Dictionary Learning Based on Structural Self-similarity and Convolution Neural Network

Ling Zhang, Wenchao Jiang[✉], and Wen Xiang

Faculty of Computer, Guangdong University of Technology, Guangzhou, China
85711585@qq.com

Abstract. Aiming at insufficient detailed description problem caused by the loss of edges during a single low-resolution (LR) image's reconstruction process, a novel algorithm for super resolution image reconstruction is proposed in this paper, which is based on fusion of internal structural self-similarity dictionary and external convolution neural network parameters learning model. Firstly, for solving training samples too scattered problem, besides external database, an internal database is constructed to learn a dictionary of the single image's structural self-similarity by multi-scale decomposition approach. Secondly, nonlocal regularization constraint is calculated on the priori knowledge, which is obtained from the internal database of the single LR image. Thirdly, similar block pairs of high and low-resolution samples in the external database are input into a convolution neural network for learning the parameters of reconstructing model. After all, combined parameters learned and the internal dictionary, the single LR image is reconstructed, and by iterative back-projection algorithm its result is improved. Experimental results show that, compared with state-of-the-art algorithms, such as Bicubic, K-SVD algorithm and SRCNN algorithm, our method is more effective and efficient.

Keywords: Super resolution · Structural self-similarity · Convolution natural network · Nonlocal regularization · Block matching

1 Introduction

Super-resolution image reconstruction refers to the technology of improving low-quality and low-resolution images to recover high-resolution images, and has important applications in military, medical, public safety and computer vision. The general way of super-resolution image reconstruction is to learn from a large number of high-resolution images to reconstruct high-frequency details of the low-resolution image [1–4]. Performances of these algorithms are not so satisfied, for they are likely to be affected by training data that are too scattering to effectively represent a given image. On the other hand, structural information of this given image is valuable for itself reconstruction but little attention has been paid so far [5–7]. Freedman et al. [8] pointed out that there were many structural self-similar blocks distributed within one single image region, so several related studies on local structural self-similarity extraction are also reported [9, 10].

However, they cannot effectively deal with irregular texture blocks, which are sparsely or infrequently appeared in a single image. Error matching between image blocks will bring a lot of fake textures and make it difficult to guarantee the good effect of reconstruction. In order to solve the problems, this paper proposes a super-resolution image reconstruction method based on the structural self-similarity of the single image. Similar structural blocks of the same scale and different scales are extracted from the single image and used to set up an internal dictionary model. Then the weights of the dictionary are learned by external sample images trained with a convolution neural network. With these data, a reconstructive model adaptive to the given single image is obtained and information of the single image is made best use of. Experimental results verify the effectiveness of our proposed algorithm when compared with other state-of-the-art approaches.

2   Parameters Learning Model Based on Convolution Neural Network

In this paper, reconstructing parameters of the single LR image are learned by a Super-Resolution Convolution Neural Network (SRCNN) [11] framework, which has been proved to have a great capability in extracting the essential features of data sets. Firstly, block pairs of Low-Resolution (LR) and High-Resolution (HR) images from the external database are matched to each other to achieve matching pairs of image blocks. Then these blocks are regarded as samples and input into SRCNN, which consists of three layers of convolution layers, including feature extracting, non-linear mapping and high-resolution image reconstructing parameters achieving, respectively. The framework of our SRCNN is shown in Fig. 1, and three convolution layers of SRCNN deep learning algorithms are expressed as the following equations:

$$Y_1 = \max\{0, W_1 \cdot X + B_1\} \tag{1}$$

$$Y_2 = \max\{0, W_2 \cdot X + B_2\} \tag{2}$$
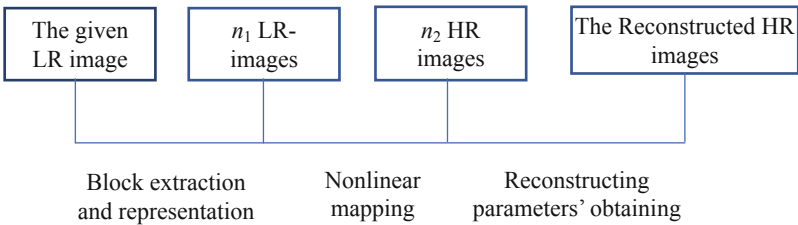
$$Y_3 = \max\{0, W_3 \cdot X + B_3\} \tag{3}$$

