

Foveated Nonlocal Self-Similarity

Alessandro Foi¹, Giacomo Boracchi²

¹ Department of Signal Processing, Tampere University of Technology, Tampere, Finland

² Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy

Abstract When we gaze a scene, our visual acuity is maximal at the fixation point (imaged by the fovea, the central part of the retina) and decreases rapidly towards the periphery of the visual field. This phenomenon is known as foveation. We investigate the role of foveation in nonlocal image filtering, installing a different form of self-similarity: the foveated self-similarity. We consider the image denoising problem as a simple means of assessing the effectiveness of descriptive models for natural images and we show that, in nonlocal image filtering, the foveated self-similarity is far more effective than the conventional windowed self-similarity. To facilitate the use of foveation in nonlocal imaging algorithms, we develop a general framework for designing foveation operators for patches by means of spatially variant blur. Within this framework, we construct several parametrized families of operators, including anisotropic ones. Strikingly, the foveation operators enabling the best denoising performance are the radial ones, in complete agreement with the orientation preference of the human visual system.

1 Introduction

Nonlocal self-similarity is among the most effective regularization priors for natural images (Katkovnik et al., 2010; Buades et al., 2010). State-of-the-art algorithms for very different imaging tasks often rely on the assumption that natural images contain a large number of mutually similar patches at different locations (Dabov et al., 2007b; Arias et al., 2011; Milanfar, 2013). To exploit the nonlocal self-similarity, similar and dissimilar patches have to be identified on the basis of a suitable patch distance, meant as inverse similarity between the patches. Choosing such distance corresponds to assuming a specific model for describing natural images and their self-similarity: the effectiveness of nonlocal algorithms depends on the validity of such underlying model.

This work was supported by the Academy of Finland (project no. 252547, Academy Research Fellow 2011-2016).

Patch similarity is typically assessed through the Euclidean distance of the pixel intensities, hence it depends on the patch size. Large patches guarantee stability of the distance with respect to degradations such as noise; however, the mutual similarity between pairs of patches typically decreases as the patch size grows. Thus, a windowed Euclidean distance is commonly used to balance these two conflicting aspects, assigning lower weights to pixels far from the patch center.

Inspired by the human visual system (HVS), we design specific foveation operators to measure the patch similarity. Foveation (Curcio et al., 1990) is a peculiarity of the HVS, which is characterized by maximal acuity at the fixation point, imaged by the fovea, the central part of the retina. Acuity rapidly decreases towards the periphery of the visual field. To reproduce such effect, foveation operators yield patches blurred by a spatially variant blur, where the blur kernels have standard deviation that increases with the spatial distance from the patch center.

We measure patch similarity by means of the foveated distance, i.e. the Euclidean distance between foveated patches. If we consider the patch center as a fixation point, the foveated distance mimics the inability of the HVS to perceive details at the periphery of the center of attention. Thus, foveation becomes an alternative to windowing as a means to attenuate the influence of periphery pixels to the patch similarity assessment. A constrained design of the foveation operators guarantees that whenever the difference between two patches can be modelled as white noise, then the foveated distance is equivalent to the distance induced by any given windowing kernel. Such is for instance the case when the two patches do not exhibit structured differences. In presence of structured differences, such as those arising in the vicinity of edges, the windowed and foveated distances are instead shown to be fundamentally distinct, with the latter providing sharper localized responses.

In the context of nonlocal image modeling, the foveated distance induces a different form of self-similarity, the *foveated self-similarity*. To quantitatively validate the foveated self-similarity as regularization prior for natural images, we investigate the image denoising problem. Indeed, the removal

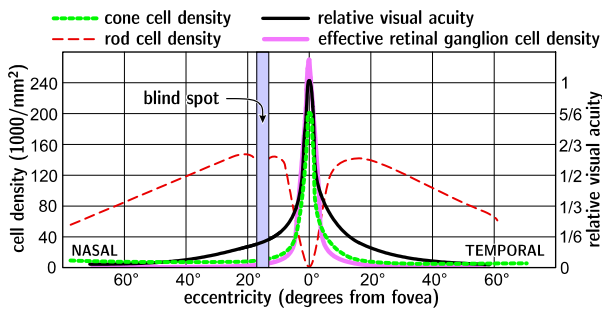


Fig. 1 Cell density and visual acuity vs eccentricity. The visual acuity as well as the cone and ganglion density peak at the fovea and decrease rapidly as we move toward the periphery of the retina. Plot based on data from Curcio et al. (1990), Garway-Heath et al. (2000), and Wertheim (1894).

of additive white Gaussian noise is the most widely used task for assessing the validity of any descriptive or generative model of natural images (e.g., Roth and Black, 2005; Weiss and Freeman, 2007; Li and Huttenlocher, 2008; Lyu and Simoncelli, 2009; Ranzato et al., 2013). As a meaningful prototype of nonlocal image filtering algorithms, we specifically consider the Nonlocal Means (NL-means) (Buades et al., 2005) and propose *Foveated NL-means*, where the foveated distance replaces the conventional windowed distance.

Extensive experiments show that Foveated NL-means can substantially outperform the standard one, in terms of both objective criteria and visual appearance: Foveated NL-means achieves better contrast and sharpness in the denoised images, particularly at edges and sharp details. We interpret the superior performance of Foveated NL-means in terms of low-level vision, as an improved ability of the foveated distance at identifying similar patches and at distinguishing different ones in noisy environments. This makes the foveated self-similarity a far more effective regularization prior for natural images than the conventional windowed self-similarity. To further investigate the role that foveation plays in assessing nonlocal self-similarity, we design a general family of foveation operators, including anisotropic ones, which embed directional blur kernels. Remarkably, the foveated distance is even more effective when the blur kernels of foveation operators follow radial directions, matching the orientation preference in the HVS (Sasaki et al., 2006; Freeman et al., 2011). Thus, foveated self-similarity reveals a connection between features of the HVS and modern imaging algorithms.

This work builds upon our preliminary conference publications: Foi and Boracchi (2012), where we first introduced isotropic foveation operators and their constrained designed for Foveated NL-means; and Foi and Boracchi (2013b,a) where we described their anisotropic extensions. Here, we develop these works in the following directions: we provide a more precise and detailed characterization of the design constraints leading to foveation operators (both isotropic and anisotropic); introduce self-map foveation operators that share the same compact support of windowing operators; illustrate the foveation operators through their singular value decomposition; present a more thorough and extensive experimental



Fig. 2 *Lena* image foveated at two different fixation points.

validation which includes a detailed analysis and optimization of the filter parameters; detail the computational overhead of replacing the windowed distance with the foveated distance; discuss several connections between the HVS and the effectiveness of foveated self-similarity, including a direct analogy between the radial bias in the HVS and the optimal orientation in anisotropic foveation, which we expose by replicating the outcomes of a recent fMRI study on human subjects within our image-denoising framework.

The remainder of the paper is structured as follows: Section 2 introduces foveation and self-similarity in the HVS and in the image-processing literature, while Section 3 presents the observation model and the NL-means algorithm. Sections 4 and 5 introduce the patch foveation and a constrained design of foveation operators; meaningful examples are then extensively developed in Section 6, including the anisotropic operators. Section 7 introduces self-map foveation operators, while Section 8 describes Foveated NL-means, which will be extensively used in the experiments of Section 9. Discussions and concluding remarks follow in Sections 10 and 11.

2 Background

2.1 Foveation

Because of the uneven size and organization of photo-receptive cells, the retinal image features various spatially variant properties (Wertheim, 1894; Curcio et al., 1990; Wandell, 1995; Eckstein, 2011; Joselevitch, 2008). Notably, the visual acuity (and hence the sharpness of the resulting retinal image) is highest at the middle of the retina, in a depression termed *fovea*, where the concentration of cones has its peak. At the periphery of the retina, vision is instead blurry. Thus, the retinal image is sharp at the center of gaze (point of fixation) and becomes progressively blurred or confused as the distance from the center increases. This phenomenon, termed *foveated vision*, *foveated imaging*, or simply *foveation*, is illustrated in Figures 1 and 2 (Wertheim, 1894; Curcio et al., 1990; Williams et al., 1996; Jennings and Charman, 1997).

The non-uniformity of the photo-receptive cells is not the unique cause of foveation. In fact, foveation is the result of a cascade of various space-variant optical, sampling, and processing contributors. When the light enters the eye, it is focused on the retina by the cornea and by the crystalline lens. This biological optical system is able to provide accurate focusing only in the vicinity of the fovea, thus the per-

foveal image (i.e. the image at the periphery of the visual field) would be blurry even if the photoreceptor density were high throughout the retina. Interestingly, this imperfect focus can be interpreted as an optical low-pass filter that precedes the coarse sampling operated by the photoreceptive cells, in order to prevent aliasing. Moreover, also ganglion cells in the human eye have non-uniform distribution and, most importantly, non-uniform number of connections depending on their eccentricity with respect to the fovea: far from the fovea there are fewer ganglions, each of which is connected to more photoreceptors, thus driving the acuity further down (Duncan and Boynton, 2003; Joselevitch, 2008). While in our work we adopt the traditional and simplest characterization of peripheral vision as blurry (i.e. low pass), we note that the complex multi-layer structure of the HVS allows more sophisticated descriptions (e.g., Strasburger et al., 2011; Balas et al., 2009; Freeman and Simoncelli, 2011; Rosenholtz et al., 2012).

2.1.1 Radial orientation bias The spatial non-uniformity of our visual acuity is also characterized by interesting anisotropic features. Several studies have investigated the HVS sensitivity to differently oriented stimuli at varying eccentricity with respect to the fixation point (such as those in Figure 3): there is consistent evidence of a *radial orientation bias* (Sasaki et al., 2006), i.e. our acuity is higher for patterns oriented along *radial lines* or *meridians* (i.e. along directions toward the center of gaze, see Figure 3, left), while lower for *tangential* patterns. For instance, we are able to detect and discriminate patterns such as the radial patterns A and C better than the tangential patterns B and D of Figure 3(right). The radial orientation bias can be found at various levels of the human and mammal visual system, from the responses of retinal ganglion cells (Levick and Thibos, 1982), up to the visual cortex through functional magnetic resonance imaging (fMRI) measurements (Freeman et al., 2011), as well as indirectly through various psychophysical discrimination and detection tasks (e.g., Toet and Levi, 1992).

There are many anatomical and physiological explanations for foveation, including the limited capacity of the optic nerve and visual cortex as well as improved low-light vision (Joselevitch, 2008). Analogous grounds have motivated also the use of foveated imaging in image processing. In particular, our image-filtering experiments reported in Sections 9 and 10 attest to the advantage of foveation, especially from anisotropic radial operators, hence corroborating a functional justification for foveation and the radial bias.

2.2 Foveation in image processing

Foveated images can be acquired by means of specific sensors that sample the visual field along a logarithmic polar grid (Weiman and Chaikin, 1979; Wodnicki et al., 1995) or other irregular photosite configurations (Wallace et al., 1994; Etienne-Cummings et al., 2000). However, in applications, foveation is most often achieved by digitally post-processing

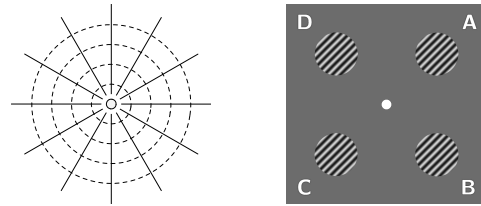


Fig. 3 Radial orientation bias: the HVS is more sensitive and acute to patterns that, like A and C, are oriented along radial lines or meridians (i.e. along directions toward the fixation point, drawn as solid lines in the diagram on the left), and less to tangential patterns such as B and D.

images acquired from standard regular-grid sensors. Techniques for a space-variant representation of images include: averaging the image intensities over areas (super-pixels) defined according to a space-variant grid (Kortum and Geisler, 1996), masking a wavelet decomposition (Wei and Li, 1998; Chang et al., 2000), and using spatially variant blur (Lee et al., 2001; Foi and Boracchi, 2012). Other space-variant techniques inspired by the topological layout of the HVS are the Reciprocal-Wedge Transform (Tong and Li, 1995) and variable-resolution models (Basu and Wiebe, 1998).

