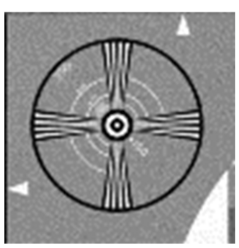
Chapter 4

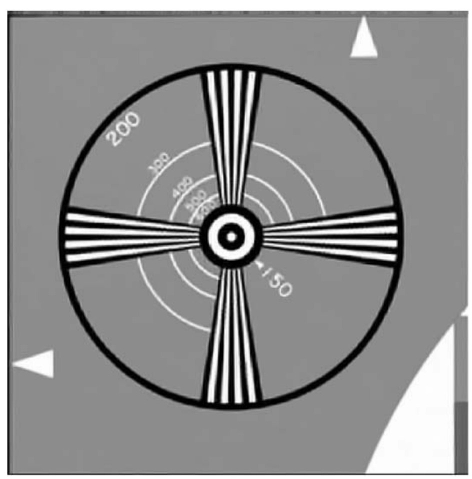
Multiframe Super-resolution
Image Reconstruction

Multi-frame SRIR (Video Enhancement)

19 frames
of size
57×49



Low
resolution
video

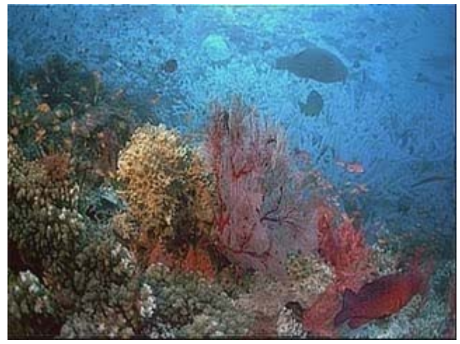


Our method

Source: https://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html

Multi-frame SRIR (Video Enhancement)

Input Video



High-definition Video

3

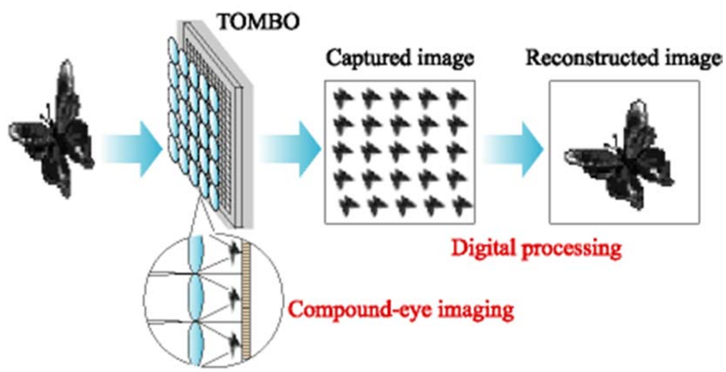
Outline

1. High-resolution Image Reconstruction
2. Video Still Enhancement Models
   - Classical
   - Tight-frame
   - Low-rank
3. Experiments

4

Super-resolution Image Reconstruction (SRIR)

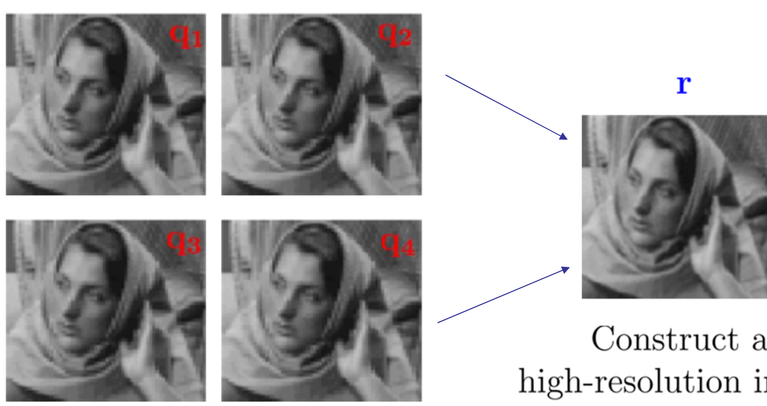
Not one, but many lens—
compound eyes



5

Super-resolution Image Reconstruction (SRIR)

Four low resolution images (64×64) of the same scene $\{\mathbf{q}_i\}_{i=1}^4$. Each shifted by exactly half pixel length.



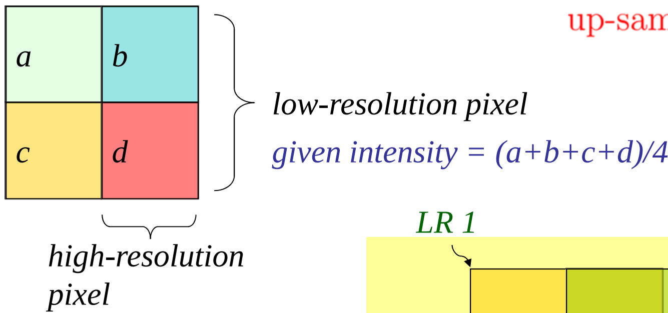
up-sample by 2

Construct a
high-resolution image
(128×128) from them

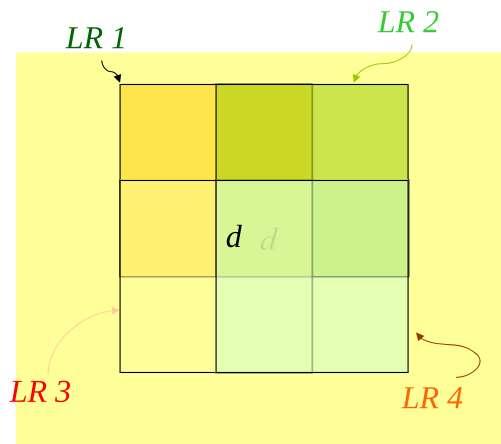
6

Modeling of SRIR

up-sample by 2



4 LR images are placed half-pixel apart



Modeling of SRIR

From the first LR image:

