# An Example-based Method in Multi-frame Super Resolution

Yu Zhu, Yanning Zhang* and Haisen Li
Shaanxi Key Laboratory of Speech and Image Information Processing
School of Computer Science, Northwestern Polytechnical University, Xi'an
E-mail: zhu_yu000@163.com ynzhang@nwpu.edu.cn lihaisen19003@gmail.com

***Abstract*—In this paper, a two-stage super resolution method that combines the multi-frame and example-based methods is proposed. Traditionally, the multi-frame super resolution method only utilizes the continuity prior and complementary information among the low-resolution(LR) images with sub-pixel misalignment. While the example-based method digs the prior from abundant training images, but performs low ability to process the severe blurring image. So in our paper, firstly, the sequence is processed by the traditional fast and robust super resolution method to enhance the definition. Then, in the second stage, the high-resolution feature (HRF)/high-resolution(HR) dictionary pairs is prepared. The near-high-resolution image acquired in the former stage is split into overlapped patches, then sparse coded to the HRF dictionary, and linear combined with the HR dictionary atoms. The experiments on the synthetic and real image sequence prove that the proposed method outstands from the other methods.**

## I. INTRODUCTION

Super resolution (SR) image reconstruction is a very active research area currently. This technology leads to a probability of acquiring the high-quality images using the low-cost rather than the expensive and high-definition image sensors, which overcomes the frequency limitation declared by the Nyquist theorem[1]. During the degradation from a high-resolution (HR) image to low-resolution (LR) images, a mass of details or high-frequency information is lost. Therefore the super resolution process is an extremely ill-posed inverse problem and many methods are developed to constrain the solution.

These methods solve two kinds of problems: multi-frame/sequence super resolution and single image super resolution.

In the multi-frame/sequence SR case, LR images with sub-pixel misalignment are utilized. It is a basic assumption that the HR could generate the identical LR images during the same degradation described by the motion and blur parameter. On this assumption the conventional Maximum Likelihood (ML) method is developed and then interpreted in the Maximum a Posteriori (MAP) framework by adding the prior regularization such as Tikhonov[2], Huber MRF[3], and Total Variation[4-5]. The main purpose of the regularization term is to distinguish the useful high-frequency component from the noise component for better results. These methods are called reconstruction-based methods which reconstruct the HR image by fusing the complementary information in LR images.

The single image SR problem is even more ill-posed due to the image number and zoom factor limitation. Conventionally simple interpolation methods are used as Bilinear and Bicubic interpolation as well as the subsequent New Edge Directed Interpolation(NEDI)[6] and Partial Differential Equation (PDE) based method[7], aiming to model the local energy diffusion during the degradation. These interpolation methods tend to generate the over smooth results with ringing or cartoon artifacts. The limitations lead to the example-based method which learns the relativity between the LR patches and HR patches from quantities of other images with the same style. Freeman[8] proposed an SR approach learning the prediction from LR patches to HR patches via a Markov random field (MRF) solved by belief propagation. Yang[9-10] proposed a sparse representation based method to recover its most likely HR patches from LR patches. But the previously mentioned methods require enormous high-resolution and low-resolution patches pairs for training in which the degradation model is fixed such as downsampling.

When fusing the multi-frame/sequence images for solution enhancement, the traditional methods focus on the high frequency dispersing into the sub-pixel misaligned images. The enhanced edges and textures could not be true necessarily. With regard to the example-based method, the procedure of HR/LR patches learning fixes the degradations such as downsampling and loses the ability to deal with other type of blurring effect. Therefore it is necessary to combine these two kinds of methods for better SR results. The similar work is involved by Glasner[11] for single image SR. They extract the similar patches in different scales from the single image and propose a multi-patches / example-learning combined method without using the multi-frame/images. But the similar patches may not supply the exact mutual high-frequency information necessarily because they describe the different local scene after all. So naturally when the LR image sequence easily got, the result will be better with LR sequence instead of multi-patches.

This paper proposes a novel method that combines the multi-frame SR with the example-learning based approach. First, the image sequence is definition enhanced via a fast and robust super resolution to a near-high-resolution image. Then as an input, the near-high-resolution image is further super resolved by the example based method using the patch representation and linear combination with respect to the HRF/HR dictionary pairs.

