Image Super-Resolution with Sparse Coding

IMAGE REPRESENTAION AND REGISTRATION

Noboru Murata

June 21, 2023

https://noboru-murata.github.io/

Introduction

- background
- single-frame vs multi-frame

Problem Formulation

- sparse representation
- single-frame super-resolution
- multi-frame super-resolution

Illustrative Example

- block matching
- double sparsity

Conclusion

Introduction

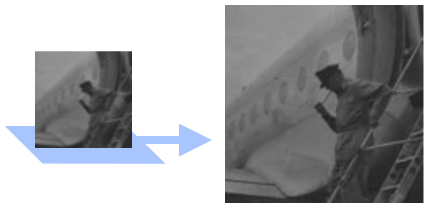
request for image processing:

- development of new devices
- re-mastering of histrical images

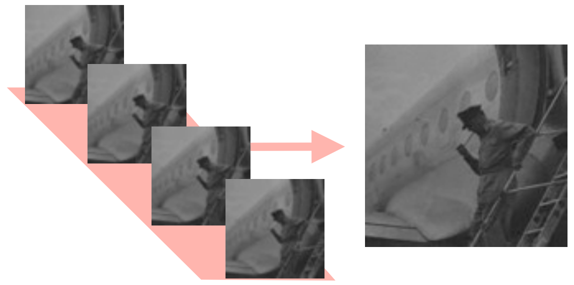
(a) LR frame.

(b) HR image.

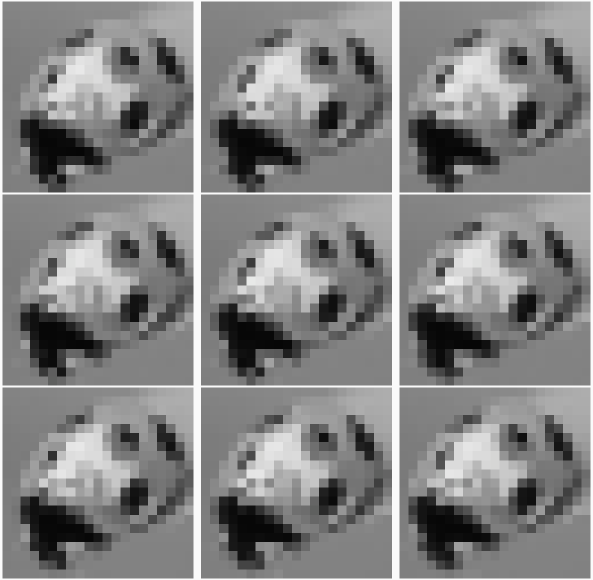
- problem:
  - restore a clear image from low-resolution images
  - consider degradation caused by
    - noise
    - blur
    - down-sampling
- typical setup:
  - single-frame: one low-resolution image
  - multi-frame: multiple low-resolution images with different degradation processes
- typical approaches:
  - model-based: e.g. random Markov field
  - example-based: e.g. sparse representation



single-frame super-resolution



multi-frame super-resolution



$\Longrightarrow$

Problem Formulation

- notation:
  - dictionary: $\mathbf{D} = (d_1, d_2, \ldots, d_k) \in \mathbb{R}^{p \times k}$
  - observation: $\boldsymbol{y} \in R^p$
  - coefficients: $\boldsymbol{\alpha} \in \mathbb{R}^k$

optimization problem

estimate appropriate α and $\mathbf{D}$:

$$\underset{\boldsymbol{\alpha}, \mathbf{D}}{\text{minimize}} \; \|\boldsymbol{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \eta \|\boldsymbol{\alpha}\|_1$$

