SINGLE IMAGE SUPER-RESOLUTION USING ADAPTIVE DOMAIN TRANSFORMATION

Abhishek Singh and Narendra Ahuja

Department of Electrical & Computer Engineering
University of Illinois at Urbana-Champaign
abhishek_singh@ieee.org, n-ahuja@illinois.edu

ABSTRACT

In this paper we propose a new image domain prior term for regularizing the super-resolution reconstruction algorithm. This term encourages preserving the local *ramp* structure around edges, in the reconstruction algorithm. *Ramp* at a pixel is defined as the steepest sequence of monotonically increasing (or decreasing) pixels among all feasible directions around the pixel. As described in previous work, ramp based modeling is a richer characterization of local image structure than conventional gradients. Our proposed ramp-preserving constraint image is obtained by first running an accurate segmentation algorithm (which is itself obtained by ramp based modeling) on the low resolution image. We then perform a domain transformation of the pixels belonging to the steepest ramps at the edge pixels, in order to preserve sharpness. The resulting non-uniformly spaced image is then upscaled to a uniform, high resolution grid, using an edge preserving non-uniform interpolation scheme. This image is then used *both* as the prior constraint as well as the initial guess for the iterative super-resolution reconstruction algorithm. Our results compare favorably to the classical back-projection algorithm as well as newer methods which use learning based gradient domain priors.

Index Terms— Super-resolution, image prior, segmentation, regularization.

1. INTRODUCTION

Super-resolving a single image is a highly ill-posed problem. Most modern approaches can be broadly categorized into two classes. One class of approaches tries to 'hallucinate' the missing information in the high resolution image using examples of pairs of low resolution (LR) and high resolution (HR) image patches, possibly using a training database [1, 2, 3], or using patches within the image [4, 5].

The second class of methods uses the fundamental reconstruction constraint of minimizing the discrepancy between the given LR image and a synthesized LR image, formed by downsampling the estimated HR image with a presumed point spread function [6, 7, 8, 9, 3]. In this paper, we are concerned with this second class of methods.

Due to the ill-posedness of the single image upscaling problem, minimizing the reconstruction error by itself is not a sufficient criterion for obtaining visually satisfactory results. A prior term is usually used for regularizing the solution. The objective function is therefore of the form,

$$J_{SR}(\hat{I}_{HR}) = \left| \left(\hat{I}_{HR} * f_{psf} \right) \downarrow - I_{LR} \right|^2 + \lambda J_{prior}(.), \quad (1)$$

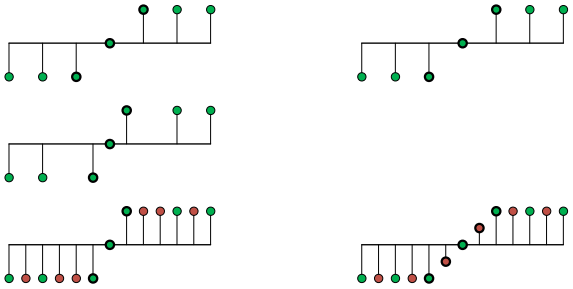
which is usually optimized using gradient descent iterations, starting with an initial guess [10, 7, 6].

The estimated HR image is highly dependent on the choice of the prior function, and also on the initial guess of the HR image used for the gradient descent iterations.

Gradient domain priors have been popularly used as edge preserving constraints [6, 7, 8, 9, 11], with the motivation of preserving the sharpness of image structures across resolutions. They encourage the gradient field of the estimated HR image to be as close as possible to a precomputed gradient field, which is usually an appropriately transformed and upscaled version of the LR gradient field.

However, the use of gradients as is conventionally done has some fundamental drawbacks. *Firstly*, gradient based linear convolutional filters implicitly assume a template for the expected edge geometry. Commonly used gradient operators operate only along the horizontal and vertical directions, and are therefore prone to errors around sharp, angular structures. *Secondly*, conventional gradient filters use a fixed size kernel. This leads to issues of geometric scale, as it imposes limitations on the size of geometric structures that the filter can detect. For instance, 3×3 filter detects structures of very different geometric scales in a 50×50 image as compared to a 200×200 image. Therefore, relating such gradient responses across resolutions (as in done by gradient domain priors in SR algorithms) may be prone to errors.

