FINDING THE RIGHT EXEMPLARS FOR RECONSTRUCTING SINGLE IMAGE SUPER-RESOLUTION

Jiahuan Zhou and Ying Wu

Department of Electrical Engineering and Computer Science
Northwestern University, 2145 Sheridan Road, Evanston, IL

ABSTRACT

Exemplar-based methods have shown their potential in synthesizing novel but visually plausible contents for image super-resolution (SR), by using the implicit knowledge conveyed by the exemplar database. In practice, however, it is common that unwanted artifacts and low quality results are produced due to the using of inappropriate exemplars. How are the "right" exemplars defined and identified? This fundamental issue has not be well addressed in these methods. This paper proposes a novel solution to this issue by learning a new distance metric in the LR space, such that affinity structure of the LR space under the new metric is as close to that of the HR space. Based on this learned best metric, appropriate exemplars can be identified. In addition, the proposed method is able to automatically determine the appropriate number of exemplars to use. Extensive experiments have shown that our method is able to handle regions with different properties and to obtain visually appealing super-resolution results with sharp details and smooth edges.

Index Terms— Super-Resolution, Metric Learning, Right Exemplars

1. INTRODUCTION

Image super-resolution (SR) refers to the estimation or inference of a high-resolution (HR) image from one or multiple low-resolution (LR) images. This is an ill-posed problem, as there may be many potential HR solutions that produce the same LR observations. Therefore, image priors or knowledge are needed to regularize and further constrain the solution. Comparing to multi-image SR, e.g. [10][4][1][13], single image SR [6][7][8][17] is much more difficult since it is less constrained. Thus the image prior knowledge is much more important to it. This paper is focused on the single image SR problem.

Based on how image priors and knowledge are represented and used, the SR approaches can be roughly categorized into two types: (1) explicit knowledge for regularization, and (2) implicit knowledge for learning. Explicit image priors [7][8][5][14] favor HR images with sharp and smooth edges but tend to over-smooth the texture regions. In contrast, implicit image priors are conveyed by a set of corresponding LR-HR image pairs as exemplars. The SR performance depends not only on the quality of the exemplars, but also on how such implicit image prior is used.

The general idea of implicit exemplar-based approaches is based on the belief that similar LR exemplars to the LR input should bring "good" HR exemplars to estimate or reconstruct the HR output. Although exemplar-based methods are able to introduce new visual contents for texture regions, it is very tricky in practice to identify the "right" exemplars to use. Using wrong exemplars will produce unwanted artifacts like non-smooth boundaries, jagged and discontinuous edges. Therefore, how to use the implicit priors by identifying the right exemplars is critical and fundamental. This issue has not been well addressed in the context of image super-resolution.

The main contributions of this paper are as follows: (1) In this paper, we propose to learn a new distance metric in the LR space, such that the affinity structure of the LR space under the learned metric is as close as possible to the affinity structure of the corresponding HR space. This guarantees the validity of directly using the same reconstruction weights in both spaces. (2) Moreover, we propose a new method to automatically determine the appropriate number of exemplars to use. (3) In addition, we integrate both learning-based method and regularization-based method into a unified way, taking advantage of the fact that both two general approaches are complementary to each other.

2. PROPOSED SOLUTION

2.1. Issues in Exemplar-based SR Methods

The typical exemplar-based SR method is based on local linear reconstruction from the set of nearest neighbors obtained from an exemplar database $\mathcal{K}$, e.g., in [6][12]. For a LR input patch $\mathbf{X}$, its k-nearest neighbors in the exemplar database $\mathcal{K}$ are obtained based on a distance to $\mathbf{X}$. Denote the set of LR exemplars by $\mathcal{N}_l = \{L_1, L_2, \ldots, L_k\}$, and $\mathcal{N}_h = \{H_1, H_2, \ldots, H_k\}$ as the corresponding set of HR exemplars. To find the best reconstruction of $\mathbf{X}$ by $\mathcal{N}_l$, conventional methods usually solve a constrained least squares estimation problem:

$$\mathbf{w}^* = \arg\min \left\| \mathbf{X} - \sum_{i \in k} w_i L_i \right\|^2 \\ s.t.\ \sum_i w_i = 1 \ \ where \ \ w = [w_1, ..., w_k]^T \tag{1}$$

