DOI: 10.1111/cgf.13942
EUROGRAPHICS 2020 / U. Assarsson and D. Panozzo
(Guest Editors)
*Volume 39 (2020), Number 2*

# Prefilters for Sharp Image Display

Luís Cláudio Gouveia Rocha† Manuel M. Oliveira† Eduardo S. L. Gastal†

Instituto de Informática – UFRGS

**(a)** *Mitchell-Netravali [MN88]* **(b)** *OQI3 [SN15]* **(c)** *Our SBS3 filter*

**Figure 1:** *Path-traced images rendered using various prefilters: (a) Mitchell-Netravali, (b) OQI3, and (c) Our SBS3. Our family of SBS3 filters strike a good balance between sharpness, aliasing and ringing, consistently yielding images that are sharper than both classic (a) and state-of-the-art linear filters (b). Note how the fine details from the wings and face of the dragon are better emphasized by SBS3 (c). This result has been computed for a baseline viewing distance of 40 cm on a 100 ppi display (better seen in the supplementary materials since PDF readers perform resampling).*

**Abstract**

*In this paper we use a simplified model of the human visual system to explain why humans tend do prefer "sharpened" digital images. From this model we then derive a family of image prefilters specifically adapted to viewing conditions and user preference, allowing for the trade-off between ringing and aliasing while maximizing image sharpness. We discuss how our filters can be applied in a variety of situations ranging from Monte Carlo rendering to image downscaling, and we show how they consistently give sharper results while having an efficient implementation and ease of use (there are no free parameters that require manual tuning). We demonstrate the effectiveness of our simple sharp prefilters through a user study that indicates a clear preference to our approach compared to the state-of-the-art.*

**CCS Concepts**

• ***Computing methodologies*** → ***Image processing;***

## 1. Introduction

A natural image maps continuous $(x,y)$ coordinates to colors. Modern display devices, on the other hand, are discrete machines, composed of millions of tiny light sources that may be individually controlled. How does one select the intensity and color of each individual light source to obtain a good representation of the original image? The standard approach is to *low-pass filter* the image function to remove "high frequencies," and then *sample* the result at a uniform grid of points [Bli89]. One then assigns the color of each sample to each corresponding light source on the display (Figure 2(b)).

This procedure is based on the sampling theorem [Sha49], which

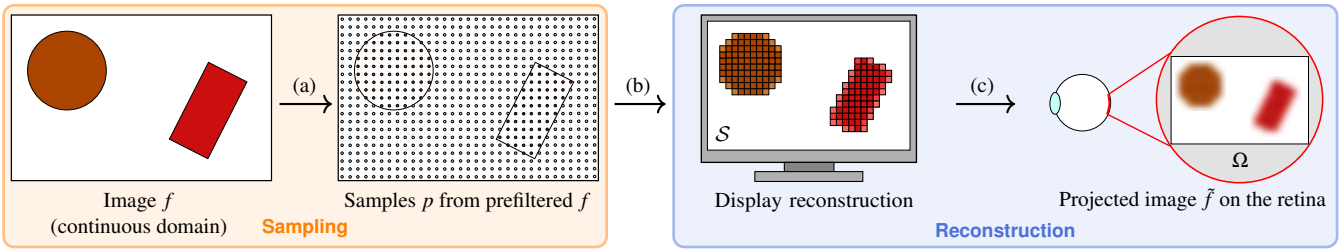



**Figure 2:** *Sampling and Reconstruction pipeline. (a) An image $f$ of a continuous domain (*a spatially-continuous image*) is converted to a discrete collection of samples $p$ ("pixels"). This process often involves a* prefiltering *of $f$ before sampling values from it. (b) A display device reconstructs a spatially-continuous image at a viewing screen $\mathcal{S}$ from the samples $p$. (c) The reconstructed image at $\mathcal{S}$ is projected into the observer's eye onto the retina $\Omega$ [FG88], generating the final reconstruction $\tilde{f}$, whose exact form depends on both the display and the observer.*

states that if a function is bandlimited and contains no frequencies higher than $W$ cycles per unit of measurement, then it is completely determined by samples spaced $1/2W$ units apart. This is a remarkable theorem, but its practical limitations have been widely known and studied for the past decades [Uns00]. In particular: *(i)* the filter that perfectly removes high frequencies without attenuating low frequencies is $\mathrm{sinc}(x) = (\sin x)/x$, which has an infinite impulse response (IIR) and is impossible to evaluate exactly in most situations; *(ii)* if otherwise one uses a non-perfect low-pass filter, it leaves behind residual high-frequency information that may lead to *aliasing* artifacts; *(iii)* even if achievable, one should not use the sinc filter since bandlimited functions often do not look good: they contain visually distracting "ringing" artifacts [ST85]; and *(iv)* to obtain a perfect reconstruction of the (bandlimited) *spatially-continuous function* (i.e., a total function of a continuous domain) from the samples, one must also use the sinc function for interpolation; otherwise, *post-aliasing* or reconstruction artifacts may appear [MN88].

Item *(iv)* in particular is a fundamental issue. It implies that, by sampling a low-pass filtered function and assigning the color of each sample to the light sources on the display, *one is implicitly assuming that reconstruction will be performed by sinc interpolation*. This is a problem because such an assumption is incorrect. To see why, we first define what we mean by "reconstruction" in this context: the optical projection of the spatially-continuous lightfield generated by the display on the retina of the observer [MN88] (illustrated on the right of Figure 2). It is performed by the display in conjunction with the optical system of the human eye, and is most definitely *not* a sinc interpolation, as can be demonstrated by a simple visual experiment we detail in Appendix A. As a result, most images generated by the standard sampling-theorem-based approaches end up excessively blurry or "soft" when *viewed* on modern displays (Section 4). This is illustrated in Figure 1(a).

To address this issue, we propose an alternative approach for computing the correct pixel intensities that should be displayed on the screen in order to maximize sharpness and high-frequency details. It is based on a simplified visual model which takes into account the low-pass filter inherent to our optical system [ZW97, AS00]. We describe how to compute the low-pass bandwidth of such a system (Appendix B) and from this we derive the mathematically optimal filters that should be used to generate sharp and detail-rich images in a variety of applications, including Monte Carlo rendering (Figure 1), image downscaling (Figure 13) and vector graphics rasterization (Figure 15). We show how this framework can be combined with existing state-of-the-art prefilters to take into account user preference and the characteristics of different applications, allowing for a trade-off between ringing and aliasing while maximizing image sharpness. We call our sharp prefilters **SBS3** since they are modeled using cubic splines.

The **contributions** of our work include:

- An alternative prefilter for computer graphics (SBS3) that generates sharp and detailed images for display, without requiring manual parameter tuning (Section 3). It is based on modern sampling theory, can be adapted for specific viewing conditions, and its spectral characteristics (Figure 5) are significantly different from traditional prefilters (Figure 6);
- An in-depth analysis and discussion of the properties and limitations of the proposed technique, and on how it relates to existing filters and theory (Section 4). We also propose a quantitative metric for measuring Sharpness, Aliasing and Ringing characteristics of linear prefilters (Section 4.1.2);
- A collection of prefilters based on SBS3 (Section 4.2), allowing for a wide range of filtering characteristics aimed at different applications and user preference;
- A list of considerations of how to apply the proposed prefilter in a variety of situations and applications (Sections 4.3, 4.4 and 5). We provide source code in the supplementary materials, making our results fully reproducible.

## 2. Background and Related Work

The reconstruction kernel used by a digital display is not a sinc. Kajiya and Ullner [KU81] argue that a *prefilter* (or *analysis filter*) should take the display's *reconstruction* (or *synthesis*) *kernel* into account in order to produce the best possible images. For the CRT display technology of the time, they approximated the reconstruction kernel with Gaussian "spots" and found optimized images by least squares. The authors reported mainly on the use of the technique for rendering text, as individual characters could be prefiltered and stored prior to display. Nonetheless, they did emphasize that the proposed method was "applicable—at least in theory—to the general problem of antialiasing arbitrary images."

† lcgrocha@inf.ufrgs.br, oliveira@inf.ufrgs.br, eslgastal@inf.ufrgs.br

We are not aware of subsequent efforts to advance the idea of Kajiya and Ullner to modern display devices. One reason for this may be the fact that, on recent LCD displays, the shape of individual light sources are tiny rectangles (Figure 2), meaning that such a display is reasonably approximated by a box reconstruction kernel [NH14]. Thus, theoretically, the optimal (in the least-squares sense) prefilter should also be a box function, as it is self-dual (Section 3.2) [Uns00]. This is a disheartening conclusion since a box filter has neither good antialiasing nor good reconstruction properties [Bli89].

Our work follows in the footsteps of Kajiya and Ullner [KU81] but goes one step further: we argue that reconstruction should be viewed as the interaction between the lightfield emitted by the display device and the human optical system (Section 3). From this, we obtain sharp prefilters for the display of detailed images, as illustrated in Figure 1(c).

### 2.1. Linear low-pass filters in graphics

Avoiding aliasing and ringing artifacts are conflicting requirements, and the design of a good low-pass filter for graphics has traditionally involved a subjective interplay among aliasing, ringing, and blurring [MN88]. The box, tent and Gaussian filters are simple and widely used, but unfortunately their frequency characteristics are also far from optimal [Bli89]. Finite approximations to the "ideal" sinc filter have been proposed based on windowing functions [Har78] and compact polynomial kernels [MN88]. More recent works design infinite impulse response filters which are evaluated efficiently using compact polynomials and recursive digital filtering [Uns00], called *generalized sampling* filters. Nehab and Hoppe [NH14] present a detailed survey on this concept, where a noteworthy representative prefilter is the O-MOMS (Optimal, Maximal Order Minimum Support) family of Blu et al. [BTU01]. In a similar way, Sacht and Nehab [SN15] derived Optimized Quasi-Interpolator filters by searching over available degrees of freedom in the polynomial kernel and recursive filter. Their cubic scheme (OQI3) is the state-of-the-art in approximation quality, with good antialiasing properties at the cost of mild ringing. In Section 4.2 we show how our SBS3 prefilter can be combined with OQI3 to adapt the filtering characteristics for specific applications.

### 2.2. Non-linear filters in graphics

Linear filters are efficient to compute and easy to analyze with spectral tools. Non-linear filters are more complex to design and understand, but are able to overcome some of the limitations imposed by the linearity constraint. Next, we list some non-linear filters relevant to our discussion.

In the context of image downscaling, several non-linear filters have been proposed to preserve image detail. For example, Kopf et al. [KSP13] use kernels that adapt themselves to image content, Öztireli and Gross [OG15] use an analytical solution for minimizing the SSIM difference between the original and downscaled image, and Weber at al. [WWA*16] use a bilateral-like filter that favors pixels that differ more from their vicinity, preserving visually-important details. We compare our prefilter against these works in Section 5.2.

In a different direction, Gastal and Oliveira [GO17] proposed a method to modify the high frequency content of the image such that no aliasing will occur after downscaling. Their method is orthogonal to ours and other resampling strategies, and it can be used together with our SBS3 prefilter for sharp and aliasing-free image display.

