EFFICIENT SPARSE REPRESENTATION BASED IMAGE SUPER RESOLUTION VIA DUAL DICTIONARY LEARNING

Haichao Zhang[†‡], *Yanning Zhang*[†] *and Thomas S. Huang*[‡]

[†] School of Computer Science, Northwestern Polytechnical University, Xi'an, China 710129
[‡] Beckman Institute, University of Illinois at Urbana-Champaign, IL, USA 61801

ABSTRACT

Super resolution is of great use in many visual media related scenarios, such as displaying low resolution contents on High-Definition TV(HD-TV). In these scenarios, the efficiency of the super resolution process is of vital importance. This paper presents a fast learning based super-resolution method. The proposed method speeds up the sparse representation based super-resolution method by learning a dual dictionary, and replaces the sparse recovery step by simple matrix multiplication, which is much more computationally efficient. Experiments demonstrate that the proposed method can generate desirable super-resolved images with significant computational advantages.

***Index Terms*—** Super-resolution, sparse representation, dual dictionary learning

1. INTRODUCTION

Image super-resolution (SR) aims to estimate a high-resolution image (HR) from a single or a set of low-resolution (LR) observations , which is an important technique for displaying low resolution media such as video data obtained from low resolution imaging devices such as cell phones, on high-resolution devices such as HD-TVs. SR has a long history and is a still very active field [1, 2, 3, 4, 5]. Conventional methods typically perform motion estimation followed by frame fusion [1, 6]. Recently, some algorithms without explicit motion estimation are proposed [3, 5] to avoid the parametric motion assumption. For single image super-resolution, the simplest method is interpolation-based method, *e.g.*, bilinear and bicubic interpolation. Such analytical interpolation methods do not exploit the underlying structures in natural images, such as edges, thus usually blurring the fine details and introducing artifacts. More advanced interpolation methods take the underlying structures in the image into consideration during interpolation. For example, Li proposed an edge directed interpolation method that the pixels coming from the same structure are given large weights while others are given small weights in interpolation [7]. Takeda *et al* developed a kernel regression framework which takes into account the underlying structural kernel for regression [8].

As an alternative, learning based method [2, 9, 10] have shown impressive results for super-resolution. Freeman *et al.* proposed to using an external database of high-low resolution patches to help with the prediction of the missing high-frequency information [2]. While this method is capable of adding some high-resolution information to the estimated super-resolved image, it also tends to introduce artifacts near structural areas, due the in lack of similar enough patches in the generic database. Chang *et al.* proposed an example based SR method by making a 'manifold' assumption for the high and low resolution image patches. Specifically, the 'manifold' assumption states that the manifold structures for the low and high resolution patches are approximately the same. Therefore, the local manifold structure estimated from the low resolution patches can be transferred to the high-resolution patches for high-resolution patch estimation. Recently, based on compressive sensing theory, Yang *et al* proposed a sparse representation based SR method [10], based on the assumption that the high resolution and low resolution patches have the same sparse representation coefficients with respect to a high-resolution dictionary and a corresponding low-resolution dictionary, respectively. In [11], they further improve this method by training compact high and low resolution dictionaries rather than using the raw image patches as in [10]. This method shows impressive SR results, with more details and less artifacts. Furthermore, SR methods exploiting self-similarities [4] and exploiting both local structural regularities as well as the image self-similarities [5] are two recent promising directions.

We focus on sparse representation based SR method in this paper. While it can generate *state-of-the-art* results, one limitation is that it has to perform the sparse recovery procedure for each patch, which is time consuming. In this paper, we aim to speed up the sparse representation based SR method while maintaining its performance. We address this problem with the method of learning a dual dictionary.

The rest of this paper is organized as follows. In Section 2, we review some related works briefly. Then we introduce the learning method for coupled dictionary together with the low resolution dual dictionary for efficient SR in Section 3. Experiments are conducted in Section 4 and we conclude this paper in Section 5.

978-1-61284-350-6/11/$26.00 ©2011 IEEE

2. RELATED WORKS

In this section, we first review the sparse representation based SR method. Then we summarize the the conventional dictionary learning method (also referred to as *primal dictionary learning*) as well as the method proposed recently for learning a dual dictionary.

