Single Textual Image Super-Resolution Using Multiple Learned Dictionaries Based Sparse Coding

Rim Walha[1,2], Fadoua Drira[1], Franck Lebourgeois[2], Christophe Garcia[2], and Adel M. Alimi[1]

[1] University of Sfax, ENIS, REGIM, BP 1173, Sfax, 3038, Tunisia
{rim.walha,fadoua.drira,adel.alimi}@ieee.org
[2] University of Lyon, INSA-Lyon, CNRS, LIRIS, UMR5205, F-69621, France
{franck.lebourgeois,christophe.garcia}@insa-lyon.fr

Abstract. In this paper, we propose a new approach based on sparse coding for single textual image Super-Resolution (SR). The proposed approach is able to build more representative dictionaries learned from a large training Low-Resolution/High-Resolution (LR/HR) patch pair database. In fact, an intelligent clustering is employed to partition such database into several clusters from which multiple coupled LR/HR dictionaries are constructed. Based on the assumption that patches of the same cluster live in the same subspace, we exploit for each local LR patch its similarity to clusters in order to adaptively select the appropriate learned dictionary over that such patch can be well sparsely represented. The obtained sparse representation is hence applied to generate a local HR patch from the corresponding HR dictionary. Experiments on textual images show that the proposed approach outperforms its counterparts in visual fidelity as well as in numerical measures.

Keywords: Super-resolution, sparse coding, multiple learned dictionaries, textual image.

1 Introduction

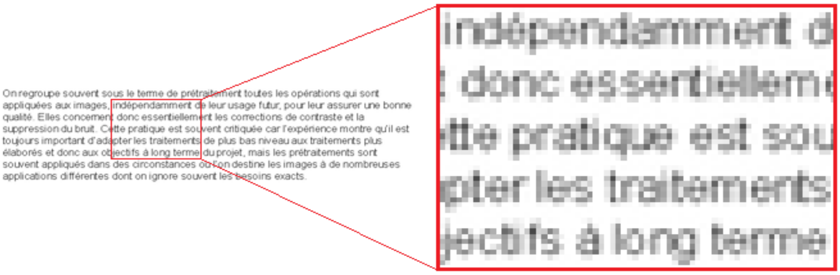
The problem of producing a HR image from an observed LR image is referred as Single Image Super-Resolution (SISR). Such problem has become an important research area due to the rapidly increasing need of high quality images in media applications. Indeed, the super-resolution could resolve some imperfections of hardware devices and also could guarantee a better utilization of the High-Definition displays capabilities. A variety of methods have been proposed in the literature to solve the SISR task. Most of these approaches have been concentrated on natural images with very limited application on the textual ones. Thus, this work aims to tackle this lack by investigating the SISR task applied on poorly resolved textual images.

The emergence of wide collections of LR scanned textual images in digital libraries introduces the need for efficient SR methods. In fact, such images are poor in visual quality and are characterized by a lack of details. For instance, text embedded in a LR image contains degraded characters which are not only disagreeable to view

A. Petrosino (Ed.): ICIAP 2013, Part II, LNCS 8157, pp. 439–448, 2013.

on a display device, but they also pose serious challenges to document recognition, search and retrieval in document images, etc. These degradations are typically produced by optical blur, spatial sampling and noise. Figure 1 shows an example of a LR textual image from which a region is enlarged in order to have a closer look.

Recently, the Sparse Coding (SC) technique has attracted increasing interest due to its effectiveness in various reconstruction tasks like SISR. In this paper, we propose a new SISR approach based on SC whose underlying idea is to suggest that an image patch can be sparsely represented from a suitable dictionary. Motivated by the important role of the dictionary in SC theory, we propose to use multiple dictionaries in order to well represent the properties of characters. Such dictionaries are learned from a large LR/HR patch pair database partitioned into several clusters by performing an intelligent clustering. Given multiple dictionaries, a reconstruction scheme is suggested to enhance the super-resolution process via exploiting the similarity to clusters and then adaptively selecting the appropriate dictionary to characterize the local LR image patch. The impact of the proposed approach is studied visually and quantitatively on LR textual images and interesting results have been achieved. The rest of this paper is organized as follows: Section 2 presents a brief review of related works on SISR task. Then, section 3 details the multiple learned dictionaries based SC approach proposed for the SR of a textual image. Experiments and comparative studies with results generated by other SR approaches are provided in section 4. Finally, conclusions and some perspectives are given in Section 5.