The most appealing application for foveated imaging is perhaps compression (Kortum and Geisler, 1996). In fact, any user gazing a screen would not notice significant differences between the fully detailed image properly displayed and the same image foveated with respect to the fixation point. Therefore, in those application scenarios where the user's fixation point can be tracked or reliably estimated (e.g., in video conferencing), images and videos can be compressed by encoding their foveated counterparts, which substantially reduce the high-frequency content at the periphery of visual field without degrading perceived quality (Geisler and Perry, 1998; Basu and Wiebe, 1998; Daly et al., 2001; Lee et al., 2001). Foveation has also inspired image (Wang and Bovik, 2001) and video (Wang and Bovik, 2005) coding algorithms, and was used in robotics to actively control calibrated binocular vision systems (Wei and Li, 1998), and to reduce the search range for point matching in uncalibrated ones (Monaco et al., 2009).

All the above methods apply foveation on the whole image: to the best of our knowledge, beside our preliminary conference publications (Foi and Boracchi, 2012, 2013b,a) and our recent work on data-driven dictionaries based on foveated principal components (Foi and Boracchi, 2014), foveation at the patch level appeared only in the fast retina keypoint descriptor (Alahi et al., 2012), which was designed mimicking the topology of ganglions in the retina. Each descriptor encodes the responses of several pairs of *receptive fields*, which are obtained by convolving the image against Gaussian kernels having standard deviation that increases with the distance from the keypoint.

In contrast, we leverage patch foveation to investigate the foveated self-similarity in natural images. In particular, ours is the first study concerning how foveation can improve filtering algorithms based on self-similarity.

2.3 Narrow visual field and saccadic eye movements

The area imaged by the fovea occupies only a narrow portion of the visual field. In order to form a complete and detailed image of the surrounding environment, the HVS resorts to *saccades*, i.e. ballistic eye movements that frequently reposition our gaze. At every saccade the whole retinal image changes and our eyes experience a different view. In fact, the eye is under steady motion also because of small fixational eye-movements such as tremor, slow drift, and microsaccades (Kowler, 2011). Due to the short duration of visual perception, the discontinuous input which follows from the motion must be linked by the brain. Stable and persistent perception of visual space requires that features in the new retinal image are associated with corresponding features in the previous retinal image (Brockmole and Irwin, 2005). In a very broad sense, the HVS operates as a patch-based image-processing system, with the complete global image described through a multitude of localized fragments.

The influence of saccadic and microsaccadic eye movements on visual perception are far from being completely understood and are subject of active investigation, mainly by the visual neuroscience and psychophysics research community. A comprehensive review is provided, e.g., by Desbordes (2007), Kowler (2011), and Martinez-Conde et al. (2013), testifying a growing consensus about the functional importance of saccades and microsaccades. In particular, it was shown that not only these movements prevent retinal image fatigue and visual fading, but play also a crucial role in controlling the fixation position (Engbert and Kliegl, 2004). Experiments (see, e.g., Donner and Hemilä, 2007; Rucci et al., 2007; McCamy et al., 2012; Poletti et al., 2013) reveal that the retinal responses caused by microsaccadic movements reduce aliasing effects in the peripheral retina, lead to improved spatiotemporal accuracy of the neural responses, and provide additional information that allows resolving fine spatial details out of optically blurred images.

It is far beyond the scope of this manuscript to investigate such intricate models of the eye movements and their interplay with low-level vision. However, the works cited above are suggestive of an intimate functional affinity between nonlocal image processing and HVS mechanisms that cause or leverage saccades and fixational eye movements (see, e.g., Olmedo-Payá et al., 2013). Since a patchwise image modeling is at the root of nonlocal methods, and because patches can be put into correspondence with the narrow visual field during fixations, our motivating idea is to endow the nonlocal methods with foveation, as this is the key feature characterizing the narrowness of the human visual field. To the best of our knowledge, this has never been investigated before, which is perhaps surprising given the burgeoning of nonlocal image processing methods during the last decade.

2.4 Nonlocal image processing

The use of nonlocal self-similarity in image processing can be traced back to fractal models of natural images (Barnsley,

1993) and fractal block coding (Jacquin, 1992), where it was argued that natural images can be compressed by expressing their self-similarity as “self-transformability on a blockwise basis”. Nevertheless, the nonlocal self-similarity gained significance later on, with the inspiring works on texture synthesis and completion (Efros and Leung, 1999; Wei and Levoy, 2000), which developed the intuition that natural images contain highly correlated patches at different, possibly far, locations. In (Zhang and Wang, 2002), the same paradigm is applied to image filtering and particularly for inpainting. In particular, it is shown that if the outer portion of two patches is similar, then the inner portion of one can be used to restore the inner portion of the other, in case this is missing, saturated or corrupted. Nonlocal filtering and, more generally, patch-based algorithms became an established paradigm, successfully applied to a broad range of imaging applications.

For the specific task of noise attenuation, nonlocality has been initially considered in (De Bonet, 1997), and later developed in (Ghazel et al., 2003; Buades et al., 2005; Awate and Whitaker, 2005). In particular, the NL-means algorithm (Buades et al., 2005) represented a breakthrough in image denoising and inspired several powerful algorithms in the following years, such as BM3D (Dabov et al., 2007b) and SAFIR (Kervrann and Boulanger, 2006). For a comprehensive overview on modern denoising approaches and algorithms we refer the reader to the review articles by Katkovnik et al. (2010) and Milanfar (2013).

Patch-based methods enforcing nonlocal self-similarity have been developed for video (Dabov et al., 2007a; Buades et al., 2008; Maggioni et al., 2012), surface (Dong et al., 2008), and point-cloud (Rosman et al., 2013) denoising, image deblurring (Kindermann et al., 2005; Dabov et al., 2008; Danielyan et al., 2012), demosaicking (Buades et al., 2009), inpainting (Arias et al., 2011), zooming (Freeman et al., 2002; Ebrahimi and Vrscaj, 2007) and super-resolution from a sequence of low-resolution images (Danielyan et al., 2008; Ebrahimi and Vrscaj, 2008b; Elad and Datsenko, 2009; Protter et al., 2009). The same *nonlocal paradigm* has been also applied for the regularization of inverse problems (Peyré, 2009b; Gilboa and Osher, 2009; Danielyan et al., 2012) and for texture synthesis (Peyré, 2009a).

3 Patch-based self-similarity

3.1 Observation model

We consider observations as noisy grayscale images $z : X \rightarrow \mathbb{R}$ that can be modeled as

$$z(x) = y(x) + \eta(x), \quad x \in X \subset \mathbb{Z}^2, \quad (1)$$

where $X \subset \mathbb{Z}^2$ is the image domain, which is assumed a regular pixel grid, $y : X \rightarrow \mathbb{R}$ is the unknown original image, and $\eta : X \rightarrow \mathbb{R}$ is i.i.d. Gaussian white noise, $\eta(\cdot) \sim \mathcal{N}(0, \sigma^2)$.

For the sake of simplicity and notation, we assume that images can be extended beyond the boundary of X to the whole \mathbb{Z}^2 through any standard padding technique¹.

¹ In our experiments we resort to symmetric padding.

3.2 Patching

Let $U \subset \mathbb{Z}^2$ be a neighborhood centered at the origin; we define the patch centered at a pixel $x \in X$ in the noisy observation z as

$$\mathbf{z}_x(u) = z(u + x), \quad u \in U. \quad (2)$$

Likewise, we can define the noise-free patch

$$\mathbf{y}_x(u) = y(u + x), \quad u \in U. \quad (3)$$

Let \mathbb{I}_X denote the space of images with domain X and \mathbb{I}_U denote the space of patches with domain U . The extraction of patches (2) can be conveniently represented by a *patching operator* \mathcal{P}_U that maps any image/point pair $(z, x) \in \mathbb{I}_X \times X$ to a patch $\mathbf{z}_x \in \mathbb{I}_U$:

$$\mathcal{P}_U : \mathbb{I}_X \times X \rightarrow \mathbb{I}_U, \quad \mathcal{P}_U[z, x] = \mathbf{z}_x. \quad (4)$$

Indeed \mathcal{P}_U can be used on any image, thus we can write (3) as $\mathbf{y}_x = \mathcal{P}_U[y, x]$. Let us also observe that the patching operator

$$\mathcal{P}_X : \mathbb{I}_X \times X \rightarrow \mathbb{I}_X \quad (5)$$

is trivially a translation operator.

3.3 Windowed patch distance

The standard approach for assessing the similarity between two patches \mathbf{z}_{x_1} and \mathbf{z}_{x_2} is based on the windowed quadratic distance

$$\begin{aligned} d^{\text{WIN}}(x_1, x_2) &= \|(\mathbf{z}_{x_1} - \mathbf{z}_{x_2})^2 \mathbf{k}\|_1 = \\ &= \sum_{u \in U} (z(u + x_1) - z(u + x_2))^2 \mathbf{k}(u), \end{aligned} \quad (6)$$

where \mathbf{k} is a non-negative *windowing kernel* defined over U , and $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the usual ℓ_1 and ℓ_2 norms, respectively. The windowing kernel adjusts the contribution of each summand in (6) depending on the position $u \in U$.

We refer to the patches $\mathbf{z}_x \sqrt{\mathbf{k}} = \mathbf{z}_x^{\text{WIN}}$ as the *windowed patches*. With this notation, (6) can be compactly written as

$$d^{\text{WIN}}(x_1, x_2) = \|\mathbf{z}_{x_1}^{\text{WIN}} - \mathbf{z}_{x_2}^{\text{WIN}}\|_2^2. \quad (7)$$

Formally, the windowed patches can be obtained from a windowing operator \mathcal{W}_U that maps any image/point pair $(z, x) \in \mathbb{I}_X \times X$ to a windowed patch $\mathbf{z}_x \sqrt{\mathbf{k}} \in \mathbb{I}_U$:

$$\mathcal{W}_U : \mathbb{I}_X \times X \rightarrow \mathbb{I}_U, \quad \mathcal{W}_U[z, x] = \mathbf{z}_x \sqrt{\mathbf{k}} = \mathbf{z}_x^{\text{WIN}}. \quad (8)$$

Naturally, \mathcal{W}_U can be seen as the left composition “ \circ ” of \mathcal{P}_U (4) with an operator $\mathcal{K} : \mathbb{I}_U \rightarrow \mathbb{I}_U$ that multiplies \mathbf{z}_x by $\sqrt{\mathbf{k}}$:

$$\mathcal{W}_U = \mathcal{P}_U \circ \mathcal{K}, \quad \mathcal{K} : \mathbf{z}_x \mapsto \mathbf{z}_x \sqrt{\mathbf{k}} = \mathbf{z}_x^{\text{WIN}}.$$

3.3.1 Expectation of windowed distance The mathematical expectation of (6) can be expressed as

$$\begin{aligned} E\{d^{\text{WIN}}(x_1, x_2)\} &= \\ &= \sum_{u \in U} (E\{z(u + x_1) - z(u + x_2)\})^2 \mathbf{k}(u) + \\ &+ \sum_{u \in U} \text{var}\{z(u + x_1) - z(u + x_2)\} \mathbf{k}(u) = \\ &= \| (E\{\mathbf{z}_{x_1} - \mathbf{z}_{x_2}\})^2 \mathbf{k}\|_1 + \|\text{var}\{\mathbf{z}_{x_1} - \mathbf{z}_{x_2}\} \mathbf{k}\|_1, \end{aligned} \quad (9)$$

where E and “var” denote the expectation and variance with respect to different realizations of the noise at a fixed spatial position. In the remainder, unless otherwise stated, we restrict ourselves to the case $x_1 \neq x_2$.

