Bayesian region selection for adaptive dictionary-based Super-Resolution

Eduardo Pérez-Pellitero[1, 2]
eduardo.perezpellitero@technicolor.com
Jordi Salvador[1]
jordi.salvador@technicolor.com
Javier Ruiz-Hidalgo[3]
j.ruiz@upc.edu
Bodo Rosenhahn[2]
rosenhahn@tnt.uni-hannover.de

[1] Image Processing Lab
Technicolor R&I
Hannover, Germany

[2] Institute for Information Processing
Leibniz Universität Hannover
Hannover, Germany

[3] Image Processing Group
Universitat Politècnica de Catalunya
Barcelona, Spain

The performance of learning-based Super-Resolution (SR) methods depends strongly on the content of the training dataset. In [3] the dictionary is built by randomly sampling raw patches from a large set of images regardless of the image to be recovered, hence relying on gathering sufficiently diverse patches so that they can generalize for any patch to be super-resolved. More recent follow-up works keep using the same strategy for the training, although these raw patches are compressed in a smaller number of patches through sparse coding techniques [4].

This paper proposes a novel sparse SR method, which focuses on adaptively selecting optimal patches for the dictionary training. Observing the fact that images usually contain a non-predictable group of different elements with their characteristic textures and edges (e.g. grass, rocks, fur, sand), we divide our training dataset into sub-image entities which we call regions, and extract descriptors in order to characterize them. The key idea is that, being these regions smaller, they have more consistent texture or edge content. For every patch to be super-resolved we find its best-fitting texture region from the training images by using the efficient local *Naive Bayes Nearest Neighbor* (NBNN), thus ensuring the obtained example pairs are highly correlated with the input low-resolution (LR) patches.

Each training image I_T is split in square regions R of size L_R. Given a patch y we find its training texture region R. Assuming a uniform region prior over R this can be achieved through a maximum likelihood decision rule:

$$\hat{R} = \operatorname*{argmax}_R p(R \mid y) = \operatorname*{argmax}_R p(y \mid R). \tag{1}$$

Let $\{f\} = f_1, f_2, \ldots, f_n$ denote the descriptors extracted from patch y. We use the Naive Bayes assumption, i.e. descriptors are independent, identically distributed:

$$\hat{R} = \operatorname*{argmax}_R \sum_{i=1}^{n} \log p(f_i \mid R). \tag{2}$$

In order to handle the high number of regions which a training set can contain, a local NBNN approach as the one in [2] is used.

Let R be some region and $\overline{R}$ the set of all other regions. The decision rule will be:

$$\hat{R} = \operatorname*{argmax}_R \sum_{i=1}^{n} \log \frac{P(f_i \mid R)}{P(f_i \mid \overline{R})} + \log \frac{P(R)}{P(\overline{R})} \tag{3}$$

After finding a region R for every patch y, we will sample patches of size L_p with a certain overlap inside the selected regions and include them in HR and LR training sets T_l and T_h, which will train sparse adapted LR and high-resolution (HR) joint dictionaries D_l and D_h respectively.

In the paper, our method is compared with two well-known SR approaches: The original LLE-based method [1] and the sparse coding SR method of [4]. In our implementation, dictionaries are adaptively trained with the proposed method and used in [4]. The direct comparison is therefore explicitly showing the improvement by using our algorithm.

In order to prove the effectivity of our adaptive training scheme, two different testing scenarios and a single training set have been used. All the images have been upscaled by 2x and 3x magnification factors, computing the luminance peak-to-noise-ratio (PSNR) and the Structural Self Similarity (SSIM) as objective quality assesments.

Comparing our method to [1], the visual improvements are easily noticeable: our method is clearly super-resolving finer details resulting in sharper images. The objective PSNR and SSIM results support this qualitative visual impression.

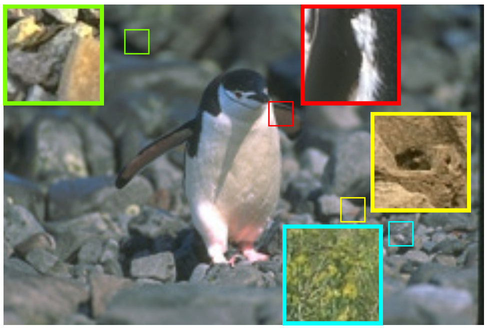

Figure 1: An example of how our method selects training regions based on Bayes rule. For every patch in the input image, a highly correlated region from the training images is found. The resulting set of regions is used to build a texture-adapted dictionary which will be used to recover the HR image.

With respect to the baseline method [4], careful visual inspection shows that ringing artifacts along the edges are mostly suppressed and certain edges are sharper. This is also suported in both PSNR and SSIM for the two magnification factors tested and in both testing set-ups. We observe that our method yields large improvement in performance when training and testing images are related (0.504 dB PSNR gain for 2x magnification factor), but there is also significant improvement for generic testing images (0.338 dB PSNR improvement for 2x magnification factor).

Our method improves performance for generic images, but also this improvement is accentuated when there exist training regions related to the testing images within the training dataset, therefore making our method specially adequate for applications where these conditions are met (e.g. video sequences, multiview scenarios, metadata-tagged images).

[1] H. Chang, D. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 275–282, 2004.

[2] S. McCann and D.G. Lowe. Local naive bayes nearest neighbor for image classification. *Computer Vision and Pattern Recognition, IEEE Conference on*, pages 3650–3656, 2012. ISSN 1063-6919.

[3] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. *IEEE Conf. on Computer Vision and Pattern Recognition, CVPR.*, pages 1–8, 2008. ISSN 1063-6919.

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