Simultaneous Multi-frame MAP Super-Resolution Video Enhancement using Spatio-temporal Priors

S. Borman, Robert Stevenson

Department of Electrical Engineering

University of Notre Dame

Notre Dame, IN 46556, USA

rls@nd.edu

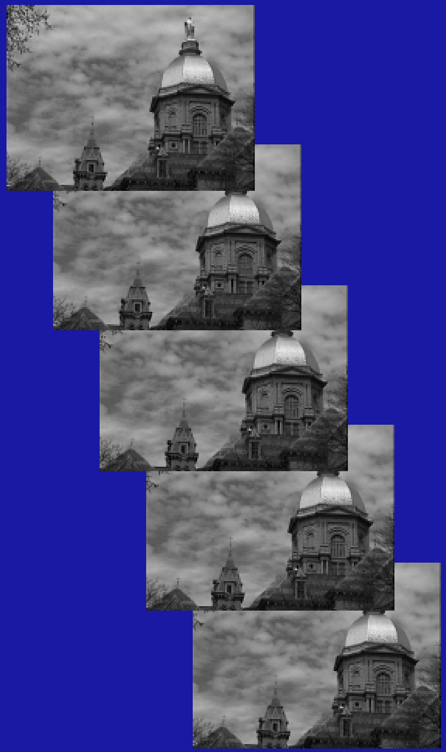
Wednesday October 27, 1999

Super-resolution video restoration

- Given a noisy, under-sampled low-resolution image sequence, restore a “super-resolved” sequence
- Use information from *multiple* images for restoration
- Sub-pixel motion contributes additional constraints
- Use *a-priori* knowledge
- Super-resolution = bandwidth extrapolation
  - Optical system diffraction limit
  - Imaging system band limit

