IMAGE INTERPOLATION USING SHEARLET BASED SPARSITY PRIORS

H. Lakshman[1], W.-Q Lim[2], H. Schwarz[1], D. Marpe[1], G. Kutyniok[2] and T. Wiegand[1,3]

[1]Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, Germany.
[2]Department of Mathematics, Technical University of Berlin, Germany.
[3]Department of Electrical Engineering, Technical University of Berlin, Germany.

ABSTRACT

This paper proposes an image interpolation algorithm exploiting sparse representation for natural images. It involves three steps: (a) obtaining an initial estimate of the high resolution image using linear methods like FIR filtering, (b) promoting sparsity in a selected dictionary through thresholding and (c) extracting high frequency information from the approximation and adding it to the initial estimate. For the sparse modeling, a shearlet dictionary is chosen to yield a multiscale directional representation. The proposed algorithm is compared to several state-of-the-art methods to assess its objective as well as subjective performance. Compared to the cubic spline interpolation method, an average PSNR gain of around 0.7 dB is observed over a dataset of 200 images.

***Index Terms*—** Interpolation, Sparity, Shearlets

1. INTRODUCTION

Image interpolation refers to generating a High Resolution (HR) image from an input Low Resolution (LR) image. Signal processing theory for band limited signals advocates sampling higher than Nyquist rate and a sinc interpolation [1, 2]. The assumption of band-limitedness does not hold for most images, due to the existence of sharp edges. However, conventional schemes adhere to this philosophy and approximate the ideal low pass filter to produce acceptable results for many practical applications. Techniques like bilinear, bicubic interpolation, etc., are some popular examples that have low computational complexity. Extending the sampling theory to shift-invariant spaces without imposing band-limit constraint has led to a generalized interpolation framework, e.g., B-spline [3] and MOMS interpolation [4] that provide improvements in image quality for a given support of basis functions. However, these linear models cannot capture the fast evolving statistics around edges. Increasing the degree of basis functions in these linear models help to capture higher order statistics but result in longer effective support in the spatial domain and hence produce artifacts like ringing around edges, etc.

To improve the linear models, directional interpolation schemes have been proposed that perform interpolation along the edge directions, e.g., NEDI [5]. It is achieved through computing local covariances in the input image and using them to adapt the interpolation at high resolution, so that the support of the interpolator is along the edges. However, the resulting images still show some artifacts. The iterative back-projection [6] technique improves image interpolation when the downsampling process is known. However, the downsampling filter may not be known in many cases, or the input image may be camera captured, where the optical anti-alias filter used within the sampling system is not known during the subsequent image processing stages. Therefore, it is desirable to design a method that does not rely directly on the knowledge of the downsampling process.

In this paper, we recognize the fact that linear models like FIR filter based interpolation are faithful in interpolating the low frequency components but distort the high frequency components. The distortion in high frequencies is considered as noise and reduced using an iterative denoising [7] algorithm that makes use of sparity priors [8]. In underdetermined problems like image interpolation, sparsity priors can be useful to exploit the geometric structure of desired solutions while satisfying problem constraints. The domain for sparse representation can be fixed or learned using training data [9]. Here, we use a fixed transform, namely, the shearlet transform [10, 11]. Shearlet elements have anisotropic directional characteristics, important for modeling image features. Unlike other fixed transforms like curvelets [12] or contourlets [13], they provide a consistent design in continuous as well as discrete domain.

2. FRAMEWORK FOR HIGH FREQUENCY SYNTHESIS

The proposed framework, depicted in Fig. 1, follows the principle of image recovery through iterative denoising [7]. This principle has been previously applied to image interpolation [14], but for a fixed observation model that requires the LR image to be the lowpass subband of a specific wavelet transform. As stated earlier, it is desirable to have a method that does not assume a specific downsampling process. Therefore, we redesign the setup for typical anti-aliased LR images.

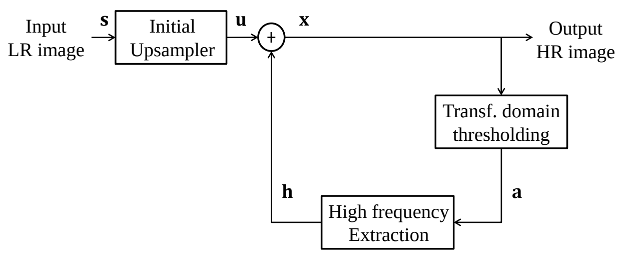


Fig. 1: *Framework for image interpolation. A linear model, e.g., FIR filter is used to produce an initial upsampled image. Then, high frequency components are extracted after a sparsity promoting step and used to refine the initial upsampled image.*