- notation:
  - X: high-resolution image
  - Y: low-resolution image
  - model of degradation process:

    $$Y = \mathbb{L}X + \varepsilon = \mathbb{S}\mathbb{H}\mathbb{W}X + \varepsilon$$

    where degradation $\mathbb{L}$ is decomposed as
    - $\mathbb{S}$: down-sampling
    - $\mathbb{H}$: blurring
    - $\mathbb{W}$: warping

    and ε is additive noise

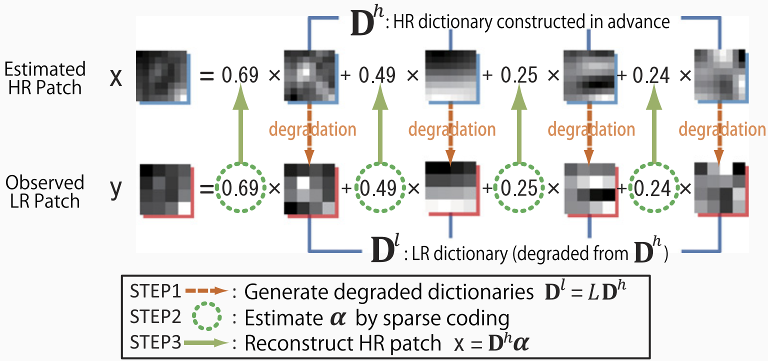
- hypothesis:
  - single observation: $Y = \mathbb{L} X + \varepsilon$
  - x: patch of X
    y: corresponding patch of Y
  - sparse representation with identical coefficients:

$$x = \mathbf{D}^h \boldsymbol{\alpha} \quad (\mathbf{D}^h : \text{high-resolution dictionary})$$
$$y = \mathbf{D}^l \boldsymbol{\alpha} \quad (\mathbf{D}^l : \text{low-resolution dictionary})$$
$$\simeq \mathbb{L} x = \mathbb{L} \mathbf{D}^h \boldsymbol{\alpha} \quad (\text{linearity assumption})$$

problem

estimate of $\boldsymbol{\alpha}$ from a low-resolution image:

$$\underset{\alpha}{\text{minimize}} \, \|y - \mathbf{D}^l \boldsymbol{\alpha}\|_2^2 + \eta \|\boldsymbol{\alpha}\|_1$$



key issue

construct good $\mathbf{D}^l$ from $\mathbf{D}^h$

- hypothesis:
  - multiple observations: $Y_1, \ldots, Y_N$

$$Y_k = \mathbb{L}_k X + \varepsilon_k, \quad k = 1, \ldots, N$$

  - $\boldsymbol{x}$: patch of X
    $\boldsymbol{y}_k$: corresponding patches of Y
  - sparse representation:

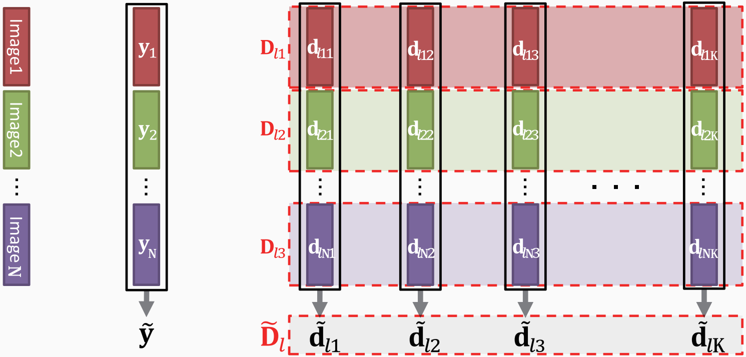
$$\begin{aligned} \boldsymbol{x} &= \mathbf{D}^h \boldsymbol{\alpha} \\ \boldsymbol{y}_k &= \mathbf{D}_k^l \boldsymbol{\alpha} \simeq \mathbb{L}_k \boldsymbol{x} = \mathbb{L}_k \mathbf{D}^h \boldsymbol{\alpha} \end{aligned}$$

- problem:
  - estimate of $\boldsymbol{\alpha}$ from multiple low-resolution images:

$$\underset{\alpha}{\text{minimize}} \, \|\tilde{\mathbf{y}} - \tilde{\mathbf{D}}^l \boldsymbol{\alpha}\|_2^2 + \eta \|\boldsymbol{\alpha}\|_1$$

where $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{y}}$ are stacked as

$$\tilde{\mathbf{D}}^l = \begin{bmatrix} \mathbf{D}_1^l \\ \vdots \\ \mathbf{D}_N^l \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}$$



