Spectral Remapping for Image Downscaling

EDUARDO S. L. GASTAL, Instituto de Informática – UFRGS
MANUEL M. OLIVEIRA, Instituto de Informática – UFRGS

(a) *Lanczos* (b) *Öztireli and Gross 2015* (c) *Weber et al. 2016* (d) *Our Spectral Remapping*

Fig. 1. Comparison of various image-downscaling techniques. (left) Reference image. (right) Downscaling to 180 × 144 pixels performed with: (a) Lanczos filtering followed by resampling using a cubic B-Spline causes structured high-frequency details from the pants, scarf, books, and most of the table cloth to be removed. (b) The technique by Öztireli and Gross introduces aliasing artifacts in those regions. (c) The technique of Weber et al. removes most of these high-frequency details, but still exhibits aliasing (e.g., see books). (d) By remapping high frequencies to the representable range of the downsampled spectrum, our approach retains the structured details. The green plots under the images are the intensity values of the highlighted pixels. The light-blue envelopes show the horizontally-compressed plot of the reference image. "Barbara" test image attributed to Allen Gersho (public domain). Please zoom in to see the details.

We present an image downscaling technique capable of appropriately representing high-frequency structured patterns. Our method breaks conventional wisdom in sampling theory—instead of discarding high-frequency information to avoid aliasing, it controls aliasing by remapping such information to the representable range of the downsampled spectrum. The resulting images provide more faithful representations of their original counterparts, retaining visually-important details that would otherwise be lost. Our technique can be used with any resampling method and works for both natural and synthetic images. We demonstrate its effectiveness on a large number of images downscaled in combination with various resampling strategies. By providing an alternative solution for a long-standing problem, our method opens up new possibilities for image processing.

CCS Concepts: • **Computing methodologies → Image processing**;

Additional Key Words and Phrases: Image downscaling, spectral remapping, frequency remapping, antialiasing, resampling, signal processing.

ACM Reference format:
Eduardo S. L. Gastal and Manuel M. Oliveira. 2017. Spectral Remapping for Image Downscaling. *ACM Trans. Graph.* 36, 4, Article 145 (July 2017), 16 pages.
https://doi.org/10.1145/3072959.3073670

Author's e-mails: {eslgastal,oliveira}@inf.ufrgs.br.

0730-0301/2017/7-ART145 $15.00
https://doi.org/10.1145/3072959.3073670

1 INTRODUCTION

Image downscaling is one of the most prevalent image-processing operations. It is present, for instance, when we (pre-)view images on the displays of smartphones and digital cameras, or browse photo collections. Unfortunately, some spatial frequencies found in the

original images might not be representable in their reduced versions. Thus, following sampling theory [Shannon 1949], the original images have traditionally been low-pass filtered before resampling. Intended to avoid aliasing, the pre-filtering stage tends to discard important visual information, such as high-frequency texture details (Fig. 1a). Representing high-frequency content in the downsized images has been a long-standing challenge in image processing, which previous techniques have not been able to conquer.

We present a *technique for image downscaling capable of representing high-frequency details*. Unlike the classic approach based on Shannon's sampling theorem, which discards high-frequency content before resampling, *our technique remaps such content to the representable range of the downsampled spectrum*. As such, it preserves structured patterns by adjusting their frequencies (Fig. 1d). While this is a form of aliasing, it is controlled by our method. This process is performed at the original image resolution, such that any resampling method, including the recent techniques of Kopf et al. [2013], Öztireli and Gross [2015], and Weber et al. [2016], can be subsequently applied without introducing aliasing or discarding important details. The resulting images better represent the original content, which may contain arbitrary high frequencies.

Our **spectral remapping** technique is based on a discrete Gabor space-frequency analysis [Gabor 1946]. The original image is decomposed into a series of overlapping patches, which are transformed to frequency domain after a pixelwise multiplication by a Gaussian window. Frequency remapping is performed for each patch. The spectrally-remapped image is then obtained from the spectra of the resulting patches by transforming them back to spatial domain, multiplying by the dual of the Gaussian window, and accumulating the overlapping patches. The algorithm is detailed in Section 4.

We demonstrate the effectiveness of our technique by applying it to a large number of natural and synthetic images, in combination with various resampling strategies. Fig. 1 compares our downscaling method to three others, applied to the reference image on the left, which has been reduced to 180×144 pixels (comparisons with additional techniques are available in the supplementary materials). Fig. 1a shows the result obtained with the use of Lanczos filtering followed by resampling using a cubic B-Spline. Note that this removes high-frequency details from the pants, scarf, books, and most of the table cloth. The technique by Öztireli and Gross [2015] introduces aliasing artifacts in those regions (Fig. 1b). The method by Weber et al. [2016] removes most of these high-frequency details, but still exhibits aliasing (see the books in Fig. 1c). Our approach remaps relevant high-frequency content to a representable lower-frequency counterpart, thus retaining the details. Fig. 1d was obtained applying our technique to the reference image, followed by resampling performed with Lanczos filtering and subsampling using a cubic B-Spline (same procedure used to generate Fig. 1a). Despite of the use of the same resampling method, all visually-important details are properly represented in our result. A direct comparison of these images illustrates the benefits of our technique.

The **contributions** of our work include:

- An image-downscaling technique that properly represents structured high-frequency content depicted in the original image (Section 4). Our technique is the first of its kind, and can be used in combination with any resampling method;
- A technique for automatically decomposing an image into a set of spatially-localized non-harmonic waves, and for computing the waves' amplitudes and phases (Section 4.1 and Appendix B);
- A technique for remapping high frequencies to the representable range of the downsampled spectrum (Section 4.2). The robustness of our approach supports the remapping of arbitrary frequencies present in the original image;
- A technique for aligning the phases of the remapped waves (Section 4.3), avoiding discontinuities in the resulting image;
- A technique for preserving the original image's energy in its spectrally-remapped version (Section 4.4 and Appendix C).

2 RELATED WORK

Image downscaling techniques have been traditionally based on sampling theory [Shannon 1949], consisting of an antialising low-pass filtering stage followed by resampling. Since the "ideal" low-pass sinc filter has infinite image-space support, several finite approximations have been used in practice to try to prevent aliasing while minimizing ringing artifacts. Examples of such filters include the Lanczos kernel, and several cubic filters [Mitchell and Netravali 1988]. Modern works also show that sampling is better seen through approximation theory [Unser 2000], and the use of compact polynomial bases with recursive digital filters leads to better and more efficient reconstruction algorithms [Nehab and Hoppe 2014; Sacht and Nehab 2015]. Nevertheless, the quality of pre-filtering and reconstruction strategies is a subjective issue involving an interplay among aliasing, blurring, and ringing [Mitchell and Netravali 1988]. Regardless of this subjectivity, by removing high-frequency content, this classic approach tends to discard important visual content (Fig. 1a).

Recently, a few alternative image-downscaling techniques have been proposed trying to preserve image details. Kopf et al. [2013] optimize the shapes of downsampling bilateral kernels to locally adjust them to image features. This approach produces crisp images, avoids ringing artifacts, and can be successfully applied to pixel art. Öztireli and Gross [2015] also model image downscaling as an optimization problem. Their technique maximizes the the structural similarity index (SSIM) [Wang et al. 2004] between the original and corresponding downscaled image. Despite of its focus on image structure, the technique does not handle structured patterns, thus introducing aliasing artifacts (Fig. 1b). Weber et al. [2016] presented a fast downscaling algorithm suited for large images and videos. It implements a variation of a joint-bilateral filter whose range kernel favors differences in local pixel neighborhoods. This technique does not handle structured patterns either. It mostly blurs them, while also introducing some aliasing (Fig. 1c). While all these techniques can preserve some image details, none of them handles structured high-frequency content, a fundamental component of visually-rich natural images. By performing frequency remapping, our approach avoids discarding important structured details, which become represented at lower frequency. The result is a more faithful representation of the original visual content (Fig. 1d). Our technique can be used with any resampling strategy, including the ones just described. Since each resampling technique has its own strengths and limitations, together they provide a suit of methods that can

be combined with our approach to obtain better results in different situations. This flexibility is illustrated in Section 6.

A separate class of image downscaling techniques try to preserve the perception of blur and noise observed in the original images. Samadani et al. [2010] estimate the amount of blur and noise found in an image, and subsequently re-synthesize them on a regularly downsampled thumbnail. Tretacoste et al. [2011] enhanced Samadani et al's "re-blurring" approach using a perceptual blur-estimation model. These techniques can produce more accurate thumbnail representations for the original content. Since downsampling is performed before the reintroduction of blur and noise, these techniques can be integrated with and benefit from our approach.

The notion of creating a new signal by analysing and shifting frequencies has already been explored in the audio-processing literature [Quatieri and McAulay 1986; Smith and Serra 1987]. For audio, however, the problem is moderately simpler, as the ear is less sensitive to errors in phase, provided that phase continuity is maintained [Smith and Serra 1987]. When working with images, phase alignment becomes a crucial factor [Oppenheim and Lim 1981].

3 MATHEMATICAL BACKGROUND

In order to properly describe our method for image downscaling, we provide a brief review of Fourier theory and Gabor analysis. For simplicity, concepts and formulations are presented in 1-D. Their extensions to 2-D are described in Appendix A. A vector s in the usual finite-dimensional vector space $\mathbb{C}^N$ is a sequence of N complex numbers whose n-th element we denote by $s(n)$, for $n = 0..N-1$. Through Fourier decomposition, it is possible to describe any such vector as a linear combination of N complex waves f_k oscillating at harmonic frequencies k/N, for $k = 0..N-1$:

$$s = \sum_{k=0}^{N-1} \hat{s}(k) f_k, \quad \text{where} \quad f_k(n) = \frac{1}{\sqrt{N}} \exp\left(2\pi i\, n\, k/N\right). \tag{1}$$

The vector $\hat{s}$ is the discrete Fourier transform (DFT), or spectrum, of s, and $i = \sqrt{-1}$. The k-th Fourier coefficient can be computed as

$$\hat{s}(k) = \langle s \mid f_k \rangle = \sum_{n=0}^{N-1} s(n) f_k^*(n), \tag{2}$$

noting that the vectors $f_0, f_1, \ldots, f_{N-1}$ define an orthonormal basis for $\mathbb{C}^N$ with respect to the usual inner product $\langle \cdot \mid \cdot \rangle$ of this space, and $f_k^*(n)$ denotes the complex conjugate of $f_k(n)$.

Let vectors ρ and θ be the power and phase spectra of s. Eq. 1 can be re-written as a sum of scaled and phase-shifted complex waves:

$$s(n) = \sum_{k=0}^{N-1} \sqrt{\rho(k)}\, \frac{1}{\sqrt{N}} \exp\left(2\pi i\, n\, k/N + i\, \theta(k)\right),$$

where $\rho(k) = |\hat{s}(k)|^2$, $\theta(k) = \operatorname{atan2}(\Im\hat{s}(k), \Re\hat{s}(k)) \in [0, 2\pi)$, and $\Im\hat{s}(k)$ and $\Re\hat{s}(k)$ are the imaginary and real parts of $\hat{s}(k)$.

3.1 Gabor Space-Frequency Decomposition

One inherent limitation of the classic Fourier decomposition just described is the fact that the basis vectors $f_0, f_1, \ldots, f_{N-1}$ are not localized in their primal-domain (e.g., time or space) representation. Thus, from the Fourier coefficients, it is not possible to determine the spatial position of, for example, high-frequency details of textured objects in an image. In his seminal work on time-frequency analysis, Gabor [1946] addressed the issue of lack of time localization by proposing a representation where a signal is expanded onto a system of vectors that are parameterized by both time and frequency. In such a system, time analysis and Fourier analysis are both special limit cases. The situation of lack of spatial localization is analogous, making Gabor analysis very relevant to image processing since the underlying structure of image patches can be described by Gabor atoms as fundamental components [Olshausen and Field 1996]. Next, we present one possible discretization of Gabor's continuous space-frequency decomposition. For a detailed description of Gabor analysis, we refer the reader to the works of Zibulski and Zeevi [1994] and Mallat [1998].

3.1.1 Discrete Gabor analysis. From now on, we consider all vectors s in $\mathbb{C}^N$ not as finite complex sequences, but instead as N-periodic complex sequences that satisfy $s(n+N) = s(n)$, $\forall n \in \mathbb{Z}$. Any such vector may be described as the linear combination of N^2 *Gabor atoms* $g_{n,k}$ defined by the translation in space and frequency of a real, symmetric, and unit-norm window vector g:

$$s = \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} \check{s}(n,k)\, g_{n,k}, \tag{3}$$

where $g(m) = g(-m)$, $\|g\| = \sqrt{\langle g \mid g \rangle} = 1$, and

$$g_{n,k}(m) = g(m-n)\, \frac{1}{\sqrt{N}}\, \exp\left(2\pi i\, m\, k/N\right). \tag{4}$$

For suitable choices of g, each Gabor coefficient $\check{s}(n,k)$ *describes how much of the signal's energy is concentrated in a neighborhood of the n-th position in space and of the k-th harmonic in frequency.* As shown by Gabor [1946] for the continuous case, the smallest possible neighborhood for energy measurement is obtained if the window is chosen as a Gaussian function, $\exp\left(-0.5\, m^2/\sigma^2\right)$. Harris [1978] presents an extensive discussion of other choices of windows. In any case, however, space and frequency locality are conflicting requirements; a consequence of the fact that the frequency of a signal with finite duration can only be defined with some uncertainty [Mallat 1998]. For the Gaussian window, decreasing the value of σ improves space resolution at the cost of poorer frequency resolution, and vice versa. We discuss our choice of σ for the image downscaling problem in Section 5.

