TRAINING-BASED SUPER-RESOLUTION ALGORITHM USING K-MEANS CLUSTERING AND DETAIL ENHANCEMENT

*Shin-Cheol Jeong and Byung Cheol Song**

The School of Electronic Engineering, Inha University
253 Yonghyun-dong, Nam-gu, Incheon, 402-751, Republic of Korea
phone: + (82) 32-860-7413, fax: + (82) 32-868-3654, email: bcsong@inha.ac.kr

ABSTRACT

This paper presents a computationally efficient learning-based super-resolution algorithm using k-means clustering and detail enhancement. Conventional learning-based super-resolution requires a huge size of dictionary for reliable performance, which brings about a tremendous memory cost as well as a burdensome matching computation. In order to overcome this problem, the proposed algorithm significantly reduces the size of the trained dictionary by properly clustering similar patches at the learning phase. Simulation results show that the proposed algorithm provides superior visual quality to the conventional algorithms, while needing much less computational complexity.

1. INTRODUCTION

Image interpolation is a key technology to display high quality up-scaled images on cutting-edge digital consumer applications such as high-definition television (HDTV), digital still camera (DSC), and digital camcorder. For several decades, a lot of single image interpolation algorithms have been developed in the literature. They can be classified into three categories: interpolation-based, reconstruction-based and super-resolution approaches. Firstly, the interpolation-based methods [1, 2] are computationally light and have simple structure in comparison with the others. However, they suffer from blurring and jagging artifacts in diagonal edges. Even the edge preserving interpolation methods as in [3-6] have a difficulty in synthesizing fine details. Secondly, reconstruction-based methods [7, 8] produce high resolution image under the constraint that the smoothed and down-sampled version of the reconstructed high resolution image is close to the input low resolution image. For example, back-projection algorithm iteratively minimizes the reconstruction error. But, those algorithms rarely avoid jagging and ringing artifacts along the strong edges. Finally, so-called super-resolution (SR) algorithms [9-13] have been developed as the most promising approach.

A typical SR image reconstruction makes use of signal processing techniques to obtain a high resolution (HR) image from multiple low-resolution (LR) images [9]. In general, success of such SR schemes depends on existence of sub-pixel motion between adjacent LR images and accurate sub-pixel estimation. However, sub-pixel motion estimation among neighbor LR images requires not only huge computational cost, but also its accuracy is not guaranteed in certain environments. In order to solve the above-mentioned problem, a lot of single image-based SR methods have been devised, e.g., example-based or learning-based SR algorithms [10-13]. They exploit the prior knowledge between the HR and its corresponding LR examples through so called learning process. Most example-based SR algorithms usually employ a dictionary composed of a large number of HR patches and their corresponding LR patches. The input LR image is split into either overlapping or non-overlapping patches. Then, for each input LR patch, either one best-matched patch or a set of the best-matched LR patches are selected from the dictionary. The corresponding HR patches are used to reconstruct the output HR image. However, most of the existing algorithms are so-called 'searching and pasting' approaches, and are therefore computationally intensive in finding the best match of LR–HR patch from a huge dictionary. Furthermore, best-matched but incorrect patches will seriously degrade the reconstruction results.

This paper achieves fast image super-resolution by reducing the size of trained dictionary. At the learning phase, the number of LR-HR patch pairs in dictionary is noticeably reduced by grouping similar LR patches using *K*-means clustering. At the synthesis phase, each input LR patch is compared with the candidate LR patches in the dictionary one-by-one. So, one best-matched patch is selected from the dictionary. Finally, the corresponding HR patch is used to reconstruct the output HR patch along with additional post-processing for detail enhancement. Thus, the reduced dictionary size makes it possible to significantly speed up SR processing and save the memory cost, while providing reasonable visual quality.

2. LEARNING-BASED SUPER-RESOLUTION

Figure 1 describes the basic concept of learning-based super-resolution that is generally composed of two phases: Off-line learning phase and on-line synthesis phase. At the learning phase, the training data, i.e., dictionary consisting of LR and HR patches is constructed. The LR and HR patch pairs are obtained from various training images. During the synthesis phase, the input LR image is super-resolved by using the dictionary. For each LR patch in the input image, its nearest neighbor LR patches are explored from the dictionary. The high frequency components of the input LR patch are synthesized using the best matched LR patches.