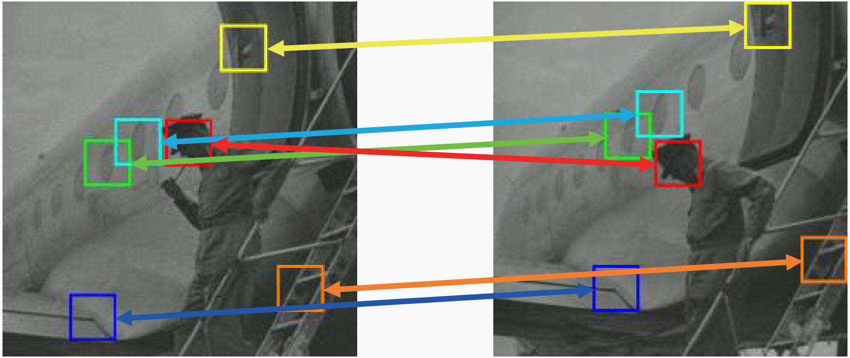
key issue

appropriately align multiple low-resolution images and dictionaries

- simple approach: (Kato, Hino, and Murata 2015)
  - block matching with reference patch

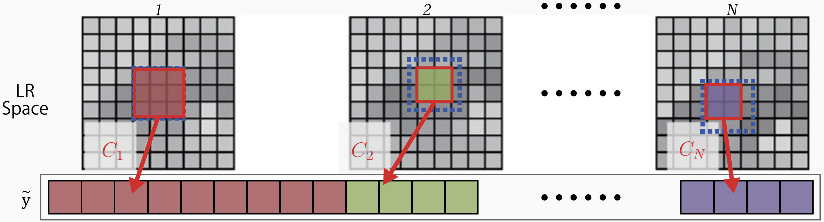
$$\text{estimate } \hat{\mathbb{L}}_k = \mathbb{S}\mathbb{H}\hat{\mathbb{W}}_k \text{ s.t. } \boldsymbol{y}_k = \hat{\mathbb{L}}_k \boldsymbol{x}$$

  - sub-pixel accuracy method (Shimizu and Okutomi 2006)



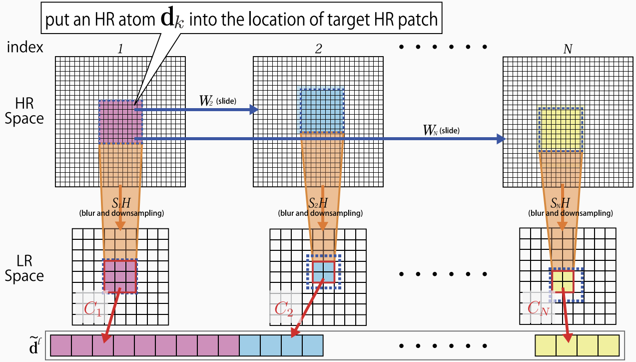
- stacked observation:

$$\tilde{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \quad \text{where} \quad y_k = \hat{\mathbb{L}}_k x = \mathbb{SH}\hat{\mathbb{W}}_k x$$



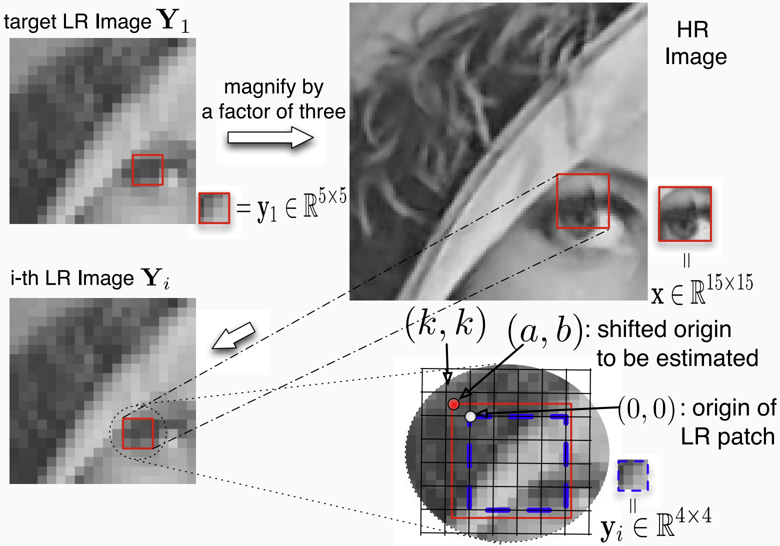
- stacked dictionary:

$$\tilde{\mathbf{D}}^l = \begin{bmatrix} \mathbf{D}_1^l \\ \vdots \\ \mathbf{D}_N^l \end{bmatrix} \quad \text{where} \quad \mathbf{D}_k^l = \hat{\mathbb{L}}_k \mathbf{D}^h = \mathbb{S}\mathbb{H}\hat{\mathbb{W}}_k \mathbf{D}^h$$



- sparse representation approach: (Kato, Hino, and Murata 2017)
    - prepare “meta-dictionary”
    - construct a dictionary
        - which is sparse combination of meta-dictionary
        - which offers sparse representation of target observations

      (Rubinstein, Zibulevsky, and Elad 2010)


target LR Image $\mathbf{Y}_1$
magnify by
a factor of three
HR
Image
$= \mathbf{y}_1 \in \mathbb{R}^{5\times 5}$
$\mathbf{x} \in \mathbb{R}^{15\times 15}$
i-th LR Image $\mathbf{Y}_i$
(k, k)
(a, b): shifted origin
to be estimated
$(0, 0)$: origin of
LR patch
$\mathbf{y}_i \in \mathbb{R}^{4\times 4}$

- notation:
  - dictionary for observation $\boldsymbol{y}_i$:

$$\mathbf{D}_i^l = \theta_{i,(0,0)}\mathbf{D}^{l(0,0)} + \theta_{i,(0,1)}\mathbf{D}^{l(0,1)} + \cdots + \theta_{i,(k,k)}\mathbf{D}^{l(k,k)}$$

  - meta-dictionary matrix:

$$\mathbf{B} = \begin{bmatrix} \mathbf{D}_1^l & & & & & & \\ & \mathbf{D}^{l(0,0)} & \cdots & \mathbf{D}^{l(k,k)} & & & \\ & & & & \ddots & & \\ & & & & \mathbf{D}^{l(0,0)} & \cdots & \mathbf{D}^{l(k,k)} \end{bmatrix}$$

  - meta-dictionary coefficient:

$$\boldsymbol{\theta} = \left[1, \theta_{2,(0,0)}, \ldots, \theta_{2,(k,k)}, \ldots, \theta_{N,(0,0)}, \ldots, \theta_{N,(k,k)}\right]^{\mathrm{T}}$$

- objective:

$$\underset{\boldsymbol{\alpha},\boldsymbol{\theta}}{\text{minimize}} \|\tilde{\boldsymbol{y}} - \mathbf{B}\,\mathrm{vec}\left(\boldsymbol{\alpha}\boldsymbol{\theta}^{\mathrm{T}}\right)\|_2^2 + \eta\|\boldsymbol{\alpha}\|_1$$

$$\text{subject to } \mathbf{E}\boldsymbol{\theta} \leq \mathbf{1},\ \boldsymbol{\theta} \geq \mathbf{0},\ \theta_1 = 1$$

where

$$\mathbf{E} = \begin{bmatrix} 1 & & & & & & \\ & 1 & \cdots & 1 & & & \\ & & & & \ddots & & \\ & & & & & 1 & \cdots & 1 \end{bmatrix}$$

