Contextual Dictionaries for Image Super Resolution

Wei Yu, Hongxun Yao, Xianming Liu, Rongrong Ji, Xiaoshuai Sun, Pengfei Xu

School of Computer Science and Technology, Harbin Institute of Technology

{w.yu, h.yao, xmliu, rrji, xssun, fxu}@hit.edu.cn

ABSTRACT

Traditional super resolution (SR) methods based on sparse representation have shown their excellent performance, however, the methods usually perform even worse when the input images and the training samples are diverse. Considering this problem, this paper presents a novel super resolution method based on sparse representation with contextual dictionary. Through adopting discriminative features instead of common features, the method train and use contextual dictionary in the SR process. Additionally, the method uses the first-order and second-order gradients of patch as representation, which ensures the neighbor information is introduced in the SR processing. The experiment results demonstrate the performance of this method has been promoted than other traditional method

Categories and Subject Descriptors

I.4.4 [**Image Processing and Computer vision**]: Restoration

General Terms

Algorithms, Experimentation

Keywords

Super resolution, **s**parse representation, contextual dictionary, K-SVD

1. INTRODUCTION

Image super resolution (SR) is drawing more research attention in recent years and has been applied to various domains, including image compression, video enhancement and surveillance, which aims at estimating a high resolution (HR) image from a low resolution (LR) one or an array of LR ones. However, the biggest challenge lies in its ill-posed nature in reconstruction process caused by insufficient information provided by low-resolution image when targeting at the high resolution one.

To overcome this problem, it is necessary to utilize some constraints or prior knowledge in super resolution process to make up the missing information. Roughly speaking, three types of knowledge are used: smooth contour assumption, reconstruction constraints, and co-occurrence priors between LR and HR images.

ICIMCS'11, August 5-7, 2011, Chengdu, Sichuan, China.

Firstly, to add smooth contour assumption corresponds to interpolation-based methods, which apply non-uniform interpolation to produce an improved high-resolution image. In [2], Zhang and Wu optimize all these interpolation-based algorithms on the local covariance of image signal and achieve state-of-the-art results. This kind of approaches is easy to implement and more direct to understand. However, it suffers blurring the discontinuities and cannot guarantee the satisfaction of the reconstruction constraint, thus performs poorly in complex images. To obtain a better result, additional steps have to be carried out to remove the blurring and noises in the LR images.

Another group of approaches aim at meeting the reconstruction constraints from low-resolution to high-resolution images, and various kinds of priors on the HR images are used in order to regularize the ill-posed inverse problem [1]. They rely on the relationships between LR and the relative HR images, but to satisfy the reconstruction constraints, they cannot ensure contour smoothness. Especially, the performance of these reconstruction-based algorithms degrades rapidly when the desired magnification factor is large or the number of available input images is small. In these cases, the result may be smooth, lacking important high-frequency details [5].

Compared with previous two groups of approaches, learning-based SR algorithms attempt to capture the co-occurrence priors between LR and HR images patch, and then generate HR images patches through a learned model. In [6], Freeman et al. brought out an example based learning strategy where the relationship between image (observation) and scene (state) pairs are modeled by a Markov Random Field (MRF) and solved by belief propagation. In [7], Sun et al. extend this approach by using the primal sketch priors, resulting more satisfying result. As recent development in sparse signal representation in computer vision, it is successfully adopted in SR problem by recovering high-dimensional signals from their low-dimension version. Yang et al., from the perspective of compressive sensing, tried to recover the sparse representation coefficients of each LR patch regarding a dictionary composed of LR patches, and then reconstructed HR patches using the recovered coefficients in terms of the corresponding dictionary composed of HR patches in [8]. It has already been shown that learning based methods achieve better performance by learning a co-occurrence model to extend the information from LR to HR images.

However, we observe that the construction of dictionary in sparse representation based SR algorithms suffers from the following problems:

1. Usage of Contextual Information: Current dictionary constructing processes merely utilize image patches' content, and ignore contextual information such as relationships between images patches. This causes the dictionary not robust and efficient from the cognitive aspect;

2. Transferring Ability: the sparse representation dictionary is built on a limited training set, trying to find a complete set of bases for most cases. Therefore it usually performs even worse when the input images and training samples are diverse. Our experiments show this problem more directly.