Edge-aware smoothing filters manage to remove some high-frequency content while preserving edge detail [TM98, FFLS08, GO11, HST13, ZSXJ14, Fat09], but are not designed as prefilters and should not be directly used for downscaling or (when applicable) image synthesis. Some operators [TM98, GO11] cannot remove high-frequency patterns when the pixel-intensity differential is large, leading to aliasing. Increasing the filter's intensity support or using alternatives formulations [ZSXJ14, XLXJ11] leads to overblurring of fine and low-intensity features [KSP13]. We note that Kopf et al. [KSP13] describe specific modifications to the edge-aware bilateral kernel specifically for image downscaling, and we compare it to our SBS3 prefilter in Section 5.2. Other techniques adapt non-linear filters specifically for rendering [RKZ12, VRM*18]. These Monte Carlo denoising techniques can be used whenever the number of samples per output pixel is not sufficient for generating the final pixel intensities. For larger numbers of samples per pixel, however, one usually falls back to classic linear low-pass prefilters (Section 2.1). Our SBS3 prefilter does not compete with denoisers and should be used in situations like the latter.

### 2.3. Subpixel rendering

Many works have used the subpixel structure of digital displays to improve the spatial resolution of rendered text [Pla00, BBD*00, FEL*11] and images [KH03, ESKD14]. Since subpixels are associated with different primary colors, these works use perceptual metrics to generate low-pass filters that reduce chromatic distortions and color fringing [ESKD14, FEL*11]. The underlying idea of displaced subpixel sampling [KH03] may also be used with our prefilters.

An important distinction between subpixel-rendering works and our work is in the choice of perceptual models. In our work, we are interested in how the eye's point spread function (PSF) is a *non-ideal* low-pass filter [AS00]. On the other hand, for subpixel rendering [Pla00, FEL*11, ESKD14], one is mostly interested in the eye's different sensitivity thresholds between luminance and chrominance [ZW97]. Thus, Platt [Pla00] employs a PSF model for the eye that is sufficient to deal with chromatic distortions, but that has shortcomings as a realistic low-pass optical function. In particular, its frequency response has a perfectly flat passband followed by a sharp decay [ESKD14], which is not representative of actual measurements and theory [AS00]. As a result, the optimal filters derived by Platt [Pla00] and subsequent works [BBD*00, FEL*11, ESKD14] are plain low-pass filters that do not provide the frequency-enhancement properties of our SBS3 prefilters (Section 4). Also, closely related to this topic is the technique of Huberman and Fattal [HF16], which aims to reduce Mach-band-like effects. Our technique is complementary to theirs.

### 2.4. Image Sharpening

To counteract the overblurring discussed in the introduction, commercial displays and graphics hardware commonly provide some form of user-controlled "image sharpness" setting. Some digital

cameras also automatically apply sharpening filters during image capture [Maî17], and professional photographers often use unsharp masking during postprocessing [FS10]. While these are in part aimed at combating the blur caused by suboptimal capture conditions (such as camera shake, lens defects or defocus), they are also used to correct for imperfect display processes. In fact, photographers empirically know that the amount of sharpening required to make an image "look good" depends on the type of display medium (e.g. print or digital) and resolution (dots or pixels per inch) [FS10].

Despite the widespread use of sharpening filters, it is not obvious how to define the specifics of what they should do. The high-level goal is to improve the representation of so-called fine details, and that usually translates to some form of high-frequency-enhancing filter, which increases the subjective perception of edge sharpness (commonly referred to as "acutance") [Maî17]. The spectral characteristics of such a (linear or non-linear) filter, however, are usually selected in an ad hoc procedure based on experimentation and rules of thumb. Photography experts for example recommend that the halo around edges—introduced when enhancing high frequencies—should be sized between 0.01 and 0.02 inches "to produce a satisfactorily sharp image without introducing visually obvious sharpening haloes" [FS10]. Similar guidelines have been proposed by computer graphics practitioners, such as the use of interpolation filters with exactly one lobe of ringing to improve perceived edge contrast [MN88].

The unrestrained application of sharpening—without a careful selection of the filtering parameters—often leads to "oversharpened" and unattractive images [FS10]. This has been observed for example in the image downscaling literature, where the use of standard post-sharpening filters after resampling may introduce objectionable artifacts [KSP13, OG15]. By using our SBS3 prefilters (which require no manual tuning of parameters) for image downscaling or image synthesis, the need for a user to apply an ad hoc post-sharpening filter is significantly diminished.

## 3. Deriving our sharp prefilter

We consider the imaging pipeline illustrated in Figure 2. The goal is to define the pixel intensities over the screen $\mathcal{S}$ that minimize the difference between the spatially-continuous image function $f$ and its reconstruction $\tilde{f}$ on the observer's retina $\Omega$. We do this by treating $f$ and $\tilde{f}$ as vectors in $L^2$ [KU81, Uns00], and computing the distance:

$$\|f-\tilde{f}\| = \left(\int_\Omega \left|f(x,y)-\tilde{f}(x,y)\right|^2 dx\,dy\right)^{1/2}. \qquad (1)$$

To evaluate this integral, $\tilde{f}$ is defined over $\Omega$ by the reconstruction model from Section 3.1, and $f$ is defined over $\Omega$ by assuming an ideal display (infinite resolution) and ideal observer (infinite resolving power). Although the $L^2$ norm is not ideal for measuring perceptual error [Str97, WBSS04], its use is widespread because it makes the problem tractable: the minimization of $\|f-\tilde{f}\|$ has a closed-form linear solution, efficient to implement and easy to analyze through Fourier theory (Section 4) [NH14]. The display+eye reconstruction model discussed next (which models how we perceive an image and is used to define $\tilde{f}$) makes the distance $\|f-\tilde{f}\|$ a perceptual metric, similar to S-CIELAB [ZW97].

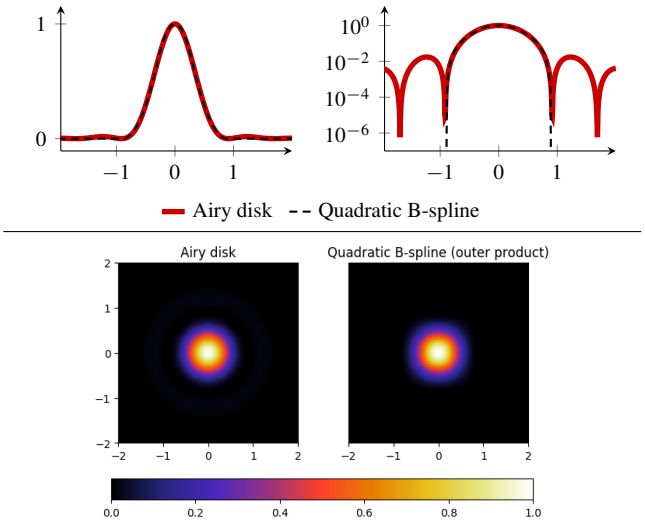



**Figure 3:** *Cross section of a bidimensional Airy disk pattern and its quadratic B-spline fit (PSNR over 40 dB). Shown in linear (top left) and log scale (top right). Also, their respective outer products preserve the similarity (lower left and lower right).*

### 3.1. The reconstruction model

For a typical LCD display, it is reasonable to represent it as a box reconstruction [NH14]. Thus, the $[i,j]$-th pixel coefficient $p[i,j]$ on the framebuffer controls the intensity of one small box-like light source. Collectively, these coefficients define a spatially-continuous, piecewise-constant, 2-D lightfield $g(u,v)$ displayed on the screen. We define continuous screen units $(u,v) \in \mathcal{S}$ with integer coordinates associated with the centers of the display's pixels:

$$g(u,v) = \sum p[i,j]\,\mathrm{rect}(u-i, v-j). \qquad (2)$$

The summation is taken over valid integer screen coordinates and $\mathrm{rect}(u,v)$ equals 1 for $(u,v) \in [-1/2, 1/2)^2$, and zero otherwise.

The lightfield $g(u,v)$ is projected into the eye and reaches the retina to form $\tilde{f}$. More precisely, assuming that the observer is focusing on the screen, $\tilde{f}$ is given by the convolution $\tilde{f} = g_\Omega * s$ [Goo05], where $g_\Omega$ is the image given by pinhole projection and $s$ is the point spread function of the visual system.

An eye free from aberrations and light scattering, with a circular pupil and focusing on the screen, can be considered a diffraction-limited system [AS00], whose theoretical PSF is an Airy disk (Figure 3). This theoretical model has been experimentally validated with laboratory tests, and holds true for pupil diameters up to 3 mm in size [WC85]. The expected human pupil size when looking at a typical LCD screen, with brightness ranging from 150 to 300 nits [AMHH08], is between 2.6 and 2.8 mm [MS44, POAR12]. As such, the Airy disk is a good model for the observer's PSF in our visualization pipeline. For computational efficiency, we replace the infinite impulse response Airy disk equation with a quadratic B-spline approximation (Figure 3), which has compact support and makes all the required computations precise and efficient in 1D and 2D [NH14].

For a constant focal distance and unchanging lighting conditions

(i.e., for a given accomodation and pupil size), the PSF has a constant size and shape on the retina $\Omega$; however, the size of the projected lightfield $g_\Omega$ depends on the distance between the observer and the screen. In particular, increasing such distance makes $g_\Omega$ smaller which, in turn, makes the PSF larger relatively to it. In the end, the relative size between $g_\Omega$ and the PSF $s$ is all that matters to compute $\tilde{f} = g_\Omega * s$. To simplify this task, we represent all quantities in screen (pixel) units $(u,v) \in \mathcal{S}$, where the size of the PSF will increase proportionally to the viewing distance. Appendix B describes how we measure the size—in screen units—of an individual's PSF, depending on viewing distance and display pixel density.

Finally, considering the full reconstruction pipeline composed of the display's box reconstruction followed by convolution with the eye's PSF, we have, in screen space,

$$\tilde{f}(u,v) = \sum p[i,j]\,\varphi(u-i, v-j). \tag{3}$$

The function $\varphi = \text{rect} * s$ is our *reconstruction kernel*. It is a piecewise cubic polynomial, whose exact shape depends on the viewing distance $D$. Figure 4 shows the equivalent 1-D kernels $\varphi(u)$ for several values of $D$, all of which have closed-form polynomial expressions. The 2-D kernels are given by the outer product $\varphi(u,v) = \varphi(u)\varphi(v)$, which makes all filtering operations separable. This is accurate because $\varphi$ is given by a convolution of separable functions rect and $s$ (the Airy disk's main lobe is virtually separable due to its striking similarity to a Gaussian—which is why it is adequately represented by a B-spline).

