Epitomic Image Super-Resolution

Yingzhen Yang,[1] **Zhangyang Wang,**[1] **Zhaowen Wang,**[2]
Shiyu Chang,[1] **Ding Liu,**[1] **Honghui Shi,**[1] **Thomas S. Huang**[1]

[1]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801
[2]Adobe Research, San Jose, CA 95110, USA
{yyang58, zwang119, chang87, dingliu2, hshi10, huang}@ifp.uiuc.edu, zhawang@adobe.com

Abstract

We propose Epitomic Image Super-Resolution (ESR) to enhance the current internal SR methods that exploit the self-similarities in the input. Instead of local nearest neighbor patch matching used in most existing internal SR methods, ESR employs epitomic patch matching that features robustness to noise, and both local and non-local patch matching. Extensive objective and subjective evaluation demonstrate the effectiveness and advantage of ESR on various images.

Introduction

Image Super-Resolution (SR) methods construct a high-resolution image from one or multiple low-resolution input images (Park, Park, and Kang 2003). SR is indeed an ill-posed problem since the information from the low-resolution input is insufficient to recover the high-resolution image. Making different assumptions on the image priors, various SR algorithms have been proposed. Exemplar-based SR methods (Freedman and Fattal 2011; Yang et al. 2012; Yang, Lin, and Cohen 2013) exhibit promising results, which exploit the mapping from low-resolution patches to their high-resolution counterparts. Exemplar-based SR methods learn such mapping from high-resolution to low-resolution primarily in three ways: through a large set of external low-resolution and high-resolution patch pairs (Yang et al. 2012; Yang, Lin, and Cohen 2013), or only relying on the self-similarities within the given low-resolution input (Freedman and Fattal 2011), or in a combined way (Wang et al. 2015). We refer to the exemplar-based SR methods that utilize external data as external SR, and the methods that only exploit the self-similarities in the input as internal SR. External SR methods typically use a large and representative set of external high-resolution and low-resolution patch pairs to predict the missing high frequency information in the high-resolution image to be constructed, and the representative external SR method (Yang et al. 2012) learns a pair of high-resolution and low-resolution dictionaries by coupled dictionary learning, from which the missing high frequency information is obtained as a sparse linear combination of the dictionary atoms. Albeit being effective in

many case, external SR methods are prone to generate unreliable high-resolution patches when the input low-resolution patches cannot be represented by the external data, and they usually produce artifacts in this case. In contrast, Internal SR methods, such as the High Frequency Transfer (HFT) method (Freedman and Fattal 2011), seek for example patches from the input image itself based on the self-similarities by local patch matching, that is, similar patches often recur within the image. By exploiting self-similarities, many singular and unique image patches that rarely appear in external data can be reliably recovered by internal SR methods.

However, the local patch matching through nearest neighbor searching used in most existing internal SR methods suffers from noise and outliers. In addition, the local matching excludes potential better global matches that are not located in the neighborhood of the target patch. We propose epitomic patch matching to replace the nearest neighbor local patch matching in HFT, and the resultant SR method is named Epitomic Image Super-Resolution (ESR). ESR effectively reduces the artifacts caused by nearest neighbor patch matching, and it enables efficient local and non-local patch matching by epitomic patch matching. The effectiveness and advantage of ESR over other internal and external SR methods are demonstrated by our extensive experimental results.

The Proposed Epitomic Super-Resolution

We first introduce the High Frequency Transfer (HFT) method (Freedman and Fattal 2011), then illustrate the details of Epitomic Super-Resolution.

The High Frequency Transfer method

According to the observation that similar features in small patches often repeat across different image scales, the High Frequency Transfer method (Freedman and Fattal 2011) searches for the high-frequency component for a target high-resolution patch by nearest neighbor patch matching across scales. For the input LR image $\mathbf{Y}$, we first obtain its initial upsampled image $\mathbf{X}'$, and a smoothed input image $\mathbf{Y}'$. For each patch of the initial upsample image, denoted by $\mathbf{X}'_{ij}$ with coordinates ij, its match $\mathbf{Y}'_{st}$ on $\mathbf{Y}'$ is obtained by local nearest neighbor matching: $(s,t) = \arg\min_{(s,t) \in \mathbf{W}_{ij}} \|\mathbf{Y}'_{st} - \mathbf{X}'_{ij}\|_F^2$, where $\mathbf{W}_{ij}$ a local win-

dow on image $\mathbf{Y}'$. Denote by $\mathbf{X}$ the high resolution image to be constructed from $\mathbf{Y}$. The high frequency band of the patch $\mathbf{Y}_{st}$, namely $\mathbf{Y}_{st} - \mathbf{Y}'_{st}$, is regarded as the high frequency band of the unknown patch $\mathbf{X}_{ij}$. Therefore, such high frequence band is pasted onto $\mathbf{X}'_{ij}$ to get $\mathbf{X}_{ij} = \mathbf{X}'_{ij} + \mathbf{Y}_{st} - \mathbf{Y}'_{st}$.

