Imperial College London

Single-Image Super-Resolution: Coupled-dictionary learning vs model-based processing

Pier Luigi Dragotti
Imperial College London

September 4, 2017

Special thanks to J. Huang and X. Deng for helping with the preparation of the slides

Outline

- This talk is about
  - Image processing applications that are useful to compare data-driven signal processing vs model-based signal processing
  - Shallow learning but not deep-learning
  - Overview of key notions used in sparsity-driven signal processing methods

Problem Statement

Real Scene Digital Image

- A visual scene is turned into a **digital** image by a camera
- Can we overcome the limitation of the camera and, given the pixels, obtain a sharper image with increased resolution?
- The problem of enhancing the resolution of a single image is known as Single-Image Super-Resolution

Problem Statement

- Single Image Super-Resolution differs from traditional Image Super-Resolution because only one low-resolution image is available

- Highly Ill-posed

- Two main approaches:
    - *Data-driven Approach*: learn the high-resolution image from a database of low-resolution (LR) and high-resolution (HR) pairs
    - *Model-Based Approach*: use priors on the property of natural images to estimate the HR image from the observed LR one.

Data-Driven SISR

- The **key insight** in the data-driven approach is that images or patches of images have a sparse representation in a (redundant) dictionary
- Example of dictionaries include wavelets, union of bases or *learned* dictionaries.

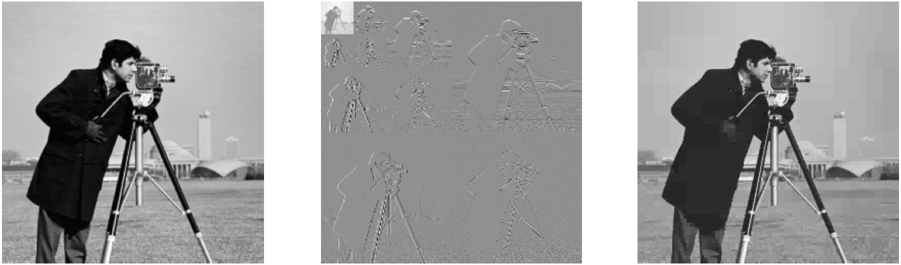
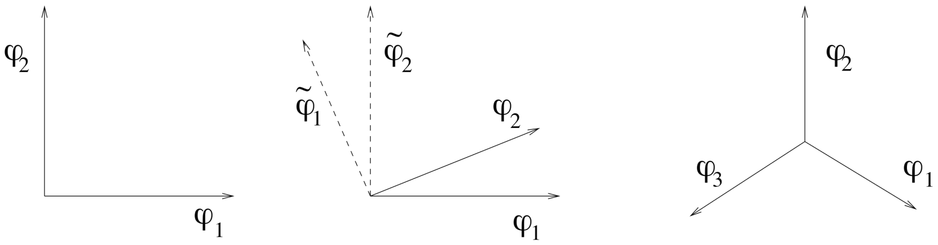


Figure: Cameraman is reconstructed using only 8% of the wavelet coefficients

Bases and Overcomplete Dictionaries



a) Orthogonal Basis b) Biorthogonal Basis c) Frame

Bases and Overcomplete Dictionaries: Matrix Interpretation

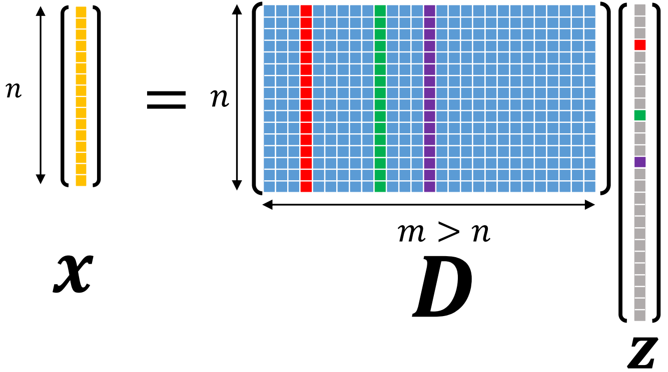
- Assume the 'atoms' $\{\varphi_i\}$ are finite dimensional column vectors of size N
- Stack them one next to the other to form the synthesis matrix M:

$$M = \begin{bmatrix} \uparrow & \cdots & \uparrow & \cdots \\ \varphi_1 & \cdots & \varphi_i & \cdots \\ \downarrow & \cdots & \downarrow & \cdots \end{bmatrix}$$

