AUT Journal of Electrical Engineering

AUT J. Elec. Eng., 53(2) (2021) 249-260
DOI: 10.22060/eej.2021.19611.5403

Learning Curvelet-based Directional Dictionaries for Single Image Super Resolution

E. Mikaeli[1*], A. Aghagolzadeh[2], M. Nooshyar[3]

[1] Instructor, Faculty of Engineering, Mohaghegh Ardabili University, Ardabil, Iran.
[2] Professor, Faculty of Electrical and Computer Engineering, Babol Noshirvani University of Technology,
[3]Associate Professor, Faculty of Engineering, Mohaghegh Ardabili University, Ardabil, Iran

ABSTRACT: Learning and reconstruction-based methods are the two main approaches to the solve single image super resolution (SISR) problem. In this paper, to exploit the advantages of both learning based and reconstruction based approaches, we propose a new SISR framework by combining them, which can effectively utilize their benefits. The external directional dictionaries (EDD) are learned from external high quality images. Additionally, we embeded the nonlocal means (NLM) filter and an isotropic total variation (TV) scheme in the reconstruction based method. We suggest a new supervised clustering scheme via curvelet based direction extraction method (CCDE) to learn the external directional dictionaries from candidate patches with sharp edges. Each input patch is coded by all the EDD. Each of the reconstructed patches under different EDD is applied with a weighted penalty to characterize the given input patch. To disclose new details, the local smoothness and nonlocal self-similarity priors are added on the recovered patch by TV scheme and NLM filter. Extensive experimental results validate the effectiveness and robustness of the proposed method comparing with the state-of- the-art algorithms in SISR methods. Our proposed schemes can retrieve more fine structures and obtain superior results than the competing methods with the scaling factors of 2 and 3.

Review History:

Received: Feb. 09, 2021
Revised: May. 06, 2021
Accepted: May. 08, 2021
Available Online: Sep. 01, 2021

Keywords:

single image super resolution
spare representation, directional features
local smoothness
nonlocal self-similarity.

1- INTRODUCTION

The goal of the SISR is to increase the resolution of a single low-resolution (LR) image, $\mathbf{y} \in R^M$, so that the reconstructed high resolution (HR) image resembles the original HR image $\mathbf{x} \in R^N$. The SISR problem can be modeled as:

$$\mathbf{y} = \mathbf{DHx} + \mathbf{n} \tag{1}$$

where $H \in R^{N\times N}$, $D \in R^{M\times N}$ and n are the blurring filter, down-sampling matrix and additive white Gaussian noise. SISR is an ill-posed inverse problem. There are various algorithms in the literature to solve this problem. In general, these algorithms are divided into three categories: Interpolation-based algorithms, reconstruction-based algorithms and learning-based algorithms [1]. The interpolation-based methods are the basic approaches for SISR. Since this kind of methods do not consider the destruction model, the reconstructed image by this type of methods has unpleasing artifacts. The solution of SISR problem by the reconstruction-based methods can be modeled as:

$$\mathbf{x} = \arg\min_{\mathbf{x}} \frac{1}{2}\left\|\mathbf{y} - \mathbf{DHx}\right\|_2^2 + \lambda R(\mathbf{x}) \tag{2}$$

where, the first term is the fidelity, $R(\mathbf{x})$ is the regularization term which imposes the suitable priors such as local smoothness, nonlocal (NL) similarity [2], edge-directed priors [3], gradient–profile prior [4] and λ in (2) is the regularization parameter [2, 5]. The local smoothness prior guarantees the similarity of the neighboring pixels and Total variation (TV) [6] is one of the smoothing classical regularizations that is used to this purpose. The repeatation of the similar patterns in different locations of the image is the basic assumption of NL regularization. In nonlocal means (NLM) [7] approach the NL property has been efficiently utilized for denoising. Recently, the learning-based algorithms by utilizing the example images or input image itself have achieved superior results. These methods can be classified into three categories: neighbor embedding-based (NE) methods, dictionary learning-based (DL) methods and deep learning-based methods. The NE-based methods [8-13] have been inspired by the manifold learning methods. In [14], first a SR convolutional neural network (SRCNN) method was proposed that learns an end-to-end mapping between the LR and HR images for SISR. Since the number of layers in this network is only three; it is difficult to learn the complex structures. A very deep super-resolution (VDSR) [15] network with 20 layers in a cascade deep network is one of the first methods that utilizes the residual learning to

*Corresponding author's email:ez.mikaiili@gmail.com

train VSDR network for SISR. In the mentioned methods, the LR input images are interpolated before utilizing them into networks; thus the obtained end-to-end mappings aren't learned from the original LR to HR images. To solve this problem, Dong et al. [16] utilized the original LR images as the input to the network. Zhang et al. in [17] proposed a super-resolution network for multiple degradation (SRMD), which its degradation maps are utilized in addition to the LR image as input. Despite the superiority of the obtained results by neural network-based methods, their learning procedure is complicated with high computation complexity. Dictionary learning algorithms [18-23] were proposed by utilizing this fact that natural images can be sparsely represented by the atoms of a dictionary. In [18] and [19], two over complete LR-HR dictionaries are learned whose atoms are the raw training image patches, and are learned by the sparsity constraints. In [20], by considering the instability of the sparse decomposition over a highly redundant dictionary, an adaptive sparse domain selection (ASDS) scheme for sparse representation using complete dictionaries, was proposed. In order to improve the ASDS, nonlocal centralized sparse representation (NCSR) model [21] was proposed with a superior performance. A statistical prediction model based on sparse representation of LR and HR image was proposed in [23]. Some fast NE-based learning methods are proposed in [9-13] . In [9] the anchored neighbor regression (ANR) method was proposed that combines NE algorithms with the sparse learned dictionaries. A+ was proposed in [10] that improves the ANR. Despite the advantages of both the learning and reconstruction based methods, there are some drawbacks. For instance, blured fine details and unexpected artifacts may appear in the estimated HR image by the reconstruction-based and learning- based methods. In this paper, to exploit the advantages of both learning based and reconstruction based approaches, we propose a new SISR framework by combining them, which can effectively utilize their benefits. In our proposed method, the external directional dictionaries (EDD) are learned by the external high quality images; also NLM filter and isotropic total variation (TV) scheme [24] have been utilized in the reconstruction based method. Despite the suitable results obtained by k-means clustering method to classify the patches, this scheme is unsupervised that different initial clusters may lead to the different final groups. To overcome this deficiency, we have developed a supervised clustering technique using explicit features of the image, thus extra details such as edges and textures can be appropriate. To this end, we have proposed directional clustering scheme. The candidate patches with sharp edges are utilized to extract their orientations via curvelet based direction extraction method [25]. The obtained patches with dominant orientations are clustered into groups according to the estimated directions and their corresponding external directional dictionaries are learned by k-svd approach [26]. This paper is the extended and completed version of the paper presented earlier in [34]. The contributions of our proposed method are listed as follows:

1)We propose a new directional clustering method via curvelet based direction extraction method (CCDE), and the external directional dictionaries (EDD) of each cluster is learned by the k-svd method.
2)Each given input patch is reconstructed under all EDD. The obtained patches are applied with weighted penalty to characterize the given input patch.
3)To enforce the local smoothness and the repeated details constraints on the recovered patches, the isotropic total variation (TV) scheme and the nonlocal means (NLM) filter are leveraged.