$$(E_1 \otimes E_1)\mathbf{r} = \mathbf{q}_1$$

where

$$E_1 = \frac{1}{2}\begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & 0 \\ \ddots & 1 & 1 & 0 & 0 & \ddots & \ddots & \\ \ddots & 0 & 0 & 1 & 1 & 0 & \ddots & \ddots \\ & \ddots & \ddots & 0 & 0 & 1 & 1 & \ddots \\ 0 & & \ddots & \ddots & \ddots & \ddots & \cdots & \cdots \end{bmatrix}_{\frac{n}{4}\times n}.$$

$$B_1\mathbf{r} = \mathbf{q}_1$$

8

Modeling of SRIR

From the second LR image:

$$(E_1 \otimes E_2)\mathbf{r} = \mathbf{q}_2$$

where

$$E_2 = \frac{1}{2}\begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & 0 \\ \ddots & 0 & 1 & 1 & \ddots & \ddots & & \\ \ddots & 0 & 0 & 0 & 1 & 1 & \ddots & \ddots \\ & \ddots & \ddots & 0 & 0 & 0 & 1 & \ddots \\ 0 & & \ddots & \ddots & \ddots & \ddots & \cdots & \cdots \end{bmatrix}_{\frac{n}{4}\times n}.$$

$$B_2\mathbf{r} = \mathbf{q}_2$$

9

Modeling of SRIR

3rd sensor: $B_3\mathbf{r} = (E_2 \otimes E_1)\mathbf{r} = \mathbf{q}_3$.

4th sensor: $B_4\mathbf{r} = (E_2 \otimes E_2)\mathbf{r} = \mathbf{q}_4$.

Combining all the LR image equations:

$$\begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \end{bmatrix} \mathbf{r} = \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \mathbf{q}_3 \\ \mathbf{q}_4 \end{bmatrix}.$$

After rearranging the rows:

$$B\mathbf{r} = \mathbf{q}$$

where B is a block-Toeplitz-Toeplitz-block matrix.

10

Outline

1. **High-resolution Image Reconstruction**
2. **Video Still Enhancement Models**
   - **Classical**
   - **Tight-frame**
   - **Low-rank**
3. **Experiments**

11

Video Still Enhancement

A 352-by-288 video from a video recorder

30 frames/second

Tight-frame method using 21 frames

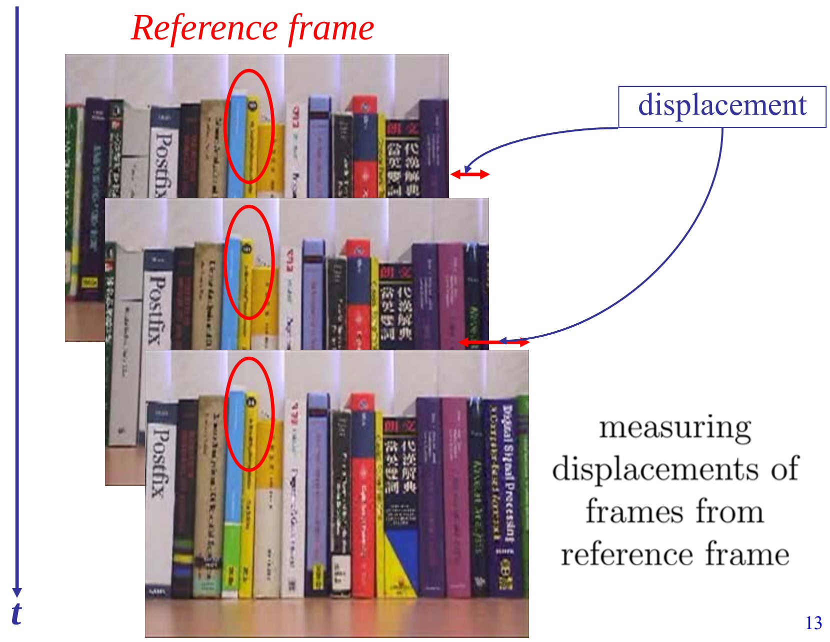
[C., Shen, & Xia, ACHA 07]

frames are not aligned at exactly half-pixel length

12

Aligning the Frames

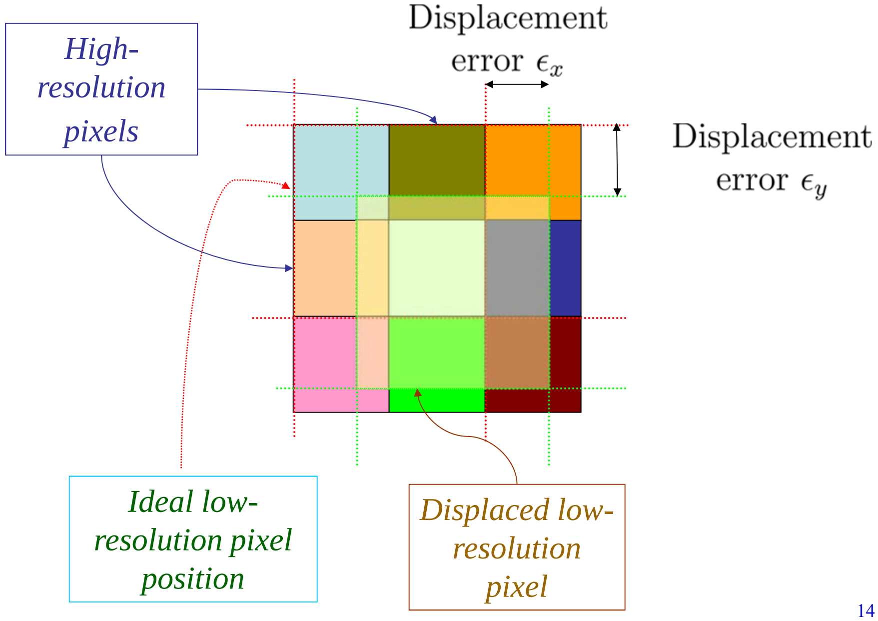
Reference frame



measuring displacements of frames from reference frame

13

Sub-pixel Displacement



14

Modeling of Video Still Enhancement

First LR image equation: $(E_x^1 \otimes E_y^1)\mathbf{r} = \mathbf{q}_1$: $B_1\mathbf{r} = \mathbf{q}_1$

$$E_j^1 = \frac{1}{2}\begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & 0 \\ \ddots & 1-\epsilon_j^1 & 1 & \epsilon_j^1 & 0 & \ddots & \ddots & \\ \ddots & 0 & 1-\epsilon_j^1 & 1 & \epsilon_j^1 & 0 & \ddots & \ddots \\ & \ddots & \ddots & 0 & 1-\epsilon_j^1 & 1 & \epsilon_j^1 & \ddots \\ 0 & & \ddots & \ddots & \ddots & \ddots & \cdots & \cdots \end{bmatrix}.$$