- If M is square and invertible then $\{\varphi_i\}_{i=1}^{N}$ is a basis (of R^N or C^N).
- If M is 'fat' but has N linearly independent columns it form a redundant or overcomplete dictionary

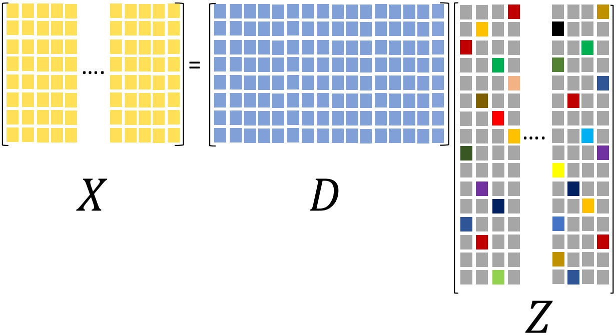
Data-Driven Super Resolution

- The **key insight** in the data-driven approach is that images or patches of images have a sparse representation in a redundant dictionary
- The dictionary is usually learned



Dictionary Learning

- Learn dictionaries by alternating between
  - Learning the sparse representations given the dictionaries (sparse coding step)
  - Update the dictionary given the sparse representations (dictionary update step)

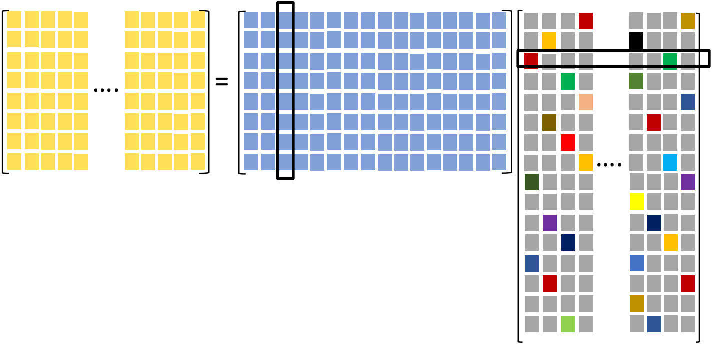


Sparse Coding Step

- Given $\boldsymbol{D}$, learn the sparse representations $\boldsymbol{z_i}$
- Sparse Representation Algorithms:
    - Greedy algorithms:
        - Matching Pursuit (MP)
        - Orthogonal Matching Pursuit (OMP)
        - ...
    - Convex Relaxation Algorithms:
        - Basis Pursuit (BP)
        - ....

Dictionary Update Step

- Given the sparse representations update the dictionary.
  - Many possible approaches, k-SVD (Aharon-Elad:06) is the most used



- Find $\boldsymbol{d}_i$ and $\boldsymbol{z}_i^T$ that minimize $\left\|\boldsymbol{E_i} - \boldsymbol{d}_i\boldsymbol{z}_i^T\right\|$
- This is achieved by taking the SDV of $\boldsymbol{E_i}$ (with a small caveat to keep $\boldsymbol{z}_i^T$ sparse)

Data-Driven Super-Resolution

Super-Resolution Model:

- One postulates that HR patches and LR patches admit a common sparse representation z_i :

$$\mathbf{x}_i^{LR} = \mathbf{D}^{LR}\mathbf{z}_i$$

$$\mathbf{x}_i^{HR} = \mathbf{D}^{HR}\mathbf{z}_i$$

$$\mathbf{x}_i^{LR} = \mathbf{A}\mathbf{x}_i^{HR} = \mathbf{A}\mathbf{D}^{\mathbf{HR}}\mathbf{z}_i = \mathbf{D}^{\mathbf{LR}}\mathbf{z}_i$$

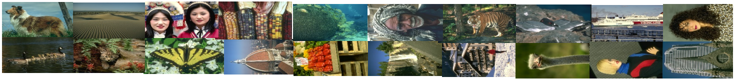
Data-Driven Super-Resolution: Algorithm

Algorithm:

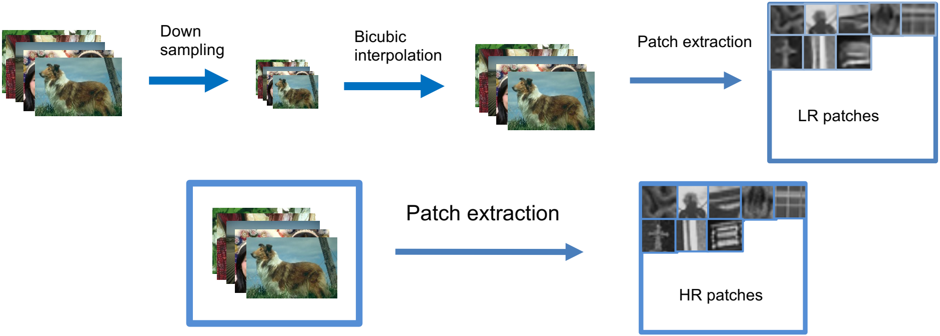
- Given $\boldsymbol{D}^{LR}$ and $\boldsymbol{D}^{HR}$ and the LR image to enhance
- Patches $\boldsymbol{x}_i^{LR}$ are extracted
- Using $\boldsymbol{D}^{LR}$ and sparse coding methods (e.g., OMP) the sparse vectors $\boldsymbol{z}_i$ are retrieved
- HR patches are then given by $\boldsymbol{x}_i^{HR} = \boldsymbol{D}^{HR}\boldsymbol{z}_i$
- Open question: How do we learn $\boldsymbol{D}^{LR}$ and $\boldsymbol{D}^{HR}$?

Data-Driven Super-Resolution: Training

Start with an external dataset of images (e.g., BSD 300 dataset)



Extract pairs of LR and HR patches



Data-Driven Super-Resolution: Training

Training:

1. Given $\boldsymbol{x}_i^{LR}$, learn $\boldsymbol{D}^{LR}$ and $\boldsymbol{z}_i$ using K-SVD

2. Given $\boldsymbol{x}_i^{HR}$ and $\boldsymbol{z}_i$ compute $\boldsymbol{D}^{HR}$ directly

Data-Driven Super-Resolution: Example

Algorithm:

- Given $\boldsymbol{D}^{LR}$ and $\boldsymbol{D}^{HR}$ and the LR image to enhance
- Patches $\boldsymbol{x}_i^{LR}$ are extracted
- Using $\boldsymbol{D}^{LR}$ and sparse coding methods the sparse vectors $\boldsymbol{z}_i$ are retrieved
- HR patches are then given by $\boldsymbol{x}_i^{HR} = \boldsymbol{D}^{HR}\boldsymbol{z}_i$

Low-resolution Image

High-resolution Image

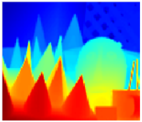
Key references include [Yang et al. 2010,2012], [Zeyde et al. 2010], [Timofte et al. 2014]

Data-Driven Super-Resolution

- Preliminary Observations:
    - Based on a (simple) modelling assumption: *sparsity*
    - Most of the complexity is shifted to the training stage

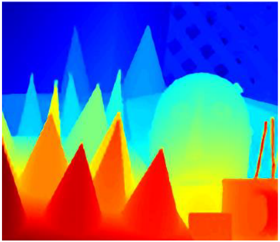
- Given more complex datasets, how much more modelling is required to achieve good performance in the application at hand?

Multimodal Depth Image Super-Resolution



LR Depth Image

HR Color Image



Estimated HR Depth Image

Depth Image Super-Resolution Model

Model:

- HR depth patches and LR depth patches admit a common sparse representation
- Colour images are made of texture + piecewise smooth 2-D signals
- The texture is **unique** but the piecewise smooth signal is in **common** with the depth image
- So HR depth, LR depth and HR colour have something in common but also unique features

Depth Image Super-Resolution Model

Model:

- HR depth, LR depth and HR colour have something in common but also unique features
- This requires a more sophisticated dictionary model
- We split the dictionary into two parts, one describes the common features and the other the features unique to each modality

$$\begin{bmatrix} \mathbf{X}^l \\ \mathbf{X}^h \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{\Psi}_c^l & \mathbf{\Psi}^l & \mathbf{0} \\ \mathbf{\Psi}_c^h & \mathbf{\Psi}^h & \mathbf{0} \\ \mathbf{\Phi}_c & \mathbf{0} & \mathbf{\Phi} \end{bmatrix} \begin{bmatrix} \mathbf{Z} \\ \mathbf{U} \\ \mathbf{V} \end{bmatrix}$$