The rest of this paper is organized as follows:

In section 2, we review the background on local smoothness modelling by anisotropic TV and curvelet based direction extraction method. In section 3, the proposed supervised clustering scheme via curvelet based direction extraction method (CCDE) is introduced and its implementation in a new SISR technique is illustrated. In section 4, the obtained results are reported and the effectiveness of our proposed method is confirmed through experiments, and finally section 5 concludes the paper. In the following of the matrixes, vectors and scalar values are shown by bold and uppercase, bold and lowercase and italic and lowercase notifications.

2- LOCAL SMOOTHNESS MODELING BY ANISOTROPIC TV AND CURVELET-BASED DIRECTION EXTRACTION

Due to the similarity of the intensities of the neighboring pixels in natural images, the output of the vertical and horizontal highpass filters are close to zero and their marginal distributions are very sharp. These marginal distributions can be modeled by Laplacian distribution [27]. By definition $\boldsymbol{D} = \left[D_v ; D_h \right]$ where, $D_V = \left[1, 0, -1\right]^T$ and $D_h = \left[1, 0, -1\right]$ are the vertical and horizontal highpass filters, the regularization term $R(\mathbf{x})$ that describes the local smoothness prior of image $\mathbf{x}$ at pixel level is modeled as follows:

$$R(\mathbf{x}) = \left\| \boldsymbol{D}\mathbf{x} \right\|_1 = \left\| D_v \mathbf{x} \right\|_1 + \left\| D_h \mathbf{x} \right\|_1 \tag{3}$$

Where, $D_v \mathbf{x}$ and $D_h \mathbf{x}$ are gradient pictures in vertical and horizontal directions that are obtained by applying the mentioned filters on image $\mathbf{x}$ through convolution operator. By utilizing the above $R(\mathbf{X})$ in (2), it can be considered as the statistical convex explanation of anisotropic TV [24].

Inspired by this fact that the patches in the natural images with sharp edges are adjusted in the limited number of directions, in [25] these directional information is exploited by applying the curvelet transform Γ on the image X:

$$\begin{aligned} &Q = \Gamma(\mathbf{X}), \\ &\left\{ \mathbf{Q}_{j,l} \,\middle|\, j = 1,...,J; l = 1,...,L_j; J = 5, L_5 = 64 \right\} \end{aligned} \tag{4}$$

where $Q_{j,l}$ is the curvelet coefficient matrix at the j th scale and l th direction. By considering the property of the directional symmetry, the fifth scale coefficients with 64 matrices are partitioned into 16 direction subsets $\{Z_k\}_{k=1}^{16}$ which are shown in Fig.1. By utilizing $\{Z_k\}_{k=1}^{16}$, the efficient directional features $\{A_k\}_{k=1}^{16}$ of the input image are defined as the following expression:

$$\mathbf{A}_k = \mathrm{P}(\Gamma^{-1}(\mathbf{H}_k \odot \mathbf{Q})), k = 1,..,16, \mathbf{H}_k \odot \mathbf{Q} = \left\{\mathbf{H}_{k,5,l}.\mathbf{Q}_{5,l} \,\middle|\, l = 1,...,64\right\}, \quad (5)$$

$$\mathbf{H}_{k,5,l} = \begin{cases} 1 \rightarrow l \in \mathbf{Z}_k \\ 0 \rightarrow O.W \end{cases}$$

where $P(.) = abs(\mathrm{Re}(.))$ and Γ^{-1} is the inverse curvelet transform. $\mathbf{v}_i = \left[\mathbf{A}_{1,i}, \mathbf{A}_{2,i}, ..., \mathbf{A}_{16,i}\right]^T$ is considered as a directional feature vector of i th pixel.

3- PROPOSED METHOD

As mentioned earlier, using dictionaries which have meaningful information of the main features of the image can lead to superior image restoration results. The external dictionaries that are learned by external high quality images, can reconstruct the features that rarely exist in the input image. In this paper, inspired by the fact that the sharp edges are oriented in one of 16 directions [25], the external dictionaries are learned via supervised clustering. By considering the SISR problem, we propose a novel scheme by combining the learning based and reconstruction based approaches which can effectively model the local smoothness and nonlocal self-similarity in a unified framework. This algorithm consists of three parts: 1) learning the directional dictionaries for external high quality training patches with sharp edges 2) obtaining the sparse representation of each input pach over the learned external directional dictionaries 3) reconstruction of the goal patch by the estimated sparse representations with weighted penalty, isotropic TV and NLM regularizations. In the following, our proposed methods for external directional dictionary learning and SISR are illustrated.

3- 1- Directional Dictionary Learning by Clustering Scheme via Curvelet-based Direction Extraction Method

In order to learn directional sub-dictionaries, we apply the curvelet based direction extraction method [25] on the set of high-quality training natural images which are provided in [18]. The directional feature, $\mathbf{A}_k, 1 \le k \le 16$ for each pixel are obtained. Each training images is decomposed into the overlapped patches with the size of $\sqrt{n} \times \sqrt{n}$. The smooth patches with intensity variances smaller than a threshold value $,\delta_1,$ are excluded. Indeed, only patches with meaningful structures are used in dictionary learning. To estimate the dominant orientation of each of the selected patches, the average energy, ε_k, for each directional features, $\mathbf{A}_k$, is calculated by (6):

$$\varepsilon_k = \sum_{i=1}^{\sqrt{n}} \sum_{j=1}^{\sqrt{n}} \frac{\mathbf{A}_{k(i,j)}}{n}, 1 \le k \le 16 \quad (6)$$

The main direction of each patch is determined by the maximum value $\varepsilon_k, 1 \le k \le 16$. Similarly [28], if the distance between the average energy of two dominant orientations is smaller than a threshold value, δ_2, the patch is considered as a stochastic patch and when this difference is larger than the threshold value, the patch has dominant orientation. Therefore, only patches with a specific orientation are selected. The selected training patches with dominant orientations are grouped into sixteen directions $\left\{\mathbf{C}^1,...,\mathbf{C}^{16}\right\}.\left\{\mathbf{\Phi}_{E,1},...,\mathbf{\Phi}_{E,16}\right\}$ are considered as external directional dictionaries (EDD) for patches in $\left\{\mathbf{C}^1,...,\mathbf{C}^{16}\right\}$,. The k-svd [26] algorithm is utilized to learn these dictionaries. The superiority of our proposed method in classifying the patches with geometrical details over k-means clustering method is shown in Fig.(2). According to this figure, we observe that the patches in each proposed clusters (Fig. 2 (a)-(h)) indicate the specific oriantation and in comparison with K-means clustering ((Fig. 2 (i)-(l)) their geometric information are more compatible with each other.

