

Super-resolution Based on Two Dictionary-Pairs

Xin Li, Xiuchang Zhu

Jiangsu Province Key Lab on Image Processing & Image Communication,
Nanjing University of Posts and Telecommunications, Nanjing, 210003, China

Tel.: +86-2583492416

E-mail: lixin@njupt.edu.cn, zhuxc@njupt.edu.cn

Received: 11 August 2013 /Accepted: 25 September 2013 /Published: 31 October 2013

Abstract: For learning-based super-resolution reconstruction, the selection and training of dictionary play an important role in improving image reconstruction quality. A super-resolution algorithm based on two dictionary-pairs is proposed in this paper. This algorithm selects image's high- and mid-frequency components as the features of high- and low-resolution patches respectively, and gets the first dictionary-pair, i.e. joint-basic high/low-resolution dictionary-pair, by means of joint training. Then it calculates the difference between original high/mid-frequency components and reconstructed high/mid-frequency components with the first dictionary-pair, and composes the second dictionary-pair, i.e. residual high- and low-resolution dictionary-pair. During super-resolution reconstruction, the high-frequency component and residual high-frequency component of low-resolution image are reconstructed respectively with the above two dictionary-pairs. The experiment results show that, the subjective and objective reconstruction quality could be effectively improved by our proposed algorithm. *Copyright © 2013 IFSA.*

Keywords: Super-resolution, Learning-based, Dictionary, Sparse representation, Neighbor embedding.

1. Introduction

Super-resolution (SR) for single-frame image is the method to recover a high-resolution (HR) image from a low-resolution (LR) image, which is useful in practices such as the restoration of historic photographs, enlargement "thumbnail" images on web pages. SR is a typical ill-posed inverse problem [1, 2] with infinite solutions.

Current SR algorithms can be classified into three types [3], i.e. interpolation-based, reconstruction-based, and learning-based. Interpolation-based algorithms [4, 5], including bilinear and bicubic interpolation, tend to generate overly smooth images. Especially if the amplification factor is large, the image edges and details will be lost, resulting in blurred image problem. The most typical reconstruction-based algorithm is iterative back-

projection (IBP) [6], which has faster convergence speed. However, its solution is not unique. So usually regularization is necessary for IBP, such as total variation [7, 8], soft edge smoothness prior [9], gradient profile prior [10] etc., according to different kinds of image's a priori information. The difficulty of this kind of algorithms is the establishment of a priori knowledge, and the drawback is the limitation of a priori. Learning-based SR [11-17] is a research hotspot in recent years. By finding the correlation between HR and LR image patches with machine learning method, high-frequency details are estimated, and finally HR image could be predicted. Freeman [12] proposed an example-based SR. This algorithm established the mappings between HR and LR image patches by hidden Markov model. Then according to the Bayesian maximum a posteriori (MAP) criterion, it uses the steepest

descent method to find a local maximum. Chang [15] proposed a SR through neighbor embedding. Based on manifold learning, this algorithm tries to find the locally linear embedding coefficients of the LR image patches, which are nearest to the target image patch, and recovers the HR image patch with the linear combination of the same coefficients. Yang [17] proposed the sparse representation based SR algorithm. By learning the correlation of sparse coefficients between HR and LR image patches, over-complete dictionary-pair comprising HR and LR image patches is trained. For each test patch, sparse representation under the over-complete dictionary is obtained by the approach of linear programming. Then finally HR image is reconstructed with these weighting coefficients.

Chang and Yang's algorithm only remove the mean value of HR image patches when training HR dictionary, so they cannot fully reflect high-frequency details of HR image. In this paper, in order for dictionary training, we select the high- and mid-frequency components as the features of HR and LR image patches respectively, and get Joint-Basic HR/LR Dictionary-Pair. According to statistical analysis, there are still much high-frequency details in the residual, after reconstructed high-frequency is subtracted from original high-frequency. So we make use of high- and mid-frequency residuals to construct Residual HR/LR Dictionary-Pair. By applying the two dictionary-pairs to LR image, the reconstruction quality could be greatly improved.

