Single-Image Super-Resolution Using Multihypothesis Prediction

Chen Chen and James E. Fowler
Department of Electrical and Computer Engineering, Geosystems Research Institute (GRI)
Mississippi State University, MS 39762 USA

***Abstract*— Single-image super-resolution driven by multihypothesis prediction is considered. The proposed strategy exploits self-similarities existing between image patches within a single image. Specifically, each patch of a low-resolution image is represented as a linear combination of spatially surrounding hypothesis patches. The coefficients of this representation are calculated using Tikhonov regularization and then used to generate a high-resolution image. Experimental results reveal that the proposed algorithm offers significantly higher-quality super-resolution than bicubic interpolation without the cost of training on an extensive training set of imagery as is typical of competing single-image techniques.**

I. Introduction

Image super-resolution (SR) has seen increasing interest within the image-processing community because it offers solutions to overcome resolution limitations of low-cost digital imaging systems and imperfect imaging environments. A popular paradigm is to synthesize a new high-resolution (HR) image by using one or more low-resolution (LR) images [1]. Existing SR algorithms in the literature can be classified as multi-image SR (e.g., [2, 3]) or example-based SR (e.g., [4–6]). In classical multi-image SR, an HR image is obtained from a set of LR images of the same scene at subpixel misalignments. However, this approach is numerically limited to only a small increase in resolution [7]. In example-based SR, the correspondences between HR and LR image patches are learned from known LR/HR image pairs in a database, and then the learned correspondences are applied to a new LR image for SR. The underlying assumption is that the missing HR details can be learned from the HR database patches. In this vein, Yang *et al*. [8] proposed a sparse coding to learn a dictionary on HR and LR images such that the LR and HR images share the same sparse representation. Finally, Glasner *et al*. [9] combined both the classical multi-image and example-based SR techniques and exploited patch redundancy within as well as across scales to reconstruct the unknown HR image.

In this paper, we propose an SR method that exploits self-similarities of image patches within a single image using the multihypothesis (MH) prediction strategy from [10]. Specifically, the MH paradigm is employed for single-image SR wherein each patch of an LR image is represented as a linear combination of spatially surrounding hypothesis patches. The coefficients of this representation are calculated using Tikhonov regularization [11] and then used to generate a high-resolution image.

The remainder of our discussion is organized as follows. In Sec. II, we overview Yang's algorithm [8] for image SR via sparse representation. In Sec. III, we present our method using MH prediction for image SR. In Sec. IV, we examine experimental results and give analysis in comparison with other algorithms. Finally, we make some concluding remarks in Sec. V.

II. Super-Resolution via Sparse Representation

In general, single-image SR aims to recover an HR image X from a given LR image Y of the same scene. Typically, the observed LR image Y is assumed to be a blurred and down-sampled version of the HR image X; i.e.,

$$Y = DLX, \tag{1}$$

where D and L are the down-sampling operator and a blurring filter, respectively.

Inspired by a recent flurry of activity in compressed sensing and sparse representation, Yang *et al*. [8] proposed SR based on an assumption that the LR and HR image patches share the same sparse representations. Specifically, given two dictionaries $\mathcal{D}_h$ and $\mathcal{D}_l$ for the HR and LR images, respectively, for each LR patch y from Y, the sparsest representation of y can be formulated as the ℓ_1 minimization problem,

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \left\| \mathcal{D}_l \alpha - y \right\|_2^2 + \lambda \left\| \alpha \right\|_1. \tag{2}$$

The HR patch $\hat{x}$ is then reconstructed using the same sparse representation vector $\hat{\alpha}$ as

$$\hat{x} = \mathcal{D}_h \hat{\alpha}. \tag{3}$$

In order to learn the dictionary pair $\{\mathcal{D}_h, \mathcal{D}_l\}$, a set of HR training patches $X^h = \{x_1, x_2, \ldots, x_n\}$ are sampled from the collected HR image database. $Y^l = \{y_1, y_2, \ldots, y_n\}$ are the corresponding LR image patches generated by (1). In [8], Yang *et al*. proposed to train a coupled dictionary so that the HR patches and the corresponding LR patches share the same sparse representation. The joint dictionary-training process is formulated as

$$\begin{aligned}\{\mathcal{D}_h, \mathcal{D}_l, Z\} = \underset{\mathcal{D}_h, \mathcal{D}_l, Z}{\operatorname{argmin}} & \frac{1}{N} \left\| X^h - \mathcal{D}_h Z \right\|_2^2 \\ & + \frac{1}{M} \left\| Y^l - \mathcal{D}_l Z \right\|_2^2 + \lambda \left(\frac{1}{N} + \frac{1}{M} \right) \left\| Z \right\|_1, \end{aligned} \tag{4}$$

where N and M are the dimensions of the HR and LR image patches in vector form, and $Z = \begin{bmatrix} z_1 & z_2 & \cdots \end{bmatrix}$ is the sparse coefficient matrix formed by placing the sparse representation vectors as columns of the matrix.