Multi-Modal Dictionary Learning

- Learn dictionaries by alternating between
  - Learning the sparse representations given the dictionaries (sparse coding step)
  - Update the dictionary given the sparse representations (dictionary update step)
    - Dictionaries are updated iteratively

$$\begin{bmatrix} \mathbf{X}^l \\ \mathbf{X}^h \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{\Psi}_c^l & \mathbf{\Psi}^l & \mathbf{0} \\ \mathbf{\Psi}_c^h & \mathbf{\Psi}^h & \mathbf{0} \\ \mathbf{\Phi}_c & \mathbf{0} & \mathbf{\Phi} \end{bmatrix} \begin{bmatrix} \mathbf{Z} \\ \mathbf{U} \\ \mathbf{V} \end{bmatrix}$$

Multi-Modal Super-Resolution: Algorithm

Algorithm:

- Given the learned dictionaries, the LR depth image to enhance and the HR colour image

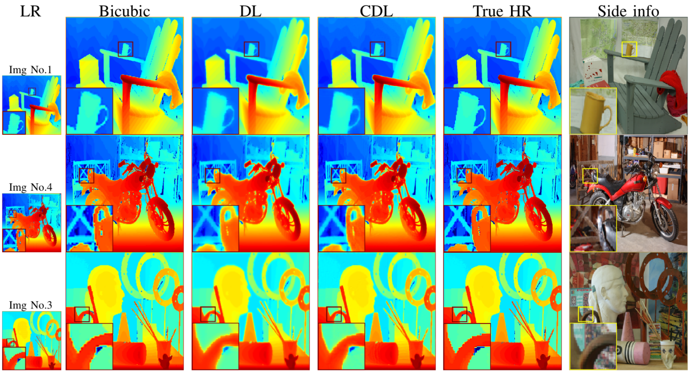
- Extract patches $\boldsymbol{x}_i^{LR}$ and $\boldsymbol{y}_i$

- Using sparse coding methods (e.g., OMP), retrieve the sparse vectors $\boldsymbol{z}_i$, $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$

- HR depth patches are then given by: $\mathbf{X}^h = \begin{bmatrix} \mathbf{\Psi}_c^h & \mathbf{\Psi}^h \end{bmatrix} \begin{bmatrix} \mathbf{Z} \\ \mathbf{U} \end{bmatrix}$

Key references include [Rodrigues et al. 2016], [Song, Deng et al. 2017], [Deng et al. 2017]

Multi-Modal Super-Resolution: Results



	Img No.1	Img No.2	Img No.3	Img No.4	Img No.5	Img No.6
Bicubic	27.63	33.31	27.27	24.30	25.00	27.89
DL	29.47	34.22	29.14	25.81	26.65	29.16
CDL	30.45	35.24	30.13	27.04	27.91	30.21

Image courtesy of M. Rodrigues

Multi-Modal Super-Resolution: Results

Super-resolving hyper-spectral images with the aid of RGB images

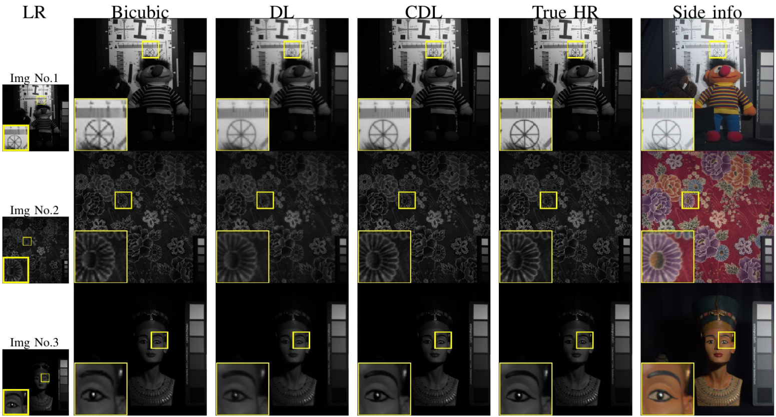
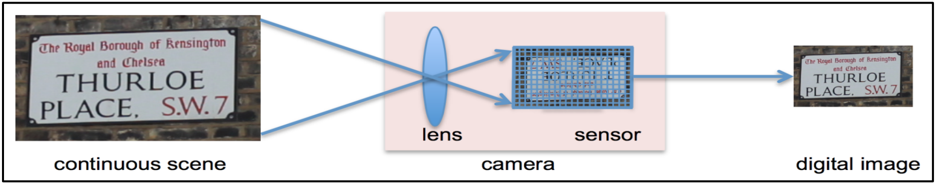
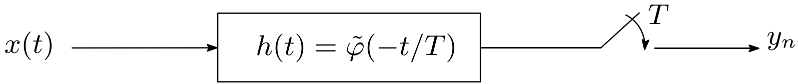


