Super-resolution Using Sub-band Constrained Total Variation

Priyam Chatterjee, Vinay P. Namboodiri and Subhasis Chaudhuri

Department of Electrical Engineering,
Indian Institute of Technology, Bombay,
Powai, Mumbai 400 076.
India.
{priyam, vinaypn, sc}@ee.iitb.ac.in

Abstract. Super-resolution of a single image is a severely ill-posed problem in computer vision. It is possible to consider solving this problem by considering a total variation based regularization framework. The choice of total variation based regularization helps in formulating an edge preserving scheme for super-resolution. However, this scheme tends to result in a piece-wise constant resultant image. To address this issue, we extend the formulation by incorporating an appropriate sub-band constraint which ensures the preservation of textural details in trade off with noise present in the observation. The proposed framework is extensively evaluated and the experimental results for the same are presented.

1 Introduction

Super-resolution is the process of increasing the spatial details in an image by computational means. In certain applications it is often not possible to obtain an image with a high level of detail. In such cases super-resolution methods become extremely necessary to provide a better observation from one or more degraded available images of the scene. Image super-resolution finds a variety of applications in video and image quality improvement (HDTV conversion), health diagnosis (from X-ray or sonographic images) and as a preprocessing step for any application where a better quality input picture is a requirement.

The problem of super-resolution can formally be stated as follows. There are p observed images y_m $(m = 1...p)$, each of size $M_1 X M_2$ which are the decimated, blurred and noisy versions of a single high resolution image z of size $N_1 X N_2$ where $N_1 = qM_1$ and $N_2 = qM_2$. If y_m is the $M_1M_2 \times 1$ lexicographically ordered vector containing pixels from the low resolution image then a vector z of size $q^2M_1M_2 \times 1$ containing pixels of the high resolution image can be formed by placing each of the $q \times q$ pixel neighborhoods sequentially so as to maintain the relationship between a low resolution pixel and its corresponding high resolution pixel. After incorporating the blur matrix and the noise vector, the image formation model is written as

$$Y_m = D(H_m * z) + n_m, m = 1, ..., p \tag{1}$$

where D is the decimation matrix of size $M_1 M_2 \times q^2 M_1 M2$, H is the blurring point spread function (PSF) and n_m is the $M_1 M_2 \times 1$ noise vector and p is the number of low resolution observations. Here we assume the blur kernel to be shift invariant and $*$ denotes the convolution operation.

The various approaches towards solving this problem can be broadly classified as being based on super-resolution using multiple images, super-resolution using a single image based on learning low-level features and super-resolution of a single image based on interpolation. In this paper, we propose a technique for super-resolution of a single image based on interpolation using total variation regularization. Here we incorporate a stronger data term for super-resolution by considering correlation among sub-bands of an image. This ensures that the resulting interpolation using total variation regularization has better texture preserving properties. Thus the contribution in this paper is a super-resolution technique of interpolation based on total variation regularization with the data terms based on the image formation model and sub-band correlation.

In the next section we discuss the related work. In section 3 we present the proposed technique. In section 4 we discuss the implementation details and the technique is validated in section 5 by comparison with various other techniques. We finally conclude in section 6.

2 Related work

A vast collection of literature is available where researchers aim to perform resolution enhancement using a wide variety of techniques. Here we discuss the different approaches

2.1 Super-resolution from multiple images

In cases where multiple degraded observations are available the low resolution (LR) images must first be registered to determine the inter-pixel shifts before placing them onto a higher dimensional grid. A survey of different registration methods is provided in [1]. The reconstruction of the image in the higher dimensional grid using maximum likelihood estimation (MLE), Maximum a-priori (MAP) estimation with priors like Gaussian Markov random field (MRF) and Huber MRFs has been demonstrated by Capel *et al.* in [2]. Using a MAP estimator and blur as a cue Rajan *et al.* perform super-resolution in [3],[4]. In [5], [6] Joshi *et al.* use zoom as a cue. They consider the linear dependency of pixels in a neighborhood and model it as a simultaneous auto-regressive process which is then used as a prior term in a regularization framework. Regularization based techniques have also been used by Chan *et al.* in [7]. In this the authors make use of the algorithm developed for image inpainting in [8]. They obtain multiple blurred, noisy LR frames from adjacent frames of a video and attempt to form a high resolution (HR) image. A total variation (TV) based regularization method is then used to perform simultaneous inpainting and deblurring to obtain a super-resolved image. A combination of TV and bilateral filter named

