15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Single Image Super Resolution Based on Content-Aware Constraint and Intensity-Order Constraint

Takashi Shibata and Atsushi Sato
NEC Corporation
1753 Shimonumabe,Nakahara-Ku,Kawasaki,Kanagawa,Japan.
t-shibata@hw.jp.nec.com, asato@ay.jp.nec.com

Abstract

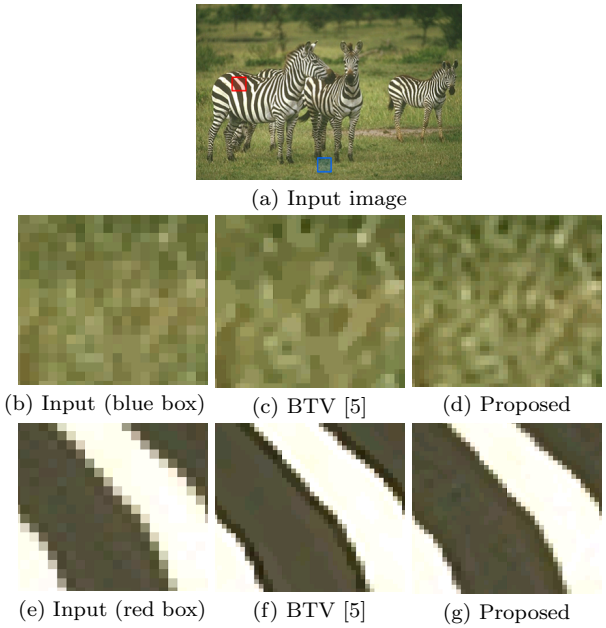
This paper presents a novel single image super resolution (SR) based on a content-aware constraint and an intensity-order constraint. The proposed method generates an SR image by minimizing an energy that consists of a data term, the content-aware constraint and the intensity-order constraint. The content-aware constraint can preserve texture patterns while reducing noise in a smooth region, while the intensity-order constraint can reduce ringing artifacts by penalizing the inconsistency of the intensity-magnitude relation. Experimental results show that the proposed method outperforms existing reconstruction-based SR methods in terms of PSNR and SSIM.

1 Introduction

Single image super resolution (SISR) is a technique to generate a high-resolution (HR) image from an input low-resolution (LR) image [15]. This technology is used for various applications ranging from video and image editing to document image up-sampling and surveillance. SISR can be considered an ill-posed problem because there is a resolution gap, i.e., an information gap, between the LR image and the HR image. Various constraints have been proposed to interpolate this gap. In general, these constraints for the SISR can be roughly classified into two types: 1) learning-based and 2) reconstruction-based.

In the learning-based approach [18, 1, 8, 3, 4, 20, 19, 16], the constraint is designed based on training images such as natural images [8], facial images [1], and character images [18, 4]. Baker et al. [1] introduced an exemplar-based constraint using the parent-structure vector. Datsenko et al. [4] further extended Baker's method by removing outlier exemplars that might lead to artifacts in the SR image. Recently, CNN-based approaches [3] and more sophisticated exemplar-based methods [20, 19] have also been proposed. In general, the learning-based approach is very sensitive against the residual between the actual degradation process and the assumed image formation model. Therefore, an accurate image formation model from the HR image to the LR image is required. However, designing such a model remains a challenge.

In the reconstruction-based approach [11, 5, 14, 15], the reconstruction-based constraints are usually designed based on generic image priors. For example, the Tikhonov-based constraint [15] is modeled using the L2 norm of the SR image intensity gradient, and



(a) Input image
(b) Input (blue box) (c) BTV [5] (d) Proposed
(e) Input (red box) (f) BTV [5] (g) Proposed

Figure 1. Results by proposed and existing methods. As shown in (c) and (f), over-smoothing and ringing artifacts are generated by BTV [5]. The proposed method can reduce these artifacts as shown in (d) and (g).

Total Variation (TV) [14] and Bilateral Total Variation (BTV) [5] are designed based on the L1 norm of the intensity gradient. Although these reconstruction-based constraints are widely used in SISR, they tend to generate two types of artifacts: 1) over-smoothing and 2) ringing. Examples of these artifacts are given in Fig. 1, where (c) shows over-smoothing and (b) shows ringing (both generated by BTV [5]).

There are two reasons these artifacts appear in reconstruction-based constraints. First, the reconstruction-based constraints [15, 5, 14] are designed so that the constraint strength is equal over the whole image regardless of their contents (edge, texture, and smooth region). Therefore, the constraints tend to generate over-smoothing artifacts in the texture region while reducing the noise in the smooth region.

