Image interpolation using Shearlet based iterative refinement

H. Lakshman, W.-Q Lim, H. Schwarz, D. Marpe, G. Kutyniok, and T. Wiegand, *Fellow, IEEE*

***Abstract*—This paper proposes an image interpolation algorithm exploiting sparse representation for natural images. It involves three main 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 iterative thresholding, and (c) extracting high frequency information from the approximation to refine 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 0.8 dB is observed over a dataset of 200 images.**

***Index Terms*—Interpolation, Sparse representation, Shearlets.**

IMAGE interpolation refers to generating a high resolution (HR) image from an input low resolution (LR) image. The resolution of an image can be defined in various ways, e.g., based on:

- the number of pixels in the image,
- the characteristics of the physical sensing device in the camera,
- the effective sharpness as perceived by a human observer.

To quantify the resolution based on the first method is simple, but the latter two are considerably more complex.

Interpolation tasks have regained attention because images/videos are being viewed on displays of different sizes, like mobile phones, tablets, laptops, PCs, etc. For example, the content for a 1080p display may be available in a 720p format and needs to be interpolated. More recently, 4K displays are becoming popular and content with a lower resolution may have to be displayed on them. It also finds many applications in computer vision, graphics, compression, editing, surveillance and texture mapping. It is vital for image browsing and video playback software. Details synthesis in image interpolation can also be used as a tool in spatial scalable video coding.

Image interpolation, due to its interdisciplinary applications, is referred to using various terms, including image upsampling, upscaling, resizing, resampling, etc., depending on the community one comes from. Many established methods are available for achieving interpolation, e.g., FIR filtering, spline based schemes, etc. These techniques are sufficient for many practical purposes, but may cause several artifacts, most commonly, blurring of the resulting HR image. The main goal of this paper is to recover sharp edges and textures, while reducing blurring, ringing, aliasing or other visual artifacts in the resulting HR images. For videos, there is an additional requirement to maintain the temporal coherence to avoid picture-to-picture flickering during playback.

Efficient image representation is at the heart of image interpolation. Natural images occupy only a small fraction of the entire space of all possible images. Images show geometric structures, like edges, and conventional Fourier or DCT domains are not well suited for accurate modeling or extraction of geometric structures, although they are very useful in compression applications.

H. Lakshman, H. Schwarz, and D. Marpe, are with the Image & Video Coding group, Fraunhofer Institute for Telecommunications – Heinrich Hertz Institute (Fraunhofer HHI), 10587 Berlin, Germany (email: firstname.lastname@hhi.fraunhofer.de). W.-Q Lim and G. Kutyniok are affiliated with the Department of Mathematics, Technical University of Berlin, 10587 Berlin, Germany. T. Wiegand is jointly affiliated with the Image Processing Department, Fraunhofer HHI, and the Image Communication Chair, Technical University of Berlin, 10587 Berlin, Germany.

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

I. State-of-the-art

To review some important mathematical principles, a categorization of various methods is provided here.

Linear methods: Signal processing theory for band limited signals, advocates sampling higher than Nyquist rate and a sinc interpolation [38, 46]. 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 or bicubic interpolation are some popular examples that have very low computational complexity. Extending the sampling theory to shift-invariant spaces without band limiting constraints has led to a generalized interpolation framework, e.g., B-spline [45] and MOMS interpolation [5] 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 the basis functions in these linear models helps to capture higher order statistics but result in longer effective support in the spatial domain and hence produce artifacts like ringing around edges.

Directional methods: To improve the linear models, directional interpolation schemes have been proposed that perform interpolation along the edge directions instead of going across the edges. Some schemes in this class use edge detectors [2, 40]. The method in *New edge directed interpolation* (NEDI) [28] computes local covariances in the input image and uses them to adapt the interpolation at the higher resolution, so that the support of the interpolator is along the

edges. However, the resulting images still show some artifacts. The iterative back projection [23] technique improves image interpolation when the downsampling process is known. Its basic idea is that the reconstructed HR image from the LR image should produce the same observed LR image when it is passed through the same blurring and downsampling process. 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 downsampling process.

Sparsity based methods: Image interpolation can be seen as an estimation problem where the input data are inadequate. Naturally, the solution to this problem is not unique due to the lack of information in the HR grid. A popular idea used in such underdetermined problems is to exploit the structure of the desired solution. For images, sparsity in transform domains has proven itself to be a very useful prior [14, 35, 36]. Sparse approximation can be viewed as approximating a signal with only a few expansion coefficients [37]. Sparsity priors have also been proposed for image interpolation, e.g., in [32, 33, 47]. The method in [33] uses a contourlet transform for sparse approximation and is designed for an observation model that assumes that the LR image is the low pass subband of a wavelet transform. It uses the same transform in a recovery framework, so it relies directly on knowledge of the downsampling process. We follow a similar recovery principle, but design a system so that it works for typical anti-aliased LR images instead of requiring a specific wavelet transform. The method in [47] involves jointly training two dictionaries for the low- and high-resolution image patches. The set of all elements that can be used in the expansion is called a dictionary. It then performs a sparsity based recovery, but involves high search complexity to determine a sparse approximation in the trained dictionary (observed to be more than 100x slower than [33]). The approach in [32] considers the case when the LR image produced by sub-sampling a HR image is aliased. The method in [9] learns a series of compact sub-dictionaries and assigns adaptively a sub-dictionary to each local patch as the sparse domain. The K-SVD algorithm proposed in [1] and its extensions are commonly used for learning an overcomplete dictionary. These methods depend on the similarity of training and test patches and number of the selected examples, which are typical issues in learning-based algorithms. Furthermore, analytically determined transforms have structures that can be exploited to produce a fast implementation, which might be hard to impose during dictionary learning.

Discussion of the proposed method: We recognize the fact that linear models such as interpolation based on FIR filters are faithful in interpolating the low frequency components but distort the high frequency components in the upsampled image. An iterative framework, based on [20, 33], is proposed that combines the output from an initial interpolator and detail components from a denoised approximation. The method used here for denoising is the so-called shrinkage or thresholding approach, i.e., by transforming the signal to a specific domain, setting the transform coefficients below a certain (absolute) value to zero and inverse transforming the coefficients to get back an approximation. The domain used for transforming is chosen so that the coefficients with large absolute values capture most of the geometric features and the coefficients with low absolute values constitute noise or finer details. To this end, multi-resolution transforms or multi-resolution directional transforms are preferred. The concepts of multi-resolution and directionality in transforms are reviewed in Sec. III, based on which a framework for details synthesis in interpolation is proposed in Sec. IV. In fact, wavelet domain thresholding has been successfully applied to many denoising problems [11, 12]. Due to the subsampling in orthogonal wavelet transforms, they are not translation invariant. But, unlike a typical compression scenario, the number of transform coefficients generated during modeling or denoising need not be the same as the number of input samples. This is exploited by removing the sub-sampling in the wavelet transform and is shown to yield better denoising results [13, 15]. Super-resolution methods that use a sequence of images can further improve the quality. However, these methods are beyond the scope of this paper and only single frame interpolation is considered.