Gabor atoms are not linearly-independent vectors, and thus do not define an orthogonal basis. But since they consist of *unitary* space translations of the window vector g, the set $\{g_{n,k}\}$ defines a *tight frame* [Zibulski and Zeevi 1994]: an overcomplete system of vectors with properties similar to an orthogonal basis. Like Fourier coefficients, Gabor coefficients for this particular expansion are computed by the inner product of s with the respective atoms, $\check{s}(n,k) = \langle s \mid g_{n,k} \rangle$. Section 5 discusses the use of non-unitary space and frequency translations for computational efficiency.

Let $g_n(m) = g(m-n)$ be the primal translation of the window and let $s \cdot g$ be the elementwise vector product. A useful characterization of Gabor decomposition is given by the identities (see Eqs. 1, 2, 4)

$$\check{s}(n,k) = \langle s \mid g_{n,k} \rangle = \langle s \mid g_n \cdot f_k \rangle = \langle s \cdot g_n \mid f_k \rangle = \widehat{s \cdot g_n}(k).$$

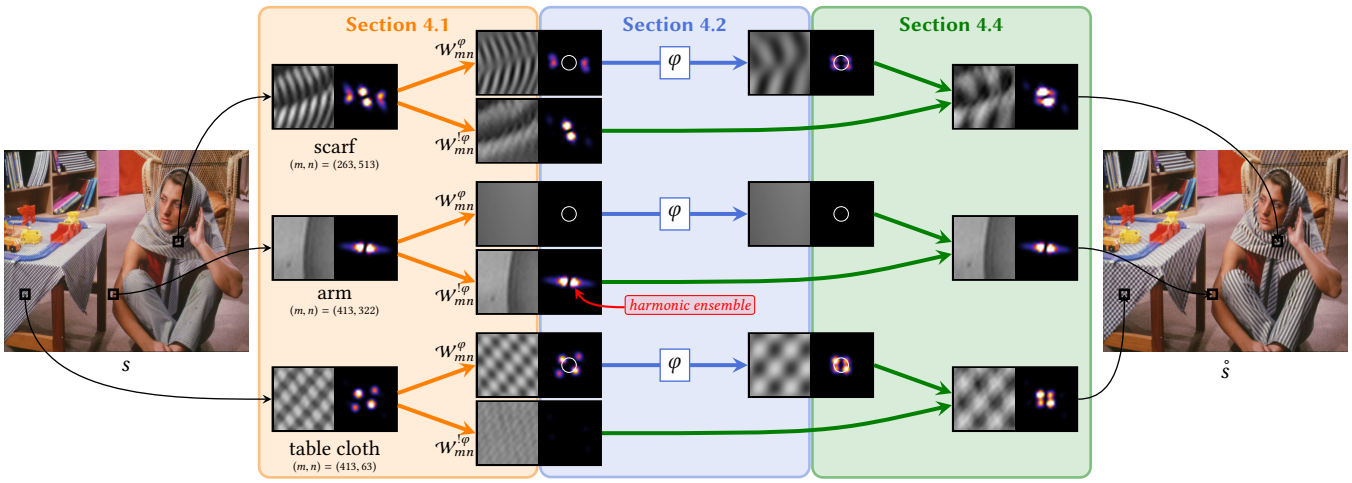

Fig. 2. Overview of our spectral-remapping technique. An image s is separated into a series of overlapping patches centered at pixels $s(m, n)$, of which three are shown (scarf, arm, and table cloth). The image to the right of each patch shows its (mean-centered) Fourier power spectrum after windowing with a Gaussian function. ***(Step 1, Section 4.1)*** Each spectrum is decomposed onto two sets of waves: those that require frequency remapping ($\mathcal{W}^{\varphi}_{mn}$), and those that do not ($\mathcal{W}^{!\varphi}_{mn}$). Waves in $\mathcal{W}^{\varphi}_{mn}$ are outside the spectral circle $\mathfrak{C}$ (in white), and thus would be non-representable after downscaling. ***(Step 2, Section 4.2)*** All waves in $\mathcal{W}^{\varphi}_{mn}$ are remapped by the function φ to a new location inside $\mathfrak{C}$. ***(Step 3, Section 4.3)*** Remapped waves' phases are re-computed in order to preserve their alignments (not shown). ***(Step 4, Section 4.4)*** Remapped waves are combined with the non-remapped residual in order to form the final spectrally-remapped image $\mathring{s}$.

Thus, for arbitrary but fixed n, Gabor coefficients $\check{s}(n, k), \forall k$, can be computed by Fourier analysis of the "windowed signal" $s \cdot g_n$.

4 SPECTRAL REMAPPING

Our technique detects relevant high-frequency and structured details and remaps them to lower frequencies representable in the downsampled spectrum. This process should maximize visual similarity between the image's original and downsampled versions. Fig. 2 provides an overview of our method, whose details and theoretical justifications are presented next.

In our mathematical notation, a 2-D digital image s, containing M rows and N columns, is a double-indexed vector whose (m, n)-th pixel we denote by $s(m, n) \in \mathbb{C}$, for $m = 0..M-1$ and $n = 0..N-1$. Fourier analysis is a separable transform, meaning that the 2-D Fourier basis is given by the outer product of the vectors defined in Eq. 1. The same consideration is valid for the Gabor decomposition. Appendix A gives the mathematical details.

4.1 Detecting Local Waves

This section introduces an algorithm for obtaining a sparse space-frequency decomposition for images. Unlike conventional Gabor analysis [Gabor 1946], which locally decomposes non-harmonic waves as a linear combination of a set of harmonic (Fourier) basis functions, our technique directly detects single non-harmonic waves of arbitrary frequencies.

Structured visual detail is characterized by periodic patterns, often with some random variations in position, orientation, or color [Portilla and Simoncelli 2000]. In a small neighborhood of a 2-D digital image s, the idealized wave decomposition consists of a small number of non-harmonic waves with well-defined frequencies. Let (m, n) be the center of such a spatial neighborhood, and let $s \cdot g_{mn}$ be a windowed portion of s around (m, n). We wish to find the set of waves $\mathcal{W}_{mn}$ that best describes $s \cdot g_{mn}$ by minimizing

$$E(\mathcal{W}_{mn}) = \left\| s \cdot g_{mn} - \sum_{(\alpha,\mathrm{a},\mathrm{b},\mathrm{c}) \in \mathcal{W}_{mn}} w_{\alpha,\mathrm{a},\mathrm{b},\mathrm{c}} \cdot g_{mn} \right\|, \tag{5}$$

where each tuple of real numbers $(\alpha, \mathrm{a}, \mathrm{b}, \mathrm{c})$ uniquely defines a bidimensional wave of amplitude α, vertical and horizontal frequencies a and b in $[-0.5, 0.5)$, and phase c in $[0, 1)$:

$$w_{\alpha,\mathrm{a},\mathrm{b},\mathrm{c}}(m', n') = \alpha \, \exp\left[\, 2\pi i \, (\mathrm{a}m' + \mathrm{b}n' + \mathrm{c}) \,\right]. \tag{6}$$

Since the waves in Eq. 5 approximately define the neighborhood around (m, n), the set $\mathcal{W}_{mn}$ provides a local description of s in terms of space-frequency non-harmonic atoms $w_{\alpha,\mathrm{a},\mathrm{b},\mathrm{c}} \cdot g_{mn}$.

4.1.1 Minimizing $E(\mathcal{W}_{mn})$. Removing the windowing by g_{mn} from Eq. 5 results in a minimization program known as *multi-tone frequency estimation* [Hannan 1973; Rife and Boorstyn 1976]. Although statistical solutions that achieve optimal asymptotic performance exist through maximum-likelihood and subspace methods, they require the estimation of the number of terms in the sum, which may be a challenging task [Quinn 1989]. We avoid this problem by using an iterative algorithm based on finding maximal peaks in the power spectrum of $s \cdot g_{mn}$ [Smith and Serra 1987], a technique that provides fast and accurate results for a variety of applications [Jacobsen and Kootsookos 2007], including image downscaling. Appendix B provides the details of this algorithm.

4.2 Remapping Waves

For each pixel (m, n) in s, we obtain a local decomposition $\mathcal{W}_{mn}$ of a neighborhood centered at (m, n). When an image is downscaled by a

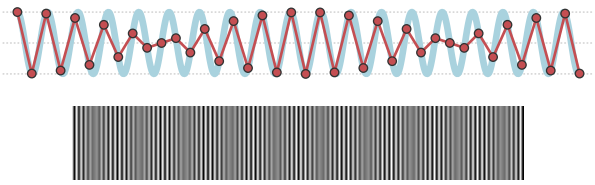

Fig. 3. Sampling waves with frequencies close to the Nyquist limit results in distracting visual artifacts. (top) Sampling of $\cos(2\pi\frac{19}{40}n)$ for integer n. (bottom) Direct 2-D display of the sampled values.

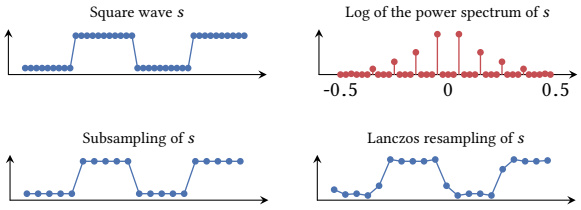

Fig. 4. Harmonic ensembles, such as the one observed in a square wave, are characterized by a spectrum composed of several correlated harmonics. For them, unlike for most non-ensemble signals, naive subsampling *does not* introduce aliasing artifacts and works better than, e.g., Lanczos resampling. Fig. 2 shows an example of a 2-D harmonic ensemble.

factor $R>1$ in both dimensions (non-uniform downscaling is similar, using two factors $R_x \neq R_y$), many waves in $\mathcal{W}_{mn}$ may not be representable anymore. As a corollary of the sampling theorem [Shannon 1949], a wave $(\alpha, \mathrm{a}, \mathrm{b}, \mathrm{c})$ in $\mathcal{W}_{mn}$ is representable in the downsampled spectrum if and only if a and b are in $[-\frac{0.5}{R}, \frac{0.5}{R})$. For each *non-representable* wave, we apply a function φ that *remaps* its frequency coordinates to a new location *inside* the representable range: $(\mathrm{a}, \mathrm{b}) \mapsto \varphi(\mathrm{a}, \mathrm{b}) \in [-\frac{0.5}{R}, \frac{0.5}{R})^2$. While there are several choices for φ, the function

$$\varphi(\mathrm{a}, \mathrm{b}) = \left(\frac{0.4}{R}\cos\vartheta, \frac{0.4}{R}\sin\vartheta\right), \quad \text{where} \quad \vartheta = \operatorname{atan2}(\mathrm{b}, \mathrm{a}), \tag{7}$$

works well for natural and synthetic images. Alternative φ functions are likely to produce different frequency-remapping results, and this subject is left for future exploration. Such a radial mapping replaces higher-frequency waves with lower-frequency ones, preserving the orientations of the original waves (i.e., ϑ), and, in turn, the local structure of the image's content.

The reason for choosing a spectral radius of $\frac{0.4}{R}$ is twofold: first, as it will be discussed shortly, some non-representable waves are not good candidates for remapping and are thus left unchanged. This means that low-pass filtering is still required before the actual downscaling to eliminate such high-frequency residues. Since it is neither desirable nor practically realizable to have a digital pre-filter with an ideal transition band [Proakis and Manolakis 2007; Sacht and Nehab 2015], coefficients corresponding to frequencies close to the Nyquist limit $\frac{0.5}{R}$ end up incorrectly attenuated. The use of $\frac{0.4}{R}$ avoids attenuating the remapped waves. A second reason for choosing $\frac{0.4}{R}$ is that the sampling of waves with frequencies too close to the Nyquist limit results in ordinates that are distracting to the human eye, as illustrated in Fig. 3. This occurs because sampling (with pre-filtering) is equivalent to an orthogonal projection that minimizes the squared norm of the space [Unser 2000], and such a norm is oblivious to human perception [Nehab and Hoppe 2014]. By remapping high frequencies to a spectral circle $\mathfrak{C}$ of radius $\frac{0.4}{R}$ instead of $\frac{0.5}{R}$, we also avoid such perceptual artifacts. One should note that other radius values smaller than the Nyquist limit could be used, and the quality of such results is a subjective matter.

Finally, to be consistent, our implementation considers a wave as non-representable if its frequency coordinates fall outside the circle $\mathfrak{C}$, *even if* still inside the theoretical Nyquist range $[-\frac{0.5}{R}, \frac{0.5}{R})^2$.

4.2.1 Selective remapping. Some non-representable waves are not good candidates for frequency remapping. In particular, waves that belong to an *ensemble of correlated harmonics* are better left unchanged. Such ensembles are often associated with abrupt intensity changes, as in sharp edges or fine impulse-like details. These ensembles do not suffer from ordinary aliasing artifacts, and are better treated by non-traditional resampling strategies, such as the recent method of Öztireli and Gross [2015], or even by direct subsampling *without* pre-filtering (Fig. 4). Thus, waves belonging to harmonic ensembles should not be remapped by φ, allowing them to be correctly handled by the chosen resampling algorithm (which is orthogonal to, and complements, our method).