Second , the constraints are evaluated only based on the intensity gradient of the SR image. In other words, there is no constraint against the intensity-magnitude relation among nearest neighbor pixels in the SR image. However, the breaking of the magnitude relation in the SR image directly generates the ringing artifacts around the edge region. In this sense, the ringing artifacts cannot be reduced by the existing reconstruction-based constraints.

To address the issue with the over-smoothing arti-

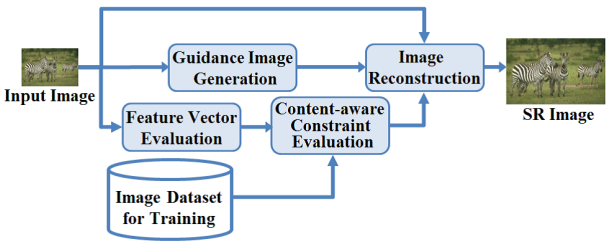


Figure 2. The overview of the proposed method.

fact, Cho et al. [2] recently presented an edge-and-texture-preserving constraint based on the constant-aware image prior. The constraint can preserve the texture pattern without over-smoothing artifacts by adaptively determining the constraint strength based on contents (edge, texture, and smooth region). However, the constraint tends to generate ringing artifacts because the constraint is basically designed based on the intensity gradient, i.e., without the intensity-magnitude relation.

In this paper, we propose a single image SR based on a content-aware constraint and an intensity-order constraint. The proposed method generates an SR image by minimizing an energy that consists of a data term, a content-aware constraint, and an intensity-order constraint. Experimental results show that the proposed method outperforms existing reconstruction-based SR methods in terms of PSNR and SSIM.

2 Proposed method

The overview of the proposed method is shown in Fig. 2. First, it generates the guidance image that is used for the intensity-order constraint by means of a guided filter [10]. Next, it computes eigenvalues of the Harris matrix for each pixel as the feature vector for the content-aware constraint. The content-aware constraint is evaluated based on a feature vector computed from the input image and a feature vector from the training images. Finally, the SR image is generated by minimizing the energy function.

In the proposed method, the energy function $E(\mathbf{x})$ consists of the data term, the content-aware constraint and the intensity-order constraint as

$$E(\mathbf{x}) = E_D(\mathbf{x}) + \lambda_1 E_C(\mathbf{x}) + \lambda_2 E_I(\mathbf{x}), \tag{1}$$

where $\mathbf{x}$ is the vectorized intensity of the SR image and λ_1 and λ_2 are the parameters to control the strength of the second and third terms, respectively. The first, second, and third terms are the data term, the content-aware constraint, and the intensity-order constraint, respectively.

The data term $E_D(\mathbf{x})$ penalizes the intensity residual between the input image and the down-sampled and blurred SR image as

$$E_D(\mathbf{x}) = \sum_i \Big[\sum_{j,k} D_{ij} B_{jk} x_k - y_i \Big], \tag{2}$$

where i, j, and k are the pixel index, D_{ij} and B_{jk} represent the down sampling and blur, and x_i and y_i are the intensity of the SR image and the input image at the i-th pixel.

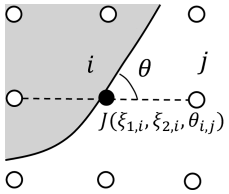


Figure 3. The perpendicular direction θ_i.

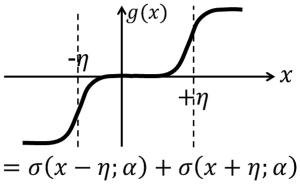


Figure 4. The outline of the smoothed step function g.

Content-aware constraint: To reduce the over-smoothing artifacts in the texture region while reducing the noise in the smooth region, we introduce the content-aware constraint $E_C(\mathbf{x})$ as

$$E_C(\mathbf{x}) = \sum_{i,j} J(\xi_{1,i}, \xi_{2,i}, \theta_{i,j}) |x_i - x_j|^{-\beta}, \tag{3}$$

where $J(\xi_{1,i}, \xi_{2,i}, \theta_{i,j})$ is the coefficient that controls the constraint strength for the intensity difference between the i-th and j-th pixels, and β is the exponent index. Here, $\xi_{1,i}$ and $\xi_{2,i}$ are the eigenvalues of the Harris matrix evaluated as feature vectors at i-th pixel, and $\theta_{i,j}$ is the angle between the perpendicular direction of the intensity gradient and the line connecting i-th and j-th pixels, as shown in Fig. 3. In the proposed method, the proposed image prior for the content-aware constraints is modeled by hyper-Laplacian distribution with the exponent β as