If we follow the simple deterministic model for the noise-free data y , the patch difference may be written as

$$\mathbf{z}_{x_1} - \mathbf{z}_{x_2} = \mathbf{y}_{x_1} - \mathbf{y}_{x_2} + \sqrt{2}\sigma n, \quad (10)$$

where $E\{\mathbf{z}_x\} = \mathbf{y}_x$ and $n(\cdot) \sim \mathcal{N}(0, 1)$ is standard Gaussian noise. Hence $E\{\mathbf{z}_{x_1} - \mathbf{z}_{x_2}\} = \mathbf{y}_{x_1} - \mathbf{y}_{x_2}$, $\text{var}\{\mathbf{z}_{x_1} - \mathbf{z}_{x_2}\} = 2\sigma^2$, and thus (9) becomes

$$E\{d^{\text{WIN}}(x_1, x_2)\} = \|(\mathbf{y}_{x_1} - \mathbf{y}_{x_2})^2 \mathbf{k}\|_1 + 2\sigma^2 \|\mathbf{k}\|_1.$$

Further, in the *ideal case of perfect nonlocal self-similarity* where noise-free patches \mathbf{y}_{x_1} and \mathbf{y}_{x_2} are identical, i.e. $E\{\mathbf{z}_{x_1}\} = E\{\mathbf{z}_{x_2}\}$, the expectation (9) reduces to

$$E\{d^{\text{WIN}}(x_1, x_2)\} = 2\sigma^2 \|\mathbf{k}\|_1. \quad (11)$$

This simple equality shall play an essential role in the mathematical development of the proposed foveation framework.

3.4 Nonlocal means

For the sake of simplicity and to elucidate our contribution, we focus our attention to the NL-means algorithm (Buades et al., 2005). In its basic implementation, the NL-means admits the following simple and elegant formulation. The denoised image \hat{y}^{WIN} consists of a weighted average of potentially all the image pixels, i.e.

$$\hat{y}^{\text{WIN}}(x_1) = \sum_{x_2 \in X} w^{\text{WIN}}(x_1, x_2) z(x_2) \quad \forall x_1 \in X, \quad (12)$$

where $\{w^{\text{WIN}}(x_1, x_2)\}_{x_2 \in X}$ is the set of adaptive weights that characterize the pixel x_1 , which are positive and sum to one. Each weight $w^{\text{WIN}}(x_1, x_2)$ is determined by the similarity between the two patches \mathbf{z}_{x_1} and \mathbf{z}_{x_2} as

$$w^{\text{WIN}}(x_1, x_2) = e^{-\frac{d^{\text{WIN}}(x_1, x_2)}{h^2}} / \sum_{x \in X} e^{-\frac{d^{\text{WIN}}(x_1, x)}{h^2}}, \quad (13)$$

where $d^{\text{WIN}}(x_1, x_2)$ is the windowed distance (6) between image patches centered at x_1 and x_2 , and $h > 0$ is a filtering parameter controlling the decay of the exponential function. Equation (13) assigns larger weights to those $z(\cdot)$ in (12) that correspond to pixels belonging to similar patches (i.e.

where the distance between patches $d^{\text{WIN}}(x_1, x_2)$ is small), regardless of their location within the image. This explains the nonlocal denomination of the algorithm. Thanks to such a weighting mechanism, the NL-means enforces the self-similarity of natural images, which, as discussed in Section 2.4 is an effective regularity prior for suppressing the noise.

Even though the summation (12) is formally over the whole image domain X , in practice, for each x_1 , only pixels x_2 belonging to a finite *search neighborhood* around x_1 are taken into consideration: this simplification is obviously justified by computational issues, but also – and perhaps most importantly – by the increased risk of mismatching when searching similar patches under noisy conditions over larger areas, as further discussed in Section 9.1.

In their original formulation of NL-means, Buades et al. (2005) suggest using a Gaussian function as windowing kernel k in (6), while in practice various windowing kernels having maximum at the patch center are also adopted (Manjon-Herrera and Buades, 2008).

The motivating idea of (6) in NL-means is very practical: the similarity between the unknown $y(x_1)$ and $y(x_2)$ can be assessed by computing the similarity of the corresponding noisy patches z_{x_1} and z_{x_2} . The influence of noise is reduced by evaluating the similarity over patches, and the decay of k reflects the extent to which similarity between $y(u+x_1)$ and $y(u+x_2)$ at a given $u \neq 0$ implicates similarity between $y(x_1)$ and $y(x_2)$.

NL-means can be seen as a modification of the bilateral filter (Yaroslavsky, 1985; Tomasi and Manduchi, 1998), where the photometric difference between two pixels $|z(x_1) - z(x_2)|^2$ is replaced by the windowed distance of two patches around them (6), and the spatial distance $|x_1 - x_2|$ between the pixels is not taken into account.

The above equations point out that the patches are compared in pixelwise manner, and that all the information from the full high-resolution image is conveyed into the similarity test (6), which indeed determines the filtering at the central pixel only. The likely variations in the high-frequency content at the patch periphery may therefore prevent the joint nonlocal filtering of otherwise mutually similar patches.

3.5 Extensions of the basic NL-means algorithm

The literature boasts several modifications to the basic NL-means algorithm. Here, we mention only the most important ones that aim at speeding up the computations, at introducing adaptive mechanisms to define the algorithm parameters, at enhancing with invariant properties the patch-similarity measure, or at modifying the filtering stage.

The computational burden of NL-means mainly resides in the computation of weights w^{WIN} (13). Hence, fast implementations and variants of NL-means adopt strategies such as the integral-images method (Wang et al., 2006; Darbon et al., 2008; Froment, 2014) or criteria for restricting the computation of weights to the most similar patches only. For instance, Buades et al. (2005) present a multiscale so-

lution where similar patches are first searched within a sub-sampled image, and then the weights are refined on the full-size noisy image, where the filtering is also performed. Similarly, Mahmoudi and Sapiro (2005) introduce a patch pre-classification step to exclude completely unrelated pixels from the search neighborhood. A different acceleration strategy consists in comparing the projections of patches onto a low-dimensional subspace, e.g., via a singular value decomposition (Orchard et al., 2008) or principal component analysis (Tasdizen, 2009). Further references concerning fast variants of NL-means are provided by Froment (2014).

Kervrann and Boulanger (2006) propose an extension of NL-means where the filtering parameters as well as the patch size vary spatially depending on the image content, while Salmon (2010) presents a numerical study on the selection of the search-neighborhood size, and a new criterion to set the weight $w^{\text{WIN}}(x_1, x_1)$ for the central pixel. These issues are addressed also by Grewenig et al. (2011). Jin et al. (2011) define adaptive weights w^{WIN} to minimize a tight upper bound on the mean squared error (MSE). Maleki et al. (2013) propose an “Anisotropic NL-means” that adaptively steers and elongates the patch support U along image edges.

The NL-means algorithm understands image self-similarity in a strict translational sense. This can be relaxed by introducing more general self-similarity measures. For instance, patches could be considered similar up to a constant additive shift or multiplicative scaling of their intensities, as to compensate different illumination in the scene. In addition, akin to fractal techniques, similar content may be sought among patches of different size. Models encompassing these forms of patch similarity were introduced by Alexander et al. (2008) and La Torre et al. (2009), showing that natural images in general contain a considerable degree of such self-similarity. Ebrahimi and Vrscay (2008a) analyzed the performance of NL-means when patches are taken at different scales. Lou et al. (2009) extended the patch similarity up to rigid transformations and scaling, by leveraging SIFT descriptors (Lowe, 2004) to preliminarily map patches in a canonical form independent of scale and orientation. Modified similarity measures that handle rotated patches were also presented by Thaipanich et al. (2010), exploiting patch clustering, by Grewenig et al. (2011), aligning rotated patches with the aid of patch centroids or structure tensors, and by Ji et al. (2009), replacing the windowed patch distance by the Euclidean difference of rotationally invariant Zernike moments of the patches. Alternative measures of patch similarity satisfying certain optimality properties under non-Gaussian noise were developed by Deledalle et al. (2012).

A generalization of the filtering stage in NL-means was introduced by Chatterjee and Milanfar (2008); in their work, the sum in (12) is replaced by a high-order kernel regression, improving the denoising in texture regions. Patchwise variants, where the pixelwise estimate (12) is replaced by the weighted sum of the whole patches, were presented by Buades et al. (2005, 2011) and Kervrann et al. (2007).

The BM3D family of algorithms (Dabov et al., 2007b, 2009) combined synergistically the use of nonlocal self-

similarity, patch-based estimates and transform domain processing, achieving the ultimate denoising performance (Chatterjee and Milanfar, 2010; Katkovnik et al., 2010; Levin et al., 2012; Milanfar, 2013).

4 Foveated patch distance

As a direct way of endowing nonlocal methods with foveation, we replace the windowed distance $d^{\text{WIN}}(x_1, x_2)$ (7), by the *foveated distance*

$$d^{\text{FOV}}(x_1, x_2) = \|\mathbf{z}_{x_1}^{\text{FOV}} - \mathbf{z}_{x_2}^{\text{FOV}}\|_2^2, \quad (14)$$

where $\mathbf{z}_x^{\text{FOV}} : U \rightarrow \mathbb{R}$ is a *foveated patch* obtained by foveating the image z at the fixation point x . This foveation is accomplished through a specially designed *patch foveation operator* $\mathcal{F} : \mathbb{I}_X \times X \rightarrow \mathbb{I}_U$ that maps any image/fixation-point pair $(z, x) \in \mathbb{I}_X \times X$ to a patch $\mathbf{z}_x^{\text{FOV}} \in \mathbb{I}_U$:

$$\mathbf{z}_x^{\text{FOV}}(u) = \mathcal{F}[z, x](u), \quad u \in U. \quad (15)$$

To reproduce the foveation properties of the HVS when fixating a point $x \in X$, $\mathcal{F}[\cdot, x]$ is expected to work as a spatially variant blurring operator with increasing blur (i.e. decreasing bandwidth) as we leave x . This means that $\mathbf{z}_x^{\text{FOV}}(u)$ is, compared to $\mathbf{z}_x(u)$, progressively blurrier as $|u|$ grows.

5 Constrained design of the foveation operator

Formally, the foveated distance d^{FOV} (14) can replace the windowed distance d^{WIN} in any nonlocal method that relies on such a pairwise patch comparison. However, without a careful design of the operator \mathcal{F} , replacing d^{WIN} with d^{FOV} would end up being ad-hoc, subject to many arbitrary parameters, and hence its impact hard to interpret. Therefore, we propose a constrained design, which associates to any windowing kernel \mathbf{k} a specific foveation operator² \mathcal{F} . In particular, we request that the windowed patch distance d^{WIN} corresponding to \mathbf{k} and the foveated patch distance d^{FOV} induced by the associated \mathcal{F} are equivalent when comparing patches whose difference can be treated as white noise. Conversely, in presence of structured differences, such as those arising in the vicinity of edges, the two distances d^{WIN} and d^{FOV} depart from each other.