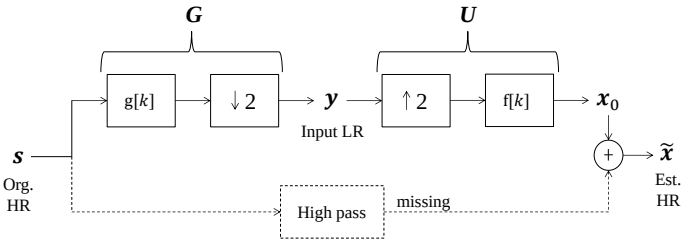


Fig. 1: Image recovery problem formulation. Notation: unknown original HR signal $\boldsymbol{s}$; given input LR signal $\boldsymbol{y}$; estimated output HR signal $\tilde{\boldsymbol{x}}$.

II. Interpolation problem formulation

We consider a setup in which the input LR image to be interpolated has been produced from an original HR image through anti-aliasing and decimation. This way, the LR image does not have evident visual artifacts, but does have a loss of information. For instance, the anti-alias low pass filter can be an optical filter in a camera or a digital filter in an image processing pipeline.

Let the (unknown) HR original signal of dimensions $N \times 1$ be denoted as $\boldsymbol{s}$. Let the (unknown) low pass filter $\mathrm{g}[k]$ followed by a decimation together be represented as a downsampling matrix $\boldsymbol{G}$ of dimension $n \times N$, where $n < N$. We are given the result $\boldsymbol{y}$ of dimension $n \times 1$ as the LR input to the interpolation system, as depicted in Fig. 1.

One way to estimate an HR signal $\tilde{\boldsymbol{x}}$ is by solving an optimization problem of the form

$$\min_{\tilde{\boldsymbol{x}}} \; D(\tilde{\boldsymbol{x}}) + \lambda \cdot R(\tilde{\boldsymbol{x}}). \tag{1}$$

where $D(\tilde{\boldsymbol{x}})$ is a fidelity term that penalizes the difference between the given LR signal $\boldsymbol{y}$ and the LR signal obtained by downsampling the estimated HR signal $\tilde{\boldsymbol{x}}$ using the

downsampler $\boldsymbol{G}$, while $R(\tilde{\boldsymbol{x}})$ is a regularizer that promotes sparsity of the estimated HR signal in a transform domain and λ is a regularization parameter. Typically, the fidelity term is chosen as an L2 norm, i.e., $D(\tilde{\boldsymbol{x}}) = \|\boldsymbol{G} \cdot \tilde{\boldsymbol{x}} - \boldsymbol{y}\|^2$, which requires the explicit knowledge of $\boldsymbol{G}$. If we need to find the sparsest solution, we need to choose the penalty function $R(\tilde{\boldsymbol{x}})$ as the L0 (pseudo) norm of the transform coefficients which is unfortunately an NP-hard problem [34]. If the penalty function $R(\tilde{\boldsymbol{x}})$ is chosen to be the L1 norm of the transform coefficients, it has been shown that it has the effect of promoting sparsity in the transform domain under certain conditions [11]. It then becomes a convex optimization problem and can be solved using general convex solvers, e.g., using interior point methods [4, 6]. However, there are simpler gradient-based algorithms for solving functions of this form and a popular method is called *iterative shrinkage/thresholding algorithm* (ISTA) [8, 18, 48]. It is also known by other names in signal processing literature, e.g., thresholded Landweber method, basis pursuit denoising [16], etc. Optimizing objective functions of this form is an active area of research and many fast algorithms, e.g., [3], are being proposed in literature. Other popular approaches include greedy techniques such as matching pursuits and orthogonal matching pursuits [31, 44].

The proposed framework follows the principle of image recovery through sparse reconstructions and iterated denoising [20, 21]. This procedure has similarities to ISTA techniques and offers some robustness to noise and transform selection. While atomic decomposition techniques (L1, greedy, etc) build a solution bottom-up, iterated denoising takes a top-down approach, starting from an initial point and pruning the signal components that it detects as noise. A detailed comparison of iterated denoising versus atomic decomposition methods for missing data estimation can be found in [19].

III. Multi-resolution directional transforms

One of the main goals of a transform representation is to determine efficient linear expansions for images. Efficiency is generally measured in terms of the number of elements needed in a linear expansion. To quantify the number of elements needed for a linear expansion, image models are employed. Commonly, images are considered as uniform 2D functions separated by singularities (e.g., edges). The singularities themselves are modeled as smooth curves. In the past decades, developments in applied harmonic analysis have provided many useful tools for signal processing. Wavelets are good at isolating singularities in 1D. Extending wavelets to 2D, makes them well adapted to capture point-singularities. But in natural images, there are mostly line- or curved- singularities (e.g., directional edges). These are also known as anisotropic features as they are dominant along certain directions. To capture such features, there has been extensive study in constructing and implementing directional transforms aiming to obtain sparse representations of such piecewise smooth data. The curvelet transform is a directional transform which can be shown to provide optimally sparse approximations of piecewise smooth images [7]. However, curvelets offer limited localization in the spatial domain since they are band limited. Also, they are based on rotations which introduce difficulties in achieving a consistent discrete implementation. Contourlets are compactly supported directional elements constructed based on directional filter banks [17]. Directional selectivity in this approach is artificially imposed by a special sampling rule of filter banks which often causes artifacts. Moreover, no theoretical guarantee exists for sparse approximations for piecewise smooth images. Recently, a novel directional representation system known as shearlets has emerged, which provides a unified treatment of continuous as well as discrete models, allowing optimally sparse representations of piecewise smooth images [25, 29]. This simplified model of natural images, which emphasizes anisotropic features, most notably edges, is found to be consistent with many models of the human visual system [26]. The framework proposed in this paper is applicable for all these transforms, although shearlets is observed to provide the best performance among the considered transforms.

Multi-resolution directional transforms can also be seen as filterbanks. The decomposition is implemented using an analysis filter bank, while the reconstruction is implemented using a synthesis filter bank. One branch of the filterbank is designed as a low pass channel that captures a coarse representation of the input signal followed by band- or high-pass channels. Each of these branches is adapted to capture signal components at different scales and directions.

Introduction to shearlets

In modeling image features that are typically anisotropic, other than the location and scale, we would like to include the orientations of the features. Therefore, a transform is built by combining a scaling operator to generate elements at different scales, an orthogonal operator to change their orientations, and a translation operator to displace these elements over the 2D plane [26]. Consider a general model for directional transforms built from a generating function $\psi(t)$ by orienting it using $\boldsymbol{O}_s$, scaling it using $\boldsymbol{A}_a$, and translating it using $\boldsymbol{T}_m$, so that

$$S(\psi) = \boldsymbol{T}_m \cdot \boldsymbol{A}_a \cdot \boldsymbol{O}_s \cdot \psi. \tag{2}$$

Below, we discuss the choice of these three operators that leads to the so-called shearlet system $S(\psi)$.

Firstly, to change the orientation of the generating function ψ, an obvious choice is a rotation operator. However, rotations destroy the integer lattice (except for trivial rotations that switch the axes). In other words, integer locations may get mapped to fractional locations after a rotation. This leads to the problem of obtaining a discrete transform that is consistent with the continuous transform (where approximation properties have been optimized). As an alternative orientation operator, consider the *shearing* matrix

$$\boldsymbol{O}_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}. \tag{3}$$

This achieves orientation changes using the slope s rather than a rotation angle. It has the advantage of leaving the integer lattice invariant when s is chosen as an integer.