2.1. Sparse Representation based Super Resolution

Inspired by the recent progress of compressive sensing, which states that a sparse signal can be recovered with high probability via a few incoherent measurements, Yang *et al.* proposed a sparse representation based SR method [10, 11], by assuming that the low resolution and high resolution patches share the same underlying sparse representations. Treating the given low resolution patches as measurements of the underlying sparse representation vector, they proposed to reconstruct the high resolution image via the sparse representation vector recovered from the low resolution measurements. Specifically, given a dictionary pair $\{D_h, D_l\}$ of high and low resolution, for a low resolution patch y_i from low resolution image y, we can solve the sparse representation problem as:

$$\begin{aligned} \hat{z}_i = \arg\min_{z_i} \|z_i\|_1 \\ \text{s.t. } \|D_l z_i - y_i\|_2^2 \leq \epsilon, \end{aligned} \tag{1}$$

or equivalently, as

$$\hat{z}_i = \arg\min_{z_i} \|D_l z_i - y_i\|_2^2 + \lambda \|z_i\|_1. \tag{2}$$

With the recovered sparse representation vector z_i, we can reconstruct the high-resolution patch as:

$$x_i = D_h z_i. \tag{3}$$

Collecting all the high resolution patches $\{x_i\}$ to their corresponding positions and performing normalization, we can get an estimation of the high resolution image x.

To better respect the same sparse representation assumption, Yang *et al.* proposed to learn a coupled dictionary via the following optimization process [11]:

$$\{D_h, D_l, Z\} = \arg\min_{D_h, D_l, Z} \|X_c - D_c Z\|_F^2 + \hat{\lambda}\|Z\|_1 \tag{4}$$

where $X_c = [\frac{1}{\sqrt{M}} X_h^\top, \frac{1}{\sqrt{N}} X_l^\top]^\top$ is the concatenated high and low resolution patches. $D_c = [\frac{1}{\sqrt{M}} D_h^\top, \frac{1}{\sqrt{N}} D_l^\top]^\top$ is the coupled high-low resolution dictionary trained from X_c. M and N are the dimensions of the high and low resolution patches. $Z = [z_1, z_2, \cdots]$ is the matrix collecting sparse representation vectors as columns.

It is shown that sparse representation based SR method can generate desirable super resolved image. However, one limitation of this method is that we have to solve the sparse representation problem (1) or (2) for each patch at each pixel location, thus is computationally expensive. It is desirable to speed up this method to improve its applicability in more realistic situations. In the following, we propose to improve the efficiency of the sparse representation based SR via the technique of dual dictionary learning.

2.2. Primal and Dual Dictionary Learning

Given training samples $X = [x_1, x_2, \cdots, x_L] \in \mathbb{R}^{d \times L}$, a conventional dictionary learning problem trains a dictionary $D = [d_1, d_2, \cdots, d_K] \in \mathbb{R}^{d \times K}$ via the following ℓ_1-norm regularized optimization problem [12, 13, 14]:

$$\begin{aligned} \{\hat{D}, \hat{Z}\} = \arg\min_{D,Z} \|X - DZ\|_F^2 + \lambda\|Z\|_1 \\ \text{s.t.} \quad \|D_i\|_2^2 \leq 1 \end{aligned} \tag{5}$$

where D_i denotes the ith column of D. This problem is convex with respect to D or Z, but not simultaneously. Typical approach for solving this problem is to minimizing over one variable each time while keeping other variables fixed. Such a scheme is alternated over all the variables until converge [12, 13, 14]. When optimizing over Z, it is a ℓ_1-norm regularized problem, which is non-differentiable. Different algorithms have developed in the past, *e.g.*, feature-sign method [12] and proximal gradient based iterative shrinkage/thresholing (IST) method [13, 14]. When optimizing over D, the problem reduces to a quadratically constrained quadratic minimization problem. The norm constraint on the atoms in D is to avoid trivial solutions to this optimization problem.