Yang's method performs well if, in fact, the input LR image patches are similar to the ones in the training set. However, one limitation of this method is that the sparse-representation problem of (2) must be solved for each patch in Y. Another limitation is that its performance relies heavily on the availability of suitable HR training images. To address these issues, we propose to exploit the self-similarities of image patches within a single image; details of our approach follow in the next section.

III. Super-Resolution Using Multihypothesis Prediction

Given a single LR image Y of size $N \times N$ and a scale factor s, we want to reconstruct the HR image X of size $sN \times sN$. In the first step, Y is magnified to $sN \times sN$ by bicubic interpolation. This interpolated, middle-resolution (MR) image is denoted as X^m. Y is then partitioned into non-overlapping patches of size $B \times B$. For each LR patch y, a corresponding $sB \times sB$ MR patch exists in X^m at the same spatial location as y in Y.

For each MR patch x^m in X^m, multiple hypothesis patches are generated from the spatially surrounding patches in a search window in X^m. We extract all the patches in the search window and place them as columns in the hypothesis matrix H of size $s^2B^2 \times K$, where K is the number of hypotheses; H thus contains all the predictions for patch x^m. This hypothesis generation is described in Fig. 1. Simultaneously with the generation of hypothesis matrix H, all the $sB \times sB$ hypothesis patches are blurred and down-sampled by a factor of s using (1) to form patches with the size of $B \times B$. All the down-sampled hypothesis patches are then collected as columns of hypothesis matrix H^l of size $B^2 \times K$ for the corresponding LR patch y.

To find a prediction that is as close to y as possible, we want to solve

$$\hat{w} = \underset{w}{\operatorname{argmin}} \left\| y - H^l w \right\|_2^2, \tag{5}$$

where w is a column vector that holds the weights for all the hypotheses in H^l. However, observing that (5) is an ill-posed least-squares problem (i.e., usually $B^2 \neq K$), we adopt the methodology of [10] and invoke Tikhonov regularization [11] which imposes an ℓ_2 penalty on the norm of w. Consequently, (5) is reformulated as

$$\hat{w} = \underset{w}{\operatorname{argmin}} \left\| y - H^l w \right\|_2^2 + \lambda_{\text{tik}} \left\| \Gamma w \right\|_2^2, \tag{6}$$

where Γ is the Tikhonov matrix, and λ_{tik} is the regularization parameter. As proposed in [12], we use a diagonal Γ in the form of $\Gamma_{jj} = \left\| y - h_j^l \right\|_2^2$, where h_j^l are the columns of H^l and $j = 1, \ldots, K$. The solution for (6) is calculated as

$$\hat{w} = \left(\left(H^l \right)^T H^l + \lambda_{\text{tik}}^2 \Gamma^T \Gamma \right)^{-1} \left(H^l \right)^T y. \tag{7}$$

With the weights calculated from (7), we form the HR patch $\hat{x}$ in estimated image $\hat{X}$ as

$$\hat{x} = H\hat{w}. \tag{8}$$

The reconstruction resulting from (8) yields an estimate, $\hat{X}$, of the HR image. To enforce a global reconstruction constraint, we further project this initial reconstructed HR image onto the solution space of $Y = DLX$ (similar to as done in [8]), computing

$$X^* = \underset{X}{\operatorname{argmin}} \left\| DLX - Y \right\|_2^2 + \lambda' \left\| X - \hat{X} \right\|_2^2. \tag{9}$$

We can then iteratively improve the reconstruction by using this X^* as the MR image X^m to repeat the hypothesis-generation and HR-image-reconstruction procedures. The entire SR process is summarized as Algorithm 1.

IV. Experimental Results

In our experiments, we magnify eight 128×128 LR grayscale images using both Yang's method and our proposed MH-based approach by a factor of $s = 2$ using the processes as described in the previous sections. To magnify the LR image by a factor of $s = 4$, we first magnify the LR image by a factor of $s = 2$, and then magnify the resulting image by again by a factor of $s = 2$ to achieve the final result. Peak signal-to-noise ratio (PSNR), root mean squared error (RMSE), and structural similarity (SSIM) [13] are used as quality measures.