2. Construction of Two Dictionary-pairs

2.1. Sparse Representation Model

Suppose the over-complete dictionary $D \in R^{n \times N}$ ($N > n$) with N atoms could sparsely represent the signal $x \in R^n$, i.e. $x = D\alpha$, where $\alpha \in R^N$ only has k ($k \ll N$) non-zero coefficients. This is the basic idea of sparse representation. In SR reconstruction, LR image is obtained by projecting HR image with $y = SHx$, where S is a down-sampling operator, and H is a blurring filter. Then image SR could be expressed by

$$\min \|\alpha\|_0 \quad \text{s.t.} \quad y = SHD\alpha \quad (1)$$

It means if sparse representation coefficients could be got by means of LR image y , then HR image will be recovered. Eq. (1) has been proved to be a NP-hard problem, so usually only approximate solutions are needed, such as Matching Pursuit (MP), Orthogonal Matching Pursuit (OMP), and other greedy algorithms. In 1999 Donoho [18] found, if the signal representation vector is sparse enough in a signal sparse decomposition problem, then l_1 -norm could replace l_0 -norm and above problem will be converted to a convex optimization problem as following:

$$\min \|\alpha\|_1 \quad \text{s.t.} \quad y = SHD\alpha \quad (2)$$

Thus it can be seen that, a proper over-complete dictionary is the key to solve the problem.

2.2. Training for Joint-Basic HR/LR Dictionary-Pair

In practice when training the dictionaries, the selected HR/LR image patch pairs may not be image patches themselves, but could be other features of image patches to show their characteristics. [12] chooses the high- and mid-frequency information as the features of HR/LR image patches respectively. [15] and [17] selected the mean value removed image patch as the feature of HR image patch, and first and second order gradients of the image patch as the feature of LR image patch. Considering people are more sensitive to high-frequency details, in this paper high- and mid-frequency information are selected to be the features of HR and LR image patches respectively, which are also used to train the first dictionary-pair, i.e. joint-basic HR/LR dictionary-pair. Let $\{P_x, P_y\}$ be the HR/LR image patch pair, where $P_x = \{x_{h1}, x_{h2}, \dots, x_{hL}\}$ are the high-frequency components of sampled HR image patches, and $P_y = \{y_{m1}, y_{m2}, \dots, y_{mL}\}$ are the mid-frequency components of the corresponding sampled LR image patches.

Our goal is to find a target dictionary pair $\{D_h, D_l\}$, which makes the corresponding image patch's high-frequency component x_{hi} and mid-frequency component y_{mi} have the same sparse representation coefficients in D_h and D_l .

$$\{D_h, \Lambda\} = \arg \min_{D_h, \Lambda} \|P_x - D_h \Lambda\| + \lambda \|\Lambda\|_1, \quad (3)$$

$$\{D_l, \Lambda\} = \arg \min_{D_l, \Lambda} \|P_y - D_l \Lambda\| + \lambda \|\Lambda\|_1, \quad (4)$$

where $\Lambda = \{\alpha_i\}_{i=1}^L$ is the coefficient matrix of sparse representation, and λ is the balance between sparsity of the solution and fidelity of the approximation to y .

To ensure the sparse representation coefficients for HR and LR dictionary are same, we combine the two dictionaries and train them simultaneously. The image patch pairs for dictionaries training are

$$P = \begin{bmatrix} \frac{1}{\sqrt{M}} \text{norm}(F_x P_x) \\ \frac{1}{\sqrt{N}} \text{norm}(F_y P_y) \end{bmatrix},$$

where $norm(\cdot)$ is normalization operator. M and N are the dimensions of the HR and LR image patches in vector form respectively, and they are used as weights to ensure the consistency in joint training.

$$\text{Let } D = \begin{bmatrix} \frac{1}{\sqrt{M}} D_h \\ \frac{1}{\sqrt{N}} D_l \end{bmatrix} \text{ be the target dictionary.}$$

Eq. (3) and (4) can be combined and rewritten as

$$\{D, \Lambda\} = \arg \min_{\Phi, \Lambda} \left\{ \|P - D\Lambda\|_2^2 + \lambda \|\Lambda\|_1 \right\} \quad (5)$$

Eq. (5) is not convex in either D or Λ . However if one is fixed, the other will be convex, and could be solved by corresponding optimization method. The training process of joint-basic HR/LR dictionary-pair is shown below in algorithm 1.