This paper is organized as follows: Section II introduced a typical example based method. Then Section III describe our example-based algorithm combined with the multi-frame SR.

Section IV gives the performance of our method on synthetic and real image sequence. And Section V is the conclusion.

## II. A TYPICAL EXAMPLE-BASED METHOD

The recent work on sparse representation and compress sensing[12-13] indicates that in SR problem the linear relationships among high-resolution signals can be precisely recovered from their low dimensional projections. Motivated by this idea, Yang etc.[9-10] propose an example-based SR method via sparse representation. The method mainly includes two aspects. The first is the joint dictionary training to create the correspondence between the LR patches and HR patches while the second is the sparse representation for the most relevant HR patches selection and linear combination by the sparse coefficients. In the dictionary training step, the HR Training images is downsampled and reupsampled by bicubic interpolation. Then a set of 1-order and 2-oder high-pass filter $f_1 = [-1, 0, 1]$ $f_2 = f_1^T$ $f_3 = [1, 0, -2, 0, 1]$ $f_4 = f_3^T$ is used for gradient feature extraction. Meanwhile, 10000 patches are extracted from both the HR images and the corresponding LR feature images. Jointly the patches are put into the dictionary training process using LASSO[14]. After the dictionary acquirement, for an input LR image to super resolved, the process is similar with the LR training image to get patches. Then each patch is sparse coded with respect to the LR dictionary while the HR patches are reconstructed by linear combining the sparse coefficients and the HR dictionary atom. In the process, the DC component is isolated. The average of each patch is subtracted in the dictionary training step and added on in the reconstruction step. The integrated procedure is shown in fig.1.

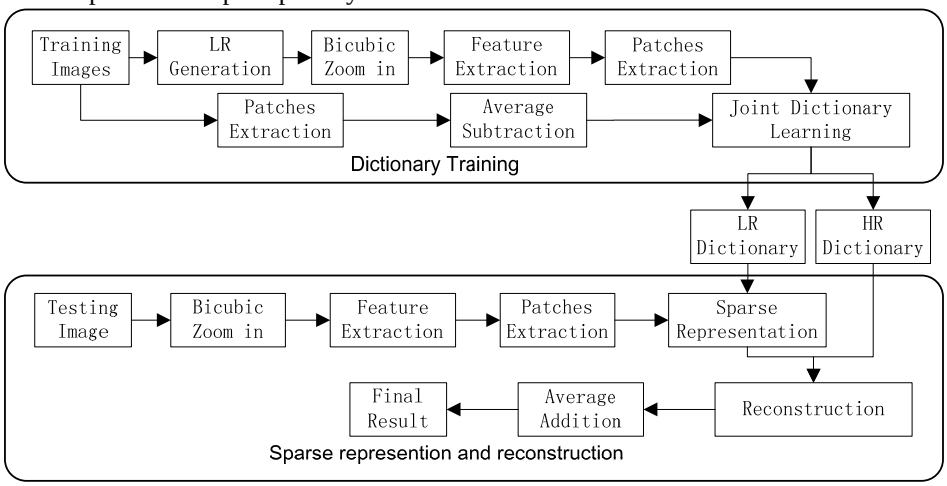



Fig. 1 The procedure of an example-based single super resolution

In Yang's method, the 1-order and 2-order gradient operator are adopted for feature extraction. This ensures that the computed coefficients fit the most relevant part of the low-resolution signal and also predicts the HR patches with more accuracy. However, the feature extraction demands that the input LR image should not be blurred or edge-defused too much. Otherwise the computation of the sparse coefficient will be severely impacted. Another limitation exists in the joint dictionary training. Given the HR image $X$ and the degraded image $Y$, from the degradation model $Y = DHX$, the downsampling operator $D$ inseparably relies on the zooming factor, which means that the trained LR/HR dictionary is definitely designed for a certain zooming factor given before the dictionary training. Here, $H$ represents a blurring filter, and $D$ the downsampling operator.