To learn the dual dictionary, one can introduce the dual matrix to D as $C = [c_1, c_2, \cdots, c_K]^\top \in \mathbb{R}^{K \times d}$. While each column of D can be regarded as atoms for synthesizing, the rows of C can be regarded as sparse filters which can be convolve with the input signal x to generate a code $z \in \mathbb{R}^K$. To learn both D and C, the following optimization problem is proposed in [15]:

$$\begin{aligned} \{\hat{D}, \hat{C}, \hat{Z}\} = \arg\min_{D,C,Z} \|X - DZ\|_F^2 \\ + \eta\|Z - CX\|_F^2 + \lambda\|Z\|_1 \\ \text{s.t.} \|D_i\|_2^2 \leq 1, \quad \|C^i\|_2^2 \leq 1, \end{aligned} \tag{6}$$

where C^i denotes the ith row of C. As one can see, compared with the primal dictionary learning problem 5, the dual dictionary learning problem 6 has an additional constraint on the distance between sparse coefficient matrix Z and the estimated one via the dual dictionary. Also, norm constraint is also applied to the learned filters, which are the rows of C.

Eqn.(6) aims to learn a dictionary D which can sparsely represent the training samples X well while learning a analysis operator C which can generate the sparse codes with small errors via simple linear operation.

3. EFFICIENT SUPER RESOLUTION WITH DUAL DICTIONARY LEARNING

The dual dictionary learning method is proposed in [15] aims to learn a linear mapping in the case of a single dictionary. Following this scheme, in this paper, we further extend it to learn the dual filters together with coupled dictionary training procedure for sparse representation based SR. Treating the dictionary D as a decoding basis or synthesis prior, we can introduce its dual during learning, which can be regarded as analysis prior or filers [16], and is also related to the works on learning sparsify filters for natural images [17]. In the following, we first propose our model for learning the dual dictionary for low resolution patches together with coupled high-low resolution dictionary learning. Then we propose an efficient learning method following.

We use the dual dictionary learning scheme for achieving fast sparse representation based super resolution. The motivation is obvious: by learning an analysis operator C dual to the dictionary D, we can approximate the sparse representation step with simple matrix multiplication, which is much more efficient. Therefore, it is desirable to introduce the dual dictionary learning ability to the coupled dictionary learning method modeled as (4). However, the formulation (6) is not proper, as it aims to learn a dual dictionary which is dual to the coupled dictionary, therefore, the approximate sparse representation is obtained as the product of the dual dictionary with the coupled high-low resolution patches. But in the case of super-resolution, the high-resolution patch is unknown and is what we are trying to estimate. For this purpose, we propose the following optimization model for learning coupled dictionaries with dual for efficient super resolution:

$$\begin{aligned} \{D_h, D_l, C_l, Z\} = \arg\min_{\{D_h, D_l, C_l, Z\}} &\|X_c - D_c Z\|_F^2 \\ + \eta\|Z - C_l X_l\|_F^2 &+ \lambda\|Z\|_1 \\ \text{s.t.} \quad \|D_i\|_2^2 \leq 1, &\quad \|C^i\|_2^2 \leq 1 \end{aligned} \tag{7}$$

where $X_c = [\frac{1}{\sqrt{M}}X_h^\top, \frac{1}{\sqrt{N}}X_l^\top]^\top \in \mathbb{R}^{(M+N)\times L}$ and $D_c = [\frac{1}{\sqrt{M}}D_h^\top, \frac{1}{\sqrt{N}}D_l^\top]^\top \in \mathbb{R}^{(M+N)\times K}$. $C_l \in \mathbb{R}^{K\times N}$ is the dual of D_l. In this formulation, the a dual dictionary corresponding to the low-resolution patches is learned together with the coupled high-low resolution primal dictionary.

By minimizing (7) over the concatenated training patches X_c, we can learn a coupled dictionary D_c and C_l–the dual of the low resolution dictionary D_l, simultaneously. Note that the reason we only learn the dual for the low resolution dictionary D_l is that we only need to perform sparse recovery (sparse analysis) on low resolution images, *i.e.* for a low resolution patch y_i,

$$z_i = C_l y_i \tag{8}$$

Then the corresponding high-resolution patch x_i can be constructed via (3).