The following five³ constraints are imposed on \mathcal{F} to meet the above goal through a practical and rigorous design. First, we introduce the constraints as properties the operator \mathcal{F} must fulfill, and then, in Section 5.2, we characterize \mathcal{F} as a spatially variant blur operator, and reformulate each constraint in terms of the blur kernels that constitute \mathcal{F} . This leads to a transparent construction of the foveation operators, which is summarized in Section 5.3.

² Because our foveation operators are always designed upon a given windowing kernel \mathbf{k} , a more precise symbol would be $\mathcal{F}_{\mathbf{k}}$. However, not to overload the notation, we omit the subscript \mathbf{k} and leave room to other decorations that are used in the later sections.

³ These five requirements include the four requirements described in our preliminary conference publications (Foi and Boracchi, 2012, 2013b,a), plus a *non-negativity* requirement, which had been always tacitly assumed but never mentioned explicitly.

5.1 The five constraints

5.1.1 Linearity $\mathcal{F} : \mathbb{I}_X \times X \rightarrow \mathbb{I}_U$ is linear with respect to \mathbb{I}_X and translation invariant with respect to X , i.e.

$$\begin{aligned} \mathcal{F}[\lambda_1 z_1 + \lambda_2 z_2, x - \tau] &= \\ &= \lambda_1 \mathcal{F}[z_1(\cdot + \tau), x] + \lambda_2 \mathcal{F}[z_2(\cdot + \tau), x], \end{aligned} \quad (16)$$

for any $\lambda_1, \lambda_2 \in \mathbb{R}$, and $\tau \in \mathbb{Z}^2$, where $z_1, z_2 \in \mathbb{I}_X$ is an arbitrary pair of images and $x \in X$ is any fixation point. This translation invariance is very natural and it simply means that if we translate of the same shift τ both the image and the fixation point, then the foveated patch does not change.

5.1.2 Non-negativity For any non-negative image the foveated patches are always non-negative, i.e.

$$\begin{aligned} \text{if } z(x) \geq 0 \quad \forall x \in X, \text{ then} \\ \mathcal{F}[z, x](u) \geq 0 \quad \forall u \in U, \forall x \in X. \end{aligned} \quad (17)$$

5.1.3 Central acuity Foveated patches are fully sharp at their center, i.e.

$$\exists \alpha > 0 : \mathcal{F}[z, x](0) = \alpha z(x) \quad \forall x \in X. \quad (18)$$

This property aims at mimicking the peak of the visual acuity at the fovea, illustrated in Figure 1. The constant α is a crucial design parameter of the foveation operator and its precise value will be determined in Section 5.2.5.

5.1.4 Flat-field preservation \mathcal{F} maps a flat image into flat patches, i.e.

$$\begin{aligned} \exists \alpha > 0 : \forall c > 0 \text{ if } z(x) = c \quad \forall x \in X, \text{ then} \\ \mathcal{F}[z, x](u) = \alpha c \quad \forall u \in U, \forall x \in X. \end{aligned} \quad (19)$$

While this property might appear as rather natural, it is perhaps striking that it is actually verified only seldom in the inner computations of image processing algorithms. For instance, any multiplication against a non-uniform windowing kernel \mathbf{k} such as in (6) prevents this property. Of course, the constant α appearing in (18) and (19) must be the same.

5.1.5 Compatibility Let $\mathbf{z}_{x_1}^{\text{FOV}}, \mathbf{z}_{x_2}^{\text{FOV}}$ be two foveated patches and $E\{\mathbf{z}_{x_1}^{\text{FOV}}\}, E\{\mathbf{z}_{x_2}^{\text{FOV}}\}$ their respective expectations. In the ideal case where the expectations are perfectly identical, $E\{\mathbf{z}_{x_1}^{\text{FOV}}\} = E\{\mathbf{z}_{x_2}^{\text{FOV}}\}$, the expectation of the foveated distance d^{FOV} (14), should match the ideal one of d^{WIN} (11), i.e.

$$\begin{aligned} \text{if } E\{\mathbf{z}_{x_1}^{\text{FOV}}\} = E\{\mathbf{z}_{x_2}^{\text{FOV}}\}, \text{ then} \\ E\left\{\|\mathbf{z}_{x_1}^{\text{FOV}} - \mathbf{z}_{x_2}^{\text{FOV}}\|_2^2\right\} = 2\sigma^2 \|\mathbf{k}\|_1. \end{aligned} \quad (20)$$

For a linear \mathcal{F} , (20) means that for any pair of identical foveated noise-free patches $\mathcal{F}[y, x_1] = \mathcal{F}[y, x_2]$, the expectation of the corresponding foveated distance $d^{\text{FOV}}(x_1, x_2)$ should be equal to the expectation (11) of the windowed distance (6) when the noise-free patches coincide, $\mathbf{y}_{x_1} = \mathbf{y}_{x_2}$.

Compatibility can be formulated also in the following stricter *pixelwise* form:

$$\text{if } E\{\mathbf{z}_{x_1}^{\text{FOV}}\} = E\{\mathbf{z}_{x_2}^{\text{FOV}}\}, \text{ then} \\ E\left\{\left(\mathbf{z}_{x_1}^{\text{FOV}}(u) - \mathbf{z}_{x_2}^{\text{FOV}}(u)\right)^2\right\} = 2\sigma^2 \mathbf{k}(u) \quad \forall u \in U. \quad (21)$$

From the very definition of the ℓ^2 norm, i.e. by summation of (21) over all $u \in U$, it is clear that (21) implies (20).

As we shall see, this stricter form of compatibility indirectly installs in the foveated patches the progressive decay of visual acuity from the center towards the periphery of the fovea illustrated in Figure 1.

From a technical point of view, compatibility allows using d^{FOV} as a direct replacement of the usual windowed distance d^{WIN} in (13), without need, e.g., of adjusting the tuning parameter h , since both distances output on comparable ranges.

5.2 Defining the foveation operator through its blur kernels

Let us analyze the implications of the above requirements.

5.2.1–2 Linearity and non-negativity (Blurring operator)

An operator \mathcal{F} satisfies the linearity requirement if and only if it can be expressed in the form

$$\mathcal{F}[z, x](u) = \sum_{\xi \in \mathbb{Z}^2} z(\xi + x) v_u(\xi - u) \quad \forall u \in U. \quad (22)$$

It means that the pixel at position u in the foveated patch $\mathbf{z}_{x_1}^{\text{FOV}} = \mathcal{F}[z, x]$ is obtained by applying a specific kernel v_u to z over a neighborhood of $x + u$. To satisfy (17), these kernels must be non-negative, $v_u \geq 0 \quad \forall u \in U$. Thus, $\{v_u\}_{u \in U}$ are blur kernels. \mathcal{F} clearly satisfies translation-invariance with respect to x (16); however, for a fixed x , $\mathcal{F}[\cdot, x]$ corresponds to a blur that is space-variant with respect to u , since $\mathcal{F}[z, x](u)$ is computed from an individual blur kernel v_u , which can be different from kernels $v_{u'}$ with $u' \neq u$.

As a consequence of (22), the foveation operator \mathcal{F} is completely determined by the collection of blur kernels $\{v_u\}_{u \in U}$; hence, the remainder of our analysis shall be based on (22) and expressed in terms of the blur kernels $\{v_u\}_{u \in U}$.

5.2.3 Central acuity (Central Dirac impulse) Central acuity holds if and only if v_0 is a scaled discrete Dirac impulse of mass $\alpha > 0$:

$$v_0(0) = \alpha, \quad v_0(x) = 0 \quad \forall x \neq 0. \quad (23)$$

5.2.4 Flat-field preservation (Constant ℓ^1 norm) From (19) and (22) we obtain

$$\alpha = \sum_{\xi \in \mathbb{Z}^2} v_u(\xi - u) = \sum_{\xi \in \mathbb{Z}^2} v_u(\xi) = \|v_u\|_1 \quad \forall u \in U, \quad (24)$$

i.e. all the blur kernels have same ℓ^1 norm equal to α .

5.2.5 Compatibility We consider first the general compatibility requirement (20), and then turn to the pixelwise compatibility (21). The latter and stricter form of compatibility is the one used for the construction of all foveation operators presented in this work.

Total ℓ^2 -norm condition The following analysis proceeds along the lines of Section 3.3.1. Analogous to (9),

$$E\left\{\left(\mathbf{z}_{x_1}^{\text{FOV}}(u) - \mathbf{z}_{x_2}^{\text{FOV}}(u)\right)^2\right\} = \\ = E\left\{\mathbf{z}_{x_1}^{\text{FOV}}(u) - \mathbf{z}_{x_2}^{\text{FOV}}(u)\right\}^2 + \text{var}\left\{\mathbf{z}_{x_1}^{\text{FOV}}(u) - \mathbf{z}_{x_2}^{\text{FOV}}(u)\right\}. \quad (25)$$

Further, similar to (10), let

$$z(\xi + x_1) - z(\xi + x_2) = E\{z(\xi + x_1) - z(\xi + x_2)\} + \\ + \text{std}\{z(\xi + x_1) - z(\xi + x_2)\} n(\xi) = \\ = (y(\xi + x_1) - y(\xi + x_2)) + \sqrt{2\sigma^2} n(\xi), \quad (26)$$

where $n(\xi) \sim \mathcal{N}(0, 1)$ is standard Gaussian noise. Hence,

$$\text{var}\left\{\mathbf{z}_{x_1}^{\text{FOV}}(u) - \mathbf{z}_{x_2}^{\text{FOV}}(u)\right\} = \\ = \text{var}\left\{\sum_{\xi \in \mathbb{Z}^2} (z(\xi + x_1) - z(\xi + x_2)) v_u(\xi - u)\right\} = \\ = \text{var}\left\{\sum_{\xi \in \mathbb{Z}^2} \sqrt{2\sigma^2} n(\xi) v_u(\xi - u)\right\}. \quad (27)$$

Treating n as uncorrelated noise, which is a safe assumption provided that x_1 is sufficiently far from x_2 , we can interchange summation and variance in (27). Finally, if $E\{\mathbf{z}_{x_1}^{\text{FOV}}\} = E\{\mathbf{z}_{x_2}^{\text{FOV}}\}$, (25) reduces to

$$E\left\{\left(\mathbf{z}_{x_1}^{\text{FOV}}(u) - \mathbf{z}_{x_2}^{\text{FOV}}(u)\right)^2\right\} = \\ = 2\sigma^2 \sum_{\xi \in \mathbb{Z}^2} \text{var}\{n(\xi)\} v_u^2(\xi - u) = 2\sigma^2 \|v_u\|_2^2. \quad (28)$$

Therefore, the compatibility requirement (20) is satisfied whenever the sum of the squared ℓ^2 norms of all blur kernels of \mathcal{F} equals the ℓ^1 norm of the windowing kernel:

$$\sum_{u \in U} \|v_u\|_2^2 = \|\mathbf{k}\|_1 = \sum_{u \in U} \mathbf{k}(u). \quad (29)$$

Kernel-wise ℓ^2 -norm condition Equality (28) also shows that the stricter pixelwise compatibility (21) holds provided that the squared ℓ^2 norm of each blur kernel coincides with the corresponding value of the windowing kernel:

$$\|v_u\|_2^2 = \mathbf{k}(u) \quad \forall u \in U. \quad (30)$$

When $u = 0$, (30) and (23) imply that $\|v_0\|_2^2 = \alpha^2 = \mathbf{k}(0)$, which determines the value of the constant α appearing in the central acuity and flat-field preservation requirements:

$$\alpha = \sqrt{\mathbf{k}(0)}.$$

5.3 Summary

We can summarize Sections 5.2.1–5.2.5 as follows.