In the light of the limitation of the example-based single image resolution, an effective solution is the multi-frame stage introduction. First, the complementary information could help enhance the definition effectively to form a proper input. And second, combined with the multi-frame SR, the LR feature dictionary is naturally substitute by the HR feature dictionary unrelated with the zooming factor. In the following section, we describe the combined method as a two-stage SR approach, the multi-frame SR acts as a preprocess for the example-based step, and the dictionary is trained only on the HR patches.

## III. EXAMPLE-BASED METHOD COMBINED WITH MULTI-FRAME IMAGE SR

In this section, we extend Yang's example-based single image SR to the multi-frame SR. As we elaborate before, first we apply the fast and robust multi-frame SR method[8]. Then we take the multi-frame SR result as an input for the example-based step.

### *A. Multi-frame SR*

Generally, for a space/time invariant blur kernel and uniform downsampling process as well as the noise model, the degradation from HR image to the LR sequence can be modeled as follows[2]:

$$Y_k = DHF_kX + V \tag{1}$$

Here $Y_k$ is the kth LR image. $F_k$ is the camera motion parameter. $H$ and $D$ represent the blurring and downsampling operator. $V$ is the noise term. In this paper the measurement error in the cost function is minimized by $L_1$ norm, which

means the noise modeled by Laplacian PDF rather than Gaussian PDF( $L_2$ norm). For more robustness, the regularization is chosen as the Bilateral filter. Then the estimation of HR tends to be an optimization problem:

$$\begin{aligned}\hat{X} &= \arg\min_X [\sum_{k=1}^{N} ||DHF_kX - Y_k||_1 \\ &+ \lambda \sum_{m=-p}^{P} \sum_{l=-p}^{P} \beta^{|m|+|l|} ||X - S_x^l S_y^m X||_1] \end{aligned} \quad (2)$$

Where $P$ is the neighbor size. the matrices (operators) $S_x^l$, and $S_y^m$shift $X$ by $l$ and $m$ pixels in horizontal and vertical directions respectively, presenting several different of derivatives. The scalar weight $\alpha$, $0 < \beta < 1$, is applied to give a spatially decay from the neighbor center to the edge. Finally, using the steepest descent method, the minimization problem can be solved efficiently.

Through the fast and robust multi-frame SR, the HR image is estimated primarily. Then we enter into the second stage, which apply the example-based method for further definition enhancement.

### *B. Joint dictionary training*

According to Yang's approarch, before the dictionary training, a large number of natural images collected from the internet are prepared, from which 10000 patch pairs are randomly selected. Then the HR patches and the LR patches are jointly put in for the HR/LR dictionary optimization. Here the LR image is downsampled and re-upsampled version of the HR image via Bicubic interpolation and then filtered by a set of high-pass filter for feature extraction. But in our method, the multi-frame SR supply the relatively high-definition image for the training input, so here we use only the HR patches and the HR feature (HRF) patches. Separately the HR/LR dictionary could be found by solving the following optimization problem:

$$D_h = \arg\min_{D_h,\alpha} ||X^h - D_h\alpha||_2^2 + \lambda||\alpha||_1 \quad (3)$$

$$D_f = \arg\min_{D_f,alpha} ||X^f - D_f\alpha||_2^2 + \lambda||\alpha||_1 \quad (4)$$

Where $D_h$ and $D_f$ are the HR/HRF dictionary. $\alpha$ is the sparse coefficient related to the dictionary. $L_1$ norm about $\alpha$ constrained the solution with sparse prior. $\lambda$ balances the error term and regularizing term. $X^h = \{x_1, x_2, ..., x_n\}$ indicates the arranged vectorized HR patches and $X^f = \{f1, f2, ..., f_n\}$ the arranged vectorized patches feature with $f_i = Fx_i$. Here $F$ represents the 1-order and 2-order gradient feature extraction operator. The equation is not convex in both the dictionary and sparse coefficient. But when one is fixed, the optimization of the other one is convex, so the two equations can be solved by alternative iteration. However, the two separate solutions may not be consistent necessarily, so the two equations are combined to enforce the same sparse representation.

$$\{D_h, D_f, \alpha\} = \arg\min_{\{D_h,D_f,\alpha\}} ||X_c - D_c\alpha||_2^2 + \lambda\left(\frac{1}{N} + \frac{1}{M}\right)||\alpha||_1 \quad (5)$$

$$X_c = \begin{bmatrix} \frac{1}{\sqrt{N}}X^h \\ \frac{1}{\sqrt{M}}X^f \end{bmatrix} D_c = \begin{bmatrix} \frac{1}{\sqrt{N}}D_h \\ \frac{1}{\sqrt{M}}D_f \end{bmatrix}, \quad (6)$$

Here $N$ and $M$ are the dimensions of the vector-formed $X^h$ and $X^f$. While the two kinds of input patches and dictionaries are categorized as one.