Bilateral TV (BTV) has been used as a regularizing term for super-resolution by Farsiu *et al.* in [9], [10], [11]. As opposed to the L_2 norm used as a data fidelity term in most cases, the authors use the L_1 norm. They present a two step algorithm where they first use the median filter to build a high resolution grid from multiple LR images. Regularization is then done to perform an iterative interpolation to deblur as well as inpaint missing pixels in the HR grid. Another algorithm is presented in [10] where the authors make use of the temporal information between frames of a low resolution video. SR is then performed under a control theoretic approach using an approximation of the Kalman Filter along with the previous framework.

2.2 Learning based Super-resolution

Considerable amount of work has also been done in the domain of super-resolution from a single image by making use of a image database where LR-HR image pairs are provided. One of the foremost papers in this approach is the work by Freeman *et al.* [12]. Here the authors model have a generative model for scenes and their rendered image, with a Markovian relationship between them. Bayesian belief propagation is used to estimate the posterior probability of the scene given an image. The priors are learnt through a database of LR-HR images. A similar approach has been proposed by Baker and Kanade [13], where low level features are recognized and the corresponding high resolution features are "hallucinated". In [14], the authors suggest a method to obtain an approximate fast one pass solution to the Markov network where again the low-level features are learnt using patches from LR-HR image database. In this paper we restrict ourselves to the case where super-resolution is performed from a single observation without the use of any such database.

2.3 Interpolation based Super-resolution

Some researchers have also applied different methods to address the issue of zooming into an image when only a single low resolution version of the scene is available. The main challenge here is to preserve edges that are present in any natural image. A variety of linear and non-linear tools are available which try to address this issue. A detailed mathematical analysis of regularization based schemes has been provided by Malgouyres and Guichard in [15]. The authors in [16] provide an interpolation method under the total variation regularization scheme. In their paper they start off with a higher resolution image formed by zero-padded interpolation of the LR image. A constrained gradient descent algorithm is presented where the authors minimize the gradient energy of the image which conforms to a linear smoothing and sampling process. The use of TV for super-resolution has also been demonstrated by Aly and Dubois in [17]. In their method they modify the data fidelity term to closely model the assumed image acquisition model. Their iterative algorithm then makes use of the back projection technique introduced by Irani and Peleg [18] for data fidelity in a regularization framework. The authors then present an algorithm that converges to a unique solution irrespective of the starting interpolated image. However, the resultant image depends upon the choice of the image formation model. The dependence of the result on the selection of the proper mathematical model that captures the downsampling process for such regularization based methods has been discussed in [19]. Jiji *et al.* in [20] propose an interpolation technique where the aliasing present in the LR image is used. They assume knowledge of the bandwidth and the amount of aliasing in a given observation and use a signal processing approach to perform super-resolution.

3 Super-resolution using TV approach

The image formation model for the low resolution image from a high resolution image is given as

$$y(x) = d(x)(h * z) + n(x) \tag{2}$$

Here $d(x)$ is the decimation matrix, h is the blur point spread function, z is the high resolution image and $n(x)$ is the noise function. Given an approximation u to the high resolution image z, and the image u_0 which is the upsampled version of the observed low resolution image, the residual error is given as

$$r(x) = u_0(x) - (h * u). \tag{3}$$

Based on the error function an objective function can be formulated the minimization of which gives the high resolution image. The objective function is given as

$$E(u) = \int \big((r(x))^2 + \alpha|\nabla u|\big) \tag{4}$$

The solution for super-resolution from a single image can be given in terms of the following objective function. Here the first term is the data term and the second term is the L_1 (TV) regularization term. The choice of TV norm has found favor in the image restoration community because it allows discontinuities in its solution. As opposed to the L_2 norm it does not smoothen the image across edges. Our motivation for the use of TV based regularization stems from its edge preserving property which is vital for super-resolution. However, the current formulation of data term and regularization term results in a solution that preserves strong edges, however, the finer details of texture are lost in the solution of the above objective function. This can be easily understood by considering the following argument. If there exists a weak edge (the magnitude of gradient is small), then the regularization constraint gives it a low weight. The data fidelity constraint, due to the averaging nature of the blurring kernel, would also not enforce the preservation of the edge. In the iterative energy minimization approach these finer details are therefore lost. In order to preserve texture details and finer details it is required to consider an additional data fidelity constraint. This constraint we formulate as a correlation constraint over various sub-bands of an image.