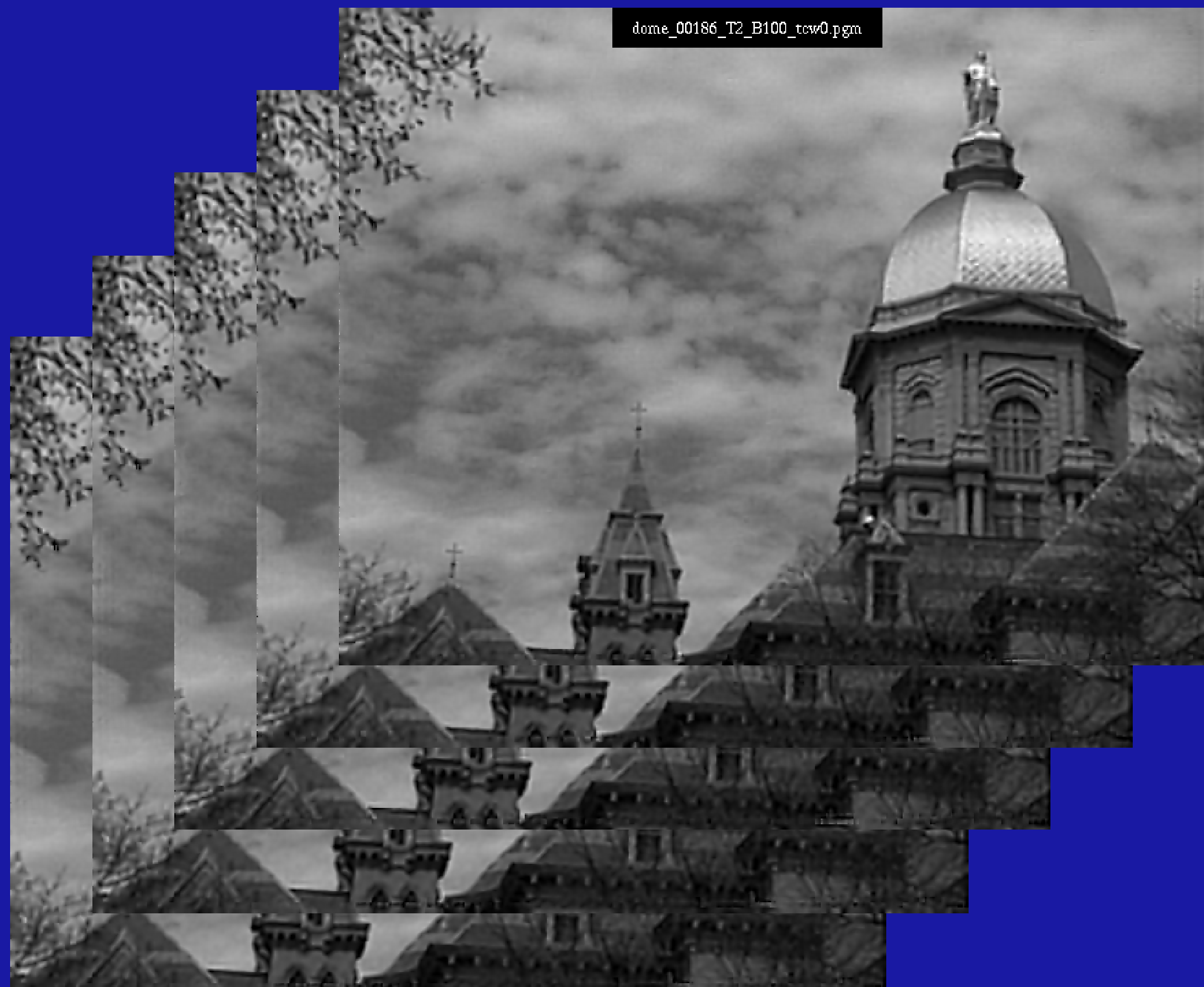
Existing approaches

- Multi-input, single output



Proposed approach

- Multi-input, multi-output



Proposed approach

- Given observed image sequence, *simultaneously* solve for all restored images
- Allows incorporation of *temporal* constraints in addition to spatial constraints
- Include confidence parameters for sub-pixel motion estimates

Summary of the problem

Given:	A short video sequence p observed LR frames N_1 rows by N_2 columns
Find:	p corresponding super-resolved images qN_1 rows by qN_2 columns, $q \in \mathbb{N}$
Approach:	Linear inverse problem observation model Stochastic regularization (ill-posedness) Bayesian MAP estimate with MRF prior

Observation model

- Relate observed LR sequence to unknown SR sequence
- Utilize linear model $\Longrightarrow$ linear inverse problem
- Observation model must include effects of:
  - motion
  - imaging degradation
- Identify two observation modes
  - temporally coincident
  - temporally non-coincident

Temporally coincident observation

- Observe LR frame $\mathbf{y}^{(l)}$ from SR frame $\mathbf{z}^{(l)}$
- Assume *known* SVPSF degradation $\mathbf{A}^{(l,l)}$

$$
\begin{array}{cccccccc}
\mathbf{z}^{(1)} & \mathbf{z}^{(2)} & \cdots\cdots & \mathbf{z}^{(l)} & \cdots\cdots & \mathbf{z}^{(p)} \\
| & | & & | & & | \\
\mathbf{A}^{(1,1)} & \mathbf{A}^{(2,2)} & & \mathbf{A}^{(l,l)} & & \mathbf{A}^{(p,p)} \\
\downarrow & \downarrow & & \downarrow & & \downarrow \\
\mathbf{y}^{(1)} & \mathbf{y}^{(2)} & \cdots\cdots & \mathbf{y}^{(l)} & \cdots\cdots & \mathbf{y}^{(p)}
\end{array}
$$

Temporally coincident observation

- Additive noise (measurement errors / errors in $\mathbf{A}^{(l,l)}$)

$$\mathbf{y}^{(l)} = \mathbf{A}^{(l,l)}\mathbf{z}^{(l)} + \mathbf{n}^{(l,l)}\ , \quad l \in P,$$

where,

$$\begin{aligned} P &= \{1, 2, \ldots, p\} \\ \mathbf{y}^{(l)}, \mathbf{n}^{(l)} &\in \mathbb{R}^{N_1 N_2 \times 1} \\ \mathbf{z}^{(l)} &\in \mathbb{R}^{q^2 N_1 N_2 \times 1} \\ \mathbf{A}^{(l,l)} &\in \mathbb{R}^{N_1 N_2 \times q^2 N_1 N_2} \end{aligned}$$