Fig. 1. Structure of SRCNN framework

In Eq. (1)–(3), matrix X represents the original single LR image, Y_i ($i = 1, 2, 3$) represents output of each convolution layer, W_i ($i = 1, 2, 3$) and B_i ($i = 1, 2, 3$) represent the neuron convolution kernel and neuron bias vector, respectively. Symbol '·' represents a convolution operation, whose result is then processed by the ReLu activation function max $\{0, x\}$. With matching pairs of image blocks, this neural network frame needs to learn parameters set $\Phi = \{W_1, W_2, W_3, B_1, B_2, B_3\}$, which are estimated by minimizing the error loss between the last output of neural network and HR image. Given a HR image Y and its corresponding LR image X, its loss function could be described by using its mean square error $L(\Phi)$, as shown in Eq. (4).

$$L(\Phi) = \frac{1}{n} \sum_{i=1}^{n} \|F(X_i, \Phi)\|^2 \tag{4}$$

Equation (4) can be solved by stochastic gradient descent and back-propagation algorithm together.

3 Extraction of Self-similarity Feature

3.1 Self-similarity on the Same and Multi-scale Images

Image self-similarity refers to as similar features available among the various regions of the entire image. Researches [12] had shown that, for a 5×5 image block in a natural image, there were a large number of image blocks in the same scale and different scales can be found in the image. A statistic shows that more than 90% of image blocks can find at least 9 similar image blocks in the same scale of itself; more than 80% of image blocks can find at least 9 similar image blocks of different scales. Based on this image similarity mechanism, we can extract a lot of redundant information of the image itself on the same scale and different scales. Frank et al. [13] pointed out that there were two characteristics in general images: one is that a large number of similar structural regions appear in the whole image; and the second is that these structural similarities can keep consistent on multiple scales of the image, as shown in Fig. 2a).
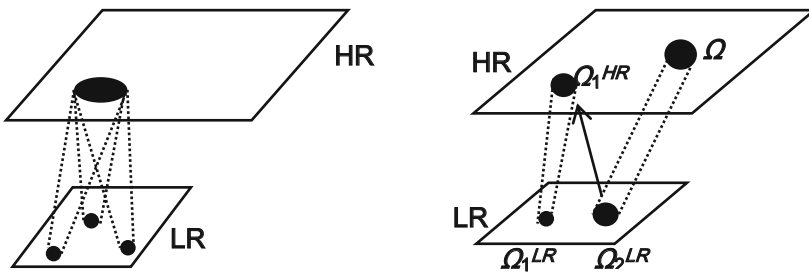
a) Similar Blocks with same scale image

b) Consistency in Different Scales HR

Fig. 2. Similar image blocks with same scale and different scales in a single image

Since there are so many structure-similar images blocks in the same scale and different scales of a single LR image, we will benefit if we make the best use of these structural similarities in its reconstruction. The basic scheme of our algorithm is shown in Fig. 2b), where HR represents a high-resolution image, and LR represents a corresponding low-resolution image thereto. Size of HR image is s times that of LR image. Suppose Ω_1^{HR} and Ω_2^{HR} represent two similar blocks with different scales in HR image, and size of Ω_2^{HR} is s times that of Ω_1^{HR}. The corresponding image blocks of Ω_1^{HR} and Ω_2^{HR} in LR image are Ω_1^{LR} and Ω_2^{LR}. In this case, Ω_1^{LR} and Ω_2^{LR} in the LR image form a pair of similar image blocks with different scales. Suppose scaling factor between the HR and LR images is the same with that between Ω_1^{HR} and Ω_2^{HR}, and then size of Ω_1^{HR} in HR image is exactly the same as Ω_2^{LR} in LR image. Accordingly, in recovering block Ω_1^{LR} to form block Ω_1^{HR} in HR image, Ω_2^{LR} could provide helpful additional information for it.

In this paper, LR and HR similarity block pairs in same scale and different scales are derived from images of the external and internal database by using a non-local block matching method. Then, these block pairs are treated as training samples for dictionary learning to reconstruct the image to be restored.