Similarly for each LR images from the video:

$$B_i\mathbf{r} = \mathbf{q}_i, \quad i = 2, 3, \ldots, m.$$

15

Modeling of SRIR

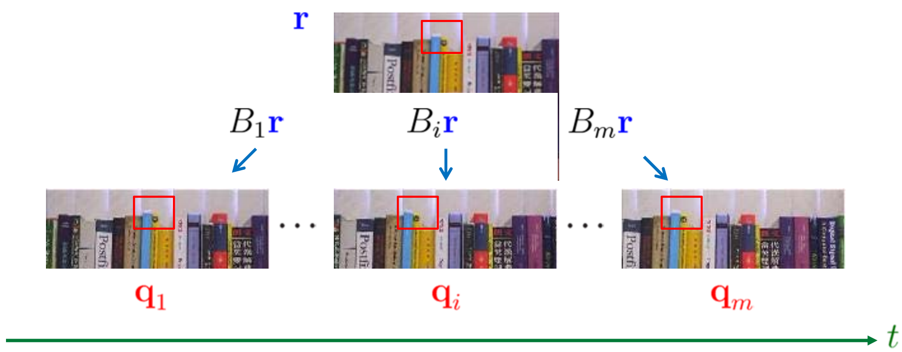
Combining all the LR image equations:

$$\begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ \vdots \\ B_m \end{bmatrix} \mathbf{r} = \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \mathbf{q}_3 \\ \vdots \\ \mathbf{q}_m \end{bmatrix}.$$

- Matrix is non-square and no structure.
- HR image $\mathbf{r}$ can be obtained by solving a least-squares problem.

16

Classical Approach [Tsai & Huang, 84]



- $\mathbf{q}_i = B_i\mathbf{r} + \mathbf{n}_i, i = 1, \ldots, m.$
- Solve, e.g.

$$\min_{\mathbf{r}} \|\nabla \mathbf{r}\|_1 + \frac{\lambda}{2}\sum_{i=1}^{m} \|B_i\mathbf{r} - \mathbf{q}_i\|_2^2.$$

17

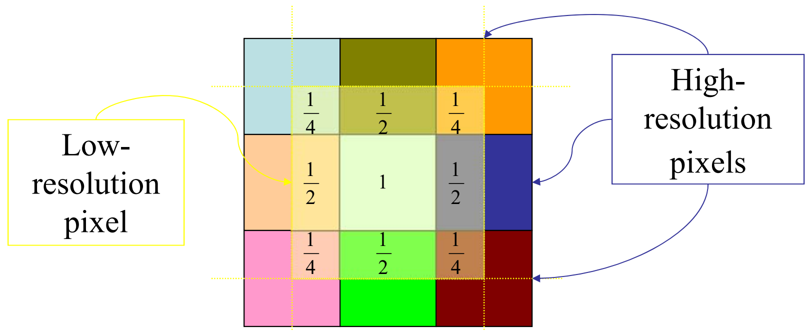
Outline

1. **High-resolution Image Reconstruction**
2. **Video Still Enhancement Models**
   - **Classical**
   - **Tight-frame**
   - **Low-rank**
3. **Experiments**

18

Tightframe Approach [C² + S², *SISC* (2003)]

4 LR images merge into 1 HR image:



HR images $\rightarrow$ LR image

Averaging process = a lowpass filter with refinement mask

$$a \equiv \frac{1}{2}\left(\ldots, 0, \frac{1}{2}, 1, \frac{1}{2}, 0, \ldots\right) \otimes \frac{1}{2}\left(\ldots, 0, \frac{1}{2}, 1, \frac{1}{2}, 0, \ldots\right)$$

19

Piecewise Linear Tight Frame

Tight-frame filters from linear spline:

$$h_0 = \frac{1}{4}[1, 2, 1], \; h_1 = \frac{\sqrt{2}}{4}[1, 0, -1], \; h_2 = \frac{1}{4}[-1, 2, -1].$$

Then

$$h_0 \longleftrightarrow H_0 = \frac{1}{4}\begin{bmatrix} 2 & 1 & & 0 & 1 \\ 1 & 2 & 1 & & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & & 1 & 2 & 1 \\ 1 & 0 & & 1 & 2 \end{bmatrix}.$$

In 2D, H_i are block-circulant-circulant-block matrices.

20

From LR Images to HR Image

$$\underbrace{\begin{bmatrix} & & \vdots & & & \\ \cdots & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \cdots & \cdots \\ \cdots & \cdots & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \cdots \\ & & \vdots & & & H_0 \end{bmatrix}}_{\text{piecewise linear tightframe transform}} \begin{bmatrix} \mathbf{q}_1 \\ \vdots \\ \vdots \\ \mathbf{q}_m \end{bmatrix}$$