Epitomic Super-Resolution

The matching of $\mathbf{X}'_{ij}$ over the image $\mathbf{Y}'$ is essential for the performance of HFT. To improve the robustness against to noise, and extend the local matching to both local and non-local matching for better matches, we propose epitomic patch matching to replace the local nearest neighbor matching in HFT. Being a generative model, the epitome of an image summarizes the raw image patches into a condensed representation of a size smaller than the original image, and it approaches this goal in a manner similar to Gaussian Mixture Models (GMMs). Please refer to (Jojic, Frey, and Kannan 2003) for more details of epitome. With the epitome $\mathbf{e}_{\mathbf{Y}'}$ learned from the smoothed input image $\mathbf{Y}'$, each patch $\mathbf{X}'_{ij}$ is matched over the epitome $\mathbf{e}_{\mathbf{Y}'}$ instead of a local window on $\mathbf{Y}'$. The location of the matching patch in the epitome $\mathbf{e}_{\mathbf{Y}'}$ for the patch $\mathbf{X}'_{ij}$ is specified by the most probable hidden mapping for $\mathbf{X}'_{ij}$: $\mathcal{T}^*_{ij} = \arg\max_{\mathcal{T}_{ij}} p\left(\mathcal{T}_{ij}|\mathbf{X}'_{ij}, \mathbf{e}\right)$. The hidden mapping $\mathcal{T}_{ij}$ specifies the location of the epitome patch from which the image patch $\mathbf{X}'_{ij}$ is generated, and it is similar to the role of hidden variables in GMMs. The top patches on $\mathbf{Y}'$ with large posterior probabilities $p\left(\mathcal{T}^*_{ij}|\cdot, \mathbf{e}\right)$ are regarded as the candidate matches for each patch $\mathbf{X}'_{ij}$, and the match $\mathbf{Y}'_{st,E}$ is the one in these K candidate patches with minimum Sum of Squared Distance (SSD) to $\mathbf{X}'_{ij}$. In this way, two matches, $\mathbf{Y}'_{st,E}$ and $\mathbf{Y}'_{st}$ are obtained for each patch $\mathbf{X}'_{ij}$ using epitomic patch matching and local nearest neighbor matching, and we can have their corresponding high frequency band $F_{st,E}$ and F_{st}. A weighted average of the two high frequency bands is treated as the final high frequency component for $\mathbf{X}'_{ij}$, i.e. $wF_{st,E} + (1-w)F_{st}$, where the weight $w = p(\mathcal{T}^*_{ij}|\mathbf{X}'_{ij}, \mathbf{e})$ denotes the probability of the most probable hidden mapping given the patch $\mathbf{X}'_{ij}$. Matching over epitome is robust to noise since each epitome patch summarizes a set of raw patches in the original image, and it is non-local matching since epitome $\mathbf{e}_{\mathbf{Y}'}$ summarizes the entire image $\mathbf{Y}'$.

Experimental Results

We compare our Epitomic Super-Resolution (ESR) to other competing methods in this section, and conduct both objective and subjective evaluation. For objective evaluation, we compare ESR to two internal SR methods, i.e. Bicubic interpolation and HFT on the Kid, Temple and Train image, and use the Peak Signal-to-Noise Ratio (PSNR) as a objective measure for the quality of the SR results. The results of objective evaluation is shown in Figure 2 with PSNR value for the three internal SR methods. We observe that ESR always achieves the highest PSNR value, revealing the advantage of robustness to noise and both local and non-
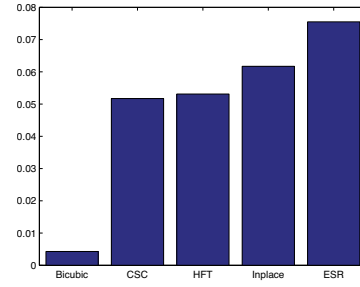

Figure 1: Subjective quality scores for different SR methods. CSC and Inplace are two external SR methods in the reference.
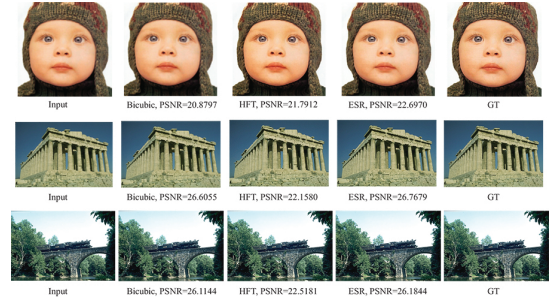

Figure 2: SR results on the Kid (upscaled by 4 times), Temple and Train images (upscaled by 3 times). From left to right: the low-resolution input, SR by Bicubic interpolation, SR by HFT, SR by ESR, and the Ground Truth (GT).

local patch matching by epitomic patch matching.

We conduct subjective evaluation on the results by different SR methods and illustrate the subjective quality scores in Figure 1. Users are required to perform a series of comparisons, and in each comparison two SR results are shown and users select the better one.

References

Freedman, G., and Fattal, R. 2011. Image and video upscaling from local self-examples. *ACM Trans. Graph.* 30(2):12:1–12:11.

Jojic, N.; Frey, B. J.; and Kannan, A. 2003. Epitomic analysis of appearance and shape. In *ICCV*, 34–43.

Park, S. C.; Park, M. K.; and Kang, M. G. 2003. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine, IEEE* 20(3):21–36.

Wang, Z.; Yang, Y.; Wang, Z.; Chang, S.; Yang, J.; and Huang, T. 2015. Learning super-resolution jointly from external and internal examples. *Image Processing, IEEE Transactions on* 24(11):4359–4371.

Yang, J.; Wang, Z.; Lin, Z.; Cohen, S.; and Huang, T. 2012. Coupled dictionary training for image super-resolution. *Image Processing, IEEE Transactions on* 21(8):3467–3478.

Yang, J.; Lin, Z.; and Cohen, S. 2013. Fast image super-resolution based on in-place example regression. In *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013*, 1059–1066.