Temporally coincident observation

- Noise model — IID Gaussian, zero mean, per-pixel variances ${\sigma_j^{(l,l)}}^2$

$$\mathcal{P}\left(\mathbf{n}^{(l,l)}\right) = \frac{1}{(2\pi)^{\frac{N_1 N_2}{2}} \left|\mathbf{K}^{(l,l)}\right|} \exp\left\{-\frac{1}{2} {\mathbf{n}^{(l,l)}}^T {\mathbf{K}^{(l,l)}}^{-1} \mathbf{n}^{(l,l)}\right\}$$

$$\mathbf{K}^{(l,l)} = \operatorname{diag}\left({\sigma_1^{(l,l)}}^2, {\sigma_2^{(l,l)}}^2, \cdots, {\sigma_{N_1 N_2}^{(l,l)}}^2\right)$$

- Observation noise assumed independent over $\{\mathbf{y}^{(l)}\}_1^p$

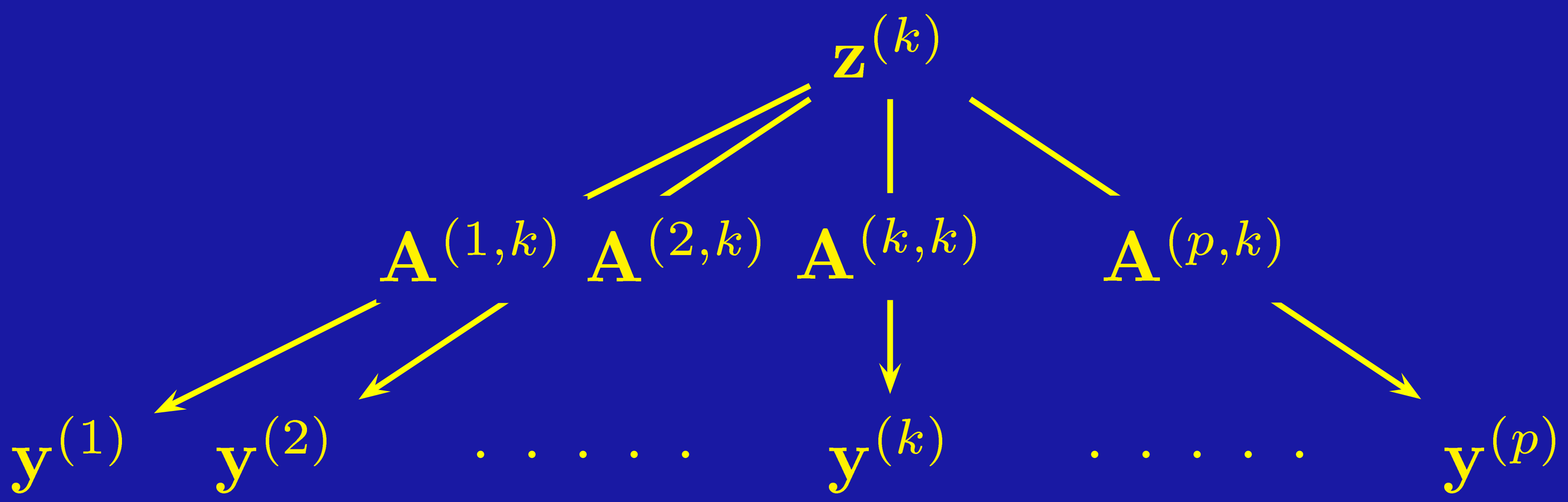
Temporally non-coincident observation

- Motion compensated observation — more constraints
- Relate $\{\mathbf{z}^{(k)}\}_1^p$ to temporally non-coincident LR frame
- Find: *motion compensating* observation matrix $\mathbf{A}^{(l,k)}$ and vector $\mathbf{n}^{(l,k)}$ s.t.

$$\mathbf{y}^{(l)} = \mathbf{A}^{(l,k)}\mathbf{z}^{(k)} + \mathbf{n}^{(l,k)}\ , \quad l \in P \setminus \{k\}$$

$$\mathcal{P}\left(\mathbf{n}^{(l,k)}\right) = \frac{1}{(2\pi)^{\frac{N_1 N_2}{2}}\left|\mathbf{K}^{(l,k)}\right|} \exp\left\{-\frac{1}{2}\mathbf{n}^{(l,k)^T}\mathbf{K}^{(l,k)^{-1}}\mathbf{n}^{(l,k)}\right\}$$