LR images are framelet coefficients of $\mathbf{r}$

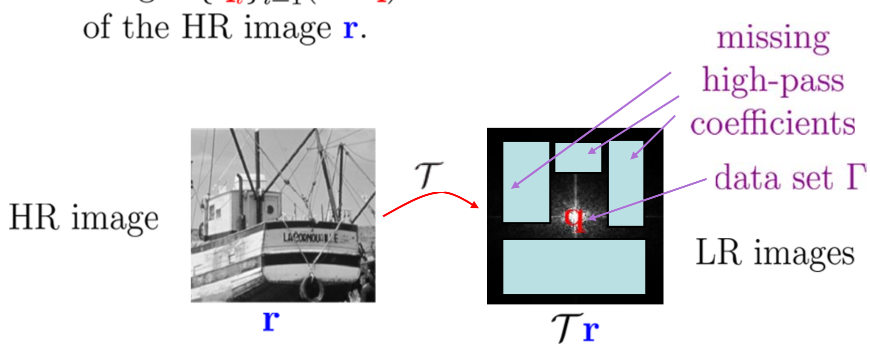
high-pass coefficients are missing

piecewise linear tightframe transform

Key Observation

reconstruting HR image = frequency domain inpainting

- Given a tightframe transform $\mathcal{T}$, the given LR images $\{\mathbf{q}_i\}_{i=1}^m (:= \mathbf{q})$ are the framelet coefficients of the HR image $\mathbf{r}$.



- SRIR: Find $\mathbf{r}$ such that $P_\Gamma \mathcal{T}\mathbf{r} = \mathbf{q}$.

Tightframe SRIR

- Given a tightframe transform $\mathcal{T}$, it solves:

$$\min_{\mathbf{r}, P_\Gamma \mathbf{c}=\mathbf{q}} \|\text{diag}(\boldsymbol{\lambda})\mathbf{c}\|_1 + \frac{1}{2}\|\mathcal{T}\mathbf{r} - \mathbf{c}\|_2^2,$$

  - $\mathbf{c}$: tightframe coefficients of HR image $\mathbf{r}$
  - fidelity term is $\|\mathcal{T}\mathbf{r} - \mathbf{c}\|_2^2$, not $\|B_i\mathbf{r} - \mathbf{q}_i\|_2^2$
  - P_Γ: restrict to the domain of LR images $\{\mathbf{q}_i\}_{i=1}^m$
- Thresholding algorithm proposed in SISC in 2003 [R. Chan, T. Chan, L. Shen, Z. Shen, $C^2 + S^2$].
- First iterative thresholding algorithm.

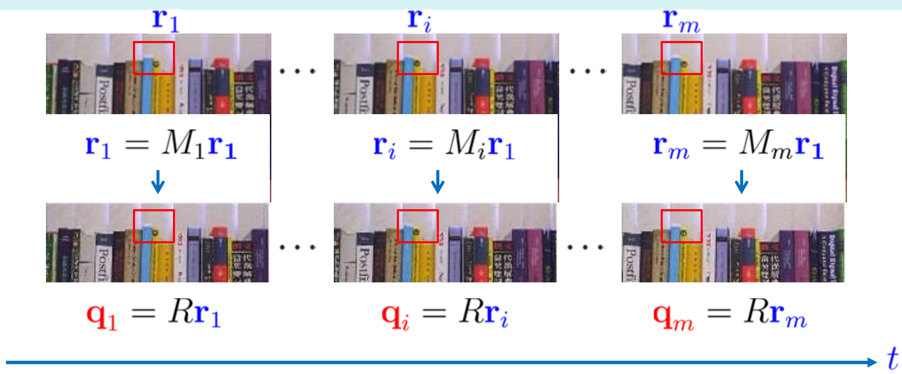
23

Outline

1. **High-resolution Image Reconstruction**
2. **Video Still Enhancement Models**
   - **Classical**
   - **Tight-frame**
   - **Low-rank**
3. **Experiments**

24

Low-rank Approach [ArXiv: 170406196]

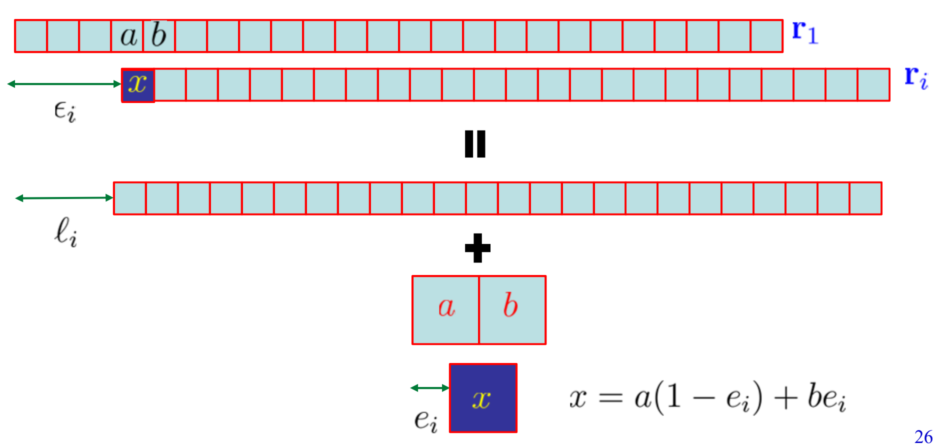


- ☐ $\mathbf{q}_i = R\mathbf{r}_i + \mathbf{n}_i, i = 1, \ldots, m.$
- ☐ $\mathbf{r}_i$ are related to $\mathbf{r}_1$ by the motion matrix M_i.
- ☐ $\{\mathbf{r}_i\}_{i=1}^m$ are all high-resolution images defined on pixels with same size.