Image courtesy of M. Rodrigues

Image Super-Resolution: Model-Based Approach

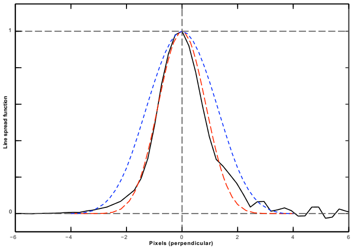




- Sampling and Resolution Enhancement are heavily connected through wavelet multi-resolution analysis
- The acquisition process can be modelled as low-pass filtering followed by sampling
- In a camera the low-pass filtering is due to the lenses and is modelled with the point spread function

Point Spread Function and Splines

(a)Original (2014 × 3039)



(b) Point Spread function

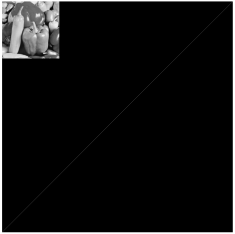
- In a camera the low-pass filtering is due to the lenses and is modelled with the point spread function
- The point spread function in a camera behaves like a spline function

Acquisition Process and Wavelet Decomposition

- The acquisition process remove the fine details of the image
- Since the low-pass filter is a spline, the acquisition process can be interpreted as a process that removes the wavelet coefficients at fine scales
- **Key insight**: Exploit the dependency across scale of the wavelet coefficients to retrieve the lost details.

(a) The high-resolution image 'Peppers'

(b) Low-pass and high-pass subbands of a 2-level 2D wavelet transform of (a)

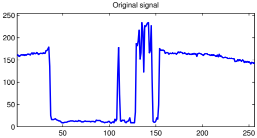


(c) We only have access to the low-pass subband of the 2-level 2D wavelet transform in (b)

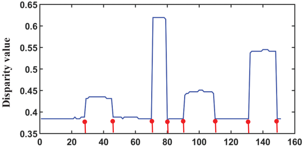
Wavelet Decomposition and Multiresolution

Wavelet Decomposition and Multiresolution

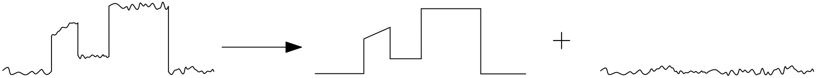
Modelling of Dependencies Across Scales



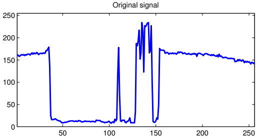
(a) Horizontal line



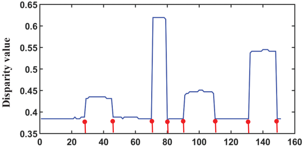
- We model lines of images as piecewise regular functions defined as the combination of a piecewise polynomial signal and a globally smooth function that lies in shift-invariant subspace:



Modelling of Dependencies Across Scales



(a) Horizontal line



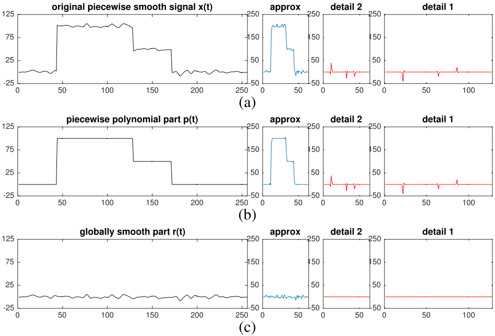
- We model lines of images as piecewise regular functions defined as the combination of a piecewise polynomial signal and a globally smooth function that lies in shift-invariant subspace:

$$x(t) = p(t) + r(t) = p(t) + \sum_n y_n \varphi(t/T - n)$$

Note that we assume: $\langle \varphi(t), \tilde{\varphi}(t-n) \rangle = \delta_n$