3-2- Single Image Super Resolution via EDD

Similar to other schemes, the initial estimation of the goal HR image $\mathbf{X}^l$, $l = 0,$ is obtained by applying the bicubic interpolation on the LR input image $\mathbf{Y}$. Our proposed SISR problem by inserting two regularization terms into (2) is formulated as follows:

$$\arg\min_{\mathbf{x}} 0.5\left\|\mathbf{y} - \mathbf{DHx}\right\|_2^2 + \lambda_1 R_{EDD} + \lambda_2 R_{NLM} \quad (7)$$

Where, the first term is the fidelity term, R_{NLM} is the NLM regularization that is used to confirm the nonlocal self-similarity prior and R_{EDD} is our proposed regularization term that is used to reconstruct the sharp and directed edges. In the following, after introducing these two regularization terms, to insert the local smoothness prior into Eq. In (7), an auxiliarly variable is considered and finally the procedure of solving this minimization problem is explained.

For each overlapped $\sqrt{\beta} \times \sqrt{\beta}$ patch $\mathbf{X}_i\}_{i=1}^n$ (n is the number of overlapped patches in the $\mathbf{X}^l$), the similar patches of $\mathbf{X}_i, \mathbf{X}_j,$ are selected according to large value of w_{ij}:

$$w_{ij} = \exp(-\frac{\| \mathbf{x}_i - \mathbf{x}_j \|^2}{h}) \quad (8)$$

Where, $\mathbf{x}_i \in R^\beta$ and $\mathbf{x}_j \in R^\beta$ are the column vectors of $\mathbf{X}_i$ and $\mathbf{X}_j$. By obtaining w_{ij}, the formulation of NLM regularization term is modeled by the following expression [7]:

$$R_{NLM}(\mathbf{x}) = \sum_{i=1}^{n} \left\| \mathbf{x}_i - \sum_{j \in \Omega(\mathbf{x}_i)} w_{ij} \mathbf{x}_j \right\|_2^2 \quad (9)$$

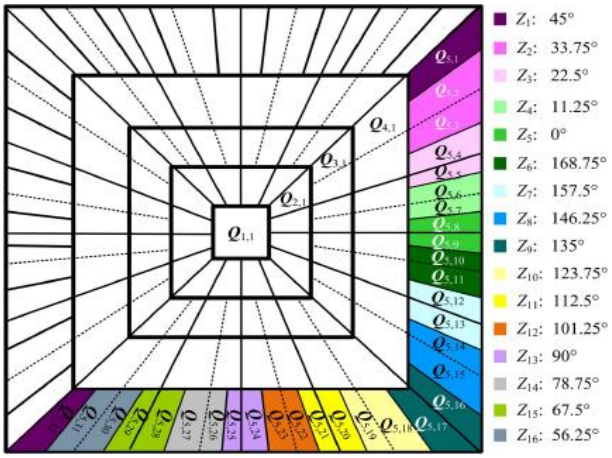

Fig.1. Curvelet coefficient matrices at different scales and its partition of 16 different direction subsets [25]

Where, $\Omega(\mathbf{x}_i)$ denotes the patches similar t. $\mathbf{x}_i$ Our proposed R_{EDD} regularization term is as below:

$$R_{EDD}(\mathbf{x}) = \left\{ \begin{array}{l} \sum_{i=1}^{n} \left\| \boldsymbol{\alpha}_{E,i}^{k} \right\|_1 \\ s.t, \dot{\mathbf{x}}_i = \sum_{k=1}^{16} c_k . (\boldsymbol{\Phi}_{E,k} \boldsymbol{\alpha}_{E,i}^{k}), 1 \le i \le n \end{array} \right\} \quad (10)$$

Where, $\boldsymbol{\alpha}_{E,k}^{i}$ is the sparse representation of the i^{th} input patch $(\mathbf{x}_i)$ over k[th] learned EDD $(\boldsymbol{\Phi}_{E,k})$ and

$\dot{\mathbf{x}}_i$ is the reconstruction of the $\mathbf{x}_i$ over all learned EDDs $\{\boldsymbol{\Phi}_{E,k}\}_{k=1}^{16}$ with the weighted penalty (c_k). c_k depends on the similarity between $\mathbf{x}_i$ and its reconstruction over $\boldsymbol{\Phi}_{E,k}$ $(\boldsymbol{\Phi}_{E,k}\boldsymbol{\alpha}_{E,i}^{k})$. For weak similarity between $\mathbf{x}_i$ and $(\boldsymbol{\Phi}_{E,k}\boldsymbol{\alpha}_{E,i}^{k})$, we force small value for c_k, that will be shown.

To depict the local smoothness, an auxiliary variable $\mathbf{u}$ that is equal to $\mathbf{x}$ is considered and (7) can be rewriten as follows:

$$\arg\min_{\mathbf{x},\boldsymbol{\alpha}_{E,i}^{k},\mathbf{u}} 0.5\left\|\mathbf{y}-\mathbf{D}\mathbf{H}\mathbf{x}\right\|_2^2 + \quad (11)$$

$$\lambda_1 \sum_{i=1}^{n} \| \boldsymbol{\alpha}_{E,i}^{k} \|_1 + \lambda_2 \sum_{i=1}^{n} \| \mathbf{x}_i - \sum_{j \in \Omega(\mathbf{x}_i)} w_{ij} \mathbf{x}_j \|_2^2$$

$$+\lambda_3 \| D\mathbf{u} \|_1, \forall k \le 16$$

$$s.t: \dot{\mathbf{x}}_i = \sum_{k=1}^{16} c_k . (\boldsymbol{\Phi}_{E,k} \boldsymbol{\alpha}_{E,i}^{k}), \ \mathbf{x} = \mathbf{u}, \ \forall i \le n$$

where, $\left\|D\mathbf{u}\right\|_1$ denotes the anisotropic TV of image $\mathbf{u}$. The minimization problem in (11) can be solved iteratively using the following steps:

u Sub problem:

By fixing $\mathbf{x}$, we can rewrite (11) as:

$$\mathbf{u}^{(l+1)} = \arg\min_{\mathbf{u}} 0.5 \| \mathbf{x}^{l} - \mathbf{u}^{l} \|_2^2 + \lambda_3 \left\|D\mathbf{u}\right\|_1, l = 1,..,L \quad (12)$$

where L is the number of iterations and $\mathbf{x}^{l}$ can be interpreted as the noisy observation of $\mathbf{u}^{l}$ at iteration l, and (12) can be regarded as anisotropic TV-based denoising problem [24]. For convenience, the superscript l has been dropped without loss of generality. The proximal operator $(prox_t(g)(\mathbf{u}))$ of a proper closed convex function g can be defined as:

$$prox_t(g)(\mathbf{u}) = \arg\min_{\mathbf{u}} \left\{ 0.5\left\|\mathbf{x}-\mathbf{u}\right\|_2^2 + tg(\mathbf{u}) \right\} \quad (13)$$

Where $t > 0$ is a scalar parameter. The $\mathbf{u}$ sub-problem can be regarded as the proximal map due to $g(\mathbf{u}) = \| D\mathbf{u} \|_1$. We utilize FIST [24] with the fixed number of iterations to solve (13).