3.2 Non-local Self-similar Block Matching

Researchers have found that [15, 16], natural images have abundant similarities in regions of texture, edge and so on. A low resolution (LR) image can restore its missing details based on this structural high-frequent similarity. It seems that exploiting the similarities between nonlocal patches distributed in different regions of the image can achieve higher image reconstruction resolution [14].

This paper presents a regularization constraint item based on non-local block similarity. Suppose X_i represents the ith block of a single LR image X, its similar blocks X_i^l ($l = 1, 2, …, L$) are firstly searched within X itself, which are then used to estimate X_i by their linear combination. The main idea of non-local constraint is that central point P_i of block X_i can be represented by the weighted average of central point P_i^l of block X_i^l ($l = 1, 2, …, L$), which could be described in Eq. (5).

$$P_i = \frac{1}{L} \sum_{l=1}^{L} w_i^l P_i^l \tag{5}$$

Suppose that each weight vector w_i is a matrix of vectors w_i^l, which consists of weight matrix B. Each P_i is made up of by P_i^l, which consists of dictionary matrix Ψ. The nonlocal regularization constraint can be expressed as Eq. (6).

$$\alpha = \|(I - B)\psi\alpha\|_2^2 \tag{6}$$

In Eq. (6), I is an identity matrix and α is the 2-norm constraint parameter of nonlocal regularization.

4 Structure Self-similarity Extraction

4.1 Low Resolution Degraded Model

LR images are caused by blurring, down-sampling or noise pollution of HR images [17]. The whole degraded process could be approximated as a linear one, as shown in Eq. (7).

$$X = HSY + n \tag{7}$$

Where Y and X are reconstructed HR image and the original LR image, respectively. H represents the down-sampling operation, S is the fuzzy operator, and n is the noise pollution matrix. In order to accurately estimate the HR image matrix Y, some priori knowledge or regular constraint items of the image need to be introduced, as shown in Eq. (8).

$$\hat{Y} = arg \min_{Y} \|X - HSY\|_F^2 + \lambda\alpha \tag{8}$$

In Eq. (8), Y is the reconstructed HR image, and $||X - HSY||_F^2$ represents error term in observation, α is the regular constraint item in Eq. (6), λ is the weighted balance parameter of regular item.

We deformalize the degraded model constrained by the Eq. (6), and on behalf of the formula (8), the final algorithm is got in Eq. (9).

$$\hat{Y} = arg \underbrace{\min}_{y} X - HSY_F^2 + \lambda\alpha_1 + \mu(I - B)\psi\alpha_2^2 \tag{9}$$

Parameter μ represents the regularization parameters. After all, we use the iterative back-projection algorithm to further enhance the image reconstruction performance.

4.2 Algorithm in Detail

Firstly, with prior knowledge of the non-local self-similarity in the original LR image, search the best match blocks of the initial super-resolution image with multi-scale method, and then take them as an internal dictionary to learn non-local regularization constraints. Depth learning is generally trained with a large amount of data, but in this case, a relatively small training set consisting of 91 images [3] is used for training. Best match blocks in LR and HR images of these training samples are found out and made up to be a lot of pairs, which would be used to compose of an external dictionary. Both these two kinds of samples are input to convolution neural network for modeling self-structure similarity of LR image. After all, with the non-local regularization constraints learned from internal dictionary, the original LR image is reconstructed. There are four steps: initial interpolation, non-local blocks matching, neural network model learning and non-local regularization constraints.

- Initial Interpolation. In this paper, cubic bilinear interpolation algorithm, the commonly used algorithm for LR image reconstruction, is selected to build its original HR image, which is later used for LR image's self-similarity extraction.

- Non-local blocks matching. The original HR image is partitioned point-by-point into blocks, which are then matched with each other to obtain structure similar blocks. There are two categories of similar image blocks, which are the ones with the same scale and the other ones with different scales. In the case of the same scale blocks set, suppose represents the ith one, we match it across the whole blocks set with the same size to search its closest similar couple block. The difference between the searched block and the current block $\hat{x}_i$ is calculated as Eq. (10).

$$e_i^l = \left\| \left(\hat{x}_i^l - \hat{x}_i \right) \right\|_2^2 \tag{10}$$

- Neural network model training. Convolution neural network model has a strong feature learning ability; therefore, we use SRCNN algorithm with a three-layer structure for dictionary training. Finally, corresponding network model parameters Φ are obtained.
- Non-local Regularization Constraints. Based on the obtained parameters of convolution neural network model, combined with nonlocal regularization and dictionary data, this section builds a reconstructed image according to Eq. (9).

