An Improved Training Sample Set Construction Method

Li Shen* and Zhai Jiaojiao
Shaanxi Information Communication Network and Security Laboratory, Xi'an University of Posts and Telecommunications, Xi'an; 710121, China

*E-mail address: 5883346@qq.com

Abstract. Sample set construction is an important step in the super-resolution reconstruction algorithm based on dictionary learning, which has an important impact on the training of dictionary and the effect of image reconstruction. When the sample set was built, the similarity between the sample blocks was not taken into account, which resulted in the redundancy of the sample set, thus increasing the time overhead of the follow-up dictionary training. To solve this problem, this paper proposes an effective method for constructing sample sets. By setting a reasonable threshold value of Euclidean distance similarity, the proposed method can ensure that the constructed sample set have structural anisotropy and diversity. When the number of pre-set blocks is the same, the training time of dictionary is reduced to about 50% of the original method, and the quality of image reconstruction is improved.

1. Introduction
The super-resolution reconstruction algorithm for multi-frame images usually considers the correlation between frames, thereby extracting prior knowledge for reconstruction. For the single image super-resolution reconstruction algorithm, the best rebuild effect is based on the learning method[1-5]. The common SR reconstruction methods based on learning include neighbourhood embedding, sparse representation, convolutional neural network (CNN) and generative adversarial network(GAN). Sparse representation (ScSR) assumes that a low resolution image block as input can be represented by a linear sparse representation of elements in an over-complete dictionary, and is reconstructed based on a high and low resolution image dictionary. The rebuild image’s quality is good, but the efficiency of calculation is low. Timofte et al.[6] combines sparse coding with neighbourhood embedding to improve the speed and effectiveness of reconstruction. After that, Dong et al.[4] realized the SR reconstruction based on the deep convolution network in 2014 to directly learn the end to end mapping relationship between the low resolution image and the high resolution image, the running efficiency and the reconstruction effect is better, but the training time is longer, and the over fitting problem can be found in the learning process. The generative adversarial network is rising in supervised learning[7]. Justin[8] and Ledig[9] introduce GAN to the problem of super-resolution reconstruction. Through the joint training of generating network and antagonistic network, the performance of super-resolution reconstruction is greatly improved, and the reconstructed image is significantly improved in visual effect.

In fact, as a reconstruction of a single image, it can also be regarded as the limit of multiple image reconstruction. It is also an ill-posed inverse problem and needs to be solved using regularization theory. As a new image representation method, sparse representation model uses a new method to solve the problem of single image super-resolution reconstruction[10-12]. One of the most effective
super-resolution reconstruction algorithm is proposed and improved by Yang[2-3]. This method applies compressed sensing theory framework to super-resolution reconstruction, the sparse representation of the image is solved by using a pair of coupled over-complete dictionary to actualize super-resolution reconstruction. Next, this paper uses the algorithm as a framework to analyze and study the super-resolution reconstruction algorithm in view of sparse representation. At the same time, considering the redundancy of the original method in the construction of the sample block set, it proposes a set of sample blocks. The method of similarity judgment for the sample blocks in the construction phase improves the structural richness of the sample block set. With the same number of preset samples, the number of effective samples for actual training is reduced, which greatly improves the training speed of the dictionary and improves the reconstruction accuracy of the image.

2. Single Image Super-resolution Reconstruction

2.1 Super-resolution reconstruction model based on coupled dictionary learning

The basic principle of the learning-based super-resolution algorithm is to obtain prior knowledge by learning high-resolution images and to use prior knowledge for the rebuilding process of low-resolution images. The precondition for the super-resolution algorithm based on coupled dictionary learning is that the main geometry of the same image with different resolution is approximate. Based on such preconditions, the algorithm first degrades the high-resolution graphics into corresponding low-resolution images through the degradation model. Although the high-resolution image is degenerated, original image and its corresponding degenerate image have approximately representation in the same dictionary. In other word, there is an approximate geometric structure for image pairs. Therefore, \( D_h \) (high-resolution dictionary) and \( D_l \) (low-resolution dictionary) can be obtained by learning training image pairs generated by degrading[11]. The observation image is learned on the \( D_l \), so that \( \alpha \) (the sparse coefficient) of the observation image on \( D_l \) is obtained. Finally, according to the invariance of the geometric structure of image blocks with different resolutions, a possible high-resolution image is reconstructed from \( \alpha \) and \( D_h \). The outline of image super-resolution algorithm based on coupled dictionary learning is shown in the figure below[13-14]:

From the above description and Figure 1, there are several key steps of the learning-based reconstruction method as follows:

- The extraction of training samples, including the features extraction of low-resolution image .
- Learning structure problems of dictionaries \( D_h \) and \( D_l \).
- The sparse representation problem based on over-complete dictionary[12], namely solving the problem of optimizing \( \alpha^* \) of low resolution image blocks on \( D_l \).
- Solve the problem of possible high-resolution image solutions by the sparse coefficients \( \alpha^* \) and the dictionary \( D_h \).