A foveation operator \mathcal{F} satisfies the five requirements 5.1.1–5.1.5 with the strict pixelwise compatibility *if and only if* it can be written as

$$\mathcal{F}[z, x](u) = \sum_{\xi \in \mathbb{Z}^2} z(\xi + x) v_u(\xi - u) \quad \forall u \in U, \quad (31)$$

with non-negative blur kernels $v_u \geq 0$, $u \in U$, such that

$$\|v_u\|_1 = \sqrt{\mathbf{k}(0)} \quad \forall u \in U, \quad (32)$$

$$\|v_u\|_2 = \sqrt{\mathbf{k}(u)} \quad \forall u \in U, \quad (33)$$

$$v_0 \text{ is a discrete Dirac impulse of mass } \sqrt{\mathbf{k}(0)}. \quad (34)$$

The foveation operators constructed in the next sections and used throughout our experiments are all based on the above design (31)–(34).

5.4 Singular value decomposition of the foveation operator

Due to the translation invariance with respect to X (16), many properties of $\mathcal{F} : \mathbb{I}_X \times X \rightarrow \mathbb{I}_U$ can be studied, without loss of generality, on the corresponding marginal operator $\mathcal{F}[\cdot, 0] : \mathbb{I}_X \rightarrow \mathbb{I}_U$. Indeed, $\mathcal{F} = \mathcal{P}_X \circ \mathcal{F}[\cdot, 0]$, where $\mathcal{P}_X : \mathbb{I}_X \times X \rightarrow \mathbb{I}_X$ is the translation operator (5). Specifically, we consider the *singular value decomposition* (SVD) of the foveation operator in the form

$$\mathcal{F} = \mathcal{P}_X \circ \mathcal{L}_{\mathcal{F}} \circ \mathcal{S}_{\mathcal{F}} \circ \mathcal{R}_{\mathcal{F}}, \quad (35)$$

where $\mathcal{L}_{\mathcal{F}} : \mathbb{I}_X \rightarrow \mathbb{I}_U$ is the partial isometry given by the inner product with an orthonormal system of *left singular functions* $l_u : X \rightarrow \mathbb{R}$, $u \in U$,

$$\mathcal{L}_{\mathcal{F}}[z] = \langle z, l_u \rangle_{u \in U},$$

$$\mathcal{L}_{\mathcal{F}}[z](u) = \sum_{\xi \in X} z(\xi) l_u(\xi), \quad u \in U,$$

$\mathcal{S}_{\mathcal{F}} : \mathbb{I}_U \rightarrow \mathbb{I}_U$ is a pointwise multiplication against non-negative singular values $s : U \rightarrow \mathbb{R}$,

$$\mathcal{S}_{\mathcal{F}}[\mathbf{z}_x] = \mathbf{z}_x s,$$

$$\mathcal{S}_{\mathcal{F}}[\mathbf{z}_x](u) = \mathbf{z}_x(u) s(u), \quad u \in U,$$

and $\mathcal{R}_{\mathcal{F}} : \mathbb{I}_U \rightarrow \mathbb{I}_U$ is a change of coordinates (global isometry) to the orthonormal basis of *right singular functions* $r_u : U \rightarrow \mathbb{R}$, $u \in U$,

$$\mathcal{R}_{\mathcal{F}}[\mathbf{z}_x] = \langle \mathbf{z}_x, r_u \rangle_{u \in U},$$

$$\mathcal{R}_{\mathcal{F}}[\mathbf{z}_x](u) = \sum_{\xi \in U} \mathbf{z}_x(\xi) r_u(\xi), \quad u \in U.$$

In practice, $\mathcal{L}_{\mathcal{F}}$, $\mathcal{S}_{\mathcal{F}}$, $\mathcal{R}_{\mathcal{F}}$ can of course be all obtained from the thin SVD of the matrix form of $\mathcal{F}[\cdot, 0]$.

In the following sections, we will adopt the SVD (35) as an effective means for analyzing and illustrating the foveation operators. Specifically, the left singular functions l_u that correspond to the largest singular values $s(u)$ describe the image structures that are best captured by the foveation operator.

6 Construction of Gaussian foveation operators

As explained in Section 2.1, foveation is the result of several cascaded space-variant processes. Taking a central-limit principle into consideration, we argue that Gaussian distributions may well approximate the final combined blurring effect of these processes, thus providing a legitimate model for the blur kernels $\{v_u\}_{u \in U}$.

Here we present the direct construction of foveation operators based on families of Gaussian blur kernels. The idea is to leverage suitable dilation and scaling so that all the blur kernels share the same ℓ^1 norm given by (32) and have ℓ^2 norms given by (33).

For the sake of generality and mathematical simplicity, this construction uses a continuous-domain variable $\xi \in \mathbb{R}^2$; any natural discretization to the integers \mathbb{Z}^2 is to be employed in the implementation. Due to such discretization, marginal numerical imprecision may eventually affect the equalities (32)–(33). While in practice the errors are negligible, in Section 7 we present a general scheme that can be used to design discrete Gaussian or non-Gaussian $\{v_u\}_{u \in U}$ meeting (32)–(33) accurate up to machine precision.

6.1 Isotropic foveation operators

Let us begin from the case where all blur kernels v_u , $u \in U$, are circularly symmetric. Such foveation is said to be *isotropic* because the blur kernels attenuate image features regardless of the features orientation and, provided also a circularly-symmetric windowing kernel \mathbf{k} (as it is customary), the attenuation strength depends only on the distance $|u|$ from the patch center.

We define g_ς as the circularly-symmetric bivariate Gaussian probability density function (PDF) with mean zero and diagonal covariance matrix $\Sigma_\varsigma = \begin{bmatrix} \varsigma^2 & 0 \\ 0 & \varsigma^2 \end{bmatrix}$:

$$g_\varsigma(\xi) = \frac{1}{2\pi\varsigma^2} e^{-\frac{|\xi|^2}{2\varsigma^2}}, \quad \xi \in \mathbb{R}^2. \quad (36)$$

The standard-deviation parameter ς determines the spread of this Gaussian PDF. Basic calculus shows that

$$\|g_\varsigma\|_1 = 1, \quad \|g_\varsigma\|_2^2 = \frac{1}{4\pi\varsigma^2}. \quad (37)$$

Therefore, to obtain a blur kernel v_u that satisfies (32) and (33) it is enough to first scale g_ς by multiplication by $\alpha = \sqrt{\mathbf{k}(0)}$ and then dilate it and scale it again by choosing ς that solves $\|\sqrt{\mathbf{k}(0)}g_\varsigma\|_2^2 = \mathbf{k}(u)$, i.e.

$$\varsigma_u^* = \frac{1}{2\sqrt{\pi}} \sqrt{\frac{\mathbf{k}(0)}{\mathbf{k}(u)}}. \quad (38)$$

Hence, we define

$$v_u(\xi) = \sqrt{\mathbf{k}(0)}g_{\varsigma_u^*}(\xi) = \frac{2\mathbf{k}(u)}{\sqrt{\mathbf{k}(0)}} e^{-2\pi|\xi|^2 \frac{\mathbf{k}(u)}{\mathbf{k}(0)}} \quad \xi \in \mathbb{R}^2. \quad (39)$$

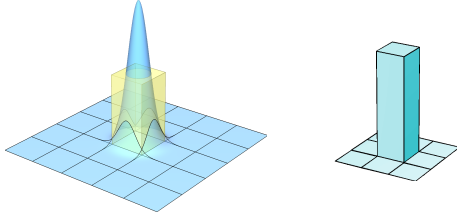


Fig. 4 Left: The scaled discrete Dirac impulse of mass $\sqrt{\mathbf{k}(0)}$ corresponds to a uniform kernel over the unit square which the Gaussian kernel v_0 (40) can at best only approximate. Interestingly, the standard deviations of these two kernels nearly coincide. Due to separability we can consider the 1-D components of the PDFs: for $u = 0$ (38) yields $\varsigma = 1/\sqrt{4\pi} \simeq 0.282$, while the standard deviation of the uniform PDF on the unit interval is $1/\sqrt{12} \simeq 0.289$. Right: after discretization (sampling and ℓ^1 renormalization), v_0 is in all practical respects equivalent to a Dirac impulse.

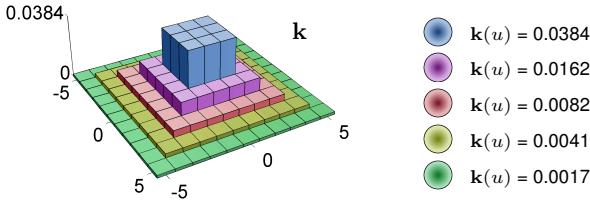


Fig. 5 A windowing kernel \mathbf{k} of size 11×11 used for computing the similarity weights in the NL-means implementation by Manjon-Herrera and Buades (2008) and the five unique values it takes over the neighborhood U .

Observe that (38) installs a direct link between the width of the blur kernel v_u and the value of the windowing kernel $\mathbf{k}(u)$. This is a consequence of the pixelwise-compatibility requirement and operates in such a way that, when $\mathbf{k}(u)$ is small (i.e. at the periphery of the patch), the blur caused by the kernel v_u is large, thus mimicking the foveation effects illustrated in Figure 1.

Finally, we need to address the central-acuity requirement (34). As a matter of fact, as illustrated in Figure 4(left), neither the central blur kernel

$$v_0(\xi) = 2\sqrt{\mathbf{k}(0)}e^{-2\pi|\xi|^2}, \quad \xi \in X, \quad (40)$$

nor any other Gaussian function can satisfy (34) exactly. In principle, it is a mere technicality to “manually” re-define v_0 according to (23) with $\alpha = \sqrt{\mathbf{k}(0)}$ (thus meeting (34) exactly), as an exception from the general form (39) for $u = 0$ only. However, the discrepancy between the discrete Dirac impulse and the discrete representation of the Gaussian v_0 (40) is in fact numerically negligible, as Figure 4(right) shows. Therefore, the central acuity requirement is satisfied in all practical respects by the foveation operator composed of blur kernels defined by (39) for all $u \in U$, including $u = 0$.