Combined observation model



$$\begin{bmatrix} \mathbf{y}^{(1)} \\ \vdots \\ \mathbf{y}^{(p)} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{(1,k)} \\ \vdots \\ \mathbf{A}^{(p,k)} \end{bmatrix} \mathbf{z}^{(k)} + \begin{bmatrix} \mathbf{n}^{(1,k)} \\ \vdots \\ \mathbf{n}^{(p,k)} \end{bmatrix}$$

$$\mathbf{y} = \mathbf{A}^{(*,k)} \mathbf{z}^{(k)} + \mathbf{n}^{(*,k)}$$

Simultaneous multi-frame observation

Simultaneous observation model for $\{\mathbf{z}^{(k)}\}_1^p$

$$\mathbf{y}^{(l)} = \mathbf{A}^{(l,k)}\mathbf{z}^{(k)} + \mathbf{n}^{(l,k)}\ , \quad l,k \in P$$

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y} \\ \vdots \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{(*,1)} & & & \\ & \mathbf{A}^{(*,2)} & & \\ & & \ddots & \\ & & & \mathbf{A}^{(*,p)} \end{bmatrix} \begin{bmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \\ \vdots \\ \mathbf{z}^{(p)} \end{bmatrix} + \begin{bmatrix} \mathbf{n}^{(*,1)} \\ \mathbf{n}^{(*,2)} \\ \vdots \\ \mathbf{n}^{(*,p)} \end{bmatrix}$$

This may be expressed in the classic form:

$$\mathbf{Y} = \mathbf{A}\mathbf{Z} + \mathbf{N}$$

Combined noise model

$$\mathcal{P}_{\mathbf{N}}(\mathbf{N}) = \frac{1}{(2\pi)^{\frac{p^2 N_1 N_2}{2}} |\mathbf{K}|} \exp\left\{-\frac{1}{2}\mathbf{N}^T \mathbf{K}^{-1} \mathbf{N}\right\}$$

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}^{(1,1)} & & & \\ & \mathbf{K}^{(2,2)} & & \\ & & \ddots & \\ & & & \mathbf{K}^{(p,p)} \end{bmatrix}$$

- $\mathbf{K}$ (diagonal) represents *a-priori* knowledge concerning observations and accuracy of motion information

Bayesian estimation of SR images

Maximize the a-posteriori probability:

$$
\begin{aligned}
\hat{\mathbf{Z}}_{\text{MAP}} &= \arg\max_{\mathbf{Z}} \left\{ \mathcal{P}\left(\mathbf{Z} | \mathbf{Y}\right) \right\} \\
&= \arg\max_{\mathbf{Z}} \left\{ \frac{\mathcal{P}\left(\mathbf{Y} | \mathbf{Z}\right) \mathcal{P}\left(\mathbf{Z}\right)}{\mathcal{P}\left(\mathbf{Y}\right)} \right\} \\
&= \arg\max_{\mathbf{Z}} \left\{ \log \mathcal{P}\left(\mathbf{Y} | \mathbf{Z}\right) + \log \mathcal{P}\left(\mathbf{Z}\right) \right\}
\end{aligned}
$$

$\mathcal{P}\left(\mathbf{Y} | \mathbf{Z}\right)$ — Likelihood function

$\mathcal{P}\left(\mathbf{Z}\right)$ — Prior

The likelihood function

- Determined by noise pdf:

$$
\begin{aligned}
\mathcal{P}\left(\mathbf{Y}|\mathbf{Z}\right) &= \mathcal{P}_{\mathbf{N}}(\mathbf{Y}-\mathbf{AZ}) \\
&= \frac{1}{(2\pi)^{\frac{q^2 N_1 N_2}{2}} |\mathbf{K}|} \cdot \\
&\qquad \exp\left\{-\frac{1}{2}(\mathbf{Y}-\mathbf{AZ})^T\mathbf{K}^{-1}(\mathbf{Y}-\mathbf{AZ})\right\}
\end{aligned}
$$

- Choose **K** according to confidence in observations
  - noise variance (temporally coincident)
  - motion estimator MAD measure (non-coincident)