On regroupe souvent sous le terme de prétraitement toutes les opérations qui sont appliquées aux images, indépendamment de leur usage futur, pour leur assurer une bonne qualité. Elles concernent donc essentiellement les corrections de contraste et la suppression du bruit. Cette pratique est souvent critiquée car l'expérience montre qu'il est toujours important d'adapter les traitements de plus bas niveau aux traitements plus élaborés et donc aux objectifs à long terme du projet, mais les prétraitements sont souvent appliqués dans des circonstances où l'on destine les images à de nombreuses applications différentes dont on ignore souvent les besoins exacts.

Fig. 1. Example of a low-resolution textual image

2 Related Works

Many methods have been proposed in the literature for the SISR task. They are broadly classified into three categories: interpolation, regularization and learning based approaches. The first category which is the most commonly used one relies on the convolution of the image with kernels such as linear, cubic or higher order. Such process is simple but it tends to generate noticeable artifacts along edges because it treats the whole image in the same way. Adaptive interpolation treating every image part differently has been introduced to improve the image's sharpness [2, 10]. However, such interpolation is very limited while generating high frequency details.

The regularization based approaches, such as [3, 8], try to find the degrading model which simulates the passage from a HR image to its corresponding LR version. Then, the HR image is reconstructed by solving the inverse problem of the degrading model and including a priori reconstruction constraints. In such reconstruction, it's difficult to find the degrading model accurately. This category of approach is still suffering from non-natural artifacts.

In order to overcome the above drawbacks of interpolation and regularization based approaches, the learning based approaches which model the relationship between LR and HR images from a training database have been proposed. Their goal is not only to maintain the sharpness of the image, but also to recover new missing HR details that are not explicitly found in the LR image and assumed to be available in the training database. Several model have been used in the literature to estimate the local image structures including the gradient profile prior [16], the Markov Random Field (MRF) [7], the neural networks [15], etc. A common drawback of these models is that they heavily rely on enormous databases of millions LR/HR patch pairs and therefore have an intensive computation. More recently, SC based approaches have been suggested for the SISR task. Using the SC principle mentioned above, more patches can be represented using a smaller training database than the above learning approaches. The dictionary is a key for the successful of a SC based approach. The authors of [14, 20] constructed prototype dictionaries by randomly sampling raw patches from training images. In [1, 9, 19], dictionary learning algorithms was developed to reduce the complexity of SC under prototype dictionaries. Recently, SC based approaches relying on multiple dictionaries are introduced [21, 22]. Such dictionaries are learned from a clustered database by imposing in advance the number of clusters. All of these listed works have been successfully applied to natural, human face and synthetic aperture radar images. Other methods have been tailored to the SR of textual images. For example, Freeman et *al.* [6] proposed an exemplar-based approach which mapped blocks of the LR image into predefined HR blocks. In [5], the authors applied this approach on textual images. Nevertheless, the results depend heavily on the examples of the training set and, more precisely, on the type font which must be known in advance. Luong and Philips [11, 12] demonstrated that the estimation of the restored pixel intensity can be based on information retrieved from the whole image, thereby exploiting the presence of repeating characters in the image. Their method started with character segmentation which is not usually evident in LR textual images. A recent method based on SC is proposed to the SR of textual image [17]. It is based on two coupled dictionaries learned from a generic LR/HR patch pair database.

In this paper, we propose a learning based approach for the SR of single textual image using SC. The following section details the proposed approach.

3 SISR Via Multiple Learned Dictionaries Based Sparse Coding

To address the SISR problem by using SC, we divide the issue into two phases. Firstly, a proposed learning phase described in section 3.1 attempts to collect a training database from which multiple dictionaries will be learned. The main idea of the proposed learning strategy based on an intelligent clustering is to find more appropriate dictionaries adapted to the properties of characters. Secondly, a proposed reconstruction phase detailed in section 3.2 tries to infer a HR version of the input LR image by designing a sparse coding based scheme via the multiple learned dictionaries. The key idea is to guide the reconstruction to adaptively select the appropriate dictionary in order to better recover each local patch.