The objective function we use is then

$$J(u) = \int_{\Omega} |\nabla u| \; dxdy + \frac{1}{2}\alpha \int_{\Omega} (u * h - u_0)^2 dxdy + \frac{1}{2}\sum_k \lambda_k \int_{\Omega_k} (\tilde{U}_k - U_{0k})^2 \; d\nu d\omega \tag{5}$$

where u denotes the HR restored image, h is the blurring kernel, u_0 is the interpolated version of the input LR image, $\tilde{U}_k$ denotes the k^{th} spectral sub-band of the estimated LR image formed under the known decimation model, U_{0k} is the k^{th} sub-band of the input LR image and λ_k is the corresponding weighing term for the regularizer. Thus an interpolation of the LR observation serves as the initial estimate of the HR image.

4 Implementation Details

The objective function given in eqn. (5) is minimized by an iterative gradient descent technique as done commonly in the literature [21]. The corresponding Euler Lagrange equation for the objective function is given by

$$\frac{\partial}{\partial x}\left(\frac{u_x}{\sqrt{{u_x}^2 + {u_y}^2}}\right) + \frac{\partial}{\partial y}\left(\frac{u_y}{\sqrt{{u_x}^2 + {u_y}^2}}\right) - \alpha(u * h - u_0)$$

$$-D^{-1}\mathscr{F}^{-1}\sum_k \lambda_k(\tilde{U}_k - U_{0k}) \; = 0, \; x, y \in \Omega \; . \tag{6}$$

$$\frac{\partial u}{\partial n} = 0 \text{ on the boundary of } \; \Omega = \partial\Omega \tag{7}$$

The resulting iterative updation process is then given by

$$u^{(n+1)} = u^{(n)} + \triangle t \left(\nabla . \frac{\nabla u}{|\nabla u|} + \alpha(u_0 - u * h) + D^{-1}\mathscr{F}^{-1}\sum_k \lambda_k(U_{0k} - \tilde{U}_k)\right)^{(n)} \tag{8}$$

where D^{-1} is the upsampling process and $\mathscr{F}^{-1}$ implies the inverse Fourier transform. Under this framework, it then becomes possible to assign different weights (λ_k) to existing errors terms in different bands. As opposed to other schemes that we have discussed before, the additional constraint term is calculated and weighed in the spectral domain. The inverse Fourier transform is then applied and finally it is scaled to match the high resolution image dimensions. It may be argued here that it follows from Parseval's theorem that calculating error power in the spatial domain and the frequency domain should be equivalent. However, the operation in the spectral domain makes it easy to split an image into separable components based on spectral contribution. The number of spectral bands k does affect the quality of super-resolution. In general, the higher the number of spectral bands, more the flexibility for preserving details. We have experimentally tried the method with 2 and 4 spectral bands. Under the absence of noise, we use a higher weight factor (λ_k) for the higher spectral bands which capture the finer details and the edges of the image. Using such a model it is then possible for us to enforce that more importance is given to data fidelity at the edges. This should ensure that the image that is formed is a sharper super-resolved image of the input LR observation under the known image decimation model. On the other hand, noise in an image can be expected to be captured in the higher frequency sub-bands, which necessitates the use of smaller weights for higher sub-bands when the input image is noisy. An appropriate choice of λ_ks would ensure a proper trade-off between the sharpness of the super-resolved image and the accentuation of the noise present. We have experimented with both noisy and noiseless cases and the results are discussed in Section 5.

5 Results

For our experiment we take the initial starting image as the bicubic interpolation of the input LR image. The images at every iteration of the restoration process are decomposed into two bands and based on the theory presented above we apply a higher weight to higher frequency component of the image. For the results shown here using a decomposition into only 2 bands we use the values $\alpha = 0.7$, $\lambda_1 = 0.6$ and $\lambda_2 = 0.8$ where a higher index of λ value implies a higher frequency band. In this experiment the higher 40% of the spectrum was assumed to capture most of the edge information of the image. We also demostrate results when a 4 band decompositon is done. In this case the frequency spectrum is equally divided into 4 bands. The parameter values in this case are $\alpha = 0.8$, $\lambda_1 = 0.4$, $\lambda_2 = 0.6$, $\lambda_3 = 1$ and $\lambda_4 = 1.2$. The results obtained using these parameter values are shown in Figure 1 and Figure 2.