In previous work done by our group, the *ramp transform* was proposed to overcome the above limitations of traditional gradient filters [12]. The ramp transform for each pixel quantifies the intensity value change caused by the steepest sequence of monotonically increasing or decreasing intensity profiles (called ramps) among all possible directions, passing through the pixel. Since the steepest ramp is searched for in *all* feasible directions around the pixel, there is no assumption made on the orientation of the image structure. In addition, the ramps are allowed to be of variable length, which lends the ability to detect structures of arbitrary width (geometric scale). This is in contrast with traditional gradient operators that only use fixed length filters along only horizontal and vertical directions. Using the analysis performed by the ramp transform, [12] also proposes a rather accurate low level image segmentation algorithm, that is able to accurately detect photometric edges in the image, without imposing any prior assumptions or models about region geometry (shape, size), region photometry (contrasts and levels of interior homogeneity of the regions), and region topology (how many regions neighbor a given region) - expectations that many segmentation algorithms fail to meet [13, 14]. As expected from these properties, the segmentation results have been shown to better localize region bound-



(a) (b)

Fig. 1. (a) *Top*: A synthetic 1D signal, with the ramp pixels shown in bold. *Middle*: Result of performing the adaptive domain transformation step, to 'compress' the ramp. *Bottom*: Result of the subsequent non-uniform edge preserving interpolation to yield a high resolution signal. The sharpness (width, in pixels) of the ramp is maintained. (b) Conventional uniform upsampling of the *top* signal to yield the *bottom* signal. The ramp width (in pixels) increases.

aries that are blurred and/or include sharp corners [12]. Ramp based structure characterization and the segmentation algorithm based on it have been shown to be beneficial for several other image analysis tasks [15, 16]

Due to the advantages the ramp based structure modeling offers over traditional gradient operators, in this paper we propose a simple SR objective function that tries to appropriately preserve the ramp based structural information in the estimated HR image. For this, we propose an image domain prior of the form,

$$J_{prior}(\hat{I}_{HR}) = \left| \hat{I}_{HR} - I_p \right|^2 \tag{2}$$

where I_p is a high resolution prior image or constraint that preserves the ramp based structural information. In order to estimate this prior image, we first perform a ramp based segmentation of the LR image using the algorithm in [12], to yield precise locations of edges. We then perform a domain transformation of the pixels belonging to the steepest ramp at each edge pixel, in order to adaptively 'compress' the ramps. The resulting non-uniformly spaced image is then upscaled to a uniform, high resolution grid, using an edge preserving non-uniform interpolation scheme. Owing to the 'compression' of the ramps in the first step, and then the subsequent upscaling, the ramp based structure tends to be preserved across the resolution change. The resulting image serves as our constraint or prior image I_p and also as the initial guess for the iterative super-resolution reconstruction algorithm. A qualitative summary of our procedure to obtain the prior image is shown in Fig. 1 for a simple 1D signal.

2. ESTIMATING THE PRIOR IMAGE

2.1. Notation

We use scripted capital letters to donate sets of appropriately defined elements. We assume that the given LR image I_{LR} is defined over a uniformly spaced domain $\Omega_{LR} \in \mathcal{Z}^2$, where $\mathcal{Z}$ is the set of integers. We define the HR domain to be $\Omega_{HR} \in \mathcal{Z}^2$. We denote vectors in the image domain using boldface lowercase letters. Lowercase italicized letters are used as indices. L denotes the upscaling factor.

2.2. Low Level Segmentation and Ramp Transform

We begin by first performing an accurate low level segmentation of in the input LR image, using the algorithm described in [12]. We refer the reader to [12] and the references [14, 13] therein for technical details of the algorithm. We summarize the outputs of the algorithm that we use in the subsequent steps of our super-resolution algorithm.