$$p(\delta x | \xi_1, \xi_2, \theta) \propto \exp\Big[- J(\xi_1, \xi_2, \theta) |\delta x|^{-\beta} \Big], \tag{4}$$

where δx is the intensity difference between nearest neighbor pixels, e.g., $|x_i - x_j|$. To evaluate the coefficient $J(\xi_1, \xi_2, \theta)$ using the training images, we first compute the eigenvalues of the Harris matrix for each pixel in each training image. Then, the coefficient $J(\xi_1, \xi_2, \theta)$ is estimated by minimizing the mean squared error of logarithmic probability $\ln p(\delta x | \xi_1, \xi_2, \theta)$. Note that the existing method [2] designs the image prior only using the variance and the fourth order moment of the absolute intensity gradient, i.e., without using the intensity gradient direction. In contrast, the proposed method evaluates the image prior based on not only the gradient intensity expressed by ξ_1 and ξ_2 but also the intensity gradient direction represented by the perpendicular direction θ. By using the perpendicular direction θ, we can preserve the sharpness of the strong edge along the normal direction while reducing the noise on the strong edge along the perpendicular direction.

Intensity-order constraint: To reduce the ringing artifact, the proposed method introduces the intensity-order constraint so that the intensity-magnitude relation between nearest neighbor pixels is preserved. The intensity-order constraint $E_I(\mathbf{x})$ penalizes the

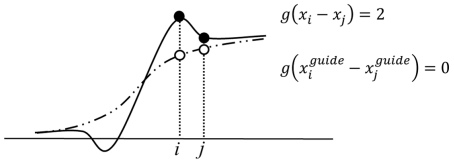


Figure 5. The effectiveness of the intensity-order constraint.

intensity-magnitude relation difference between the nearest neighbor pixels as

$$E_I(\mathbf{x}) = \sum_{i,j} \phi_i \left[g(x_i - x_j) - g(x_i^{guide} - x_j^{guide}) \right]^2, \quad (5)$$

where x_i^{guide} is the guidance image intensity at the i-th pixel, $g(\cdot)$ is the smoothed step function whose outline is shown in Fig.4, and ϕ_i is the weight of the intensity-order constraint at the i-th pixel. In the proposed method, the weight ϕ_i is a binary mask that only 1 around the strong edge region. The function $g(\cdot)$ represents the intensity-magnitude relation using sigmoid function with gain α as

$$g(x) = \sigma(x - \eta; \alpha) + \sigma(x + \eta; \alpha), \quad (6)$$

where η controls the width of the step size.

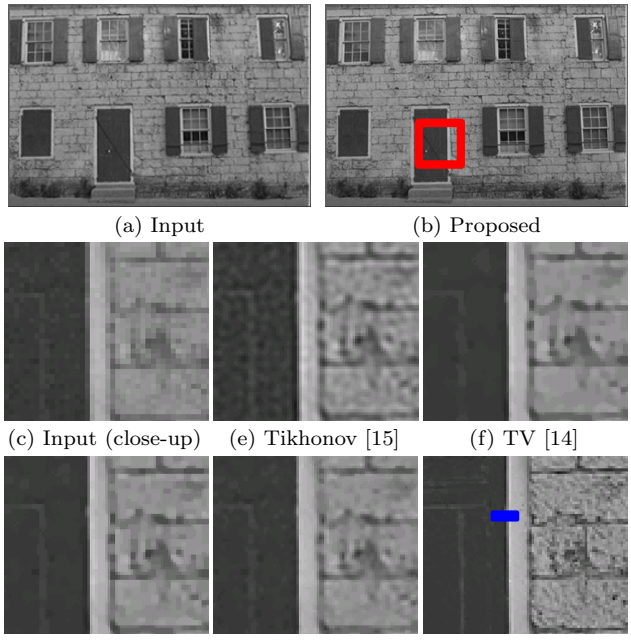
The effectiveness of the intensity-order constraint is shown in Fig. 5, where the SR image intensity and the guidance image intensity are shown as solid and broken lines, respectively. Note that there is a ringing artifact between the i-th and j-th pixels in the SR image. As shown, the output of the smoothed step function $g(\cdot)$ for the SR image (i.e., $g(x_i - x_j)$) is 2, while that for the guidance image (i.e., $g(x_i^{guide} - x_j^{guide})$) is 0. Therefore, in this case, the proposed intensity-order constraint $E_I(\mathbf{x})$ penalizes the SR image that includes the ringing artifacts.