In Figure 1 we can see that the total variational deblurring performed on the bicubic reconstruction sharpens the image at edges but at the cost of loss of the texture. This is not the case for Figure 1(d) & (e). This can be noted from the presence of texture in the hat and the hair, even though the overall reconstruction remains sharp. This is specifically what we wanted to achieve by our method. A similar effect can be seen in Figure 2 where the result from our method yields a better texture than that of TV based deblurring. This is visible at the terrain and finer details on the tank. This proves that band splitting and differential weighing of the bands indeed perform better as far as restoration of texture is concerned.

We also compare our results with the alias-free interpolation method [20] proposed by Jiji *et al.* and the results are presented in Fig. 3. The alias-free interpolation method performs super-resolution by introducing high frequency components. Since, this method does not assume any priors or additional data set, the assumptions are comparable to our method. Here, it is assumed that the aliasing is present in the top ten percent of the spectrum and it performs reconstruction based on samples from other parts of the spectrum. The input image is a 64x64 LR image and we perform 2X Zoom using both the methods. The parameters used for TV based method are same as those used in the previous

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

Fig. 1. SR using proposed TV based approach for 2× zoom : (a) Input LR image, (b) bicubic interpolated image used as the initial estimate, (c) TV based deblurring of (b), (d) SR using modified TV based approach using only 2 bands, (e) reconstruction using 4 bands.

experiment on a 128x128 LR image. We show the results for the proposed method using two bands and four bands respectively. The results shown in Fig. 3 (c) and (d), show that the proposed method adds more coherent high frequency components and the resultant images are sharper as compared to the alias free interpolation method. The result with four bands are better as compared to the result with two frequency bands, indicating the effectiveness of multiple bands.

We next compare our results with image super-resolution methods described in [22] and [23]. In [22], the authors suggest a multi-image super-resolution method. However, the method based on delaunay triangulation based approximation can also be used for interpolation with a single image. In [23], the authors consider the use of kernel regression for image upscaling. We compare our proposed method with these methods and the results are shown in Fig. 4. Fig. 4(a) shows the bicubic interpolated Lena image for 3 × zoom factor which is used

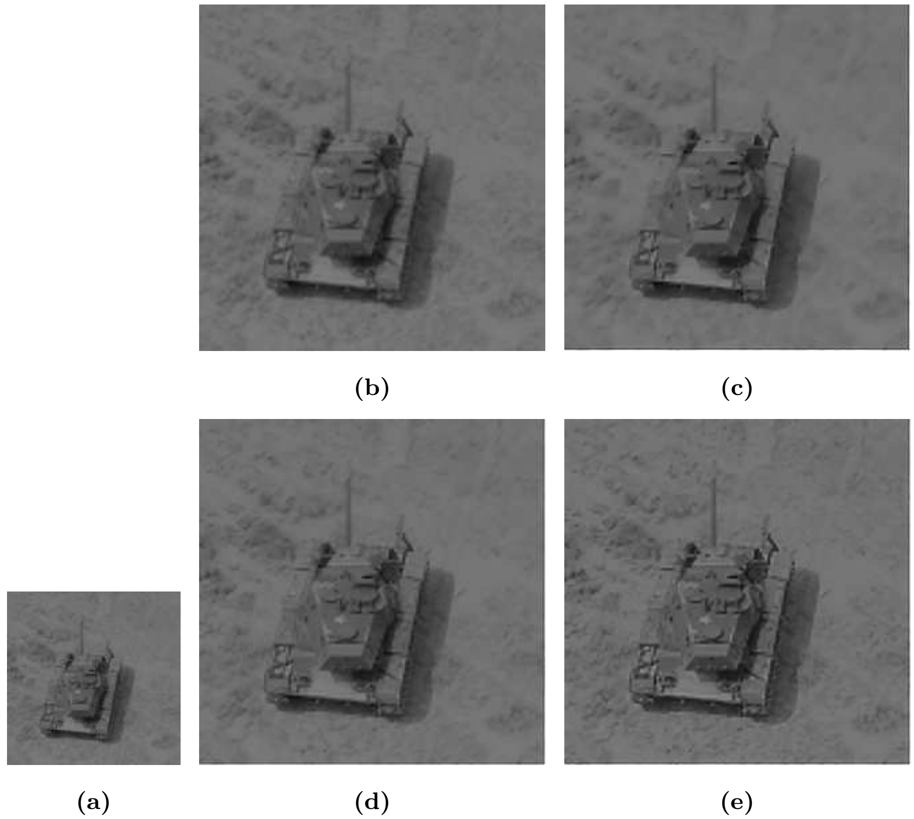


Fig. 2. SR using proposed TV based approach for 2× zoom : (a) Input LR image, (b) bicubic interpolated image used as the initial estimate, (c) TV based deblurring of (b), (d) SR using modified TV based approach using only 2 bands, (e) reconstruction using 4 bands.