For a viewing distance $D = 40$ cm and an LCD display with 100 pixels per inch (ppi), the reconstruction kernel is given by the following piecewise cubic polynomial:

$$\varphi(u) = \begin{cases} c_1|u|^2 + 1 & 0 \le |u| < d_1; \\ c_2|u|^3 + c_3|u|^2 + c_4|u| + c_5 & d_1 \le |u| < d_2; \\ c_6|u|^3 + c_7|u|^2 + c_8|u| + c_9 & d_2 \le |u| < d_3; \\ c_{10}|u|^3 + c_{11}|u|^2 + c_{12}|u| + c_{13} & d_3 \le |u| < d_4; \\ 0 & \text{otherwise;} \end{cases} \tag{4}$$

where

$$\begin{aligned} c_1 &= -2.06052, & c_2 &= 2.63514, & c_3 &= -3.6554, \\ c_4 &= 0.322987, & c_5 &= 0.978197, & c_6 &= 1.7501, \\ c_7 &= -2.62514, & c_8 &= -0.0813455, & c_9 &= 1.03109, \\ c_{10} &= -0.875048, & c_{11} &= 3.6554, & c_{12} &= -5.08999, \\ c_{13} &= 2.36253, & d_1 &= 0.202514, & d_2 &= 0.392458, \\ d_3 &= 0.797486, & d_4 &= 1.392460. \end{aligned}$$

This function was obtained through the convolution of the rect function with the PSF model described in Appendix B–Eq. (16). It was used for generating all results shown in the paper and supplementary materials (unless otherwise stated), and it is the one we recommend for general use when the exact viewing conditions are unknown. We chose $D = 40$ cm at $P = 0.25$mm pixel size (100 ppi) because this $D/P$ ratio is a good representative for the average viewing condition (considering that this is the ratio where the perceived pixels lose their box shape (Figure 4), and that display manufacturers intend to produce devices with pixels that are imperceptible to the user at the average viewing distance). Section 4.4 discusses some practical considerations for working with arbitrary viewing conditions.

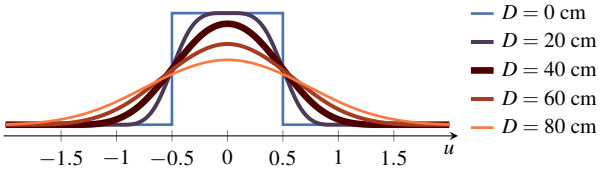



**Figure 4:** *The simulated reconstruction kernels $\varphi$ for viewing distances ranging from 0 to 80 cm, computed for a 100 ppi display. All of these are piecewise cubic polynomials which have closed-form expressions. The values of $u \in \mathcal{S}$ are in continuous screen units, as described in the text.*

### 3.2. The prefilter

Given $\varphi$ and Eq. (3), we find the coefficients $p[i,j]$ (the pixel intensities to be displayed on the screen) that minimize the reconstruction error $\|f - \tilde{f}\|$. This has a closed-form solution [Uns00, NH14]:

$$p[i,j] = \int_{\mathcal{S}} f(u,v)\,\psi(u-i, v-j)\,du\,dv. \tag{5}$$

The function $\psi$ is the so-called *dual* of $\varphi$. In Eq. (5), it is convolved with the image $f$ to generate the pixel intensities $p[i,j]$ to be displayed on the screen. Thus, $\psi$ is our **SBS3 prefilter**.

Since $\varphi(u,v) = \varphi(u)\varphi(v)$ is separable, $\psi(u,v) = \psi(u)\psi(v)$ is also separable and

$$\psi(u) = \sum q[k]\,\varphi(u-k). \tag{6}$$

The discrete sequence $q$ is an infinite impulse response digital filter best characterized through its Z-transform [PM06]:

$$Q(z) = 1 \Big/ \Big(\sum a_\varphi[k]\,z^{-k}\Big), \tag{7}$$

where $a_\varphi[k]$ is the sampled autocorrelation of $\varphi$:

$$a_\varphi[k] = \int_{\mathcal{S}} \varphi(u)\,\varphi(u-k)\,du. \tag{8}$$

Since our reconstruction kernel $\varphi$ is a compactly-supported polynomial spline, its autocorrelation sequence $a_\varphi$ is finite and easily computed by analytical integration.

### 3.3. Sharp filtering with SBS3

For efficient prefiltering with $\psi$, Eq. (5) and Eq. (6) imply:

$$p[i,j] = \sum_k \sum_m q[k]q[m] \int_{\mathcal{S}} f(u,v)\,\varphi(u-i-k, v-j-m)\,du\,dv. \tag{9}$$

Conceptually, the SBS3 prefilter works as follows:

1. Convolve the input $f(u,v)$ with the filter $\varphi$, to obtain $f_\varphi(u,v)$;
2. Sample $f_\varphi$ at the valid pixel positions $[i,j]$, to obtain $\mathbf{f}_\varphi[i,j]$;
3. Compute the optimal image $p[i,j] = \sum_k \sum_m q[k]q[m]\mathbf{f}_\varphi[k,m]$.

Steps 1-2 constitute filtering with a piecewise polynomial kernel $\varphi$. Since $\varphi$ is compactly supported and separable, the convolution integral (in Eq. (9)) can be computed very efficiently. The summations in Step 3 are also efficient and may be executed as two separable applications of causal/anticausal recursive filters [Uns00, PM06], which are $\mathcal{O}(n)$ operations in the number of pixels $n$.

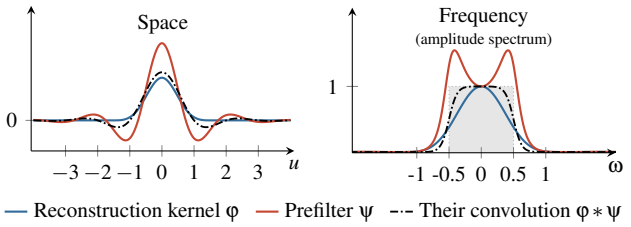



**Figure 5:** *Our recontruction kernel* φ *and its associated prefilter* ψ *for* $D = 40$ *cm (normalized to unit area for comparison). The dashed line represents their Fourier pointwise product, i.e., their spatial convolution. The frequency axis* ω *is given in cycles per screen pixel and the light-gray rectangle represents the Nyquist region.*

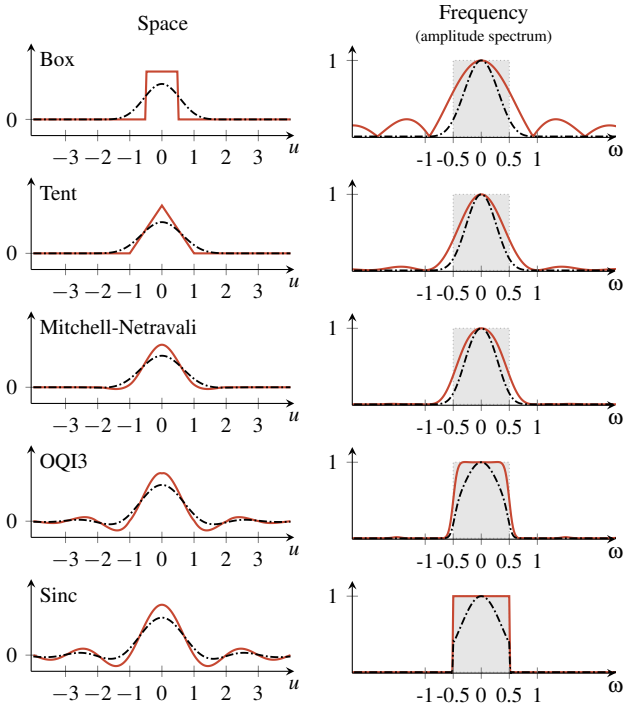



**Figure 6:** *A variety of existing Shannon-type filters (red curves) and their interaction with the kernel* φ *modeling the display + eye reconstruction (dashed lines). In all cases, the dashed line is far from a box-like response in frequency, which means the observer will perceive the image as overblurred.*

## 4. Analysis and Discussion

Figure 5 shows the shape of our reconstruction kernel φ from Eq. (4) and of its dual ψ, together with their Fourier transforms. Observe how φ is a low-pass filter, modeling the blur from the eye's PSF [ZW97, AS00]. Also note how our prefilter ψ acts in the reverse direction, *enhancing* frequencies that would otherwise be blurred by φ. The combined effect of ψ and φ has a frequency response which adequately preserves details across the Nyquist portion of the spectrum, as shown by the dashed black line on the right plot

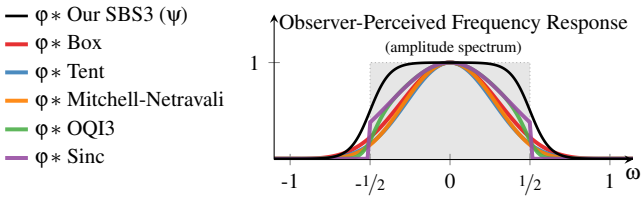



**Figure 7:** *Observer-perceived frequency responses. Our SBS3 prefilter* ψ *(black line) preserves more detail across the Nyquist range. This results in sharper images when compared to the ones generated by existing filters.*

of Figure 5. As such, an observer looking at a version of image $f$ processed with ψ will perceive a good representation of $f$.

It is instructive to perform the same analysis on existing Shannon-type filters. Figure 6 shows the spatial and frequency response for the box, tent, Mitchell-Netravali [MN88], Optimized cubic Quasi-Interpolator (OQI3) [SN15], and sinc.† As before, the dashed black lines represent the observer-perceived combined effect of the respective filter (which acts as a prefilter) in conjunction with the reconstruction kernel φ (which models the display + the eye's PSF). From the Fourier domain plots on the right column of Figure 6, it is clear that the resulting frequency responses have excessive attenuation in the Nyquist region. As a result, the images processed by these filters will be perceived as *overblurred*.

Contrasting Figures 5 and 6, one can see that our prefilter ψ is unique in the sense that it appropriately enhances frequencies in order to minimize the blurring introduced by the human optical system. Figure 7 overlays all the simulated observer-perceived frequency responses for easier comparison. Note how our sharp prefilter ψ better preserves the frequencies in the Nyquist region. The frequency responses of the other filters are similar to each other, and cause significant attenuation in the Nyquist region.

### 4.1. Aliasing Analysis

The cost of sharpness is a greater tendency for aliasing. Aliasing exclusively occurs as a consequence of sampling (with *post*-aliasing only ocurring during reconstruction [MN88]). To measure the possible effects of this phenomenon, we employ the same theoretical analysis proposed by Nehab and Hoppe [NH14].

The **purple** curve in each plot of Figure 8 illustrates the frequency response of the full sampling-and-reconstruction pipeline on a broad-spectrum function $f$ (i.e., this is a worst-case scenario—a function containing all possible frequencies). The pipeline is composed of **(1)** convolving $f$ with a specified filter, which performs a pointwise multiplication of the spectrums; **(2)** sampling the result at integer positions, which convolves the spectrum with a Dirac comb (performing a periodic summation); and **(3)** performing final

† Note that OQI3 and Mitchell-Netravali were designed to be reconstruction filters. Even so, because OQI3 (and also Mitchell-Netravali, to a lesser degree) approximates the sinc filter—which does ideal prefiltering *and* reconstruction—they both do a great job on prefiltering tasks.