### *C. Sparse representation and construction*

In the reconstruction step, the previous estimation of HR image is used for calculate the sparse coefficients with regarded to $D_f$. Then the final HR patches can be acquired by linear combining the coefficients with the atoms in $D_h$. Tightly connected to the compress sensing theory, each patch is sparse coded for the most relevant linear combination of HR patches. Originally, the sparse representation tends to be an $L_0$ norm optimization problem, which is proved to be NP-hard. Relaxed to the $L_1$ norm optimization, it can be solved efficiently by the linear programming.

Given the dictionary pair $D_h$ and $D_f$, the following equation indicates the necessary constraint for the sparse coding in the local patches.

$$\begin{aligned} min||\alpha||_1 \quad s.t. \quad ||D_f\alpha - Fy||_2^2 &\leq \varepsilon_1 \\ ||PD_h\alpha - \omega||_2^2 &\leq \varepsilon_2 \end{aligned} \quad (7)$$

Here $\alpha$ is the sparse coefficient. $F$ is the linear feature extraction operator which converts the original brightness to the gradient. $P$ is the overlapped region extraction operator, while $\omega$ is the previous reconstruction value on the overlap. The first-order and second-order 1D filter is used for the feature extraction as $f_1 = [-1, 0, 1]$ , $f_2 = f_1^T$ , $f_3 = [1, 0, -2, 0, 1]$, $f_4 = f_3^T$ in the sparse representation as well as the dictionary learning. The first constraint indicates that the error between the feature and the sparse representation should be minimized. The second constraint guarantees the continuity between adjacent patches. In practice, the desired HR patches are reconstructed by averaging the overlapping patches. Then the HR patches are reconstructed by the linear combination $X = D_h\alpha$ .The integrated flow chart is shown as Fig.2.

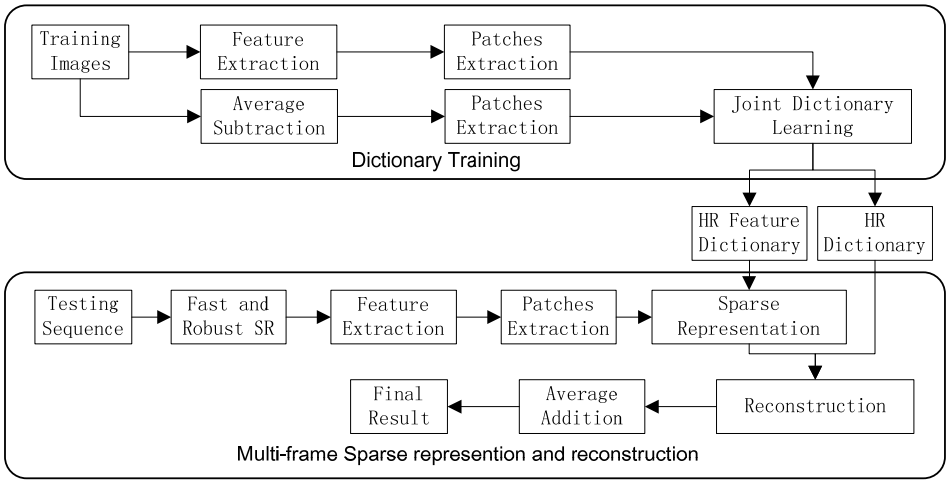



Fig.2 The procedure of the proposed method

## IV. EXPERIMENT AND ANALYSIS

In this section, several experiments are conducted to prove the effect of the proposed method. As above elaboration, in our experiments, the input sequence is firstly processed by the fast and robust SR algorithm. In this stage, the Bliteral filter decaying parameters $\beta = 0.7$, neighbor size $P = 2$ and the iteration step $\beta = 1$. While the blurring factor $H$ is regarded as the Gaussian convolution with the PSF size $3 \times 3$ and variance $\sigma = 1$. The iteration stops until the convergence. In the second stage, we use 70 HR images and randomly select 10000 $5 \times 5$patches to implement the dictionary training. In the reconstruction step, the patches are extracted with 4 pixels overlap. The balance factor between sparsity and error term $\lambda = 0.15$ in the sparse representation.

