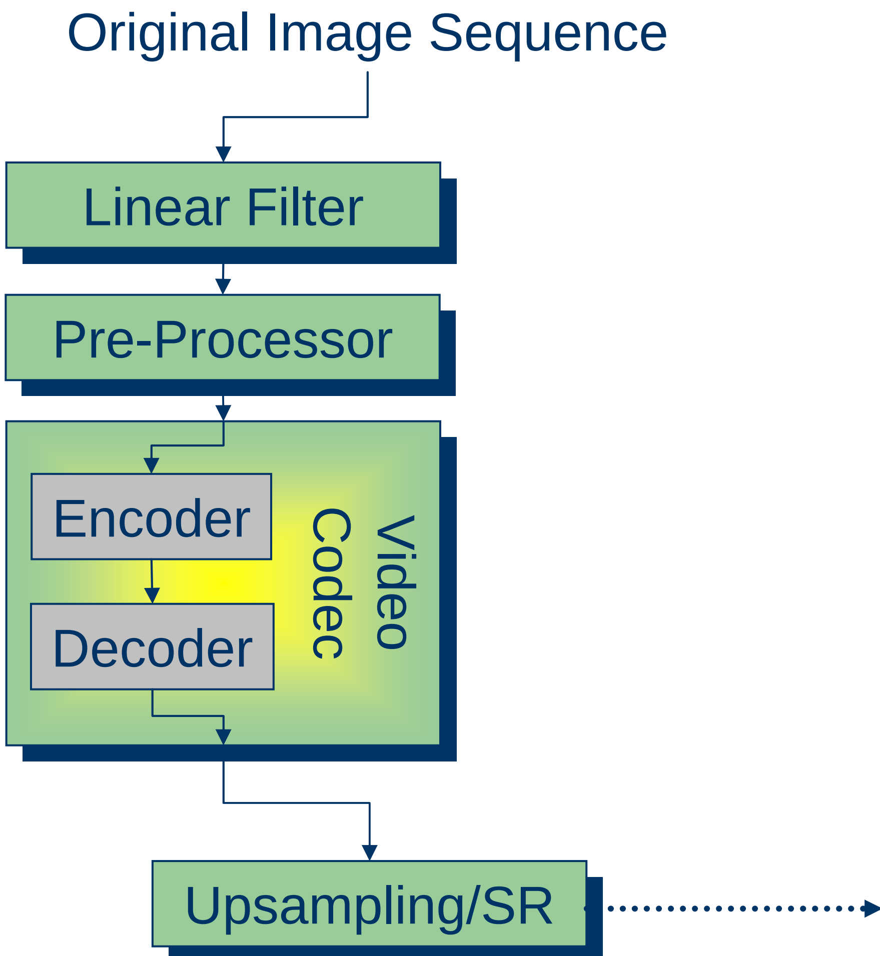
One SR frame from the sequence

L. Alvarez, R. Molina and A.K. Katsaggelos "Multichannel reconstruction of video sequences from low-resolution and compressed observations", Proc. 8th Iberoamerican Congress on Pattern Recognition (CIARP'2003), 2003.

SR for Compression

- Simulation conditions
  - Post-processor implements upsampling operation
    - Super-resolution method

      -- or --
    - Adaptive filter
  - Pre-processor transforms original image data
    - Account for linear filter
    - Account for encoder structure
    - Full knowledge of post-processor



Simulations

(b)

- Simulations
  - Rolling tomatoes sequence
  - Code at lower resolution and upsample at the decoder
  - Upsample procedure is fixed
  - Results
    - Original sequence encoded at native 1080p
      - PSNR 38dB
      - Bitrate 2.5Mbps
    - Sequence encoded at 720p and upsampled at the decoder
      - PSNR 38dB
      - Bitrate 1.7Mbps
    - Rate savings of ~30%.