4.3   Algorithm Enhancement

We use the iterative back-projection algorithm to enhance our reconstructed image, which is based on a down-sampling image degradation model with sub-pixel displacement [18] Firstly, multi-frames of LR image are sampled in sequence and registered, and then errors between the LR image simulation and its observation results are iteratively back-projected to the HR image. Suppose that there are K sequential observation LR images, described as $f_k (m_1, m_2)$ with resolution $M_1 \times M_2$. Size of the estimated HR image $f (n_1, n_2)$ is enlarged by s times, which means resolution of the estimated HR image $N_1 \times N_2 = (sM_1) \times sM_2$). Using the Iterative Back-Project (IBP) method to estimate the HR image can be described as Eq. (11).

$$\hat{f}^{n+1}(n_1, n_2) = \hat{f}^n(n_1, n_2) + \sum_{m_1, m_2} \left(g_k(m_1, m_2) - \hat{g}_k^n((m_1, m_2)) \right) \times h^{BP}(m_1, m_2; n_1, n_2)$$

$$\tag{11}$$

In Eq. (11), $\hat{g}_k^n$ represents the kth simulation result of LR image in the nth iteration, generated by the actual displacement information of LR images. $h^{BP}(m_1, m_2; n_1, n_2)$ is the back-projection kernel, which determines how error affects the HR image construction during each iteration. We use a down-sampling rate of $s = 3$, and the displacement of sub-pixel (x, y) are $(0, 0)$, …, $(1, 3)$, respectively. We get 8 LR observation images and corresponding simulation images, and calculate errors between these two kinds of images. At last, we obtain the HR image according to the Eq. (11).

4.4   Algorithm Implementation

The proposed self-similar similarity convolution neural network algorithm is divided into two processes, training and reconstruction, respectively.

In order to clarify the algorithm in this paper more clearly, the algorithm flow chart is shown in Fig. 3.
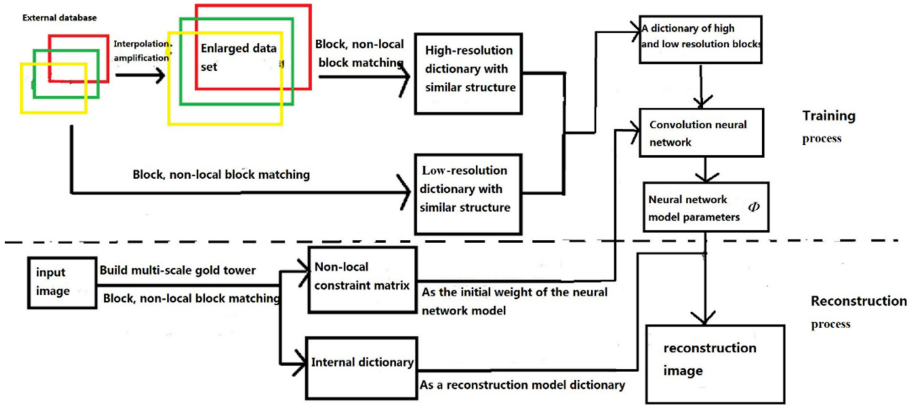
Fig. 3. Algorithm flow chart

4.5 Non-local Regularization Constraints Example

In this process, non-local regularization constraints of the single LR image are obtained by its structure self-similarity blocks' representation. According to Eq. (6), non-constraints are iteratively calculated by weighted average matrix made up by each block's similar representation, and a dictionary, which is consisted of by similar blocks' central points. A simple example of this process is shown in Fig. 4*a*). Four clustering results of structure blocks are also shown in Fig. 4*b*), which represent non-local regularization constraints of the example image.