Two kinds of input data are selected in our experiment, synthetic LR images from an HR image and the real video sequence. The former data is used for evaluate the capacity of the detail recovery while the latter outstand the advantage of the proposed method on blurring sequence recovery. Finally the RMS error criterion is applied for the evaluation. The definition of RMS is given by

$$RMS = \sqrt{\frac{\sum_{i,j}(x(i,j) - x_{ref}(i,j))^2}{M \times N}} \tag{8}$$

Where $M$, $N$ represent the number of the rows and columns, $x(i,j)$ and $x_{ref}(i,j)$ is the $(i,j)th$ pixel value of the processed image and reference image.

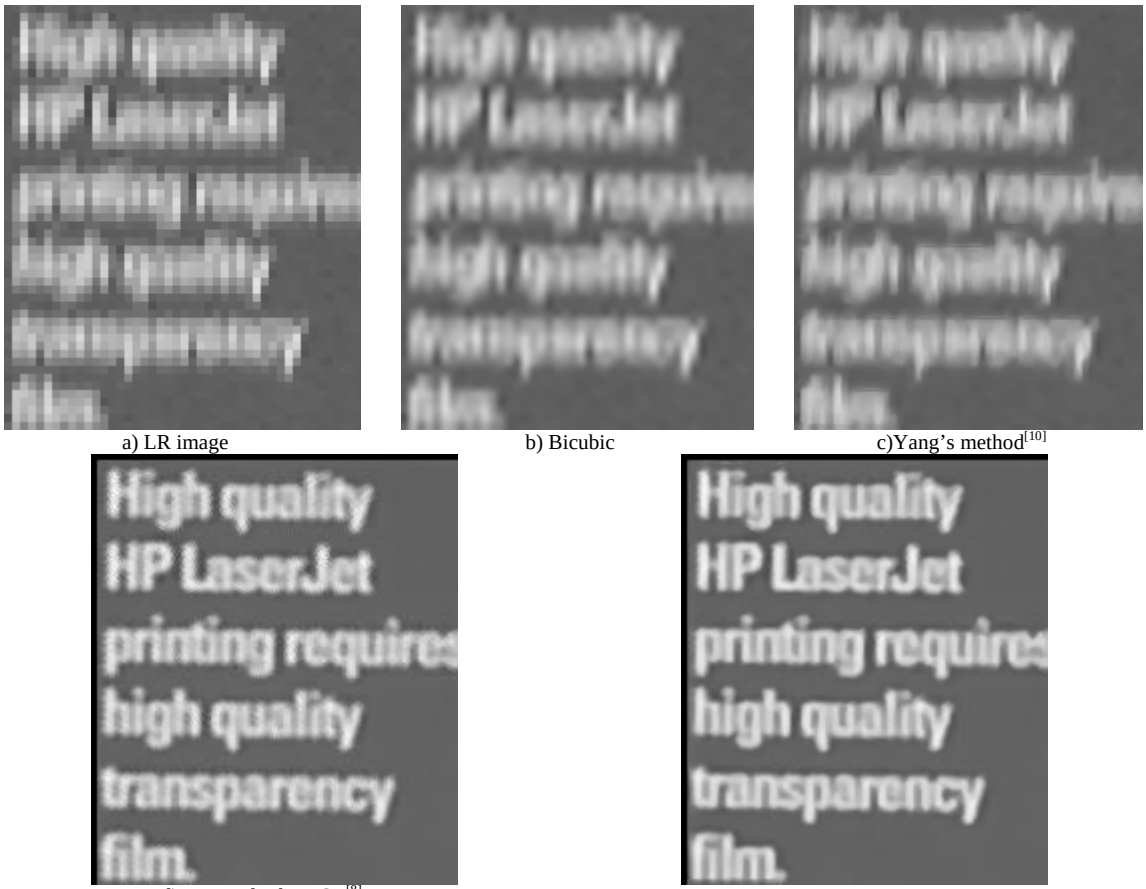
First we apply the proposed method to a set of synthetic satellite images (10 frames) which are downsampled from an high definition image with random translational motion added. Then we make a comparison with Yang's method and the multi-frame SR proved it is effect on the texture recovery (see Fig.3). Whether from the vision or PSNR criterion, the proposed method outstands from the other methods a lot.

