A Appendix

This appendix presents additional experimental results on super-resolution (A.1) and demonstrates an application of the proposed method to removal of artifacts in JPEG encoded images (A.2).

A.1 Additional Results

To gain a better understanding of the effects of individual system components on the final results, a set of experiments has been performed with several variations of the proposed method (henceforth referred to as ‘the original system’ or ‘O’), which were constructed by replacing sub-systems of the original system with related alternatives. The variations considered here include (S1) replacing the part for constructing the sparse basis set by KMP, (S2) combining SVR and NIP (i.e., to post-process the results of our earlier SVR-based method based on the NIP), (S3) optimizing the post-processing cost functional (3) throughout the whole image region (instead of only the pixels at the vicinity of major edges), (S4) replacing the deviation penalty term in (3) by the reconstruction constraint, which leads to a MAP framework (cf. Sec. 2.2), (S5) adding the reconstruction constraint to the cost functional (3) (i.e., using the reconstruction constraint, deviation penalty, and NIP for the post-processing), and (S6) optimizing the parameters of the proposed system in a larger search space.

As demonstrated in Fig. 2, the combination of KMP and gradient descent can provide a significantly better optimization of basis points over the other alternatives considered in the current paper including KMP and gradient descent. Comparison of Fig. A.1(S1) with (O) (the results of S1 and the original system, respectively) shows that this can actually lead to improved reconstructions of images: detailed visual inspection along textured area reveals that the proposed method can generate sharper images than KMP (plus NIP). Furthermore, as observed in the visor of Lena, the original system resulted in more coherent enhancements of edges. On the other hand, since the training time of this system is much shorter than that of the original system (on the order of minutes; the testing time remains the same), it could be regarded as an alternative for on-line applications.

As shown clearly in the first and the third image of Fig. A.1(S2), the performance of SVR-based super-resolution has also been improved by adopting the NIP-based post-processing. However, still the results of the original system turned out to be much sharper and more coherent.

Fig. A.1(S3) shows the results of S3, which are sharper and cleaner than the results of the other methods (cf. the first and the third images). However, overall it leaded to degraded results as it flattened out texture details. This is clearly visible in the the crown of the hat in the Lena image and in waves of the fourth image.

---

1The evaluation of the whole training set in each KMP step is computationally very demanding as it requires $O(M \times l^2)$-time. Accordingly, for each KMP step, we pre-selected 1,000 candidates based on the cost functional $C$ (cf. Sec. 2.1). The single best basis point is then chosen based on the KMP among only these candidates.

2Since the original SVR-based method does not generate multiple candidates for each pixel, we used the candidates generated by the proposed method to facilitate the optimization during the post-processing.

3Numerically indexed equations and figures (i.e., without the ‘A’ symbol) are those appeared in the main paper.

4The NIP term in (3) utilizes the $1 \times 2$ derivative kernel $[1,-1]$ (along each of 8 directions) whose only null space is the set of constant images. Accordingly, it tends to flatten the image. On the other hand, this phenomenon is rarely observed in the original NIP framework of [Tappen et al., 2003] as they use the $1 \times 3$ derivative kernel $[1,0,-1]$ whose null space is larger than ours. However, it should be noted that in this case, the NIP enforces only the constancy of every other pixels. Furthermore, the down-sampling operator (the three-tap Gaussian kernel) used in the reconstruction constraint [Tappen et al., 2003] has the effect of smoothing out high-frequency oscillations of signals. Accordingly, both of these constraints do not prevent the generation of
In S4, we minimize a new post-processing cost functional:

\[
U(\{x\} | \{z\}) = \sum_{(j,i \in N_B(j))} \left( \frac{|x_j - x_i|}{\sigma_N} \right) + \sum_j \left( \frac{B(x_{(j,N_B(j))}) - z_{T_B(j)}}{\sigma_Q} \right)^2, \tag{A.1}
\]

where \( \{z\} \) are the pixel values of input low-resolution image, \( B \) is the down-sampling operator (blurring and interpolation, implemented based on a local convolution kernel), \( T_B(j) \) gives the locations of all low-resolution pixels which are influenced through \( B \) by the high-resolution pixel at \( j \), and \( N_B(j) \) gives the indices of all high-resolution variables which jointly with \( x_j \) influence on the values of the low-resolution pixels \( z_{T_B(j)} \). The parameters \( \sigma_N \), \( \sigma_Q \), and \( \alpha \) are optimized based on a set of validation images.