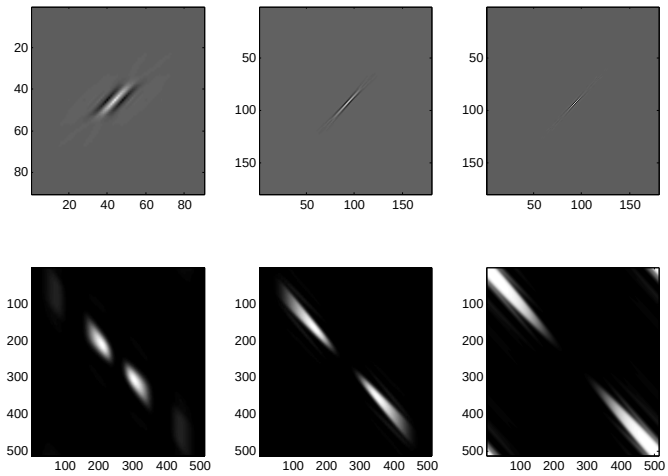


Fig. 2: Example of shearlet elements for three scales generated using [39] (top row: spatial domain, bottom row: frequency domain). They are directional and band pass in nature with increasing center frequencies from left to right.

Next, the scaling operator is considered. Equal scaling along both axes will not be able to capture anisotropic features, hence different scaling for the axes is required. Consider the case when one axis is scaled by the factor a and the other by $a^{1/2}$, so that

$$\boldsymbol{A}_a = \begin{bmatrix} a & 0 \\ 0 & a^{1/2} \end{bmatrix}. \tag{4}$$

Although other ratios for scaling the axes are possible, this choice, known as parabolic scaling, optimizes the approximation properties for the piecewise smooth image model considered.

Finally, a translation operator is defined that shifts the generating function

$$\boldsymbol{T}_m \, \psi(t) \rightarrow \psi(t-m). \tag{5}$$

The conditions on the generating function ψ so that the shearlet system $S(\psi)$ can represent any square-integrable function are known as admissibility conditions [26].

Directional elements capture high frequencies along certain directions and are not good at representing the low frequencies. Therefore, in general, a low pass filter is used to extract the low frequency part and the directional elements are operated on the remaining signal, leading to the so-called cone-adapted shearlet transform. By varying the parameters of the shearlet system, different properties can be achieved, e.g., compact support [24], orthonormality [26], etc. However, a shearlet system with compact support that is also orthonormal is, most likely, not achievable [22]. Nevertheless, compactly supported shearlet systems have good frame properties, i.e., they are close to being a tight frame.

Fig. 2 shows examples of practical filters (shearlet) at a certain orientation and three different scales.

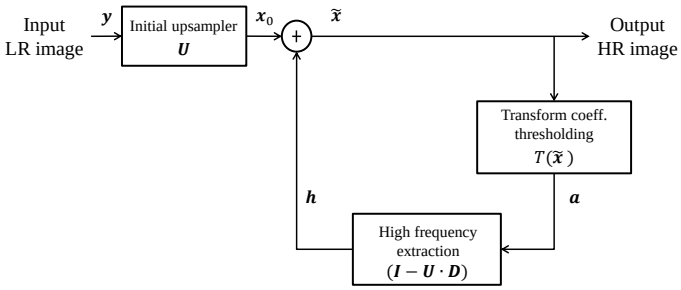


Fig. 3: Framework for image interpolation. A linear model, e.g., FIR filter is used to produce an initial upsampled image. Then, high frequency components are extracted from a sparse approximation and used to refine the initial upsampled image.

IV. Proposed Framework for High Frequency Synthesis

The proposed framework, depicted in Fig. 3, uses the iterated denoising principle. It involves:

- **Sparsity constraint:** promoting sparsity, e.g., in a multi-resolution directional transform domain to improve regularity along edges, and,
- **Data constraint:** enforcing constraints according to known data.

The problem considered in [20] is that of filling missing samples in an image, where enforcing known data constraints is achieved by replacing input samples at the known locations after the sparsity promoting step. However, in the context of image interpolation, the available LR input image constitutes the known data. The iterated denoising principle has been applied to image interpolation using contourlets in [33], however, utilizing the knowledge of the LR image generation during the HR image estimation. Specifically, the LR image was produced through the low pass subband of a specific wavelet transform and the same transform was used during the HR image estimation to enforce the known data constraint. It is a goal of the proposed approach to interpolate a given LR image without the knowledge of the exact method generating the LR image. Therefore, the iterative procedure is redesigned so that the input LR image can be used as the known data constraint, instead of requiring the low pass subband of a specific wavelet transform.

Initial upsampling: The first stage of the proposed framework involves a conventional FIR filter based interpolation of the LR signal $\boldsymbol{y} \in \mathbb{R}^n$ to produce an initial HR estimate $\boldsymbol{x}_0 \in \mathbb{R}^N$. It can be expressed in a vector notation as

$$\boldsymbol{x}_0 = \boldsymbol{U} \cdot \boldsymbol{y}, \tag{6}$$

where the rows of the upsampler $\boldsymbol{U}$ specify the filter coefficients used to generate the samples of $\boldsymbol{x}_0$. This process can also be seen as a zero insertion in the spatial domain followed by a low pass filter to remove the spectral replication due to the zero insertion. Since the coefficients in $\boldsymbol{U}$ act as a low pass filter, some high pass details would be missing/distorted in the initial HR estimate compared to the HR original. Therefore, the initial HR estimate is seen as a noisy version of an

unknown HR original and then refined in an iterative manner. The refined HR signal is denoted as $\tilde{\boldsymbol{x}}$, which, during the first iteration, is set as $\tilde{\boldsymbol{x}}_1 = \boldsymbol{x}_0$.

Sparsity promoting: As stated earlier, a dictionary consisting of multi-resolution directional transform elements is considered. Promoting sparsity in such a dictionary results in regular directional structures in the approximated signal. Denoting the iteration number of refinement as k, the sparsity promoting step operates as follows:

- the signal $\tilde{\boldsymbol{x}}_k$ is forward transformed to the selected domain (resulting in directional components in different scales),
- the transform coefficients are hard-thresholded, and
- inverse transformed to generate an approximation $\boldsymbol{a}_k$.

The overall operation is written compactly as, $\boldsymbol{a}_k = T(\tilde{\boldsymbol{x}}_k)$. This denoising step is closely related to techniques such as ISTA for L1 regularization but has some differences [19].