a) Ground truth b) LR image RMS 23.7422 c) Bicubic RMS 21.6849

d) Yang's method[10] RMS 17.2144 e) Fast and robust SR[8] RMS 19.0667 f) Proposed method RMS 13.9991

Fig.3 Experiment result on the synthetic data

In Fig.4, the proposed method is applied on the actual text sequence. The sequence contains 30 frames with severe blurring effect. From the results we can see the ordinary single image SR method (Bicubic or Yang's method) will totally fail when implemented on the blurring sequence. Meanwhile, due to the registration error, the multi-frame SR performs with jagged effect around the edges, and the proposed method well overcomes the difficulty.




a) LR image b) Bicubic c)Yang's method[10]

d) Fast and robust SR[8] e)Proposed method

Fig.4 Experiment result on the read sequence data

## V. CONCLUSIONS

In this paper, we proposed a two stage super resolution method that combines the multi-frame and example-based methods. Firstly the fast and robust super resolution method is applied to the multi-frame images to enhance the definition. Then as the input of next stage, the near-high-resolution image is directly used for the example-based super resolution, which prepares the HRF/HR rather than LRF/HR dictionary pairs to estimate the high frequency of the input. Then the input is super resolved on the patch level by sparse

representation and linear combination with respect to the HRF and HR dictionary separately. The experiments on the synthetic and actual image sequence show that the proposed method has remarkable improvements over whether multi-frame method or example-based single-frame algorithms.

## ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No.60872145, No.60903126), National High Technology Research and Development Program(863) of China (No.2009AA01Z315), China Postdoctoral Special Science Foundation(No.201003685), China Postdoctoral Science Foundation(No.20090451397).The Cultivation Fund of the Key Scientific and Technical Innovation Project , Ministry of Education of China(No. 708085)

## REFERENCES

[1] S. C. Park, M. K. Park, and M. G. Kang, "Super-Resolution Image Reconstruction: A Technical Overview," *IEEE Signal Processing Magazine*, vol. 20, pp. 1646-1658, 2003.

[2] M. Elad and A. Feuer, "Restoration of a single super-resolution image from several blurred, noisy, and undersampled measured images," *IEEE Transactions on Image Processing*, vol. 6, pp. 1646-1658, 1997.

[3] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Bayesian Methods for Image Super-Resolution," *The Computer Journal*, vol. 52, pp. 101-113, 2009.

[4] T. F. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *IEEE Transactions on Image Processing*, vol. 10, pp. 231-241, 2001.

[5] D. Capel and A. Zisserman, "Super-resolution enhancement of text image sequences," in *International Conference on Pattern Recognition*, 2000, pp. 600-605 vol.1.

[6] L. Xin and M. T. Orchard, "New edge-directed interpolation," *IEEE Transactions on Image Processing*, vol. 10, pp. 1521-1527, 2001.

[7] Z. Ning, W. Jing, and W. Zhongqian, "Image Zooming Based on Partial Differential Equations," *Journal of Computer Aided Design & Computer Graphics*, vol. 17, pp. 1941-1945, 2005.

[8] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, pp. 1327-1344, 2004.

[9] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Computer Vision and Pattern Recognition*, Anchorage, AK, United states, 2008.

[10] J. Yang, J.Wright, H. T. S., and Y. Ma, "Image Super-Resolution Via Sparse Representation," *IEEE Transactions on Image Processing*, vol. 19, pp. 2861-2873, 2010.

[11] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *International Conference on Computer Vision*, 2009, pp. 349-356.

[12] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289-1306, 2006.

[13] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed Sensing and Redundant Dictionaries," *IEEE Transactions on Information Theory*, vol. 54, pp. 2210-2219, 2008.

[14] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267-288, 1994.