Algorithm 1: training algorithm for joint-basic HR/LR dictionary-pair

- 1) Input: Training image patch pairs P , dictionary size K .
- 2) Initialization: Initialize dictionary $D \in R^{(M+N) \times K}$ with a Gaussian random matrix, where $D_h \in R^{M \times K}$ and $D_l \in R^{N \times K}$. Note that each column of dictionary D should be normalized.
- 3) Iterations until convergence:
 - a) Fix D , update Λ with linear programming.

$$\Lambda = \arg \min_{\Lambda} \left\{ \|P - D\Lambda\|_2^2 + \lambda \|\Lambda\|_1 \right\} \quad (6)$$

- b) Fix Λ , update D

$$D = \arg \min_{\Phi} \|P - D\Lambda\|_2^2$$

s.t.

$$\|D_i\|_2 \leq 1, i = 1, 2, \dots, K, \quad (7)$$

where D_i is the column vector of dictionary D , i.e. every atom of it. This is a Quadratically Constrained Quadratic Programming problem, and many optimization packages could solve the problem. Lasso algorithm [19] is adopted in this paper.

- 4) Output: joint-basic HR/LR dictionary-pair $\{D_h, D_l\}$.

2.3. Construction of Residual HR/LR Dictionary-Pair

With the joint-basic HR/LR dictionary-pair trained in last section, HR image could be

reconstructed by learning-based SR algorithms. We get the residual mid- and high-frequency images, by subtracting the reconstructed mid- and high-frequency information from the original mid- and high-frequency information. As shown in Fig. 1, the difference between reconstructed mid-frequency and original mid-frequency is small, but some high-frequency details can still be seen in the residual high-frequency image. So we make use of residual mid- and high-frequency information to further predict the image's high-frequency details.

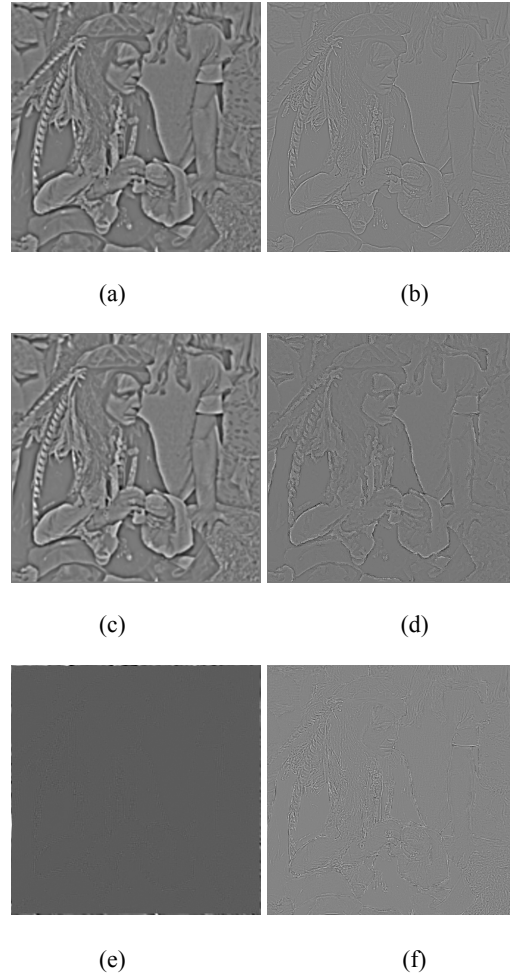


Fig. 1. Residual mid- and high-frequency image.

- (a) mid-frequency image,
- (b) high-frequency image,
- (c) reconstructed mid-frequency image,
- (d) reconstructed high-frequency image,
- (e) residual mid-frequency image,
- (f) residual high-frequency image.

Define residual HR image patches $Re P_x = \{Re x_{h1}, Re x_{h2}, \dots, Re x_{hT}\}$, and residual LR image patches $Re P_y = \{Re y_{m1}, Re y_{m2}, \dots, Re y_{mT}\}$. Kim [20] calculated the sum of absolute difference for residual HR image patches and residual LR image patches respectively, and found strong correlations between them. Therefore in this paper, residual HR

and LR image patches are used to construct the second dictionary-pair, i.e. residual HR/LR dictionary-pair, in order to improve the reconstruction quality further.

If residual patches are normalized with the training method in algorithm 1, mean value will be lost. As a result, residual HR/LR dictionary-pair should be directly composed of residual HR and LR image patches without joint training, i.e. $\{\text{Re } D_h, \text{Re } D_l\} = \{\text{Re } P_x, \text{Re } P_y\}$.