3.1 Dictionaries Learning Phase

The SC relies on the use of a dictionary. In the case of SC based SR, a HR dictionary and a LR dictionary are generally needed. Rather than generating two coupled dictionaries [17, 19, 20], we propose to learn multiple coupled dictionaries from examples of character images. This allows us to find dictionaries more adapted to the properties of characters. In this work, several high-quality character images are created by discretizing vector fonts via the graphic library FreeType [23]. For each character, we produce a large variety of sizes, styles (italic, non-italic) and fonts (serif, sans-serif) currently used in textual documents, signs, bills, etc. Therefore, a collection of HR character images is obtained. From each image, we extract several HR patches p_h^k that are localized along the edges of character. In fact, shapes, orientations and positions of edges are very interesting references to describe characters. After that, we generate for each HR patch p_h^k the corresponding LR patch p_l^k by blurring, downsampling and scaling-up via bicubic interpolation. This leads to the collection of a generic LR/HR patch-pairs $\{p_l^k, {}_h^k\}_k$database.

Training database contains numerous patterns which are very different due to their shapes, sizes, orientations and positions in the image patches. Single HR dictionary learned from all these data is not guaranteed to be suitable for the SR task even if it has large number of atoms. Moreover, this leads to an intensive computation complexity. To overcome these limitations and to take full advantage of large training database, we propose to learn multiple dictionaries from a clustered database. We haven't prior knowledge about the database to guide the clustering process such as the number of clusters whose correct choice is often ambiguous. Therefore, unlike [22], we partition the HR training database into several clusters in unsupervised fashion by using an "intelligent" version of the K-means method, referred as iK-means [13]. This clustering method determines automatically the number of clusters and initial cluster centers for K-Means using the anomalous pattern algorithm. The application of such clustering method on our database provides K final cluster centers $\{p_{C^i}, i = 1..K\}$ and K clusters $\{C^i, i = 1..K\}$ gathering similar patches in the same group.

Given the HR and the corresponding LR image patches of each cluster C^i, we turn to learn two coupled LR/HR dictionaries $\{D_l^i, D_h^i\}$ by using joint SC method [19]. The goal of this learning method, is to have the same sparse representation for each LR/HR image patch pair. Figure 2 displays some examples of the coupled LR/HR learned dictionaries whose atoms are shown as 7 × 7 pixel image. We can observe that each coupled dictionaries can describe the intrinsic geometrical structure of the corresponding training cluster. Thus, the proposed learning strategy can provide more appropriate dictionaries representing each cluster. The production of these dictionaries enables the reconstruction phase described in the following subsection.

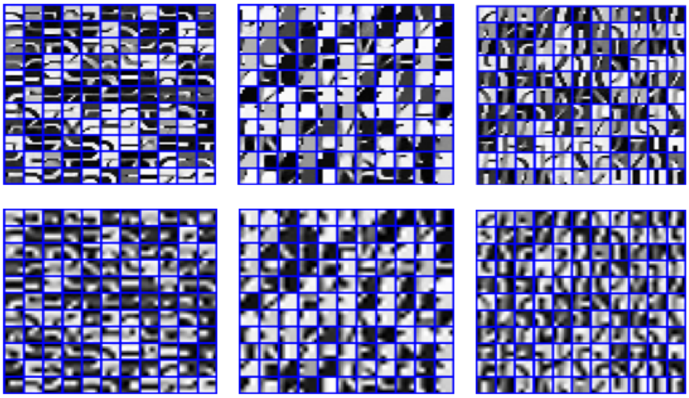


Fig. 2. Examples of coupled LR/HR dictionaries generated by the proposed learning phase. Top row: The HR dictionaries. Bottom row: The corresponding LR dictionaries.

3.2 Reconstruction Phase

The reconstruction phase aims to recover a HR image from the input LR image. It consists of two consecutive steps: local reconstruction and global reconstruction. First, the local reconstruction that is based on the sparse coding theory is applied to recover lost high-frequency for local details. Second, the global reconstruction is performed to remove possible artifacts generated by the first step.