In our experience, inequality (8) correlates well with the occurrence of a harmonic ensemble related to a wave $(\alpha, \mathrm{a}, \mathrm{b}, \mathrm{c})$. This is an approximate, heuristic estimator, constructed by analysing ensembles from numerous natural images. Thus, we do *not* remap (a, b) to $\varphi(\mathrm{a}, \mathrm{b})$ if the energy of the wave, measured by its amplitude α, is less than or equal to the signal's energy at frequency $\varphi(\mathrm{a}, \mathrm{b})$, measured by its Fourier coefficient relative to the non-harmonic complex wave $f_{\varphi(\mathrm{a},\mathrm{b})} = w_{1,\varphi(\mathrm{a},\mathrm{b}),0}$, and increased by 0.6 dB:

$$\alpha \leq \frac{10^{0.06}}{\langle 1 \,|\, g_{mn}\rangle}\left|\langle \bar{s}_{mn} \cdot g_{mn} \,|\, f_{\varphi(\mathrm{a},\mathrm{b})}\rangle\right|. \tag{8}$$

The division by $\langle 1 \,|\, g_{mn}\rangle = \hat{g}_{mn}(0,0)$ disconsiders the energy originating from the window. Similarly, using the mean-centered signal $\bar{s}_{mn}$ instead of s disconsiders the energy contribution of the "0-th harmonic", making (8) invariant to the mean intensity of the image [Mairal et al. 2014]. $\bar{s}_{mn} = s - \langle s \,|\, g_{mn}\rangle / \langle 1 \,|\, g_{mn}\rangle$ is obtained by subtracting from s its windowed weighted average. This preserves all waves $(\alpha, \mathrm{a}, \mathrm{b}, \mathrm{c})$ from $s \cdot g_{mn}$ in $\bar{s}_{mn} \cdot g_{mn}$, except for its 0-th harmonic $(\alpha, 0, 0, 0)$, which becomes $(0, 0, 0, 0)$.

4.3 Aligning Remapped Waves

The phase spectrum θ defined in Section 3 carries much of the essential information of a signal s, especially for visual data such as digital images [Oppenheim and Lim 1981]. The phase shifts in θ define the alignments among waves, determining the locations of constructive and destructive interferences that build-up s.

In our space-frequency decomposition, θ is encoded in the phase values c of the waves $(\alpha, \mathrm{a}, \mathrm{b}, \mathrm{c})$ in $\mathcal{W}_{mn}$. Since many of those waves have their frequencies remapped by φ, their corresponding phases must also be re-computed in order to preserve their alignments. We do this by first detecting which pairs of remapped waves need to be aligned, and then solving a linear system for the remapped phases, so that all important alignments are preserved.

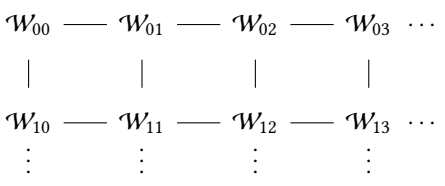

Fig. 5. Graph topology defined by the sets $\mathcal{W}_{mn}$.

4.3.1 Detecting alignment. Two waves $\mathrm{w}_1 = (\alpha_1, \mathrm{a}_1, \mathrm{b}_1, \mathrm{c}_1)$ and $\mathrm{w}_2 = (\alpha_2, \mathrm{a}_2, \mathrm{b}_2, \mathrm{c}_2)$ have an alignment at (m, n) if their values match in a 3×3 pixel neighborhood $\mathcal{N}_{mn}$ around (m, n). Note that waves only provide a locally-valid description of s (due to the windowing) and one should not require perfect alignment. Thus, we define a positive measure $\mu_{mn}(\mathrm{w}_1, \mathrm{w}_2) = \exp\left(-\delta^2/\lambda^2\right)$ in $[0, 1]$ (a Gaussian function on δ, with variance controlled by λ) for their degree of alignment at (m, n), where

$$\delta = \frac{1}{\min(\alpha_1, \alpha_2)} \sqrt{\sum_{(m', n') \in \mathcal{N}_{mn}} \left| w_{\mathrm{w}_1}(m', n') - w_{\mathrm{w}_2}(m', n') \right|^2}.$$

$w_{\mathrm{w}_j}(m', n')$ is a short notation for $w_{\alpha_j, \mathrm{a}_j, \mathrm{b}_j, \mathrm{c}_j}(m', n')$ (Eq. 6). The factor outside the sum makes the alignment measure μ proportional to the amplitude of the waves (note that there is an inverse relationship between δ and μ due to the exponential function). This is done in order to consider the alignment between larger amplitude waves as more important, seeing that μ is used as a weight in the least-squares system defined by Eq. 10. We use $\lambda = 0.5$, as it produces good results in practice.

4.3.2 Computing remapped phases. For all pixels (m, n) in s, the sets $\mathcal{W}_{mn}$ define a graph topology based on direct pixel neighborhoods (Fig. 5). Let w_1 and w_2 be two distinct waves in the union $\mathcal{W}_{mn} \cup \mathcal{W}_{m'n'}$ of any two sets $\mathcal{W}_{mn}$ and $\mathcal{W}_{m'n'}$ that are connected by an edge in this graph, and such that both w_1 and w_2 require frequency remapping to, respectively, $\varphi(\mathrm{w}_1)$ and $\varphi(\mathrm{w}_2)$. This abuse of notation in the argument of φ should be interpreted as $\varphi(\mathrm{w}_j) = (\alpha_j, \varphi(\mathrm{a}_j, \mathrm{b}_j), \mathrm{c}_j)$. If an alignment exists between w_1 and w_2, we would like to similarly create an alignment between $\varphi(\mathrm{w}_1)$ and $\varphi(\mathrm{w}_2)$ as this will preserve the continuity of the pixels' intensities for the reconstructed waves (described in Section 4.4). Thus, we require the remapped waves to match at point $(\bar{m}, \bar{n})$ halfway between (m, n) and (m', n'), the centers of the neighborhoods of $\mathcal{W}_{mn}$ and $\mathcal{W}_{m'n'}$ (recall that spectral remapping is performed at the original image resolution):

$$w_{\varphi(\mathrm{w}_1)}(\bar{m}, \bar{n}) = w_{\varphi(\mathrm{w}_2)}(\bar{m}, \bar{n}), \text{ where } \begin{matrix} \bar{m} = (m + m')/2, \\ \bar{n} = (n + n')/2. \end{matrix} \quad (9)$$

Finding the phase values of all waves that simultaneously solve this equation for all possible pairs w_1 and w_2 is a highly nonlinear problem, especially due to the oscillating nature of the complex exponential originating from Eq. 6. Furthermore, an exact solution may not exist. Thus, we instead find an approximate solution by making the assumption that any two waves w_1 and w_2 that have a large alignment measure $\mu_{\bar{m}\bar{n}}(\mathrm{w}_1, \mathrm{w}_2)$ will most likely have similar amplitudes. Indeed, assuming $\alpha_1 = \alpha_2$ greatly simplifies the issue, as per Eq. 6, Eq. 9 becomes $(\mathrm{a}_1^\varphi \bar{m} + \mathrm{b}_1^\varphi \bar{n} + \mathrm{c}_1^\varphi) \overset{\mathrm{mod}\ 1}{=} (\mathrm{a}_2^\varphi \bar{m} + \mathrm{b}_2^\varphi \bar{n} + \mathrm{c}_2^\varphi)$, where $(\mathrm{a}_j^\varphi, \mathrm{b}_j^\varphi) = \varphi(\mathrm{a}_j, \mathrm{b}_j)$ are the *known remapped frequencies* and c_j^φ are the *desired remapped phases*. Since the validity of this approximation depends on the magnitude of the alignment measure, we model the remapping-phase problem as a weighted least-squares system, resulting in the following functional which should be minimized with respect to the desired remapped phases c_1^φ and c_2^φ:

$$\left[\mu_{\bar{m}\bar{n}}(\mathrm{w}_1, \mathrm{w}_2) \left| (\mathrm{a}_1^\varphi \bar{m} + \mathrm{b}_1^\varphi \bar{n} + \mathrm{c}_1^\varphi) - (\mathrm{a}_2^\varphi \bar{m} + \mathrm{b}_2^\varphi \bar{n} + \mathrm{c}_2^\varphi) \right| \right]^2. \quad (10)$$

The minimization of the sum of all such functionals for all possible pairs of waves w_1 and w_2 that require remapping is a quadratic problem on their phases, solvable through a sparse system of linear equations. To reduce the system's size as well as to not force the alignment of waves with a small value of μ, we clamp $\mu_{\bar{m}\bar{n}}(\mathrm{w}_1, \mathrm{w}_2)$ to zero whenever $\delta > 3\lambda$.

4.4 Image Reconstruction with Remapped Waves

After frequency and phase remapping, one can synthesize *a new version $\mathring{s}$ of the signal s with the remapped waves*. We call $\mathring{s}$ the **spectrally-remapped image** obtained from s. Thus, let $\mathcal{W}_{mn}^{\varphi}$ be the subset of *remapped* waves $\subset \mathcal{W}_{mn}$ (from $s \cdot g_{mn}$), and let $\mathcal{W}_{mn}^{!\varphi} = \mathcal{W}_{mn} - \mathcal{W}_{mn}^{\varphi}$ be its complement. Furthermore, for a wave $\mathrm{w} = (\alpha, \mathrm{a}, \mathrm{b}, \mathrm{c})$, let $\varphi(\mathrm{w})$ denote the version of w with frequencies remapped to $\varphi(\mathrm{a}, \mathrm{b})$ and with phase computed by the least-squares system described in Section 4.3. We write $\mathring{s}$ as the sum of vector q, obtained from the original waves ($\mathcal{W}_{mn}^{!\varphi}$), and a scaled version of vector r, constructed from the remapped waves ($\mathcal{W}_{mn}^{\varphi}$):

$$\mathring{s} = q + \beta r$$

where, for $g_{mn}^2 \overset{\mathrm{def}}{=} g_{mn} \cdot g_{mn}$,

$$q = \sum_{\forall m, n} \sum_{\forall \mathrm{w} \in \mathcal{W}_{mn}^{!\varphi}} w_{\mathrm{w}} \cdot g_{mn}^2, \quad r = \sum_{\forall m, n} \sum_{\forall \mathrm{w} \in \mathcal{W}_{mn}^{\varphi}} w_{\varphi(\mathrm{w})} \cdot g_{mn}^2.$$

In practice, q is computed as the residual obtained after removing from s all the remapped waves:

$$q = s - \sum_{\forall m, n} \sum_{\forall \mathrm{w} \in \mathcal{W}_{mn}^{\varphi}} w_{\mathrm{w}} \cdot g_{mn}^2.$$

These summations are the reconstructions of the space-frequency decompositions defined by the sets $\mathcal{W}_{mn}^{!\varphi}$ and $\mathcal{W}_{mn}^{\varphi}$, similar to Eq. 3. Note that undoing the windowing effect in $w_{\mathrm{w}} \cdot g_{mn}$ requires an elementwise multiplication by the dual $\tilde{g}_{mn}$ of g_{mn} (see Appendix D). For a normalized window with unitary translations, $\tilde{g}_{mn} = g_{mn}$. This explains the g_{mn}^2 terms in the expressions for q and r. The value of $\beta > 0$ is chosen as to preserve the norm of the original signal, resulting in an isometry $s \mapsto \mathring{s}$ where $\|\mathring{s}\| = \|q + \beta r\| = \|s\|$. The computation of β is discussed in Appendix C.

Fig. 6 illustrates the image-reconstruction process with remapped waves, and Fig. 7 shows the result of downscaling using four resampling algorithms. The spectrally-remapped image of Fig. 1 (left) is shown in Fig. 2 (right) ($\mathring{s}$). The supplementary materials include the spectrally-remapped images of all examples shown in the paper.

Since our image-downscaling process consists of frequency remapping followed by resampling, the spectral-remapping process should

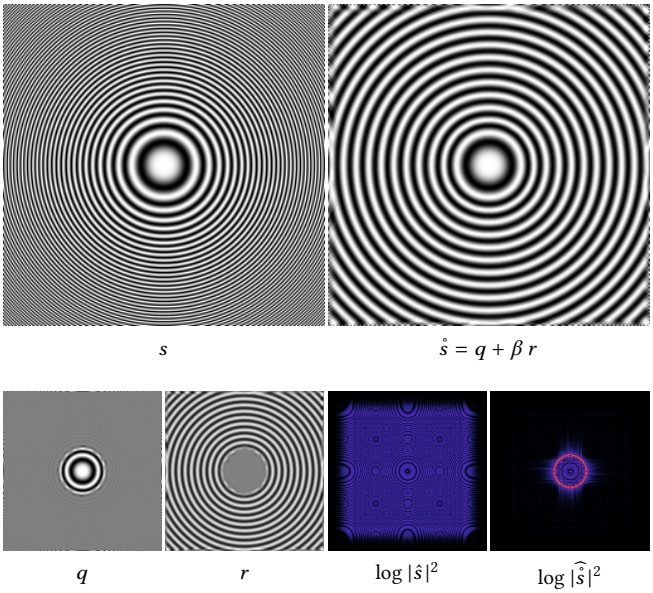

Fig. 6. Original image s (a chirp) defined by the equation $cos(x^2 + y^2)$ and centered at (0, 0), and its spectrally-remapped version $\mathring{s}$. The spectrum $\widehat{\mathring{s}}$ clearly shows all non-representable waves in $\hat{s}$ effectively remapped inside the spectral circle $\mathfrak{C}$ ($R = 4$ and $\sigma = 3.5$). Vector r is shown normalized to the range [0, 1].