Known data constraint: Then, we enforce the known LR data constraint. It is done by assuming that the initial upsampled signal $\boldsymbol{x}_0$ is equal to the low pass channel of a two-channel filterbank, depicted in Fig. 1. The missing high pass channel is generated by using the approximated signal $\boldsymbol{a}_k$. Hence, it is required to separate the signal $\boldsymbol{a}_k$ into low pass and high pass channels. At this stage we face the issue of the unknown downsampler that generated the input LR signal $\boldsymbol{y}$. A blind deconvolution would be necessary to jointly estimate the unknown downsampler and undo its effect, which is very difficult. Instead, a downsampler $\boldsymbol{D}$ is chosen so that the product $\boldsymbol{P} = \boldsymbol{U} \cdot \boldsymbol{D}$ acts as a projection matrix, i.e., $\boldsymbol{P}^2 = \boldsymbol{P}$. Then, enforcing the known data constraint can be implemented by only considering the components of $\boldsymbol{a}_k$ that do not fall on the low pass projection space, i.e., using the high pass components of $\boldsymbol{a}_k$ for refinement. However, there could be a mismatch between the utilized $\boldsymbol{D}$ and the actual external downsampler that produced the LR signal. This will be experimentally studied in Sec. V-D by fixing the upsampler and downsampler of the proposed system, but varying the actual external downsampler to produce different LR inputs to the proposed system and recording the performance variation.

Summarizing, we can write the low pass $\boldsymbol{l}_k$ and the high pass $\boldsymbol{h}_k$ decomposition of the approximated signal $\boldsymbol{a}_k$ as

$$\begin{aligned} \boldsymbol{l}_k &= \boldsymbol{U} \cdot \boldsymbol{D} \cdot \boldsymbol{a}_k, \\ \boldsymbol{h}_k &= (\boldsymbol{I} - \boldsymbol{U} \cdot \boldsymbol{D}) \cdot \boldsymbol{a}_k. \end{aligned} \tag{7}$$

Refinement step: The high pass component $\boldsymbol{h}_k$ is used for refinement by adding it to the initial HR estimate $\boldsymbol{x}_0$, to produce a refined HR estimate $\tilde{\boldsymbol{x}}_{k+1}$, i.e.,

$$\tilde{\boldsymbol{x}}_{k+1} \longleftarrow \boldsymbol{x}_0 + \boldsymbol{h}_k. \tag{8}$$

For the first iteration, the vector $\boldsymbol{h}_0$ is initialized to zero, therefore, $\tilde{\boldsymbol{x}}_1 = \boldsymbol{x}_0$.

By combining Eq. 6 through Eq. 8, the overall system connecting the input LR signal $\boldsymbol{y} \in \mathbb{R}^n$ to the refined HR signal $\tilde{\boldsymbol{x}}_{k+1} \in \mathbb{R}^N$ can now be expressed as

$$\tilde{\boldsymbol{x}}_{k+1} \longleftarrow \boldsymbol{U} \cdot \boldsymbol{y} + (\boldsymbol{I} - \boldsymbol{U} \cdot \boldsymbol{D}) \cdot T(\tilde{\boldsymbol{x}}_k). \tag{9}$$

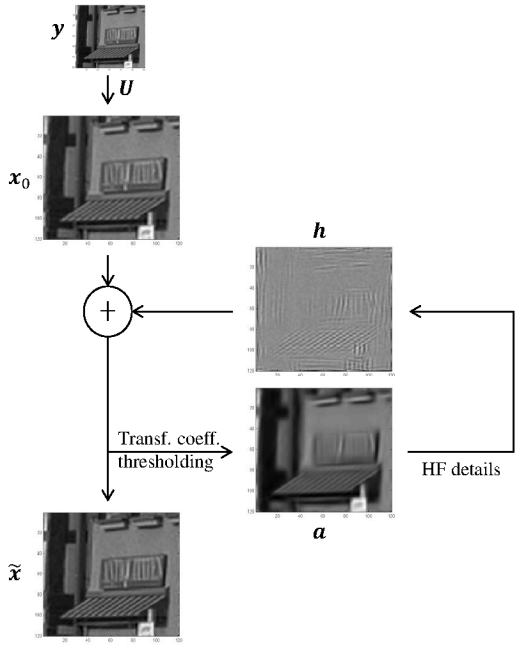


Fig. 4: Example images at each stage of processing. Figure shows the quality of the initial HR estimate, the result of transform domain thresholding and the estimated high pass details. Notice that the diagonal lines become slightly sharper after adding the estimated details.

The iterative procedure is repeated for a certain maximum number of iterations and $\tilde{\boldsymbol{x}}_{k+1}$ after the last iteration is taken as the output HR image.

Fig. 4 depicts example images during different stages of the proposed approach. It can be seen that the initial upsampled image is blurry around the diagonal edges. The step of transform domain thresholding retains only the dominant information. After adding the high frequency part, the resulting image looks slightly sharper.

V. Simulation Results

The proposed algorithm is tested for both subjective and objective performance. For a subjective evaluation, original images are directly used as LR inputs and the HR outputs are inspected for visual quality/artifacts. Using original images as LR inputs avoids downsampling artifacts in inputs. However, for an objective evaluation, we require a reference HR image. To this end, a 11-tap FIR anti-alias filter, that is tested in the ITU-T/ISO-IEC evaluations of Scalable Video Coding [10], is used before decimation to generate an LR image and the original image is used as the reference HR image to measure the PSNR. The coefficients of the 11-tap filter for 2x downsampling are $[2, -2, -9, 3, 40, 60, 40, 3, -9, -2, 2]/128$. In all the experiments, this filter remains unknown to the proposed interpolation system. Additionally, in Sec. V-D, the proposed system is kept fixed and the external downsamplers are varied to record the performance variation.

There are many free parameters to be chosen in the proposed method, such as the initial upsampling filter, number of scales and directions in the transform, thresholds levels for hard

Symbol	Interpolation filter coefficients
u2	$[1, 1]/2$
u4	$[-1, 9, 9, -1]/16$
u6	$[1, -5, 20, 20, -5, 1]/32$
u8	$[-1, 4, -11, 40, 40, -11, 4, -1]/64$
u12	$[-1, 4, -10, 22, -48, 161, 161, -48, 22, -10, 4, -1]/256$

TABLE I: Set of FIR filters considered for initial interpolation.

Symbol	N-tap	Anti-aliasing filter coefficients
d3	3	$[1, 2, \cdots]/4$
d9	9	$[-1, 0, 9, 16, \cdots]/32$
d13	13	$[1, 0, -5, 0, 20, 32, \cdots]/64$
d17	17	$[-1, 0, 4, 0, -11, 40, 64, \cdots]/64$
d25	25	$[-1, 0, 4, 0, -10, 0, 22, 0, -48, 0, 161, 256, \cdots]/256$

TABLE II: Set of FIR filters considered for anti-aliasing in high frequency extraction. Dots denote repetition of coefficients with mirror symmetry.

thresholding in the transform domain, etc. A joint optimization of all these internal parameters involves a large search space. Hence, a simpler approach is followed here, where we first select an initial set of parameters and optimize some free parameters keeping the others fixed, for 2x upsampling. The optimization of free parameters is conducted using a training set (16 images) and the final performance is evaluated on a test set (200 images). The training and test sets are disjoint. Throughout the optimization, the proposed method with the chosen parameter set is compared to a system with an 8-tap FIR filter without any iterative refinement to record the average PSNR gain in the training set. Although a 12-tap filter provides a higher PSNR, it is not preferred as a reference, since some ringing artifacts can be noticed in the 12-tap filter results.