The starting point of the local reconstruction is the input LR image. The classical way of beginning a SR process is by up-sampling the LR image via bicubic interpolation. The interpolated image is considered as the LR image to be processed. LR patches are crawled in raster scan with overlapping between adjacent patches. Instead of representing each LR patch y_j by a set of features [20, 21], we rely in our setting directly on pixels values composing the patch from which the mean pixel value m_j is subtracted. Based on the assumption that patches of the same cluster live in the same subspace, we seek for each local LR patch y_j which training cluster it corresponds to. Specifically, K Euclidean distances are calculated between the LR patch and the K cluster centers $\{p_{C^i}, i = 1..K\}$. The label i of the nearest cluster is then found according to: $\min\|y_j - p_{C^i}\|_2^2$. Via this strategy, we guide the local reconstruction of y_j to choose the coupled LR/HR dictionaries learned from the selected nearest cluster. Based on the SC theory, the LR patch y_j can be coded as a sparse linear combination of atoms from the chosen LR dictionary D_l^i. This can be mathematically written as:

$$(P_0) : \min_{\alpha}\|\alpha\|_0 \quad s.t. \quad \|y_j - D_l^i\alpha\|_2^2 \leq \rho \tag{1}$$

where α is the sparsest representation of y in D_l^i and ρ characterizing an allowable reconstruction error. Because of solving the optimization problem (P_0) is often difficult, Chen et *al.* [4] proposed that as long as α is sufficiently sparse, the problem (P_0) can be substituted by instead minimizing the l_1-norm as follows :

$$(P_1) : \min_{\alpha} \|\alpha\|_1 \quad s.t. \quad \left\|y_j - D_l^i\alpha\right\|_2^2 \le \rho \tag{2}$$

Several algorithms have been proposed in the literature to solve (P_1) [9]. In our implementation, the feature-sign search algorithm is selected because of its efficiency and significant speedup. According to the appropriate LR dictionary D_l^i, it finds the optimal solution $\tilde{\alpha}$ that is then applied to generate a local HR patch x_j from the corresponding HR dictionary D_h^i based on : $x_j = D_h^i\alpha + m_j$. Subsequently, the initial HR image X_0 is obtained by simply averaging the values in the overlapped regions to enforce compatibility between adjacent patches. After that, the global reconstruction is applied on X_0 to eliminate the local reconstruction errors and to ensure consistency with the LR input image. It is based on the assumption that the observed LR image Y consists of a blurred and downsampled version of a HR image X of the same scene: $DHX = Y$, where D and H represents respectively a downsampling operator and a blurring filter. In order to enforce such assumption in the global reconstruction, the following optimization problem should be solved:

$$\tilde{X} = \arg min_X \|X - X_0\| \quad s.t. \ DHX = Y \tag{3}$$

To finalize this task, we use Back-Projection method originally developed in computer tomography and applied to SR in [19, 20]. Finally, a bilateral filtering is performed on the recovered HR image to better preserve edges and hence to further enhance the global reconstruction. Because the human visual system is more sensitive to the illuminance changes, our algorithm is applied only to the illuminance channel in the case of color images. So, we just predict color channels using bicubic interpolation. Based on the above description, the entire proposed reconstruction phase can be summarized as algorithm 1 and an overview of the proposed approach is depicted in figure 3.

Algorithm 1: SISR via multiple learned dictionaries based SC.

```
Input: a LR image $Y$, the multiple coupled LR/HR dictionaries $\{D_l^i, D_h^i, i = 1..K\}$
learned from $K$ clusters of the training database.
1. For each LR patch $y_j$ of $Y$ crawled in raster scan from the upper-left corner with
   overlapping in each direction,
   (a) Subtract the mean pixel value $m_j$ from $y_j$.
   (b) Calculate $K$ Euclidean distances between the LR patch $y_j$ and the $K$ cluster
       centers $\{p_{C^i}, i = 1..K\}$.
   (c) Find the label $i$ of the nearest cluster according to: $\min\left\|y_j - p_{C^i}\right\|_2^2$ in order to
       select the appropriate coupled LR/HR dictionaries $\{D_l^i, D_h^i\}$.
   (d) Perform the SC under the LR learned dictionary $D_l^i$ to find the solution $\alpha$ of
       (2).
   (e) Compute the local HR version $x_j$ of the LR patch $y_j$ by $x_j = D_h^i\alpha + m_j$.
   End.
2. Merge the overlapped HR patches to generate the initial HR image $X_0$.
3. Perform the global reconstruction by solving the optimization problem (3) and then
   applying a bilateral filtering to generate the HR image $\tilde{X}$.
Output: HR image $\tilde{X}$.
```