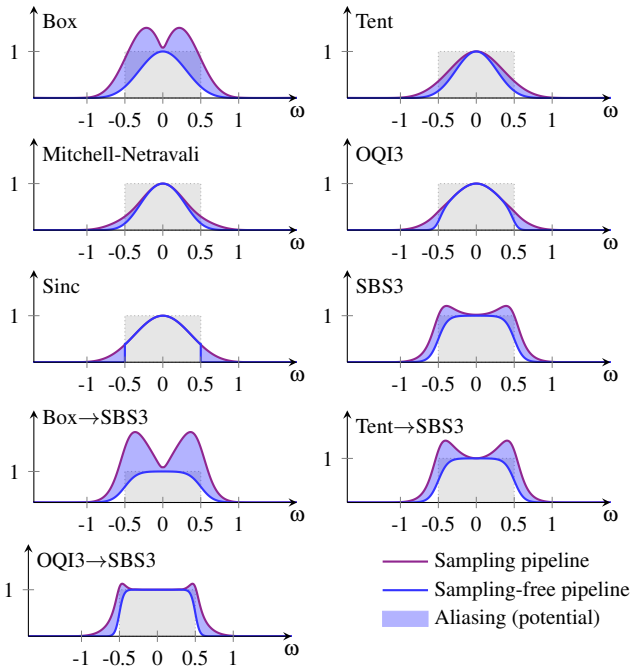



**Figure 8:** *Aliasing Analysis. The purple curves in each plot represent the frequency response of the full sampling-and-reconstruction pipeline on a broad-spectrum function $f$. The blue curves represent the effect of the same pipeline but* without *sampling. By definition, the blue frequency response is* uncorrupted *by aliasing and represents the pure effect of the prefilter and reconstruction kernel. The shaded area between the curves measures the potential for the ocurrence of aliasing at each frequency.*

reconstruction by convolution with the kernel $\varphi$ (also a pointwise multiplication of the spectrums), to simulate the observer-perceived final frequency response. Similarly, the **blue** curve in each plot illustrates the effect of the same pipeline but *without* sampling, i.e., without step (2). By definition, the blue frequency response is *uncorrupted* by aliasing and represents the pure effect of the prefilter and reconstruction kernel. Thus, the shaded area between the blue and purple curves measures the potential for the ocurrence of aliasing at each frequency (the shaded areas located outside the Nyquist interval measure the potential for *post*-aliasing, which is dependent on the reconstruction kernel).

The ideal zero-aliasing frequency response is one where the purple curve matches the blue curve. The ideal blue curve is likewise one which matches the flat and broad spectrum of the input $f$, i.e., a constant unity-intensity line across *all* frequencies ($-\infty$ to $+\infty$). The closer the blue curve is to 1, the sharper is the final reconstruction (since more frequencies are preserved from the input). Conversely, the closer the blue curve is to 0, the blurrier the reconstruction. As seen in Figure 8, our prefilter SBS3 has by far the best-performing sharpness (blue) curve, but it is also more prone to aliasing than, for instance, Mitchell-Netravali or OQI3. The most severe aliasing, however, comes from the box filter (top left).

It *is* possible to avoid aliasing while also avoiding the loss of sharpness, but that comes at the cost of a bigger problem: *ringing*. It occurs because of the instantaneous cutoff in the frequency response of the prefilter (necessary to eliminate aliasing while keeping things sharp). The only way to diminish ringing is to soften the transition at the cutoff, which inevitably leads back to either aliasing or blurring. Thus, ringing, blurring, and aliasing are highly-interdependent phenomena [MN88]: reducing the effect of one impacts the others.

#### 4.1.1. Aliasing vs Ringing vs Blurring

Since, *a priori*, these three artifacts will be present in some degree, it is worth discussing which one is the bigger evil. In particular, blurring is often taken to be the worst of all three, as the human visual system is "more sensitive to high spatial frequency errors" [KU81], making fuzzy edges highly objectionable. Our user study corroborates this observation (Section 5.2.2). Furthermore, aliasing and ringing can sometimes be employed in our favor: when downscaling an image, for example, instead of low-pass filtering step-like edges, which could incur in unwanted blurring, simply subsampling it (thus introducing aliasing) gives the best results, completely preserving the step edge in the downsampled version [GO17]. Similarly, one lobe of ringing has been observed to increase the perceived contrast of edges, improving the overall sharpness of the image [MN88]. Note that our prefilter $\psi$ has lobes which boost high frequencies, thus compensating for the blurriness caused by the eye's PSF.

It is thus reasonable to seek a prefilter which avoids blurring (maximizes sharpness) as much as possible, and at the same time achieves a good balance between aliasing and ringing. Since that is a subjective and application-dependent balance, it is useful to have means of moving between the two extremes. We thus describe (Section 4.2) a simple way for changing the characteristics of our SBS3 prefilter $\psi$ by *combining* it with other existing prefilters, while still preserving its unique sharpness. We use this procedure to define a collection of prefilters based on SBS3, composed of *(i)* $\psi_{\text{OQI3}\to\text{SBS3}}$ for when one wants to move away from aliasing (but towards more ringing); *(ii)* $\psi_{\text{Box}\to\text{SBS3}}$ for moving away from ringing (but towards more aliasing); and *(iii)* our original $\psi$, which strikes a good balance between aliasing and ringing for the general case. All of these alternatives are efficient and preserve the linear-time computational complexity of the technique. We additionaly define $\psi_{\text{Tent}\to\text{SBS3}}$, which is equivalent to $\psi$ in sharpness and ringing, but facilitates the integration of our approach with e.g. black-box Monte Carlo rendering software, as discussed in Section 5.3.

#### 4.1.2. Measuring Sharpness, Aliasing and Ringing

The sharpness, aliasing and ringing characteristics of our collection of SBS3 filters are summarized in Table 1. These quantitative metrics were obtained as follows: sharpness (S) measures how close the reconstruction spectrum (blue curves in Figure 8) approximates an input broadband spectrum (a constant spectrum with amplitude 1) in the frequency interval $[-2,2)$; aliasing (A) is measured as the area of the blue shaded regions shown in Figure 8; and ringing (R) is measured by the area of the second and higher negative lobes in the impulse response (thus measured in the spatial domain) of the prefilters. We ignore the first negative lobe when computing (R) since, as pointed out by Mitchell and Netravali [MN88], the first negative lobe helps to improve sharpness. Table 1 shows that

| Sampling prefilter | (S)harpness | (A)liasing | (R)inging | Observation |
|---|---|---|---|---|
| Box | S=1.136 | A=1.000 | R=0.000 | |
| Tent | S=1.000 | A=0.267 | R=0.000 | |
| Gaussian 2x2 ($\sigma = 1/3$) | S=1.099 | A=0.422 | R=0.000 | |
| Gaussian 3x3 ($\sigma = 3/6$) | S=0.922 | A=0.152 | R=0.000 | |
| Gaussian 4x4 ($\sigma = 2/3$) | S=0.777 | A=0.070 | R=0.000 | |
| Mitchell-Netravali | S=1.010 | A=0.172 | R=0.000 | |
| OQI3 | S=1.146 | A=0.180 | R=0.090 | |
| Sinc | S=1.162 | A=0.168 | R=1.000 | |
| SBS3 | S=1.514 | A=0.451 | R=0.074 | Good Compromise |
| OQI3→SBS3 | S=1.514 | A=0.339 | R=0.398 | Reduce (A) |
| Box→SBS3 | S=1.526 | A=1.606 | R=0.052 | Reduce (R) |
| Tent→SBS3 | S=1.514 | A=0.609 | R=0.071 | |

**Table 1:** *Sharpness, aliasing, and ringing characteristics of various prefilters. Higher values of sharpness are preferred, while lower values of aliasing and ringing are better. Values normalized by the scores of Tent (sharpness), Box (aliasing), and Sinc (ringing). Our SBS3 prefilter offers a good compromise among sharpness, aliasing, and ringing (also see Figure 9).*

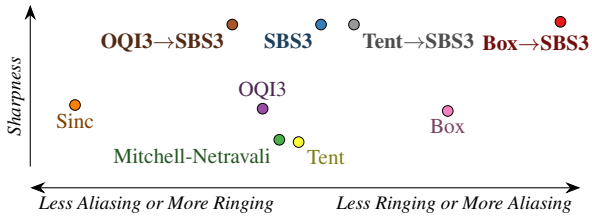



**Figure 9:** *Relationship among various prefilters considering their sharpness, aliasing, and ringing characteristics. The names of our collection of SBS3 prefilters are in bold.*

our SBS3 prefilter $\psi$ offers a good compromise among sharpness, aliasing, and ringing. Its variants $\psi_{\text{Box}\to\text{SBS3}}$ and $\psi_{\text{OQI3}\to\text{SBS3}}$ can be used if one wants to favor reduction in ringing, and aliasing, respectively, while preserving sharpness.

The S-A-R indices are used to visualize the relationship among the filters by positioning them in a 2-D space (Figure 9), where the vertical axis is given by S and the horizontal axis is given by A minus R (representing the trade-off between aliasing and ringing). Note that such a visualization has to be taken with a grain of salt since the 3-D S-A-R space is not embedable in 2-D without some loss of information. Nonetheless, Figure 9 portrays the fact that our collection of SBS3 prefilters are considerably sharper than existing options, while allowing for a wide range of choices in the aliasing-ringing range based on user preference and application characteristics.

### 4.2. Oblique Projections

To compute Eq. (5) one must know $f$ almost everywhere in its domain. Unfortunately, in many situations $f$ is not known, and we only have access to its measurements with some prefilter $\eta$:

$$p_\eta[m,n] = \int f(u,v)\,\eta(u-m,v-n)\,du\,dv. \tag{10}$$

This happens, for example, when one is given the pixels $p_\eta$ generated by a black-box Monte Carlo renderer, with the accompanying information that they were computed using, e.g., a tent filter $\eta$. In any case, when $f$ is unknown it is not possible to minimize $\|f-\tilde{f}\|$. An alternative is to seek a reconstruction $\tilde{f}$ which results in the same measurements $(p_\eta)$ if prefiltered and sampled again with $\eta$. This is called an *oblique projection* due to the geometry of the underlying inner products [UA94], and it also has a closed-form solution, which involves convolving the coefficients $p_\eta$ with the (discrete) convolutional inverse of the cross-correlation of $\varphi$ and $\eta$:

$$a_{\varphi,\eta}[k] = \int \varphi(u)\,\eta(u-k)\,du. \tag{11}$$

In our cases of interest $\eta$ has compact support, and the computational complexity of oblique projection is linear on the number of samples.

#### 4.2.1. A collection of sharp prefilters based on SBS3

If we call $V_\varphi$ and $V_\eta$ the function spaces spanned by shifts of $\varphi$ and $\eta$, we empirically observe that the oblique projection $\tilde{f} \in V_\varphi$ is also close to $V_\eta$. Thus, it retains important characteristics from both spaces. We use this fact to define a collection of prefilters which combine the sharpness of our SBS3 with the good properties of other prefilters from the literature.