a) Original LR image   *b*) Four structure blocks clustering results
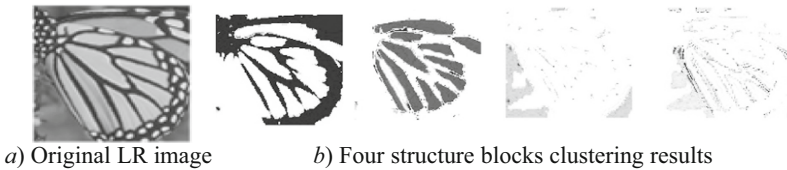
Fig. 4. An example of non-local regularization constraints

5   Experimental Results and Analysis

5.1   Experimental Setup

In order to verify the validity of our proposed method, three international public SR databases are used, which are Set5, Set14 and Urban100, and three-layer convolution neural network is used to for model learning. The first layer has 9×9 size and 64 convolution kernels and neurons, the second layer has 1×1 and 32, and the third layer has 5×5 and 1. In the experiment, Bicubic interpolation, K-SVD and convolution neural networks are selected as contrast analysis approaches to compare with the performance of our proposed method, based on indicator of Peak Signal to Noise Ratio (PSNR).

5.2   Experimental Setup

In order to evaluate the quality of image reconstruction, we compare the performance of these methods on PSNR. Taking three images in database Set14 as an example, the reconstructed results with four approaches are shown in Fig. 5 under the condition of magnification of 3 times. From them, local information restored by ours is clearer and more delicate, and global reconstructed images are more approaching to the original images.

a) Original image *b*) Bi-interpolation   *c*) K-SVD   *d*) SRCNN   *e*) ours
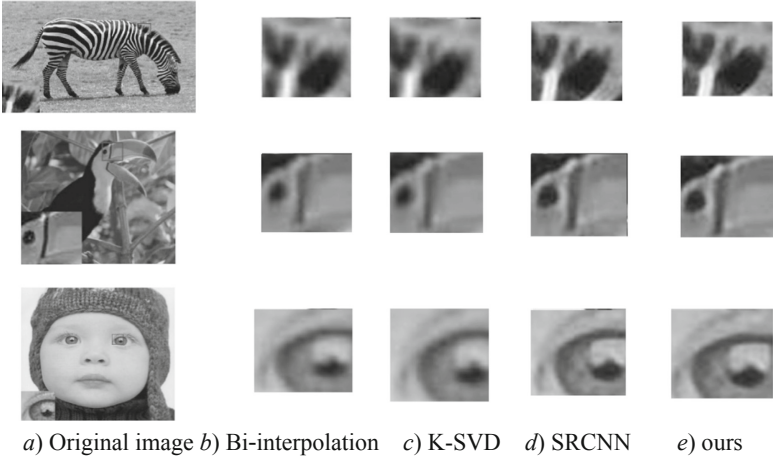
Fig. 5. Three reconstructed results from Set14 with up scaling factor 3

PSNR of each approach is shown in Table 1, 2 and 3. Bicubic interpolation method takes the lowest place, only 22.101 db and the best contrast algorithm can reach 40.642 db, while our proposed method can reach the highest 42.204 db.

Table 1. PSNR comparison of four sr methods with up-scaling factor as 2

Image name	Four SR methods			
	Bicubic	*K-SVD*[19]	*SRCNN* [11]	*Our method*
Zebra	30.63	31.73	33.29	34.94
Baby	37.07	38.25	38.30	39.42
Butterfly	27.43	30.65	32.20	33.33
Woman	32.14	34.49	34.94	36.05
Head	34.86	35.59	35.64	36.86
Bird	36.81	39.93	40.64	42.20

[] is the method's reference Number

Table 2. PSNR comparison of four SR methods with up-scaling factor as 3

Image name	Four SR methods			
	Bicubic	*K-SVD* [19]	*SRCNN* [11]	*Our method*
Zebra	26.63	26.92	28.87	29.94
Baby	33.91	35.08	35.01	36.22
Butterfly	24.04	25.94	27.58	28.84
Woman	28.56	30.37	30.92	32.55
Head	32.88	33.56	33.55	35.86
Bird	32.58	34.57	34.91	36.20

[] is the method's reference Number

Table 3. PSNR comparison of four SR methods with up-scaling factor as 4

Image name	Four SR methods			
	Bicubic	*K-SVD* [19]*	*SRCNN* [11]*	*Our method*
Zebra	24.08	24.92	25.97	27.03
Baby	31.78	33.06	32.98	34.21
Butterfly	22.10	23.57	25.07	26.24
Woman	26.46	27.89	28.21	30.42
Head	31.59	32.21	32.19	34.06
bird	30.18	31.71	31.98	33.27

[] is the method's reference Number

Our method has an average declining rate 9.17% in PSNR, higher than other methods.