6.1.1 Illustration Let us give a concrete illustration of the construction of a foveation operator and particularly of its blur kernels for a given windowing kernel \mathbf{k} . For this example, we take the windowing kernel \mathbf{k} used in the NL-means implementation by Manjon-Herrera and Buades (2008) for a neighborhood U of size 11×11 pixels, shown in Figure

5. Due to its symmetries, \mathbf{k} takes only a very limited number of distinct values $\mathbf{k}(u)$, $u \in U$, as reported in the figure. For each distinct value of $\mathbf{k}(u)$, (38) yields a distinct value of the standard-deviation parameter ς_u^* and correspondingly (39) defines a distinct blur kernel v_u . These blur kernels are shown in Figure 6 along with their frequency response. Each of these blur kernels is based on a discrete Gaussian PDF $g_{\varsigma_u^*}$ of size $(2 \lceil 3\varsigma_u^* \rceil + 1) \times (2 \lceil 3\varsigma_u^* \rceil + 1)$, $\lceil \cdot \rceil$ being the ceiling function, where radius $\lceil 3\varsigma_u^* \rceil$ is based on a *three-sigma* rule. Observe that v_0 is practically a scaled Dirac impulse, confirming central acuity. The bandwidth of v_u decreases with $\mathbf{k}(u)$, as clearly visualized. Also note that the frequency responses of the five blur kernels all attain their maximum at the origin and that the value of this maximum is $\sqrt{\mathbf{k}(0)} = 0.196$, as implied by (32). The squared ℓ^2 -norm values $\|v_u\|_2^2$ for these five kernels are 0.0379, 0.0231, 0.0090, 0.0041, 0.0017, respectively, which are nearly equal to the corresponding values of $\mathbf{k}(u)$, as by (33). The minor differences are due to the discretization of the Gaussian PDF g_ς used in our implementation, while the ℓ^2 -norm condition in (37) assumes continuous domain variables. In Section 7 we present a method that delivers discrete blur kernels whose ℓ^1 and ℓ^2 norms satisfy (32)–(33) with perfect accuracy.

Figure 7 compares the foveated patches $\mathbf{z}_x^{\text{FOV}}$ (15) against the windowed patches $\mathbf{z}_x \sqrt{\mathbf{k}}$ (8). Observe how $\mathbf{z}_x^{\text{FOV}}$ exhibit effective smoothing at their periphery while not altering the main structures in the patch. Because of (8) and (30), the corresponding pixels in a foveated and windowed patch have exactly the same noise variance:

$$\text{var}\{\mathbf{z}_x^{\text{FOV}}(u)\} = \text{var}\{\mathbf{z}_x^{\text{WIN}}(u)\} = \sigma^2 \mathbf{k}(u) \quad \forall u \in U. \quad (41)$$

Thus, foveated patches typically enjoy an intrinsically higher signal-to-noise ratio (SNR) (e.g., in the sense of Mosseri et al., 2013) than their windowed counterparts.

6.1.2 Limiting cases It is interesting to consider the limiting cases where the window \mathbf{k} becomes a Dirac impulse or a uniform kernel over the patch domain. Curiously, in either case, the foveated distance approaches the windowed distance.

Dirac window When \mathbf{k} is scaled Dirac impulse (i.e. $\mathbf{k}(0) > 0$ and $\mathbf{k}(u) = 0 \forall u \neq 0$), by (38) the standard deviation ς of the Gaussian PDFs g_ς should be infinite for any $u \neq 0$. Hence, they can be treated as uniform distributions over any image domain X . For a finite X with area $|X| < \infty$, assuming symmetric padding of z outside of X , we obtain

$$\mathbf{z}_x^{\text{FOV}}(u) = \sqrt{\mathbf{k}(0)} \frac{1}{|X|} \sum_{\xi \in X} z(\xi), \quad \forall u \neq 0, \forall x \in X,$$

with $\frac{1}{|X|} \sum_{\xi \in X} z(\xi)$ being the mean of z over X . Therefore, $\mathbf{z}_{x_1}^{\text{FOV}}(u) - \mathbf{z}_{x_2}^{\text{FOV}}(u) = 0 \forall x_1, x_2$ and $\forall u \neq 0$. Hence, due to (31) and (34), $d^{\text{FOV}}(x_1, x_2)$ depends solely on pixels at x_1 and x_2 , exactly like $d^{\text{WIN}}(x_1, x_2)$, and thus the two distances coincide:

$$d^{\text{FOV}}(x_1, x_2) = \mathbf{k}(0) (z(x_1) - z(x_2))^2 = d^{\text{WIN}}(x_1, x_2).$$

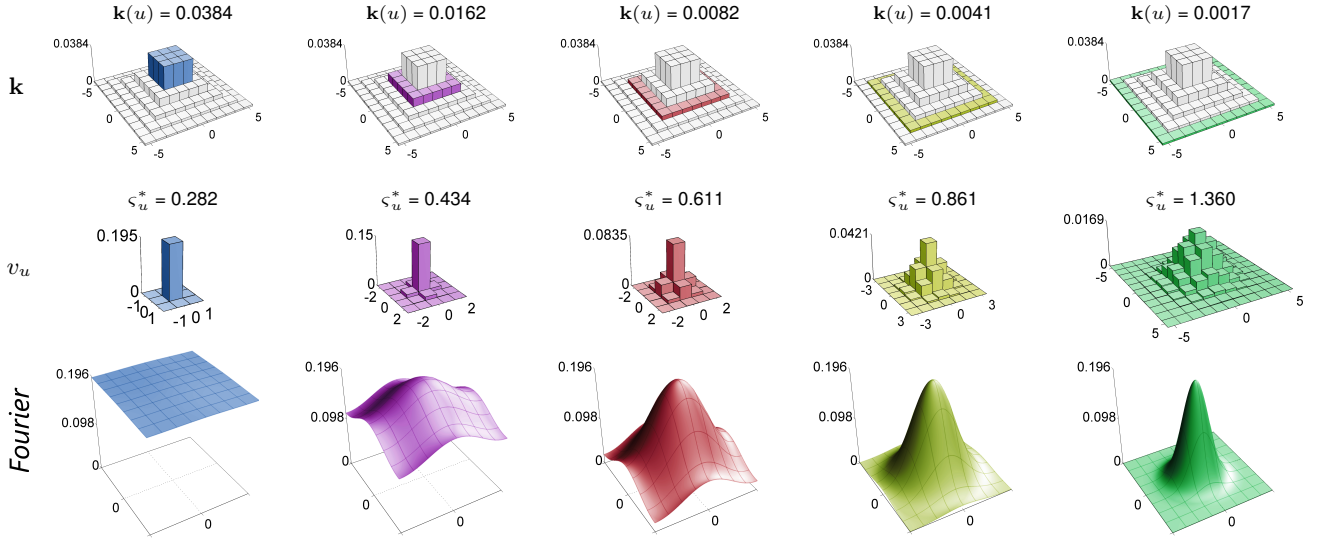


Fig. 6 The five blur kernels v_u , $u \in U$, (middle row) corresponding to the five unique values of the windowing kernel \mathbf{k} (top row) and their respective frequency response (bottom row).

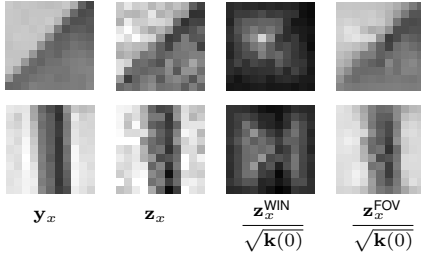


Fig. 7 Two illustrative examples of how patches are transformed when computing the windowed distance d^{WIN} (6) and the foveated distance d^{FOV} (14): \mathbf{z}_x denotes the patch in the noisy image, while $\mathbf{z}_x \sqrt{\mathbf{k}} = \mathbf{z}_x^{\text{WIN}}$ and $\mathbf{z}_x^{\text{FOV}}$ denote, respectively, the patches used for computing the quadratic windowed distance and foveated distance, for the same image and fixation point. The patch \mathbf{y}_x , i.e. the noise-free counterpart of \mathbf{z}_x , is also shown for reference. Dividing $\mathbf{z}_x^{\text{WIN}}$ and $\mathbf{z}_x^{\text{FOV}}$ by $\sqrt{\mathbf{k}(0)}$ ensures that these patches are displayed on the same intensity range of \mathbf{z}_x and \mathbf{y}_x . In both examples the standard deviation of the noise in \mathbf{z}_x is $\sigma = 20$.

Uniform window When \mathbf{k} is uniform over the patch domain U (i.e. $\mathbf{k}(u) = \mathbf{k}(0) \forall u \in U$), again by (38) we have that all the blur kernels v_u (39) coincide with the central one v_0 (40):

$$v_u = v_0, \quad \forall u \in U.$$

Hence, based on the arguments at the end of Section 6.1, all blur kernels v_u are practically Dirac impulses scaled by $\sqrt{\mathbf{k}(0)}$. Then, any foveated patch $\mathbf{z}_x^{\text{FOV}}$ coincides with the corresponding scaled patch $\mathbf{z}_x \sqrt{\mathbf{k}(0)}$; further, since \mathbf{k} is uniform, they also coincide with the windowed patch $\mathbf{z}_x \sqrt{\mathbf{k}}$ (8):

$$\mathbf{z}_x^{\text{FOV}} = \sqrt{\mathbf{k}(0)} \mathbf{z}_x = \sqrt{\mathbf{k}} \mathbf{z}_x = \mathbf{z}_x^{\text{WIN}}.$$

Therefore, also in this limiting case, the foveated distance coincides with the corresponding windowed distance.

6.2 Anisotropic foveation operators

Anisotropic foveation operators generalize the isotropic ones described in Section 6.1 by utilizing, in place of the circularly-symmetric Gaussian PDF g_ζ (36), an elliptical Gaussian PDF $g_\zeta^{\rho, \vartheta}$ whose covariance matrix depends not only on $\zeta > 0$, but also on a parameter $\rho > 0$ that determines the elongation of the PDF, and on an angular parameter $\vartheta \in \mathbb{R}$ that controls the orientation of the axes of the elliptical PDF. Specifically, $g_\zeta^{\rho, \vartheta}$ is the elliptical Gaussian PDF with covariance matrix

$$\Sigma_\zeta^{\rho, \vartheta} = \zeta_\vartheta^2 \mathbf{R}_\vartheta \mathbf{D}_\rho \mathbf{R}_\vartheta^T,$$

where the $\mathbf{D}_\rho = \begin{bmatrix} \rho & 0 \\ 0 & 1/\rho \end{bmatrix}$ determines the PDF elongation and $\mathbf{R}_\vartheta = \begin{bmatrix} \cos(\vartheta) & -\sin(\vartheta) \\ \sin(\vartheta) & \cos(\vartheta) \end{bmatrix}$ is a rotation of angle ϑ . Therefore

$$g_\zeta^{\rho, \vartheta}(\xi) = \frac{1}{2\pi\zeta^2} e^{-\frac{1}{2}\xi^T (\Sigma_\zeta^{\rho, \vartheta})^{-1} \xi}, \quad \xi \in \mathbb{R}^2,$$

where ξ in the exponential is a column-vector representation of ξ . Clearly, $\rho = 1$ corresponds to the circularly-symmetric case considered in Section 6.1 and it follows that $g_\zeta^{1, \vartheta} = g_\zeta$ for any $\vartheta \in \mathbb{R}$ and any $\zeta > 0$. Thus, \mathbf{D}_ρ and \mathbf{R}_ϑ determine the deformation of g_ζ into $g_\zeta^{\rho, \vartheta}$. Because $\det \mathbf{R}_\vartheta = \det \mathbf{D}_\rho = 1$, any integral of $g_\zeta^{\rho, \vartheta}$ (including Minkowski norms) is invariant with respect to any such rotation and elliptical elongation, which means that $g_\zeta^{\rho, \vartheta}$ conforms to the same norms as the circularly symmetric g_ζ (37):

$$\|g_\zeta^{\rho, \vartheta}\|_1 = 1, \quad \|g_\zeta^{\rho, \vartheta}\|_2^2 = \frac{1}{4\pi\zeta^2} \quad \forall \rho > 0, \forall \vartheta \in \mathbb{R}. \quad (42)$$