$\boldsymbol{\alpha}_{E,i}^{k}$ Sub problems:

By fixing $\mathbf{x}$ and $\mathbf{u}$, minimizing (11) with respect to $\boldsymbol{\alpha}_{E,i}^{k}$ leads to:

$$\arg\min_{\boldsymbol{\alpha}_{E,i}^{k}} 0.5\left\|\mathbf{y}_i - \mathbf{D}\mathbf{H}(\boldsymbol{\Phi}_{E,k}\boldsymbol{\alpha}_{E,i}^{k})\right\|_2^2 + \quad (14)$$

$$\lambda_1 \| \boldsymbol{\alpha}_{E,i}^{k} \|_1, \ 1 \le i \le n, \ 1 \le k \le 16$$

where, $(\mathbf{y}_i)$ is the input of LR patch, $(\boldsymbol{\Phi}_{E,k}\boldsymbol{\alpha}_{E,i}^{k})$ is its corresponding HR patch over $\boldsymbol{\Phi}_{E,k}$. The sub-problem in (14) can be solved efficiently by the orthogonal matching pursuit (OMP) algorithm. By considering the subject in (11), the reconstructed patch $\dot{\mathbf{x}}_i$ over all EDDs $\{\boldsymbol{\Phi}_{E,k}\}_{k=1}^{16}$ with the recovery weight c_k is expressed as:

Table 1.EDD_based image super resolution algorithm.

Input : LR input image, directional external learned dictionaries $\{\mathbf{\Phi}_{E,1},...,\mathbf{\Phi}_{E,16}\}$, **Output:** HR image $\boldsymbol{x}$

Initialization:

set l=0, set $x^{(0)}$ as the initial estimation of HR target image by bicubic interpolation;
set initial regularization parameters $\delta_1 = 30, \delta_2 = 1.5, \lambda_3 = 5, \lambda_1 = 0.01, \gamma_1 = 2, \gamma_2 = 0.01$

Repeat

$\mathbf{x}^{(l+1)} = \mathbf{x}^{(l)}$

Reconstruction $u^{(l+1)}$ via anisotropic TV-based denoising problem by considering $x^{(l)}$ as the noisy observation of it**.**

a. Definition: $g(u) = \| Du \|_1$ **b** . $u^{(l+1)} = prox_{\lambda_3}(g)(x^{(l)})$

Update $\mathbf{x}^{(l+1)}$

1. **Reconstruction** $\dot{\mathbf{x}}_i\}_{i=1}^{n}$ over all directional dictionaries with the weighted penalty

   for each patch $\mathbf{x}_i\}_{i=1}^{n}$

   reconstruct $\boldsymbol{\alpha}_{E,i}^{k}$ by computing Eq.(14) over learned EDDs $\mathbf{\Phi}_{E,k}\}_{k=1}^{16}$, **reconstruct** $\dot{\mathbf{x}}_i$ by computing Eq. (15).

   end for
2. **Update** $\boldsymbol{x}_i$ by computing Eq.(17)
3. **Reconstruct** $\mathbf{x}^{(l+1)}$ by averaging all reconstructed $\mathbf{x}_i^{l+1}$

l=l+1; **until** $l > \max Iter$ or $\frac{\| \mathbf{x}^l - \mathbf{x}^{l+1} \|_2^2}{N} < \varepsilon$

$$\dot{\mathbf{x}}_i = \sum_k c_k \mathbf{\Phi}_{E,k} \boldsymbol{\alpha}_{E,i}^{k} \rightarrow c_k = \frac{1}{16-1}\left[1 - \frac{\| \mathbf{\Phi}_{E,k} \boldsymbol{\alpha}_{E,i}^{k} - \mathbf{x}_i \|_2^2}{\sum_{k=1}^{16} \| \mathbf{\Phi}_{E,k} \boldsymbol{\alpha}_{E,i}^{k} - \mathbf{x}_i \|_2^2}\right] \quad (15)$$

By (15), it is obvious that for a weak similarity between $\mathbf{x}_i$ and $(\mathbf{\Phi}_{E,k}\boldsymbol{\alpha}_{E,i}^{k})$, we force small value for c_k.

$\mathbf{x}$ Sub problem:

By fixing the obtained $\boldsymbol{\alpha}_{E,i}^{k}$ and $\mathbf{u}$, the $\mathbf{x}$ sub problem is written for all the patches $\mathbf{x}_i$ as follows:

$$\ddot{\mathbf{x}}_i = \arg\min_{\mathbf{x}_i} \left\| \mathbf{x}_i - \sum_{j \in \Omega(\mathbf{x}_i)} w_{ij} \mathbf{x}_j \right\|_2^2 + \gamma_1 \left\| \mathbf{x}_i - \mathbf{u}_i \right\|_2^2 + \gamma_2 \left\| \mathbf{x}_i - \dot{\mathbf{x}}_i \right\|_2^2 \quad (16)$$

where, $\mathbf{u}_i$ and $\dot{\mathbf{x}}_i$ are the i^{th} patch of $\mathbf{u}$ and the reconstructed patch over all EDDs with the recovery weight.

The first term in (16) indicates the nonlocal self-similarity property which is used to reconstruct the more noticeable repeated features in the image. The second term ensures the local consistency prior in the image which is used to confirm the similarity of the intensities of the neighboring pixels and the third term preserves the local geometry of each patch which is used to reconstruct the sharp edges in the image. The $\mathbf{x}_i$ sub problem in (16) has a closed form solution as follows:

$$\ddot{\mathbf{x}}_i = \left[(\gamma_1 + \gamma_2 + 1)\mathbf{I}\right]^{-1} . \left[\mathbf{S}_j (\mathbf{w}_i)^T + \gamma_1 \mathbf{u}_i + \gamma_2 \dot{\mathbf{x}}_i\right] \quad (17)$$

Where $\mathbf{w}_i = \left[w_{i1},...,w_{i|\Omega(\mathbf{x}_i)|}\right]$ that its elements is the obtained by (8) and $\mathbf{S}_j \in R^{\beta \times |\Omega(\mathbf{x}_i)|}$ is a matrix, which its columns are the similar patches to $\mathbf{x}_i$. Finally, the goal image $\mathbf{X}$ is reconstructed by replacing all the reconstructed patches in their corresponding positions and their average in the overlapped areas. The mentioned seperated sub problems are solved iteratively, until a stopping criteria is satisfied. The detailed description of the proposed SISR using EDD is summerized in Table 1.

Table 2. Ablation analysis in terms of average PSNR/SSIM on Set14 database.