Image prior model

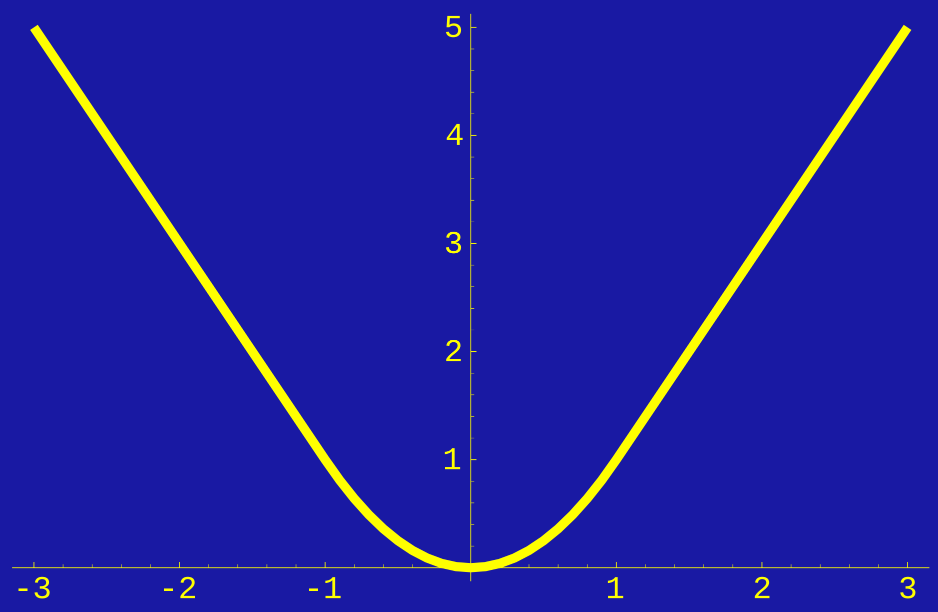
- Markov random field (MRF)
- Local interaction $\Longleftrightarrow$ global pdf
- Pdf is Gibbsian (Hammersley-Clifford)

$$\mathcal{P}(\mathbf{Z}) = \frac{1}{k_p} \exp\left\{ -\frac{1}{\beta} \sum_{c \in \mathcal{C}} \rho_\alpha \left(\partial_c \mathbf{Z} \right) \right\}$$

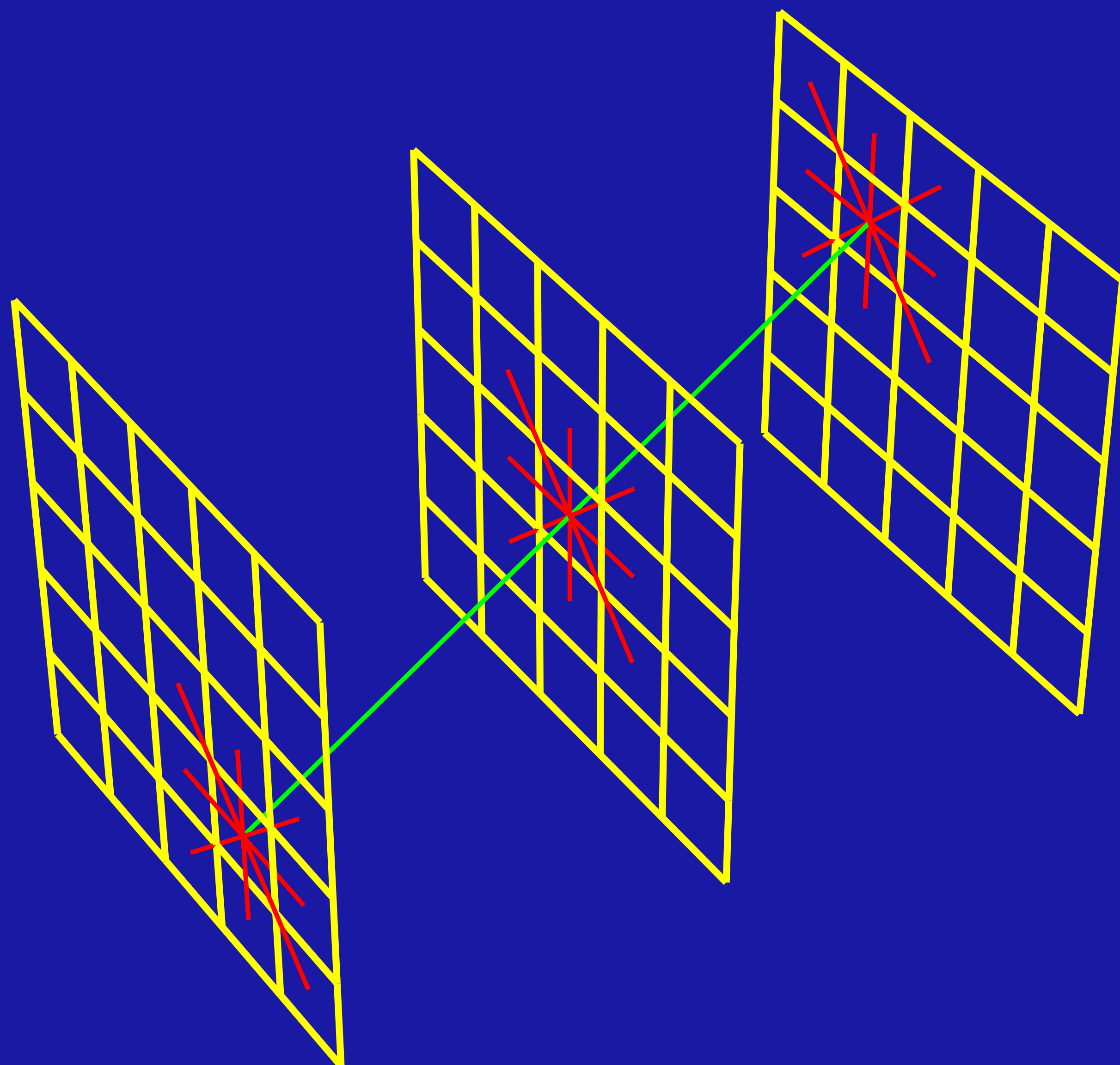
- Spatio-temporal smoothness
- Convex nonlinear activity penalization function $\rho_\alpha(\cdot)$