6   Conclusions

The proposed algorithm considers the reconstruction of a single super-resolution of image based on self-structure similarity within the image. The algorithm derives self-similarity of the training samples through the scale decomposition of the image, and makes full use of the structural self-similarity of the input image to solve the problem that training samples are too scattered for representing the LR image. The intrinsic structure self-similarity of the image is obtained through the nonlocal regularization constraint. Finally, the iterative back-projection algorithm is used to further optimize the reconstructive effect. Compared with state-of-the-art algorithms such as Bicubic, KSVD and SRCNN, the proposed algorithm can achieve better reconstructive performance.

References

1. Li, Z.H.X., He, H., Wang, R., et al.: Single image super-resolution bidirectional group sparsity and directional features. Image Process. **9**(24), 2874–2888 (2015)
2. Dong, C., Loy, C.C., He, K., et al.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **2**(38), 295–307 (2016)
3. Yang, J., Wright, J., Huang, T., et al.: Image super-resolution via spare representation. IEEE Trans. Image Process. **19**(11), 2861–2873 (2010)
4. Timofte, R., Smet, V., Gool, L.: Anchored neighborhood regression for fast example-based super-resolution. In: IEEE International Conference on Computer Vision, Sydney, pp. 1920–1927. IEEE (2013)
5. Timofte, R., De Smet, V., Van Gool, L.: A + : adjusted anchored neighborhood regression for fast super-resolution. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9006, pp. 111–126. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16817-3_8
6. Yang, C.Y., Yang, M.H.: Fast direct super-resolution by simple functions. In: IEEE Interaction Conference on Computer Vision, Sydney, pp. 561–568. IEEE (2013)
7. Dai, D., Timoft, R., Vangool, L.: Jointly optimized regressors for image super-resolution. Comput. Graph. Forum. **34**(2), 95–104 (2015)
8. Freedman, G., Fattal, R.: Image and video upscaling from local self-examples. ACM Trans. Graph. **30**(2), 12 (2011)
9. Protter, M., Elad, M., Takeda, H., Milanfar, P.: Generalizing the nonlocal-means to super-resolution reconstruction. IEEE Trans. Image Process. **18**(1), 36–51 (2009)
10. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Non-local sparse models for image restoration. In: IEEE International Conference on Computer Vision. Kyoto, Japan, pp. 2272–2279. IEEE (2009)
11. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 184–199. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10593-2_13
12. Gkasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: International Conference on Computer Vision, Kyoto, pp. 349–356. IEEE (2009)
13. Candocia, F.M., Principe, J.C.: Super-resolution of images based on local correlations. IEEE Interact. Neural Netw. **2**(10), 372–380 (1999)
14. Dong, W.S., Zhang, L., Shi, G.M., et al.: Nonlocally centralized sparse representation for image restoration. IEEE Trans. Image Process. **4**(22), 1620–1630 (2013)
15. You, X., Xue, W., et al.: Single image super-resolution with non-local balanced low-rank matrix restoration. In: International Conference on Pattern Recognition, Cancun, vol. 10, no. 23, pp. 1255–1260. IEEE (2016)
16. Xu, J., Zhang, L., Zuo, W., et al.: Patch group based nonlocal self-similarity prior learning for image denoising. In: Proceedings of IEEE Conference on Computer Vision, Santiago, pp. 244–252. IEEE (2015)

17. Tekalp, A.M.K., Sezan, M.I.: Hight-resolution image reconstruction from lower-resolution image sequences and space varying image restoration. In: Proceedings of the IEEE International Conference on Acoustics. Speech and Signal Processing, San Francisco, pp. 169–172 (1992)
18. Lu, Y., Imanura, M.: Pyramid-based super-resolution of the under sampled and subpixel shifted image sequence. Int. J. Syst. Technol. **12**, 254–263 (2002)