Settings	Different combinations of Patch_based, Group_based and NLM			
anisotropic TV	✓	✓		✓
NLM	✓		✓	✓
CCDE		✓	✓	✓
PSNR/SSIM	28.26/0.7873	28.54/0.8086	29.30/0.8212	29.51/0.8231

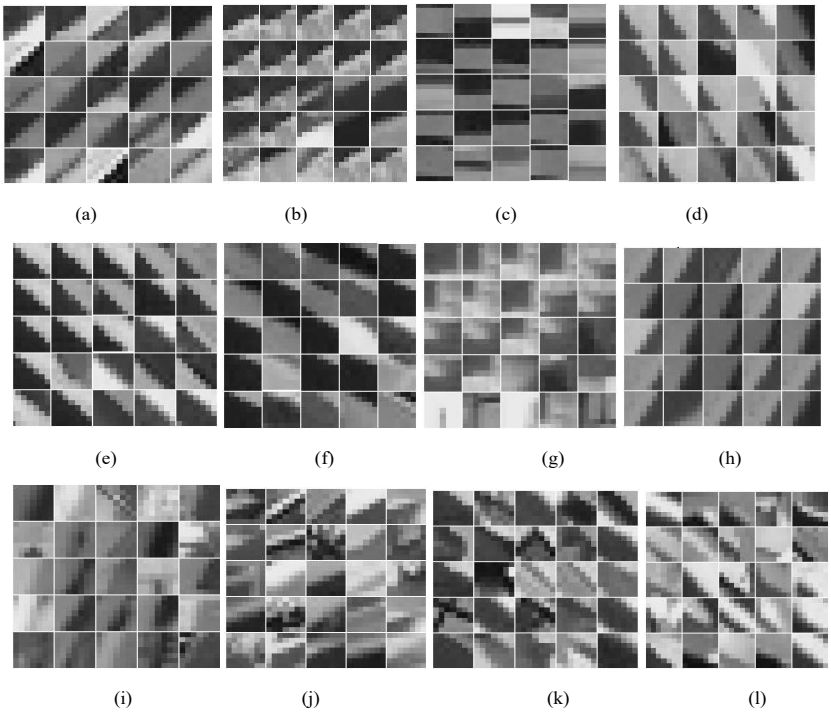

Fig.2. obtained clustered patches via the proposed EDD method and k_means clustering method: (a) subset , patches, (b) subset , patches, (c) subset , patches, (d) subset , patches, (e) subset , patches, (f) subset , patches (g) subset , patches, (h) subset , patches, (i) cluster 1, (j) cluster 2, (k) cluster 3, (l) cluster 4.

4- EXPERIMENTAL RESULTS

All the experiments have been implemented using MATLAB 2014a, an Intel ® Core™ i7, 2.4GHz processor with an 8 GB of RAM and a 64-bit Windows operating system. Four benchmark datasets are used for testing in our expriments: Set5 [29], Set14 [30], BSDS100 [31] and Urban100 [32]. In our experiments, the LR images are obtained by applying a 7×7 Gaussian kernel with the standard deviation of 1.6 on the HR test images and downsampled by the scaling factors of 2 and 3. In all experiments, the size of the HR patch is 7×7 with a 4 pixels overlap. The performance of the proposed methods are evaluated by the peak signal to noise ratio (PSNR) and structural similarity (SSIM) [33] between the luminance components of test images and the corresponding reconstructed ones. The basic parameters in our proposed algorithm are set as follows: $\delta_1=30,\delta_2=1.5,\lambda_3=5,\lambda_1=0.01,\gamma_1=2,\gamma_2=0.01$. h=75 and maxIter=150. The number of similar patches is set to 12. The searching window and the size of learned dictionaries are set to 20×20 and 500.

4-1- Ablation Analysis

We demonstrate the effect of each regularization terms in our proposed method, by using an ablation analysis. This analysis, in terms of the average PSNR/SSIM on Set14 dataset, is shown in Table.2. The PSNR/SSIM values are limited to (28.26/0.7873) with anisotropic TV and NLM terms. The anisotropic TV and the proposed EDD scheme, change the results to (28.54/0.8086). For (NLM term and EDD simotanously), the quantative results are improved to (29.30/0.8212).

Combination of the NLM with anisotropic TV term and EDD achieves the best results (29.51/0.8231) . In Fig.3, we compared visually the effect of the various settings on some test images. The anisotropic TV and NLM lead to smooth results and the finer structures cannot be reconstructed. The anisotropic TV and the proposed EDD scheme fail to recover the texures and its reconstructed images have some ringing artifacts. NLM and EDD scheme obtain much better results with fine structures. Combination of all the regularization terms, achieves the best pleasing textures, details and edges with low artifacts.

Table 3. PSNR (dB) / SSIM results for various SISR methods with the scaling factors of 2 and 3. (The best results are shown in red color and the second best performance are appeared by blue color)

Method	Scale	Set5	Set14	BSDS100	Urban100
Bicubic	×3	28.78/0.8308	26.38/0.7271	26.33/0.6918	23.52/0.6862
A+	×3	29.75/0.8420	27.35/0.7549	27.15/0.7226	24.23/0.7413
NCSR	×3	32.46/0.9041	28.82/0.8165	28.14/0.7930	26.39/0.8150
FSRCNN	×3	26.23/0.8124	24.44/0.7106	24.86/0.6832	22.05/0.6745
SRCNN	×3	32.05/0.8944	28.80/0.8074	28.13/0.7736	25.70/0.7770
VDSR	×3	33.25/0.9150	29.46/0.8244	28.57/0.7893	26.60/0.8136
SRMD	×3	33.85/0.9242	29.91/0.8334	28.89/0.7980	26.60/0.8192
Proposed	×3	33.32/0.9167	29.51/0.8231	28.93/0.7950	27.20/0.8211
Bicubic	×2	28.91/0.8354	0.7359/28.12	27.42/0.7543	25.83/0.7250
A+	×2	30.46/0.8636	29.44/0.8063	28.15/0.7771	26.34/0.7726
ASDS	×2	34.32/0.9250	30.83/0.8635	29.52/0.8522	28.69/0.8483
NCSR	×2	34.51/0.9293	30.77/0.8619	28.96/0.8537	28.48/0.8462
SRCNN	×2	34.12/0.9122	30.14/0.8542	28.55/0.8324	27.45/0.8227
Proposed	×2	35.03/0.9314	30.98/08684	30.12/0.8602	29.31/0.8623

