

Image Quality Enhancement for Single-Image Super Resolution Based on Local Similarities and Support Vector Regression

Atsushi YAGUCHI^{†a)}, Nonmember, Tadaaki HOSAKA^{††}, and Takayuki HAMAMOTO[†], Members

SUMMARY In reconstruction-based super resolution, a high-resolution image is estimated using multiple low-resolution images with sub-pixel misalignments. Therefore, when only one low-resolution image is available, it is generally difficult to obtain a favorable image. This letter proposes a method for overcoming this difficulty for single-image super resolution. In our method, after interpolating pixel values at sub-pixel locations on a patch-by-patch basis by support vector regression, in which learning samples are collected within the given image based on local similarities, we solve the regularized reconstruction problem with a sufficient number of constraints. Evaluation experiments were performed for artificial and natural images, and the obtained high-resolution images indicate the high-frequency components favorably along with improved PSNRs.

key words: *super resolution, local similarity, support vector regression (SVR), total variation (TV)*

1. Introduction

Recently, super resolution, a type of magnification technology that generates a clear high-resolution (HR) image by restoring high frequency components, has attracted considerable attention. A representative framework is reconstruction-based super resolution [1], in which a HR image is obtained as a solution of the minimization problem for the cost function representing errors between the low resolution (LR) images predicted by the observation model and the input LR images. The traditional iterative back-projection (IBP) method [2] and maximum *a posteriori* estimation [1] have often been used for this purpose. Since these approaches require multiple LR images having sub-pixel misalignments in order to increase constraints with respect to a HR image, applying the reconstruction-based framework to a single image is generally difficult. Another framework is example-based super resolution [3], [4], which estimates a HR image (in most cases, on a patch-by-patch basis) based on the relationships between HR and LR images learned from a prepared image database. Although this approach can be applied to a single image by the help of an external database, selecting the (sub)optimal images for the database is generally difficult.

For single-image super resolution, Glansner et al. [5] focused on the tendency that similar local patches appear

repeatedly in an image, and utilized only a given LR image and its scale-converted images as the database. Furthermore, they utilized the IBP method to increase the consistency between predicted and original LR images. Their framework is considered to be significant for single image super resolution.

This letter proposes a method to enhance the quality of single-image super resolution by applying the regularized reconstruction-based super resolution. To compensate for the shortage of constraints attributed to the lack of input LR images, we interpolate pixel values at sub-pixel locations on a patch-by-patch basis using support vector regression (SVR), in which learning samples are collected within the given image based on local similarities pointed out in [5]. Then, we solve the optimization problem defined by the given LR image and the interpolated pixels with a regularization constraint. We also investigate the effects of two types of regularization constraints: the L_2 -norm of the Laplacian filtered image and the total-variation (TV) norm.

2. Proposed Method

2.1 Reconstruction-Based Super Resolution

We use a simple observation model which generates a LR image from the HR image, and this is defined as

$$l_i = \mathbf{B}_i^T \cdot \mathbf{H}, \quad (i = 1, 2, \dots, M) \quad (1)$$

where l_i represents the pixel value of the LR image at pixel i ($= 1, 2, \dots, M$; M is the number of pixels in the LR image), and \mathbf{H} and \mathbf{B}_i are vector forms of the HR image and the blur kernel for pixel i , respectively. The HR image is, in principle, reconstructed such that the differences between the LR image predicted by this model and the given LR image are minimized. Although this can be achieved by the IBP method [2], only minimizing this prediction error unfortunately tends to yield excess high-frequency components in the resulting HR image because this is an essentially ill-posed problem. Therefore, as in super resolution using multiple LR images, the cost function with a regularization constraint is introduced as

$$F(\mathbf{H}) = \sum_{i=1}^M [\mathbf{B}_i^T \cdot \mathbf{H} - l_i]^2 + \alpha C(\mathbf{H}), \quad (2)$$

where the first term measures errors between the original and predicted LR images; $C(\mathbf{H})$ in the second term usually

Manuscript received June 12, 2010.

Manuscript revised October 5, 2010.

[†]The authors are with the Graduate School of Engineering, Tokyo University of Science, Tokyo, 102-0073 Japan.

^{††}The author is with the Faculty of Engineering Division I, Tokyo University of Science, Tokyo, 102-0073 Japan.

a) E-mail: atsushi@isl.ee.kagu.tus.ac.jp

DOI: 10.1587/transfun.E94.A.552

represents the regularization constraint reflected by the prior knowledge, which suppresses the excess high-frequency components included in the HR image; and α is the parameter balancing these two terms.

When the above regularized reconstruction-based method is simply applied to a single image, the number of constraints in the first term of Eq. (2) is not sufficient for precise reconstruction of a HR image. For instance, even in the case of a magnification factor of 2, the number of constraints is only a quarter of the HR pixels, which tends to yield a blurred HR image, as with interpolation methods. Thus, we increase the constraints in the error term by interpolating pixel values at sub-pixel locations using the SVR based on local similarities.