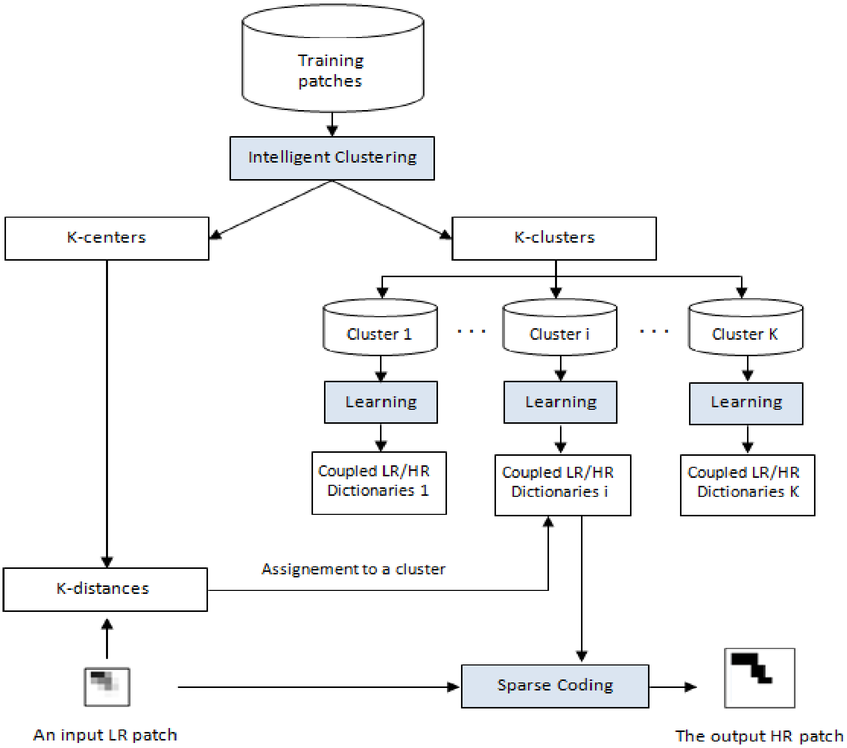


Fig. 3. Overview of the proposed approach

4 Experiments and Evaluations

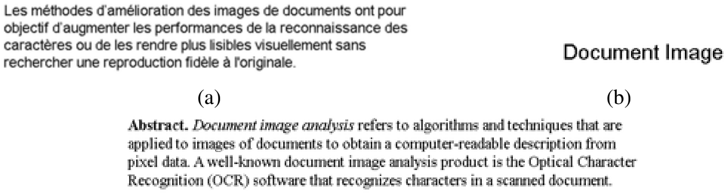
In this section, experimental SR results achieved by applying the proposed SISR method and other SR methods on different LR textual images are given. Results are evaluated both visually and quantitatively on text image quality.

In the proposed learning phase, we generate 2480 high-quality character images from which 124000 HR patches of size 7 × 7 and the corresponding LR patches are collected to form the LR/HR patch-pairs training database. By performing the intelligent clustering iK-means method [13], this database is divided into 13 clusters. The multiple LR/HR dictionaries learned from the clustered database are of size 49 × 128. In fact, each dictionary has 128 atoms and each atom is represented by 49 pixel values. In the proposed reconstruction phase, a sliding window of size 7 × 7 with 5 pixels overlap is selected to scan the input LR image interpolated by bicubic interpolation and then to construct the HR image.

To investigate the performance of the proposed approach, we compare it with bilinear interpolation, bicubic interpolation and other recently published SISR methods based on SC [17, 19] that rely on two coupled LR/HR dictionaries. Such dictionaries (each one is composed by 512 atoms) are learned from the same database of high-quality character images prepared in this study and used in [17] and in Yang's method

[19] which is hence adapted to textual image SR. Tests are performed for the magnification factor 2 on the LR textual images shown in figure 4. These images, which are generated by blurring and down-sampling by a factor of 2 of the HR ground truth images, contain different texts size (10, 12, 14), style (italic, non-italic, bold, non-bold) and font (serif, sans-serif). Results are evaluated quantitatively based on the widely used metrics in image processing for recovery including the Root Mean Square Error (RMSE), the Peak SNR (PSNR) and the Structural SIMilarity index (SSIM) [18]. Table 1 compares the measurement values of images recovered by different SR methods. According to this table, the proposed approach achieves the best results in terms of all these measurements. More precisely, we conclude that under the same training database, our method performs better than the other sparse coding based methods involved in this study.