3.1. Proximal Gradient for Dictionary Learning

We aim to learn a coupled dictionary D_c, a dual dictionary for low resolution patches C_l and the sparse representation matrix Z by solving the optimization problem (7). Similar to (5), problem (7) is not convex for all the variables simultaneously, but is convex with respect to one variable while keeping others fixed. Specifically, we address each of the variables Z, D and C in the sequel following similar approaches as in [15].

Z-subproblem: Sparse Coding
Fixing D_c and C_l, problem (7) reduces to:

$$\begin{aligned} Z = \arg\min_Z &\|X_c - D_c Z\|_F^2 \\ &+ \eta\|Z - C_l X_l\|_F^2 + \lambda\|Z\|_1, \end{aligned} \tag{9}$$

which gives the following updating formula for Z given the previous estimation Z^k [15]:

$$\begin{aligned} Z^{k+1} =& T_{\tau/2\sigma_Z}\left[\left(1 - \frac{\eta}{\sigma_Z}\right)Z^k\right. \\ &\left. + \frac{1}{\sigma_Z}\left(D_c^\top(X_c - D_c Z^k) + \eta C X_l\right)\right], \end{aligned} \tag{10}$$

where $T_\lambda[X]$ is the soft-thresholding operator defined element-wise on matrix X as:

$$(T_\lambda[X])_{i,j} = \text{sign}(X_{i,j}) \max\left\{|X_{i,j}| - \lambda, 0\right\}. \tag{11}$$

D_c-subproblem: Coupled Primal Dictionary Updating
Fixing Z and C_l, problem 7 reduces to:

$$\begin{aligned} D_c = \arg\min_{D_c} &\|X_c - D_c Z\|_F^2 \\ &\|D_i\|_2^2 \leq 1. \end{aligned} \tag{12}$$

The coupled dictionary can be updated by taking a gradient step and then project onto the unit ball as:

$$D^{k+1} = \pi_{D_c}\left(D_c^k + \frac{1}{\sigma_{D_c}}(X_c - D_c^k Z)Z^\top\right). \tag{13}$$

C_l-subproblem: Low-res Dual Dictionary Updating
To optimize C_l, the optimization problem becomes:

$$\begin{aligned} C_l = \arg\min_{C_l\}} &\eta\|Z - C_l X_l\|_F^2 \\ \text{s.t.} \quad &\|C^i\|_2^2 \leq 1, \end{aligned} \tag{14}$$

which has a similar form as (12). Therefore, C_l can be updated by taking a gradient descent step and then perform projecting for each row of C_l onto the unit ball. Specifically, C_l can be updated as:

$$C_l^{k+1} = \pi_{C_l}\left(C_l^k + \frac{1}{\sigma_{C_l}}(Z - C_l^k X_l)X_l^\top\right). \tag{15}$$

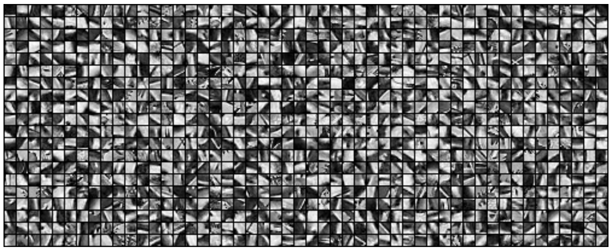


Fig. 1. High resolution dictionary D_h trained via (7) using 8000 patches randomly sampled from natural images. Only the first 1000 out of the 1024 atoms are shown here.

$\pi(x) = \frac{x}{\max\{1,\|x\|\}}$ is the projection onto the unit ball. π_D denotes the operation of applying the projection to each column of D while π_C applying projection on each row of C.

In the above updating equations, the parameters σ_Z, σ_{D_c} and σ_{C_l} are the step-sizes corresponding to sparse representation matrix Z, coupled dictionary D_c and the dual low resolution dictionary C_l respectively. To achieve fast convergence in dictionary learning, appropriate choices of these step-sizes are crucial. Detailed discussions are made in [15]. Moreover, the accelerated first order method for proximal gradient descent can be used for further improving convergence [18, 19].