Modelling of Dependencies Across Scales

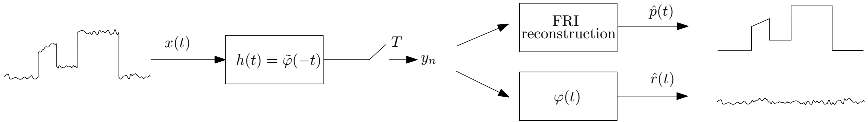
In the wavelet domain, the detail coefficients are only due to the piecewise polynomial signal



$$
\begin{aligned}
x(t) &= p(t) + r(t) \\
&= \underbrace{\sum_{n=-\infty}^{\infty} y^{p}_{J,n}\varphi_{J,n}(t) + \sum_{m=-\infty}^{J}\sum_{n=-\infty}^{\infty} d^{p}_{m,n}\psi_{m,n}(t)}_{p(t)} \\
&\quad + \underbrace{\sum_{n=-\infty}^{\infty} y^{r}_{J,n}\varphi_{J,n}(t)}_{r(t)} \\
&= \sum_{n=-\infty}^{\infty} \underbrace{(y^{p}_{J,n}+y^{r}_{J,n})}_{y_{J,n}}\varphi_{J,n}(t) + \sum_{m=-\infty}^{J}\sum_{n=-\infty}^{\infty} d^{p}_{m,n}\psi_{m,n}(t).
\end{aligned}
$$

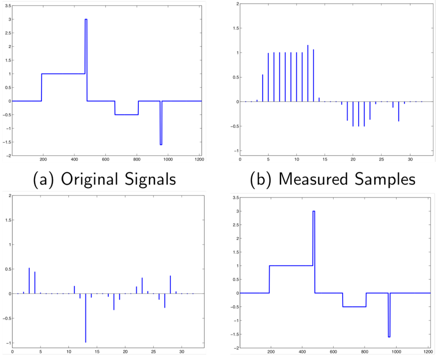
Reconstruction of Piecewise Smooth Signals

Key Insight:
- The residual can be recovered using traditional linear reconstruction methods
- Piecewise polynomial signals are continuous sparse signals and can be recovered using sparse sampling theory (i.e., finite rate of innovation theory [DragottiVB:07, UriguenBD:13])



Note that we assume: $\langle \varphi(t), \tilde{\varphi}(t-n) \rangle = \delta_n$

Exact Reconstruction of Piecewise Polynomial Signals

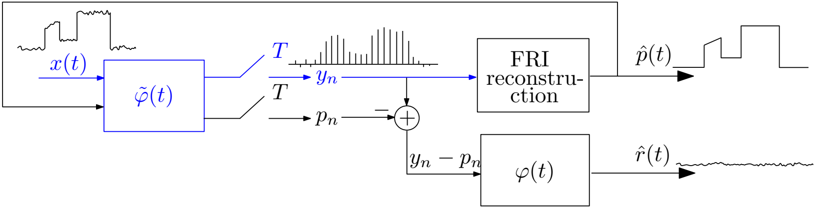


(a) Original Signals
(b) Measured Samples
(c) Finite Difference
(d) Reconstructed Signal

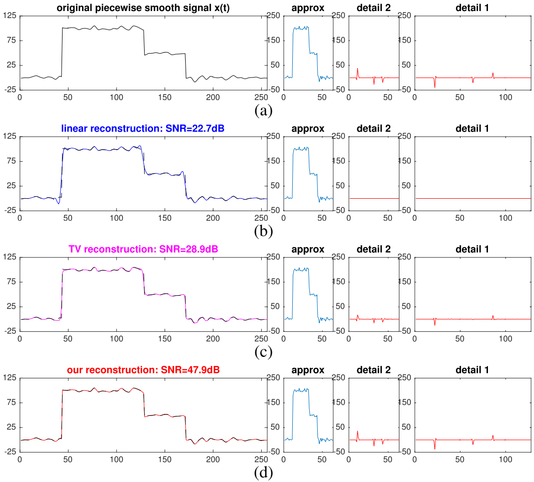
Piecewise polynomial signals are continuous sparse signals and can be recovered using sparse sampling theory (i.e., finite rate of innovation theory [DragottiVB:07, UriguenBD:13])

Reconstruction of Piecewise Smooth Signals

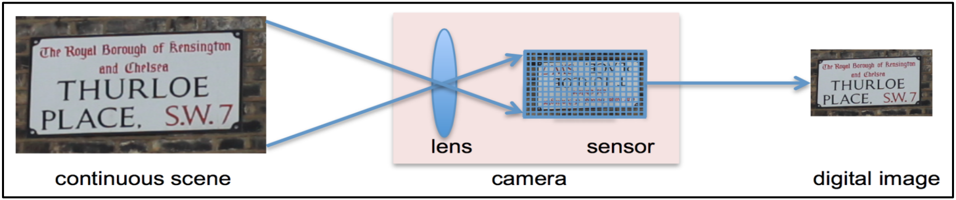
- remove the contribution of the reconstructed polynomial part $\hat{p}(t)$ from the samples y_n.
- reconstruct the residual $\hat{r}(t)$ by classical linear reconstruction.