Hence, like in the isotropic case, to satisfy (32) and (33) it suffices to multiply $g_\zeta^{\rho, \vartheta}$ by $\sqrt{\mathbf{k}(0)}$ and set $\zeta = \zeta_u^*$ exactly as in (38), which gives

$$\left\| \sqrt{\mathbf{k}(0)} g_{\zeta_u^*}^{\rho, \vartheta} \right\|_2^2 = \mathbf{k}(u). \quad (43)$$

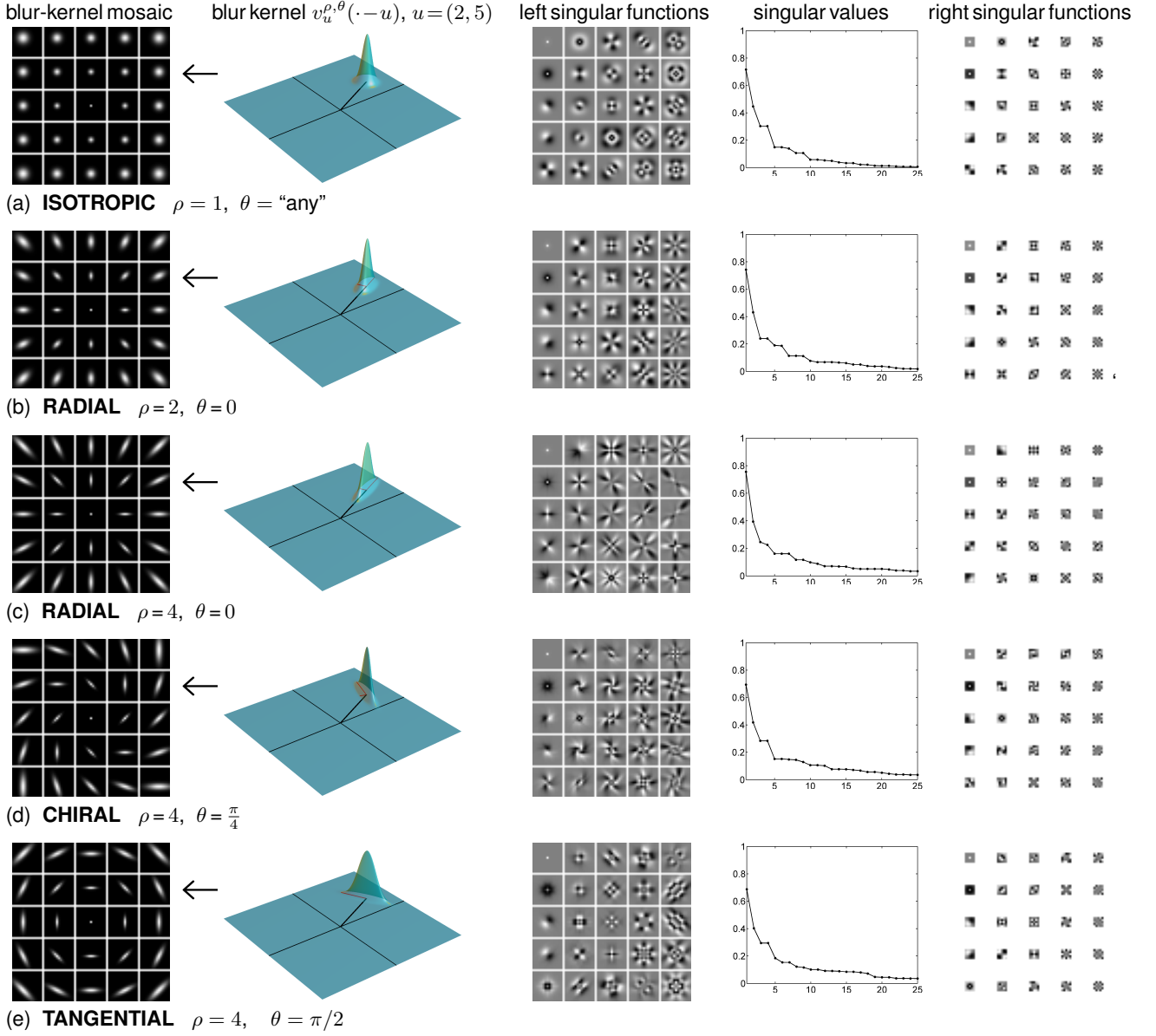


Fig. 8 Illustration of an isotropic foveation operator \mathcal{F} (a), and four anisotropic foveation operators $\mathcal{F}_{\rho, \theta}$ (b)–(e) for different values of the parameters ρ and θ . Each operator is displayed (leftmost column) by the mosaic of the 25 blur kernels $v_u^{\rho, \theta}(\cdot - u)$ that produce the pixels of a 5×5 foveated patch. The subimages in each mosaic are placed at the corresponding position $u \in U$ of the pixels in the foveated patch. Note the relative displacement of the kernels due to the $-u$ shift in their argument. For the sake of visualization, blur kernels in the mosaic are displayed after intensity normalization; in reality, the peak of the kernels gets lower as the spread increases, because of the ℓ^1 condition (42). Next to each mosaic, we depict one of its blur kernels $v_u^{\rho, \theta}(\cdot - u)$, as indicated by the arrow. Here one can visualize the angle $\vartheta = \angle u + \theta$. In the right half of the figure, we show the decomposition (35) of the foveation operator into left singular functions, singular values, and right singular functions. The singular functions are sorted columnwise according to the decreasing magnitude of the corresponding singular value, and are displayed after intensity normalization. The 5×5 support of the right singular functions coincides with that of the patch and it is hence smaller than that of the blur kernels and of the left singular functions.

Both (42) and (43) hold for any ρ and ϑ . Thus, one can arbitrarily assign different combinations of these parameters to different $u \in U$, without invalidating the ℓ^1 (32) and ℓ^2 (33) conditions. In fact, sophisticated models of HVS acuity suggest that ρ ought depend on $|u|$ and ϑ on $\angle u$. However, we focus on a specific simplified design, where ρ is constant and $\vartheta = \angle u + \theta$, being $\theta \in \mathbb{R}$ an angular offset. These choices

lead to a class of *anisotropic foveation operators*

$$\mathcal{F}_{\rho, \theta}[z, x](u) = \sum_{\xi \in \mathbb{Z}^2} z(\xi + x) v_u^{\rho, \theta}(\xi - u), \quad \forall u \in U \quad (44)$$

defined through the blur kernels

$$v_u^{\rho, \theta} = \begin{cases} \sqrt{\mathbf{k}(0)} g_{\frac{1}{2\sqrt{\pi}}}^{\rho, \angle u + \theta} \sqrt{\frac{\mathbf{k}(0)}{\mathbf{k}(u)}}, & u \neq 0, \\ \sqrt{\mathbf{k}(0)} g_{\frac{1}{2\sqrt{\pi}}}, & u = 0. \end{cases} \quad (45)$$

The bottom equality in (45) addresses the special case $u = 0$ for which $\angle u$ cannot be defined, consequently enforcing the circularly-symmetric v_0 (40). By construction, the anisotropic foveation operators $\mathcal{F}_{\rho,\theta}$ (44) satisfy the constraints set in Section 5 for any combination of $\rho > 0$ and $\theta \in \mathbb{R}$.

Specific choices of θ and ρ yield different operators that can be classified as follows. When $\theta = 0$ and $\rho > 1$, the major axes of the blur kernels are directed along the meridians toward the patch center, yielding *radial* foveation operators. Conversely, if $\theta = \pi/2$ and $\rho > 1$, it is the minor axes of the blur kernels that are directed towards the patch center and we obtain *tangential* foveation operators. Intermediate cases $0 < \theta < \pi/2$ yield (right-hand) *chiral* foveation operators. Due to elliptical symmetry of the blur kernels, the operators $\mathcal{F}_{\rho,\theta}$, $\mathcal{F}_{1/\rho,\theta+\pi/2+j\pi}$ and $\mathcal{F}_{\rho,\theta+j\pi}$ coincide for any $j \in \mathbb{Z}$. The larger $\max\{\rho, 1/\rho\}$ is, the more elongated the kernels are along their major axis, while when $\rho = 1$, $\mathcal{F}_{\rho,\theta}$ coincides with the *isotropic* foveation operator of Section 6.1.

Figure 8 shows the blur kernels that constitute a few anisotropic foveation operators corresponding to a same windowing kernel \mathbf{k} of size 5×5 for different combinations of ρ and θ values. In the right half of the figure we illustrate the SVD of $\mathcal{F}_{\rho,\theta}$, showing left singular (analysis) functions, singular values, and right singular (synthesis) functions. The functions are sorted columnwise, according to the decreasing magnitude of the corresponding singular value. Thus, the first few left singular functions visualize the specific image patterns that are best captured in the foveated patch.

In Figure 8 we also see that left and right singular functions have supports of different size. Right singular functions synthesize the foveated patch, hence they have the same size as the windowing kernel \mathbf{k} . Left singular functions have a larger support, since they embed the analysis offered by the blur kernels $v_u^{\rho,\theta}(\cdot - u)$, whose centers range within U but take non-zero values beyond this set. The next section investigates a special construction of the foveation operator that ensures that blur kernels, as well as both left and right singular functions, are defined on the same support of \mathbf{k} .

7 Foveation operators with self-mapping in patch space

7.1 Motivation

For any given $u \neq 0$, the Gaussian kernel $v_u^{\rho,\theta}$ is strictly positive everywhere, i.e. $v_u^{\rho,\theta}(\xi) > 0$ for any $\xi \in \mathbb{R}^2$ (or, upon discretization of the kernel domain, for any $\xi \in \mathbb{Z}^2$). Formally, this implies that for computing $\mathbf{z}_x^{\text{FOV}}(u)$ one has to sample the observations $z(\cdot + x)$ also outside of the set U that defines the patch \mathbf{z}_x (2). However, the exponential decay of $v_u^{\rho,\theta}$ (45) means that, in practice, only a small neighborhood of u is necessary for computing $\mathbf{z}_x^{\text{FOV}}(u)$: for example, if like in Section 6.1.1 we follow a *three-sigma* rule, it suffices to sample $z(\cdot + x)$ over a disc Ω_u of radius $3\sqrt{\max\{\rho, 1/\rho\}}\zeta_u^*$ around u , which encompasses the three-sigma ellipse for $v_u^{\rho,\theta}(\cdot - u)$. Hence, in (22) and (44) we can replace the summation domain \mathbb{Z}^2 with Ω_u without affecting the result. This

notwithstanding, it is clear that when u is at the boundary of U , where $\mathbf{k}(u)$ is small and the foveation blur is strong, the disc Ω_u may exceed U , i.e. $\Omega_u \setminus U \neq \emptyset$. In other words, the foveated patch $\mathbf{z}_x^{\text{FOV}}$ obtained from a Gaussian foveation operator requires more samples of z than does the windowed patch $\sqrt{\mathbf{k}}\mathbf{z}_x$, for which it suffices to sample $z(\cdot + x)$ over U .