Motion Matrices in 1D

If ith HR frame $\mathbf{r}_i$ has only horizontal displacement

$$\epsilon_i = \lfloor \epsilon_i \rfloor + e_i := \ell_i + e_i$$



$$x = a(1 - e_i) + be_i$$

Decomposition of the Motion Matrices

- Hence

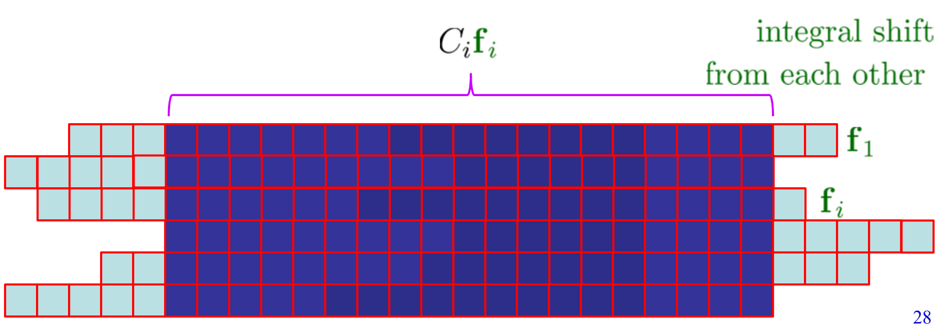
$$M_i = T(e_i)S^{\ell_i}$$

- S^{ℓ_i} is the shift operator S applied ℓ_i times.
- $$T(e) = \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & & & 0 \\ \ddots & 1-e & e & 0 & \ddots & & \\ \ddots & 0 & 1-e & e & 0 & \ddots & \\ & \ddots & 0 & 1-e & e & 0 & \ddots \\ 0 & & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$
- Both $T(e_i)$ and S^{ℓ_i} are Toeplitz

27

Creating Low Rank Structure

- Then $\mathbf{r}_i = M_i\mathbf{r}_1 = T(e_i)S^{\ell_i}\mathbf{r}_1 := T_i\mathbf{f}_i$. $\quad$ $S^{\ell_i}\mathbf{r}_1 = \mathbf{f}_i$
- $\mathbf{q}_i = RT_i\mathbf{f}_i + \mathbf{n}_i, i = 1, \ldots, m$.
- All $\{\mathbf{f}_i\}_{i=1}^m$ defined on the same HR grid.
- Let $C_i\mathbf{f}_i$ constrain $\mathbf{f}_i$ onto the overlap part.



28

Variational Formulation

- ❑ $[C_1\mathbf{f}_1, \ldots, C_m\mathbf{f}_m]$ should have low rank.
- ❑ In ideal case, rank = 1.
- ❑ $\min_{\mathbf{f}_i} \operatorname{rank}[C_1\mathbf{f}_1, \ldots, C_m\mathbf{f}_m] + \frac{\lambda}{2}\sum_{i=1}^{m} \|RT_i\mathbf{f}_i - \mathbf{q}_i\|_2^2.$
- ❑ Convexifying it by nuclear norm:

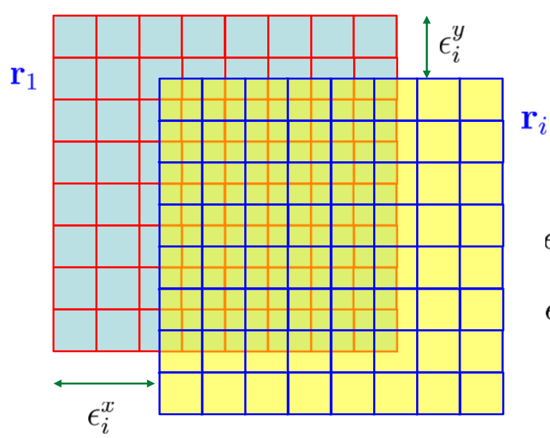
$$\min_{\mathbf{f}_i} \|[C_1\mathbf{f}_1, \ldots, C_m\mathbf{f}_m]\|_* + \frac{\lambda}{2}\sum_{i=1}^{m} \|RT_i\mathbf{f}_i - \mathbf{q}_i\|_2^2.$$

- ❑ Nuclear norm: $\|U\|_* = \sum_j \sigma_j(U) = \|\boldsymbol{\sigma}(U)\|_1$.

29

Motion Matrices in 2D

General 2D motion matrices given by tensor product.



$$\epsilon_i^x = \lfloor \epsilon_i^x \rfloor + e_i^x := \ell_i^x + e_i^x$$

$$\epsilon_i^y = \lfloor \epsilon_i^y \rfloor + e_i^y := \ell_i^y + e_i^y$$

$$M_i = T(e_i^x)S^{\ell_i^x} \otimes T(e_i^y)S^{\ell_i^y}$$

30

Creating Low Rank Structure

$$\begin{aligned} M_i &= T(e_i^x)S^{\ell_i^x} \otimes T(e_i^y)S^{\ell_i^y} \\ &= (T(e_i^x) \otimes T(e_i^y))(S^{\ell_i^x} \otimes S^{\ell_i^y}) \\ &:= T(e_i)S^{\ell_i} \end{aligned}$$

- Both $T(e_i)$ and S^{ℓ_i} are BTTB matrices.
- Then $\mathbf{r}_i = M_i\mathbf{r}_1 = T(e_i)S^{\ell_i}\mathbf{r}_1 := T_i\mathbf{f}_i$. $S^{\ell_i}\mathbf{r}_1 = \mathbf{f}_i$
- $\mathbf{q}_i = RT_i\mathbf{f}_i + \mathbf{n}_i, i = 1, \ldots, m$.
- Let $C_i\mathbf{f}_i$ constrain $\mathbf{f}_i$ onto the overlap part.
- $[C_1\mathbf{f}_1, \ldots, C_m\mathbf{f}_m]$ should have low rank.

Low-rank Super-resolution Model

$$\min_{\mathbf{f}_i} \|[C_1\mathbf{f}_1, \ldots, C_m\mathbf{f}_m]\|_* + \frac{\lambda}{2}\sum_{i=1}^{m} \|RT_i\mathbf{f}_i - \mathbf{q}_i\|_2^2.$$

- ℓ_1-ℓ_2 problem conveniently solved by ADMM (SVD thresholding) algorithm.
- In each ADMM iteration:
  - Computing SVD of an n-by-m matrix: $O(nm^2)$.
  - Solving m linear systems of size n-by-n.
  - Nice structure in C_i, R and T_i to be exploited.