Given the LR image as input, the segmentation algorithm of [12] yields the following:

1) A set of pixel locations $\mathcal{E} = \{\mathbf{e}_i\}_{i=1}^{N_e} \in \Omega_{LR}$ in the LR domain that contain the edge locations of the LR image. N_e is the total number of edge pixels.

2) For each edge pixel $\mathbf{e}$, a set of ramp pixels $\mathcal{R}_{\mathbf{e}} = \{\mathbf{r}_j(\mathbf{e})\}_{j=1}^{N_r(\mathbf{e})}$, that constitute the steepest ramp passing through $\mathbf{e}$. $N_r(\mathbf{e})$ denotes the width (in pixels) of this steepest ramp through the edge pixel $\mathbf{e}$.

3) The (normalized) intensity value change $\Delta I(\mathbf{e}) \in (0, 1]$ caused by the steepest ramp passing through the edge pixel $\mathbf{e}$.

2.3. Adaptive Domain Transformation

Using the above information yielded by the low level segmentation algorithm, we now carry out an adaptive domain transformation of the ramps corresponding to all the edge pixels.

We create displacement vectors $\mathbf{d}(\mathbf{x})$ for the vectors $\mathbf{x} \in \Omega_{LR}$ in the low resolution uniform grid. For each ramp pixel $\mathbf{r}_j(\mathbf{e})$, we define the displacement vector $\mathbf{d}(\mathbf{r}_j(\mathbf{e}))$ to be a vector in the direction of $\mathbf{e}$ from $\mathbf{r}_j(\mathbf{e})$, as follows:

$$\mathbf{d}(\mathbf{r}_j(\mathbf{e})) = \beta(\mathbf{e})(\mathbf{e} - \mathbf{r}_j(\mathbf{e})), \quad j = 1, ..., N_r(\mathbf{e}). \tag{3}$$

$\beta(\mathbf{e}) \in [0, 1)$ is an adaptive scaling parameter, and controls the magnitude of displacement for each ramp pixel. We define it to be,

$$\beta(\mathbf{e}) = \frac{L-1}{L} \left(\Delta I(\mathbf{e})\right)^{\alpha}. \tag{4}$$

The dependence of $\beta(\mathbf{e})$ on the ramp intensity value change $\Delta I(\mathbf{e})$ allows for a greater compression of the ramps that correspond to high contrast edges. The parameter α controls the nature of this dependence. We set this parameter experimentally to 0.5 in our experiments.

The factor $\frac{L-1}{L}$ ensures that the distance between neighboring ramp pixels in the transformed domain remains at least $1/L$. This, in turn, ensures that the distance between the neighboring ramp pixels after the subsequent upscaling step (as described in the next section) remains at least 1.

For pixels that do not belong to the edge ramps, the displacement vectors are set to zero.

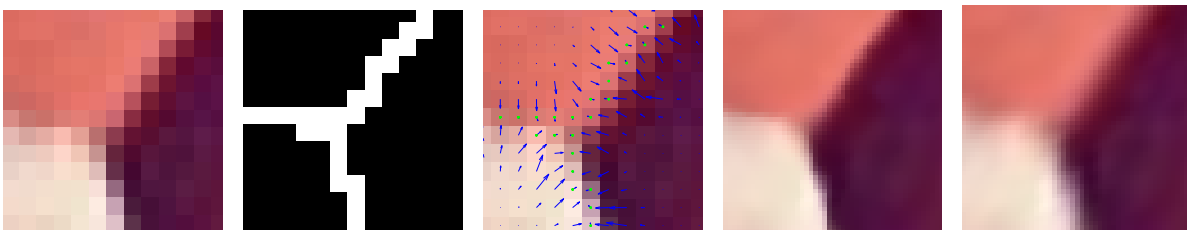
$$\mathbf{d}(\mathbf{x}) = 0, \forall \mathbf{x} \notin \mathcal{E}. \tag{5}$$

After computing all displacement vectors, we smooth the resultant displacement field $\mathbf{d}$ with a Gaussian filter. Figure 2(c) shows an example of the displacement field obtained for a small image patch.