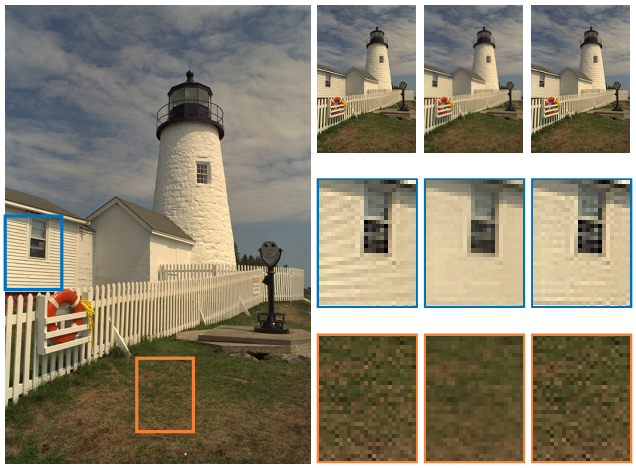
Choosing $V_\eta$ as the OQI3 space of Sacht and Nehab [SN15], one obtains the oblique-projection prefilter $\psi_{\text{OQI3}\to\text{SBS3}}$ which is as sharp as SBS3 but with improved aliasing characteristics. This is seen in the plot of Figure 8 and in Table 1 (numbers highlighted in blue). Reducing aliasing while being sharp means that $\psi_{\text{OQI3}\to\text{SBS3}}$ rings more. As such, it should be applied in situations where aliasing is less preferable than ringing, such as when downscaling images with repeating patterns (Figure 10).

Choosing $V_\eta$ as the space spanned by box kernels allows one to move in the reverse direction. The $\psi_{\text{Box}\to\text{SBS3}}$ prefilter considerably reduces ringing and should be used in situations where the function $f$ has sharp, step-like discontinuities (where ringing is most problematic). A representative application is the rasterization of vector graphics, illustrated in Figure 15. An additional advantage of $\psi_{\text{Box}\to\text{SBS3}}$ is that it is evaluated simply by box prefiltering (followed by efficient digital filtering), thus being trivial to implement and to integrate into existing pipelines.

Finally, choosing $V_\eta$ as the space spanned by tent kernels results in a prefilter $\psi_{\text{Tent}\to\text{SBS3}}$ which is very similar to SBS3 ($\psi$) in sharpness and ringing, but suffers more from aliasing. It is therefore an option for obtaining a balance similar to SBS3 when $f$ is only known from its measurements in a tent basis. A good example of its use is to post-process a common variety of Monte Carlo renderings generated with the tent, as illustrated in Figure 16.

### 4.3. Clamping and negative light

Filters with negative lobes provide higher contrast at the edges due to their "controlled ringing" characteristics [ST85]; this nevertheless comes at the cost of possibly having negative and overshot pixel values. The straightforward solution is to simply clamp invalid values to the $[0,1]$ range. For overshot values this procedure works satisfactorily since humans are less sensitive to errors in high luminance regions [KU81]. Furthermore, with the increasing availability of high-dynamic range displays, the intensity of overshot values may actually be representable and doesn't need to be clamped.



**(a)** *Original* **(b)** *SBS3* **(c)** *OQI3* **(d)** *OQI3* $\longrightarrow$*SBS3*

**Figure 10:** *Comparison between prefiltering and downscaling (by a factor of 4) the same image using SBS3, OQI3 and the oblique projection from OQI3 to SBS3. While SBS3 (b) is sharp (see grass in the third row), it suffers from aliasing (second row). With OQI3 (c) the converse occurs: it removes aliasing but the image looks too blurry. Oblique projection from OQI3 to SBS3 (d) has simultaneously low aliasing and good sharpness properties. Check the supplementary material for the images in their intended (native) resolution.*

In a theoretical sense, clamping values after sampling and prior to display means that the reconstructed image (the one arriving on the retina of the viewer) will not be the one that minimizes the reconstruction error. How close this clamped image will be to the best reconstruction is briefly discussed by Kajiya and Ullner [KU81]. For orthogonal basis functions (like the sinc), the Pythagorean theorem implies that it is safe to clamp negative values to zero, as this results in an image as close as possible to the best reconstruction we can have. For bi-orthogonal bases like our φ-ψ pair, the best non-negative reconstruction is not necessarily the one achieved through clamping.

To solve this positivity constraint, Kajiya and Ullner propose an iterative, non-linear optimization to minimize the distance $\|\tilde{f} - f\|$ while constraining the coefficients $p[i, j]$ to be non-negative. Nehab and Hoppe [NH14] go a step further and account also for the overshot values. They propose a minimization program over the coefficients of the digital filter $q$ (Eq. (6)) which guarantees that the reconstructed image minimizes the reconstruction residual while being within the valid range. Both solutions have in common the need for intricate non-linear optimizations, which takes away the advantages of using linear filters. Although one should be aware of this fact, we believe that, in practice, clamping does not pose much of a problem. Natural images and 3D renderings, with adequate exposure settings, shouldn't have many pixels in the extremes of the representable dynamic range. For vector graphics, on the other hand, black to white transitions are common. In this case we recommend the use of our $\psi_{\text{Box}\to\text{SBS3}}$ prefilter, which significantly reduces ringing and consequently the occurence of negative or overshot pixels. We include several examples in our supplementary materials.

### 4.4. Arbitrary Viewing Conditions

The size of the observer's PSF (in screen units) increases proportionally to the ratio between the viewing distance and display pixel size ($D/P$ in Eq. (15)). A larger PSF has a narrower frequency passband, attenuating frequencies more and more as, for example, one moves away from the screen. As more information gets blurred, our prefilter ψ (red curve in Figure 5) must work harder to boost detail in order to prevent the loss of image sharpness. This process becomes progressively unstable as the frequency response of the inverse filter $Q(z)$, in Eq. (7), tends to infinity. Long before that happens, the limited dynamic range of the display device (Section 4.3) prevents us from showing the necessary high-intensity values on the screen that in theory would minimize the reconstruction error.

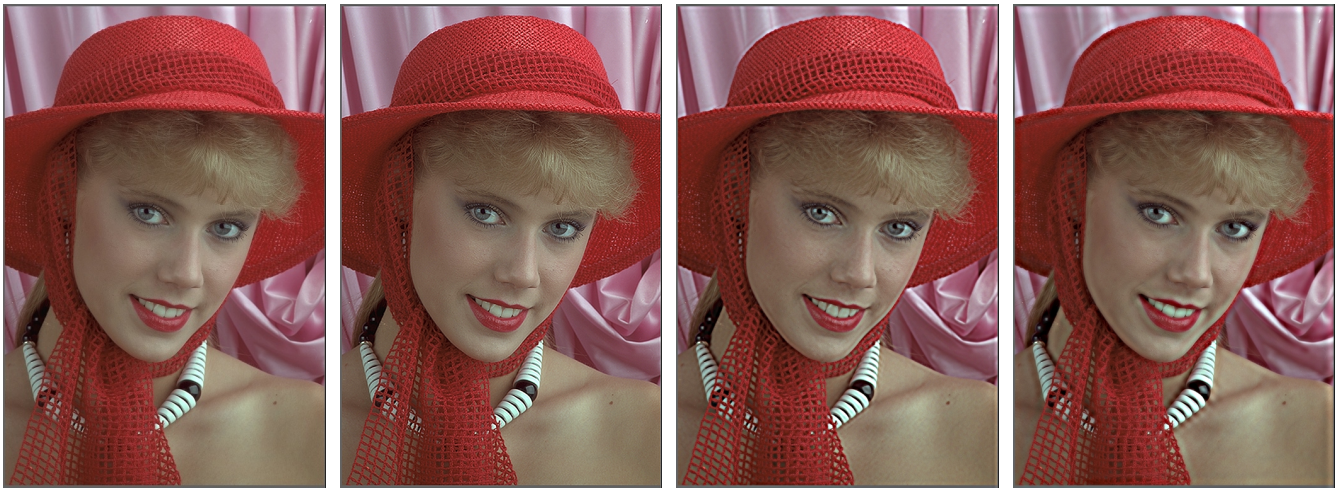
One must therefore stop the over-boosting of frequencies that inevitably get annihilated by the PSF for large viewing distances. This is achieved by finding a regularized solution to the inverse problem in $Q(z)$, one which defines a *quasi*-dual prefilter [Uns96]. The frequency response of this regularized solution matches the exact solution as one approaches the origin, but tapers off for increasing frequencies in order to avoid overshoot. Figure 11 shows one example of regularized solutions computed for increasing viewing distances.

In general, however, the exact viewing conditions may be either unknown or highly variable, and it may be impractical to repeatedly recompute an image for a variety of situations. In this case we recommend the use of $D = 40$ cm and $P = 0.25$ mm as the baseline for computing a prefilter which gives good results for average viewing conditions on digital displays. All results shown in the paper and supplementary materials were computed for this baseline, unless otherwise stated.

## 5. Applications

We showcase a variety of applications where our SBS3 prefilters yield results consistently sharper than the state-of-the-art. We compare against the classic box, tent, and Mitchell-Netravali [MN88] filters; in addition to the Optimized cubic Quasi-Interpolator (OQI3) [SN15], included as a good representative of modern generalized sampling filters [Uns00]. For image downscaling, we include in the comparisons the least-squares image resizing method of Munõz et al. [MnBU01], and the non-linear filters of Öztireli and Gross [OG15], Weber et al. [WWA*16], and Kopf et al. [KSP13]. We do not include results from the downscaling technique of Gastal and Oliveira [GO17] since it is orthogonal to and can be combined with any resampling strategy, including our SBS3 prefilter.

Our images are meant to be viewed at their native resolution. We magnify some regions to emphasize the enhancement of sharpness and the balances between aliasing and ringing, but please keep in mind that zoomed-in results may incorrectly appear "oversharpened" (for example, an edge perceived as overshot when zoomed in may look naturally sharp when viewed in native resolution due to the attenuation of the viewer's PSF). As such, we encourage the readers



**(a)** *Original photograph* **(b)** *Prefiltered for D = 80 cm* **(c)** *Prefiltered for D = 200 cm* **(d)** *Prefiltered for D = 400 cm*

**Figure 11:** *Prefiltering an image to a variety of viewing distances using SBS3 (Section 4.4). These results have been computed for a 100 ppi display and the observation distances D listed under each image. Zooming out of the digital version of this document may be used to simulate larger viewing distances, but this causes extra filtering by the PDF viewer. Thus, for best results images should be displayed at 1:1 pixel zoom and seen from the actual distance indicated in the corresponding caption (raw image files may be found in the supplementary material).*

to inspect the original-size results included in the supplementary materials.

### 5.1. Monte Carlo Rendering

In Monte Carlo rendering, "image reconstruction" commonly refers to the final computation of pixel intensities from samples collected by the renderer [PJH16]. The reader should note that such a task actually corresponds to the *sampling* stage in Figure 2(a). The spatially-continuous image $f$ is the scene being rendered.