Consider an input LR signal of dimension $N \times 1$, represented as a vector $\mathbf{s} \in \mathbb{R}^N$. The first stage of the proposed framework involves a conventional upsampling, e.g., using an FIR filter based interpolation, to produce an initial approximation $\mathbf{u} \in \mathbb{R}^M$, where $M > N$. It can be expressed in vector notation as,

$$\mathbf{u} = \mathbf{U} \cdot \mathbf{s}, \tag{1}$$

where, the rows of the matrix $\mathbf{U}$ specify the filter coefficients used to generate the samples of $\mathbf{u}$. Then, an iterative procedure is followed, in which an estimated refinement signal $\mathbf{h}^{(i)}$, where i represents an iteration number, is added to the upsampled signal $\mathbf{u}$ to produce a refined signal $\mathbf{x}^{(i)}$, i.e.,

$$\mathbf{x}^{(i)} = \mathbf{u} + \mathbf{h}^{(i)}, \tag{2}$$

with $\mathbf{h}^{(0)}$ initialized to the zero vector. Next, a sparsity promoting denoising step operates on $\mathbf{x}^{(i)}$ to produce an approximation $\mathbf{a}^{(i)}$. It is realized via a forward transform, hard-thresholding and inverse transform. However, we assume that the low frequency components of $\mathbb{R}^M$ are faithfully upsampled by the filter in $\mathbf{U}$ and hence do not want to alter those components, but use the high frequency components of the denoised signal $\mathbf{a}^{(i)}$ to refine the upsampled picture. To this end, the low pass components of the signal $\mathbf{a}^{(i)}$ are determined by downsampling it to $\mathbb{R}^N$ and upsampling it back to $\mathbb{R}^M$. Denoting the downsampler as $\mathbf{D}$, the low pass components are computed as $\mathbf{U} \cdot \mathbf{D} \cdot \mathbf{a}^{(i)}$. The high frequency part $\mathbf{h}^{(i+1)}$ is determined by subtracting the low pass components, i.e.,

$$\mathbf{h}^{(i+1)} = \mathbf{a}^{(i)} - \mathbf{P} \cdot \mathbf{a}^{(i)}, \tag{3}$$

where, $\mathbf{P} = \mathbf{U} \cdot \mathbf{D}$. Ideally, using sinc filters in $\mathbf{U}$ and $\mathbf{D}$ results in $\mathbf{P}$ being an orthogonal projection as required in the convergence analysis in [7]. However, it is experimentally found that FIR filter approximations in $\mathbf{U}$ and $\mathbf{D}$ are sufficient for the purpose of high frequency extraction in the current setup.

The refinement procedure is repeated for a predefined number of iterations and the samples in $\mathbf{x}^{(i)}$ after the last stage form the output HR image.

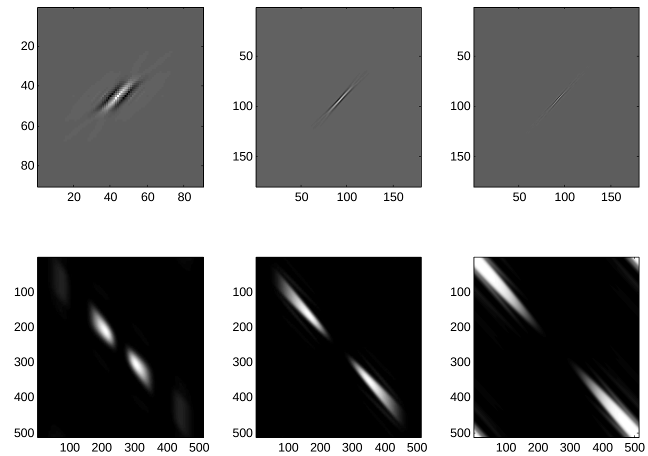


Fig. 2: *Top row: shearlet filters at a certain orientation for three scales. Bottom row: corresponding frequency responses.*

3. SHEARLET REPRESENTATION

There has been extensive study in constructing and implementing directional transforms aiming at obtaining sparse representations of piecewise smooth data. The curvelet transform is a directional transform which can be shown to provide optimally sparse approximations of piecewise smooth images [12]. However, curvelets offer limited localization in the spatial domain since they are band-limited. Contourlets are compactly supported directional elements constructed based on directional filter banks [13]. Directional selectivity in this approach is artificially imposed by a special sampling rule of filter banks which often causes artifacts. Moreover, there exists no theoretical guarantee for sparse approximations of piecewise smooth images.