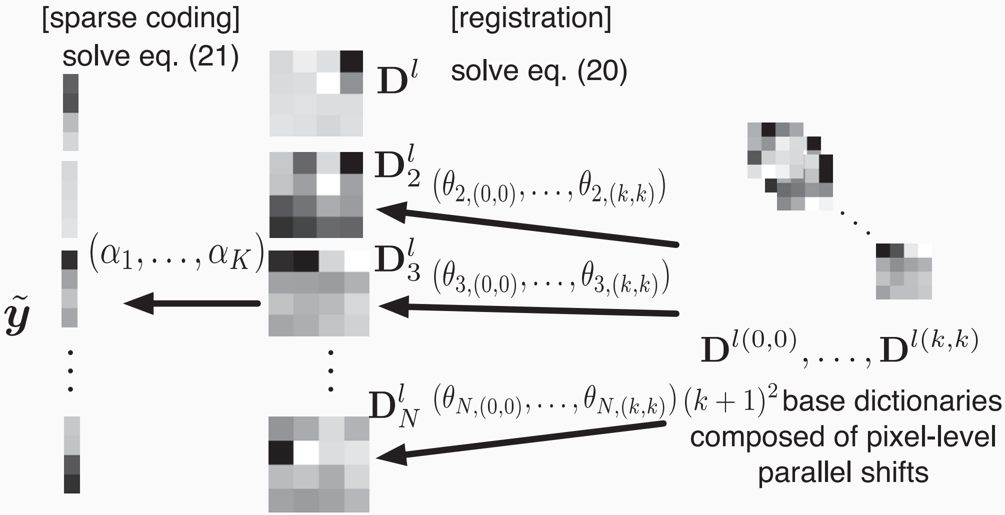
- optimization for registration:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\tilde{\mathbf{y}} - \mathbf{B}(\mathbf{I} \otimes \boldsymbol{\alpha})\boldsymbol{\theta}\|_2^2$$

subject to $\mathbf{E}\boldsymbol{\theta} \leq \mathbf{1},\ \boldsymbol{\theta} \geq \mathbf{0}$

- optimization for sparse representation:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\tilde{\mathbf{y}} - \mathbf{B}(\boldsymbol{\theta} \otimes \mathbf{I})\boldsymbol{\alpha}\|_2^2 + \eta\|\boldsymbol{\alpha}\|_1 \tag{21}$$


[sparse coding]
solve eq. (21)
[registration]
solve eq. (20)
$\mathbf{D}^l$
$\mathbf{D}^l_2$ $(\theta_{2,(0,0)}, \ldots, \theta_{2,(k,k)})$
$\mathbf{D}^l_3$ $(\theta_{3,(0,0)}, \ldots, \theta_{3,(k,k)})$
$\mathbf{D}^l_N$ $(\theta_{N,(0,0)}, \ldots, \theta_{N,(k,k)})$
$(\alpha_1, \ldots, \alpha_K)$
$\tilde{\boldsymbol{y}}$
$\mathbf{D}^{l(0,0)}, \ldots, \mathbf{D}^{l(k,k)}$
$(k+1)^2$ base dictionaries
composed of pixel-level
parallel shifts

Illustrative Example

comparison with existing works:

- ASDS (Dong et al. 2011): single-frame, sparse representation
- MF-JDL (**Wang_etal2011**): multi-frame, sparse representation
- BTV (Farsiu et al. 2004): multi-frame, model-based
- LABTV (Li et al. 2010): multi-frame, model-based
- Proposed (Kato, Hino, and Murata 2015): multi-frame, sparse representation, block-matching


a) ASDS.


(b) MF-JDL.


(c) BTV.


d) LABTV.


(e) Proposed.


(f) Original HR image.


a) ASDS.


(b) MF-JDL.


(c) BTV.

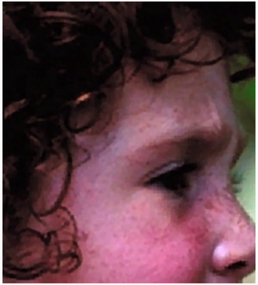

d) LABTV.


(e) Proposed.


(f) Original HR image.

(a) ASDS.

(b) MF-JDL.



(c) BTV.

(d) LABTV.

(e) Proposed.

(f) Original HR image.


(a) ASDS.


(b) MF-JDL.


(c) BTV.


(d) LABTV.


(e) Proposed.


(f) Original HR image.

(a) ASDS.

(b) MF-JDL.

(c) BTV.

(d) LABTV.

(e) Proposed.