Huber penalty function

$$\rho_\alpha(x) = \begin{cases} x^2 & |x| \le \alpha \\ 2\alpha|x| - \alpha^2 & |x| > \alpha \end{cases}$$



Spatio-temporal clique structure



Spatio-temporal clique structure

Spatial activity measure — spatial derivatives

$$\partial_1 \mathbf{z}^{(k)} = z^{(k)}_{n_1-1,n_2} - 2z^{(k)}_{n_1,n_2} + z^{(k)}_{n_1+1,n_2}$$

$$\partial_2 \mathbf{z}^{(k)} = z^{(k)}_{n_1,n_2-1} - 2z^{(k)}_{n_1,n_2} + z^{(k)}_{n_1,n_2+1}$$

$$\partial_3 \mathbf{z}^{(k)} = \tfrac{1}{2}\, z^{(k)}_{n_1+1,n_2-1} - z^{(k)}_{n_1,n_2} + \tfrac{1}{2}\, z^{(k)}_{n_1-1,n_2+1}$$

$$\partial_4 \mathbf{z}^{(k)} = \tfrac{1}{2}\, z^{(k)}_{n_1-1,n_2-1} - z^{(k)}_{n_1,n_2} + \tfrac{1}{2}\, z^{(k)}_{n_1+1,n_2+1}$$

Temporal smoothness — time derivative along motion

$$\partial_5 \mathbf{z}^{(k)} = z^{(k-1)}_{n_1+\delta_1,n_2+\delta_2} - 2z^{(k)}_{n_1,n_2} + z^{(k+1)}_{n_1+\Delta_1,n_2+\Delta_2}$$

Objective function

- Combine likelihood and prior terms

$$\hat{\mathbf{Z}}_{\text{MAP}} = \arg\max_{\mathbf{Z}} \left\{ -\frac{1}{2}(\mathbf{Y}-\mathbf{AZ})^T\mathbf{K}^{-1}(\mathbf{Y}-\mathbf{AZ}) - \frac{1}{\beta}\sum_{c\in\mathcal{C}} \rho_\alpha(\partial_c \mathbf{Z}) \right\}$$

$$\hat{\mathbf{Z}}_{\text{MAP}} = \arg\min_{\mathbf{Z}} \left\{ \frac{1}{2}(\mathbf{Y}-\mathbf{AZ})^T\mathbf{K}^{-1}(\mathbf{Y}-\mathbf{AZ}) + \frac{1}{\beta}\sum_{c\in\mathcal{C}} \rho_\alpha(\partial_c \mathbf{Z}) \right\}$$