4. EXPERIMENTS

In this section, we conduct several experiments to evaluate the effectiveness of the proposed method, both in terms estimation quality as well as computational efficiency. Following [11], we use 3×3 patches for low resolution images, which are upsampled to 6×6 and convolved with the following 1-D derivative filters for the extraction of more salient features from the low resolution image:

$$f_1 = [-1, 0, 1], \qquad f_2 = f_1^\top,$$
$$f_3 = [1, 0, -2, 0, 1], \quad f_4 = f_3^\top.$$

Their outputs are concatenated with the corresponding mean-subtracted high-resolution patches for dictionary training. We evaluate the proposed method for single image SR with magnification factor of 3. For the low resolution image, there is an overlap of 1 pixel between adjacent patches, which corresponds to overlap of 3 pixels for the 9×9 high-resolution patches. We train the coupled dictionary with the low resolution dual via the procedure presented above using $L = 8000$ patches randomly sampled from natural images. A coupled dictionary D_c with 1024 atoms together with the low-resolution dual dictionary C_l are learned via (7). The learned high resolution dictionary D_h is shown in Figure 1.

We compare the SR performance of the proposed method with Nearest Neighbor Interpolation (NN), Bicubic Interpolation (BI), the original Sparse representation based SR method (SSR) [11] and proposed Dual dictionary based SR method

Table 1. Super resolution result comparison of estimation quality (magnification factor: 3).

Methods		NN	BI	SSR	DSR
girl	RMSE	6.945	5.910	5.659	5.768
	SSIM	0.743	0.776	0.797	0.792
flower	RMSE	4.511	3.520	3.298	3.480
	SSIM	0.843	0.881	0.892	0.883
koala	RMSE	9.263	7.643	7.219	7.349
	SSIM	0.770	0.815	0.838	0.831
castle	RMSE	13.442	12.474	12.057	12.304
	SSIM	0.764	0.785	0.801	0.791

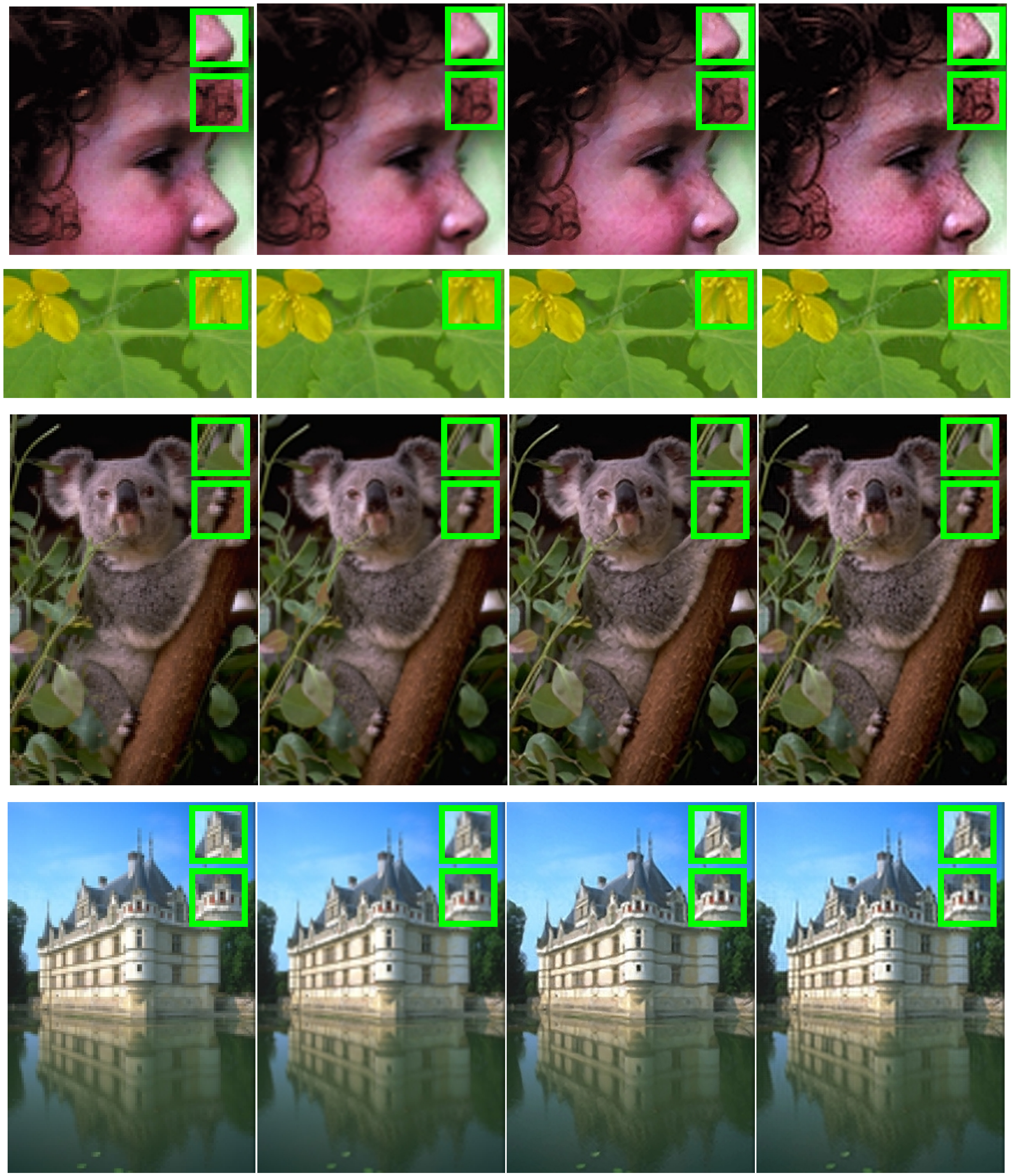
Table 2. Super resolution result comparison of time complexity (magnification factor: 3).