Research renderers like PBRTv3 and Mitsuba implement only a few different prefilters: Mitchell-Netravali, Catmull-Rom, Lanczos, box, tent, and Gaussian. In the industry, a similar situation occurs: the documentations for Renderman [Ren19] and Arnold [Arn19] recommend the use of a Gaussian prefilter with a width of 2 pixels ($\sigma = 1/3$), which is almost as blurry as a tent (S=1.099 vs S=1.000) and with much more aliasing (A=0.422 vs A=0.267) (refer to Table 1). Furthermore, standard references [PJH16, HvDM*13] discuss only the classic prefiltering solutions, with only indications of modern alternatives. This is understandable, given that simple approaches such as Gaussian prefiltering give good-enough results when one is not aware of better options.

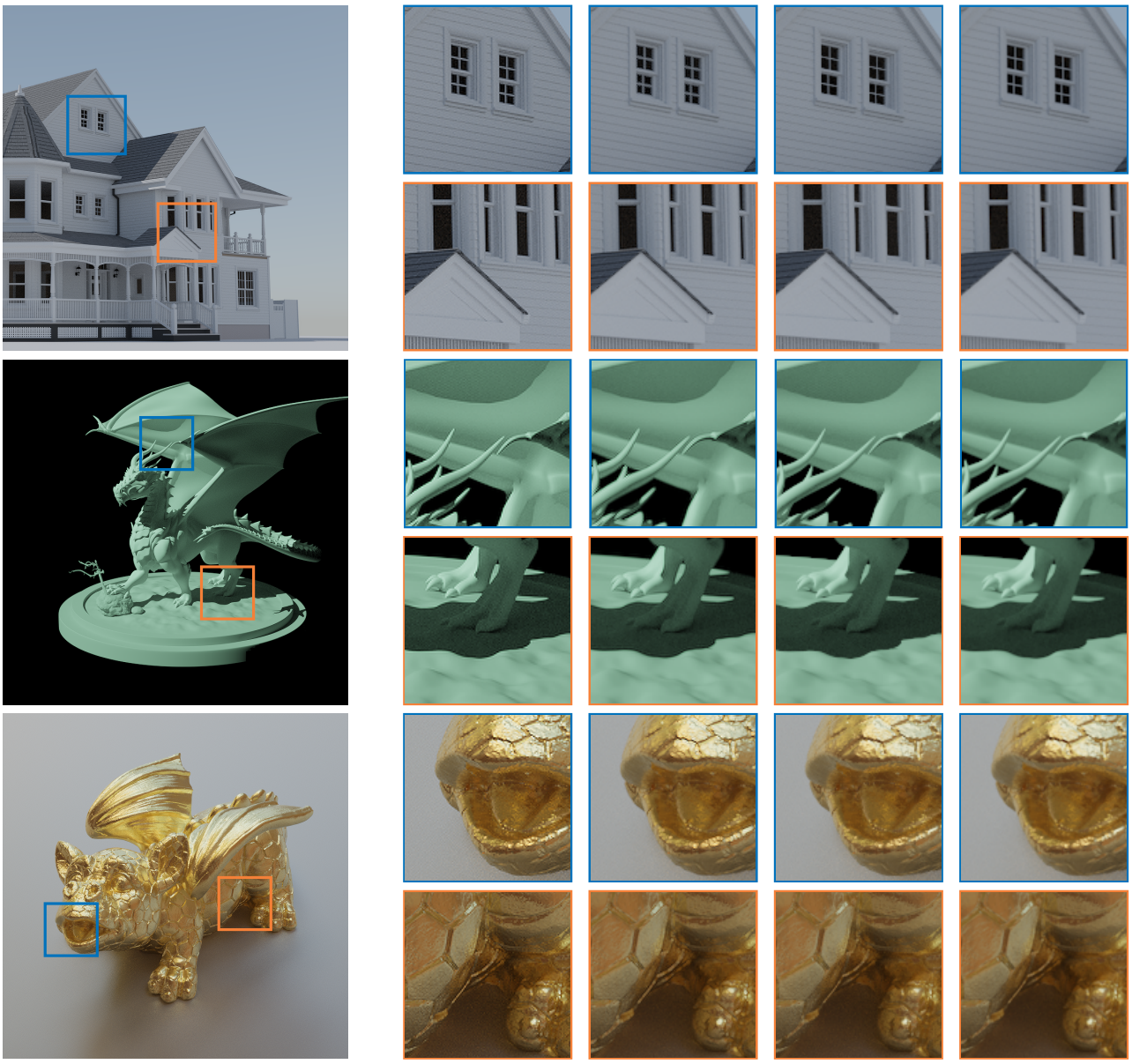
Our sharp prefilters may be efficiently employed for rendering. This is done by computing the integrals in Eq. (9) with stochastic sampling [DW85]. As suggested by Nehab and Hoppe [NH14], we share the subpixel samples between overlapping shifts of $\varphi$ to reduce variance. The resulting sequence of coefficients is then convolved (along the rows and columns) with the recursive digital filter $q$, Eq. (7), in linear time, to produce the final image. The exact same steps can be applied for rasterizing vector graphics, as one also computes pixel values from stochastic sampling, the only difference being the function being integrated. This pipeline is easily incorporated into any rendering system, as $\varphi$ (Eq. (9)) is a simple piecewise-polynomial filter, and the recursive-filtering step may be done outside the renderer by post-processing the output image.

Figure 12 compares the performance of the box, Mitchell-Netravali, OQI3, and SBS3 filters in rendering three different path-traced scenes. The same set of stochastic samples were precomputed and used by all approaches to generate the images. Notice how our SBS3 technique preserves fine details and emphasizes edges with subtle outlines, pre-compensating for the blurring caused by the observer's PSF.

Our technique is orthogonal to Monte Carlo denoising methods, which focus on generating good images from an insufficient number of samples. When a large number of samples is available but the image still looks blurry, denoising cannot improve the image quality—something that our SBS3 prefilter is designed to do. Furthermore, our technique could be used to generate sharp ground-truth images for training recent machine learning denoising techniques [BVM*17]. In the other direction, given a suitable model for the denoising kernel of a particular denoising method, it is perfectly reasonable to employ the oblique projection operator (Section 4.2) to compute the equivalent sharp images in the SBS3 space.

### 5.2. Image Downscaling

Image downscaling can be implemented and interpreted in countless different ways [UAE95, LEU98, MnBU01, NH14, OG15, WWA*16]. A finite collection of pixels gives an inherently limited view of what the original spatially-continuous information actually was. As such, for downscaling (or upscaling), one must assume some model for the relation between the pixels $p$ and the original function $f$.



**(a)** *Our SBS3 rendering* **(b)** *Our SBS3 (zoom)* **(c)** *OQI3* **(d)** *Box* **(e)** *Mitchell-Netravali*

**Figure 12:** *Performance of different prefilters for path-traced rendering. Fine details like the bars of the window (top row) and scratches/carving on the surfaces (fifth and sixth rows) are blurred by OQI3 (c), Box (d) and Mitchell-Netravali (e); but are preserved by SBS3 (b). Moreover, SBS3's controlled ringing on the edges makes shadows (third row) and contours (second, fourth row) better outlined. Notice how OQI3's ringing is more perceptible around the shadow region.*

We analyze the performance of our SBS3 prefilter for image downscaling by choosing the simplest possible model for the input image: that its pixels were obtained by sampling the original (and unknown) function $f$ at unit intervals, $p[n] = f(n), n \in \mathbb{Z}$. Since $f$ may not have been bandlimited before this sampling, we avoid making any assumptions of how it behaved between samples. As such, we define a reconstruction $f^*$ as a sequence of Dirac deltas located at the sample positions and scaled by the associated sample values: $f^*(x) = \sum_{n\in\mathbb{Z}} p[n]\,\delta(x-n)$. For a downscaling factor $\tau > 1$, it is now possible to find the orthogonal projection of the rescaled $f^*(\tau x)$ onto the SBS3 space. Plugging this $f^*$ into Eq. (5) yields

$$p_{\text{downscaled}}[m] = \sum_{n\in\mathbb{Z}} p[n]\,\psi(n/\tau - m), \tag{12}$$

which represents the correlation (convolution) between the discrete pixels of the input image and the $\tau$-scaled continuous SBS3 prefilter $\psi$. Since $\psi$ has infinite support, it is computationally preferable to substitute Eq. (6) into Eq. (12) to reach

$$p_{\text{downscaled}}[m] = \sum_{k\in\mathbb{Z}} q[k]\, c[m+k], \tag{13}$$

where

$$c[i] = \sum_{n\in\mathbb{Z}} p[n]\, \varphi(n/\tau - i). \tag{14}$$

If $\tau$ is an integer, then $c$ is given by *(i)* the discrete correlation between the input image's pixels $p$ and a *sampling* of the compactly-supported polynomial kernel $\varphi$; followed by *(ii)* decimation by a factor of $\tau$. If $\tau$ is not an integer, computing Eq. (14) is slightly more intricate: one has to evaluate $\varphi$ at arbitrary locations $n/\tau - i$, for $n, i \in \mathbb{Z}$. After $c$ has been computed, the final downscaled pixel values are obtained by efficient recursive filtering with $q$ (Eq. (13)).

#### 5.2.1. Visual Quality Evaluation

We used the aforementioned downscaling procedure to evaluate all linear prefilters: SBS3, OQI3, Mitchell-Netravali, and box. For the downscaling techniques of Öztireli and Gross [OG15], Weber et al. [WWA*16], Munõz et al. [MnBU01], and Kopf et al. [KSP13], we used the source code kindly provided by the original authors.

Figure 13(a) shows a portrait picture with some fine to medium-scale details. Our SBS3 prefilter in (b) manages to downscale this image by a factor of 5 in each dimension while preserving details of the hat, scarf and necklace. Notice also the sharpness of the eyes and the outline of the face in our result (b), a region which is significantly blurrier in the output of the other filters (c-g). The result of the Öztireli-Gross downscaling method in (c) preserves some of the high-frequency variations of the hat's texture, but converts it to aliasing-like noise. All of these techniques can be computed in real time, except the filter of Kopf et al. [KSP13] which takes several seconds for computation.

While the notion that "sharper equals better" may be subject to individual preference, previous works on perceptual image downscaling [KSP13,OG15] verify through user studies that their techniques—described as "sharper" [KSP13]—are preferred over others. Our user study, described next, also points in this direction.