Numerical Results



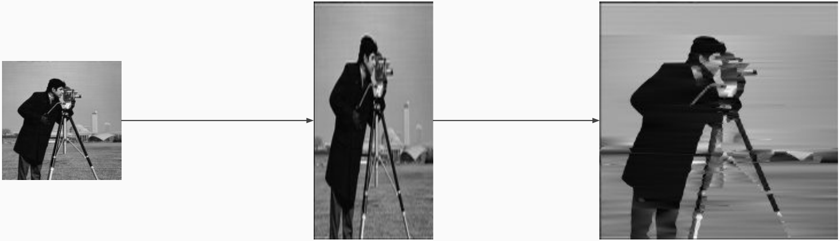
FRESH: FRI-based Single-Image Super-Resolution



- Algorithm capable of **increasing the resolution of digital images up to 4X**.
- Based on applying the 1-D resolution enhancement algorithm along several directions of the image
- The upsampled images are merged using wavelet theory
- Self-learning further improves performance
- **Accurately retrieve fine details** lost during the acquisition process.

[WeiD:TIP16]

FRESH: FRI-Based Single-Image Super-Resolution



Input image

128x128 pixels

Linear upsampling along columns

256x128 pixels

FRI upsampling along rows

256x256 pixels

FRESH: FRI-Based Single-Image Super-Resolution

Input image

128x128 pixels

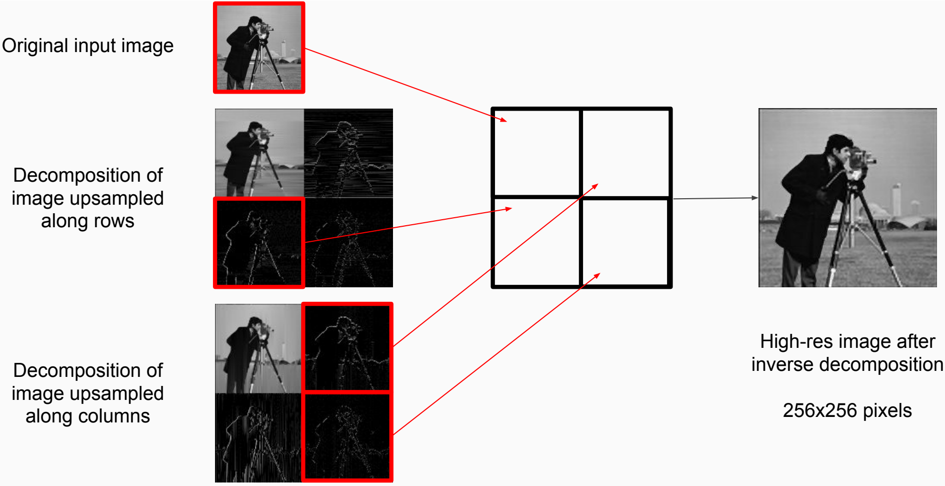
Linear upsampling along rows

128x256 pixels

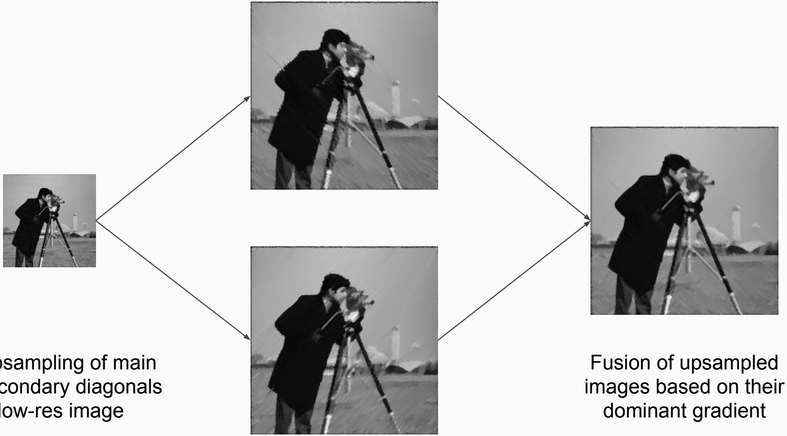
FRI upsampling along columns

256x256 pixels

FRESH: FRI-based Single-Image Super-Resolution



FRESH: FRI-based Single-Image Super-Resolution



FRI upsampling of main and secondary diagonals of low-res image

Fusion of upsampled images based on their dominant gradient

FRESH Results: Real Data

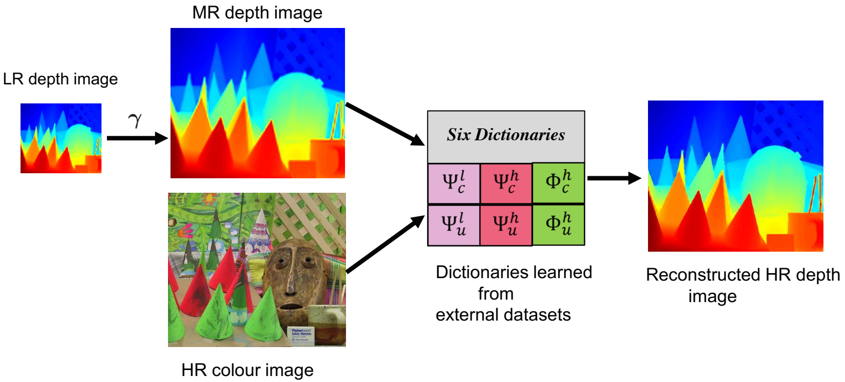
Low-res input
64 x64 pixels

Final result
256x256 pixels

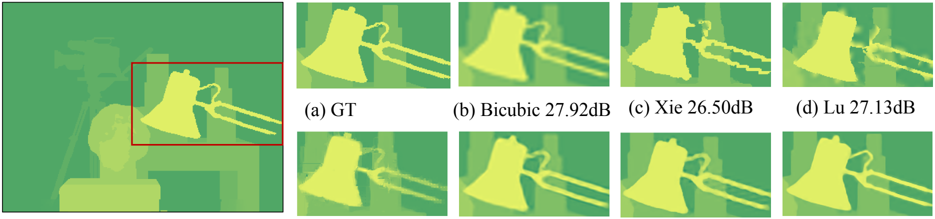
Single-Image Super-Resolution: Numerical Comparisons

(a) Original | (b) Linear (25.9dB) | (c) Timofte (27.3dB) | (d) FRESH (27.7dB)

Multimodal Depth Super-Resolution



Multimodal Depth Super-Resolution: Numerical Comparisons



(a) GT (b) Bicubic 27.92dB (c) Xie 26.50dB (d) Lu 27.13dB

(e) Ferstl 28.04dB (f) Timofte 31.13dB (g) Song 32.27dB (h) Ours 33.78dB

- Method (h) is a combination of FRESH and multimodal dictionary learning
- Method (g) is based on deep learning

Conclusions