Images	size	SSR	DSR	Speedup
girl	86×85	12.066	**1.017**	11.86
flower	57×110	10.110	**0.788**	12.83
koala	480×321	30.652	**2.449**	12.52
castle	480×321	32.198	**2.421**	13.30

(DSR). The results of the four different algorithms on several test images are shown in Figure 2. As can be seen from Figure 2, the proposed method can generate similar results to the original SSR method, which are visually much better than conventional methods such as bicubic interpolation.

To quantitatively compare the speed and accuracy of the SSR method with proposed DSR method, we summarize the quantitative comparison results in in Table 1 and Table 2. The Root Mean Square Error (RMSE) and Structural SIMilarity (SSIM) [20] results are summarized in Table 1 while the computation time costs during reconstruction process are presented in Table 2. As can be seen from these two tables, the proposed method can generate SR results with desirable quality in terms of RMSE and SSIM (Table 1) while enjoys a significant improvement in computational efficiency compared to the original sparse representation SR method (Table 2). This indicates the potential application of the proposed method in many real scenarios where realtime or near-realtime performance is a must.

It is noteworthy to point out that while some encouragement results can be obtained via such a simple linear model, there are still some limitations. The linear model is limited in the modeling power thus can not capture more complex non-linearity relations for the mapping from image patches to sparse codes, which explains the slightly worse results compared with the original method. Furthermore, some artifacts can occur near the structures in the super-resolved images, due to the lack of the smoothing ability as in the ℓ_1-norm regularized regression model used in the original sparse representation based SR method.



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

Fig. 2. Single image super resolution results ‘girl’, ‘flower’, ‘koala’ and ‘castle’ images with magnification factor of 3. (a) Nearest neighbor interpolation, (b) Bicubic interpolation, (c) origina sparse representation based SR method [11] and (d) proposed dual dictionary based sparse representation SR method.

5. CONCLUSION

An efficient sparse representation based super resolution method is proposed in this paper. The main improvement lies in the approximate sparse coding procedure via an efficient linear predictive model. This linear model can approximate the sparse codes of a given low-resolution image patch by multiply it with a matrix, which is regarded as the dual to the low resolution dictionary. A coupled dictionary training procedure with the dual of the low resolution dictionary is developed with accelerated proximal gradient descent method. With the learned low resolution dual dictionary, we can replace the time-consuming sparse representation procedure with efficient matrix multiplication, thus speeds up the overall super resolution process. Experiments on single image super resolution tasks demonstrated the effectiveness of the proposed method. For future work, we would like to examine more complex productive models which can capture the non-linearities in sparse coding to further improve the super resolution results while maintain the efficiency.