In this paper, we propose a contextual dictionary assumption based on these two considerations. The contextual here refers to the local relationships between patches rather than using entire samples to represent target. We demonstrate that using contextual related sub-set of patches to reconstruct HR images outperforms traditional methods, partly because the constructed dictionary will reduce the useless bases and represent discriminative features more efficiently. This could also be supported in a vision manner, that the discriminative features are more powerful and intuitive than common features. Another contribution of our work is we facilitate a cross dataset Super Resolution reconstruction by emphasizing the discriminative bases via contextual dictionary.

Two datasets are used for validation: 1) Common used benchmark images and 2) 500 Natural images from Corel dataset. Compared with traditional approaches adopts the common bases, a significant improvement is shown in our experiments.

2. SPARSE REPRESENTATION IN SUPER RESOLUTION

The key idea of Super Resolution via sparse representation is to find a mapping between coefficients of LR patches and HR ones, based on separate learned dictionaries. In this section, we will formulate this process and propose a potential solution to improve the quality of HR images.

2.1. Super Resolution via Sparse Representation

Given a dictionary $D \in R^{n\times k}$ with $k > n$, implying D is an over-complete dictionary. And suppose a signal $x \in R^n$ can be represented sparsely over this dictionary, that is, x can be represented as a sparse linear combination with respect to D:

$$x = D\alpha \quad with \quad \|\alpha\|_0 << k \tag{1}$$

where $\alpha \in R^k$ is a coefficient vector, $\|\alpha\|_0$ is the zero-norm of α which stands for the count of the nonzero entries.

In the problem of Super Resolution, image patches are formulated as signals. We treat the HR patches as the high-dimensional signals, and the LR patches as low-dimensional signals. Therefore, the sparse representation can be perfectly applied to the image super resolution, which is an inverse problem basically.

Generally, given a known input image f, it can be divided into two parts v and n, that is:

$$f = v + n, \tag{2}$$

where v is the structural component, which contains important geometric information and composed by the sub-image of the smooth slice region; n contains the texture or noise part of the oscillation characteristics, which made up by the similar cyclical and repetitive small scale component of the image. Considering the practical performance and the image component n, the representation of patch x may either be exact $x = D\alpha$, or approximate $x \approx D\alpha$.

2.2. Further View of Sparse Representation SR

According to [8-9], sparse representation based SR problems typically follow this criterion:

$$\min \lambda \|\alpha\|_1 + \frac{1}{2}\left\|\widetilde{D}\alpha - \tilde{y}\right\|_2^2 \tag{3}$$

From the cognitive perspective, the dictionary D can be divided into D_d and D_c which represents the discriminative and common shared bases for sparse reconstruction. Thus, the upper bound of Eq. (3) can be estimated as:

$$\left\|\widetilde{D}\alpha - \tilde{y}\right\|_2^2 = \left\|[D_d, D_c]\alpha - \tilde{y}\right\|_2^2 \le \left\|D_d\alpha - \tilde{y}\right\|_2^2 + \left\|D_c\alpha - \tilde{y}\right\|_2^2 \tag{4}$$

Therefore, it is reasonable to relax the optimization as:

$$\min \lambda \|\alpha\|_1 + \frac{1}{2}\left\|D_d\alpha - \tilde{y}\right\|_2^2 + \left\|D_c\alpha - \tilde{y}\right\|_2^2 \tag{5}$$

Actually, it is difficult to find a complete set of common bases to cover all cases such that the reconstruction problem is able to be perfectly solved. However, it is reasonable to find a correlated discriminative set of bases D_d which could represent image details more precisely under the useful contextual information approximately. This is because the common bases D_c will introduce some useless components which reduce the performance of sparse representation.

In this paper, to solve this problem, we propose a contextual dictionary for super resolution. The motivation is to find a more compact and discriminative dictionary D_d within a contextual related set of patches, in order to make the optimization convergent more effectively. Then the optimization becomes:

$$\min \lambda \|\alpha\|_1 + \frac{1}{2}\left\|D_d\alpha - \tilde{y}\right\|_2^2 \tag{6}$$

It is easier to achieve. Besides, we want to illustrate the rule for sparse representation based super resolution: using contextual related subset of dictionary will work better than trying with a complete set of bases.