For the super-resolution reconstruction algorithm, the model is designed as follows:

\[
\min \| \alpha \|_0 \quad \text{s.t.} \quad \| Fx_i - FD_l \alpha_i \|_2^2 \leq \varepsilon
\]  

(1)

In the above formula, \( F \) is a feature extraction filter. The principle of selection of the feature extraction filter is to extract major features that can remarkable represent the image. Because the human eye is more sensitive to high-frequency information, it often chooses a high-pass filter as a feature extraction filter, which is beneficial to recover high-frequency information losing in the process of reconstruction. In addition, Sun[15] used a Gaussian derivative filter to extract the outline of low-resolution blocks, and Liu[16] selected first and second-order gradients in the sparse representation algorithm to extract low-resolution image features. Since the \( l_0 \) problem is a NP-hard problem, it translates into an easy-to-solve \( l_1 \) norm problem:

\[
\min \| \alpha \|_1 \quad \text{s.t.} \quad \| FD_l \alpha - Fx \|_2^2 \leq \varepsilon
\]  

(2)

The Lagrangian multiplier method gets the equivalent formula:

\[
\min \| FD_l \alpha - Fx \|_2^2 + \lambda \| \alpha \|_1
\]  

(3)
In the above formula, $\lambda$ is the balance factor, which is used to adjust the sparsity and reconstruction accuracy of the equation. In order to ensure the adjacent blocks’ compatibility, range of the overlapping area needs to be limited. The specific formula is as follows:

$$\min ||\alpha||_1 \text{ s.t. } ||D_{h} \alpha - Fx||_2^2 \leq \varepsilon_1$$

$$||P_{w} - \omega||_2^2 \leq \varepsilon_2$$  \hspace{1cm} (4)

The matrix $P$ in the formula (4) is used to extract the overlapping area of the current reconstructed image block. Formula (4) can be simplified as follows:

$$\min_{\alpha} \frac{1}{2} ||D_{h} \alpha - x||_2^2 + \lambda ||\alpha||_1$$  \hspace{1cm} (5)

In the formula (4): $D = \begin{bmatrix} F_{D_{h}} \\ \beta P_{D_{h}} \end{bmatrix}$, $y = \begin{bmatrix} Fx \\ \beta \omega \end{bmatrix}$. The parameter $\beta$ controls the balance between similarity and neighborhood compatibility between high and low sample blocks. According to formula (5), an optimized solution $\alpha^*$ is obtained, and the high-resolution image block reconstruction result is: $y = D_{h} \alpha^*$. 

---

**Figure 1.** Single image super resolution method based on coupled dictionary training.
2.2 Problems with the construction of training sample sets

For the case of constructing a universal dictionary, the sample structure will try to select sample images containing rich structures to ensure the richness of over-complete dictionary atoms. Similarly, if for a certain application, the sample can be selected when selecting apply related samples for training[17]. For face recognition, when the sample is selected, the person’s face should be selected for sample set construction. At the same time, in order to better mine the internal information of the image and reduce the computational complexity, the image should be divided into image blocks for processing and operation. The specific steps are as follows:

Assume that the input high-resolution training image set is \( \{X^I_h\}_j \). First, each image in the training set is subjected to dimension reduction processing, that is, the blur and downsampling operation: \( X^I_i = B S X^h \), where B and S respectively represent the blur and the lower Sampling operation, where noise effects are ignored. In order to ensure consistency between the image pair during dictionary training, the low-resolution images obtained by dimension reduction need to be interpolated and magnified, so that the low-resolution images maintain similar structure and size to the original images.

The interpolated low-resolution training set is \( \{X^I_i\}_j \). Next, an extraction operation of a patch is performed, including \( p_h \) (high-resolution image block) and \( p_l \) (low-resolution image block). The high resolution image block \( p^k_h \) is extracted from the position \( K \) of the high resolution image, and the size is \( \sqrt{n} \times \sqrt{n} \). To better represent the texture of the image, let \( p^k_h = p^k_h - \bar{p}^k_h \) (where \( \bar{p}^k_h \) is the mean of \( p^k_h \)), and merge \( p^k_h \) and \( p^T_h \) (where \( p^T_h \) is a transposed matrix of \( p^k_h \)) into a high resolution image block \( p^k_h = \{p^k_h, p^T_h\} \). When the \( p_l \) is extracted, the low-resolution sample is first subjected to filter processing in order to extract various features of the low-resolution image. Because first-order gradient and second-order gradient of image can effectively express the features of the image, it is often used to represent the characteristics of the image. These four filters are shown as follows:

\[
\begin{align*}
    f_1 &= [-1,0,1]; \\
    f_2 &= f_1^T; \\
    f_3 &= [1,0,-2,0,1]; \\
    f_4 &= f_3^T
\end{align*}
\]