Acknowledgement This work is supported by the U.S. Army Research Laboratory and U.S. Army Research Office under grant number W911NF-09-1-0383. This work is also supported by National Natural Science Foundation (No.60872145, No.60903126), Cultivation Fund from Ministry of Education (No.708085), National High Technology Program (No.2009AA01Z315) and Postdoctoral (Special) Science Foundation (No.20090451397, No.201003685) of China.

6. REFERENCES

[1] R.Y. Tsai and Thomas S. Huang, "Multiframe image restoration and registration," *Advances in Computer Vision and Image Processing*, pp. 317–339, 1984.

[2] W.T Freemann, T.R Jones, and E.C. Pasztor, "Fast and robust multiframe super resolution," *IEEE Computer Graphics and Applications*, , no. 2, pp. 56–65, 2002.

[3] Matan Protter, Michael Elad, Hiroyuki Takeda, and Peyman Milanfar, "Generalizing the non-local-means to super-resolution reconstruction," *IEEE TIP*, vol. 18, no. 1, pp. 36–51, 2009.

[4] Daniel Glasner, Shai Bagon, and Michal Irani, "Super-resolution from a single image," in *ICCV*, 2009.

[5] Haichao Zhang, Jianchao Yang, Yanning Zhang, and Thomas S. Huang, "Non-local kernel regression for image and video restoration," in *ECCV*, 2010.

[6] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE TIP*, vol. 13, no. 10, pp. 1327–1344, 2004.

[7] Xin Li and Michael T. Orchard, "New edge-directed interpolation," *IEEE Transactions on Image Processing*, vol. 10, pp. 1521–1527, 2001.

[8] Hiroyuki Takeda, Student Member, Sina Farsiu, Peyman Milanfar, and Senior Member, "Kernel regression for image processing and reconstruction," *IEEE Transactions on Image Processing*, vol. 16, pp. 349–366, 2007.

[9] Hong Chang, Dit-Yan Yeung, and Yimin Xiong, "Super-resolution through neighbor embedding," in *CVPR*, 2004.

[10] Jianchao Yang, John Wright, Thomas S. Huang, and Yi Ma, "Image super resolution as sparse representation of raw image patches," in *CVPR*, 2008.

[11] Jianchao Yang, John Wright, Thomas S. Huang, and Yi Ma, "Image super-resolution via sparse representation," *IEEE TIP*, vol. 19, no. 11, pp. 2861 – 2873, 2010.

[12] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y. Ng, "Efficient sparse coding algorithms," in *In NIPS*. 2007, pp. 801–808, NIPS.

[13] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro, "Online dictionary learning for sparse coding," in *ICML*, 2009.

[14] Rodolphe Jenatton, Julien Mairal, Guillaume Obozinski, and Francis Bach, "Proximal methods for hierarchical sparse coding," in *ICML*, 2010.

[15] Curzio Basso, Matteo Santoro, Alessandro Verri, and Silvia Villa, "Paddle: Proximal algorithm for dual dictionaries learning," Tech. Rep., Nov., 2010.

[16] Michael Elad, Peyman Milanfar, and Ron Rubinstein, "Analysis versus synthesis in signal priors," , no. 3, pp. 947–968, 2007.

[17] Stefan Roth and Michael J. Black, "Fields of experts: A framework for learning image priors," in *CVPR*, 2005.

[18] Y. Nesterov, "Smooth minimization of non-smooth function," *Math Program*, , no. 1, pp. 127–152, 2005.

[19] Stephen Becker, Jérôme Bobin, and Emmanuel J. Cands, "Nesta: A fast and accurate first-order method for sparse recovery," Tech. Rep., 2009.

[20] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, Student Member, Eero P. Simoncelli, and Senior Member, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 2004.