Initial upsampler and Downsampler for high frequency extraction: In the first stage of the proposed framework, the input LR image is upsampled using $\boldsymbol{U}$. The rows of $\boldsymbol{U}$ are filled with FIR filter coefficients so that the samples in the HR grid corresponding to zero phase shift in the LR grid are copied directly and the required fractional shifts are produced using FIR filters. To this end, for 2x upsampling, five different filters are considered which are given in Tab. I.

Next, a downsampler $\boldsymbol{D}$ is designed to enforce the known data constraint. Ideally, a sinc filter for $\boldsymbol{U}$ and $\boldsymbol{D}$ results in $\boldsymbol{P} = \boldsymbol{U} \cdot \boldsymbol{D}$ being a projection operator [42]. However, it will be shown in Sec. V-A that FIR filter approximations in $\boldsymbol{U}$ and $\boldsymbol{D}$ are sufficient for the purpose of high frequency extraction in the current setup. To this end, five different anti-alias filters are evaluated for 2x downsampling, given in Tab. II. All the considered filters are odd-length and symmetric, hence they do not induce any phase shift.

Directional transform parameters: A compactly supported shearlet transform [29, 39] is chosen for the multi-resolution directional representation. The initial configuration used for the shearlet transform is: 1 low pass component, 2^3 directional band pass components and 2^3 directional high pass components. These settings can be compactly represented in an array as $[0, 3, 3]$, where the entries of the array are interpreted as exponents of two. The number of entries in the array denotes the number of scales used. For instance, $[0, a]$ represents a configuration consisting of two scales: one low pass component and 2^a directional high pass components. The configuration $[0, a, b]$ represents three scales: one low pass component, 2^a directional band pass components, and 2^b directional high pass components. The computation of shearlet transform coefficients and the reconstruction are carried out as multiplications in the Fourier domain instead of convolutions in the spatial domain to reduce the computational complexity. 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 multiplicative factor of 0.6 in each iteration. The proposed framework is also tested with the contourlet transform. For a direct comparison of the contourlet and shearlet dictionaries, the upsampling and downsampling filters in the proposed framework are kept fixed and only the dictionaries are switched. The threshold values for the contourlet case are taken from [33].

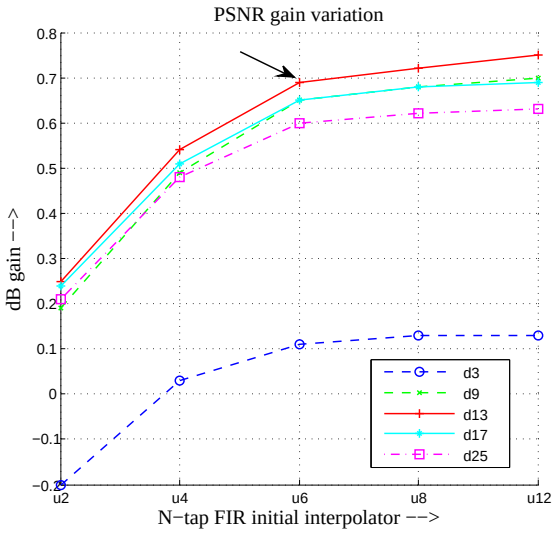


Fig. 5: Influence of initial upsampler; and downsampler for HF extraction. Average PSNR difference (dB) between reference (8-tap FIR interpolator) and test (proposed refinement approach with different combinations of $\boldsymbol{U}$ and $\boldsymbol{D}$) for a dataset of 16 training images. PSNR improvements for initial interpolation filters beyond 6-tap are rather small.

A. *Influence of initial interpolator & high frequency extractor*

The influence of $\boldsymbol{U}$ and $\boldsymbol{D}$ on the final HR result is studied here. To this end, each interpolation filter from the set {u2, u4, ..., u12} is combined with a downsampling filter from the set {d3, d9, ..., d25} and 25 HR results are produced for each LR input, i.e., the entire product space is tested. Fig. 5 shows the test results for each tested parameter combination, in the form of average PSNR difference to the 8-tap FIR (u8) reference system. In the y-axis, the 0 dB gain level represents a PSNR that is the same as the reference system. It can be seen that the 3-tap anti-alias filter d3 is not well suited for the system, because it leaves too much aliasing. The remaining anti-alias filters from the set perform relatively well. The best PSNR

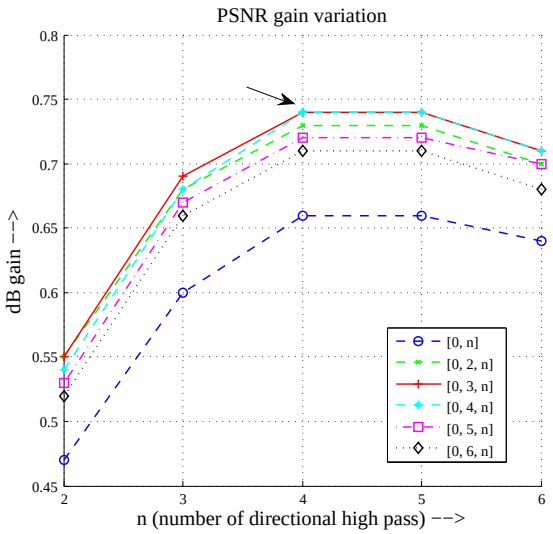


Fig. 6: Influence of the number of scales and directions. Each tested configuration is represented in the array notation introduced. The configuration [0, 3, 4], i.e., splitting the signal into one low pass, 2^3 directional band pass and 2^4 directional high pass components, is observed to give the best results.

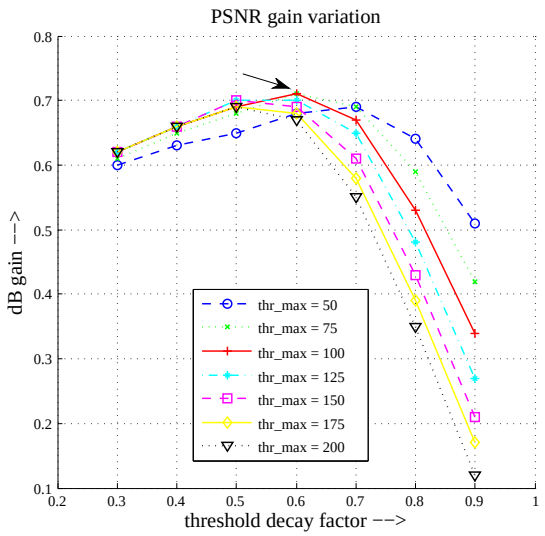


Fig. 7: Threshold selection experiments. The threshold for the first refinement iteration is denoted as thr_max and decreased exponentially in each iteration by a decay factor (x-axis). The PSNR gain compared to an 8-tap FIR interpolator is recorded (y-axis). The system shows best gains for thr_max between 75 to 125 with a decay factor around 0.5 to 0.6.

performance is observed when the 13-tap anti-alias filter is combined with a 12-tap interpolator, giving 0.75 dB gain over the reference 8-tap FIR interpolator. However, PSNR improvements for interpolation filters beyond 6-tap are rather small and the 12-tap interpolation filter might introduce ringing artifacts in the initial upsampled image. Therefore, the combination of the 6-tap interpolation filter and the 13-tap downsampling filter is chosen for further investigation.