 ISSN 2076-1465

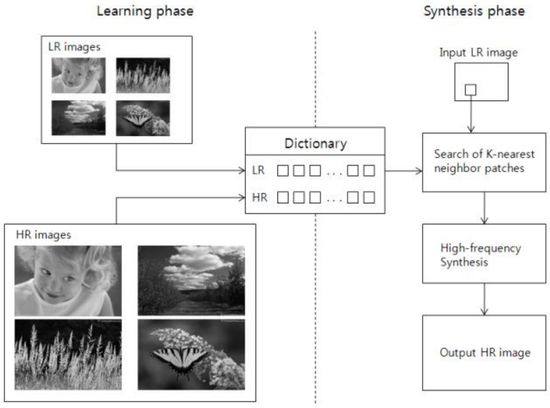


Fig. 1. Conventional learning-based super-resolution.

Freeman *et al*. [10] embedded two matching criteria into a Markov network. One is that the LR patch from the dictionary should be similar to the input observed patch, while the other criterion is that the contents of the corresponding HR patch should be consistent with its neighbors. Chang *et al*. [12] presented a neighbor embedding-based SR algorithm which assumes that generation of a high-resolution image patch depends on multiple nearest neighbors in the dictionary. The algorithm finds the optimal reconstruction weights of the nearest neighbor patches and then estimates a proper HR patch by applying the weight to the corresponding HR patches.

The performance of those learning-based SR algorithms highly rely on matching accuracy of an input LR patch with candidate LR patches in the dictionary. In order to improve the accuracy of matching, a sufficient number of LR-HR patch pairs must be included in the dictionary. Usually, existing learning-based SR methods require hundreds of thousands of training examples for reliable performance. However, such a dictionary size causes tremendous memory cost for storing the training samples as well as awfully large computational complexity in matching process. Therefore, it makes conventional learning-based SR impractical in implementation and restrictive in applications. In order to overcome this problem, we propose a fast learning-based SR algorithm with reduced dictionary based on *K*-means clustering.

3. THE PROPOSED ALGORITHM

The proposed algorithm accomplishes fast HR image reconstruction without degradation by reducing dictionary size at the learning phase. Figure 2 describes the overview of the proposed algorithm. The learning phase of the proposed algorithm includes pre-processing, patch extraction, dictionary size reduction, and ordinary dictionary generation steps. In addition, the residue dictionary is designed to compensate for some high frequency (HF) components lost during dictionary size reduction. The synthesis phase is composed of pre-processing, patch extraction, HF synthesis using ordinary dictionary, and residual HF synthesis using residue dictionary.

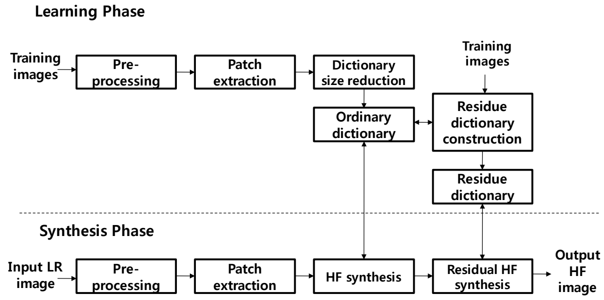


Fig. 2. The proposed algorithm.

3.1 Learning Phase 1: Patch Extraction

Prior to learning process, training images should be appropriately pre-processed to achieve effective dictionary construction (see Figure 3). Firstly, each HR image I_H is blurred and sub-sampled to generate a LR image I_L. And then, I_L is again up-scaled using simple linear interpolation such as bi-linear interpolation or cubic convolution to produce an image I_{UP} having the same resolution as I_H.