Its closed-form solution is given by:

$$\mathbf{w} = \frac{C^{-1}\mathbf{1}}{\mathbf{1}^T C^{-1}\mathbf{1}}, \ \ where \ \ C_{ij} = (\mathbf{X} - L_i)^T (\mathbf{X} - L_j) \tag{2}$$

Once the estimated reconstruction weights $\mathbf{w}$ for LR exemplars have been solved, they are directly applied to the corresponding HR exemplars to reconstruct the HR image patch, i.e., the desired HR patch $\mathbf{Y}$ is reconstructed by $\mathbf{Y} = \sum_i w_i H_i$. However, there are two important issues that have not been well addressed:

(1) The treatment of transferring the LR reconstruction weights $\mathbf{w}$ to HR exemplars for the HR reconstruction is taken for granted, but not grounded. In general, the nearest neighbors are identified based on the Euclidean distance in the LR space. However, some image patches that are close in the LR space may not be as close in the HR space. Therefore the metric in the LR space needs to be adjusted (see Sec. 2.3).

(2) The determination of the number of exemplars is not well answered. If we use a large number of exemplars, it is likely that some of them are actually false positives. They will be very detrimental to the HR reconstruction. If we use a small number, the reconstruction results may not be accurate. Thus, a solution is needed to determine a suitable number of exemplars (see Sec. 2.4).

2.2. Overview of Proposed Objective

Although this paper is focused on the exemplar-based implicit prior, to make it self-contained, it is necessary to give an overview of our entire SR formulation. By unifying both explicit regularization and implicit exemplar-based learning, our SR method has a general objective:

$$I^* = \arg\min_I \left\| I^l - (I * K) \downarrow \right\|_2^2 \\ + \mu \left\| \sum w_i I_i(I^l, \mathcal{K}) - I \right\|_M^2 + \lambda \|I\|_G, \tag{3}$$

where I is the HR output, I^l is the LR input, K is a spatial smooth filter, $*$ and $\downarrow$ represent the convolution and downsampling operations, respectively. This general objective consists of three components:

(1) *Data fidelity*. $\left\| I^l - (I * K) \downarrow \right\|_2^2$ is the fidelity, or likelihood term. L_2 norm is used to measure the distance because L_2 norm punishes more on reconstruction error.

(2) *Implicit prior*. $\left\| \sum w_i I_i(I^l, \mathcal{K}) - I \right\|_M^2$ is a new implicit prior term we propose and the major contribution of this paper. $\mathcal{K}$ is the LR-HR exemplar database that represents the implicit prior, and $\{I_i(I^l, \mathcal{K})\}$ is the set of identified HR exemplars from the exemplar database $\mathcal{K}$ by I^l, based on the distance metric M. This set of identified HR exemplars is linearly weighted by $\{w_i\}$ and combined to regularize the SR solution I. μ and λ are parameters used to balance the weights between terms. The details will be discussed in Sec.2.3.

(3) *Explicit prior*. $\|I\|_G$ represents the explicit prior term. Specifically, we adopt an edge smoothness model named *Soft-Cuts* [8] in this paper, which is a generation of Geocuts metric [2] for regularizing the image smoothness. Minimizing L_1 norm of $\|I\|_G$ is edge-preserving.

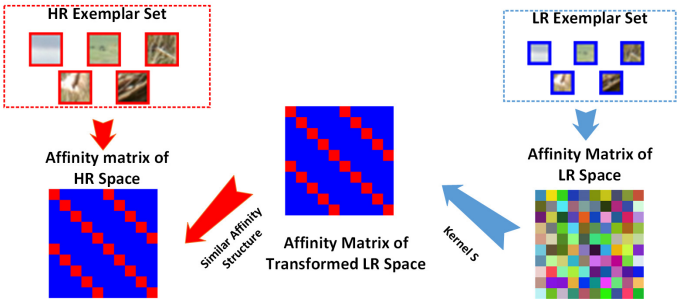

Fig. 1. The proposed metric learning method correctly refines the affinity structure of the LR space to make it much more similar to the affinity structure of the HR space. So the LR exemplars will share the same similarity relationships as that in the HR space with respect to the learned metric.