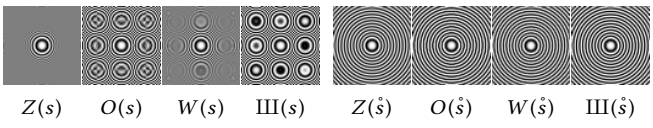

Fig. 7. Downscaling of the original image s and of its spectrally-remapped version $\mathring{s}$ (Fig. 6) to 64×64 pixels ($R = 4$ and $\sigma = 3.5$), using algorithms listed in Section 5.3. The downscaling obtained from $\mathring{s}$ provide more faithful representations of the image s.

be specific for the desired downscaling factor R. Note that depending on the frequency content of the input image, using a spectrally-remapped image obtained for a scaling factor R to create a downscaled image by a different factor R' may result in aliasing artifacts or unnecessary trade of higher by lower frequencies. Thus, if $R' > R$ and the spectrally-remapped image contains frequencies in the interval $(0.4/R', 0.4/R]$, the result should exhibit aliasing. If, however, $R' < R$, the resulting image might unnecessarily replace higher frequencies by lower ones.

5 IMPORTANT COMPUTATIONAL ASPECTS

The decomposition defined by Eq. 3 is highly redundant. Indeed, for a real signal, such as a digital image, N real elements in s are converted to N^2 complex coefficients in $\check{s}$. One question is whether it is possible to obtain a *non*-redundant space-frequency decomposition (with exactly N coefficients in $\check{s}$), since this would significantly reduce the required computational effort. Unfortunately, this is only possible for windows g that are poorly localized in either space or frequency, as described by the Balian-Low theorem [Heil 2007]. Thus, for the Gaussian window (which is optimally localized in both space and frequency), some redundancy is necessary and provides stability to the reconstruction of s from $\check{s}$.

In practice, we reduce the computational cost of Eq. 3 by subsampling the space-frequency plane [Mallat 1998], yet still providing the necessary redundancy. Thus, let T and L be natural numbers that satisfy $TL > N$, and such that T divides N (mirror padding of s works around this restriction). We decompose s as

$$s = \sum_{n=0}^{T-1} \sum_{k=0}^{L-1} \check{s}(n,k)\, \tilde{g}_{n\tau,k} = \sum_{n=0}^{T-1} \sum_{k=0}^{L-1} \langle s \,|\, g_{n\tau,k} \rangle\, \tilde{g}_{n\tau,k}, \qquad (11)$$

where $\tau = N/T$ is the integer step size used for the spatial translations of the atoms $g_{n\tau,k}$, whose harmonics are now multiples of the fundamental frequency $1/L$ instead of $1/N$:

$$g_{n\tau,k}(m) = g(m - n\,\tau)\, \frac{1}{\sqrt{L}}\, \exp\left(2\pi i\, m\, k/L\right).$$

As the vectors $g_{n\tau,k}$ do not form a tight frame, the expansion in Eq. 11 must use a *dual frame* composed of the vectors $\tilde{g}_{n\tau,k}$ [Zibulski and Zeevi 1994]. Appendix D shows how to compute $\tilde{g}_{n\tau,k}$. From Eq. 11, a signal s containing N elements can be represented by T sets of L complex coefficients each: a Gabor decomposition obtained by windowing s with T equally-spaced Gaussian windows, and sampling their elementwise product at L equally-spaced frequencies. This generalizes the decomposition in Eq. 3, for which $T = L = N$. If $TL = \gamma N$, $1 < \gamma < N$, the redundancy in the representation of s reduces by a factor of N/γ. The analysis of 2-D signals is similar.

The choices of T and L depend on the space-frequency spread of the Gaussian $g(m) = \exp\left(-0.5\, m^2/\sigma^2\right)$ [Mallat 1998], which is controlled by its standard deviation σ. For image downscaling, σ in turn depends on the downscaling factor R, since R dictates the spectral radius $\frac{0.4}{R}$ of $\mathfrak{C}$, which separates representable from non-representable waves (see Section 4.2). The frequency spread of g should be selected in order to properly resolve waves whose frequencies lie *outside* the spectral circle $\mathfrak{C}$. According to the guidelines described by Harris [1978], we select σ such that $\hat{g}$'s 6-dB bandwidth is small enough to avoid spectral leakage between conjugate frequencies outside $\mathfrak{C}$ (see relation (16) in Appendix B). In practice, this means that σ **should be larger than** $R/2$ **pixels**.

The simple choice of $\sigma = R/2$ is not necessarily the best option, unfortunately, as the space-frequency spread of the Gaussian must also be specifically adapted to each signal s. If waves in a local portion of s have nearby frequencies, then the value of σ must be increased to improve frequency resolution [Harris 1978; Mallat 1998]. Thus, the rule of thumb is to start with $\sigma = R/2$ and increase its value until a satisfactory decomposition is obtained. The values of σ used to generate the results in the paper are listed under each figure, except for Fig. 1, for which $\sigma = 3$ and $R = 4$.

Given a choice of σ and noting that $g(m)$ is practically zero for $|m| > 4\sigma$, we set L to the odd integer nearest to 8σ. Thus, we may truncate the window vector g at L elements, where $m \in \mathbb{Z}$ ranges from $-\lfloor L/2 \rfloor$ to $\lfloor L/2 \rfloor$ (approximately -4σ to 4σ). To guarantee that $\|g\| = 1$, we construct the discrete window g by sampling $g(m) = \exp\left(-0.5\, m^2/\sigma^2\right)$ at the specified values of m and then normalizing the obtained vector. Similarly, for numerical accuracy we

use a redundancy factor of $\gamma = 8$ for the subsampled transform in Eq. 11, and set T to the smallest integer that satisfies $TL > 8N$.

5.1 Complexity Analysis

For fixed n, we compute all L coefficients $\langle s \,|\, g_{n\tau,k} \rangle$ using a fast Fourier transform (FFT). Since $n = 0..T-1$, the time complexity of the method is $O(TL \log L) = O(N \log \sigma)$. The linear system from Section 4.3 has $O(T) = O(N/\sigma)$ non-zero ele- ments, which we solve using a sparse QR factorization [Davis 2011].

5.2 Color

Natural images have correlated color channels. We decorrelate them with principal component analysis [Reinhard et al. 2001], and perform spectral remapping only on the coefficients η_{v_1} associated with the maximal direction v_1 of color variation. Thus, $s_{\text{rgb}} = [s_\text{r}\ s_\text{g}\ s_\text{b}]$ is mapped to $\eta_{\text{pca}} = [\eta_{v_1}\ \eta_{v_2}\ \eta_{v_3}]$, then the monochromatic coefficients η_{v_1} undergo spectral-remapping to $\mathring{\eta}_{v_1}$, and $\mathring{s}_{\text{rgb}}$ is obtained from $\mathring{\eta}_{\text{pca}} = [\mathring{\eta}_{v_1}\ \eta_{v_2}\ \eta_{v_3}]$ by undoing the principal- component basis change. This procedure reduces the computational cost of processing multichannel images and is used for all results shown in the paper. As humans are more sensitive to contrast in lightness, it is possible to use CIELAB and work only with the L* channel; however, the principal-component approach works better in corner cases, such as constant-lightness synthetic color images.

5.3 Downscaling Algorithms

Spectral remapping is performed at the original image resolution. Subsequently, a suitable resampling strategy may be employed to downscale the image, without the risk of introducing aliasing or losing high-frequency structured detail. The resampling algorithms used in the paper are: Lanczos (Z), Cubic O-MOMS (D) [Blu et al. 2001; Nehab and Hoppe 2014], Kopf et al. [2013] (K), Öztireli and Gross [2015] (O), Weber et al. [2016] (W), and nearest-neighbor subsampling (III). For Lanczos we use a 5-lobed kernel which has a passband loss of 0.8 dB at the spectral radius $\frac{0.4}{R}$ of $\mathfrak{C}$. For classic antialiasing, however, we recommend the use of generalized-sampling strategies with efficient digital filters [Nehab and Hoppe 2014], such as the Cubic O-MOMS (loss of 0.4 dB at $\frac{0.4}{R}$). Subsampling (III) has 0 dB loss at *all* frequencies and thus may introduce aliasing due to high-frequency residues. The other listed algorithms are non-linear and cannot have their frequency response characterized. In general, choosing a resampling strategy is a subjective choice, and each algorithm provides best results for different kinds of images.

6 RESULTS

We have successfully applied our technique to a large number of natural and synthetic images. Here we show a few representative examples obtained through uniform downscaling (for a reported downscaling factor R, the total number of pixels is reduced by R^2). Non-uniform scaling is straightforward, as discussed in Section 4.2. In all comparisons with other techniques, the images were generated using software provided by their own authors. Finally, we encourage the readers to zoom in on the PDF (full-resolution images embedded) and explore the supplementary materials for these and other examples in their original sizes.

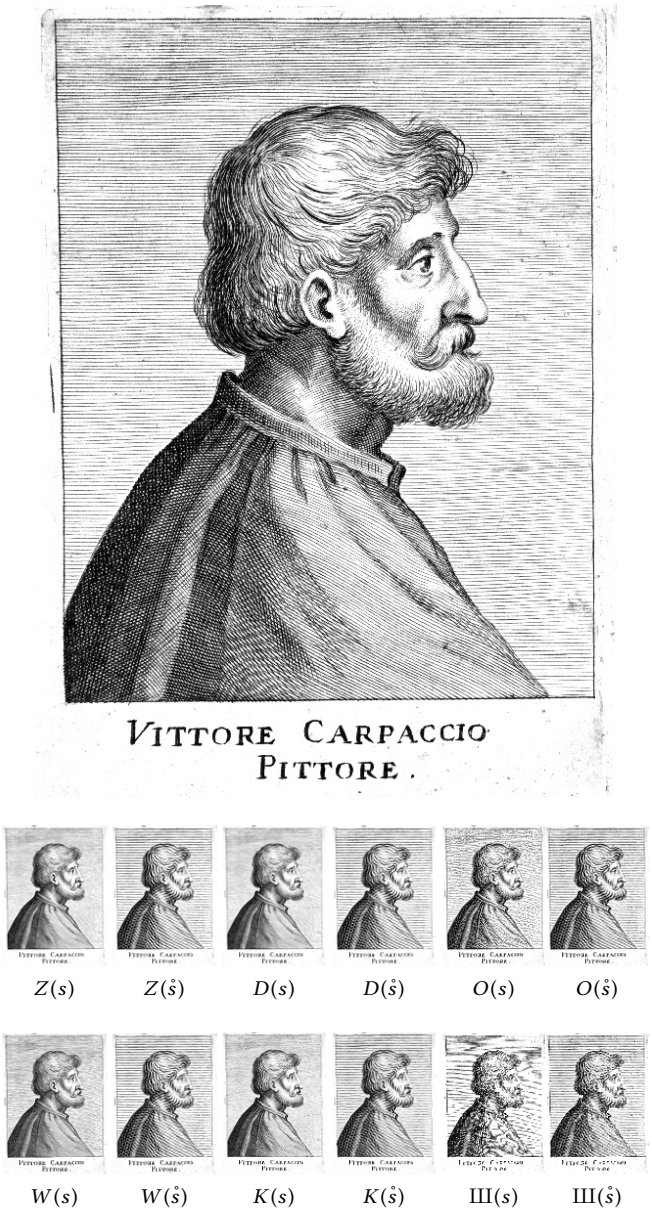

Fig. 8. Several downscaling techniques applied to the image s (top, public domain) and to its spectrally-remapped version $\mathring{s}$. Resampling with: Lanczos (L), O-MOMS (D), Öztireli and Gross (O), Weber et al. (W), Kopf et al. (K), and Subsampling (III). Downscaling to 112×150 pixels ($R = 6$ and $\sigma = 4.5$).

Fig. 1 illustrates the use of our technique to downscale the image on the left to 180×144 pixels (using $R = 4$, $\sigma = 3$). Note that our method provides a faithful representation for the image's high-frequency content: the woman's pants, scarf, books, and table cloth. While this kind of result cannot be obtained with previous techniques, our approach successfully represents the original textures.

Our method handles both natural and synthetic images. Fig. 6 illustrates the spectral-remapping process for a radial chirp centered

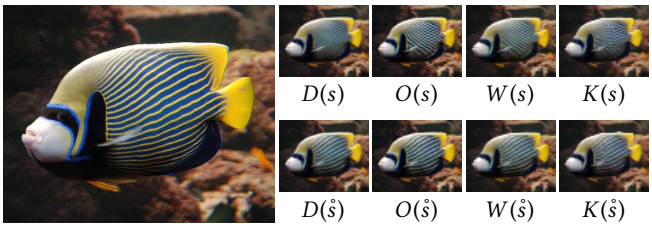

Fig. 9. A fish with wiggly stripes downscaled to 63×50 pixels ($R = 8$ and $\sigma = 6$). (top) Applying the resampling methods to the original image (s) results in alisasing artifacts (stripes in wrong direction and/or checkerboard pattern). (bottom) The same methods produce appropriate results with our spectrally-remapped image ($\mathring{s}$). Original photograph © Monceau on Flickr. Used with permission.