The dictionary should possess various HF details lost by image degradation process and specific features to index them. The HF image I_{HF} is obtained by subtracting I_{UP} from I_H, and mid frequency (MF) image I_{MF} stands for high pass filtered version of I_{UP}. I_{MF} is employed as the features for indexing. Note that the HF layer I_{HF} is the target information to be recovered by the proposed algorithm. They indicate lost HF and MF layers for predicting them, respectively. As a result, we extract and store HR and LR patches from I_{HF} and I_{MF}, respectively. Those patches are properly overlapped with neighbour patches for local smoothness. Without loss of generality, we assume that the relationship between I_{HF} and I_{MF} is independent of the local image contrast. So, we normalize the contrasts of LR and HR patches by dividing them by the energy of the LR patch. Here, the energy stands for the L_1-norm.

Finally, so-called primitive patches including edges or textures are chosen and they only belong to the dictionary. In other words, the proposed synthesis may be applied only for the primitive regions. Maximum response filter [15] is used to extract primitives.

3.2 Learning Phase 2: Dictionary Size Reduction

Now, we need to effectively reduce the number of LR-HR patch pairs in the dictionary so as to mitigate memory cost and computational burden in synthesis. This process is very significant in that the number of training examples in the dictionary generally dominates the performance of learning-based SR. Most of all, the small number of the samples can improve practicality of the proposed SR algorithm.

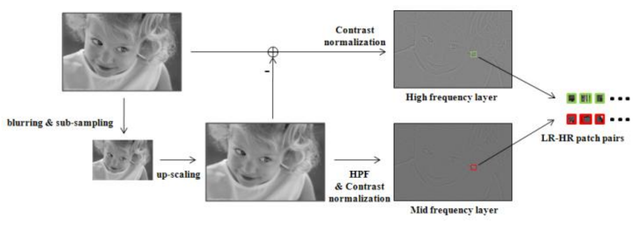


Fig. 3. Pre-processing for dictionary construction.

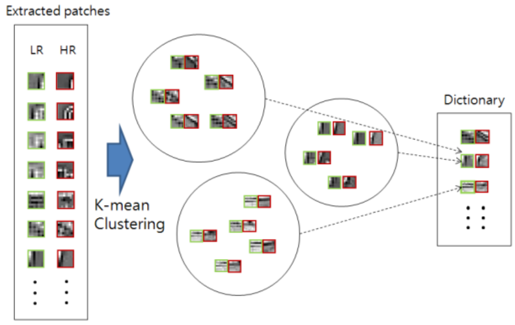


Fig. 4. Dictionary size reduction.

So, we group adjacent LR-HR patch pairs into a single patch pair. We adopt *K*-means clustering to gather similar patches. Figure 4 illustrates this clustering process. LR patches which are close each other in terms of L_2-norm distances are clustered into a single group and similarly the corresponding HR patches are clustered. Finally, the center points of each cluster become new LR and HR patches belonging to the ordinary dictionary. In practice, we can determine *K* considering memory cost and computational complexity.

3.3 Learning Phase 3: Residue Dictionary Design

The above-mentioned dictionary size reduction sometimes causes blurring artifact because HF components can be weakened during averaging HR patches. Kim *et al.* showed that if the residue patch, i.e., the difference between the actual HR and estimated HR patches is well-learned, the learned dictionary of residues may provide improved visual quality. So, in order to compensate for lost HF components, we employ additional post-processing where a proper residue patch for each LR patch is explored from the learned residue dictionary and it is added to the HF patch synthesized using the ordinary dictionary (see Fig. 2). Figure 5 describes the training procedure to construct so-called residue dictionary. Note that the training images for residue dictionary are different from those used for ordinary dictionary. Firstly, the HF and MF images are generated from pre-processing. Next, the best matched MF patch to each training MF patch is found with its corresponding HF patch. Then, the HF residue patches between the original HF patches and the estimated HF patches are produced.

Similarly, the MF residue patches between the input MF patches and the matched MF patches in the ordinary dictionary are produced. Finally, the MF and HF residue patches are clustered at the same fashion as subsection 3.2, and the residue dictionary is finally obtained.