32

Nuclear Norm [Candes, Recht, 09; Recht, Fazel, Parrilo, 10]

$$\min_{\mathbf{f}_i} \|[R_1\mathbf{f}_1, \ldots, R_m\mathbf{f}_m]\|_* + \frac{\lambda}{2}\sum_{i=1}^{m} \|DT_i\mathbf{f}_i - \mathbf{q}_i\|_2^2.$$

- Nuclear norm: $\|U\|_* = \sum_j \sigma_j(U) = \|\boldsymbol{\sigma}(U)\|_1$.
- An ℓ^1-ℓ^2 model. Can be solved by ADMM:
  - Auxiliary variables: $\mathbf{v}_i = R_i\mathbf{f}_i$.
  - $\mathbf{f}_i$-subproblem: $(\beta R_i^t R_i + \lambda T_i^t D^t D T_i)\, \mathbf{f}_i^{j+1} = \mathbf{b}^j$.
  - $\mathbf{v}_i$-subproblem: $[\mathbf{v}_1{}^{j+1}, \ldots, \mathbf{v}_m{}^{j+1}] = \mathrm{SVT}_{\frac{1}{\beta}}(U^j)$.

33

Motion Estimation

- Local motion estimation:
  - Optical flow: track motion for each pixel in each frame
  - Pull images $\mathbf{q}_j$, $j = 1, \cdots, m$, pixel by pixel back to same grid as $\mathbf{q}_{\text{ref}}$ to get $\tilde{\mathbf{q}}_j$.
- Global motion estimation:
  - Affine motion between $\mathbf{q}_{\text{ref}}$ and $\tilde{\mathbf{q}}_j$ estimated by least-squares.

34

Local Motion Estimation [C. Gilliam and T. Blu, 2015]

- *All-pass optical flow:* motion = all-pass filter
- Given filters $\{\phi_j\}_{j=0}^{N}$ and $\phi := \phi_0 + \sum_{j=1}^{N-1} c_j \phi_j$, find

$$\min_{\{c_i\}} \sum_{k,l\in W} |\phi(k,l)\mathbf{q}_j(x-k,y-l)-\phi(-k,-l)\mathbf{q}_{\text{ref}}(x-k,y-l)|^2.$$

- The optical flow of $\mathbf{q}_j$ at (x,y) is

$$F_j(x,y) = \left(\frac{2\sum_{k,l} k\phi(k,l)}{\sum_{k,l}\phi(k,l)}, \frac{2\sum_{k,l} l\phi(k,l)}{\sum_{k,l}\phi(k,l)} \right).$$

- Pull back image $\tilde{\mathbf{q}}_j(x,y) = \mathbf{q}_j(F_j(x,y))$.

35

Global Motion Estimation

- There may still be a small perturbation between $\tilde{\mathbf{q}}_j$ and the reference frame $\mathbf{q}_{\text{ref}}$.
- Assume small perturbation is affine:

$$\tilde{\mathbf{q}}_j = \begin{bmatrix} \rho_{0,j} & \rho_{2,j} \\ \rho_{1,j} & \rho_{3,j} \end{bmatrix} \mathbf{q}_{\text{ref}} + \begin{bmatrix} \rho_{5,j} \\ \rho_{6,j} \end{bmatrix}, \; j = 1,2,\ldots,m.$$

- $\{\rho_{i,j}\}_{i=1}^{4}$ give the rotation, and $\rho_{5,j}$ and $\rho_{6,j}$ give the displacements.
- $\{\rho_{i,j}\}_{i=1}^{6}$ are obtained by least-squares.

[C., Shen and Xia, ACHA, 2007]

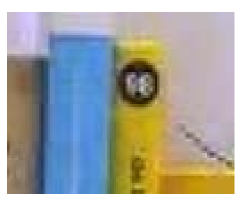
36

Outline

1. **High-resolution Image Reconstruction**
2. **Video Still Enhancement Models**
   - o **Classical**
   - o **Tight-frame**
   - o **Low-rank**
3. **Experiments**

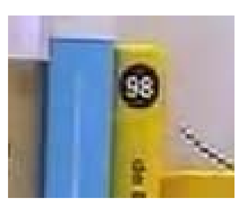
37

Bookshelf Video Upsampled by 2

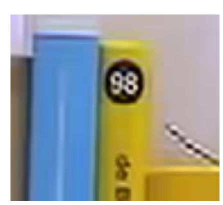


Single frame with bilinear interpolation

21 frames with TV regularization



21 frames with tightframe



21 frames with nuclear norm

38

Disk Video Upsampled by 2

19 frames
of size
57×49

LR video

MAP (2015)

tightframe (2007)

reference frame

sparse direct'l ('10)