as the initial condition in our algorithm. Fig. 4(b) shows the result from the delaunay triangulation based method [22] and fig. 4(c) shows the result from the kernel regression based method. Fig. 4(d) shows the result from the proposed approach. It can be seen that the results from the proposed approach has better texture preserving properties as compared to the others. This can be more clearly seen by comparing the texture in the hat and hair areas of the image.

We also attempt to try our method where the image is corrupted by zero mean additive noise. We use a Gaussian noise of variance 25 ($[0, 255]$ being the range of pixel values in the image). In our reconstruction process we do not make use of any information about the nature of the noise. Our theory builds on the assumption that this noise remains limited to higher frequency bands of the image. Hence we use a lower scaling factor for the higher frequency band. The results obtained using a 2 band decomposition with parameter values of

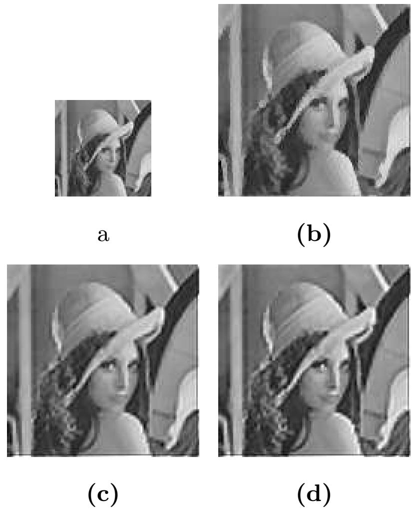


Fig. 3. Comparison of proposed method with alias-free interpolation method [20] approach for 2× zoom : (a) Input LR image, (b) Image interpolated using alias-free Interpolation method [20], (c) SR using modified TV based approach using only 2 bands, (e) reconstruction using 4 bands.

$\alpha = 0.7$, $\lambda_1 = 0.8$ and $\lambda_2 = 0.6$ are shown in Figure 5(d). For the case where we make use of a 4 band decomposition the parameters used for denoising are $\alpha = 0.6$, $\lambda_1 = 1$, $\lambda_2 = 0.8$, $\lambda_3 = 0.6$ and $\lambda_4 = 0.4$. The result obtained using this configuration is shown in Figure 5(e).

It can be seen from Figure 5(b) that the bicubic interpolation does not perform any denoising, as is expected. Total variation based regularization performed on the bicubic interpolated image, in Figure 5(c), reduces a lot of noise but the resultant image is smooth and lacks texture. The modified approach yields a sharper reconstruction but the presence of noise is clearly visible. Apart from an enhancement of details, an improvement in the PSNR is also observed for the proposed method. More denoising will lead to a smoother reconstruction, as can be expected from the method. The reason for this is that spectral bands which capture sharp edges and texture will also contain most of the noise and hence it becomes difficult to distinguish between noise and texture using the present method.

6 Conclusion

The total variation based regularization scheme has been widely used by researchers for image restoration. It has been known to be a very good tool for denoising and deblurring. In our work we make use of the advantages of this method and apply it to perform super-resolution. We show that though in itself