3 Experiments

In order to evaluate the performance of the proposed method, we performed experiments on natural images [12, 13]. Images from the BSD dataset [13] were used as the training images to evaluate the content-aware constraint. All experiments were performed on a PC with a 3.33 GHz CPU and 92.0 GB RAM. We set the parameters as $\lambda_1 = 3$, $\lambda_2 = 0.1$, $\beta = 2/3$, $\eta = 8$, and $\alpha = 0.5$. These parameters were empirically determined using training images that were not used for the following evaluation.

We first evaluated the performance of the proposed method and existing reconstruction-based SR methods including Tikhonov-based constraint [15], TV [14], and BTV [5]. Parameters for the existing methods were empirically determined using the training images. We used 24 images from the Kodak image dataset [12] as the test images. In the experiments, the test images were generated by blurring the ground truth image with Gaussian kernel (PSF width: 1.0 [pix]) and by down-sampling the blurred image with Gaussian white noise (40 [dB]). The magnification rate was set to two.

Examples of the results obtained by the proposed and existing methods are given in Fig. 6. As shown in



(a) Input (b) Proposed

(c) Input (close-up) (e) Tikhonov [15] (f) TV [14]

(g) BTV [5] (h) Proposed (i) ground truth

Figure 6. Result by proposed and existing method. (×2)

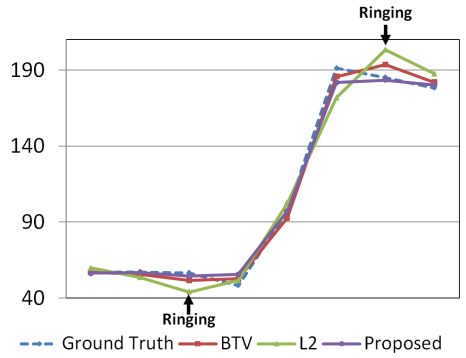


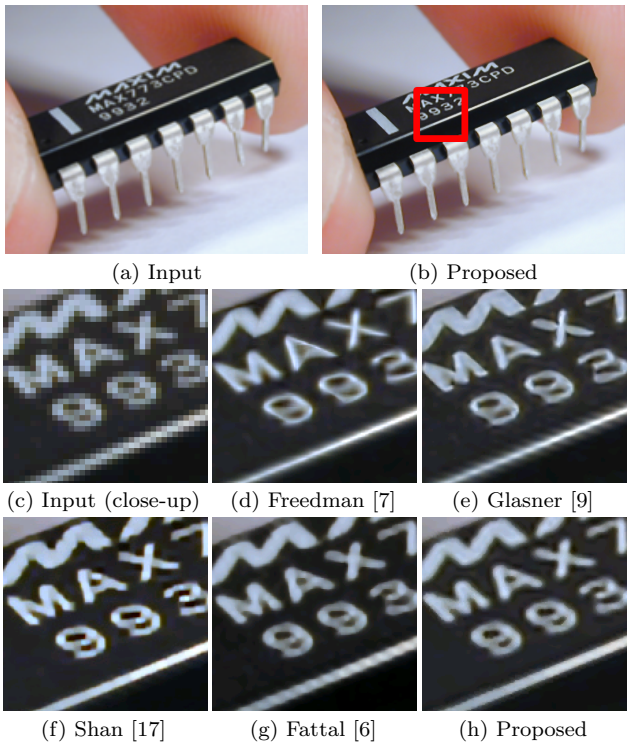
Figure 7. Edge profile of the proposed and the existing methods. (blue line in Fig. 6 (i))

Table 1. Mean of PSNR and SSIM.

	Tikhonov	TV	BTV	Proposed
PSNR	28.4	29.2	29.3	**29.6**
SSIM	0.913	0.953	0.954	**0.955**

Fig. 6 (e), the Tikhonov-based constraint generated severe ringing artifacts around the strong edge between the door and the wall. Over-smoothing artifacts were also generated by the existing methods [15, 14, 5] in the texture region of the stone wall, as shown in Fig. 6 (e), (f), and (g). In contrast, Fig. 6 (h) shows that the proposed method can generate the SR image without the ringing while preserving the texture. To evaluate the effectiveness of the proposed constraints more specifically, the edge profile around the reconstructed strong edge (the blue line in Fig. 6 (i)) by the proposed and existing methods are shown in Fig. 7. We found that the existing methods [15, 5] generated ringing artifacts around the edge, whereas the proposed method, by using the intensity-order constraint, could effectively reduce the ringing artifacts.