2.2 Pixel Interpolation Using the SVR

From the viewpoint of example-based super resolution, we aim at utilizing the relation between a given LR image and its down-sampled image to estimate the HR image from the given LR image. The pixel interpolation process (Fig. 1) is described as follows:

* Generate the down-sampled (factor of 2, in our study) image from the original LR image. To retain sufficient number of learning samples, we generate four kinds of down-sampled images by simply throwing away three quarters of pixels. Moreover, we also use the rotated (angles of 90°, 180°, 270°) and axisymmetric version of these images (the total number of available images is 32).

1. Select a small patch of 3×3 pixels in the original LR image (represented as source patch in Fig. 1).
2. Find a similar patch from the down-sampled images in terms of a similarity calculated by the Gaussian-weighted sum of squared difference.
3. Focus on the corresponding enlarged patch of 5×5 pixels in the original LR image.
4. Collect N similar patches in the same way; learn the regression functions explaining the value of each triangular-marked (Δ) pixel $\mathbf{y} = (y_1, y_2, \dots, y_8)$ from the nine pixel values of circular-marked (\circ) pixels $\mathbf{x} = (x_1, x_2, \dots, x_9)$ (Although this is a multiple output regression problem, we simplify this to a single output regression by dealing with each output $y_i (i = 1, 2, \dots, 8)$ independently, as in the literature of super resolution for a video sequence [6]). When N is less than the threshold $S (=10, \text{ in this study})$, return to step 1 based on the idea that a small number of learning samples do not provide reliable regression functions. Writing the N training samples as $\Omega_i = \{(\mathbf{x}^1, y_i^1), (\mathbf{x}^2, y_i^2), \dots, (\mathbf{x}^N, y_i^N)\} (i = 1, 2, \dots, 8)$, the coefficients of the regression function $y_i = \mathbf{a}_i^T \boldsymbol{\varphi}(\mathbf{x}) + b_i$ ($\boldsymbol{\varphi}(\mathbf{x})$ is the kernel function mapping the vector \mathbf{x} to a higher-dimensional space) are estimated by the SVR [7].
5. Predict values of triangular-marked pixels in the HR image (corresponding to the pixel values at sub-pixel

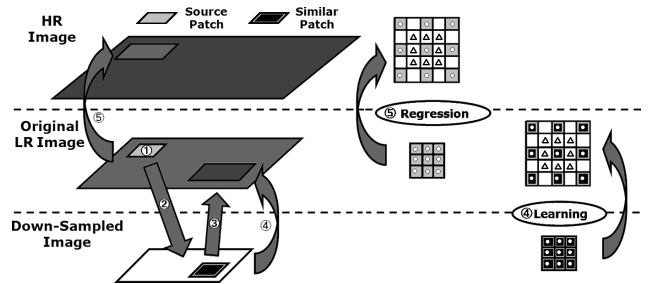


Fig. 1 Process of pixel interpolation: the numerals in the figure correspond to process numbers in Sect. 2.2.

locations in the LR image) using the obtained regression functions.

* Repeat steps 1–5 for all patches in the LR image.

Note that this interpolation process cannot provide a complete image. Since a source patch is selected on a pixel to pixel basis as mentioned above, interpolated pixels in the HR space can be overlapped. On the other hand, some locations are not interpolated due to the lack of similar patches for learning the regression function. A complete image is obtained as a solution of the optimization problem in which the desired image should be consistent with the original and interpolated pixels.

2.3 Regularization Constraints and Optimization

In the optimization, the local diversity with respect to regression accuracy and the number of interpolated pixels should be appropriately accommodated. The regularization constraint introduced in Eq. (2) can also serve this purpose. Therefore, we define the cost function as

$$F(\mathbf{H}) = \sum_{i=1}^M [\mathbf{B}_i^T \cdot \mathbf{H} - l_i]^2 + w \sum_{i=1}^m [\mathbf{B}_i^T \cdot \mathbf{H} - \hat{l}_i]^2 + \alpha C(\mathbf{H}), \quad (3)$$

where \hat{l}_i denotes interpolated values, m is the number of interpolated pixels, and w represents a weight to control the balance between the original and interpolated pixels. Moreover, we investigate the effects of two types of regularization constraints $C(\mathbf{H})$: the L_2 -norm of the Laplacian filtered image and the TV norm [8]. The constraint by the TV norm is expressed as

$$C_{TV}(\mathbf{H}) = \sum_i \sqrt{|g_x(i)|^2 + |g_y(i)|^2} + \beta, \quad (4)$$

where $g_x(i)$ and $g_y(i)$ represent gradients at pixel i in the horizontal and vertical directions, respectively. The positive parameter β is added to ensure differentiability.