Recently, a novel directional representation system, the so-called shearlets, has emerged which provides a unified treatment of continuous as well as discrete models, allowing optimally sparse representations of piecewise smooth images [10]. One of the distinctive features of shearlets is that the directional selectivity is achieved by shearing in place of rotation; this is, in fact, decisive for a clear link between the continuous and discrete world which stems from the fact that the shear matrix can preserve the integer lattice. Furthermore, shearlets offer a high degree of localization in the spatial domain since they can be compactly supported.

Shearlets are originally defined in the continuous domain and they are generated by dilating, shearing and translating a fixed generating function. One of the fundamental properties of shearlets is that they provide nearly optimal sparse approximation of functions that belongs to the class of cartoon-like images, a standard model for images with edges [12]. (see [10, 11] for more details). Furthermore, shearlets can be faithfully discretized, which leads to discrete shearlet filters. Fig. 2 illustrates discrete shearlet filters at a certain orientation and three different scales, along with their frequency responses.

Image name	NEDI [5]	Bicubic	Cubic spline	8-tap FIR	Contourlet [13]	Proposed
bikes	24.86	25.33	25.66	25.86	26.20	26.93
building2	21.64	22.56	22.79	22.99	23.22	23.47
buildings	22.01	22.51	22.69	22.88	23.08	23.39
caps	33.38	33.61	33.77	34.02	34.21	34.81
coinsinfountain	28.31	29.20	29.00	29.66	30.17	30.49
flowersonih35	21.41	22.39	22.52	22.78	23.08	23.30
house	29.37	29.94	30.20	30.33	30.45	30.84
lighthouse2	27.16	27.81	28.05	28.15	28.35	28.59
monarch	29.55	30.32	30.75	30.95	31.25	31.93
ocean	30.15	30.62	30.67	30.89	31.05	31.25
paintedhouse	26.78	27.39	27.62	27.76	27.92	28.40
parrots	32.58	32.89	33.30	33.51	33.51	34.19
plane	29.14	30.17	30.28	30.55	30.63	31.41
rapids	27.10	27.89	28.12	28.32	28.57	28.97
sailing1	26.34	27.17	27.36	27.48	27.63	27.83
Average	27.32	27.99	28.19	28.41	28.62	29.05
Diff. to proposed	**-1.73**	**-1.06**	**-0.86**	**-0.64**	**-0.43**	**–**

Table 1: *PSNR results in dB for 2x upsampling comparing six interpolation methods. Three linear approaches (bicubic, cubic spline & 8-tap FIR) and two non-linear approaches (NEDI [5] & contourlet [13]) are compared to the proposed technique. Average PSNR over 15 test images and the PSNR difference to the proposed approach are summarized.*

4. SIMULATION RESULTS

The proposed algorithm is tested for both objective and subjective performance. The subjective quality of the interpolated image is evaluated without downsampling the original images to avoid artifacts due to downsampling. For an objective evaluation, a high resolution reference image is required. To this end, the original is used as the reference HR image and a 11-tap anti-alias filter, that is used in ITU-T/MPEG evaluations of Scalable Video Coding [15], is employed to generate a LR image. The coefficients of the 11-tap filter for 2x downsampling are $[2, -2, -9, 3, 40, 60, 40, 3, -9, -2, 2]/128$.

In the first stage of the proposed framework, the input LR image is upsampled using an 8-tap FIR filter whose coefficients are f8 = $[-1, 4, -11, 40, 40, -11, 4, , -1]/64$ for the 2x interpolation case. The shearlet modeling is carried out at 4 different scales, with 8 directional filters for the first two scales and 16 directional filters for the last two scales. The stages of sparsity enforcement and high frequency extraction are repeated 8 times. The threshold value for hard-thresholding the shearlet coefficients is set to 100 and decreased by a factor of 0.7 in each iteration.

The performance of the proposed method is compared to various linear and non-linear methods. In linear methods, bicubic, cubic spline and f8 filter are considered. Among the non-linear models, the NEDI [5] technique and a contourlet based approach [13, 14] are considered. For a direct comparison of contourlet and shearlet dictionaries, the upsampling and downsampling filters in the proposed framework are kept fixed and the dictionaries are switched. The thresholds for the contourlet case are taken from [14].