For Yang's sparse-representation SR, LR patches of size 5×5 with an overlap of 4 pixels between adjacent patches are used in all experiments as suggested in [8]. The implementation[1] is from the authors and is used with their pre-trained 1024×1024 dictionary.

For SR using MH prediction, two iterations (*MaxIter* $= 2$) are used with LR patches fixed to size 4×4 in Y, and search-window size fixed to 2 for hypothesis generation for magnifying the image by a factor of $s = 2$. We note that more iterations and a larger search-window size could improve performance, but we found that two iterations and a search-window size half of the patch size gave satisfactory results. To achieve fast reconstruction for $s = 4$, we use only one iteration (*MaxIter* $= 1$) in each constituent $s = 2$ magnification process. For the Lagrange parameter in (6), we set $\lambda_{\text{tik}} = 0.01$ in all our experiments.

Quantitative results are presented in Table I, while visual comparisons of the reconstructed HR images using various algorithms are shown in Figs. 2 and 3 for upscale factors of $s = 2$ and $s = 4$, respectively. From the results, we can see that Yang's sparse-representation SR yields superior image quality as compared to our proposed method; this is due largely to the fact that a meaningful HR training set is used. However, our method, which exploits the self-similarities of image patches within a single-image, effectively avoids the complex dictionary-training procedure required by Yang's approach. Our proposed technique, on the other, outperforms

[1] http://www.ifp.illinois.edu/~jyang29/ScSR.htm

bicubic interpolation by about 1 dB PSNR on average for an upscale factor of 2. In addition, in terms of reconstruction time as shown in Table II, our proposed approach runs much faster than Yang's approach for $s = 4$.

V. Conclusion

In this paper, an algorithm for single-image SR based on MH prediction was proposed. The proposed strategy exploited self-similarities existing between image patches within a single image. The fact that no HR training set is required for SR based on this MH prediction makes it more practical than a competing SR based on sparse representation since there is no guarantee that a relevant HR training set is available for low-resolution input images in all situations.

References

[1] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, May 2003.

[2] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, October 2004.

[3] D. Capel and A. Zisserman, "Super-resolution from multiple views using learnt image models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, Kauai, HI, December 2001, pp. 627–634.

[4] S. Baker and T. Kanade, "Hallucinating faces," in *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, March 2000, pp. 83–88.

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

[6] W. T. Freeman and E. C. Pasztor, "Learning low-level vision," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2, Kerkyra, Greece, September 1999, pp. 1182–1189.

[7] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, September 2002.

[8] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, November 2010.

[9] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proceedings of the IEEE International Conference on Computer Vision*, Kyoto, Japan, September 2009, pp. 349–356.

[10] C. Chen, E. W. Tramel, and J. E. Fowler, "Compressed-sensing recovery of images and video using multihypothesis predictions," in *Proceedings of the 45th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2011, pp. 1193–1198.

[11] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington, D.C.: V. H. Winston & Sons, 1977.

[12] E. W. Tramel and J. E. Fowler, "Video compressed sensing with multihypothesis," in *Proceedings of the Data Compression Conference*, J. A. Storer and M. W. Marcellin, Eds., Snowbird, UT, March 2011, pp. 193–202.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

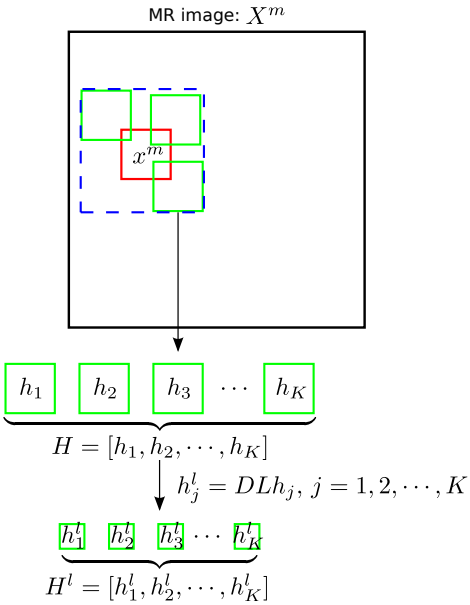

Fig. 1. Hypothesis generation within a search window.