Optimization

- Convex prior ensures convexity
- $\Longrightarrow$ Existence, uniqueness
- Convex, non-quadratic optimization
- Gradient descent optimization applicable
- Tractable even though dimensions are high

Original low resolution frames

Zero order hold

Cubic spline interpolation

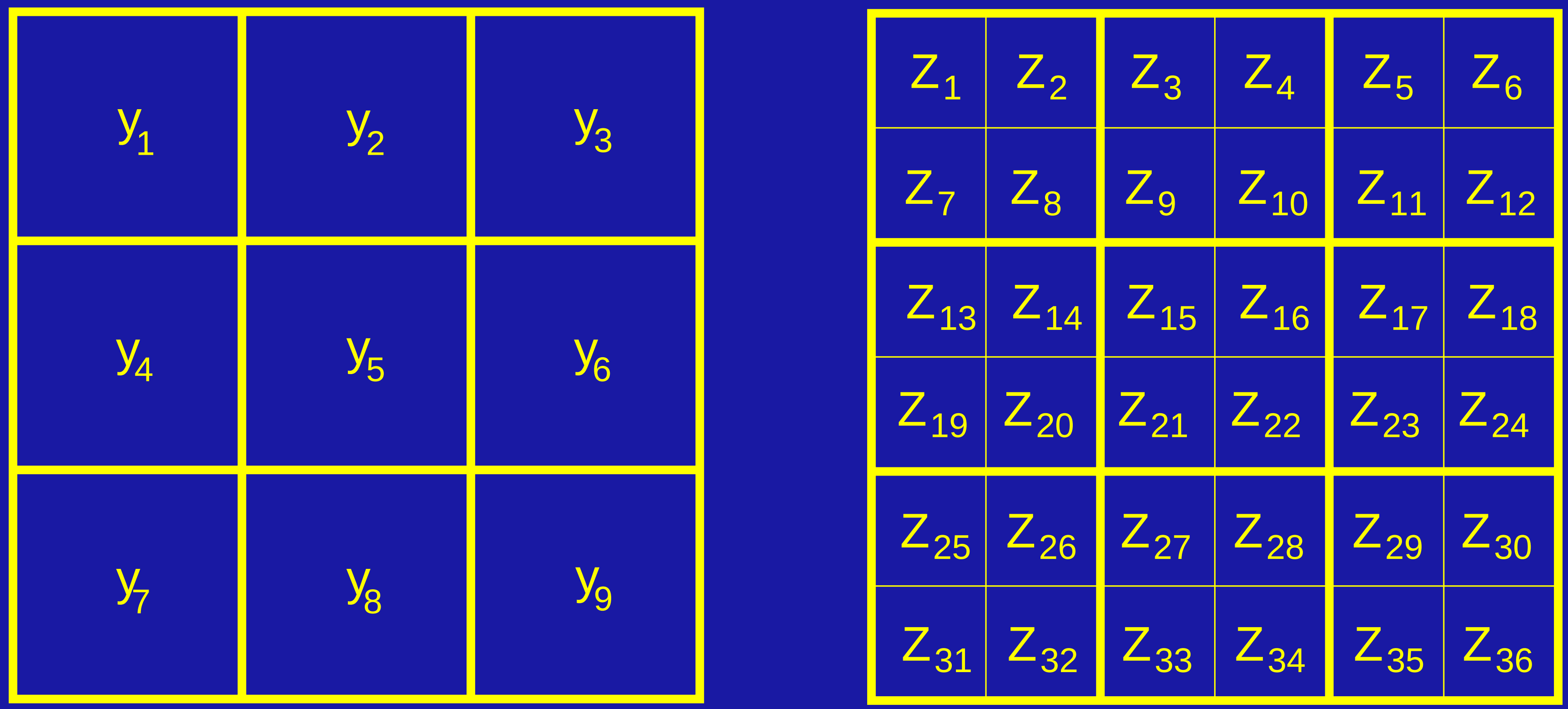
Super-resolution restoration

Super-resolution vs cubic spline

Summary and extensions

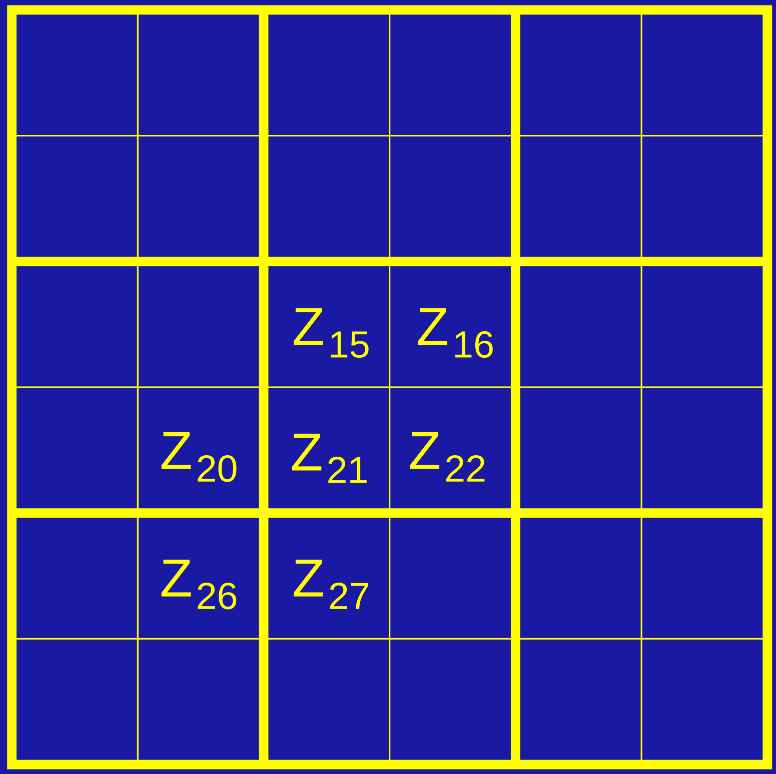
- Summary
  - Simultaneous multi-frame approach enables inclusion of temporal smoothness constraints