nuclear

Source: https://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html

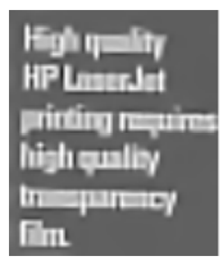
39

Text Video Upsampled by 2

21 frames
of size
57×49

LR video

MAP



tightframe

reference frame

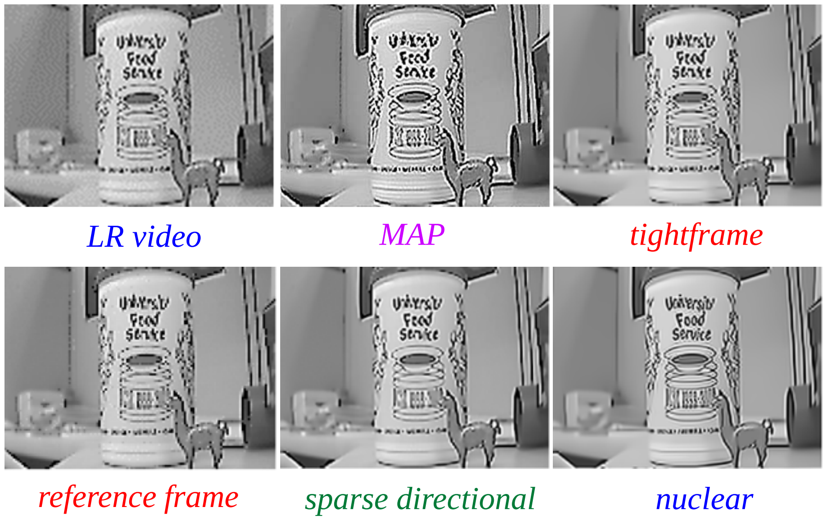
sparse directional

nuclear

Source: https://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html

40

Alpaca Video Upsampled by 2

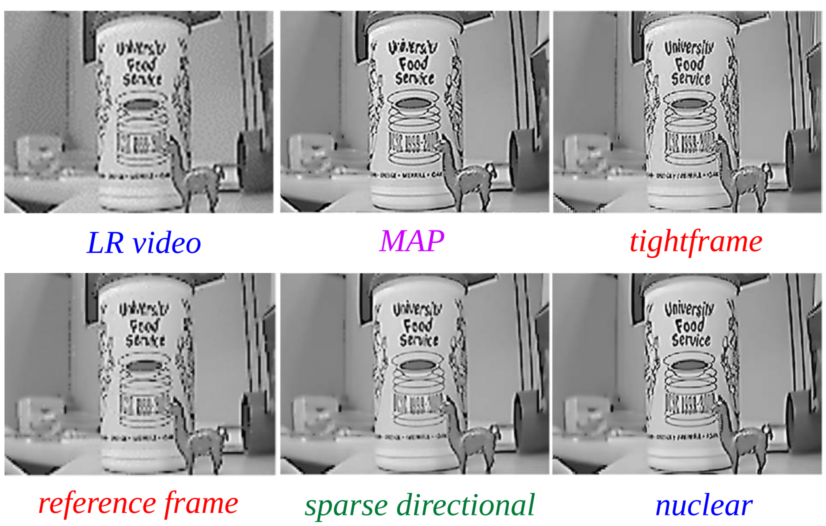


21 frames of size 96×128

Source: https://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html

41

Alpaca Video Upsampled by 4



21 frames of size 96×128

Source: https://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html

42

Alpaca Image Upsampled by 4

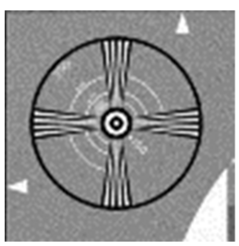
reference image

nuclear

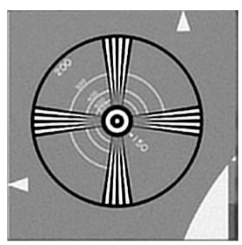
43

EIA Video Upsampled by 4

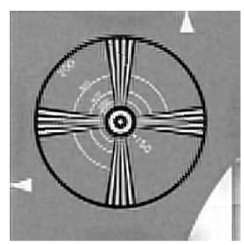
19 frames of size 57×49



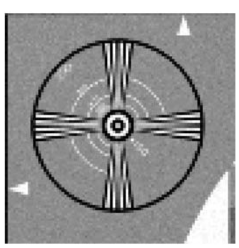
LR video



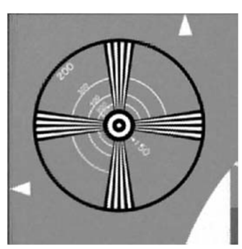
MAP



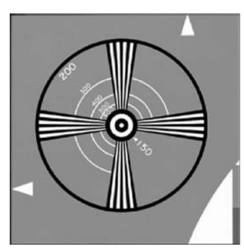
tightframe



reference frame



sparse directional

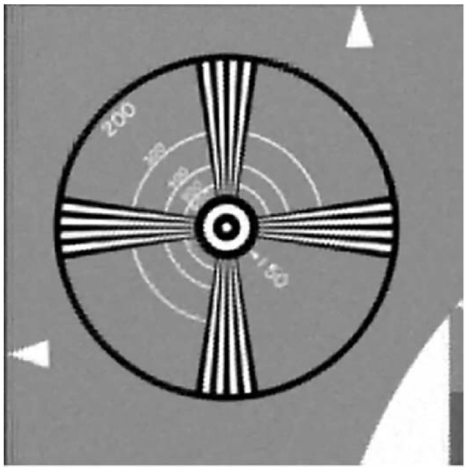


nuclear

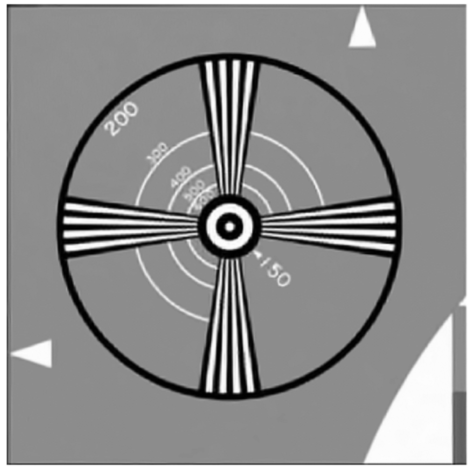
Source: https://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html

44

EIA Image Upsampled by 4



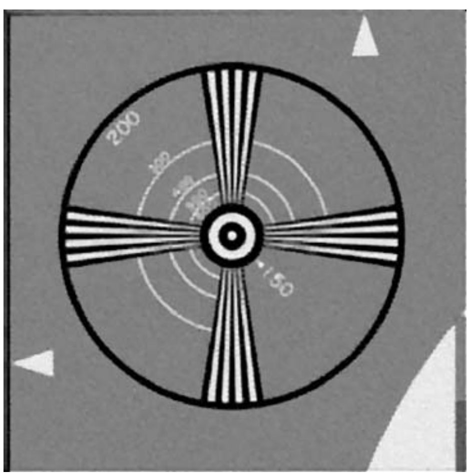
sparse directional



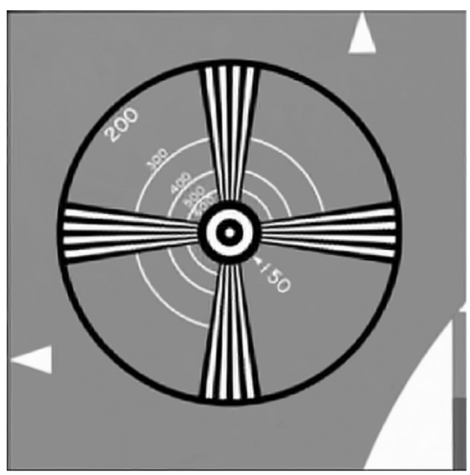
nuclear

45

EIA Image Upsampled by 4



L1-robust, Farsiu et al. (2004)
10 x 10 (.jpg)