Segall, A., M. Elad, P. Milanfar, R. Webb and C. Fogg, “Improved High-Definition Video by Encoding at an Intermediate Resolution," Proc SPIE Conf on VCIP, Jan. 18-22, 2004, San Jose, CA

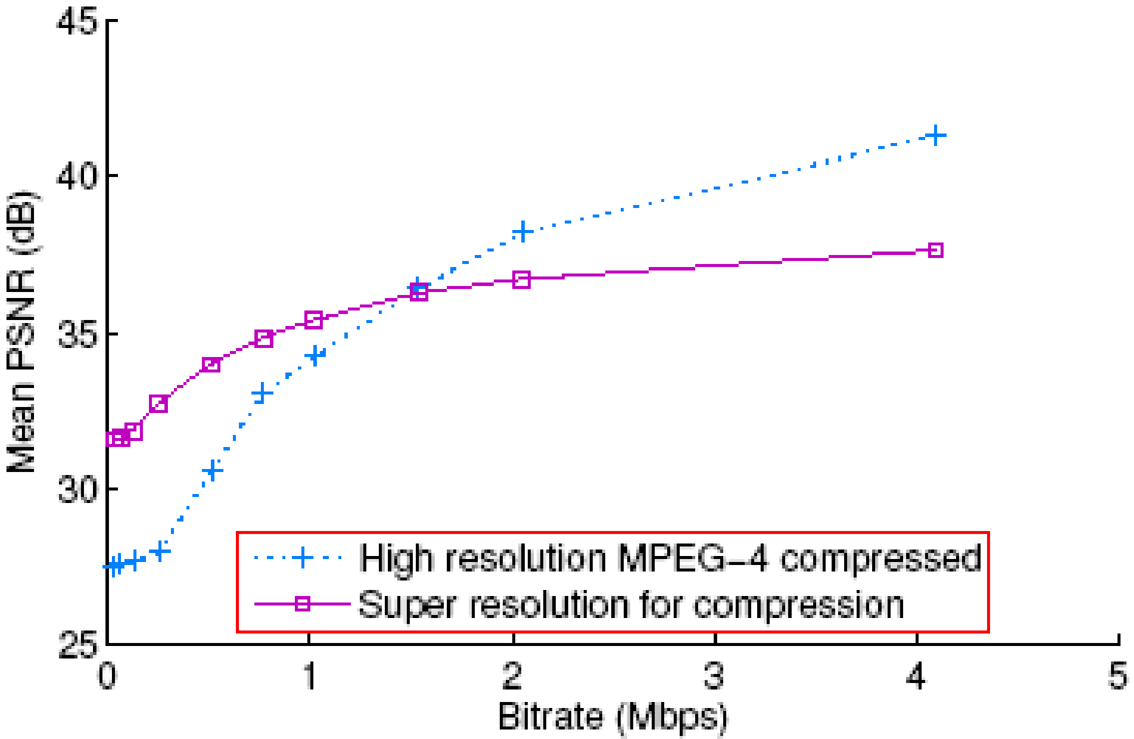
Simulations

Sequence coded at native 1080p

Sequence coded at intermediate resolution

Portion of one frame from the rolling tomatoes sequence. Notice the severity of the blocking errors.