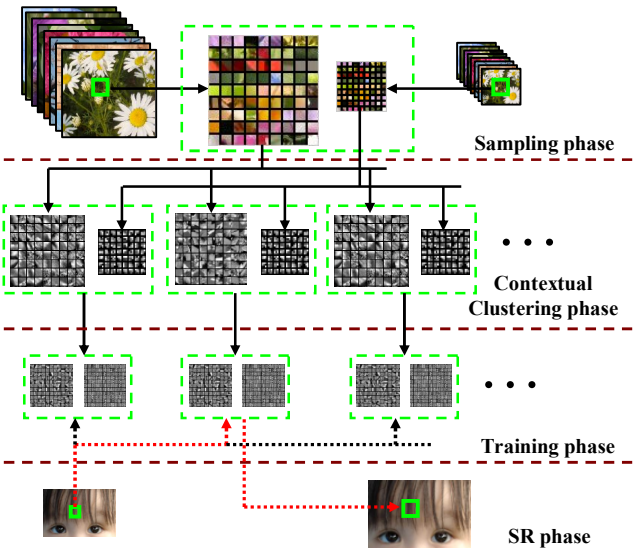

Figure 1. The framework of Image Super Resolution based on Contextual Dictionary.

3. CONTEXTUAL DICTIONARY

Based on the above discussion and analysis, we propose a novel strategy for super-resolution, which uses contextual information sufficiently in this paper. Fig.1 shows the basic framework of the proposed method including 4 phases: Patch Sampling, Contextual Clustering, Dictionary Training and SR phases.

The *Contextual Dictionary* refers to the one constructed from related patches, and we simply judge the relation by calculating the similarities between different patches.

Given training image set:

$$\{I_h, I_l\}, I_h = \{i_{h,1}, i_{h,2}, ..., i_{h,K}\}, I_l = \{i_{l,1}, i_{l,2}, ..., i_{l,K}\}$$

with I_h being the high resolution images and I_l the low resolution images, a random sampling is firstly used to collect a patch collection, represented as:

$$P = \{(p_{hi}, p_{li}), i = 1, 2, ..., N\}.$$

To reflect the contextual information, a K-means clustering is performed on P based on patch similarities, and obtain contextual patch set:

$$P = \mathrm{U}\{P_i\}_{i=1}^{M}$$

And for each P_i, a similar dictionary learning process as [11] is utilized to train a separate dictionary D_i, the component of contextual dictionary D, following optimization of Eq. (6).

In the super-resolution stage, each input image is segmented into patches, and for each patch, a most suitable dictionary D_i is selected from D to reconstruct the high resolution patch. The whole algorithm is shown as follow:

Algorithm: Contextual Dictionary for SR

Input: training images $\{I_h, I_l\}$, a LR image Y.

Output: the HR image X through SR

1. Randomly Sampling on $\{I_h, I_l\}$, and get patch set:

$$P = \{(p_{hi}, p_{li}), i = 1, 2, ..., N\},$$

HR size 9*9 and LR size 3*3

2. Contextual Clustering:

Perform K-means on P: $P = \mathrm{U}\{P_i\}_{i=1}^{M}$

3. ***Foreach*** $P_i \in P$

4. Generating dictionary D_i=$\{D_{i,h}, D_{i,l}\}$as [8]

5. $D = D \mathrm{U} D_i$

6. ***End***

7. Sample input Y into 3*3 patches: P_Y

8. ***Foreach*** patch $p_{i,y}$ in P_Y

9. Select a most related P_m

10. Construct patch $p_{i,x}$ using D_m=$\{D_{m,h}, D_{m,l}\}$

11. Put patch $p_{i,x}$ into HR image X

12. ***End***

13. **Return** X

---

4. EXPERIMENTS

4.1 Preliminary

In our experiments, we choose a magnify factor of 3 for the LR images, so that our result can compare with other's under the same condition. We set LR patch size as 3×3, and set HR patch size as 9×9 correspondingly. We sampled 800,000 patch pairs randomly from natural images downloaded from internet, including fruit, tree and flowers, to train the dictionaries for HR image patches and LR image patches.

In order to reduce the computation cost, we choose only illumination channel, since we believe observers are more sensitive to illumination for color images. For enhancing the contextual characteristic, we also introduce first-order and second-order gradients of patches as the representation for LR patches on illumination channel in order to integrate the neighbor information intro patches[12-13].

During the clustering process, we classify the patch pairs into 4 sub-sets based on the LR patches' feature, which reduce the time cost significantly in dictionary training. Figure 2 is one example of D_h used in our experiment.