nuclear

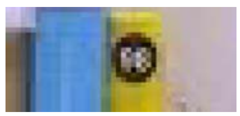
46

Book-shelf Upsampled by 2

LR video

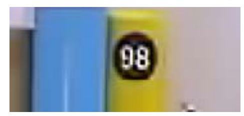
MAP

tightframe



reference frame

sparse directional



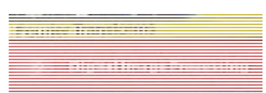
nuclear

47

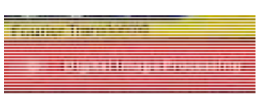
Bookshelf Video Upsampled by 2

LR video

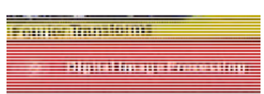
MAP



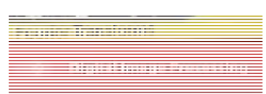
tightframe



reference frame



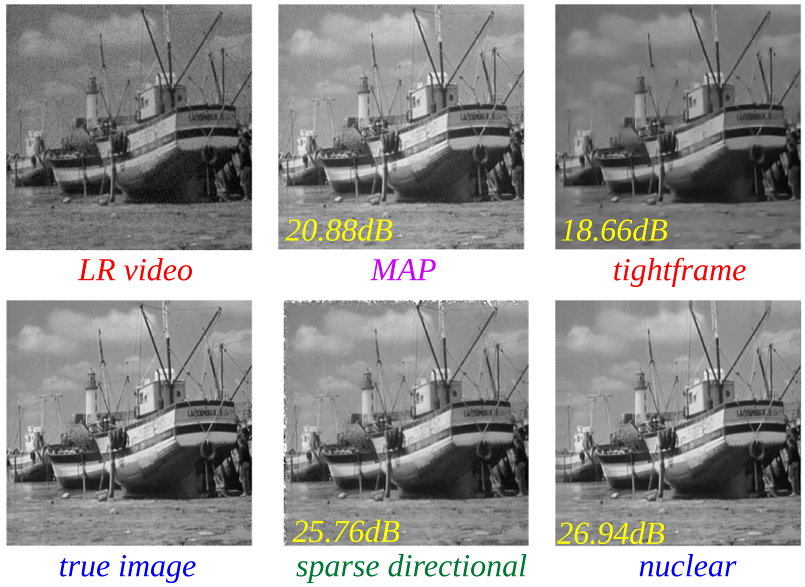
sparse directional



nuclear

48

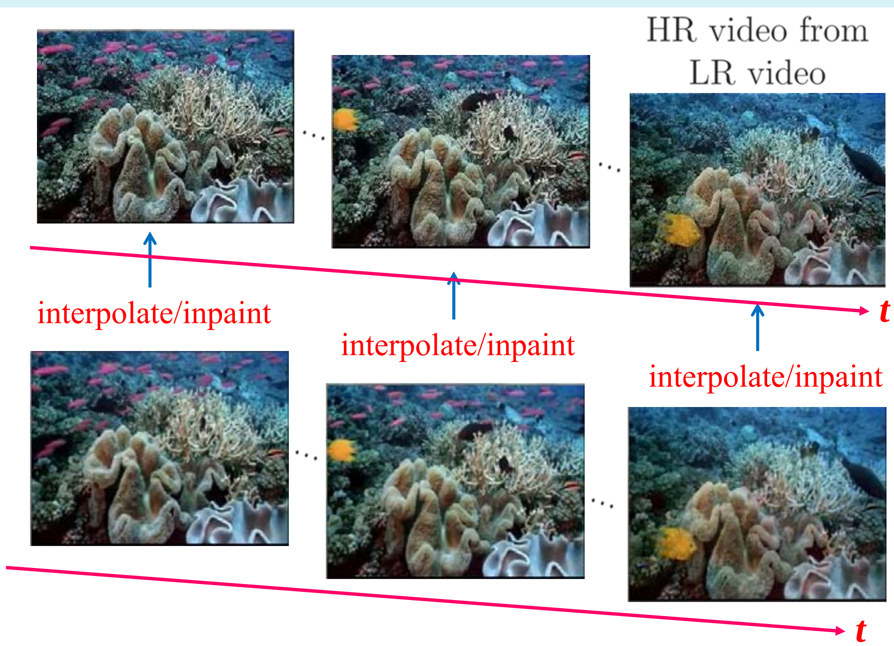
Boat Video Upsampled by 2



17 frames of size 256 $\times$ 256 with change of lighting and noise

49

Single Frame Upsampling

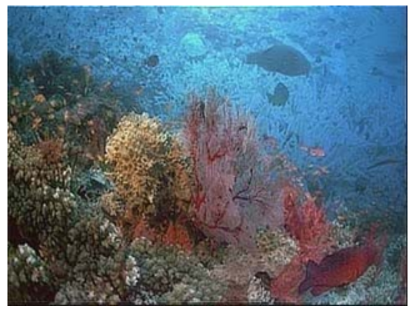


50


Input Video


Qian *et al.* Siggraph 09: $\|\Psi(\nabla \mathbf{f})\|_1$


Upsampled by bicubic


Level 6 Tightframe

51


Input Video


Qian *et al.* Siggraph 09


Upsampled by bicubic


Level 6 Tightframe

No thresholding = completely parameter free

52

Concluding Remarks

- New approach for video still enhancement
- Decompose motion matrix into product of shift matrix (for integral shift) and Toeplitz matrix (for fractional shift)
- Enforce low rank structure using the integral shift matrix
- Resulting convexified functional can be solved efficiently via ADMM

R. Zhao and R. Chan, arXiv:1704.06196