RD Performance



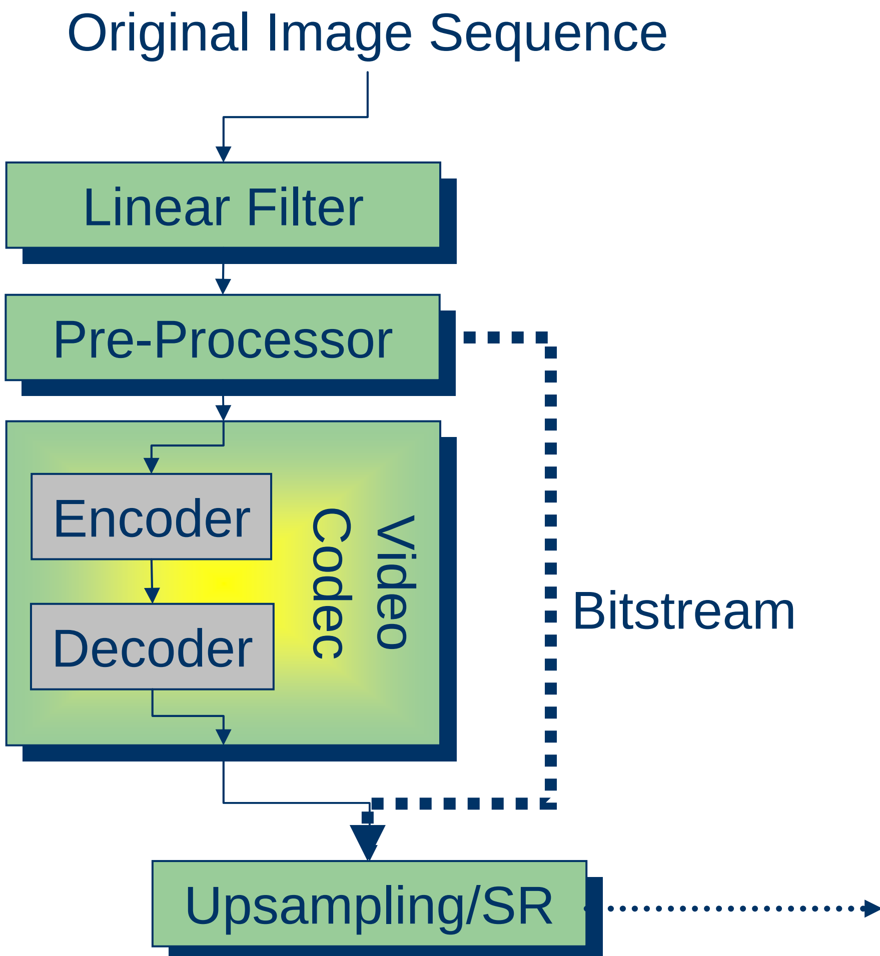
A. K. Katsaggelos, R. Molina, and J. Mateos, *SR for Images and Video*, Claypool, 2006 (forthcoming)

SR for Compression

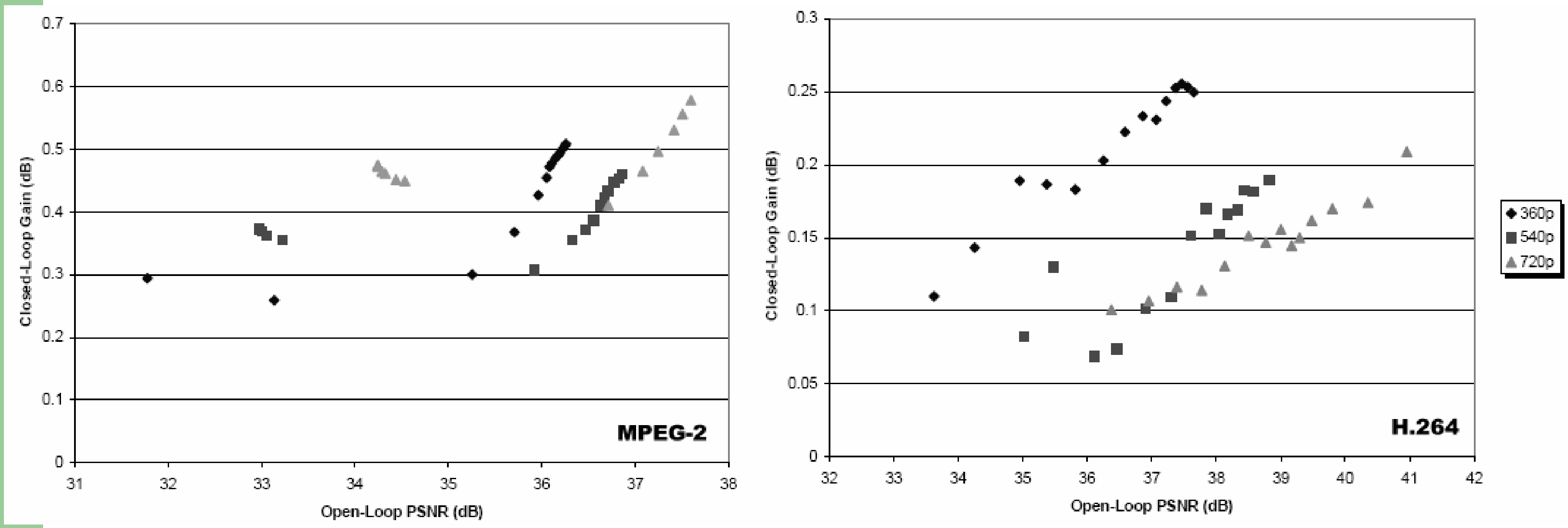
Closed Loop System

- Simulation conditions
  - Post-processor implements upsampling operation
    - Super-resolution method