2.3. Learning to Find "Similar" LR Patches

In this paper, we propose to learn an appropriate Mahalanobis distance in the LR space as the similarity metric. Since the visual features for LR images are not reliable due to the missing of visual details, the similarity metric in the LR space needs external prior information to drive. Therefore we consider taking advantage of its corresponding HR space, because reliable visual features can be extracted from HR images with details. Another major advantage is that some sophisticated yet efficient features which are not suitable for LR scale can be readily used in the HR space. Therefore in this paper, a mixture of color and texture features is used for the representation in the HR space. The conventional color Gaussian Mixture Model (GMM) and the perceptual image quality assessment criteria proposed in [15] are adopted for feature representation. So the Euclidean distance can be easily used in the HR space as the similarity metric $\mathcal{D}_P(h_i, h_j)$.

Assume pair-wise training data $\{(l_i, h_i)\}_{i=1}^n$ are the LR-HR exemplar databases where $\{l_i\}_{i=1}^n$ represent LR exemplars and $\{h_i\}_{i=1}^n$ are the corresponding HR exemplars. With the proposed similarity metric $\mathcal{D}_P(h_i, h_j)$ in the HR space, the affinities for all pairs of the HR patches are computed by $v_{ij} = \exp\left(-\frac{\mathcal{D}_P(h_i, h_j)}{2\sigma_{hr}}\right)$. To reflect the relative connectivity,

we normalize the HR affinity matrix and obtain $P = [p_{ij}]$, where

$$p_{ij} = \frac{v_{ij}}{\sum_{k \neq i} v_{ik}}, \;\; p_{ii} = 0 \tag{4}$$

such that p_{ij} represents the nearest neighbor probability from one data point to another.

In the LR space, the distance metric is defined by a Mahalanobis distance M:

$$\begin{aligned} \mathcal{D}_M(l_i, l_j) &= (Al_i - Al_j)^T (Al_i - Al_j) \\ &= (l_i - l_j)^T M (l_i - l_j), \end{aligned} \tag{5}$$

where $M = A^T A$ is positive semi-definite and called the transformation kernel. M changes the affinity structure of the LR space. Similarly, we can construct the affinity among the LR patches $\{l_i\}_{i=1}^N$ with respect to M. This leads to a matrix $U = [u_{ij}]$ and its normalization $Q = [q_{ij}]$, where

$$\begin{aligned} u_{ij} &= \exp\left(-\tfrac{(l_i - l_j)^T M (l_i - l_j)}{2\sigma_{lr}}\right), and \\ q_{ij} &= \tfrac{u_{ij}}{\sum_{k \neq i} u_{ik}}, \;\; q_{ii} = 0 \end{aligned} \tag{6}$$

Affinity matrix Q is determined by the unknown Mahalanobis kernel M which needs to be learned. Therefore the objective of metric learning is to find the best M, such that the affinity structure of the LR space is as close to that in the HR space as possible. Choosing the KL divergence as distance measure, we get:

$$\begin{aligned} &M^* = \arg\min_M \sum_{i,j} KL[p_{ij} | q_{ij}] \\ &s.t. \;\; M \preceq PSD \end{aligned} \tag{7}$$

Eqn.(7) can be solved via gradient-based optimization algorithms. Denoted by $f(M) \triangleq \sum_{i,j} KL[p_{ij} | q_{ij}]$, we obtain

$$f(M) = \sum_{ij} p_{ij} \log p_{ij} - \sum_{ij} p_{ij} \log q_{ij} \tag{8}$$

Differentiating $f(M)$ with respect to the transformation kernel M yields a gradient rule that we can use to minimize the objective function:

$$\begin{aligned} \nabla f(M) &= \tfrac{1}{2\sigma_{lr}} \sum_{ij} (p_{ij} - q_{ij}) (l_i - l_j) (l_i - l_j)^T \\ M^{t+1} &\leftarrow M^t - \epsilon \nabla f(M^t) \end{aligned} \tag{9}$$

In order to guarantee M to be PSD, we need to project M back to the PSD cone by eliminating its negative components.

$$M^\circ = \sum_k \max(\lambda_k, 0) v_k v_k^T \tag{10}$$

where λ_k is the eigenvalue of M, and v_k is its corresponding eigenvector.

As local geometry is characterized by the similarity between one patch and other patches, the affinity structure in the HR space is preserved in the transformed LR space so the LR exemplars will share the same affinity relationships as that in the HR space with respect to the learned metric. Therefore once we obtain the reconstruction weights $\mathbf{w}$ in the transformed LR space, they can be transferred to the HR space for reconstructing an HR patch.