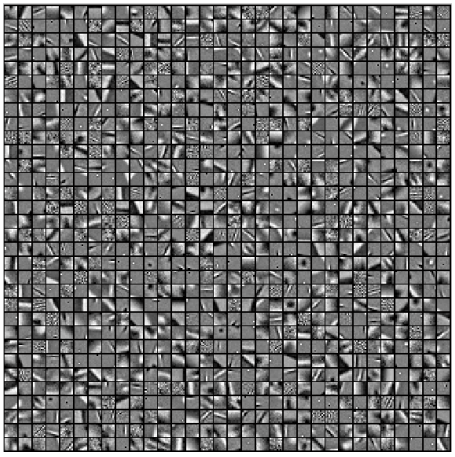

Figure 2. A HR dictionary of our method.

4.2 Experiments Results

We randomly selected 500 natural images from Corel image data set for super-resolution processing, including bears, eagles, horses, helicopters and so on. We use SSIM as quantitative evaluation metric, which is classic measurement of the similarity between two images. The average SSIM value of our method is 0.8239, which is better than 0.8231 of [8]. Visual comparisons are shown in Figure 3. Table 1 reports some results. We also trim some un-frequent bases and use these trimmed dictionary to reconstruct HR images, indicates as Trim Bases in Table 1. SR_SP is the performance reported in [8], and we also repeat their experiments on our larger dataset, indicated as SR_ SP2.

Table 1 Average SSIM of Image Super Resolution

	Our Method	Trim Bases	SR_SP	SR_SP2
Average SSIM	**0.8778**	0.8769	0.8806	0.8664

It is obvious we achieve similar performance as [8] using less patches and less time. And compared with the performance of

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

Figure. 3 The above images magnified by a factor of 3. (a): bicubic interpolation, (b): original image, (c): our method, (d): SR method based on single dictionary [8], (e): input image.

repeated experiments of [8] on our dataset, our contextual dictionary obtained better SSIM. Besides, by trimming some less used bases on contextual dictionary, the performance drops little. The convincing performance demonstrates our contextual assumption and show the possibility of better SR solutions.

5. CONCLUSION

In this paper, we analyzed the problem in current image super resolution using sparse representation, and make a contextual assumption. A contextual dictionary strategy for image SR is further proposed. This method solve the performance degradation when the training images and testing are not the same. The robust and accurate results shows that our method outperforms the state-of-the-art approaches in various situations, and demonstrate our contextual assumption.

6. ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (No. 61071180) and NEC cooperative project (No. Lc04-20101201-03).

7. REFERENCES

[1] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super-resolution. *IEEE TIP*, 2004

[2] X. Zhang and X. Wu, Image interpolation by adaptive 2D autoregressive modeling and soft-decision estimation. *IEEE Trans. Image Process*, vol. 17, no. 6, pp.887-896, Jun. 2008.

[3] Haifeng Li, Hongkai Xiong, Liang Qian. Image super-resolution with sparse representation prior on primitive patches. *VCIP*, 2010.

[4] Yu-Wing Tai, Shuaicheng Liu, Michael S. Brown and Stephen Lin. Super Resolution using Edge Prior and Single Image Detail Synthesis. *CVPR*, 2010.

[5] S. Baker and T. Kanade. Limits on super-resolution and how to break them. IEEE TPAMI, 24(9):1167-1183, 2002

[6] Freeman, W., Pasztor, E., and Carmichael, O., Learning low-level vision. *IJCV*, 2000

[7] Sun, J., Zheng, N.-N., Tao, H., and Shum, H.-Y., Image hallucination with primal sketch priors. *CVPR*, 2005

[8] Yang. J., Wright. J., Huang. T., and Ma. Y., Image super-resolution as sparse representation of raw image patches. *CVPR*, 2008

[9] D.L. Donoho. For most large under determined systems of linear equations, the minimal l1–norm solution is also the sparsest solution. *Comm. On pure and Applied Math*, Vol. 59, No. 6, 2006

[11] Ron Rubinstein, Michael Zibulevsky and Michael Elad. Double Sparsity: Learning Sparse Dictionaries for Sparse Signal Approximation. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, VOL. 58, NO.3, MARCH 2010

[12] Rongrong Ji, Hongxun Yao, Xiaoshuai Sun. Toward Semantic Embedding in Visual Vocabulary. *CVPR*, 2010

[13] H. Chang, D. –Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. *CVPR*, 2004