2.4. Edge-Oriented Non-Uniform Interpolation

The displacement field $\mathbf{d}$ therefore defines a non-uniformly spaced domain Ω_{NU} from the uniformly spaced domain Ω_{LR} as,

$$\Omega_{NU} = \{\mathbf{y} : \mathbf{y} = \mathbf{x} + \mathbf{d}(\mathbf{x}), \forall \mathbf{x} \in \Omega_{LR}\}. \tag{6}$$



(a) (b) (c) (d) (e)

Fig. 2. (a) An input LR image patch. (b) Edge pixels obtained using the segmentation algorithm of [12]. (c) Displacement field (blue arrows) obtained around the edge pixels (green dots). (d) Result of non-uniform edge preserving interpolation into the HR domain. This serves as our prior constraint image. (e) Result of simple bicubic interpolation of (a). Our prior image (d) encourages sharp boundaries through preservation of ramp structures.

Using the pixel locations in Ω_{NU} and the corresponding pixels values in I_{LR}, we now interpolate I_{LR} over the uniform HR domain Ω_{HR}. The resulting image I_p would serve as our prior constraint image and as the initial guess for the iterative SR algorithm.

Although there exist several popular techniques for interpolating an image from non-uniform samples [17, 18], we choose a rather simple, one-pass filtering procedure for obtaining I_p. We compute $I_p(\mathbf{z})$ using the relation,

$$I_p(\mathbf{z}) = \frac{G_{\Sigma_i}(\|L\mathbf{y}_i - \mathbf{z}\|)I_{LR}(\mathbf{x}_i)}{G_{\Sigma_i}(\|L\mathbf{y}_i - \mathbf{z}\|)}, \tag{7}$$

where $\mathbf{z} \in \Omega_{HR}, \mathbf{y}_i \in \Omega_{NU}, \mathbf{x}_i \in \Omega_{LR}$ and $\mathbf{y}_i = \mathbf{x}_i + \mathbf{d}(\mathbf{x}_i)$. We choose $G_{\Sigma_i}(.)$ to be a 2D Gaussian kernel with an adaptive anisotropic covariance structure Σ_i.

We choose Σ_i to be adaptive in order to interpolate in an edge oriented manner. We use the already computed displacement field $\mathbf{d}$ to appropriately steer the anisotropic Gaussian interpolation kernel. Let θ_i be the angle made by the vector $\mathbf{d}(\mathbf{x}_i)$, with the first (horizontal) axis. Since the vector $\mathbf{d}(\mathbf{x}_i)$ points towards the edge and along the direction of the steepest edge ramp, we should interpolate in the orthogonal direction. We therefore choose Σ_i to be of the form,

$$\Sigma_i = \begin{cases} \begin{bmatrix} \sigma + s & 0 \\ 0 & \sigma - s \end{bmatrix} \cdot R(\frac{\pi}{2} + \theta_i) & \text{if } \mathbf{d}(\mathbf{x}_i) \neq 0 \\ \sigma \mathbb{1}_{\{2\times 2\}} & \text{if } \mathbf{d}(\mathbf{x}_i) = 0 \end{cases}. \tag{8}$$

$R(.)$ is a rotation matrix, and $\mathbb{1}_{\{2\times 2\}}$ is a 2×2 identity matrix. The scalar s controls the eigenvalue spread of the resultant covariance matrix. The variance parameter σ is chosen based on the desired upsampling factor L. For $L = 4$ as in our experiments, we choose $\sigma = 2$, and $s = 1$.