B. Selection of the number of scales and directions in transform

Next, the influence of the number of scales and directions for thresholding the estimated HR image in the transform domain is studied. The tested configurations are compactly represented in the same array format described earlier. PSNR results using the proposed system in the tested configurations are compared to the reference 8-tap FIR (u8) system and the observed average gains are shown in Fig. 6. It can be seen that the configuration $[0, 3, 4]$, i.e., one low pass, 8 directional band pass and 16 directional high pass components, provides the best performance among the tested transforms (0.74 dB improvement over reference).

In fact, for a 2x upsampling, we expect that only around half the frequency components need refinement, for which, using two scales should be sufficient. However, it can be seen from Fig. 6 that the three scale configurations, namely, $[0, 2, \mathrm{n}], [0, 3, \mathrm{n}], \cdots, [0, 6, \mathrm{n}]$, perform better than the two scale configuration $[0, \mathrm{n}]$. It suggests that an intermediate scale provides a soft transition from low- to high- frequency components for refinement. In other experiments (not shown in figure), it is observed that using more than three scales for 2x upsampling does not increase the gain further.

C. Threshold selection for sparse approximation

The effect of thresholding in the shearlet domain on the final interpolation quality is hard to express analytically. To this end, two parameters for heuristic optimization are identified: (a) threshold for the first iteration of refinement, denoted as thr_max, and (b) a multiplicative decay factor to decrease the threshold in each iteration. The maximum number of iterations is set as 8 to limit the overall computational complexity. For instance, thr_max = 200 and decay = 0.7 generates the following thresholds: $\{200, 200 \times 0.7, 200 \times 0.7^2, \cdots, 200 \times 0.7^7\}$. The low pass components of the shearlet transform are not thresholded and the same threshold value is used for the remaining components, although a band wise optimization of thresholds may further improve the performance. PSNR results of the proposed method with chosen parameters are compared to the reference 8-tap FIR (u8) system and the PSNR gain is computed. Average PSNR gains on the training set is plotted in Fig. 7. It can be observed that thr_max = 75, 100 and 125 perform well with a decay factor of 0.5 or 0.6. The combination of thr_max = 100 and decay = 0.6, which is the same as our initial setting, is selected for the final evaluation on the test set.

D. Influence of external downsamplers to generate LR images

With the system parameters fixed, the influence of the external downsampling filter used to generate an LR input from the HR original is studied in this experiment. To this end, six different downsampling filters (approximately halfband cut-off) are used and six LR images are generated for each HR original. The test is conducted such that the proposed method remains fixed and is unaware of the actual external downsampler that has been used to generate the LR input. As

External downsampler to produce LR input	Proposed vs. 8-tap FIR
$[-1, 0, 9, 16, 9, 0, -1]/32$	0.66 dB
$[-2, 0, 64, 132, 64, 0, -2]/256$	0.58 dB
$[1, 0, -5, 0, 20, 32, 20, 0, -5, 0, 1]/64$	0.67 dB
$[1, 0, -11, 0, 74, 128, 74, 0, -11, 0, 1]/256$	0.66 dB
$[-1, 0, 4, 0, -17, 0, 78, 128, 78, 0, -17, 0, 4, 0, -1]/256$	0.66 dB
$[1, 0, -2, 0, 7, 0, -21, 0, 79, 128, 79, 0, -21, 0, 7, 0, -2, 0, 1]/256$	0.60 dB

TABLE III: Influence of using different downsampling filters to generate LR images. For each HR image, six different LR images are generated using 2x downsampling filters given in the first column. It can be seen that the proposed method achieves stable results and the external downsampling filter does not greatly influence the gains.

Image name	Bicubic	Directional	Cubic spline	8-tap	12-tap	Contourlet	Shearlet
bikes	26.68	26.20	27.02	27.23	27.32	27.63	28.38
building2	23.83	22.89	24.08	24.28	24.34	24.58	24.84
buildings	23.85	23.32	24.06	24.23	24.29	24.51	24.78
caps	35.60	35.38	35.78	36.06	36.13	36.33	37.03
coinsinfountain	30.56	29.60	30.44	31.08	31.16	31.62	32.08
flowersonih35	23.74	22.76	23.87	24.13	24.19	24.47	24.71
house	31.09	30.62	31.38	31.52	31.60	31.73	32.14
lighthouse2	29.19	28.55	29.44	29.55	29.61	29.78	30.07
monarch	31.87	31.04	32.37	32.59	32.71	33.03	33.85
ocean	32.17	31.70	32.23	32.47	32.52	32.62	32.93
paintedhouse	28.23	27.64	28.50	28.65	28.71	28.90	29.35
parrots	34.82	34.39	35.36	35.59	35.70	35.88	36.59
plane	31.47	30.32	31.59	31.86	31.92	32.30	32.78
rapids	29.42	28.73	29.67	29.91	29.98	30.18	30.66
sailing1	28.60	27.77	28.81	28.92	28.97	29.14	29.34
stream	24.73	24.03	24.93	25.08	25.14	25.29	25.50
Average (Train)	29.12	28.43	29.35	29.57	29.64	29.87	30.32
PSNR diff. (Train)	-1.20	-1.88	-0.97	-0.74	-0.67	-0.44	-
PSNR diff. (Test)	**-1.09**	**-1.86**	**-0.81**	**-0.63**	**-0.56**	**-0.47**	**-**

TABLE IV: PSNR results in dB for 2x interpolation comparing seven methods. Three linear approaches (bicubic, cubic spline, and 8-tap FIR) and two non-linear approaches (Directional [28] and contourlet [17]) are compared to the proposed technique. The PSNR difference over 16 training and 200 test images are summarized.

a reference, the 8-tap FIR filter (u8) is used to interpolate the same LR image and the resulting PSNR is measured. Then, the PSNR difference to the reference result is recorded. The average PSNR gain on the training set is summarized in Tab. III. It can be seen from the result that the gains from the proposed technique do not vary much when changing the downsampling filters, as long as there is not much aliasing in the generated LR images.

E. Final results on training and test set

The performance of the proposed method is compared to various linear and non-linear methods. Among linear methods: bicubic interpolation (u4), 8-tap filter (u8), 12-tap filter (u12) and cubic spline interpolation are considered. The cubic spline approach is implemented as an IIR prefilter to compute spline coefficients followed by a 4-tap FIR filter for interpolation. Among the non-linear models, a directional interpolation (NEDI [28]) technique is considered. The proposed framework is tested with contourlet and shearlet transforms. The parameters for the contourlet case are taken from [33].

The objective performance numbers of the overall system with the selected parameter settings are summarized in Tab. IV for the training and test set. As can be seen, the proposed approach consistently achieves a higher PSNR compared to the other methods tested. On an average, a PSNR improvement of 0.74 dB is achieved compared to the 8-tap filter for the considered training images. As a test set, 200 images from the Berkeley Segmentation Dataset [30] are used. Average PSNR improvements are recorded in the last row of Tab. IV. Compared to the 8-tap FIR filter, an average gain of about 0.63 dB is observed. The maximum gain and the minimum gain in the test set, compared to the 8-tap filter, are observed to be 3.13 dB and 0.14 dB, respectively. The average gains observed on the test set are close to the training set numbers.