Next, to evaluate the performance of the proposed and existing methods quantitatively, we measured peak signal-to-noise ratio (PSNR) and structural sim-



(a) Input (b) Proposed

(c) Input (close-up) (d) Freedman [7] (e) Glasner [9]

(f) Shan [17] (g) Fattal [6] (h) Proposed

Figure 8. Result by proposed and existing method. (×4)

ilarity (SSIM) [21] in the SR images. The average PSNR and SSIM of the 24 test images (shown in Table 1) reveal that the proposed method outperformed the existing methods in terms of both PSNR and SSIM.

Finally, we evaluated other SISR methods, with examples shown in Fig. 8. Here, the magnification rate is four. As shown in Fig. 8 (e) and (g), the jaggy artifacts are generated in the edge region by Fattal's and Glasner's methods [9, 6]. The characters are also blurred by Shan's and Freedman's methods [7, 17], as shown in Fig. 8 (d) and (f). In contrast, Fig. 8 (h) shows that the proposed method can generate the SR image without the jaggy or blur artifacts.

4 Conclusion

We proposed a single image SR method based on the content-aware constraint and the intensity-order constraint. The proposed method generates an SR image by minimizing the energy that consists of the data term, the content-aware constraint, and the intensity-order constraint. Experimental results showed that the proposed method outperformed existing reconstruction-based SR methods with standard constraints in terms of PSNR and SSIM.

References

[1] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Tran. on Pattern Analysis and Machine Intelligence (TPAMI)*, 2002.

[2] T. S. Cho, N. Joshi, C. L. Zitnick, S. B. Kang, R. Szeliski, and W. T. Freeman. A content-aware image prior. *In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[3] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. *In Proc. of European Conf. on Computer Vision (ECCV)*, 2014.

[4] M. Elad and D. Datsenko. Example-based regularization deployed to super-resolution reconstruction of a single image. *The Computer Journal*, 2009.

[5] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Trans. on Image Processing (TIP)*, 2004.

[6] R. Fattal. Image upsampling via imposed edge statistics. *ACM Transactions on Graphics (TOG)*, 2007.

[7] G. Freedman and R. Fattal. Image and video upscaling from local self-examples. *ACM Transactions on Graphics (TOG)*, 2011.

[8] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 2002.

[9] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. *In Proc. of IEEE Int. Conf. on Computer Vision(ICCV)*, 2009.

[10] K. He, J. Sun, and X. Tang. Guided image filtering. *In Proc. of European Conf. on Computer Vision (ECCV)*, 2010.

[11] M. Ishii, T. Shibata, and A. Sato. Artifact-free image reconstruction for satellite imagery. *In Proc. of IEEE Conf. on Geoscience and Remote Sensing Symposium (IGARSS)*, 2016.

[12] X. Li. Demosaicing by successive approximation. *IEEE Transactions on Image Processing (TIP)*, 14, 2005.

[13] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *In Proc. of IEEE Int. Conf. on Computer Vision(ICCV)*, 2001.

[14] M. K. Ng, H. Shen, E. Y. Lam, and L. Zhang. A total variation regularization based super-resolution reconstruction algorithm for digital video. *EURASIP Journal on Advances in Signal Processing*, 2007.

[15] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 2003.

[16] S. Senda, T. Shibata, and A. Iketani. Example-based super resolution to achieve fine magnification of low-resolution images. *NEC Technical Journal*, 7(2):81, 2012.

[17] Q. Shan, Z. Li, J. Jia, and C.-K. Tang. Fast image/video upsampling. *ACM Transactions on Graphics (TOG)*, 2008.

[18] T. Shibata, A. Iketani, and S. Senda. Single image super resolution reconstruction in perturbed exemplar sub-space. *In Proc. of Asian Conf. on Computer Vision (ACCV)*, 2012.

[19] R. Timofte, V. De Smet, and L. Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. *In Proc. of Asian Conf. on Computer Vision (ACCV)*, 2014.

[20] R. Timofte, R. Rothe, and L. Van Gool. Seven ways to improve example-based single image super resolution. *In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[21] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, 2003.