2.4. Determining the Best Number of Exemplars

Another important issue in exemplar-based SR approach is to determine the appropriate number of exemplars, once the initial set of "similar" exemplars are retrieved. Most existing methods, e.g. [6][12][11], simply use a fixed number or use a fixed similarity threshold. However, the drawback is obvious. Too few exemplars can not describe the input completely, while too many exemplars will include noises to ruin the reconstruction quality. Moreover, for different inputs, the numbers of needed exemplars can be quite different.

Inspired by [3][16], we propose an algorithm that utilizes Robust Principal Component Analysis (R-PCA) to automatically eliminate the false positive exemplars. Align the vectorized nearest neighbors $\{\mathcal{N}_l\}^k$ to form a new matrix N, the structure information S and the noise information E can be easily separated and extracted from N by the proposed algorithm, which are crucial for determining and eliminating the false positive exemplars. The Lagrangian formulation of the optimization problem is:

$$\min_{S,E} \|S\|_* + \lambda \|E\|_0 \;\; subj \;\; S + E = N \tag{11}$$

This convex problem can be easily solved by several different methods proposed in the literature[3]. Therefore appropriate "right" exemplars can be determined by the obtained error matrix E.

3. EXPERIMENTS

3.1. Exemplar Database and Model Parameters

The LR-HR exemplar database $\{(l_i, h_i)\}_{i=1}^N$ is generated from *The Berkeley Segmentation Dataset* (**BSD300**), which contains 300 natural images from various scenes. In order to obtain a clean and size-manageable database, we implement a pruning processing to eliminate the homogeneous and redundant patches.

The number of exemplars k can be automatically determined by the proposed method in Sec.2.4. The parameters σ_{hr} and σ_{lr} in Sec.2.3 are the standard variance of the HR and LR exemplars respectively. In experiments, 5×5 patch with an overlap of 2 pixel between adjacent patches are used in the LR image. Therefore the patch size in the HR image is $5N \times 5N$ if the magnification factor is N.

Images	*Paddy*(Fig.2)	*Parthenon*(Fig.3)	*Girl*(Fig.5)
BC	2.323	5.347	4.479
NE[6]	2.326	5.395	5.234
Soft[8]	2.368	5.284	4.634
SP[17]	2.321	5.255	4.478
AS[9]	2.317	5.213	4.452
Ours	**2.312**	**5.154**	**4.344**

Table 1. The RMS errors of SR results by different SR methods W.R.T the ground truth images.

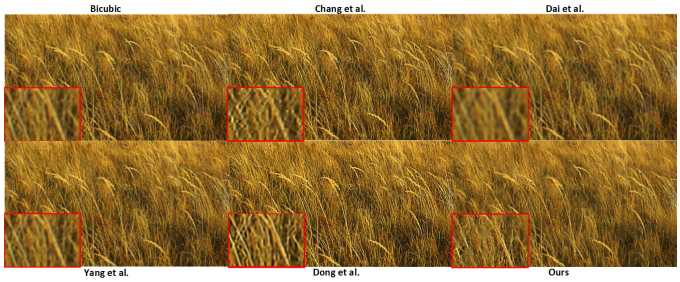

Fig. 2. SR results on a paddy field image with magnification factor 2.

3.2. The Effect of the Learned Metric

To verify the effectiveness of the learned metric on identifying the exemplars, we present an example in Fig.4.

3.3. Super-Resolution Results and Comparison

We compare our proposed method with several state-of-the-art SR methods. Dai's regularization-based SR algorithm with soft edge prior in [8], Chang's neighbor embedding method in [6], Yang's sparse coding algorithm in [17], and Dong's ASDS method in [9] are chosen as the baseline methods. In addition, bi-cubic interpolation result is also provided.

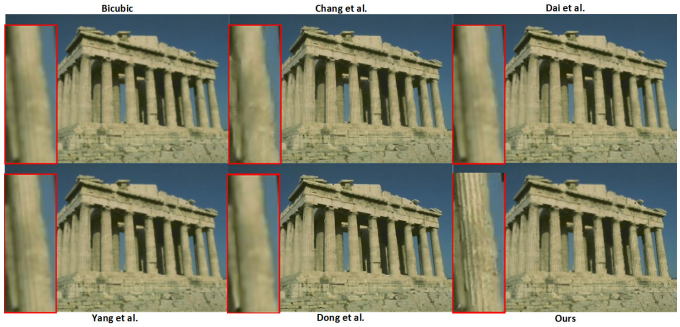

Fig. 3. SR results on the Parthenon image with magnification factor 2.