Subjective evaluation

Fig. 8 shows two input LR images, (a) and (b), and output HR images produced using directional, cubic spline and the proposed interpolation technique. Directional interpolation results, (c) and (f), have some jaggedness for regions with strong edges and show some artifacts. The cubic spline results, (d) and (g), do not have any strong artifacts but show blurring of edges. HR images produced using the proposed approach, (e) and (h), are sharper and do not exhibit any noticeable artifacts. Fig. 9 shows two more input LR images, (a) texture and (b) text areas, and their corresponding output HR images. The texture in (e) appears slightly sharper than other methods, and the text in (h) seems to be sharper than the other results. It can also be seen that, even for intricate textures, the proposed method produces results without evident artifacts.

(a) (b)

(c) Directional [28] (d) Cubic spline (e) Proposed

(f) Directional [28] (g) Cubic spline (h) Proposed

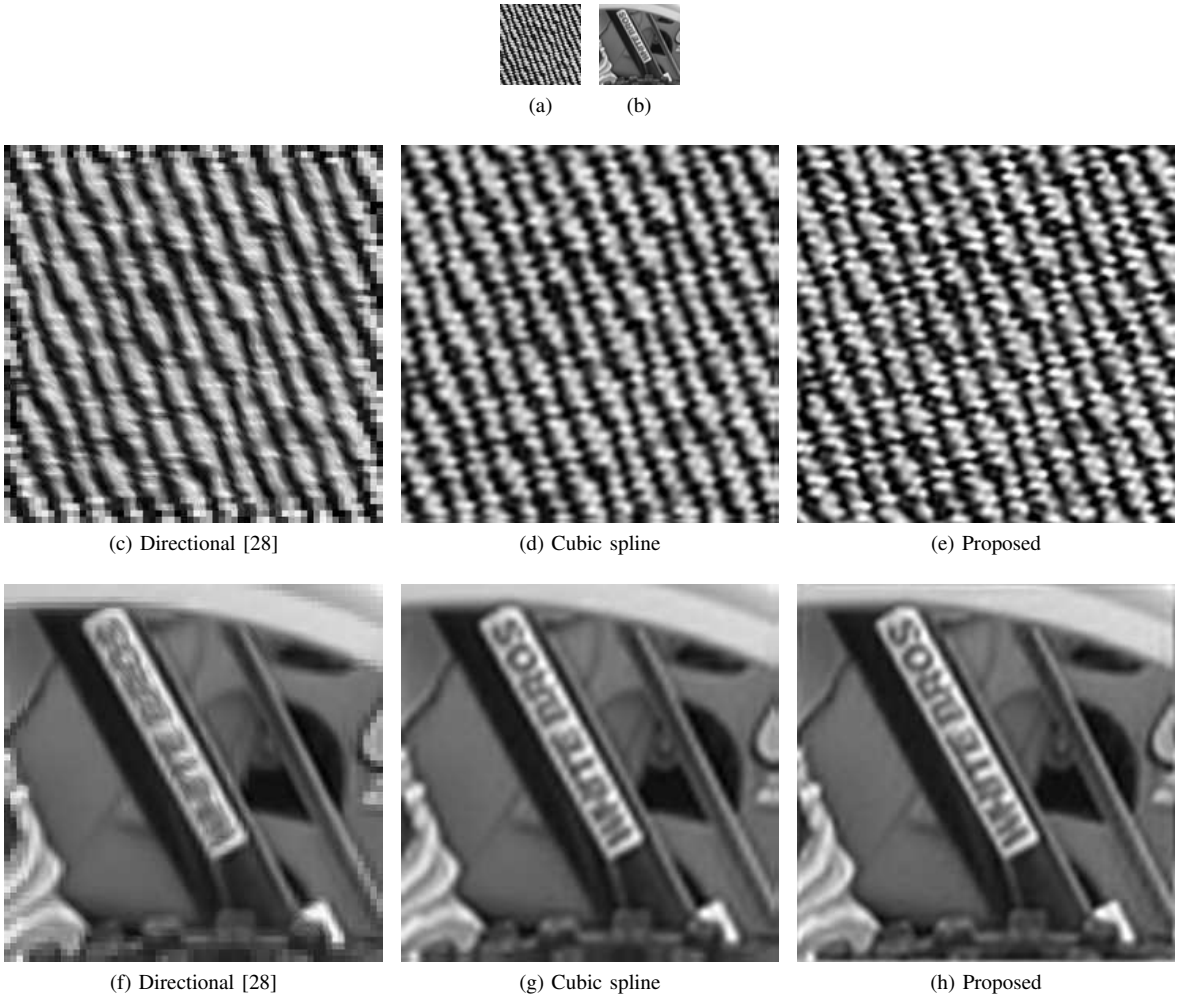
Fig. 8: Example 4x interpolation results. Input patches of size 64×64 in (a) and (b) are upsampled to 256×256. In (c) the diagonal stripes show jaggedness, in (d) the diagonal stripes are blurred. In (f) some artifacts can be noticed, in (g) the numbers and the rectangular frame below are blurry. The results of the proposed approach, (e) and (h), appear slightly sharper without evident artifacts.

One of the main drawbacks of the proposed approach is the high computational complexity. The complexity of the proposed approach is much higher than that of typical FIR interpolators, but of the same order of magnitude as other non-linear methods such as the contourlet scheme [33] and about 1.5x faster than the directional interpolation approach of [28]. Some important parameters that can be tuned for reducing the complexity are: the number of iterations for sparse approximation, the number of scales, the number of orientations for the directional filtering, etc. The filtering operations and element-wise thresholding involved in the proposed approach are amenable to parallel implementation.

VI. Summary and discussion

The problem of image interpolation is closely related to image modeling, i.e., we "select" a particular HR image that fits our model from a set of images that satisfy the given LR data. Unlike many other forms of data, images can show abrupt variations, e.g., across edges, which introduces challenges in modeling. In this paper, a framework for image interpolation that combines low frequencies from a linear method and high frequencies from a sparse approximation was presented. The key idea is to keep the support of the FIR filter short to avoid ringing artifacts in the initial upsampling and attack the problem of blurriness of the resulting image using a high pass estimate, through a sparse approximation in a multi-resolution directional dictionary.

In this paper, we evaluated linear methods such as bicubic, 6-tap, 8-tap, and 12-tap filters, as well as spline based methods. In the non-linear category, a directional interpolation method was evaluated, along with the proposed method using contourlet and shearlet dictionaries. All the tested approaches perform well for smooth image regions, with the main differences being observed at edges and in textured areas. The linear methods have only a small number of free parameters and once a set of parameters has been chosen, the performance variation from image-to-image is relatively small. The non-linear methods have a higher number of free parameters, hence a more careful setting is required. Some quantitative methods were provided for parameter selection in the proposed



(a) (b)

(c) Directional [28] (d) Cubic spline (e) Proposed

(f) Directional [28] (g) Cubic spline (h) Proposed

Fig. 9: More interpolation results for subjective evaluation. Results of 4x interpolation of LR inputs (a) texture and (b) text areas.

approach. With the final set of selected parameters, an average PSNR gain of around 0.63 dB was observed compared to a 8-tap filter over a test set of 200 images. The maximum gain was around 3.13 dB, which is significant. Additionally, many LR image regions with different characteristics were interpolated and subjectively evaluated. The proposed method showed improvements in subjective quality compared to other approaches and no evident artifacts were observed, even for complex regions.