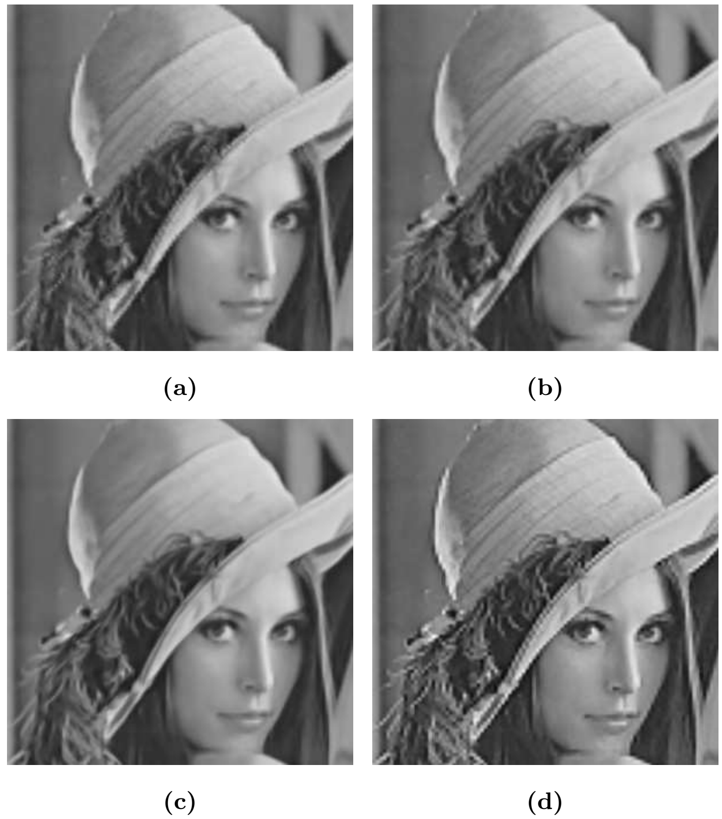
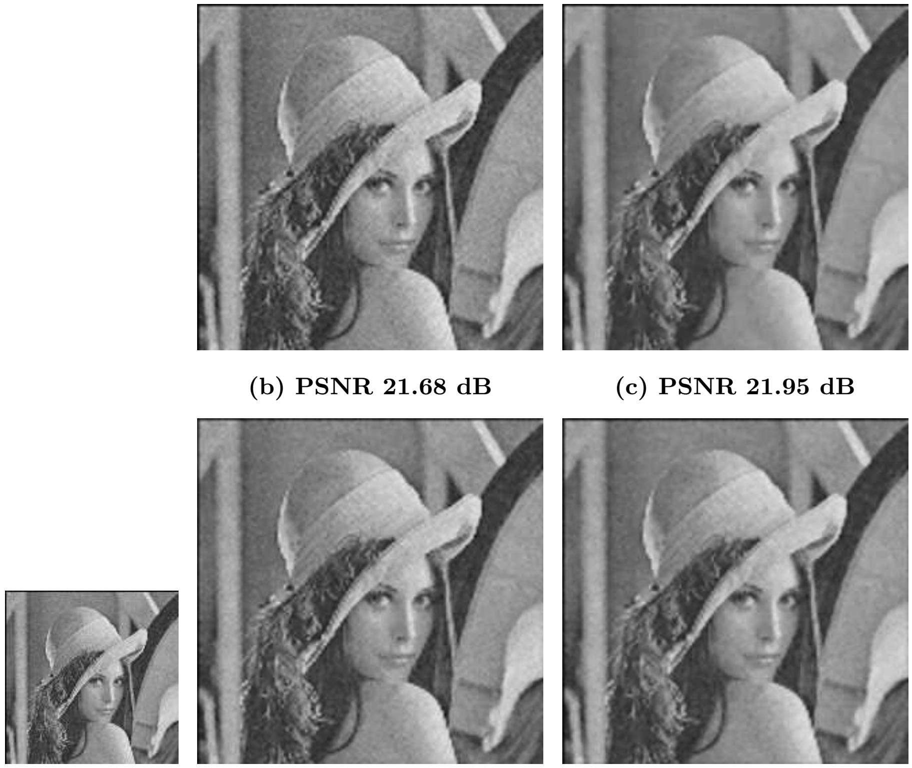


Fig. 4. SR using proposed TV based approach for 3× zoom : (a) bicubic interpolated image used as the initial estimate, (b) Image interpolated using delaunay triangulation [22], (c) Image interpolated using kernel regression method [23] (d) Reconstruction using proposed method with 4 bands.

the method results in a good interpolated image, it fails to reconstruct texture and other finer details in an image. This is obtained through decomposition of the image into multiple spectral bands and enforcing differential data fidelity in each band. As shown and discussed before, in the presence of noise this method provides a trade-off between the desired sharpness of the edges and the amount of denoising achieved. This is due to the inherent limitation of the TV denoising process which cannot tell finer texture apart from noise. Further investigation is thus necessary to make the method more robust in the presence of noise.