3. Proposed SR Algorithm Based on Two Dictionary-pairs

According to above discussion, SR reconstruction problem needs to solve Eq. (2). If apply the dictionary-pair $\{D_h, D_l\}$ trained in section 2.2 into Eq. (2), Then Eq. (2) can be rewritten as

$$\hat{\alpha} = \arg \min_{\alpha} \left\{ \|y_m - D_l \alpha\|_2^2 + \lambda \|\alpha\|_1 \right\}, \quad (8)$$

where y_m is the mid-frequency component of LR image patches.

Firstly high-frequency component is reconstructed by use of the first dictionary-pair $\hat{x}_h = D_h \hat{\alpha}$. To get better quality of HR image patches, then residual HR/LR dictionary-pair is used for the second reconstruction. Since residual HR/LR dictionary-pair is only combination of residual patches, our proposal is to use simple neighbor embedding (NE) to get the residual high-frequency component $\text{Re } \hat{x}_h$ of the image. Then by superimposing the high-frequency component and residual high-frequency component on LR image patches, we will get the final HR image patches.

To reduce the blocky artifacts caused by patch by patch processing, we make certain number of pixels in adjacent patches overlapped when extracting patches, and then combine HR image patches to get HR image \hat{X} . Moreover IBP algorithm is adopted for global optimization constraints.

Algorithm 2: SR algorithm based on two dictionary-pairs

1) Input: trained joint-basic HR/LR dictionary-pair $\{D_h, D_l\}$, residual HR/LR dictionary-pair $\{\text{Re } D_h, \text{Re } D_l\}$, and LR image Y .

2) Initialization: get the same size LR image with HR image by bicubic interpolation, extract the corresponding image patch y , and ensure that the adjacent image patches have overlapping pixels in each direction.

3) Iteration until all the image patches are processed:

a) Calculate the mid-frequency component y_m of LR image patch y . Get the optimized sparsity coefficients according to Eq. (8).

b) Get the image's reconstructed high-frequency component $\hat{x}_h = D_h \hat{\alpha}$ and reconstructed mid-frequency component $\hat{y}_m = D_l \hat{\alpha}$.

c) Calculate residual mid-frequency component $\text{Re } y_m = y_m - \hat{y}_m$. Get residual high-frequency component $\text{Re } \hat{x}_h$ with neighbor embedding algorithm.

d) Get HR image patch $\hat{x} = y + \hat{x}_h + \text{Re } \hat{x}_h$.

Output of interaction: HR image \hat{X} consolidated from HR image patches \hat{x} .

4) For \hat{X} , solve the global optimization constraint with IBP algorithm.

$$X^* = \arg \min_x \left\{ \|Y - SHX\|_2^2 + \zeta \|X - \hat{X}\|_2^2 \right\} \quad (9)$$

5) Final Output: HR image X^* , solved by our proposed SR algorithm based on two dictionary-pairs.

4. Experimental Results and Analysis

A series of experiments are done to verify the effectiveness of our proposed algorithm as following. HR images are obtained by three times amplifying LR images. The sizes of HR/LR image patches and residual HR/LR image patches are all 5×5 . The sizes of joint-basic HR/LR dictionaries, i.e. the number of atoms in a dictionary, are 512. And the sizes of residual HR/LR dictionaries are 10000.

It's shown in Fig. 2 and Fig. 3 that, bicubic interpolation makes image blurred, and method [15] could get good HR image. For our proposal, even if only joint-basic HR/LR dictionary-pair is used, it still could get better reconstruction quality than method [15]. The reason is method [15] only removes the mean value, while our algorithm selects high-frequency component as the feature of HR image, resulting in better reflection of image's high-frequency information. Furthermore, if our proposed two dictionary-pairs algorithm is applied to SR, even better image details will be obtained. Both subjective and objective quality could be evidently improved.

5. Conclusions

A SR algorithm based on two dictionary-pairs is proposed in this paper. Firstly, this algorithm selects image's high- and mid-frequency components to compose HR/LR image patch pair, in order for the training of joint-basic HR/LR dictionary-pair. Secondly, to construct residual HR/LR dictionary-pair, it calculates the residual high- and mid-frequency components with the reconstructed high- and mid-frequency components.