The objective performance numbers in terms of peak signal to noise ratio (PSNR) are summarized in Tab. 1 for 15 commonly used test images. As can be seen, the proposed approach consistently achieves a higher PSNR result. On an average, a PSNR improvement of 0.64 dB is achieved compared to the 8-tap linear model for the considered test images. In order to further show the capability of the proposed approach, it is tested on a large set of 200 images from the Berkeley Segmentation Dataset [16]. Average PSNR improvements of 1.76 dB, 0.95 dB, 0.69 dB, 0.52 dB, 0.44 dB compared to NEDI, bicubic, cubic spline, f8 filter and contourlet, respectively, are observed.

Fig. 3 shows some input LR images (a, b, c) and output HR images produced using NEDI, cubic spline and the proposed technique. NEDI results (d, g, j) have some jaggedness for regions with strong edges and show some artifacts. The cubic spline results (e, h, k) do not have any strong artifacts but show blurring of edges. HR images produced using the propsosed approach (f, i, l) are sharper and do not exhibit noticeable artifacts.

5. CONCLUSION

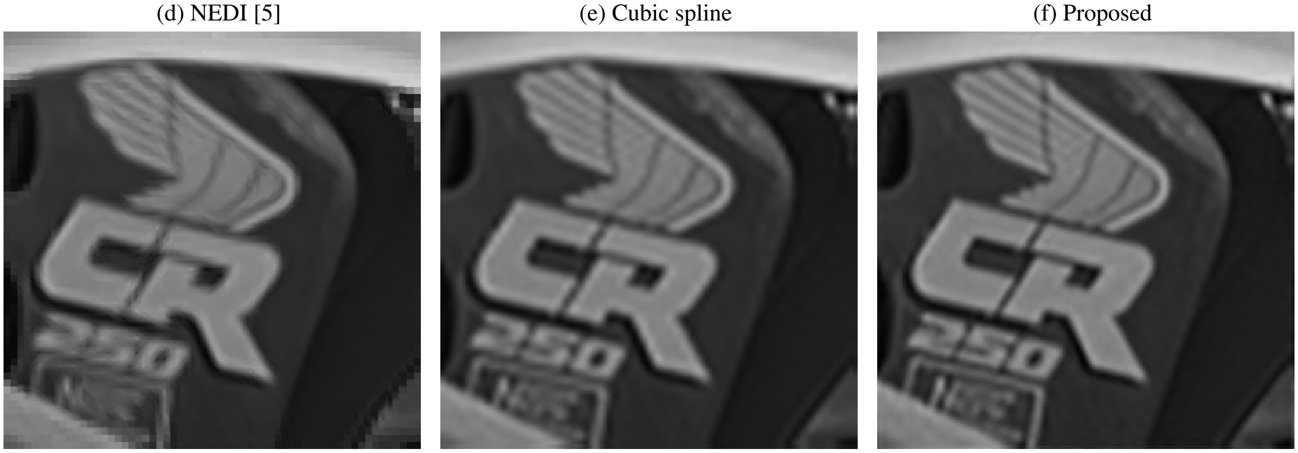
A framework for image interpolation that combines low frequencies from a linear method and high frequencies from a sparse representation is presented. The key idea is identify dominant structures through a sparse representation in a dictionary composed of shearlet atoms. Considerable objective and subjective gains are observed with the proposed method. Some important parameters that can be tuned for reducing the complexity are the number of iterations for refinement, number of scales and number of directions for subband filtering.

Acknowledgements

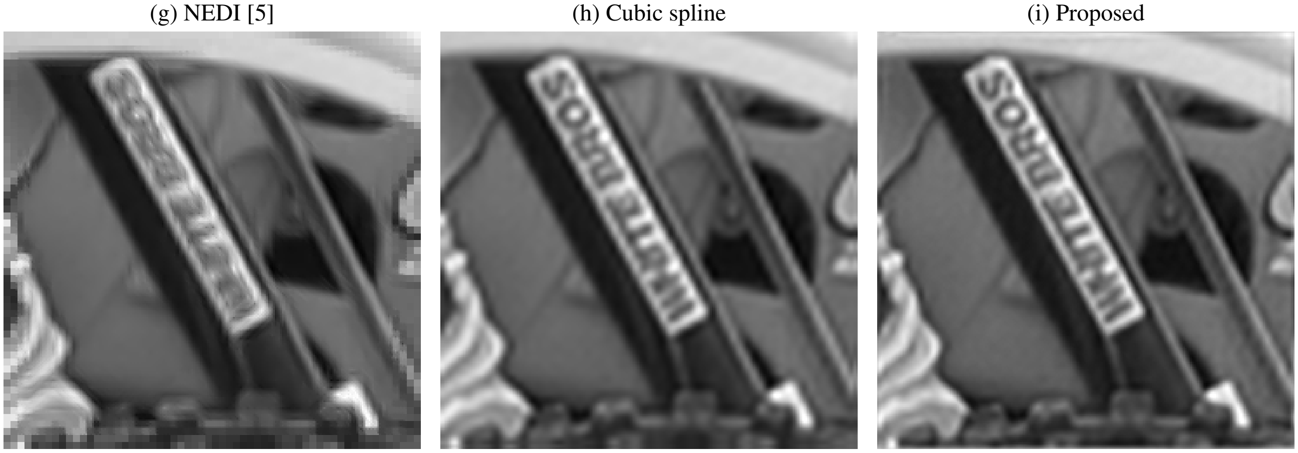
G. Kutyniok acknowledges support by the Einstein Foundation Berlin, by Deutsche Forschungsgemeinschaft (DFG) Grant SPP-1324 KU 1446/13 and DFG Grant KU 1446/14, by the DFG Collaborative Research Center TRR 109 “Discretization in Geometry and Dynamics”, and by the DFG Research Center Matheon “Mathematics for key technologies” in Berlin, Germany. W.-Q Lim acknowledges support by DFG Grant SPP-1324 KU 1446/13.