The above formulation ensures that if $\mathbf{d}(\mathbf{x}_i) \neq 0$ (that is, if the pixel at $\mathbf{x}_i$ has been displaced), then the larger eigenvector of the covariance matrix Σ_i is orthogonal to $\mathbf{d}(\mathbf{x}_i)$, and therefore orthogonal to the direction of the steepest edge ramp. This orients the kernel to interpolate parallel to the edge. Given the displacement field $\mathbf{d}$, it therefore becomes trivial to incorporate edge awareness in the interpolation procedure, obviating the need for more elaborate schemes [19]. Such an interpolation scheme is akin to structure-aware filtering methods such as in [20] and [21].

Fig. 2(d) shows an example of our prior image I_p obtained from the initial image patch of Fig. 2(a), after a 4X interpolation over the non-uniform grid created by the displacement field in Fig. 2(c). We can see from this example that I_p maintains sharper edges, as compared to using classical methods like uniform bicubic interpolation (Fig. 2(e)).

2.5. SR Reconstruction

Our final SR result $\hat{I}_{HR}$ is estimated by using the thus computed prior image I_p in the regularized backprojection algorithm. Our cost function is therefore,

$$J_{SR}(\hat{I}_{HR}) = \left|\left(\hat{I}_{HR} * f_{psf}\right) \downarrow - I_{LR}\right|^2 + \lambda \left|\hat{I}_{HR} - I_p\right|^2. \tag{9}$$

This is optimized using a gradient descent strategy:

$$\hat{I}_{HR}^{(t+1)} = \hat{I}_{HR}^{(t)} - \mu \nabla J_{SR}(\hat{I}_{HR}^{(t)}), \tag{10}$$

where μ is an appropriately chosen step size. We use our prior image I_p as the initial guess for the above algorithm.

$$\hat{I}_{HR}^{(0)} = I_p. \tag{11}$$

Note that the gradient $\nabla J_{SR}(\hat{I}_{HR})$ in (10) is with respect to the HR estimate $\hat{I}_{HR}^{(t)}$. Therefore, unlike the prior terms in the gradient domain [7], since our prior is in the image domain, we do not need to compute image derivatives or laplacians while implementing the iterative scheme of (10).

3. RESULTS

We use an upscaling factor of $L = 4$ in all our experiments. We use a 7×7 Gaussian function of variance 1.6 as f_{psf} for simulating the downsampling process. We choose the step size parameter to be $\mu = 0.5$. We experimentally set the regularization parameter to $\lambda = 0.6$, as this gave the most visually pleasing results.

Fig. 3 shows our results and comparisons on four images - *Penguin, Lighthouse, Flower, Bride*. The columns show results of nearest neighbor interpolation (NN), bicubic interpolation, iterative backprojection (IBP) [10], gradient profile prior (GPP) [1] [7], our proposed method, and the ground truth images.

We can see that our results do not exhibit the 'ringing' artifacts around edges as commonly seen in the IBP [10] results. Our results

[1] The source code for the method in [11] contains the GPP [7] implementation. http://yuwing.kaist.ac.kr/projects/superresolution/index.htm