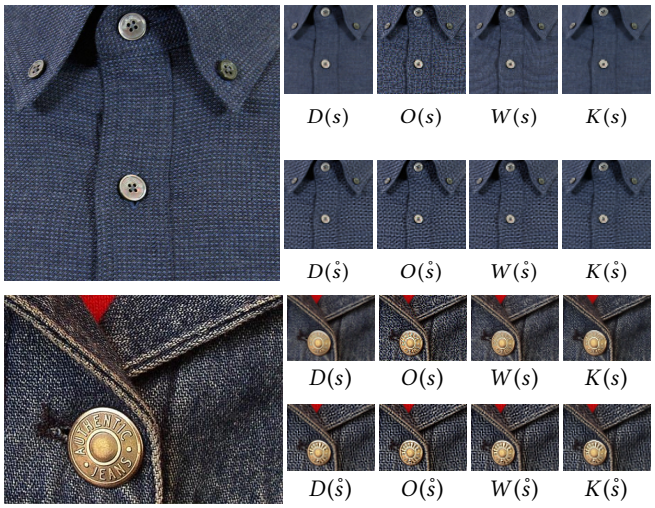

Fig. 10. Retaining very-high frequency details of different kinds of fabrics. (top) A shirt downscaled to 75×75 pixels ($R = 8$ and $\sigma = 8$). (bottom) Some jeans reduced to 80×60 pixels ($R = 5$ and $\sigma = 5$). Our spectral-remapping technique retains the structure of these fabrics, even in such extremely low resolutions. Images from MSRA Salient Object Set [Liu et al. 2007].

at $(0, 0)$ and defined by the equation $cos(x^2 + y^2)$. The corresponding spectrally-remapped image $\mathring{s} = q + \beta r$ is shown on the top right. q is the residue image consisting of all non-remappable waves. In this example, q only contains frequencies inside the spectral circle $\mathfrak{C}$. r consists of all the remapped waves with frequencies on the border of $\mathfrak{C}$, making it appear as a circular wave with constant frequency $\frac{0.4}{R}$. A visual inspection of the power spectra $|\hat{s}|^2$ and $|\hat{\mathring{s}}|^2$ shows that s's high-frequency content has been completely remapped inside $\mathfrak{C}$. Fig. 7 compares the downscaled results of both s and $\mathring{s}$ using four resampling methods: Lanczos (L), Öztireli and Gross (O), Weber et al. (W), and Subsampling (III). When applied to s, Lanczos resampling removes the high-frequency details, while the other techniques introduce severe aliasing artifacts. These same techniques applied to our spectrally-remapped image $\mathring{s}$ produce more faithful representations of the original image s.

Fig. 8 shows a hand-drawn portrait of the Italian painter Vittore Carpaccio. The small images at the bottom contain 112×150 pixels (including a relatively thick white border around the drawing itself). The capital letters under the resized images indicate the used resampling algorithm following the convention in Section 5.3. The symbol inside parentheses indicate whether the resampling was applied to the original image (s) or to our spectrally-remapped image ($\mathring{s}$). The inspection of these thumbnails is quite instructive, as they reveal the robustness of our technique as well as its independence from resampling methods. For instance, Lanczos resampling applied to the original image ($Z(s)$) smooths out salient features, such as the hatching from the background and details in Vittore's hair and beard. However, when applied to the spectrally-remapped image ($Z(\mathring{s})$), Lanczos resampling nicely preserves these features. The same observation applies to Cubic O-MOMS (D). The technique by Öztireli and Gross ($O(s)$) shows some high-frequency content, but the resulting image looks noisy. When applied to $\mathring{s}$ ($O(\mathring{s})$), however, all features are properly preserved. The technique by Weber et al. ($W(s)$) produces a result somewhat in between Lanczos's and Öztireli and Gross's. Again, when applied to our spectrally-remapped image, all relevant features are nicely represented. The result produced by the technique of Kopf et al. ($K(s)$) resembles $Z(s)$ and $D(s)$. Like the previous ones, when used with $\mathring{s}$, it preserves the key aspects of the original hatching. Even subsampling, which tends to introduce severe aliasing artifacts (III(s)), can represent relevant features reasonably well when combined with our technique (III($\mathring{s}$)). For the remaining examples, we limit the comparisons to four resampling methods: Cubic O-MOMS [Nehab and Hoppe 2014], Öztireli and Gross [2015], Weber et al. [2016], and Kopf et al. [2013]. Subsampling tends to produce lower-quality results, and Lanczos's results are somewhat similar to Cubic O-MOMS's.

Fig. 9 shows a fish with a wiggly-striped pattern. The input image, seen on the left, was reduced to only 63×50 pixels. On the top right, one sees the results produced by the four selected resampling methods applied directly to the original image. All of them exhibit severe aliasing artifacts, even Cubic O-MOMS—which makes use of an antialising pre-filter. Such artifacts cause the stripes to be perceived in a different direction and/or as a checkerboard pattern. At the bottom right, we show the results obtained with our technique. Note that the stripes have been properly preserved in all cases.

Fig. 10 illustrates how our technique properly represents even very-high-frequency details, such as the weaving pattern of different kinds of fabrics. The top image shows a shirt, which has been downscaled to 75×75 pixels. The image at the bottom shows some jeans, which has been reduced to 80×60 pixels. While previous techniques cannot preserve the original textures on their own, our spectral remapping (bottom rows on the right) retains the structure of the fabrics. Note, however, that in order to preserve the texture in such extremely low resolutions, our method may make the texture appear bigger relative to other elements (e.g., the buttons). Thus, due to subjective reasons, some people may prefer the downscaled images with texture removed ($D(s)$) or with noise-like texture ($O(s)$).

Our technique works for all kinds of high-frequency details and not only for stripes. Fig. 11 shows a mosaic made of tiny pieces, defining a complex texture (a myriad of thin piece boundaries). The original image has been downscaled to 200×133 pixels, making it

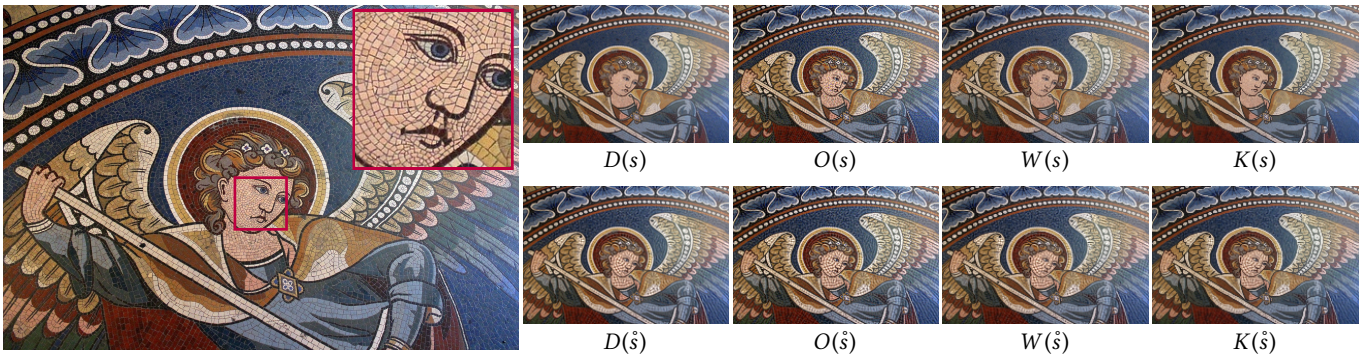

Fig. 11. A mosaic formed by tiny pieces (tesserae) and downscaled to 200×133 pixels ($R = 8$ and $\sigma = 6$). (top) Resampling applied to the original image (s). $D(s)$, $W(s)$, and $K(s)$ cannot capture the fine mosaic texture. $O(s)$ represents it as noise-like structures. (bottom) When applied to a spectrally-remapped image ($\mathring{s}$), all these methods properly represent the texture (see the angel's face). Original photograph courtesy of Lawrence Rice (used with permission).

extremely hard for previous techniques to properly represent such details. In fact, the results for $D(s)$, $W(s)$, and $K(s)$ do not capture the boundaries of these fine structures. The result for $O(s)$ represents them as noise. Using these same techniques with a spectrally-remapped image ($\mathring{s}$), all these methods properly represent the mosaic texture (see the angel's face for all results).

Fig. 12 shows a scene with various high-frequency periodic patterns: the blinds, the small checkerboard on the bear's hat, and the stripes on the face of the dog (sun shining through the blinds). The original image has been reduced to only 60×80 pixels, thus providing a very hard test for all techniques. As one can observe, none of the resampling methods could handle this scene on their own (top right). While Cubic O-MOMS introduces less aliasing in comparison to the others, it does so at the cost of blurring the image. The other techniques introduce severe aliasing in the blinds region, are unable to represent the sunny stripes on the dogs face, and either remove or alias the checkerboard. Our spectral-remapping approach avoids aliasing when combined with all of these techniques. As an important observation in favor of the robustness and effectiveness of our approach, we are able to represent the tiny checkerboard pattern even when resampling with Cubic O-MOMS ($D(\mathring{s})$).

Fig. 13 shows three challenging examples exploring different kinds of high-frequency contents and patterns. (top) A shirt with stripes downscaled to 200×150 pixels. Previous downscaling techniques either remove the stripes ($D(s)$, and $K(s)$), or introduce aliasing artifacts ($O(s)$, $W(s)$, $K(s)$). (middle) "Brothers", hand drawing by artist Tyler Hobbs, reduced to 135×171 pixels. Its thin features (hatching) and blue-noise-like background are not adequately handled by previous techniques. (bottom) Picture of a lizard scaled down to 256×341. Note the high-frequency details on its belly and legs, which cannot be captured by previous approaches. The insets provide zoomed in views of important parts of these images and illustrate the ability of our spectral-remapping approach to faithfully retain these various forms of fine details, even in such small resolutions. Please zoom in to see the details.

Our technique handles structured patterns at multiple scales, as well as arbitrary downscaling factors. Fig. 14 shows a photograph of a Romanesco broccoli, which has been reduced to 102×68 pixels. Due to the broccoli's fractal nature, this example illustrates our method

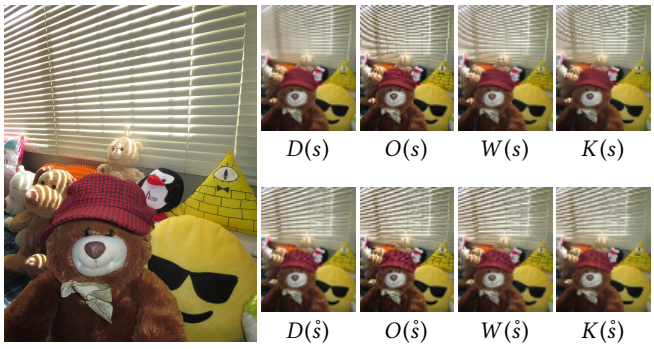

Fig. 12. Scene with various kinds of high-frequency patterns: blinds, checkboard on the bear's hat, and sun stripes on the dog's face. Original image reduced to 60×80 pixels ($R = 20$, $\sigma = 12$). None of the resampling methods are able to handle this scene on their own (top row). Our spectral remapping even allows for the checkerboard to be retained by O-MOMS ($D(\mathring{s})$).

applied to the same structure at multiple scales simultaneously, properly preserving them all. Fig. 15 shows a light-house image downscaled by six different factors ($R \in \{1.35, 2.25, 2.84, 4.2, 5.94, 11\}$) using our technique in combination with Lanczos resampling ($Z(\mathring{s})$). $\sigma = 4$ for all R values, except for $R = 11$, in which case $\sigma = 6$ to guarantee that $\sigma > R/2$. Note how it can properly represent the high-frequency details (e.g., fence, texture of the grass and light-house, parallel lines on the wall of the first house) at all downscaled images, including the smallest one ($R = 11$) with only 47×70 pixels. For comparison, the two images marked with red dots were downscaled ($R = 4.2$) without spectral remapping and exhibit aliasing artifacts or excessive blurring (e.g., see the house on the left).

Fig. 17 shows a camera test-resolution chart downsized from $3,783 \times 2,856$ to 252×190 pixels ($R = 15$ and $\sigma = 15$). Note how our technique preserves fine structured patterns that would otherwise appear blurred or subject to strong aliasing by other techniques.

6.1 Discussion

Despite its apparent complexity (due to the rigor of the underlying formalism used in its derivation), the actual implementation of our

Fig. 13. Challenging examples exploring different kinds of high-frequency contents and patterns. (top) A shirt with stripes downscaled to 200×150 pixels ($R = 8$ and $\sigma = 6$). (middle) "Brothers", hand drawing by artist Tyler Hobbs (used with permission), reduced to 135×171 pixels ($R = 9$ and $\sigma = 9.75$). (bottom) Picture of a lizard (courtesy of William Warby) scaled down to 256×341 ($R = 6$ and $\sigma = 8$). Note the high-frequency details on its belly and legs. None of these examples can be handled by previous approaches. Our spectral-remapping technique retains these fine details even at such a coarse resolution. We recommend viewing the original-sized images in the supplementary materials.

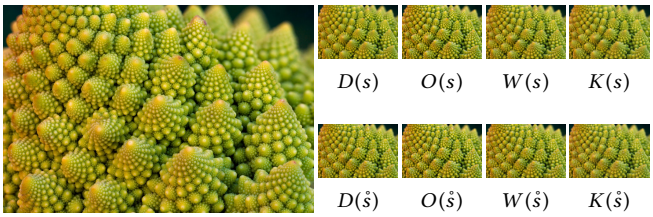

Fig. 14. Photograph of a Romanesco broccoli exhibiting a fractal structure, downscaled to 102×68 pixels ($R = 10$ and $\sigma = 7$). Original photograph, "Romanesco Vortex", by Keld Bach (used with permission).