#### 5.2.2. User Study

We conducted a user study to evaluate how users perceive our filter in comparison to other techniques. The test consisted in showing the users a sequence of combinations of an input image, in its original size, along with two downscaled versions of it, generated by different prefiltering techniques, to which the user is asked to choose the one she/he thinks "better represents the full-sized image". We used the same 13 image dataset used by Kopf et al. [KSP13]. For each user we displayed in random order all pairs from $\{\text{Box}, \text{Mitchell-Netravali}, \text{OQI3}, \text{Oztireli}\} \times \{\text{Ours}\}$, which gives a total of $4 \cdot 13 = 52$ pairs of downscaled images evaluated per user. We chose two "classical" filters: box, for its simplicity and widespread use, and Mitchell-Netravali for its "overblurring" behavior, as we wanted to validate whether in any situation users prefer blurrier images. OQI3 [SN15] and Öztireli and Gross [OG15]

**(a)** *SBS3* **(b)** *Öztireli-Gross* **(c)** *Munõz et al.*

**(d)** *OQI3* **(e)** *Weber et al.* **(f)** *Kopf et al.*

**Figure 13:** *Downscaling the image of Figure 11 by a factor of 5 (103x154). Note how details of the hat and scarf, as well as the face and eyes, are better outlined due to SBS3's sharpness. (These images are better seen in the supplementary materials to avoid the unwanted resampling done by the PDF reader).*

were chosen as state-of-the-art techniques for which we had access to source code. We did not compare against Weber et al. [WWA*16] as their technique is not suited for small images. The technique of Kopf et al. [KSP13] did not perform better than Öztireli and Gross' technique in their user tests, thus we also removed it from the study in order to reduce the total number of pairs shown to the user. To check consistency of the choices, all users were shown each pair twice along the test, switching the left-right order of the downscaled images The answers of a participant for an image pair are considered inconsistent if he/she indicated preference for both images (one each time the pair was presented). A total of 35 subjects saw $2 \cdot 52 = 104$ pairs of images on the same monitor (1080p, 144 ppi). As recomended by [KSP13] we discarded all responses from subjects with more than 80% inconsistency and, moreover, we discarded individual inconsistent choices.

The preference counts are summarized in Figure 14. Subjects had an average inconsistency rate of 33%, with the most inconsistent one scoring 61%, and the most consistent scoring 3%. Among the consistent choices, there is a clear preference for our technique over the others.

We found that when presented with Mitchell-Netravali vs. SBS3, subjects are highly consistent in preferring our algorithm. This shows that our sharp images are preferred over overblurred ones. In all pair-wise comparisons involving our filter and Mitchel-Netravali, Öztireli and Gross, OQI3, or Box, our results were consistently preferred. In all cases, the number of inconsistent answers superseded the other technique multiple times. Finally, we found no apparent correlation between the image content (text, fine details, edges) and our technique scoring higher or lower, or the choices being more or less inconsistent.

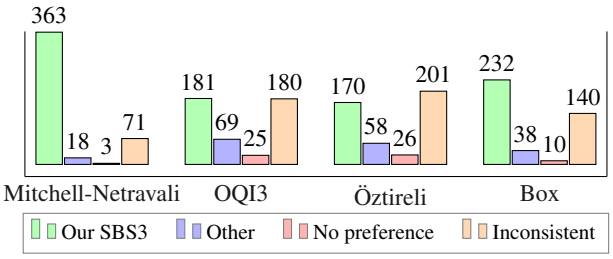
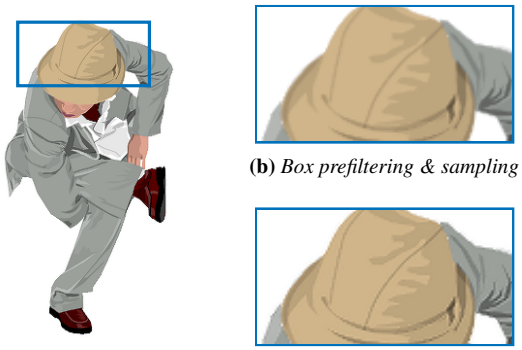



**Figure 14:** *Counts of how many times images downscaled using our technique were preferred over others in the user study. As can be seen, our SBS3 prefilter (green bars) is consistently preferred over both classic and state-of-the-art filters (blue bars).*



**(a)** *Vector graphics rasterized with our* $\psi_{Box \to SBS3}$

**(b)** *Box prefiltering & sampling*

**(c)** *Box→SBS3 (zoom of (a))*

**Figure 15:** *Vector graphics illustration rasterized through stochastic sampling using a box prefilter (b), followed by oblique projection to SBS3 (a,c). Please refer to the supplementary materials to inspect the images in their native size (PDF readers perform interpolation).*

### 5.3. Oblique Projection Applications

Our collection of prefilters allows for the trade-off between aliasing and ringing. Figure 10 illustrates an image-downscaling task performed with our $\psi_{OQI3 \to SBS3}$ prefilter, which defines an oblique projection from the OQI3 space [SN15] to SBS3. As seen in (d), this defines a prefilter which combines the sharpness of our SBS3 space with the good anti-aliasing properties of OQI3.

Figure 15 shows our $\psi_{Box \to SBS3}$ prefilter applied to the task of vector graphics rasterization. It defines an oblique projection from the space spanned by box kernels to SBS3 and, as such, combines the no-ringing property of the box space with the sharpness of the SBS3 space. Notice how the resulting image in (a,c) is much sharper than the one obtained by simple box filtering (shown in (b)), but has jagged edges caused by the box prefilter that are sharpened by the digital filter.

Figure 16(a,b) displays an image rendered with the Tungsten software, which we obtained online (https://benedikt-bitterli.me/resources/). It was generated with a tent prefilter. We post-process this image to perform the oblique projection Tent→SBS3, resulting in a significantly sharper and detail-rich rendering, shown in (c)

**(a)** *Original rendering with tent* (b) Zoom of (a) (c) Tent→SBS3

**Figure 16:** *Image rendered from stochastic samples with a tent filter (a,b) and obliquely projected to our SBS3 space (c). Note how edges are much sharper with the use of SBS3, and also how fine texture details—originally blurred in (a,b)—are enhanced in (c).*

(uncropped results available in the supplementary materials). If the scene had been rendered directly with our SBS3 prefilter (Section 5.1), the resulting image would be as-sharp-as (c) but with better anti-aliasing properties, as attested by the metrics in Table 1. However, by performing the oblique projection Tent→SBS3 as a post-processing, one can easily integrate our prefilter with any black-box rendering software. The supplementary materials include video examples showing the application of Tent→SBS3 to short clips rendered by the Blender Foundation, in addition to many more example images.

## 6. Conclusions

We presented an extended view of the image sampling and reconstruction process, where we take into account the observer for designing sharp prefilters for image display. We discussed how to generate optimal images based on this new pipeline, in addition to applications which we think are greatly improved by employing our technique. Our method is general and allows for the definition of a collection of filters based on our SBS3. The proposed filters offer a wide range of options for balancing the trade-off between aliasing and ringing, while keeping the resulting images sharp. We demonstrated this fact through an in-depth analysis and discussion of the properties of the proposed SBS3 family and how it relates to existing filters and theory.

As future work, we are currently investigating the use of non-linear filters to address the fundamental trade-off between aliasing, ringing and blurring. In particular, the challenge of how to prefilter and sample a signal without overblurring, without incurring aliasing, and with no apparent ringing (which is perceptually objectionable), is still an important open problem. The integration of SBS3 with Monte Carlo denoising techniques and the use of the oblique projection operator to mix and match filters (Section 5.3) are also interesting directions of future exploration.

Our work provides an alternative look into the problem of image sampling and reconstruction, which is pervasive in graphics and image processing. As such, we believe it has the potential to stimulate new research, and improvements on applications and display systems.

**Acknowledgements**

We would like to thank the anonymous reviewers for their comments, and the authors of [KSP13,NH14,OG15,WWA*16,SN15,MnBU01] for kindly providing the implementation of their algorithms. This work was sponsored by CNPq-Brazil (fellowships and grants 436932/2018-0, 131102/2018-4, 312975/2018-0, 423673/2016-5), and financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

**Appendix A:** Display-Eye reconstruction is not a sinc

We present a simple experiment which demonstrates that the combined reconstruction performed by a display device and the human visual system is *not* a sinc interpolation. Consider a cosine wave in screen space with a frequency of $19/40$ cycles per pixel, as illustrated here:

Since the frequency of this wave is below the Nyquist limit of 0.5 cycles per pixel, this continuous wave can be reconstructed *exactly* from its samples taken at integer positions, using sinc interpolation, as illustrated below:

By mapping the values of these samples to the intensity values of the pixel columns of a 2-D image, one obtains the following result:

If this image is displayed at exact 1:1 zoom (please see the supplementary materials since PDF readers do interpolation), then theory says that one should observe a perfectly continuous cosine wave **if** reconstruction is done with a sinc. Instead, when looking at this image we see a conspicuous "beating" pattern [Bli89], more precisely known as post-aliasing or reconstruction artifacts [MN88].

This demonstrates that the joint reconstruction done by the display device and the human visual system is *not* sinc interpolation.

**Appendix B:** Measuring the Human PSF in screen units

As discussed in Section 3.1, under constant conditions the eye's PSF has a fixed shape and size on the retina $\Omega$. What does change however is its *relative* size compared to the image projected on $\Omega$: the farther away the display, the smaller the final, projected image will be. More precisely, the image projected on the retina gets smaller as the *viewing distance* $D$ increases and also as the *pixel size* $P$ decreases (as this causes the image itself to be smaller on the display). Thus, the relative size between the projected image and the PSF can be computed in terms of these two quantities. To simplify implementation, we work in screen (pixel) units $(u,v) \in \mathcal{S}$, where the image's size is constant but the PSF's size will vary with $D$ and $P$. Furthermore, this variation is readily seen to be linear [Goo05]. As such, we perform a simple experiment to measure the approximate size of the PSF for some known value of $D$ and $P$, from which its size in any other situation is immediately defined.

An alternating sequence of black and white pixels (0.5 cycles per pixel, a frequency we refer to as $\hat{\omega}$) is displayed on a monitor with $P = 0.25$mm (100 dpi display). We empirically observe that this sequence starts to be perceived as a continuous gray line when viewed by an average person at a distance of about $D = 120$cm. Under these conditions, the eye's PSF removes the high-frequency information of the black-and-white sequence, and consequently its frequency response is (close to) zero at frequency $\hat{\omega}$. This fact and the PSF's shape is all that is necessary to compute the PSF's size.