In order to have a closer look, we enlarge some regions in the LR images (b) and (c) and in the reconstructed images as shown in figure 5. Compared to the original LR images in figure 5(a), we can notice clearly significant improvements in visual quality in the results produced by each SR method. Indeed, the letters are much better readable and blur is heavily reduced. On the other hand, we can observe that images recovered by our SR method (figure 5(e)) are clearer and sharper at edges and have better visual quality than those produced by interpolation methods which generate blur effects and Yang's method that generates significant artifacts appearing near edges.



Les méthodes d'amélioration des images de documents ont pour objectif d'augmenter les performances de la reconnaissance des caractères ou de les rendre plus lisibles visuellement sans rechercher une reproduction fidèle à l'originale.

(a)

Document Image

(b)

Abstract. *Document image analysis* refers to algorithms and techniques that are applied to images of documents to obtain a computer-readable description from pixel data. A well-known document image analysis product is the Optical Character Recognition (OCR) software that recognizes characters in a scanned document.

(c)

Fig. 4. Illustration of low-resolution textual images

Table 1. RMSE, PSNR and SSIM results of textual images recovered by different SR methods

Image	Measures	Bilinear interpolation	Bicubic interpolation	Yang et *al.* [19]	Walha et *al.* [17]	Proposed method
(a)	RMSE	43.254	39.581	34.541	32.355	**28.034**
	PSNR	15.410	16.180	17.364	17.931	**19.176**
	SSIM	0.760	0.805	0.825	0.864	**0.914**
(b)	RMSE	28.033	25.784	24.171	23.533	**20.928**
	PSNR	19.177	19.903	20.464	20.697	**21.715**
	SSIM	0.887	0.907	0.902	0.918	**0.943**
(c)	RMSE	45.614	42.739	38.488	37.190	**34.124**
	PSNR	14.948	15.514	16.424	16.722	**17.469**
	SSIM	0.701	0.746	0.771	0.805	**0.852**

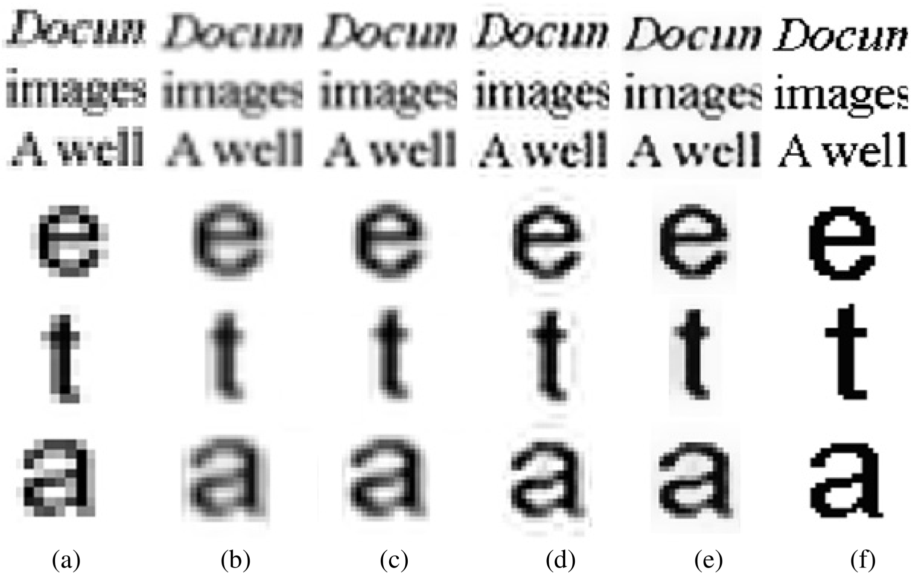


Fig. 5. Visual comparison of enlarged regions from LR and textual images recovered by different SR methods. (a) LR image. (b) Bilinear interpolation. (c) Bicubic interpolation. (d) Yang's method [19]. (e) Proposed method. (f) HR ground truth image.