Fig. 15. Downscaling the same image ("Lighthouse" by Alan Fink, on the left) to several sizes using our spectral remapping and Lanczos resampling ($Z(\mathring{s})$). For these images, $R \in \{1.35, 2.25, 2.84, 4.2, 5.94, 11\}$. Note how structured details are correctly preserved. For comparison, the two images marked with red dots have not been spectrally-remapped before downscaling and exhibit aliasing or excessive blurring. Zoom in to see the details.

method is fairly straightforward. The most complex components are the detection of local waves, and the computation of remapped phases. The detection of local waves only involves standard operations, such as fitting a quadratic function, and finding its maximum (Algorithm 1, Appendix B). The computation of the remapped phases, in turn, requires solving a sparse linear system. Our method is, nevertheless, computationally more intensive than traditional image-downscaling techniques (i.e., low-pass filtering + resampling) and the recent techniques of Öztireli and Gross [2015] and Weber et al. [2016]. Its cost is driven by the Gabor space-frequency decomposition, which is $O(TL \log L)$. However, since each one of the T fast Fourier transforms (in Eq. 11) and T decompositions $\mathcal{W}_{mn}$ (Appendix B) can be computed independently, our technique can be parallelized, bringing its execution time to a theoretical $O(L \log L)$. Our current implementation (written in Julia) has not been optimized nor parallelized, taking the order of seconds to minutes to spectrally-remap an image, depending on its size and spectral contents. For instance, on a 4.2 GHz CPU, it takes 0.75 second to process the chirp image (Fig. 6), 5.9 seconds to process the image in Fig. 1, and 39.8 seconds for Fig. 11. Furthermore, by adjusting the redundancy γ of the Gabor transform, it is possible to control the balance

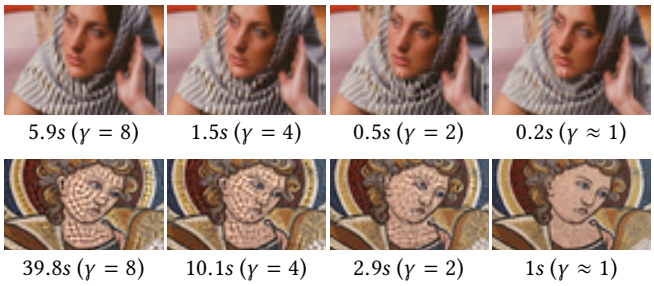

Fig. 16. The redundancy factor γ controls the trade-off between computation time and spectral-remapping numerical accuracy.

between computation time and spectral-remapping numerical accuracy (Fig. 16). With proper optimization and parallelization, our technique might achieve real-time rates. Finally, while the ability to preserve high-frequency details is certainly desirable, there might be situations for which one may prefer not to represent them, due to artistic or subjective reasons.

6.1.1 Limitations. While our technique produces good results in general, it may occasionally over-emphasize some directional component of a structured texture at the expense of others. This is illustrated in Fig. 18, for which the downscaled version of a tiled roof appears as a diagonal texture. Synthetic circular test patterns with spatial frequency increasing towards the center present a challenge for all downscaling algorithms. This is illustrated in Fig. 19. For this example, the original image on the left was reduced from $4{,}725 \times 4{,}725$ to 236×236 pixels ($R = 20$ and $\sigma = 20$). Previous approaches either discard high-frequency information ($D(s)$ and $W(s)$), and/or introduce aliasing artifacts ($O(s)$ and $K(s)$). Our wave alignment, as described in Section 4.3, also fails in this case due to the impossibility of properly aligning such a circular wave of constant remapped frequency $\frac{0.4}{R}$ (middle row). However, by manually introducing a cut in the graph illustrated in Fig. 5 (see supplementary materials for details), our wave alignment system returns a solution that retains the convergence of the stripes at the cost of a visual discontinuity, as shown by the results for "$\mathring{s}$ with cut" (last row). Neither our spectral-remapping nor existing approaches produce optimal results for this challenging example.

7 CONCLUSION AND FUTURE WORK

We presented an image-downscaling technique capable of appropriately representing high-frequency structured patterns. Unlike conventional techniques based on sampling theory, which discard high-frequency information to avoid aliasing, our approach remaps such frequencies to the representable range of the downsampled spectrum. This creates an intermediate image with the same spatial resolution as the original one, which is then resampled to the target dimensions. The resampling process causes the remapped frequencies to be represented as high frequencies in the downscaled image. As such, our approach generates more faithful representations of the original content, retaining visually-important details that would otherwise be lost, or subject to undesirable aliasing artifacts. To achieve such goals, we have introduced new techniques

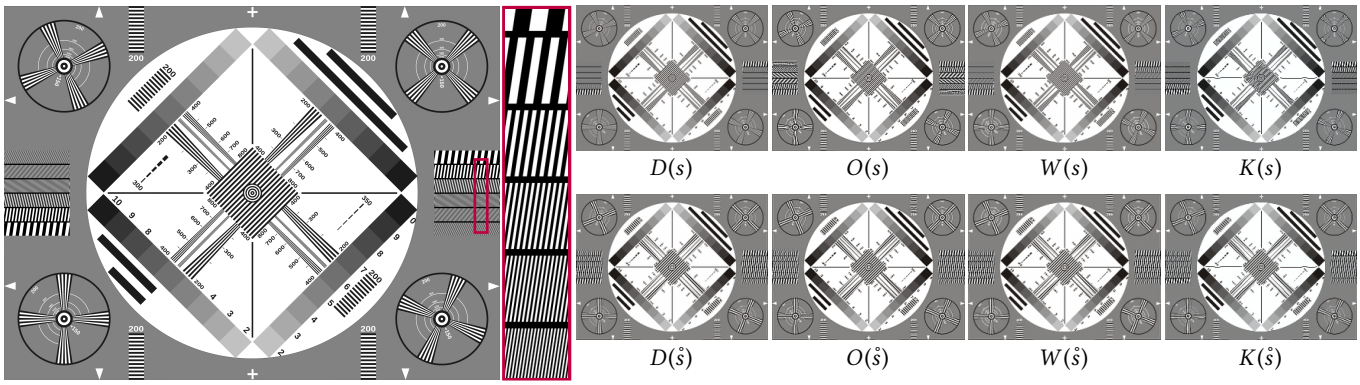

Fig. 17. A camera test-resolution chart (public domain) downscaled from 3,783 × 2,856 to 252 × 190 pixels ($R = 15$ and $\sigma = 15$). Our technique preserves fine structured patterns that would otherwise appear blurred or subject to strong aliasing by other techniques. Due to the low target resolution, our technique maps the converging lines inside the circles to sets of parallel antialiased lines (better seen on the supplementary materials).

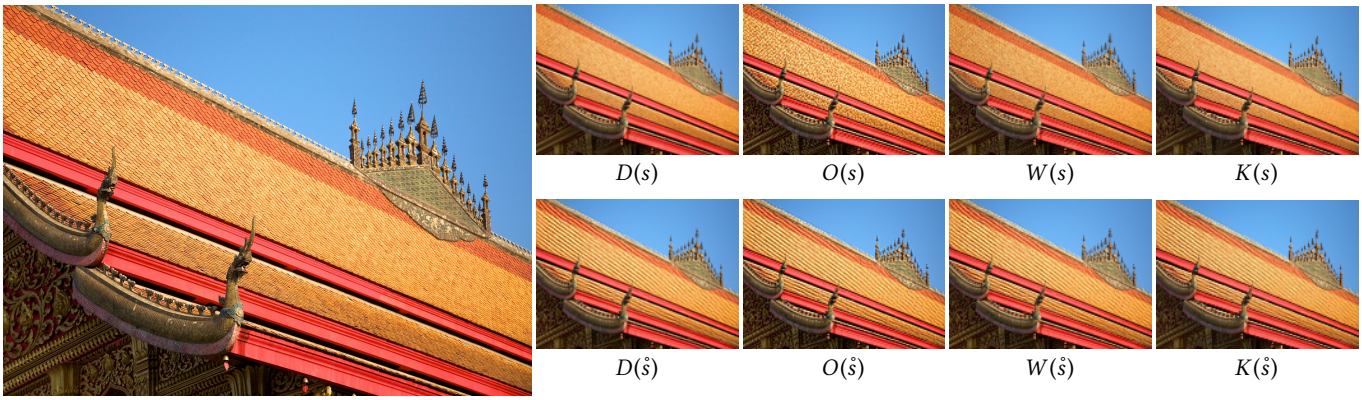

Fig. 18. Roof photograph downscaled from 1,911 × 1,412 to 159 × 118 pixels ($R = 12$ and $\sigma = 10$). For this example, our technique has over-emphasized the diagonal-like pattern observed in the tiled roof. Original photograph from Chi King on Flickr.

and algorithms for estimating the frequency, amplitude, and phase of spatially-localized waves; for performing frequency remapping; for re-aligning the phases of the remapped waves; and for preserving the original image's brightness in its downscaled version. Our technique handles arbitrary-frequency content, is applicable to both natural and synthetic images, and can be used in combination with any resampling method. We have demonstrated its effectiveness on a large number of images downscaled in combination with different resampling strategies.

Our technique is completely automatic. However, as discussed in Section 5, some local regions in the input image s may contain nearby frequencies that require remapping. In this case, best results are obtained with $\sigma > R/2$. While a larger σ value allows for better frequency localization, its corresponding larger Gaussian window affects spatial localization and execution time. Thus, we would like to adaptively compute the smallest σ required to distinguish relevant nearby frequencies for each local region in s. Video downscaling is also an interesting direction for future exploration. While directly applying our technique to individual video frames should be straightforward, we speculate that maintaining temporal coherence for some dynamic non-rigid objects (e.g., a waving flag) may be challenging and require inter-frame phase alignment.

Our technique provides an alternative solution for a long-standing problem in image processing. As such, it has the potential to open up opportunities for further theoretical advances, as well as for the development of new applications.

ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers for their insightful comments, and the authors of [Kopf et al. 2013; Nehab and Hoppe 2014; Öztireli and Gross 2015; Weber et al. 2016] for kindly providing the implementation of their downsampling algorithms. This work was sponsored by CNPq-Brazil (fellowships and grants 306196/2014-0, 403591/2014-7, and 150550/2015-4).

REFERENCES

E. H. Adelson. 1995. Checkershadow Illusion. http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html. (1995). Accessed: 2017-01-05.

T. Blu, P. Thcvenaz, and M. Unser. 2001. MOMS: Maximal-order interpolation of minimal support. *IEEE TIP* 10, 7 (2001), 1069–1080. https://doi.org/10.1109/83.931101

T. A. Davis. 2011. Algorithm 915, SuiteSparseQR: Multifrontal Multithreaded Rank-revealing Sparse QR Factorization. *ACM Trans. Math. Softw.* 38, 1 (2011), 8:1–8:22.

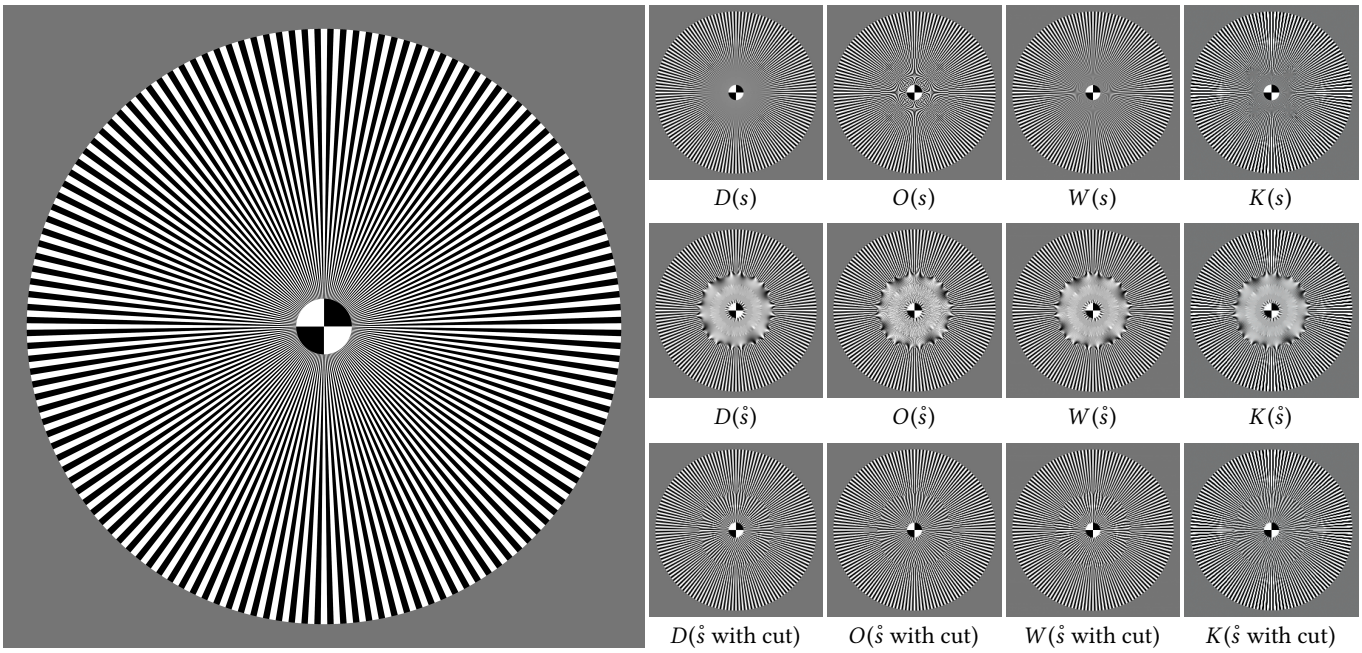

Fig. 19. A synthetic circular test pattern with spatial frequency increasing towards the center presents a challenge for all downscaling algorithms. The image on the left was reduced from 4,725 × 4,725 to 236 × 236 pixels ($R = 20$ and $\sigma = 20$). Previous approaches either discard high-frequency information ($D(s)$ and $W(s)$), and/or introduce aliasing artifacts ($O(s)$ and $K(s)$). Our wave alignment, as described in Section 4.3, also fails in this case (middle row). By manually introducing a cut in the graph illustrated in Fig. 5 (see supplementary materials for details), our wave alignment system returns a solution that retains the convergence of the stripes at the cost of a visual discontinuity—"$\mathring{s}$ with cut" (last row).