References

1. Zitova, B., Flusser, J.: Image registration methods : A survey. Image and Vision Computing **21** (2003) 977–1000



(b) PSNR 21.68 dB (c) PSNR 21.95 dB

(a) (d) PSNR 22.36 dB (e) PSNR 22.13 dB

Fig. 5. SR using proposed TV based approach for 2× zoom for a noisy case: (a) Input LR image, (b) bicubic interpolated image, (c) TV based denoising of (b), (d) SR using modified TV based approach using only 2 bands, (e) reconstruction using 4 bands.

2. Capel, D., Zisserman, A.: Computer vision applied to super resolution. IEEE Signal Processing Magazine **20** (2003) 75–86
3. Rajan, D., Chaudhuri, S., Joshi, M.V.: Multi-objective super resolution: Concepts and examples. IEEE Signal Processing Magazine **20** (2003) 49–61
4. Rajan, D., Chaudhuri, S.: Simultaneous estimation of super-resolved scene and depth map from low resolution defocused observations. IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003) 1102–1117
5. Joshi, M., Chaudhuri, S., Rajkiran, P.: Super-resolution imaging: Use of zoom as a cue. Image and Vision Computing **22** (2004) 1185–1196
6. Joshi, M., Chaudhuri, S., Rajkiran, P.: A learning based method for image super-resolution from zoomed observations. IEEE Trans. Systems, Man & Cybernetics, Part-B **35** (2005) 527–537
7. Chan, T.F., Ng, M.K., Yau, A.C., Yip, A.M.: Superresolution image reconstruction using fast inpainting algorithms. Technical report, University of California, Los Angeles (2006)

8. Chan, T.F., Yip, A.M., Park, F.E.: Simultaneous total variation image inpainting and blind deconvolution. Technical report, University of California, Los Angeles (2004)
9. Farsiu, S., Robinson, D., Elad, M., Milanfar, P.: Robust shift and add approach to super-resolution. In: Proc. of the SPIE on Applications of Digital Signal and Image Processing. Volume 5203., San Diego, USA (2003) 121–130
10. Farsiu, S., Elad, M., Milanfar, P.: A practical approach to super-resolution. In: Proc. of the SPIE: Visual Communications and Image Processing. Volume 6077., San Jose, USA (2006) 24–38
11. Farsiu, S., Robinson, D., Elad, M., Milanfar, P.: Advances and challenges in super-resolution. International Journal of Imaging Systems and Technology **14** (2004) 47–57
12. Freeman, W.T., Pasztor, E.C., Carmichael, O.T.: Learning low-level vision. International Journal of Computer Vision **40** (2000) 25–47
13. Baker, S., Kanade, T.: Limits on super-resolution and how to break them. IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002) 1167–1183
14. Freeman, W., Jones, T., Pasztor, E.: Example-based super-resolution. IEEE Transactions on Computer Graphics and Applications **22** (2002) 56–65
15. Malgouyres, F., Guichard, F.: Edge direction preserving image zooming: A mathematical and numerical analysis. SIAM Journal on Numerical Analysis **39** (2002) 1–37
16. Guichard, F., Malgouyres, F.: Total variation based interpolation. In: Proceedings of the European Signal Processing Conference. Volume 3., Amsterdam, The Netherlands, The Netherlands, Elsevier North-Holland, Inc. (1998) 1741–1744
17. Aly, H., Dubois, E.: Image up-sampling using total-variation regularization. IEEE Transaction on Image Processing **14** (2005) 1646–1659
18. Irani, M., Peleg, S.: Improving resolution by image registration. CVGIP: Graphical Model and Image Processing **53** (1991) 231–239
19. Aly, H., Dubois, E.: Specification of the observation model for regularized image up-sampling. IEEE Transaction on Image Processing **14** (2005) 567–576
20. Jiji, C., Neethu, P., Chaudhuri, S.: Alias-free interpolation. In: Proceedings of 9th ECCV - Part IV. Lecture Notes in Computer Science, Graz, Austria, Springer (2006) 255–266
21. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear Total Variation Based Noise Removal Algorithms. Physica D **60** (1992) 259–268
22. Lertrattanapanich, S., Bose, N.K.: High resolution image formation from low resolution frames using delaunay triangulation. IEEE Transactions on Image Processing **11** (2002) 1427–1441
23. Takeda, H., Farsiu, S., Milanfar, P.: Kernel regression for image processing and reconstruction. accepted for publication in IEEE Transactions on Image Processing (2006)