      -- or --
    - Adaptive filter
  - Pre-processor transforms original image data
    - Account for linear filter
    - Account for encoder structure
    - Full knowledge of post-processor
    - ***Transmits filter parameters in the bit-stream***



Pre-for-Post: Simulations



Comparison of the open-loop and closed-loop system for the rolling tomatoes sequence. A closed-loop system allows the pre-processor to signal information to the post-processor.

Summary / Conclusions

- A plethora of traditional applications
- A paradigm was introduced that intimately couples SR and compression
- Offers a plethora of new possibilities
- Might represent a new direction in developing a new video compression standard!

Selected Publications

1. B. C. Tom and A. K. Katsaggelos, "Reconstruction of a High Resolution Image from Multiple-Degraded and Misregistered Low-Resolution Images", *Proc. 1994 SPIE Conf. on Visual Communications and Image Processing*, SPIE Vol. 2308, pp. 971-981, Chicago, IL, Sept. 1994.
2. B. C. Tom and A. K. Katsaggelos, "Reconstruction of a High-Resolution Image by Simultaneous Registration, Restoration, and Interpolation of Low-Resolution Images," *Proc. 1995 IEEE International Conf. on Image Processing*, pp. II-539-542, Washington, DC, Oct. 1995.
3. B. C. S. Tom, K. T. Lay, and A. K. Katsaggelos, "Multi-Channel Image Identification and Restoration Using the Expectation-Maximization Algorithm," *Optical Engineering*, Special Issue on "Visual Communications and Image Processing", vol. 35, no. 1, pp. 241-254, Jan. 1996.
4. B. C. Tom and A. K. Katsaggelos, "Resolution Enhancement of Monochrome and Color Video Using Motion Compensation," *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 278-287, Feb. 2001.
5. C.A. Segall, A.K. Katsaggelos, R. Molina and J. Mateos, “Super-Resolution from Compressed Video,” in *Super-Resolution Imaging*, S. Chaudhuri, editor, Kluwer Academic Publishers: Boston, MA, p. 211-242, 2001.
6. C. A. Segall, R. Molina, and A. K. Katsaggelos, " High Resolution Images from a Low-Resolution Compressed Video," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp.37-48, May 2003.
7. R. Molina, J. Abad, M. Vega, and A. K. Katsaggelos, "Parameter Estimation in Bayesian High-Resolution Image Reconstruction with Multisensors," *IEEE Trans. Image Processing*, vol.12, no.12, pp.1642-1654, Dec. 2003.
8. S.C. Park, M.G. Kang, C.A. Segall, and A.K. Katsaggelos, “Spatially Adaptive High-Resolution Image Reconstruction of Low-Resolution and Compressed Observations,” *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 573-585, Apr. 2004.
9. C.A. Segall, A.K. Katsaggelos, R. Molina and J. Mateos, “Bayesian Resolution Enhancement of Compressed Video,” *IEEE Trans. on Image Processing*, vol. 13, no. 7, pp. 898-911, July 2004.
10. L.D. Alvarez, J. Mateos, R. Molina, and A.K. Katsaggelos, "High Resolution Images from Compressed Low Resolution Video: Motion Estimation and Observable Pixels," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 58-66, August 2004.
11. N. A. Woods, N. P. Galatsanos, and A. K. Katsaggelos, "Stochastic Methods for Joint Registration and Interpolation and Multiple Under Sampled Images", to appear, *IEEE Trans. Image Processing*, Jan 2006.
12. A. K. Katsaggelos, R. Molina, and J. Mateos, *SR for Images and Video*, Claypool, 2006 (forthcoming)