D. Gabor. 1946. Theory of Communication. Part 1: The Analysis of Information. *Journal of the IEEE* 93, 26 (Nov. 1946), 429–441. https://doi.org/10.1049/ji-3-2.1946.0074

E. J. Hannan. 1973. The Estimation of Frequency. *Journal of Applied Probability* 10, 3 (1973), 510–519.

F.J. Harris. 1978. On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform. *Proc.of the IEEE* 66, 1 (1978), 51–83.

C. Heil. 2007. History and Evolution of the Density Theorem for Gabor Frames. *Journ. of Fourier Anal. and Applic.* 13, 2 (2007), 113–166.

E. Jacobsen and P. Kootsookos. 2007. Fast, Accurate Frequency Estimators. *IEEE Signal Process. Mag.* 24, 3 (2007), 123–125.

J. Kopf, A. Shamir, and P. Peers. 2013. Content-adaptive Image Downscaling. *ACM TOG* 32, 6, Article 173 (Nov. 2013), 8 pages. https://doi.org/10.1145/2508363.2508370

T. Liu, J. Sun, N. N. Zheng, X. Tang, and H. Y. Shum. 2007. Learning to Detect A Salient Object. In *IEEE CVPR*. 1–8.

J. Mairal, F. Bach, and J. Ponce. 2014. Sparse Modeling for Image and Vision Processing. *Foundations and Trends in C.G. and Vision* 8, 2-3 (2014), 85–283.

S. G. Mallat. 1998. *A Wavelet Tour of Signal Processing* (2nd ed.). Academic Press.

D. P. Mitchell and A. N. Netravali. 1988. Reconstruction Filters in Computer-graphics. In *Proc. SIGGRAPH '88*. 221–228. https://doi.org/10.1145/54852.378514

D. Nehab and H. Hoppe. 2014. A Fresh Look at Generalized Sampling. *Foundations and Trends in C.G. and Vision* 8, 1 (2014), 1–84. https://doi.org/10.1561/0600000053

B. A. Olshausen and D. J. Field. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381 (1996), 607–609.

A. V. Oppenheim and J. S. Lim. 1981. The importance of phase in signals. *Proc. IEEE* 69, 5 (1981), 529–541.

A. Cengiz Öztireli and Markus Gross. 2015. Perceptually Based Downscaling of Images. *ACM TOG* 34, 4, Article 77 (July 2015), 10 pages. https://doi.org/10.1145/2766891

J. Portilla and E. P. Simoncelli. 2000. A Parametric Texture Model based on Joint Statistics of Complex Wavelet Coefficients. *Int'l Journal of Computer Vision* 40, 1 (2000), 49–71.

J. G. Proakis and D. K. Manolakis. 2007. *Digital Signal Processing: Principles, Algorithms, and Applications.* Pearson Education India.

T. Quatieri and Rl McAulay. 1986. Speech Transformations Based on a Sinusoidal Representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 34, 6 (1986), 1449–1464.

B. G. Quinn. 1989. Estimating the Number of Terms in a Sinusoidal Regression. *Journal of Time Series Analysis* 10, 1 (1989), 71–75.

E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley. 2001. Color Transfer Between Images. *IEEE Comput. Graph. Appl.* 21, 5 (2001), 34–41.

D. C. Rife and R. R. Boorstyn. 1976. Multiple Tone Parameter Estimation From Discrete Time Observations. *Bell System Technical Journal* 55 (1976), 1389–1410.

L. Sacht and D. Nehab. 2015. Optimized quasi-interpolators for image reconstruction. *IEEE TIP* 24, 12 (2015), 5249–5259. https://doi.org/10.1109/TIP.2015.2478385

R. Samadani, T. A. Mauer, D. M. Berfanger, and J. H. Clark. 2010. Image Thumbnails That Represent Blur and Noise. *IEEE TIP* 19, 2 (2010), 363–373.

C. E. Shannon. 1949. Communication in the Presence of Noise. *Proc. Institute of Radio Engineers* 37, 1 (1949), 10–21.

J. O. Smith and X. Serra. 1987. PARSHL: An Analysis / Synthesis Program for Non-Harmonic Sounds Based on a Sinusoidal Representation. In *Proc. of the Int'l Computer Music Conference.*

M. Trentacoste, R. Mantiuk, and W. Heidrich. 2011. Blur-Aware Image Downsampling. *Computer Graphics Forum* (2011). https://doi.org/10.1111/j.1467-8659.2011.01894.x

M. Unser. 2000. Sampling-50 Years after Shannon. *Proc. IEEE* 88, 4 (2000), 569–587.

Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE TIP* 13, 4 (2004), 600–612.

N. Weber, M. Waechter, S. C. Amend, S. Guthe, and M. Goesele. 2016. Rapid, Detail-preserving Image Downscaling. *ACM TOG* 35, 6, Article 205 (Nov. 2016), 6 pages.

M. Zibulski and Y. Y. Zeevi. 1994. Frame Analysis of the Discrete Gabor-Scheme. *IEEE Trans. on Sig. Proc.* 42, 4 (1994), 942–945. https://doi.org/10.1109/78.285657

A 2-D FOURIER AND GABOR ANALYSIS

Denote $f \otimes g$ the outer product of vectors: $(f \otimes g)(k,l) = f(k)g(l)$. The spectrum $\hat{s}$ of the 2-D image s defined in Section 4 is given by

$$\hat{s}(k,l) = \langle s \mid f_k^M \otimes f_l^N \rangle, \quad \text{for} \quad k = 0..M-1,\ l = 0..N-1;$$

where $f_k^{\square}(n) = \frac{1}{\sqrt{\square}} \exp(2\pi i\, n\, k/\square)$. Similarly, its Gabor decomposition $\check{s}$ is a quadruple-indexed vector:

$$\check{s}(m,n,k,l) = \langle s \cdot g_{mn} \mid f_k^M \otimes f_l^N \rangle = \widehat{s \cdot g_{mn}}(k,l),$$

where $g_{mn} = g_m \otimes g_n$ is the 2-D Gaussian window.

B ALGORITHM FOR DETECTING LOCAL WAVES

Our goal is to find the set $\mathcal{W}_{mn}$ that minimizes $E(\mathcal{W}_{mn})$ in Eq. 5. A perfect decomposition $\mathcal{W}^\star_{mn}$ yields $E(\mathcal{W}^\star_{mn}) = 0$, which gives

$$s \cdot g_{mn} \;=\; \sum_{(\alpha,\mathrm{a},\mathrm{b},\mathrm{c})\in\mathcal{W}^\star_{mn}} w_{\alpha,\mathrm{a},\mathrm{b},\mathrm{c}} \cdot g_{mn}.$$

We may expand both sides of this equality in the Fourier basis, resulting in the well-known equivalence between the convolution product $*$ and windowing [Harris 1978]:

$$\widehat{s \cdot g_{mn}} \;=\; \sum_{(\alpha,\mathrm{a},\mathrm{b},\mathrm{c})\in\mathcal{W}^\star_{mn}} \widehat{w_{\alpha,\mathrm{a},\mathrm{b},\mathrm{c}} \cdot g_{mn}} \;=\; \sum_{(\alpha,\mathrm{a},\mathrm{b},\mathrm{c})\in\mathcal{W}^\star_{mn}} \widehat{w}_{\alpha,\mathrm{a},\mathrm{b},\mathrm{c}} * \widehat{g}_{mn}. \quad (12)$$

For a moment, let us consider all sequences as continuous functions, since this simplifies the math. In this case, $\widehat{w}_{\alpha,\mathrm{a},\mathrm{b},\mathrm{c}}$ is the Fourier transform of the wave function $w_{\alpha,\mathrm{a},\mathrm{b},\mathrm{c}}$, given by

$$\widehat{w}_{\alpha,\mathrm{a},\mathrm{b},\mathrm{c}}(\xi,\nu) = \alpha\,\delta(\mathrm{a}-\xi)\,\delta(\mathrm{b}-\nu)\,\exp(2\pi i\mathrm{c}),$$

where δ is the Dirac delta. The convolution in Eq. 12 becomes

$$G_{\mathrm{ab}}(\xi,\nu) = (\widehat{w}_{\alpha,\mathrm{a},\mathrm{b},\mathrm{c}} * \widehat{g}_{mn})(\xi,\nu) = \widehat{g}_{mn}(\mathrm{a}-\xi,\mathrm{b}-\nu)\,\alpha\exp(2\pi i\mathrm{c}), \quad (13)$$

where $\widehat{g}_{mn}$ is the Fourier transform of the Gaussian window g, whose absolute value $|\widehat{g}_{mn}(\mathrm{a}-\xi,\mathrm{b}-\nu)|$ is also a Gaussian with peak at (a,b). Thus, the right-hand side of Eq. 12 decomposes $\widehat{s\cdot g_{mn}}$ as a sum of functions G_{ab} (each corresponding to a wave $(\alpha,\mathrm{a},\mathrm{b},\mathrm{c})$), such that $|G_{\mathrm{ab}}(\xi,\nu)|$ has a Gaussian peak at $(\xi,\nu) = (\mathrm{a},\mathrm{b})$. This conclusion leads to a simple **algorithm for finding the frequencies** (a,b) **of the waves in** $\mathcal{W}^\star_{mn}$, which consists of *searching for Gaussian peaks in* $|\widehat{s\cdot g_{mn}}|$ [Jacobsen and Kootsookos 2007; Smith and Serra 1987].

B.1 The Decomposition Algorithm

Back to periodic complex sequences s, let $S_{mn} = \widehat{s\cdot g_{mn}}$ be the DFT of $s\cdot g_{mn}$. $S_{mn}(k,l)$ is a double-indexed vector with the same number of elements as s, where $k = 0..M-1$ and $l = 0..N-1$. Starting with an empty set $\mathcal{W}_{mn} = \varnothing$, we iteratively find the waves that should be added to it. Algorithm 1 summarizes the process, whose steps are detailed next.

Step 1. Find the integer indices (k',l') where $|S_{mn}(k',l')|$ is maximal. Note that we are interested in detecting the *non-representable* waves that exist in $\mathcal{W}^\star_{mn}$, since these are the ones that require frequency remapping (see Section 4.2). Thus, we limit the search to indices (k',l') associated with non-representable harmonics whose frequencies fall outside of the spectral circle $\mathfrak{C}$ (see Section 4.2).

Step 2. Find the frequencies $(\mathrm{a}',\mathrm{b}')$ such that $G_{\mathrm{a}'\mathrm{b}'}$ gives the best possible fit to the neighborhood of $|S_{mn}|$ around (k',l'). Since $|G_{\mathrm{a}'\mathrm{b}'}|$ is Gaussian, its logarithm is a parabola with a maximum at $(\mathrm{a}',\mathrm{b}')$. Thus, inspired by Smith and Serra [1987], we find the quadratic surface $Q(k,l)$ that gives the best least-squares fit to $\log(|S_{mn}|)$ in the 3×3 neighborhood centered at (k',l'). $(\mathrm{a}',\mathrm{b}')$ is then the frequency location of the maximum of Q (if it exists; see the numerical considerations below). Furthermore, the height of the surface at its maximum directly gives the amplitude of the wave:

$$\alpha' = \frac{\exp(Q(\mathrm{a}',\mathrm{b}'))}{\hat{g}_{mn}(0,0)}. \quad (14)$$

Algorithm 1 Detecting Local Waves

```
W_mn ← ∅                                   # Initialize set of waves locally describing s
S_mn ← DFT(s̄_mn · g_mn)
α_max ← 0
while |W_mn| < 10 do
    # Find indices of maximum value in |S_mn|
    (k', l') ← arg max_{(k',l')∉𝔠} |S_mn(k', l')|                    # Step 1
    # Compute parameters (α', a', b', c') of the wave
    fit quadratic Q to log(|S_mn|) around (k', l')                  # Step 2
    break if Hessian of Q is not negative definite
    compute frequencies (a', b') ← arg max_{(a',b')} Q(a', b')
    break if (a', b') falls outside 3 × 3 neighborhood of (k', l')
    compute amplitude α' using Eq. 17
    break if α' < (1/4) α_max; else α_max ← max(α', α_max)
    compute phase c' using Eq. 15                                   # Step 3
    # Remove detected wave from S_mn
    S_mn ← S_mn − ŵ_{α',a',b',c'} · g_mn                            # Step 4
    break if the energy ||S_mn||^2 did not decrease
    W_mn ← W_mn ∪ {(α', a', b', c')}                                # Add wave to W_mn
end while
return W_mn
```

The exponential is needed as Q is fitted to the logarithm of $|G_{\mathrm{a}'\mathrm{b}'}|$. The division by $\hat{g}_{mn}(0,0)$ compensates for the multiplication of α by the Gaussian $\hat{g}_{mn}$ in the rightmost part of Eq. 13. Note that $|G_{\mathrm{a}'\mathrm{b}'}(\mathrm{a}',\mathrm{b}')| = \hat{g}_{mn}(0,0)\,\alpha = \exp(Q(\mathrm{a}',\mathrm{b}'))$.