(a) LR1 (b) LR2 (c) LR3

(d) NEDI [5] (e) Cubic spline (f) Proposed



(g) NEDI [5] (h) Cubic spline (i) Proposed



(j) NEDI [5] (k) Cubic spline (l) Proposed

Fig. 3: *Example 4x upsampling results (via successive application of 2x upsampling). Input patches of size* 64×64 *are upsampled to* 256×256*. Left column results show jaggedness and other artifacts. Center column results are blurred. Right column results (proposed approach) appear slightly sharper without evident artifacts.*

6. REFERENCES

[1] C. E. Shannon, “Communication in the presence of noise,” *Proc. Institute of Radio Engineers*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[2] M. Unser, “Sampling - 50 Years After Shannon”, *Proc. IEEE*, vol. 88, no. 4, pp. 569-587, Apr. 2000.

[3] M. Unser, “Splines: A Perfect Fit for Signal and Image Processing”, *IEEE Signal Proc. Magazine*, vol. 16, no. 6, pp. 22-38, Nov. 1999.

[4] T. Blu, P. Thévenaz and M. Unser, “MOMS: Maximal-Order Interpolation of Minimal Support”, *IEEE Trans. on Image Proc.,* vol. 10, no. 7, Jul. 2001.

[5] X. Li and M. T. Orchard, “New edge-directed interpolation.” *IEEE Trans. on Image Proc.*, 10(10):1521–1527, 2001.

[6] M. Irani and S. Peleg, “Improving resolution by image registration,” *Graphical Models and Image Processing*, 53:231–239, 1991.

[7] O. G. Guleryuz, “Nonlinear Approximation Based Image Recovery Using Adaptive Sparse Reconstructions and Iterated Denoising: Part I - Theory,” *IEEE Trans. on Image Proc.*, March 2006.

[8] B. A. Olshausen and D. J. Field, “Natural image statistics and efficient coding,” *Network: Computation in Neural Systems*, no. 7, pp. 333–339, 1996.

[9] M. Aharon, M. Elad and A.M. Bruckstein, “The K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation,” *IEEE Trans. On Signal Proc.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[10] G. Kutyniok and W.-Q Lim “Compactly Supported Shearlets are Optimally Sparse”, *J. Approx. Theory*, Vol. 163, pp. 1564-1589, 2011.

[11] W.-Q Lim, “Nonseparable Shearlet Transform”, to appear in *IEEE Trans. on Image Proc.*, 2013.

[12] E. J. Candès and D. L. Donoho, “New tight frames of curvelets and optimal representations of objects with C2 singularities,” *Comm. Pure Appl. Math.*, vol. 57, no 2, pp. 219-266, 2004.

[13] M. N. Do and M. Vetterli, “The contourlet transform: an efficient directional multiresolution image representation,” *IEEE Trans. on Image Proc.*, vol. 14, no. 12, Dec. 2005.

[14] N. Mueller, Y. Lu, and M. N. Do, “Image interpolation using multiscale geometric representations,” *Proc. SPIE Conf. on Electronic Imaging*, San Jose, USA, 2007.

[15] J. Dong, Y. He, Y. Ye, “Downsampling filters for anchor generation for scalable extensions of HEVC,” *MPEG-M23485*, San Jose, Feb 2012.

[16] D. Martin and C. Fowlkes and D. Tal and J. Malik, “A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics,” *Proc. 8th Int'l Conf. Computer Vision*, vol. 2, pp. 416–423, July 2001. `http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/`.