Due to the high connectivity of latent variables in the reconstruction constraint, the optimization of (A.1) based on BP is infeasible.\(^5\) We therefore adopted the stochastic relaxation [Geman and Geman, 1984] instead.\(^6\) As shown in Fig. A.1(S4), this system generated visually distracting artifacts. This can be explained by the fact that once the images containing these artifacts are reduced (based on \( B \)) down to the scale of the low-resolution images, the artifacts disappear and accordingly are not penalized by the reconstruction constraint. Actually, the reconstruction costs (the second term in (A.1)) of the images obtained by applying \( B \) to the images of Figure A.1(S4) were much smaller than those of the results of applying \( B \) to Fig. A.1(O), even though the images in the latter have higher PSNR values (PSNRs of different systems are plotted in Figs. A.2 and A.3). This problem can be resolved by introducing the deviation penalty term of (3) to (A.1) as the second regularizer (S5). An advantage of this system over the original system is that that the MAP framework provides it with a better mathematical basis. Indeed, the system demonstrated an improved performance on the original system in terms of PSNR (on average 0.06dB improvement). However, as shown in Fig. A.1(f), the results of S5 are not significantly visually more plausible than those of the original system. Furthermore, the running time of the new system (which also uses the stochastic relaxation based optimization) is much longer than that of the original system (on the order of hours per image). A preference between the two systems may depend on specific application requirements.

For the construction of the original system, we have traded the complexity off with the quality of super-resolution by pre-fixing some parameters (i.e., \( l_b \), \( N \), and \( L \) are fixed at 300, 25(5 × 5), and 49(7 × 7), respectively). Extensive optimization of these parameters is impractical especially because of \( l_b \): one has to construct a kernel matrix of size \( l_b \times l_b \) and invert a matrix of \( l_b \times l_b \) in training, and to perform \( l_b \) kernel evaluations for each pattern in testing. However, an insight into how the potential optimal system might behave could be gained by optimizing the parameters in a rather large space (S6). Figures A.1(S6) and A.2 show the results of the system where \( l_b \) is set to 1,000 while the search space of \( N \) and \( L \) are not constrained. The optimal values of \( N \) and \( L \) were identified as 9 and 9, respectively. Comparison of the PSNR values of the results of S6 with those of the original system indicates that optimizing the parameters in a larger search space does result in an improved system. However, the visual quality of the results of two systems turned out to be not so significantly different from each other. Furthermore, the testing time of the new system is more than 10 times longer than that of the original system on average.

\(^5\)Sharp alternations of pixel values. Although not frequent (since the configuration space is constrained: candidates are \( 2 \times 2 \)-size image patches rather than individual pixels (cf. Sec. 2.2)), this can sometimes result in noisy patterns (cf. the second image of Fig. 5(d)).

\(^6\)It should be noted that (A.1) corresponds to an energy functional in a Markov network where each node (variable) \( j \) has \( |N_B(j)| \) neighbors, which is lower-bounded by the size of the kernel \( m \). Optimizing (A.1) based on BP then, requires calculating the maximum among (at least) \( N^{m-1} \) messages where \( N \) is the number of candidates.