We extract low-resolution image blocks \( \{p^k_l, p^T_l, p^{kT}_l, p^{T^T}_l\}_j \) of size \( \sqrt{n} \times \sqrt{n} \) from the same position \( K \) of the four low-resolution images after filtering. It makes \( p^k_l = \{p^k_l, p^T_l, p^{kT}_l, p^{T^T}_l\} \) and \( p^T_l = \{p^k_l, p^T_l, p^{kT}_l, p^{T^T}_l\} \), and merge \( p^T_l \) and \( p^{T^T}_l \) into a low resolution image block \( p^k_l = \{p^k_l, p^T_l\} \). Finally, the \( p^k_h \) and the \( p^k_l \) are combined as a dictionary training sample block pair set \( P = \{p^k_h, p^k_l\}_k \).

Through the above method to complete the construction of the sample block pair which is a training sample for the next dictionary construction. This method needs to record the location information in a combined calculation for separating the dictionary \( D \) into \( D_h \) and \( D_l \) in subsequent operations.

3. An improved method for constructing sample set

The construction method of the sample set is introduced in the previous section. The sample set is combined by segmenting the image pair which needs to extract its features. There is a problem with sample sets constructed using this method: Similarity between blocks is not considered. For an image, there are many similar blocks in it, which is one of the prerequisites for the image to be sparsely represented. In the construction of the sample set, it is hoped that the samples can be as rich as possible and represent different features as much as possible. The dictionary trained by such sample sets can be more abundant and can express images more precisely and concisely. It can be said that the construction of training samples directly affects the time complexity of dictionary training and the accuracy of reconstructed images.

In this paper, in order to ensure the structural heterosexuality of the sample block set, we will judge the similarity between the newly added low-resolution sample block and the existing low-resolution sample block by calculating the Euclidean distance. The sample block set constructed by this method
ensures the high efficiency of the follow-up process which including the training of dictionaries and reconstruction of image.

**Note:** This method only makes validity judgments on low-resolution image blocks, and does not judge high-resolution image blocks. However, when adding a sample block, in order to ensure the consistency of the each block pair, the same addition processing or no addition processing is performed. The following algorithm will not be described too much.

Assume that the sample block set has been constructed as $X_l$, and $X_l = \{p_l^k\}_k$, $k$ is the number of sample blocks in the constructed sample block set, $p_l^{k+1}$ is a new sample block and the judgment threshold value is $\sigma$. The pretreatment method is as follows:

1. Calculate the number $k$ of vectors and the vector size $n$ of the constructed sample block set: $[n, k] = \text{size}(X_l)$
2. Calculate the Euclidean distance of the pre-added sample block $p_l^{k+1}$ and each sample in the constructed sample block set: $\{d_j\}_{j=1,...,k} = \{\|p_l^j - p_l^{k+1}\|_2\}_{j=1,...,k}$
3. Extract the minimum distance: $d_{\text{min}} = \min_{j=1,...,k}\{d_j\}$
4. Make threshold value $\sigma = n$. (The threshold can be adjusted according to the situation. In this paper, the size of the vector is used as the threshold. It is considered that the difference between the pixels is less than or equal to 1 and it is similar between the two vectors.)
5. Decision procedure: if $d_{\text{min}} \leq \sigma$, it is considered that a sample block with a similar structure to the pre-added sample block already exists in the constructed sample block, the pre-added sample $p_l^{k+1}$ is not added.

In the process of constructing the sample set, the algorithm is used to judge every new sample block.

4. **Experiment and data analysis**

In order to verify that the improved method proposed in this paper improves dictionary training efficiency and image reconstruction accuracy, the following experiments were performed in this section.

Select 10 high-resolution images as training samples, set the number of dictionary atoms to 256, the number of iterations set by the sparse coefficient solving algorithm to be 50, and set the threshold $\sigma$ to the vector size of the low-resolution image block. The number of preset sample blocks for building a sample block set is set to 200, 400, 600, 800, and 1000, respectively. The original method and the proposed method are used to construct the sample block set, record the dictionary training time, and compare the final reconstruction effect of the image. The above methods are used for three times reconstruction of low resolution images.

The high resolution image training sample set is shown in Figure 2. The input low resolution image is shown in Figure 3.