  - Utilize covariance matrix **K** to represent confidence in observations and motion estimates
- Extensions
  - Generalize to arbitrary length sequences
  - Arbitrary positive definite **K**

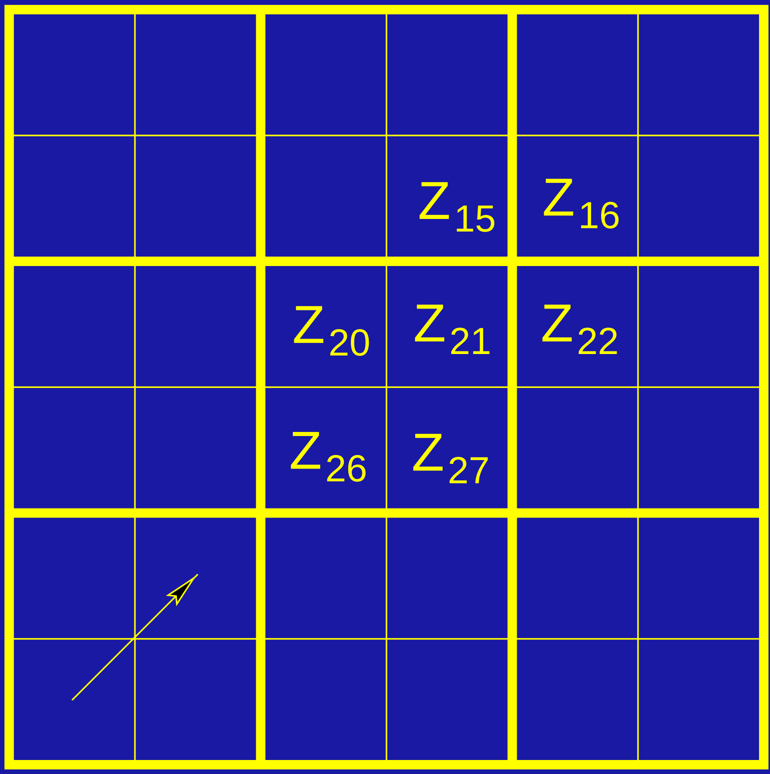
Temporally coincident observation



Observed image : $y_5 = \frac{1}{4}(z_{15} + z_{16} + z_{21} + z_{22})$

Temporally non-coincident observation





Original image : $y_{o_5} = \frac{1}{4}(z_{15} + z_{16} + z_{21} + z_{22})$

Translated image : $y_{t_5} = \frac{1}{4}(z_{20} + z_{21} + z_{26} + z_{27})$