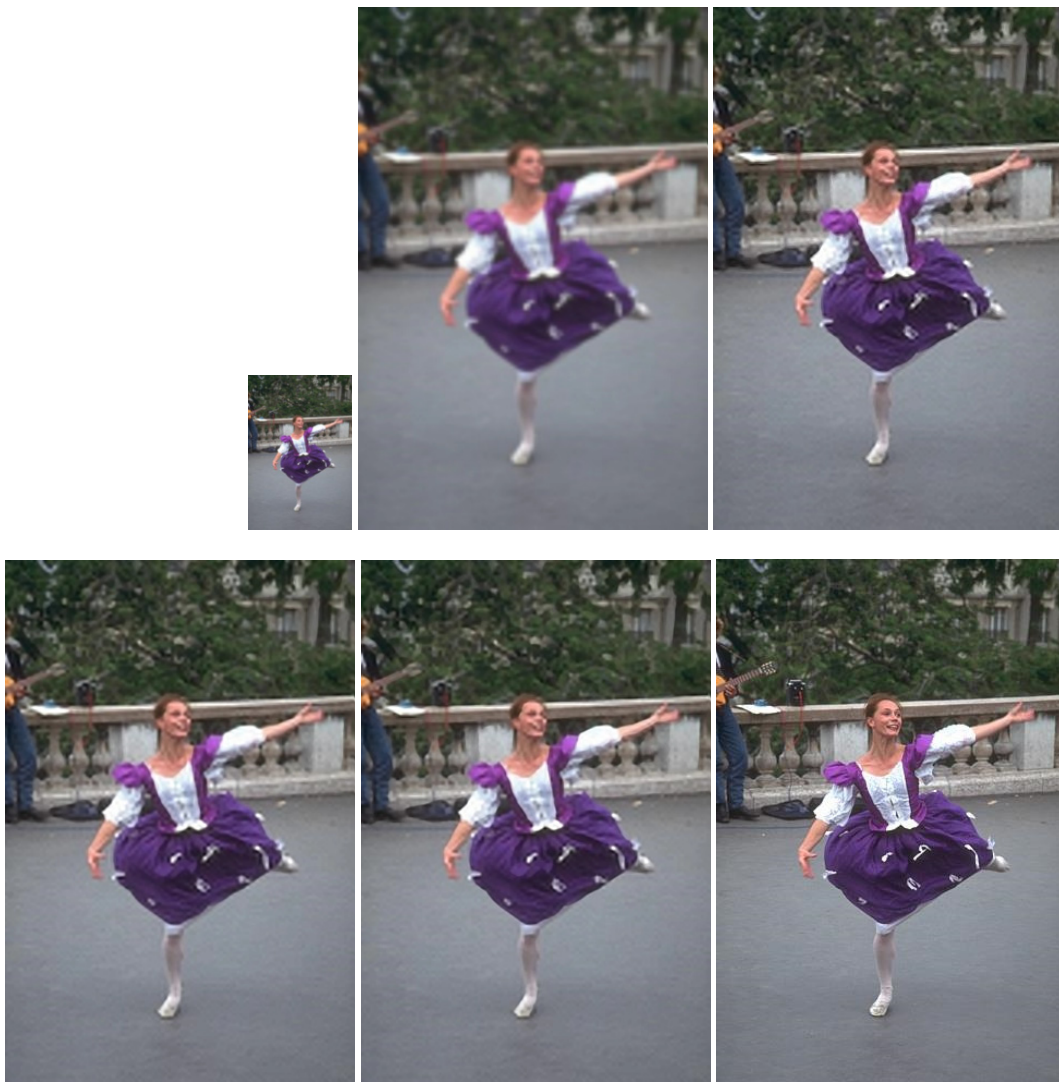


Fig. 2. The SR results (scaling factor 3) on image *Dancer*.
 Top row: LR image, bicubic interpolation (PSNR = 28.43 dB, SSIM = 0.8234),
 method [15] (PSNR = 29.73 dB, SSIM = 0.8526).
 Bottom row: single dictionary-pair (PSNR = 30.59 dB, SSIM = 0.8751),
 two dictionary-pairs (PSNR = 31.11 dB, SSIM = 0.8882), Original HR image.