5 Conclusions and Perspectives

We conclude this paper by summarizing the main contributions of this work. In fact, an intelligent clustering of the collected training database is incorporated into the learning of multiple coupled LR/HR dictionaries. Such learning strategy doesn't require prior knowledge about the number of clusters or the number of dictionaries. It thus differs from that of the existing SISR methods which impose in advance such parameters. Given multiple learned dictionaries, a reconstruction scheme is proposed to adaptively select the appropriate dictionary and then to better recover each local patch. The performance of the proposed approach is evaluated visually and quantitatively on textual images and compared to other SR methods. Experimental results show the effectiveness of the proposed approach.

The suggested approach is not limited to textual images and can be employed for the SR of other type of images if we use another training database. An alternative extension to our work is to apply the proposed approach in the SR of video.

References

1. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. 54(11), 4311–4322 (2006)

2. Battiato, S., Gallo, G., Stanco, F.: Smart interpolation by anisotropic diffusion. In: ICIAP, pp. 572–577 (2003)
3. Ben-Ezra, M., Lin, Z.C., Wilburn, B.: Penrose pixels: Super-resolution in the detector layout domain. In: ICCV (2007)
4. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM Review 43(1), 129–159 (2001)
5. Dalley, G., Freeman, W., Marks, J.: Single-frame text super-resolution: a bayesian approach. In: ICIP, pp. 3295–3298 (2004)
6. Freeman, W., Jones, T., Pasztor, E.: Example-based super-resolution. IEEE Computer Graphics and Applications 22(2), 55–65 (2002)
7. Freeman, W.T., Pasztor, E.C., Carmichael, O.T.: Learning low-level vision. In: IJCV (2000)
8. Irani, M., Peleg, S.: Motion analysis for image enhancement: Resolution, occlusion and transparency. JVCL 4(4), 324–335 (1993)
9. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: Advances in Neural Information Processing Systems (NIPS) (2007)
10. Li, X., Orchard, M.T.: New edge-directed interpolation. IEEE Trans. Image Process. 10, 1521–1527 (2001)
11. Luong, H., Philips, W.: Non-local text image reconstruction. In: ICDAR, vol. 1, pp. 546–550 (2007)
12. Luong, H., Philips, W.: Robust reconstruction of low-resolution document images by exploiting repetitive character behavior. IJDAR 11(1), 39–51 (2008)
13. Mirkin, B.G.: Clustering for data mining: a data recovery approach, vol. 3. CRC Press (2005)
14. Mairal, J., Sapiro, G., Elad, M.: Learning multiscale sparse representations for image and video restoration. SIAM Multiscale Model. Simul. 7(1), 214–241 (2008)
15. Staelin, C., Greig, D., Fischer, M., Maurer, R.: Neural network image scaling using spatial errors. In: Tech. Rep. HPL-2003-26R1, HP Labs (2003)
16. Sun, J., Xu, Z., Shum, H.: Image super-resolution using gradient profile prior. In: CVPR (2008)
17. Walha, R., Drira, F., Lebourgeois, F., Alimi, A.M.: Super-resolution of single text image by sparse representation. In: Proc. of Workshop on Document Analysis and Recognition, pp. 22–29 (2012)
18. Wang, Z., Bovik, A.C., Sheikh, H.R.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. 13(4), 600–612 (2004)
19. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution via sparse representation. IEEE Trans. Image Process. 19(11), 2861–2873 (2010)
20. Yang, J., Wright, J., Huang, T., Ma, Y.: Image Super-Resolution as Sparse Representation of Raw Image Patches. In: CVPR (2008)
21. Yang, S., Wang, M., Chen, Y., Sun, Y.: Single-Image Super-Resolution Reconstruction via Learned Geometric Dictionaries and Clustered Sparse Coding. IEEE Trans. Image Process. 21(9) (2012)
22. Yang, S.Y., Liu, Z.Z., Jiao, L.C.: Multitask dictionary learning and sparse representation based single-image super-resolution reconstruction. Neurocomputing 74(17), 3193–3203 (2011)
23. `http://www.freetype.org`