The outputs of the regression step are utilized as the initial configurations. The final results are then obtained as MAP estimates generated by the serial Gibbs sampler which bases on the distribution

\[
P(\{x\} | \{z\}) = \frac{\exp(-U(\{x\} | \{z\})/T)}{C(T)}, \tag{A.2}
\]

where \( C(T) \) is a normalization function and the annealing schedule \( T \) is defined as a decreasing function of iteration number \( k \) (cf. [Geman and Geman, 1984]):

\[
T(k) = \frac{Q}{\log(1+k)}, \quad 1 \leq k \leq K. \tag{A.3}
\]

Here, \( K \) is the total number of iterations which was determined such that all variables are visited (in raster order) 1,000 times while \( Q \) was empirically determined as 0.1.
Figure A.1: Additional super-resolution examples: (I) interpolation (magnification factor 3), (S1) KMP and NIP, (S2) SVR and NIP, and (S3) NIP applied to the entire image.

A.2 Application to JPEG artifact removal

One definite advantage of the proposed method and in general of learning-based approaches is that, in principle, the generic learning part can be applied to any problem when suitable examples of input and target output images are available. As an example, we present an application of the proposed method to artifact removal of JPEG-encoded images.

The basic idea of JPEG-encoding is to divide the image into disjoint blocks and then individually transform (based on the discrete cosine transform (DCT)), quantize, and encode each block. This approach well exploits effectiveness of the DCT for the compression of small image blocks and its efficiency in hardware im-
plementations. However, at low bit rates, decoded images exhibit block artifacts (discontinuities that appear between the boundaries of the blocks). The underlying idea of the proposed artifact removal method is to cast the problem into the super-resolution: once blurred, the boundaries between DCT blocks can be reduced in JPEG images [Lee et al., 2005, Reeve III and Lim, 1984] and the problem becomes estimating missing high-frequency details. This can be resolved by replacing the interpolation part with blurring in the proposed method (cf. [Kim and Kwon, 2008] for more details).

The application scenario of the proposed method is then to train a model specialized to each small interval of compression factors such that the whole range of compression factors is covered by several models. However, to facilitate the demonstration here, we focus on a specific quantization table (which determines the compression ratio) for JPEG encoding (Q2 in Table 2 of [Nosratinia, 2001]) such that both in training and in testing, the JPEG images are obtained with this quantization table. This table has been used in many published works (e.g., [Lee et al., 2005, Zakhor, 1992, Sun and Cham, 2007]) and accordingly for the standard Lena image, the performance of the proposed method can be compared with these methods. Figure A.4 shows examples of artifact removal.

### A.3 Notes and Acknowledgments

- For the gradient descent in the optimization of basis points, we used the code provided by Carl E. Rasmussen, which is available at [http://www.kyb.tuebingen.mpg.de/bs/people/carl/code/minimize/](http://www.kyb.tuebingen.mpg.de/bs/people/carl/code/minimize/).
- A Matlab demo of the proposed super-resolution algorithm is available at [http://www.kyb.tuebingen.mpg.de/bs/people/kimki/supres/supres.htm](http://www.kyb.tuebingen.mpg.de/bs/people/kimki/supres/supres.htm).
- The code used in generating the results of [Alter et al., 2005] is courtesy of Jacques Froment.
- The ideas of some systems presented in A.1 are originated from comments of the reviewers.

### References


Figure A.1 (continued): (S4) reconstruction constraint and NIP, (S5) reconstruction constraint, deviation penalty, and NIP, (S6) proposed method (optimized in a large parameter space), (O) original system, and (G) original high-resolution images.
Figure A.2: Performance of different super-resolutions algorithms: (S1) KMP and NIP, (S2) SVR and NIP, (S3) NIP applied to whole image region, (S4) reconstruction constraint and NIP, (S5) reconstruction constraint, deviation penalty, and NIP, and (S6) proposed method optimized in a large parameter space.

Figure A.3: Performance of different example-based superresolution algorithms.
Figure A.4: Examples of JPEG artifact removal. (a) and (g) input JPEG images, (b) and (h) blurred JPEG images, (c) and (i) re-application of JPEG [Nosratinia, 2001], (d) and (j) adapted total variation-based method [Alter et al., 2005], (e) and (k) proposed method, and (f) and (l) original images. Increases of PSNRs from the input JPEG images (displayed below each image) were calculated based on the complete images. The results of the proposed method are much sharper than the results of the re-application of JPEG and less blocky and as sharp as the results of the adapted TV.