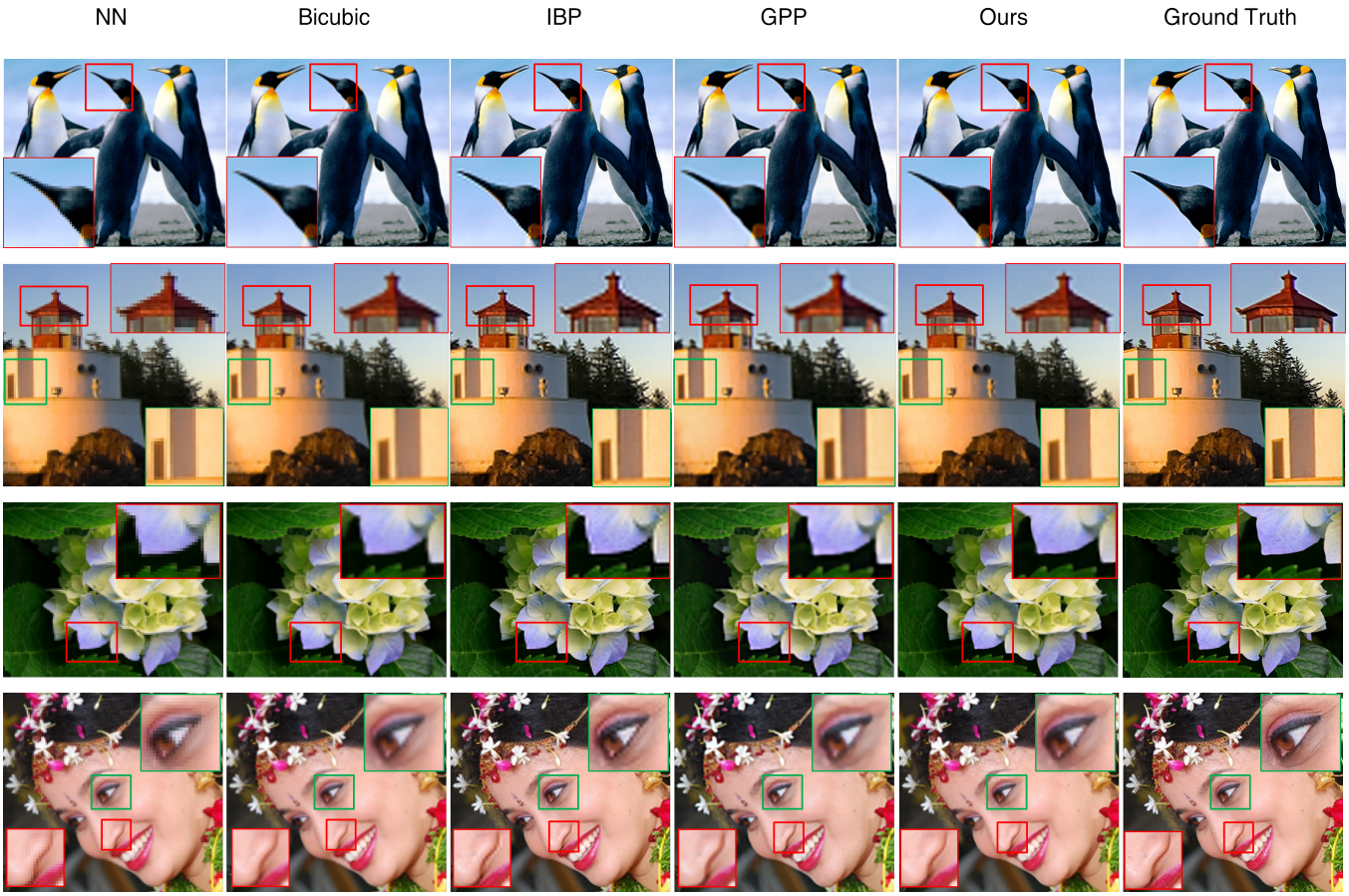


Fig. 3. Super-resolution results on the *Penguin, Lighthouse, Flower* and *Bride* images. The columns from left to right show nearest neighbor interpolation (NN), bicubic interpolation, iterative backprojection algorithm (IBP) [10], gradient profile prior (GPP) [7], our results, and the ground truth images. (Best viewed if zoomed in)

Table 1. Comparison of the mean squared error (MSE) obtained by bicubic interpolation, IBP [10], and our method.

Image	Bicubic	IBP [10]	Ours
Penguin	15.4572	13.7490	**13.3392**
Lighthouse	17.2851	16.1000	**16.0032**
Flower	10.3590	8.7514	**8.3666**
Bride	14.7352	13.1533	**13.0427**

are also sharper than GPP [7], which uses a gradient domain prior learnt from training examples. Our results are more closer to the ground truth than the GPP method along curved edges (nose curvature in the *Bride* image) and pointed structures (beak of the *Penguin* image), perhaps due to the advantages of the ramp based structure model over gradient based modeling, as discussed earlier.