4-2- Comparison with Competitive Techniques

We compare the proposed SISR method both subjectively and objectively with the Bicubic, A+ [10], NCSR [21], SRCNN [14], FSRCNN [16], VDSR[15] and SRDM [17]. However, in 2×magnification, our proposed methods are compared with Bicubic, A+ [10], ASDS [20], NCSR [21] and SRCNN [14]. PSNR/ SSIM for all methods and for two scaling factors of 2 and 3 are reported in Table 3. For the scaling factor of 3, it can be seen that our proposed method outperforms all the current non-CNN-based methods (A+, ASDS, NCSR), SRCNN , FSRCNN and VDSR on the most benchmarked datasets. Furthermore, the proposed method achieves the highest PSNR/SSIM and PSNR over the SRDM on Urban100 dataset and BSDS100. The average PSNR/ SSIM gains of SR-EDD over Bicubic, A+, NCSR, FSRCNN, SRCNN, VDSR and SRMD on Urban100 dataset can be up to 3.68/0.1349, 2.97/0.0798, 0.81/0.0061, 5.15/0.1466, 1.5/0.0441, 0.6/0.0075 and 0.6/0.0019. The average PSNR/ SSIM gains of SR_EDD over Bicubic, A+, NCSR, ASDS, FSRCNN, SRCNN, FSRCNN and SRMD on BSDS100 dataset can be up to 2.6/0.1052, 1.78/0.0744, 0.79/0.004, 4.07/0.1138, 0.8/0.0234, 0.36/0.0077 and 0.04/-0.001. In 2 ×magnificaction, our SR_EDD algorithm achieves the best results compared to the other methods on all datasets. For instance, the average PSNR of 29.31 dB and the average SSIM of 0.8623 offers the best SR performance on Urban100. ASDS have the second best performance. The proposed SR_CCDE has an improvement of 0.62dB in PSNR and 0.014 in SSIM over the average results of ASDS. The visual comparison between our proposed algorithm and the other methods for the scaling factor of 3 for test images 'baby' and '106024' are shown in Figs. 4 and 5. The objective comparisons show that Bicubic interpolation generates the worst results. A+, FSRCNN and SRCNN generate results with obvious unpleasant blurring artifacts. NCSR method can generate results with fair edges and textures with blurring and artifacts. VDSR obtains better results but can't recover more details and can't reduce the blurring artifacts. By alleviating the ringing and blurring effects around some edges, SR_ EDD algorithm obtains superior results than VDSR. SR_ EDD with sharper edges and more fine details achieve the superior results than all methods except for SRMD. VDSR and SRMD for generating the necessary models with complicated structures, utilize high quality images with 291 training images and DIV2K dataset with 800 training images. While our proposed method (for dictionary learning step) rely only on 91 training images, the performance of our proposed method is comparable with the VDSR and SRDM in most of the datasets. The sharper edges and finer structures illustrate the high performance of the proposed method which is imputed for using directional dictionaries for preserving the sharp edges and the proper merging the local smoothness and nonlocal self-similarity regularizations in the proposed model.

4-4- Discussions: Comparisons Between the Different Sizes of Dictionary, Algorithm Convergence and Computational Complexity

To investigate the effects of the number of the atoms of the learned dictionaries on the recovered results, we applied dictionaries with various sizes on the benchmark datasets. In Fig. 6 (left), the PSNR values for different sizes of atoms (200, 500, 700,1000) of the learned dictionary on Set5 are shown. As can be seen from this figure, the larger atoms lead to better results. In Fig 6 (right), PSNR (dB) values for the proposed CCDE-based method of test images for different iteration numbers with scaling factor $d=3$ are shown. These experiments confirm the good convergence of the proposed method which by incresing the iteration numbers, their PSNR converges to a fixed values. The complexity of the

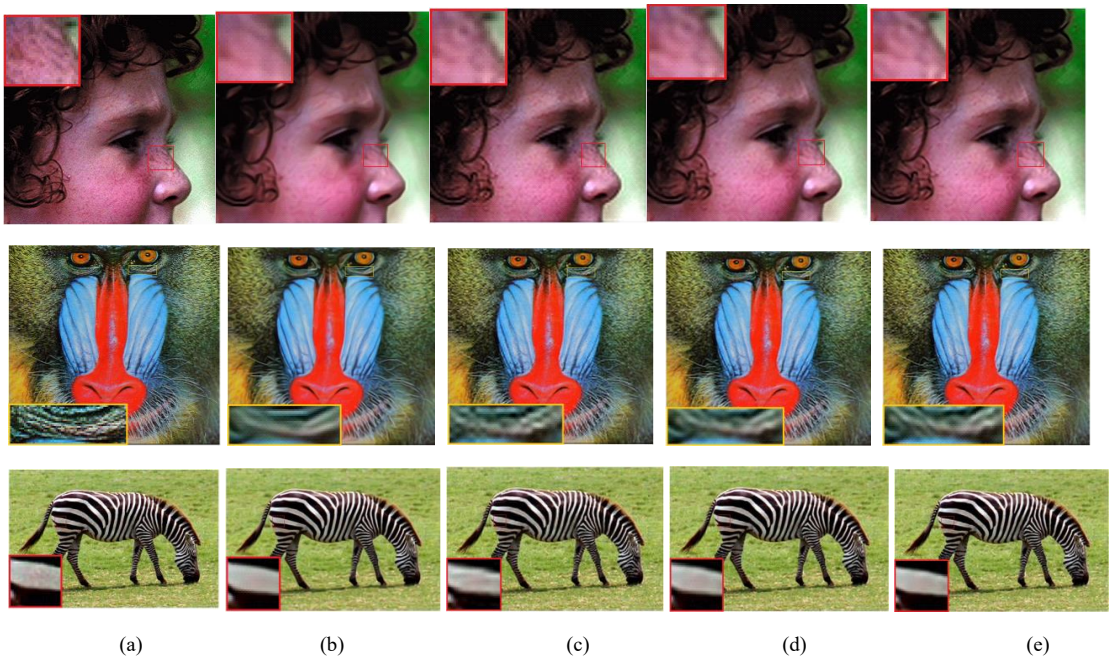

Fig.3. Investigation of the effect of each scheme in the proposed methods, on images of Set14 dataset from top to bottom. Columns from (a) to (e) represent the original HR images, obtained images with anisotropic TV and NLM, obtained images with anisotropic TV and proposed EDD, obtained images with NLM and proposed EDD and obtained images with TV, NLM and proposed EDD algorithms, respectively.

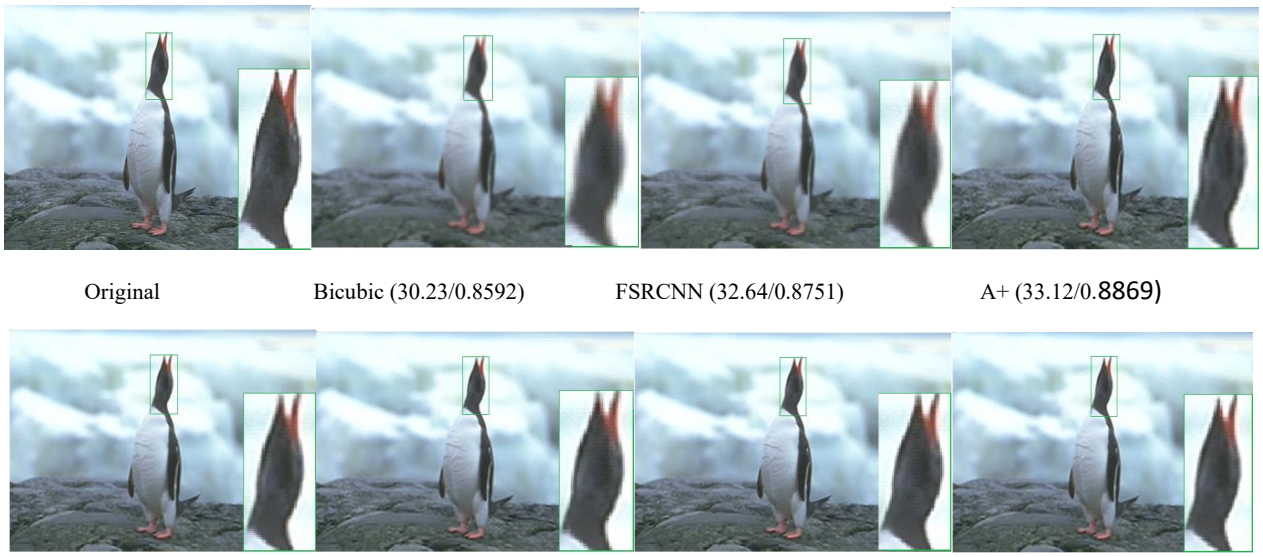

Fig.4. Visual comparisons between different SR methods for the image *'106024'* of BSDS100 dataset with scaling factor of 3.