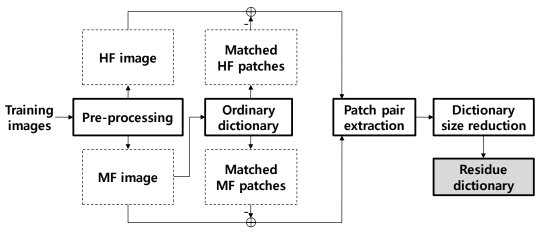


Fig. 5. Construction of residue dictionary.

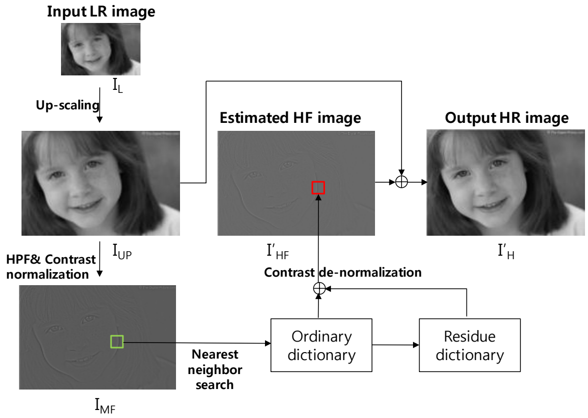


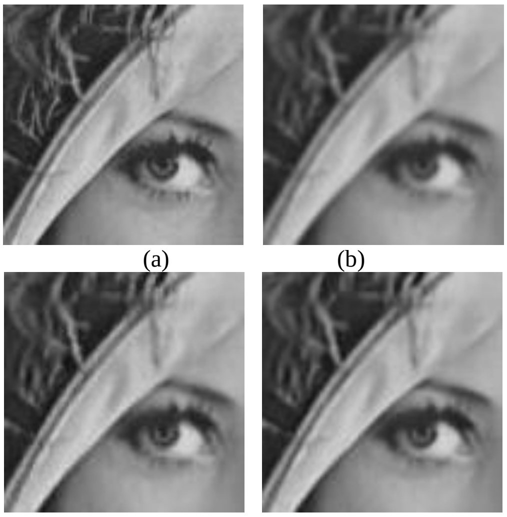
Fig. 6. Synthesis phase.

3.4 Synthesis Phase

The input LR image is first up-scaled using bi-linear interpolation, and then LR patches are extracted from MF layer of the input image as in the learning phase. Each input LR patch is compared with the candidate LR patches in the ordinary dictionary to find the best match in L_2-norm distance. Note that the proposed algorithm selects the first nearest neighbor patch only unlike the conventional learning-based SR algorithms using multiple nearest patches [13]. In our algorithm, the nearest neighbor patch may correspond to the average of multiple adjacent patches because similar patches are already clustered in the learning phase. Therefore, even though a single best-matched patch is used for HF synthesis, we can obtain a similar effect to synthesis using multiple nearest patches as in [13].

Next, the HR patch corresponding to the best matched LR patch is de-normalized by multiplying with the energy of the input LR patch. This process is applied to all the input patches. Only for pixels in overlapped regions, averaging is performed. Finally, we obtain a synthesized HR image by adding the high frequency image I'_{HF} to the initially up-scaled image I_{UP} (see Figure 6). Note that the input to the residue dictionary is the difference between the input LR patch and the best-matched LR patch chosen from the ordinary dictionary.

Fig. 7. 16 training images (upper 16) and two test images (lower).



(a) (b)

(c) (d)

Fig. 8. A part of up-scaled images of *lena* using 10,000 training examples (a) original, (b) bi-cubic (c) Fan's (d) proposed.