Fig. 3. The super-resolution results (scaling factor 3) on image *Butterfly*.
 Top row: LR image, bicubic interpolation (PSNR= 29.57 dB, SSIM = 0.8847),
 method [15] (PSNR = 31.57 dB, SSIM= 0.9095).
 Bottom row: single dictionary-pair (PSNR = 32.09 dB, SSIM = 0.9207),
 two dictionary-pairs (PSNR = 32.66 dB, SSIM = 0.9320), Original HR image.

So SR reconstruction in our algorithm is based on two dictionary-pairs. The experiment results show that, our algorithm achieves much better results than many state-of-the-art algorithms in terms of PSNR, SSIM and visual perception.

Acknowledgements

This work is supported by the National Natural Science Fund of China (No. 61071091, 61071167), the Postgraduate Innovation Program of Scientific Research of Jiangsu Province (No. CXLX11_0406).

References

- [1]. M. Bertero and P. Boccacci, Introduction to Inverse Problems in Imaging, *IOP*, Bristol, U. K., 1998.
- [2]. S. Baker and T. Kanade, Limits on super-resolution and how to break them, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 9, Sep. 2002, pp. 1167-1183.
- [3]. J. Van Ouwerkerk, Image super-resolution survey, *Image and Vision Comput.*, Vol. 24, No. 10, 2006, pp. 1039–1052.
- [4]. V. Caselles, J. Morel and C. Sbert, An axiomatic approach to image interpolation, *IEEE TIP*, Vol. 7, No. 3, March 1998, pp. 376-386.
- [5]. P. Thevenaz, T. Blu and M. Unser, Image Interpolation and Resampling, *Academic Press, USA*, 2000.
- [6]. M. Irani and S. Peleg, Improving resolution by image registration, *CVGIP: Graphical Models and Imaging Processing*, Vol. 53, No. 3, 1991, pp. 231–239.
- [7]. L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D: Nonlinear Phenomena*, Vol. 60, No. 1-4, 1992, pp. 259–268.
- [8]. T. Chan, S. Esedoglu, F. Park and A. Yip, Recent developments in total variation image restoration, in *Mathematical Models of Computer Vision*, N. Paragios, Y. Chen and O. Faugeras, Eds., *Springer Verlag*, New York, 2005.
- [9]. S. Dai, M. Han, W. Xu, Y. Wu and Y. Gong, Soft edge smoothness prior for alpha channel super resolution, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Classification (CVPR)*, 2007, pp. 1–8.
- [10]. J. Sun, Z. Xu and H. Shum, Image super-resolution using gradient profile prior, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.
- [11]. W. T. Freeman, E. C. Pasztor and O. T. Carmichael, Learning low-level vision, *International Journal of Computer Vision*, Vol. 40, No. 1, 2000, pp. 25–47.
- [12]. W. T. Freeman, T. R. Jones and E. C. Pasztor, Example-based super-resolution, *IEEE Computer Graphics and Applications*, Vol. 22, 2002, pp.56–65.
- [13]. Q. Wang, X. Tang and H. Shum, Patch based blind image super resolution, in *Proceedings of the ICCV*, 2005.
- [14]. C. Liu, H. Y. Shum and W. T. Freeman, Face hallucination: theory and practice, *IJCV*, Vol. 75, No. 1, October 2007, pp. 115-134.
- [15]. H. Chang, D.-Y. Yeung and Y. Xiong, Super-resolution through neighbor embedding, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Classification (CVPR)*, Vol. 1, 2004, pp. 275–282.
- [16]. J. Yang, J. Wright, T. Huang, and Y. Ma, Image super-resolution as sparse representation of raw image patches, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.
- [17]. J. Yang, J. Wright, T. S. Huang and Y. Ma, Image super-resolution via sparse representation, *IEEE Transactions on Image Processing*, Vol. 19, No. 11, 2010, pp. 1–8.
- [18]. Chen S, Donoho D, Saunders M, Atomic decomposition by basis pursuits, *SIAM Journal on Scientific Computing*, Vol. 20, 1999, pp. 33-61.
- [19]. H. Lee, A. Battle, R. Raina and A. Y. Ng, Efficient sparse coding algorithms, in *Advances in Neural Information Processing Systems (NIPS)*, 2007.
- [20]. C. Kim, K. Choi, J. B. Ra, Improvement on learning-based super-resolution by adopting residual information and patch reliability, in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 1197-1200.