CCDE-based method comes from five sources: NLM weight, u Sub problem, α_E Sub problem and x Sub problem. For an image with N pixels, computing NLM weight matrix has $O\left(Nc^2\left(n+16\right)l\right)$ operations, where c^2 is the size of the searching window, n is the image patch size, l is the number of selected nonlocal neighbors. To solve the $\boldsymbol{u}$ Sub problem, FIST is used which is computationally efficient and its computing time is about 5 seconds. The training times for the CNN-based methods depends on the execution platform. In Table 4, the time of the dictionary training and the utilized training data for our proposed method and some other methods are shown. Indeed, without complicated training procedure compared with CNN-based methods, our proposed method leads to pleasant results over most of the methods.

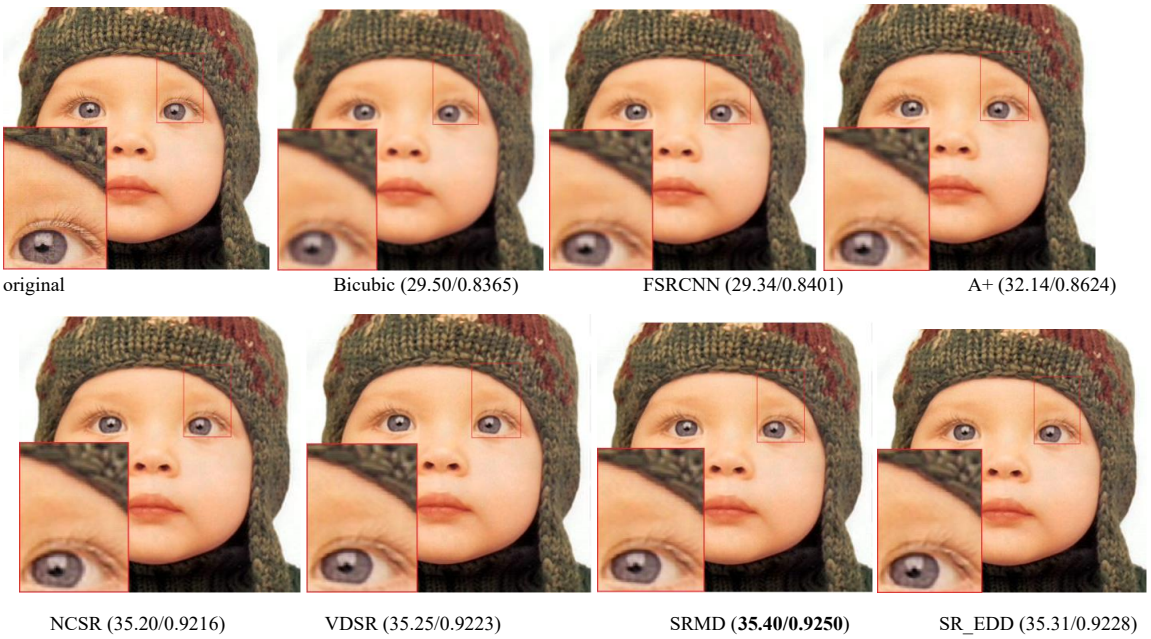

Fig.5. Visual comparisons between different SR methods for image' *baby'* of Set5 dataset with scaling factor of 3

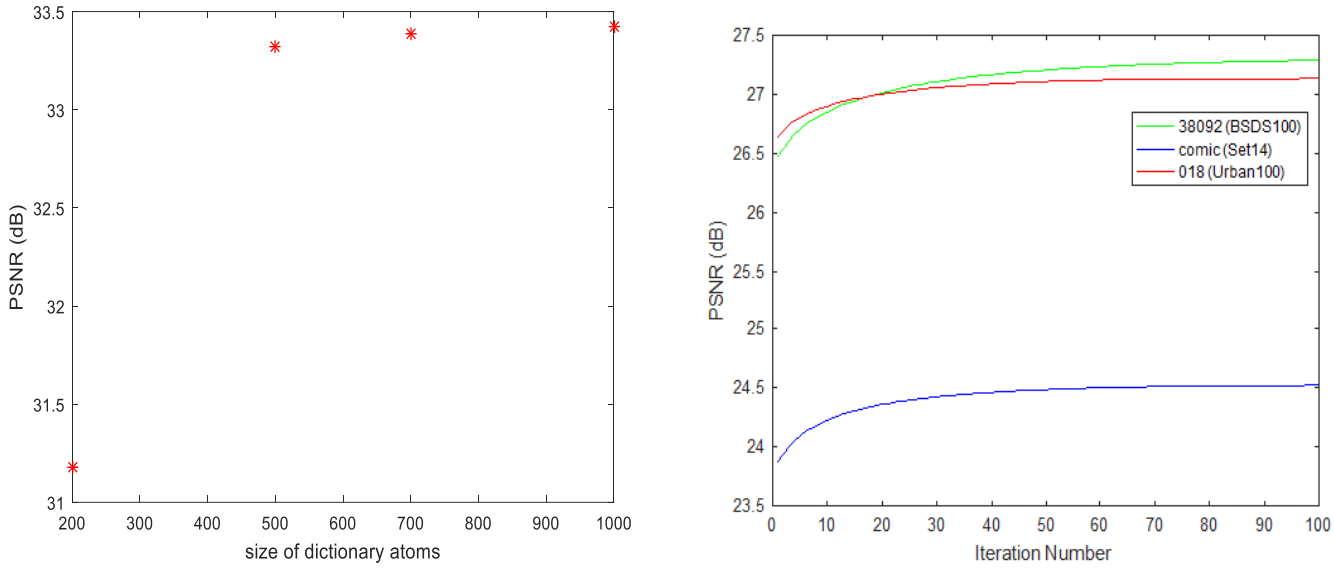

Fig. 6. (left), comparison average PSNR of our proposed method for different sizes of atoms and (right) advancement of PSNR values attained by the proposed method for test images for different iteration number for scaling factor 3 on Set5 dataset.