For the optimization of L_2 regularization, since the cost function becomes wholly quadratic, the convergence of minimization for Eq. (3) is relatively fast with the conjugate-gradient procedure. However, it is likely that ringing artifacts will appear around the edge and step edges will be over

smoothed. In contrast, the TV regularization can generally maintain the step edge and suppress the ringing artifacts. However, solving the optimization problem is generally difficult due to the nonlinearity of differential of $C_{TV}(\mathbf{H})$. In this study, we solve it by using the naive gradient descent procedure, which requires many iterations and precise adjustment of the update rate.

3. Experimental Results

As shown in Fig. 2, evaluation experiments were performed for an artificial image ((a), 256×256) and natural images ((b), (c), 128×128). Our proposed methods (represented as Proposed (L₂) and Proposed (TV)) are compared with the bicubic interpolation, the version with no pixel interpolation by the SVR (NPI (L₂) and NPI (TV)), and the IBP method with pixel interpolation (IBP+PI). In this study, the magnification factor is 2. The parameter w was set as 0.005 for all test images, which was empirically determined so that the best performance can be achieved in most cases. The number of interpolated pixels m was (a) 240496, (b) 29888, (c) 44248, including overlapped pixels.

Figure 3 shows trimmed HR images for Fig. 2(a) and the PSNRs are summarized in Table 1. The efficacy of the regularization can be confirmed by the fact that our proposed methods indicate higher PSNR than IBP+PI which does not impose the regularization constraint. This improvement is attributed to the suppression of excess high-frequency components by the regularization constraint. Especially, Proposed (TV) can suppress the development of ringing artifacts which appear in IBP+PI and Proposed (L₂), and Proposed (TV) exhibits the best performance for these three test images.

The effectiveness of pixel interpolation by the SVR is also investigated. In the result of NPI (L₂), the high-frequency components around the edges are not restored satisfactorily, and ringing artifacts appear greatly because of the lack of constraints. The result of Proposed (L₂) indicates sharp edges, and its PSNR is improved compared with NPI (L₂). Although we observe only a little differences by appearance between NPI (TV) and Proposed (TV), the improvement of PSNR is confirmed.

An example of computational time is as follows: The processing time of our method (Proposed (TV)) is 21.86 s for the image of Fig. 2(b) on a personal computer with a Core i7-960 Processor 3.20 GHz and 12 GB RAM. The improvement of this execution speed can be achieved by implementation on a GPU architecture, which is one of our future works.

4. Conclusions and Future Work

We enhanced the quality of single-image super resolution by regularized reconstruction-based methods in which pixel values at sub-pixel locations are interpolated based on local

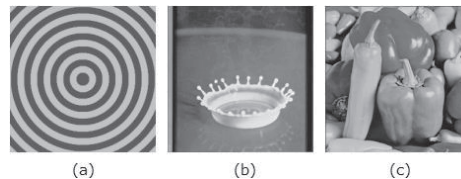


Fig. 2 Test images.

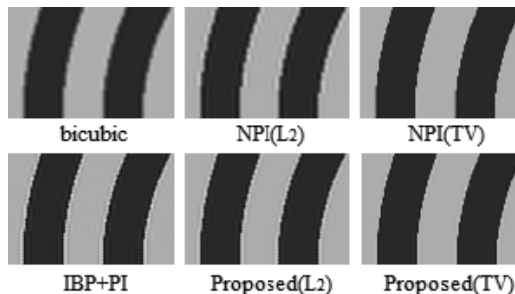


Fig. 3 Experimental results.

Table 1 Quantitative evaluation based on PSNR (dB).

	bicubic	NPI (L ₂)	NPI (TV)	IBP+PI	Prop.(L ₂)	Prop.(TV)
(a)	26.62	27.15	30.82	29.33	29.53	31.42
(b)	34.25	35.64	36.43	36.78	37.55	37.94
(c)	31.28	32.43	33.17	33.12	33.67	33.84

similarities and SVR. The experimental results confirm the validity of the proposed method. The value of w was currently fixed for all interpolated pixels. Investigation for adaptive determination of it depending on the regression accuracy is under way.

References

- [1] S.C. Park, M.K. Park, and M.G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol.20, no.3, pp.21–36, 2003.
- [2] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol.53, no.3, pp.231–239, 1991.
- [3] W.T. Freeman, T.R. Jones, and E.C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol.22, no.2, pp.56–65, 2002.
- [4] K.I. Kim and Y. Kwon, "Example-based learning for single image super-resolution," *Proc. DAGM*, pp.456–465, 2008.
- [5] M. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," *Proc. ICCV*, 2009.
- [6] K.S. Ni, S. Kumar, N. Vasconcelos, and T.Q. Nguyen, "Single image superresolution based on support vector regression," *International Conference on Acoustics, Speech, and Signal Processing*, 2006.
- [7] V. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, 1998.
- [8] M.K. Ng, H. Shen, E.Y. Lam, and L. Zhang, "A total variation regularization based super-resolution reconstruction algorithm for digital video," *EURASIP J. Advances in Signal Processing*, no.74585, pp.1–16, 2007.