Algorithm 1 SR Using MH Prediction

Input: Y (LR image), D (down-sampling operator), L (blurring filter), s (scale factor), *MaxIter* (maximum number of iterations).
Output: SR image X^*.
Initialization: $i = 1$, $X^m = \textbf{Bicubic}(Y)$ (initial MR image).
for $i = 1 \rightarrow$ *MaxIter* **do**
  for each patch $x^m \in X^m$ **do**
    (1) Generate hypothesis matrix H for x^m
    (2) Generate hypothesis matrix H^l for LR patch y via (1)
    (3) Solve the optimization problem defined in (6) for the weights vector w
    (4) Generate the HR image patch $\hat{x} = Hw$ in estimated HR image $\hat{X}$
  end for
  Using gradient descent, find the closest image to $\hat{X}$ which satisfies the constraint defined in (9):

$$X^* = \underset{X}{\operatorname{argmin}} \|DLX - Y\|_2^2 + \lambda' \left\|X - \hat{X}\right\|_2^2.$$

  Update $X^m \leftarrow X^*$
end for

TABLE I

PSNR (DB), RMSE, AND SSIM FOR $s = 2$ AND $s = 4$ SCALE FACTOR

	$s = 2$			$s = 4$		
Algorithm	PSNR	RMSE	SSIM	PSNR	RMSE	SSIM
Lenna						
Bicubic	31.22	7.00	0.9282	28.85	9.21	0.9167
Yang	**33.50**	**5.39**	**0.9536**	**30.28**	**7.81**	**0.942**
Proposed	32.83	5.82	0.9487	29.87	8.18	0.9391
Barbara						
Bicubic	29.57	8.47	0.8714	23.60	16.84	0.8140
Yang	**30.93**	**7.25**	**0.8975**	**23.96**	**16.15**	**0.8398**
Proposed	30.63	7.50	0.8920	23.88	16.31	0.8353
Goldhill						
Bicubic	30.91	7.26	0.8760	27.66	10.55	0.8527
Yang	**32.13**	**6.31**	**0.9083**	**28.27**	**9.84**	**0.8855**
Proposed	31.61	6.70	0.8999	28.08	10.06	0.8808
Mandrill						
Bicubic	25.75	13.16	0.7445	21.08	22.51	0.6897
Yang	**26.42**	**12.18**	**0.8070**	**21.34**	**21.85**	**0.7541**
Proposed	26.30	12.35	0.7988	21.31	21.93	0.7525
Peppers						
Bicubic	31.17	7.05	0.9486	27.93	10.23	0.9385
Yang	**33.64**	**5.30**	**0.9684**	**29.25**	**8.82**	**0.9578**
Proposed	32.69	5.92	0.9608	28.65	9.42	0.9513
Couple						
Bicubic	28.41	9.69	0.8470	25.38	13.72	0.8213
Yang	**29.71**	**8.34**	**0.8879**	**26.12**	**12.64**	**0.8633**
Proposed	29.29	8.75	0.8779	25.91	12.91	0.8568
Man						
Bicubic	29.85	8.20	0.8828	26.86	11.57	0.8635
Yang	**31.79**	**6.56**	**0.9217**	**28.01**	**10.20**	**0.9028**
Proposed	31.08	7.12	0.9112	27.59	10.64	0.8958
Boat						
Bicubic	28.38	9.72	0.8658	25.54	13.48	0.8443
Yang	**30.01**	**8.05**	**0.9070**	**26.49**	**12.14**	**0.8862**
Proposed	29.42	8.62	0.8965	26.18	12.51	0.8796
Average						
Bicubic	29.41	8.82	0.8705	25.86	13.51	0.8426
Yang	**31.02**	**7.42**	**0.9064**	**26.72**	**12.43**	**0.8790**
Proposed	30.48	7.85	0.8982	26.43	12.74	0.8739

(a) LR input

(b) HR output

Fig. 2. (a) The LR input image. (b) Results of the Lenna image magnified by a factor of $s = 2$. Top-row (left to right): the original HR image, bicubic interpolation (RMSE: 7.0); bottom-row (left to right): Yang's method [8] (RMSE: 5.39), our method (RMSE: 5.82).

TABLE II

SR RECONSTRUCTION TIME FOR THE 128×128 LENNA IMAGE ON A QUADCORE 2.67-GHZ MACHINE FOR UPSCALE FACTOR OF $s = 4$

Magnification	*Algorithm*	*Time (sec.)*
$s = 2$	Yang's (2 $\times$)	162.71
	Proposed	198.94
$s = 4$	Yang's (4 $\times$)	844.38
	Proposed (4 $\times$)	410.45

Fig. 3. Results of the Lenna image magnified by a factor of $s = 4$. Top-row (left to right): the original HR image, bicubic interpolation (RMSE: 9.21); bottom-row (left to right): Yang's method [8] (RMSE: 7.81), our method (RMSE: 8.18).