Table 4. Dictionary training time comparison and utilized training datasets for various SISR methods

methods	A+	SRCNN	FSRCNN	VDSR	SRMD	EDD_NLM
Dict. Training	1304s	Days	Hours	Days	Days	2490s
Train. Data	Yang91	Yang91	Yang91	G200+Yang91	DIV2K	Yang91

5- CONCLUSION

In this paper, we first introduced a new supervised clustering scheme to learn directional sub-dictionaries by applying the curvelet based directional extraction method on the set of high-quality training natural images, called CCDE. Afterwards a new scheme for high fidelity SISR was proposed. The proposed CCDE-based method efficiently preserves the sharp edges of patches in 16 main directions from $0°$ to $167.5°$ with step size of $11.25°$ and describes the local smoothness of patches and the nonlocal self-similarity, simultaneously in a unified manner. Each input patch is reconstructed by all the learned directional dictionaries and then to keep its dominant direction, a weighted average of reconstructed patches is used. The local smoothness of the patches is considered as an isotropic TV- based denoising problem, the nonlocal means (NLM) regularization is utilized to measure the similarity weight between patches according to the pixel values. Experimental results show that the superiority of the proposed methods over the state of the art methods for scaling factor 2 and 3 in PSNR and SSIM measures.

REFERENCES

[1] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, L. Zhang, Image super-resolution: The techniques, applications, and future, Signal Processing, 128 (2016) 389-408.

[2] M. Protter, M. Elad, H. Takeda, P. Milanfar, Generalizing the Nonlocal-Means to Super-Resolution Reconstruction, IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 18 (2009) 36-51.

[3] L. Wang, S. Xiang, G. Meng, H. Wu, C. Pan, Edge-Directed Single-Image Super-Resolution Via Adaptive Gradient Magnitude Self-Interpolation, IEEE Transactions on Circuits and Systems for Video Technology, 23(8) (2013) 1289-1299.

[4] Q. Song, R. Xiong, D. Liu, Z. Xiong, F. Wu, W. Gao, Fast Image Super-Resolution via Local Adaptive Gradient Field Sharpening Transform, IEEE Transactions on Image Processing, 27(4) (2018) 1966-1980.

[5] D. Tao, J. Cheng, X. Lin, J. Yu, Local structure preserving discriminative projections for RGB-D sensor-based scene classification, Information Sciences, 320 (2015) 383-394.

[6] A. Marquina, S.J. Osher, Image Super-Resolution by TV-Regularization and Bregman Iteration, Journal of Scientific Computing, 37(3) (2008) 367-382.

[7] A. Buades, B. Coll, J. Morel, A non-local algorithm for image denoising, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 60-65 vol. 62.

[8] C. Hong, Y. Dit-Yan, X. Yimin, Super-resolution through neighbor embedding, in: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., 2004, pp. I-I.

[9] R. Timofte, V. De, L.V. Gool, Anchored Neighborhood Regression for Fast Example-Based Super-Resolution, in: 2013 IEEE International Conference on Computer Vision, 2013, pp. 1920-1927.

[10] R. Timofte, V.D. Smet, L.V. Gool, A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution, in, 2014, pp. 111-126.

[11] C. Yang, M. Yang, Fast Direct Super-Resolution by Simple Functions, in: 2013 IEEE International Conference on Computer Vision, 2013, pp. 561-568.

[12] K. Zhang, B. Wang, W. Zuo, H. Zhang, L. Zhang, Joint Learning of Multiple Regressors for Single Image Super-Resolution, IEEE Signal Processing Letters, 23 (2015) 1-1.

[13] Y. Zhang, Y. Zhang, J. Zhang, Q. Dai, CCR: Clustering and Collaborative Representation for Fast Single Image Super-Resolution, IEEE Transactions on Multimedia, 18(3) (2016) 405-417.

[14] C. Dong, C.C. Loy, K. He, X. Tang, Learning a Deep Convolutional Network for Image Super-Resolution, 2014.

[15] J. Kim, J.K. Lee, K.M. Lee, Accurate Image Super-Resolution Using Very Deep Convolutional Networks, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1646-1654.

[16] C. Dong, C.C. Loy, X. Tang, Accelerating the Super-Resolution Convolutional Neural Network, 2016.

[17] K. Zhang, W. Zuo, L. Zhang, Learning a Single Convolutional Super-Resolution Network for Multiple Degradations, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 3262-3271.

[18] Y. Jianchao, J. Wright, T. Huang, M. Yi, Image super-resolution as sparse representation of raw image patches, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1-8.

[19] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image Super-Resolution Via Sparse Representation, IEEE Transactions on Image Processing, 19(11) (2010) 2861-2873.

[20] W. Dong, L. Zhang, G. Shi, X. Wu, Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization, IEEE Transactions on Image Processing, 20(7) (2011) 1838-1857.

[21] W. Dong, L. Zhang, G. Shi, X. Li, Nonlocally Centralized Sparse Representation for Image Restoration, IEEE Transactions on Image Processing, 22(4) (2013) 1620-1630.

[22] J. Li, J. Wu, H. Deng, J. Liu, A self-learning image super-resolution method via sparse representation and non-local similarity, Neurocomputing, 184 (2016) 196-206.

[23] T. Peleg, M. Elad, A Statistical Prediction Model Based on Sparse Representations for Single Image Super-Resolution, IEEE Transactions on Image Processing, 23(6) (2014) 2569-2582.

[24] A. Beck, M. Teboulle, Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems, IEEE Transactions on Image Processing, 18(11) (2009) 2419-2434.

[25] X. Li, H. He, R. Wang, D. Tao, Single Image Super-Resolution via Directional Group Sparsity and Directional Features, IEEE transactions on image processing : a

publication of the IEEE Signal Processing Society, 24 (2015).

[26] M. Aharon, M. Elad, A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Transactions on Signal Processing, 54(11) (2006) 4311-4322.

[27] J. Zhang, D. Zhao, R. Xiong, S. Ma, W. Gao, Image Restoration Using Joint Statistical Modeling in a Space-Transform Domain, IEEE Transactions on Circuits and Systems for Video Technology, 24(6) (2014) 915-928.

[28] S. Yang, M. Wang, Y. Chen, Y. Sun, Single-Image Super-Resolution Reconstruction via Learned Geometric Dictionaries and Clustered Sparse Coding, IEEE Transactions on Image Processing, 21(9) (2012) 4016-4028.

[29] M. Bevilacqua, A. Roumy, C. Guillemot, M.-L. Alberi-Morel, Low-Complexity Single Image Super-Resolution Based on Nonnegative Neighbor Embedding, (2012).

[30] R. Zeyde, M. Elad, M. Protter, On Single Image Scale-Up Using Sparse-Representations, 2010.

[31] J. Huang, A. Singh, N. Ahuja, Single image super-resolution from transformed self-exemplars, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 5197-5206.

[32] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, 2001, pp. 416-423 vol.412.

[33] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image Quality Assessment: From Error Visibility to Structural Similarity, Image Processing, IEEE Transactions on, 13 (2004) 600-612.

[34] E. Mikaeli, A. Aghagolzadeh, M. Azghani, "Single Image Super Resolution via curvelet based directional dictionaries," 11th Iranian Conf. on *Machine Vision and Image Processing (MVIP), Qom, Iran, 2020.*

HOW TO CITE THIS ARTICLE

E. Mikaeli, A. Aghagolzadeh, M. Nooshyar, Learning Curvelet-based Directional Dictionaries for Single Image Super Resolution , AUT J. Elec. Eng., 53(2) (2021) 249-260.

DOI: 10.22060/eej.2021.19611.5403