(f) Original HR image.

comparison with existing works:

- ASDS (Dong et al. 2011): single-frame, sparse representation
- MF-JDL (**Wang_etal2011**): multi-frame, sparse representation
- BTV (Farsiu et al. 2004): multi-frame, model-based
- LABTV (Li et al. 2010): multi-frame, model-based
- MF-SC (Kato, Hino, and Murata 2015): multi-frame, sparse representation, block-matching
- Proposed (Kato, Hino, and Murata 2017): multi-frame, sparse representation, double sparsity

(a) Observed LR image

(b) Original HR image

(c) MF-SC

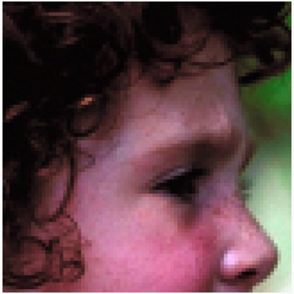
(d) Proposed


(a) Observed LR image

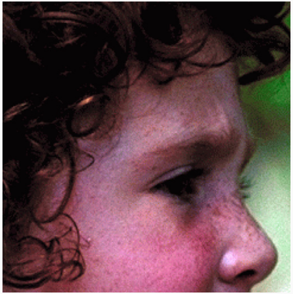

(b) Original HR image


(c) MF-SC


(d) Proposed



(a) Observed LR image



(b) Original HR image

(c) MF-SC

(d) Proposed

(a) Observed LR image

(b) Original HR image

(c) MF-SC

(d) Proposed

(a) Observed LR image

(b) Original HR image

(c) MF-SC

(d) Proposed

	SF-JDL	ASDS	MF-JDL	BTV	LABTV	MF-SC	Proposed
MacArthur	34.33	35.63	35.18	34.39	34.40	34.79	35.63
	(2.69)	(178.08)	(133.78)	(61.72)	(96.17)	(27.70)	(61.74)
Samurai	25.97	26.66	26.12	26.16	26.07	25.90	26.49
	(2.50)	(211.65)	(138.38)	(62.13)	(96.24)	(30.75)	(59.86)

$$\mathrm{PSNR[dB]} = 10\log_{10}\frac{255^2}{\mathrm{MSE}}$$

computational times[sec] (in parentheses)

Conclusion

we have investigated

- multi-frame super resolution method based on sparse representation
- registration performance of sub-pixel block matching and double sparsity

practical applications would be

- old or historic movies
- medical images

which consist of a number of low-resolution images

- Dong, Weisheng et al. (Jan. 28, 2011). "Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization." In: *IEEE Transactions on Image Processing* 20 (7), pp. 1838–1857. DOI: 10.1109/TIP.2011.2108306. PMID: 21278019.
- Farsiu, Sina et al. (Oct. 2004). "Fast and robust multiframe super resolution." In: *IEEE Transactions on Image Processing* 13 (10), pp. 1327–1344. DOI: 10.1109/TIP.2004.834669. PMID: 15462143.
- Kato, Toshiyuki, Hideitsu Hino, and Noboru Murata (Mar. 9, 2015). "Multi-frame image super resolution based on sparse coding." In: *Neural Networks* 66, pp. 64–78. DOI: 10.1016/j.neunet.2015.02.009.
- — (May 31, 2017). "Double sparsity for multi-frame super resolution." In: *Neurocomputing* 240, pp. 115–126. DOI: 10.1016/j.neucom.2017.02.043.
- Li, Xuelong et al. (Feb. 2010). "A multi-frame image super-resolution method." In: *Signal Processing* 90 (2), pp. 405–414. DOI: 10.1016/j.sigpro.2009.05.028.

Rubinstein, Ron, Michael Zibulevsky, and Michael Elad (Mar. 2010). **“Double Sparsity: Learning Sparse Dictionaries for Sparse Signal Approximation.”** In: *IEEE Transactions on Signal Processing* 58 (3), pp. 1553–1564. DOI: 10.1109/TSP.2009.2036477.

Shimizu, Masao and Masatoshi Okutomi (May 2006). **“Multi-Parameter Simultaneous Estimation on Area-Based Matching.”** In: *International Journal of Computer Vision* 67 (3), pp. 327–342. DOI: 10.1007/s11263-006-5632-3.