We also compute the mean squared error (MSE) between our results and the ground truth, and compare it to those obtained though bicubic interpolation and IBP [10]. Table 1 tabulates the MSE values. Our results are slightly better than those of IBP [10]. Our parameters, however, have been set for best visual quality which is not always accurately reflected by MSE values.

4. CONCLUSION

In this paper, we have presented a prior for single image super-resolution, that is aimed at preserving ramp structures along the edges of the image. The motivation for our work stems from the advantages of ramp based characterization of low level image structure, as compared to conventional gradient filters. We have shown that the ramp based characterization is indeed useful for SR problems. We have obtained visually better results than the classical IBP method [10] as well as the recent gradient based GPP method [7] which uses training data to learn an 'optimum' transformation between LR gradients and the HR counterpart. Learning can be incorporated in the proposed ramp based prior as well, and it can perhaps yield a more principled way of obtaining the displacement field for computing our prior image, as compared to the scheme proposed here. We would be looking into these extensions in the near future.

5. ACKNOWLEDGMENTS

This work was supported by US Office of Naval Research grant N00014-12-1-0259, the Joan and Lalit Bahl Fellowship and the Computational Science & Engineering Fellowship at UIUC.

6. REFERENCES

[1] W.T. Freeman and E.C. Pasztor, "Learning low-level vision," in *ICCV*, 1999.

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

[3] J. Yang, J. Wright, T.S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Proc.*, 2010.

[4] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *ICCV*, 2009.

[5] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph.*, 2010.

[6] R. Fattal, "Image upsampling via imposed edge statistics," *ACM Trans. Graph.*, 2007.

[7] J. Sun, J. Sun, Z. Xu, and H-Y. Shum, "Gradient profile prior and its applications in image super-resolution and enhancement," *IEEE Trans. Image Proc.*, 2011.

[8] H.A. Aly and E. Dubois, "Image up-sampling using total-variation regularization with a new observation model," *IEEE Trans. Image Proc.*, 2005.

[9] L. Wang, S. Xiang, G. Meng, Huai-yu Wu, and C. Pan, "Edge-directed single image super-resolution via adaptive gradient magnitude self-interpolation," *IEEE Trans. CSVT*, 2013.

[10] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP*, vol. 53, no. 3, pp. 231–239, Apr. 1991.

[11] Y-W. Tai, S. Liu, M. Brown, and S. Lin, "Super resolution using edge prior and single image detail synthesis," in *CVPR*, 2010.

[12] E. Akbas and N. Ahuja, "From ramp discontinuities to segmentation tree," in *ACCV*, 2009.

[13] H. Arora and N. Ahuja, "Analysis of ramp discontinuity model for multiscale image segmentation," in *ICPR*, 2006.

[14] N. Ahuja, "A transform for multiscale image segmentation by integrated edge and region detection," *IEEE TPAMI*, vol. 18, no. 12, pp. 1211–1235, 1996.

[15] A. Singh and N. Ahuja, "Exploiting ramp structures for improving optical flow estimation," in *ICPR*, 2012.

[16] M. Nam and N. Ahuja, "Learning human preferences to sharpen images," in *ICPR*, 2012.

[17] K. Grchenig and H. Schwab, "Fast local reconstruction methods for nonuniform sampling in shift invariant spaces," *SIAM Journal on Matrix Analysis and Applications*, 2003.

[18] A. Aldroubi and K. Grchenig, "Nonuniform sampling and reconstruction in shift-invariant spaces," *SIAM Review*, 2001.

[19] X. Li and M. Orchard, "New edge-directed interpolation," *IEEE Trans. on Image Proc.*, 2001.

[20] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. Image Proc.*, 2007.

[21] Tuan Q. Pham, Lucas J. van Vliet, and Klamer Schutte, "Robust fusion of irregularly sampled data using adaptive normalized convolution," *EURASIP J. Appl. Signal Proc.*, 2006.