- The notion of sparsity is still essential to develop and understand both model-based or data-driven methods
- Data-driven algorithms based on shallow (or deep) learning still use a modelling assumption (sparsity)
- Model-based approaches are competitive when they can reflect closely the nature of the data (e.g., depth images), but lack flexibility
- Model-based algorithm can always be combined with data-driven methods to yield algorithms with best performance

References

On Model-based Single-Image Image Super-Resolution

- X. Wei and P.L. Dragotti, FRESH -FRI-based single image super-resolution algorithm, IEEE Trans on Image Processing, Vol.25(8), pp. 3723-3735, August 2016.

On Data-Driven Single-Image Super Resolution

- J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” IEEE Trans. Image Process., vol. 19, no. 11, pp. 2861–2873, 2010
- J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, “Coupled dictionary training for image super-resolution,” IEEE Trans. Image Process., vol. 21, no. 8, pp. 3467–3478, 2012.
- R. Zeyde, M. Elad, and M. Protter, “On single image scale-up using sparse-representations,” in Curves and Surfaces. Springer, 2010, pp. 711–730.
- R. Timofte, V. De Smet, and L. Van Gool, “A+: Adjusted anchored neighborhood regression for fast superresolution,” in Proc. Asian Conf. Comput. Vision. Springer, 2014, pp. 111–126.

References (cont'd)

On Coupled Dictionary Learning for Multi-Modal Data

- P. Song, J. F. Mota, N. Deligiannis, and M. R. Rodrigues, "Coupled dictionary learning for multimodal image super-resolution," in IEEE Global Conf. Signal Inform. Process, pp. 162–166, 2016.
- P. Song, X. Deng, J F. C. Mota, N. Deligiannis, P. L. Dragotti and M. R. Rodrigues, "Multimodal Image Super-resolution via Joint Sparse Representations induced by Coupled Dictionaries", IEEE Trans. on Image Processing, under review, 2017