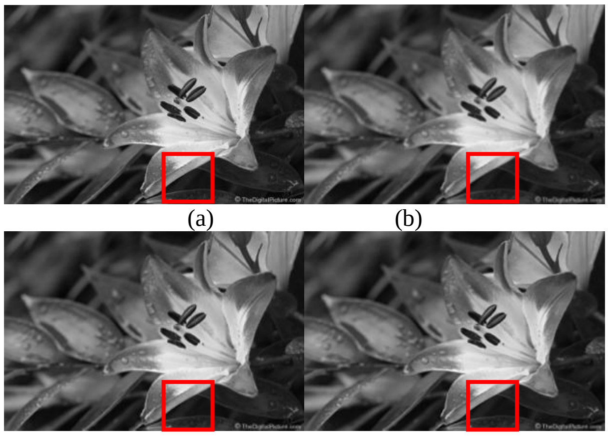
4. EXPERIMENTAL RESULTS

In order to fairly evaluate the performance of our algorithm, we compared the proposed algorithm with bi-cubic interpolation and Fan's algorithm, i.e., one of existing SR algorithm [13] for various test images. We employed four well-known 512×512 still images: *lena*, *pirate*, *barbara*, and *baboon*, and two 500×333 images: *flower* and *bee* (see lower two image of Figure 7). For training, we used 16 digital camera images of 500×333 (also see Figure 7) downloaded from http://www.the-digital-picture.com/Gallery/.

This paper considered a scaling ratio of 1/2. So, the LR images were produced from the corresponding HR images by using anti-aliasing filtering and down-sampling. The patch size was set to 7×7 and the patches were overlapped every 4 pixels. Initial ordinary dictionary was constructed with about 100,000 primitive patch pairs extracted from the upper eight training images in Fig. 7. Then, we reduced the size of the initial dictionary by 1/10 and 1/20, respectively as in subsection 3.2. Here, we utilized bilinear interpolation and Laplacian filter for initial up-scaling and high pass filtering, respectively. The residue dictionary was learned using lower eight images in Fig. 7. The dictionary size was set to a half of the ordinary dictionary size so as to reduce the memory cost. Also, we restricted the number of iteration of *K*-means clustering to 10 as a termination condition.

Table 1. PSNR comparison [dB]. The numbers in parenthesis indicates the reduction ratios of the dictionary size.

Test images	bi-cubic	Fan's (1/20)	Fan's (1/10)	Proposed (1/20)	Proposed (1/10)
lena	33.15	34.35	34.46	34.89	34.86
barbara	24.89	25.18	25.22	25.33	25.32
baboon	24.25	24.64	24.69	24.91	24.90
pirate	30.34	31.20	31.32	31.68	31.72
flower	32.43	33.67	33.73	34.37	34.50
bee	28.23	29.16	29.36	29.68	29.70



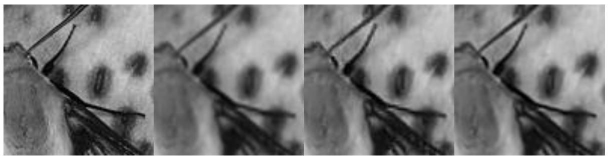
(a) (b)

(c) (d)

Fig. 9. Up-scaled images of *flower* image using 10,000 training examples (a) original (b) bi-cubic (c) Fan's (d) proposed.

(a) (b) (c) (d)

Fig. 10. Magnified images of the red box in Fig. 9 (a) original (b) bi-cubic (c) Fan's (d) proposed.



(a) (b) (c) (d)

Fig. 11. Results for a part of *bee* image in Fig. 7 (a) original (b) bi-cubic (c) Fan's (d) proposed.

Actually, in order to store 100,000 patch pairs in dictionary, the proposed algorithm requires a large memory size of about 15MBytes (MB). Note that our algorithm reduces such a huge dictionary size up to 1.5MB (1/10) or 0.75MB (1/20). On the other hand, we obtained the dictionary for Fan's algorithm by regularly sampling the same number of LR-HR patch pairs as our algorithm. As a result, the total size of dictionary of the proposed algorithm is equivalent to that of Fan's algorithm.

Table 1 shows PSNR comparison results for various test images. For reduction ratio of 1/10, the proposed algorithm provides higher PSNRs of 2 dB than bi-cubic at maximum for *flower* image. For the same reduction ratio, the proposed algorithm shows higher PSNRs of about 0.8 dB than Fan's algorithm for *flower* image. Note that even when the dictionary size becomes much smaller, i.e., reduction ratio of 1/20, the proposed algorithm still maintains higher PSNRs than the bi-cubic and Fan's algorithm.