Step 3. At this point, to reconstruct the wave $(\alpha',\mathrm{a}',\mathrm{b}',\mathrm{c}')$, the only missing parameter is its phase $\mathrm{c}' \in [0,1)$. To obtain it, we use the following program which has a closed form solution in c', where $W_{mn} = \widehat{w_{\alpha',\mathrm{a}',\mathrm{b}',0}\cdot g_{mn}}$ (note the zero phase in wave w):

$$\begin{aligned}
&\min_{\mathrm{c}'}\ \|S_{mn} - \widehat{w_{\alpha',\mathrm{a}',\mathrm{b}',\mathrm{c}'}\cdot g_{mn}}\|^2 && (15)\\
\equiv\ &\min_{\mathrm{c}'}\ \|S_{mn} - \exp(2\pi i\mathrm{c}')\,W_{mn}\|^2 && \text{(by Eqs. 12 and 13)}\\
\equiv\ &\min_{\mathrm{c}'}\ \|S_{mn}\|^2 + \|W_{mn}\|^2 - 2\,\Re\exp(2\pi i\mathrm{c}')\,\langle W_{mn}\,|\,S_{mn}\rangle\\
\equiv\ &\max_{\mathrm{c}'}\ \Re\exp(2\pi i\mathrm{c}')\,\langle W_{mn}\,|\,S_{mn}\rangle\\
\equiv\ &\max_{\mathrm{c}'}\ \cos(2\pi\mathrm{c}')\,\Re\langle W_{mn}\,|\,S_{mn}\rangle - \sin(2\pi\mathrm{c}')\,\Im\langle W_{mn}\,|\,S_{mn}\rangle\\
\equiv\ &\max_{\mathrm{c}'}\ F(\mathrm{c}').
\end{aligned}$$

The objective F is continuous for c' in $[0,1]$, and thus attains a maximum in this interval. It is also differentiable on the whole real line, with critical points c_1 and c_2 in $[0,1)$ given by

$$\begin{aligned}
c_1 &= \mathrm{atan2}(-\,\Im\langle W_{mn}\,|\,S_{mn}\rangle,\ \ \Re\langle W_{mn}\,|\,S_{mn}\rangle)/2\pi \mod 1,\\
c_2 &= \mathrm{atan2}(\ \ \Im\langle W_{mn}\,|\,S_{mn}\rangle,\ -\Re\langle W_{mn}\,|\,S_{mn}\rangle)/2\pi \mod 1.
\end{aligned}$$

The solution is $\mathrm{c}' = c_1$ if $F(c_1) > F(c_2)$, or $\mathrm{c}' = c_2$ otherwise.

Step 4. Add the detected wave $(\alpha',\mathrm{a}',\mathrm{b}',\mathrm{c}')$ to $\mathcal{W}_{mn}$. If the stopping criterion has *not* been reached (see below), we remove

such wave from the spectrum by the update rule $S_{mn} \leftarrow S_{mn} - \widehat{w_{\alpha',a',b',c'} \cdot g_{mn}}$, and restart from Step 1. In the end, the set $\mathcal{W}_{mn}$ will contain a suitable decomposition of $s \cdot g_{mn}$, according to Eq. 5.

B.1.1 Stopping Criteria and Numerical Considerations. The waves with largest amplitudes contribute the most to the image's local contrast, a principal stimulus sensed by the human visual system [Adelson 1995]. Since Step 1 of the decomposition algorithm searches for maximal coefficients in $|S_{mn}|$, larger amplitude waves are, luckily, the first to be detected. So, we stop the search whenever the current wave's amplitude is smaller than $\frac{1}{4}$-th of the maximum detected amplitude so far (since it and all subsequent waves will contribute little energy to the local contrast). Furthermore, to have an upper bound on the total computation time, we limit the size of each set $\mathcal{W}_{mn}$ to a maximum of 10 waves, a number that works well in practice. Recall that each set $\mathcal{W}_{mn}$ only provides a local description of s (i.e., a description to a windowed portion of s around pixel $s(m,n)$), and the total number *of sets* grows with the number of pixels in s.

We *stop the search* for Gaussian peaks in S_{mn} if: (i) the Hessian of Q is *not* negative definite, meaning that Q has no maximum and thus is not a good estimator for a Gaussian peak; (ii) Q has a maximum, but it falls outside of the 3×3 neighborhood of (k', l'), meaning that (k', l') is probably a small noise peak; or (iii) if the energy of S_{mn} *fails to decrease* after the wave's removal (i.e., $\|S_{mn} - \widehat{w_{\alpha',a',b',c'} \cdot g_{mn}}\|^2 \geq \|S_{mn}\|^2$), meaning that the decomposition given by $w_{\alpha',a',b',c'}$ is highly non-orthogonal and badly conditioned, in which case we do not add $w_{\alpha',a',b',c'}$ to $\mathcal{W}_{mn}$. Note that whenever the waves in $\mathcal{W}^\star_{mn}$ form an orthogonal system, then the energy necessarily decreases since $\|S_{mn} - \widehat{w \cdot g_{mn}}\|^2 = \|S_{mn}\|^2 - \|\widehat{w \cdot g_{mn}}\|^2 < \|S_{mn}\|^2$. Finally, we start the decomposition algorithm with $S_{mn} = \widehat{\bar{s}_{mn} \cdot g_{mn}}$, where $\bar{s}_{mn}$ is the mean-centered signal from Eq. 8. This removes the spectral energy of the 0-th harmonic (mean intensity of the image) [Mairal et al. 2014], leading to better numerical stability.

Finally, digital images have pixels that are real, not complex. Thus, waves in $\mathcal{W}^\star_{mn}$ necessarily occur as conjugate pairs:

$$(\alpha, \mathrm{a}, \mathrm{b}, \mathrm{c}) \in \mathcal{W}^\star_{mn} \iff (\alpha, -\mathrm{a}, -\mathrm{b}, -\mathrm{c}) \in \mathcal{W}^\star_{mn}. \tag{16}$$

Since the DFT is defined as the periodic summation of the (sampled) continuous transform in Eq. 13, there may occur noticeable "self-leakage between positive and negative frequencies" [Harris 1978]. Thus, as the periodic copies of the Gaussian $G_{-\mathrm{a}-\mathrm{b}}$ overlap with the center of G_{ab}, they may drastically affect the computed value for the peak of G_{ab}, and vice versa. Since the peaks of $G_{-\mathrm{a}-\mathrm{b}}$ and G_{ab} are spaced apart by $(-2\mathrm{a}, -2\mathrm{b})$, the nine periodic copies of $G_{-\mathrm{a}-\mathrm{b}}$ closest to G_{ab} are spaced from G_{ab} by $(-2\mathrm{a} + \delta_\mathrm{a}, -2\mathrm{b} + \delta_\mathrm{b})$, for $\delta_\mathrm{a}, \delta_\mathrm{b} \in \{-1, 0, 1\}$. Thus, for computing α', we replace Eq. 14 by

$$\alpha' = \frac{\exp\left(Q(\mathrm{a}', \mathrm{b}')\right)}{\hat{g}_{mn}(0,0) + \sum_{\delta_\mathrm{a}, \delta_\mathrm{b} \in \{-1,0,1\}} |\hat{g}_{mn}(-2\mathrm{a}' + \delta_\mathrm{a}, -2\mathrm{b}' + \delta_\mathrm{b})|}. \tag{17}$$

The partial sum in the denominator gives a good approximation of the true periodization of Eq. 13, and effectively corrects any noticeable discrepancies in the detected waves' amplitudes. As such, it greatly improves the quality of the reconstructed images. Note the absolute values in the terms of the sum, since, except for $\hat{g}_{mn}(0,0)$, all values of $\hat{g}_{mn}$ have an imaginary part.

C COMPUTING β

We wish to find $\beta > 0$ such that $\|\mathring{s}\| = \|s\|$ for $\mathring{s} = q + \beta r$. Since the norm $\|\cdot\| = \sqrt{\langle \cdot \,|\, \cdot \rangle}$ is a positive function:

$$\|q + \beta r\| = \|s\| \iff C + B\beta + A\beta^2 = 0, \tag{18}$$

where $A = \|r\|^2$, $B = 2\,\Re\langle q \,|\, r\rangle$, and $C = \|q\|^2 - \|s\|^2$. The solution of this quadratic equation falls into a few cases.

Case 1 ($C = 0$). If non-representable and non-ensemble waves do not exist in s, then $\mathcal{W}_{mn}$ is empty $\forall\ (m,n)$, resulting in $A = B = C = 0$. Thus, any value of β gives $\|\mathring{s}\| = \|s\|$ since $\mathring{s} = s$.

Case 2 ($C < 0$). If $C < 0$, then $AC < 0$ and the quadratic in Eq. 18 has two real roots β_1 and β_2 since $\Delta = B^2 - 4AC > 0$ and

$$\beta_1 = \frac{-B + \sqrt{\Delta}}{2A}, \quad \beta_2 = \frac{-B - \sqrt{\Delta}}{2A}.$$

Furthermore, β_1 is positive since $\sqrt{\Delta} > \sqrt{B^2} = |B|$ and thus $-B + \sqrt{\Delta} > -B + |B| \geq 0$. β_2 is negative since the quadratic in Eq. 18 may be written as $A(\beta - \beta_1)(\beta - \beta_2)$, which implies that $A\beta_1\beta_2 = C < 0$. Thus, the solution we are looking for is $\beta = \beta_1$.

Case 3 ($C > 0$). C will never be positive in exact arithmetic since we stop the algorithm described in Appendix B whenever $\|S_{mn}\|$ fails to decrease, meaning that necessarily $\|q\|^2 < \|s\|^2$. In floating point arithmetic, however, C may become positive. In this case, the wave decomposition is most likely noise, and thus we set $\mathring{s} = s$.

D DUAL-FRAME COMPUTATION

Since the vectors $g_{n\tau,k}$ in Eq. 11 define a non-tight frame, the synthesis of s from the coefficients $\langle s \,|\, g_{n\tau,k}\rangle$ requires the use of a dual frame $\tilde{g}_{n\tau,k}$. Fortunately, such vectors are similarly given by spatial and frequency translations of a *dual window* $\tilde{g}$:

$$\tilde{g}_{n\tau,k}(m) = \tilde{g}(m - n\,\tau)\,\frac{1}{\sqrt{L}}\,\exp\left(2\pi i\, m\, k/L\right). \tag{19}$$

Denote $\tilde{g}_{n\tau}(m) = \tilde{g}(m - n\,\tau)$ and $f_k = \frac{1}{\sqrt{L}}\,\exp\left(2\pi i\, m\, k/L\right)$. We may re-write Eq. 11 as:

$$\begin{aligned} s &= \sum_{n=0}^{T-1}\sum_{k=0}^{L-1} \langle s \,|\, g_{n\tau,k}\rangle\, \tilde{g}_{n\tau,k} = \sum_{n=0}^{T-1}\sum_{k=0}^{L-1} \langle s \,|\, g_{n\tau}\cdot f_k\rangle\, \tilde{g}_{n\tau}\cdot f_k \\ &= \sum_{n=0}^{T-1} \tilde{g}_{n\tau}\cdot \sum_{k=0}^{L-1} \langle s\cdot g_{n\tau} \,|\, f_k\rangle\, f_k = \sum_{n=0}^{T-1} s\cdot g_{n\tau}\cdot \tilde{g}_{n\tau}. \end{aligned} \tag{20}$$

Similarly, given the knowledge of a tight frame $h_{n\tau,k} = \tilde{h}_{n\tau,k}$ (defined by the spatial and frequency translations of a *self-dual* window $\tilde{h} = h$), we may also write

$$s = \sum_{n=0}^{T-1}\sum_{k=0}^{L-1} \langle s \,|\, h_{n\tau,k}\rangle\, h_{n\tau,k} = \sum_{n=0}^{T-1} s\cdot h_{n\tau}\cdot h_{n\tau}. \tag{21}$$

One way of obtaining h from g is [Mallat 1998, Thm. 5.8]:

$$h(m) = \frac{g(m)}{\sqrt{\sum_{n\in\Lambda_m} g(m - n\,\tau)^2}}, \quad m = -\lfloor L/2\rfloor .. \lfloor L/2\rfloor,$$

where the index set $\Lambda_m = \{n \in \mathbb{Z} : -\lfloor L/2\rfloor \leq m - n\tau \leq \lfloor L/2\rfloor\}$. Finally, since Eqs. 20 and 21 imply $(g\cdot\tilde{g})_{n\tau} = g_{n\tau}\cdot\tilde{g}_{n\tau} = h_{n\tau}\cdot h_{n\tau} = (h\cdot h)_{n\tau} = (h^2)_{n\tau}$, we obtain $\tilde{g} = h^2 \div g$, where $\div$ denotes elementwise vector division.