REFERENCES

[1] 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.

[2] J. Allebach and P. W. Wong, "Edge-directed interpolation," in Proc. Int. Conf. Image Process., 1996, vol. 3, pp. 707–710.

[3] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imag. Sci., 2(1):183–202, 2009.

[4] A. Ben-Tal and A. Nemirovski, "Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications," MPS/SIAM Ser. Optim., SIAM, Philadelphia, 2001.

[5] T. Blu, P. Thévenaz, M. Unser, "MOMS: Maximal-Order Interpolation of Minimal Support," IEEE Trans. on Image Proc., vol. 10, no. 7, pp. 1069 – 1080, Jul. 2001.

[6] S. Boyd and L. Vandenberghe, "Convex Optimization," Cambridge University Press, Mar. 2004.

[7] 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.

[8] I. Daubechies, M. Defriese, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," Commun. Pure Appl. Math, LVII:1413–1457, 2004.

[9] W. Dong, L. Zhang, G. Shi, X. Wu, "Image Deblurring and Super-resolution by Adaptive Sparse Domain Selection and Adaptive Regularization," IEEE Trans. on Image Proc., vol. 20, no. 7, pp. 1838–1857, Jul. 2011.

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

[11] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," Biometrika, vol. 81, pp. 425-455, 1994.

[12] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian and D. Picard, "Wavelet Shrinkage: Asymptopia," J. Roy. Statist. Soc., vol. 57, no. 2, pp. 301–369, 1995.

[13] R. R. Coifman and D.L. Donoho, "Translation-Invariant De-Noising, in Wavelets and Statistics," A. Antoniadis and G. Oppenheim, Eds. San Diego, CA: Springer-Verlag, Lecture notes 1995.

[14] D. Field, "What is the goal of sensory coding?," Neural Computation, vol. 6, pp. 559–601, 1994.

[15] J. E. Fowler, "The redundant discrete wavelet transform and additive noise," IEEE Signal Proc. Letters, vol. 12, issue 9, pp. 629-632, 2005.

[16] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," SIAM Journal on Scientific Computing, vol. 20, pp. 33–61, 1998.

[17] 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.

[18] M. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," IEEE Trans. Image Process., 12(8), pp.906–916, Aug. 2003.

[19] O. G. Guleryuz, "On Missing Data Prediction using Sparse Signal Models: A Comparison of Atomic Decomposition with Iterated Denoising," Proc. of the SPIE, 5914, 1G, 2005.

[20] O. G. Guleryuz, "Nonlinear Approximation Based Image Recovery Using Adaptive Sparse Reconstructions and Iterated Denoising: Part I - Theory," IEEE Trans. on Image Proc., Mar. 2006.

[21] O. G. Guleryuz, "Nonlinear Approximation Based Image Recovery Using Adaptive Sparse Reconstructions and Iterated Denoising: Part II - Adaptive Algorithms," IEEE Trans. on Image Proc., Mar. 2006.

[22] R. Houska, "The nonexistence of shearlet scaling functions," Appl. Comput. Harmon. Anal., vol. 32, pp. 28–44, 2012.

[23] M. Irani and S. Peleg, "Improving resolution by image registration," Graphical Models and Image Proc., 53:231–239, 1991.

[24] P. Kittipoom, G. Kutyniok and W.-Q Lim, "Construction of Compactly Supported Shearlet Frames", Constr. Approx., vol. 35, pp. 21–72, 2012.

[25] G. Kutyniok and W.-Q Lim "Compactly Supported Shearlets are Optimally Sparse", J. Approx. Theory, vol. 163, pp. 1564-1589, 2011.

[26] G. Kutyniok and D. Labate, "Shearlets - Multiscale Analysis for Multivariate Data" Birkhäuser, 2012.

[27] H. Lakshman, W.-Q Lim, H. Schwarz, D. Marpe, G. Kutyniok, T. Wiegand, "Image Interpolation using Shearlet based Sparsity Priors," to apper in IEEE Int. Conf. on Image Proc., Melbourne, Australia, Sep. 2013.

[28] X. Li and M. T. Orchard, "New edge-directed interpolation." IEEE Trans. on Image Proc., vol. 10, 2001.

[29] W.-Q Lim, "Nonseparable Shearlet Transform", to appear in IEEE Trans. on Image Proc., 2013.

[30] 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, Jul. 2001. http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/.

[31] S. Mallat and Z. Zhang, "Matching pursuit in a time-frequency dictionary," IEEE Trans. Signal Process., vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[32] S. Mallat, and G. Yu, "Super-Resolution With Sparse Mixing Estimators," IEEE Trans. on Image Proc., vol. 19, no. 11, Nov. 2010.

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

[34] B. K. Natarajan, 'Sparse Approximate Solutions to Linear Systems," SIAM Journal on Computing, vol. 24, no. 2, pp. 227–234, Apr. 1995.

[35] B. Olshausen and D. Field, "Natural image statistics and efficient coding," Network: Computation in Neural Systems, no. 7, pp. 333–339, 1996.

[36] B. Olshausen and D. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?," Vision Research, vol. 37, pp. 3311–3325, 1997.

[37] Lecture notes: http://eeweb.poly.edu/iselesni/lecture_notes/sparsity_intro/index.html

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

[39] ShearLab: http://www.shearlab.org/index_software.html

[40] H. Shi and R. Ward, "Canny edge based image expansion," in Proc. IEEE Int. Symp. Circuits Syst., vol. 1, pp. 785–788, 2002.

[41] G. Strang and G. Fix, "A Fourier analysis of the finite element variational method," (in Cremonese), in Constructive Aspect of Functional Analysis, G. Geymonat, Ed. Rome, Italy: Edizioni Cremonese, pp. 796 – 830, 1971.

[42] G. Strang, "Introduction to Linear Algebra," Wellesley-Cambridge Press and SIAM, 4^{th} Edition.

[43] P. Thevenaz, T. Blu, and M. Unser, "Interpolation revisited," IEEE Trans. on Medical Imaging, vol. 19, no. 7, pp. 739 –758 , Jul. 2000.

[44] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," IEEE Trans. Inform. Theory, vol. 50, pp. 2231–2242, 2004.

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

[46] M. Unser, "Sampling - 50 Years After Shannon," Proc. IEEE, vol. 88, no. 4, pp. 569-587, Apr. 2000.

[47] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image Super-resolution via Sparse Representation," IEEE Trans. on Image Proc., pp. 2861–2873, vol. 19, no. 11, May 2010.

[48] M. Zibulevsky and M. Elad, "L1-L2 optimization in signal and image processing – Iterative shrinkage and beyond," IEEE Sig. Proc. Magazine, pp. 76–88, May 2010.