Figure 8 shows the interpolated images of *lena* when the dictionary size is only 10,000. We can observe that the proposed algorithm provides better visual quality than the bi-cubic and Fan's algorithm. Note that Fan's algorithm shows some annoying artifacts in diagonal edge due to mismatching of LR patches as the dictionary size decreases. Also, we can see such artifacts in Figures 9, 10 and 11. Note that the proposed algorithm not only provides much better details than the previous schemes, but also shows high visual quality close to original images in Fig. 10 and Fig. 11.

5. CONCLUDING REMARKS

This paper proposed a fast image super-resolution algorithm which reduced the size of trained dictionary with minimal performance degradation. At the learning phase, the number of sample patch pairs in dictionary is noticeably reduced by grouping similar patches using K-means clustering. At the synthesis phase, one best-matched patch for each input low resolution patch is selected from the dictionary. Finally, the corresponding high frequency patch is used to reconstruct the output high resolution patch. Thus, the proposed algorithm realizes fast image super-resolution with the reduced dictionary size, while providing reasonable visual quality.

ACKNOWLEDGMENT

This work was sponsored by ETRI System Semiconductor Industry Promotion Center, Human Resource Development Project for SoC Convergence.

REFERENCES

[1] R.G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transaction on Acoustics, Speech and Signal processing*, vol. 29, no. 6, Dec. 1981, pp. 1153-1160.

[2] H.S. Hou and H.C. Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Transaction Signal processing*, vol. 26, no. 6, Dec. 1978, pp. 1153-1160.

[3] J.Allebach and P.W. Wong, "Edge-directed interpolation," *Proc. Int. Conf. Image Processing*, Sept. 1996, pp.707-710.

[4] X. Li and M. Orchard, "New edge directed interpolation," *IEEE Transaction on Signal Processing*, vol. 10, no. 10, Oct. 2001, pp. 1520-1527.

[5] Q. Wang and R. Kreidieh, "A new orientation-adaptive interpolation method," *IEEE Trans. Image Processing*, vol. 16, no. 4, Apr. 2007, pp. 889-900.

[6] S. M. Kwak, J. H. Moon, and J. K. Han, "Modified cubic convolution scaler for edge-directed nonuniform data," *Optical Engineering*, vol. 46, no. 10, Oct. 2007, p. 107001.

[7] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion and transparency," *J. Visual Commun. Image Represent*, 1993, pp. 324-335.

[8] C. Liu and H.Y. Shum, "Fundamental limits of reconstruction based super-resolution algorithms under local translation," *IEEE Trans. Pattern Anal. Machine Intell*, vol. 26, no. 1, Jan. 2004, pp. 83-97.

[9] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, 2003, pp. 21-36.

[10] W. T. Freeman, T.R. Jones and E.C. Pasztor, "Example based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, Oct. 2002, pp. 56-65.

[11] J. Sun, N.N. Zheng, H. Tao and H.Y. Shum, "Image hallucination with primal sketch prior," *Comp. Vision Pattern Recog*, 2005, pp. 729-736.

[12] H. Chang and D. Yeung, "Super-resolution through neighbor embedding," *Proc. IEEE Conf. Comp. Vision Pattern Recog.*, 2004, pp.275-282.

[13] W. Fan and D. Yeung, "Image hallucination using neighbor embedding over visual primitive Manifolds," *Comp. Vision Pattern Recog*, 2007, pp. 1-7.

[14] Z. Lin and J. He, "Limits of learning-based superresolution algorithms," *Int. J. Computer Vision*, vol. 80, no. 3, 2008, pp. 406-420.

[15] M. Varma and A. Zisserman, "A statiscal approach to texture classification from single images," *Int. J. Computer Vision*, vol. 62, no. 1-2, 2005, pp. 61-81.

[16] C. Kim, K, Choi and J. Ra, "Improvement on learning-based super-resolution by adopting residual information and patch reliability," *IEEE Int. Conf. Image Processing*, 2009, pp. 1197-1200.