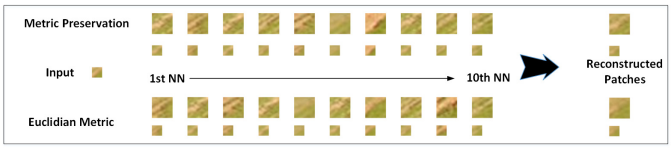

Fig. 4. Comparison example of finding exemplars by Euclidean metric and our learned Mahalanobis metric is shown.

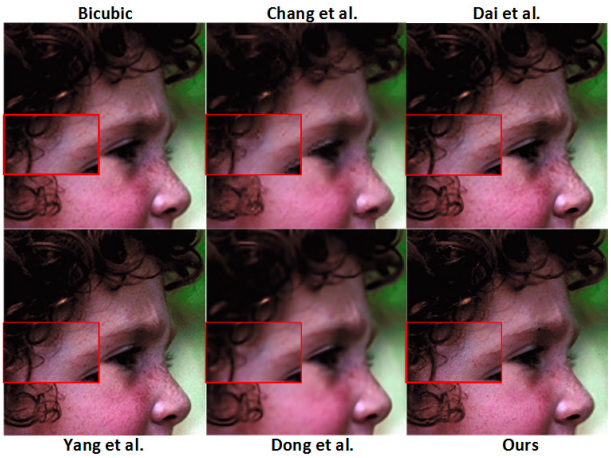

Fig. 5. SR results on a girl image with magnification factor 2.

Some experiment results on various types of natural images (Fig.2, Fig.3 and Fig.5) are presented. All the SR results are under magnification factor 2. In all the cases, our method presents sharper edges and clearer texture details. There are observable and significant improvements in the textures, such as the engraving on the pillar, the ears of rice and the freckles on the face. And the salient edges are well preserved and sharpened in our results. The quantitative comparison results in terms of the RMS errors are shown in Table 1.

4. CONCLUSIONS

We propose a novel exemplar-based image super-resolution method that is able to automatically identify the "right" exemplars. This method learns a best metric in the LR space, driven by the metric in the HR space. Such a metric brings the affinity structures of the LR and HR spaces as close as possible, such that the local reconstruction can be directly transferred from the LR space to the HR space for SR reconstruction. In addition, our method also automatically determines the best number of exemplars to use, rather than simply using a fixed number or a fixed threshold. By unifying both explicit and implicit priors, our SR approach is able to introduce new visual contents, and at the same time to generate smoother and sharper edges in region boundaries.

5. REFERENCES

[1] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(9):1167–1183, 2002.

[2] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 26–33. IEEE, 2003.

[3] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[4] D. Capel. *Image mosaicing and super-resolution*. Springer, 2004.

[5] F. Champagnat, C. Kulcsár, and G. Le Besnerais. Continuous super-resolution for recovery of 1-d image features: Algorithm and performance modeling. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 916–926. IEEE, 2006.

[6] H. Chang, D.-Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2004.

[7] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong. Soft edge smoothness prior for alpha channel super resolution. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[8] S. Dai, M. Han, W. Xu, Y. Wu, Y. Gong, and A. K. Katsaggelos. Softcuts: a soft edge smoothness prior for color image super-resolution. *Image Processing, IEEE Transactions on*, 18(5):969–981, 2009.

[9] W. Dong, D. Zhang, G. Shi, and X. Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *Image Processing, IEEE Transactions on*, 20(7):1838–1857, 2011.

[10] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *Image processing, IEEE Transactions on*, 13(10):1327–1344, 2004.

[11] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *Computer Graphics and Applications, IEEE*, 22(2):56–65, 2002.

[12] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 349–356. IEEE, 2009.

[13] Z. Lin and H.-Y. Shum. Fundamental limits of reconstruction-based superresolution algorithms under local translation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(1):83–97, 2004.

[14] A. Sánchez-Beato and G. Pajares. Noniterative interpolation-based super-resolution minimizing aliasing in the reconstructed image. *Image Processing, IEEE Transactions on*, 17(10):1817–1826, 2008.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.

[16] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in neural information processing systems*, pages 2080–2088, 2009.

[17] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *Image Processing, IEEE Transactions on*, 19(11):2861–2873, 2010.