In order to simplify the Fourier analysis in this derivation, we replace the quadratic B-spline approximation of the PSF in Figure 3 with an equivalent Gaussian model (remember that B-splines of increasing order converge to a Gaussian). Thus, taking the PSF as a Gaussian function, its Fourier transform (also a Gaussian) should have a standard deviation of $\hat{\omega}/3$ units in order to reach close to zero frequency response at $\hat{\omega}$. In the spatial domain such a PSF has a standard deviation of $3/\pi$ pixels. Thus, given a linear pinhole projection model, the standard deviation for an arbitrary viewing distance $D$ (in cm) and pixel size $P$ (in mm) is given by

$$\sigma = \frac{3}{\pi}\frac{(D/120)}{(P/0.25)} = \frac{3}{\pi}\frac{D}{P}\frac{0.25}{120} \text{ (in pixels).} \tag{15}$$

The equivalent quadratic B-spline that models the PSF for this deviation is (for $\alpha = 0.535/\sigma$ obtained by numerical optimization):

$$s(u) = \frac{1}{3}\begin{cases} -4|\alpha u|^2 + 3 & \text{if } \ 0 \le |\alpha u| < 1/2; \\ 2|\alpha u|^2 - 6|\alpha u| + 9/2 & \text{if } 1/2 \le |\alpha u| < 3/2; \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

## References

[AMHH08] AKENINE-MOLLER T., HAINES E., HOFFMAN N.: *Real-Time Rendering, Third Edition*. A K Peters/CRC Press, Wellesley, Mass, July 2008. 4

[Arn19] Arnold Documentation: Pixel Filter. https://docs.arnoldrenderer.com/display/AFHUG/Pixel+Filter, 2019. Accessed: 2019-01-13. 10

[AS00] ATCHISON D. A., SMITH G.: *Optics of the Human Eye*. Butterworth-Heinemann, 2000. 2, 3, 4, 6

[BBD*00] BETRISEY C., BLINN J. F., DRESEVIC B., HILL B., HITCHCOCK G., KEELY B., MITCHELL D. P., PLATT J. C., WHITTED T.: 20.4: Displaced Filtering for Patterned Displays. *SID Symposium Digest of Technical Papers 31*, 1 (2000), 296–299. 3

[Bli89] BLINN J. F.: Jim Blinn's corner-return of the Jaggy (high frequency filtering). *IEEE Computer Graphics and Applications 9*, 2 (Mar. 1989), 82–89. 1, 3, 14

[BTU01] BLU T., THÉVENAZ P., UNSER M.: MOMS: Maximal-order interpolation of minimal support. *IEEE TIP 10*, 7 (July 2001). 3

[BVM*17] BAKO S., VOGELS T., MCWILLIAMS B., MEYER M., NOVÁK J., HARVILL A., SEN P., DEROSE T., ROUSSELLE F.: Kernel-predicting convolutional networks for denoising monte carlo renderings. *ACM Trans. Graph. 36*, 4 (July 2017). 10

[DW85] DIPPÉ M. A. Z., WOLD E. H.: Antialiasing through stochastic sampling. *SIGGRAPH Comput. Graph. 19*, 3 (July 1985), 69–78. 10

[ESKD14] ENGELHARDT T., SCHMIDT T.-W., KAUTZ J., DACHSBACHER C.: Low-Cost Subpixel Rendering for Diverse Displays. *Computer Graphics Forum 33*, 1 (Feb. 2014), 199–209. 3

[Fat09] FATTAL R.: Edge-avoiding wavelets and their applications. *ACM Trans. Graph. 28*, 3 (2009), 1–10. 3

[FEL*11] FARRELL J., ELDAR S., LARSON K., MATSKEWICH T., WANDELL B.: Optimizing subpixel rendering using a perceptual metric. *Journal of the Society for Information Display 19*, 8 (2011), 513–519. 3

[FFLS08] FARBMAN Z., FATTAL R., LISCHINSKI D., SZELISKI R.: Edge-preserving Decompositions for Multi-scale Tone and Detail Manipulation. In *ACM SIGGRAPH 2008 Papers* (New York, NY, USA, 2008), SIGGRAPH '08, ACM, pp. 67:1–67:10. 3

[FG88] FERWERDA J. A., GREENBERG D. P.: A psychophysical approach to assessing the quality of antialiased images. *IEEE Comput. Graph. Appl. 8*, 5 (Sept. 1988), 85–95. 2

[FS10] FRASER B., SCHEWE J.: *Real World Image Sharpening with Adobe Photoshop, Camera Raw, and Lightroom*, 2nd ed ed. Real World. Peachpit Press, Berkeley, Calif, 2010. 4

[GO11] GASTAL E. S. L., OLIVEIRA M. M.: Domain Transform for Edge-Aware Image and Video Processing. *ACM Trans. Graph. 30*, 4 (2011), 69:1–69:12. 00184. 3

[GO17] GASTAL E. S. L., OLIVEIRA M. M.: Spectral remapping for image downscaling. *ACM Trans. Graph. 36*, 4 (2017), 145:1–145:16. Proceedings of SIGGRAPH 2017. 3, 7, 9

[Goo05] GOODMAN J. W.: *Introduction to Fourier optics*. Roberts and Company Publishers, 2005. 4, 14

[Har78] HARRIS F.: On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE 66*, 1 (1978), 51–83. 3

[HF16] HUBERMAN I., FATTAL R.: Reducing lateral visual biases in displays. *Computer Graphics Forum 35*, 8 (2016), 19–31. 3

[HST13] HE K., SUN J., TANG X.: Guided Image Filtering. *IEEE TPAMI 35*, 6 (June 2013), 1397–1409. 3

[HvDM*13] HUGHES J. F., VAN DAM A., MCGUIRE M., SKLAR D. F., FOLEY J. D., FEINER S. K., AKELEY K.: *Computer graphics: principles and practice (3rd ed.)*. Addison-Wesley Professional, July 2013. 10

[KH03] KLOMPENHOUWER M. A., HAAN G. D.: Subpixel image scaling for color-matrix displays. *Journal of the Society for Information Display 11*, 1 (2003), 99–108. 3

[KSP13] KOPF J., SHAMIR A., PEERS P.: Content-adaptive image downscaling. *ACM Trans. Graph. 32*, 6 (2013). 3, 4, 9, 12, 14

[KU81] KAJIYA J., ULLNER M.: Filtering High Quality Text for Display on Raster Scan Devices. In *Proceedings of the 8th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1981), SIGGRAPH '81, ACM, pp. 7–15. 2, 3, 4, 7, 8, 9

[LEU98] LEE C., EDEN M., UNSER M.: High-quality image resizing using oblique projection operators. *IEEE TIP 7*, 5 (May 1998), 679–692. 10

[Maî17] MAÎTRE H.: *From Photon to Pixel: The Digital Camera Handbook*, 2 ed. Wiley-ISTE, Hoboken, NJ, Apr. 2017. 4

[MN88] MITCHELL D. P., NETRAVALI A. N.: Reconstruction filters in computer-graphics. In *ACM Siggraph Computer Graphics* (1988), vol. 22, ACM, pp. 221–228. 1, 2, 3, 4, 6, 7, 9, 14

[MnBU01] MUÑOZ A., BLU T., UNSER M.: Least-squares image resizing using finite differences. *IEEE TIP 10*, 9 (2001). 9, 10, 12, 14

[MS44] MOON P., SPENCER D. E.: On the Stiles-Crawford Effect. *JOSA 34*, 6 (June 1944), 319–329. 4

[NH14] NEHAB D., HOPPE H.: A fresh look at generalized sampling. *Found. Trends. Comput. Graph. Vis. 8*, 1 (Mar. 2014), 1–84. 3, 4, 5, 6, 9, 10, 14

[OG15] ÖZTIRELI A. C., GROSS M.: Perceptually based downscaling of images. *ACM Trans. Graph. 34*, 4 (July 2015), 77:1–77:10. 3, 4, 9, 10, 12, 14

[PJH16] PHARR M., JAKOB W., HUMPHREYS G.: *Physically Based Rendering: From Theory to Implementation*, 3rd ed. 2016. 10

[Pla00] PLATT J. C.: Optimal filtering for patterned displays. *IEEE Signal Processing Letters 7*, 7 (July 2000), 179–181. 3

[PM06] PROAKIS J. G., MANOLAKIS D. K.: *Digital Signal Processing*, 4 edition ed. Pearson, Upper Saddle River, NJ, Apr. 2006. 5

[POAR12] PAMPLONA V. F., OLIVEIRA M. M., ALIAGA D. G., RASKAR R.: Tailored Displays to Compensate for Visual Aberrations. *ACM Trans. Graph. 31*, 4 (July 2012), 81:1–81:12. 4

[Ren19] Renderman 22 Documentation: Filtering. https://rmanwiki.pixar.com/display/REN22/Filtering, 2019. Accessed: 2019-01-13. 10

[RKZ12] ROUSSELLE F., KNAUS C., ZWICKER M.: Adaptive Rendering with Non-local Means Filtering. *ACM Trans. Graph. 31*, 6 (Nov. 2012), 195:1–195:11. 3

[Sha49] SHANNON C. E.: Communication in the presence of noise. *Proceedings of the IRE 37*, 1 (1949), 10–21. 1

[SN15] SACHT L., NEHAB D.: Optimized Quasi-Interpolators for Image Reconstruction. *IEEE TIP 24* (Dec. 2015). 1, 3, 6, 8, 9, 12, 13, 14

[ST85] SCHREIBER W. F., TROXEL D. E.: Transformation Between Continuous and Discrete Representations of Images: A Perceptual Approach. *IEEE TPAMI 7*, 2 (Mar. 1985), 178–186. 2, 8

[Str97] STRANG G.: The Search for a Good Basis. In *Numerical Analysis*. Chapman and Hall/CRC, 1997. 4

[TM98] TOMASI C., MANDUCHI R.: Bilateral filtering for gray and color images. In *ICCV* (1998), vol. 98, p. 2. 3

[UA94] UNSER M., ALDROUBI A.: A general sampling theory for non-ideal acquisition devices. *IEEE Transactions on Signal Processing 42*, 11 (Nov 1994), 2915–2925. 8

[UAE95] UNSER M., ALDROUBI A., EDEN M.: Enlargement or reduction of digital images with minimum loss of information. *IEEE TIP 4*, 3 (March 1995), 247–258. 10

[Uns96] UNSER M.: Quasi-orthogonality and quasi-projections. *Applied and Computational Harmonic Analysis 3*, 3 (1996), 201 – 214. 9

[Uns00] UNSER M.: Sampling-50 years after shannon. *Proceedings of the IEEE 88*, 4 (April 2000), 569–587. 2, 3, 4, 5, 9

[VRM*18] VOGELS T., ROUSSELLE F., MCWILLIAMS B., RÖTHLIN G., HARVILL A., ADLER D., MEYER M., NOVÁK J.: Denoising with kernel prediction and asymmetric loss functions. *ACM Trans. Graph. 37*, 4 (2018), 124:1–124:15. 3

[WBSS04] WANG Z., BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image Quality Assessment: From Error Visibility to Structural Similarity. *Trans. Img. Proc. 13*, 4 (Apr. 2004), 600–612. 4

[WC85] WALSH G., CHARMAN W. N.: Measurement of the Axial Wavefront Aberration of the Human Eye. *Ophthalmic and Physiological Optics 5*, 1 (1985), 23–31. 4

[WWA*16] WEBER N., WAECHTER M., AMEND S. C., GUTHE S., GOESELE M.: Rapid, detail-preserving image downscaling. *ACM Trans. Graph.* (September 2016). 3, 9, 10, 12, 14

[XLXJ11] XU L., LU C., XU Y., JIA J.: Image Smoothing via L0 Gradient Minimization. *ACM Trans. Graph. 30*, 6 (Dec. 2011), 1–12. 3

[ZSXJ14] ZHANG Q., SHEN X., XU L., JIA J.: Rolling guidance filter. In *ECCV 2014*. Springer, 2014, pp. 815–830. 3

[ZW97] ZHANG X., WANDELL B. A.: A spatial extension of CIELAB for digital color-image reproduction. *Journal of the Society for Information Display 5*, 1 (1997), 61–63. 2, 3, 4, 6