![Figure 2](image1.jpg)  ![Figure 3](image2.jpg)
Table 1. Dictionary training time using different training sample construction methods (min)

| The number of preset sample blocks | Original method dictionary training time | This article method dictionary training time |
|-----------------------------------|------------------------------------------|---------------------------------------------|
| 200                               | 3.13562                                  | 2.7165                                      |
| 400                               | 15.2236                                  | 7.58139                                     |
| 600                               | 30.9964                                  | 12.809                                      |
| 800                               | 42.2843                                  | 27.6002                                     |
| 1000                              | 58.9998                                  | 31.9742                                     |

From the table 1, it can be seen that for the same number of pre-set sample blocks, the dictionary training time obtained by optimizing the constructed sample block set using this method is greatly reduced, which is approximately half of the original training time. Figure 4 shows the reconstructed image and detail comparison diagrams obtained using the different sample block set construction methods. The number of preset sample blocks used in a)-d) is 1000. The number of preset sample blocks used in e)-h) is 400.

Figure 4. Reconstructing images using different sample block set construction methods
From the reconstruction effect graph of Figure 4, it can be seen that the visual effect of the reconstructed image using this method is better than that of the original method, and the false reconstruction effect at the reconstruction image edge is reduced.

5. Conclusion
This paper first gives a brief introduction to super-resolution algorithm based on over-complete dictionary learning, and focuses on the image reconstruction algorithm proposed by Yang based on coupled dictionary learning. By analyzing the redundancy of the original method in the method of sample set construction, this paper proposes an improved method of building block sets, which improves the effectiveness of sample block sets by judging the similarity between sample blocks. Experiments show that for the same number of preset blocks, the improved method proposed in this paper greatly improves the training speed of the dictionary and also improves the quality of the reconstructed image.

References
[1] Yegani.F, Nazzal.M. and Ozkaramanli.H. 2015 Image super-resolution via sparse representation over multiple learned dictionaries based on edge sharpness and gradient phase angle. Signal, Image and Video Processing. 9 pp:285-293.
[2] Yang.J.C, Wright.J, Huang.T.S, et al. 2010 Image super-resolution via sparse representation. IEEE Transactions on Image Processing. 19 (11):2861-2873.
[3] Yang.J.C, Wang.Z.W and Lin.Z. 2012 Coupled Dictionary Training for Image Super Resolution, IEEE Transaction on Image Processing. 21 pp: 3467-3478.
[4] Dong.C, Loy.C.C, He.K.M, et al. 2014 Learning a deep convolutional network for image super-resolution. Proceedings of European Conference on Computer Vision. pp:184-199.
[5] Chang.H, Yeung.D.Y and Xiong.Y. 2013 Super-resolution through neighbor embedding. Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. pp: 275-282.
[6] Timofte R, De Smet V and Van Gool L. 2013 Anchored neighborhood regression for fast example-based super-resolution. Proceedings of 2013 IEEE International Conference on Computer Vision. pp:1920-1927.
[7] Radford A, Metz L and Chintala S. 2015 Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. Computer Science.
[8] Justin J, Alexandre A and Li F F. 2016 Perceptual Losses for Real-Time Style Transfer and Super-Resolution. European Conference on Computer Vision. pp:694-711.
[9] Ledig C, Theis L, Huszar F, et al. 2016 Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. arXiv preprint arXiv:1609.04802.
[10] Nayak Rajashree and Patra Dipti 2018 New single-image super-resolution reconstruction using MRF model. Neurocomputing. 293:108-129.
[11] Liling Zhao, Quansen Sun and Zelin Zhang. 2017 Single Image Super-Resolution Based on Deep Learning Features and Dictionary Model, Recent Advantages of Computer Vision based on Chinese Conference on Computer Vision (CCCV). 5: 17126 - 17135
[12] Yaolan Zhang and Yijun Liu. 2017 Single image super-resolution reconstruction method based on LC-KSVD algorithm, AIP Conference Proceedings. Volume 1839, Issue 1, id.020095
[13] Leslie N.Smith and Michael Elad 2013 Improving Dictionary Learning: Multiple Dictionary Updates and Coefficient Reuse. IEEE Signal Processing Letters. Vol.20, No.1.
[14] Chang Liu and Yi Jun Liu. 2017 Adaptive super resolution algorithm based on RBM dictionary learning, Wireless Communications, Signal Processing and Networking. pp:2699-2703.
[15] Sun.J Zheng.N.N and Tao.H 2003 Image hallucination with primal sketch priors. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 2:729-736.
[16] Weirong Liu 2013 Multi-morphology image super-resolution via sparse representation. Neurocomputing. 120 pp: 645-654.
[17] Alvarez.R Valentin.P.V Shkvarko.Y and Reyes.R 2017 Image Super-Resolution via Block Extraction and Sparse Representation. IEEE Latin America Transactions. 15(10), 1977-1982.