7.2 Self-map foveation operators

Let us introduce *self-map foveation operators* that compute each foveated patch $\mathbf{z}_x^{\text{FOV}}$ using data exclusively from the patch \mathbf{z}_x , without sampling $z(\cdot + x)$ outside of U . Self-map foveation operators can be expressed as the left composition of the patching operator $\mathcal{P}_U : \mathbb{I}_X \times X \rightarrow \mathbb{I}_U$ (4) with a space-variant anisotropic blur operator $\mathcal{B}_{\rho,\theta} : \mathbb{I}_U \rightarrow \mathbb{I}_U$ that maps any patch \mathbf{z}_x (2) to a foveated patch $\mathbf{z}_x^{\text{FOV}}$:

$$\mathcal{F}_{\rho,\theta}^{\text{self}} = \mathcal{P}_U \circ \mathcal{B}_{\rho,\theta}, \quad \mathcal{B}_{\rho,\theta} : \mathbf{z}_x \mapsto \mathbf{z}_x^{\text{FOV}}. \quad (46)$$

More precisely, $\mathcal{F}_{\rho,\theta}^{\text{self}} : \mathbb{I}_X \times X \rightarrow \mathbb{I}_U$ maps any image/fixation-point pair $(z, x) \in \mathbb{I}_X \times X$ to a patch $\mathbf{z}_x^{\text{FOV}} \in \mathbb{I}_U$

$$\mathbf{z}_x^{\text{FOV}} = \mathcal{F}_{\rho,\theta}^{\text{self}}[z, x] = \mathcal{B}_{\rho,\theta}[\mathcal{P}_U[z, x]] = \mathcal{B}_{\rho,\theta}[\mathbf{z}_x],$$

and such that, for any $u \in U$,

$$\begin{aligned} \mathbf{z}_x^{\text{FOV}}(u) &= \mathcal{F}_{\rho,\theta}^{\text{self}}[z, x](u) = \mathcal{B}_{\rho,\theta}[\mathbf{z}_x](u) = \\ &= \sum_{\xi \in U} z(\xi + x) \bar{v}_u^{\rho,\theta}(\xi - u), \end{aligned} \quad (47)$$

where the blur kernels $\bar{v}_u^{\rho,\theta}(\cdot - u)$ are all supported on U , i.e.

$$\xi \notin U \implies \bar{v}_u^{\rho,\theta}(\xi - u) = 0.$$

Observe that the summation in (47) is computed over $\xi \in U$ only, as opposed to (44) where the summation is over $\xi \in \mathbb{Z}^2$.

7.3 Construction of compatible self-map foveation operators

The self-map foveation operator $\mathcal{F}_{\rho,\theta}^{\text{self}}$ is constructed by defining a family of blur kernels $\{\bar{v}_u^{\rho,\theta}\}_{u \in U}$ that obey the requirements set in Section 5.3 and whose supports are entirely contained in U .

Let $\bar{g}_{\zeta,u}^{\rho,\theta}$, $\zeta > 0$, be the truncated PDF supported on U that is obtained by first restricting the Gaussian PDF $g_{\zeta}^{\rho,\theta}(\cdot - u)$ on U and then by ℓ^1 re-normalization:

$$\bar{g}_{\zeta,u}^{\rho,\theta}(\xi - u) = \begin{cases} g_{\zeta}^{\rho,\theta}(\xi - u) / \sum_{\xi \in U} g_{\zeta}^{\rho,\theta}(\xi - u) & \xi \in U, \\ 0 & \xi \notin U. \end{cases} \quad (48)$$

For any fixed $u \in U$, the limiting cases of (48) for $\zeta \rightarrow 0$ and $\zeta \rightarrow +\infty$ are, respectively, the Dirac impulse and the uniform kernel on U . Therefore, we can extend the definition of $\bar{g}_{\zeta,u}^{\rho,\theta}$ to any $\zeta \in [0, +\infty]$ by adding these two limiting functions:

$$\bar{g}_{0,u}^{\rho,\theta}(\xi - u) = \begin{cases} 1 & \xi = u, \\ 0 & \xi \neq u, \end{cases} \quad \bar{g}_{\infty,u}^{\rho,\theta}(\xi - u) = \begin{cases} \frac{1}{|U|} & \xi \in U, \\ 0 & \xi \notin U, \end{cases}$$

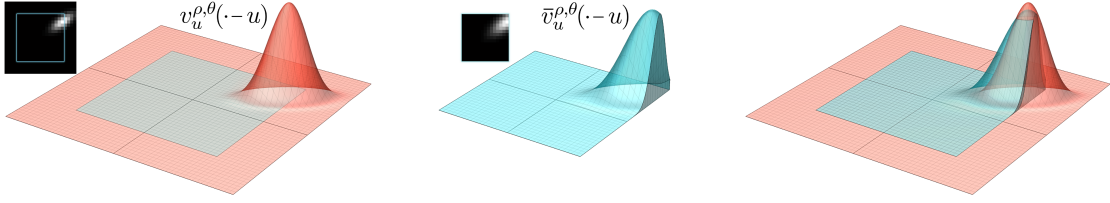


Fig. 9 Left: a radial blur kernel $v_u^{\rho, \vartheta}(\cdot - u)$ from a radial foveation operator ($\rho = 2.5$, $\theta = 0$). The neighborhood U is marked by the inner square. Center: the blur kernel $\bar{v}_u^{\rho, \vartheta}(\cdot - u)$ from the corresponding self-map operator, which is fully supported on U . By construction, their ℓ^1 and ℓ^2 norms coincide: $\|\bar{v}_u^{\rho, \vartheta}\|_1 = \|v_u^{\rho, \vartheta}\|_1$, $\|\bar{v}_u^{\rho, \vartheta}\|_2 = \|v_u^{\rho, \vartheta}\|_2$. To compensate the truncation while preserving the norms, the standard-deviation parameter $\zeta_u^{*, \vartheta}$ of $\bar{v}_u^{\rho, \vartheta}$ has to be larger than the standard deviation ζ_u^* of $v_u^{\rho, \vartheta}$, as exposed by a direct comparison (right).

where $|U| = \sum_{u \in U} 1$ denotes the size of the patch U . From this definition follows that $\|\bar{g}_{\zeta, u}^{\rho, \vartheta}\|_1 = 1$ for any $\rho > 0$, $\vartheta \in \mathbb{R}$, and – most importantly – for any standard-deviation parameter $\zeta \in [0, +\infty]$. However, because of truncation and re-normalization, the ℓ^2 -norm condition in (43) may be not satisfied by (48) and therefore compatibility would not be guaranteed if we were to define $\zeta = \zeta_u^*$ as in (38).

We tackle the pixelwise compatibility through the minimization problems

$$\zeta_u^{\rho, \vartheta} = \operatorname{argmin}_{\zeta \in [0, +\infty]} \left| \mathbf{k}(0) \|\bar{g}_{\zeta, u}^{\rho, \vartheta}\|_2^2 - \mathbf{k}(u) \right|, \quad \forall u \in U \quad (49)$$

where, for each $u \in U$, we adjust ζ to meet (43). Standard numerical methods for nonlinear optimization (Conn et al., 2009) can be used for solving (49)⁴. Hence, we define the restricted Gaussian kernels $\bar{v}_u^{\rho, \vartheta}$ as

$$\bar{v}_u^{\rho, \vartheta} = \sqrt{\mathbf{k}(0)} \bar{g}_{\zeta_u^{\rho, \vartheta}, u}^{\rho, \vartheta} \quad \forall u \in U.$$

Let us observe that $\|\bar{g}_{\zeta, u}^{\rho, \vartheta}\|_2^2$ is subject to the inequalities

$$1 \geq \|\bar{g}_{\zeta, u}^{\rho, \vartheta}\|_2^2 \geq \frac{1}{|U|} \quad \forall \zeta \in [0, +\infty], \quad (50)$$

with the extreme values 1 and $\frac{1}{|U|}$ attained by $\bar{g}_{0, u}^{\rho, \vartheta}$ and $\bar{g}_{\infty, u}^{\rho, \vartheta}$, respectively. Therefore the problem (49) can be minimized to zero, i.e. exactly fulfilling (43), only provided that $\frac{1}{|U|} \leq \frac{\mathbf{k}(u)}{\mathbf{k}(0)} \leq 1$, that is if $\mathbf{k}(u)$ does not decay too fast as $|u|$ grows and if $\mathbf{k}(0)$ is the maximum of \mathbf{k} over U .

The optimization (49) is directly applicable also to generic blur kernels of (non self-map) isotropic and anisotropic foveation operators, resolving the possible discrepancies between the continuous and discrete formulation of the Gaussian PDFs $g_{\zeta}^{\rho, \vartheta}$, so that (32)–(33) hold accurately up to machine precision.

⁴ In particular, the recursive sequence $\{\zeta_n\}_{n=0}^{+\infty}$ defined by

$$\zeta_0 = \frac{1}{2\sqrt{\pi}} \sqrt{\frac{\mathbf{k}(0)}{\mathbf{k}(u)}}, \quad \zeta_{n+1} = \zeta_n \left\| \bar{g}_{\zeta_n, u}^{\rho, \vartheta} \right\|_2 \sqrt{\frac{\mathbf{k}(0)}{\mathbf{k}(u)}}$$

converges monotonically to the solution $\zeta_u^{*, \vartheta}$ of (49), with geometric rate for any $\mathbf{k}(u) < \mathbf{k}(0)$ (contraction mapping).

7.3.1 Singular value decomposition of self-map operators
Due to (46), the SVD (35) of self-map operators can be simplified into

$$\mathcal{F}_{\rho, \theta}^{\text{self}} = \mathcal{P}_U \circ \mathcal{L}_{\mathcal{F}} \circ \mathcal{S}_{\mathcal{F}} \circ \mathcal{R}_{\mathcal{F}},$$

where $\mathcal{L}_{\mathcal{F}} : \mathbb{I}_U \rightarrow \mathbb{I}_U$ is a change of coordinates (global isometry) to the orthonormal basis of *left singular functions* $l_u : U \rightarrow \mathbb{R}$, $u \in U$,

$$\begin{aligned} \mathcal{L}_{\mathcal{F}}[\mathbf{z}_x] &= \langle \mathbf{z}_x, l_u \rangle_{u \in U}, \\ \mathcal{L}_{\mathcal{F}}[\mathbf{z}_x](u) &= \sum_{\xi \in U} \mathbf{z}_x(\xi) l_u(\xi), \quad u \in U, \end{aligned}$$

and $\mathcal{S}_{\mathcal{F}}$ and $\mathcal{R}_{\mathcal{F}}$ are defined as in Section 5.4. Thus, both the left and the right singular functions are defined over U and it is natural to wonder whether they coincide. The answer is negative, since the operator $\mathcal{B}_{\rho, \theta}$ is not self-adjoint. This is a direct consequence of the space-variant blur inherent to foveation, i.e. that the foveation operator embeds blur kernels that can be different one from the other. Indeed, we can always find a pair of points $u_1, u_2 \in U$ such that⁵ $\bar{v}_{u_1}^{\rho, \vartheta}(u_2) \neq \bar{v}_{u_2}^{\rho, \vartheta}(u_1)$.

7.3.2 Illustration Figure 9 shows a blur kernel $v_u^{\rho, \vartheta}$ from a radial foveation operator $\mathcal{F}_{\rho, \theta}$ (left) and the blur kernel $\bar{v}_u^{\rho, \vartheta}$ from the corresponding self-map operator $\mathcal{F}_{\rho, \theta}^{\text{self}}$ (center), $\rho = 2.5$, $\theta = 0$. By construction, the two blur kernels share the same ℓ^1 and ℓ^2 norms:

$$\begin{aligned} \|v_u^{\rho, \vartheta}\|_1 &= \|\bar{v}_u^{\rho, \vartheta}\|_1 = \sqrt{\mathbf{k}(0)}, \\ \|v_u^{\rho, \vartheta}\|_2 &= \|\bar{v}_u^{\rho, \vartheta}\|_2 = \sqrt{\mathbf{k}(u)}. \end{aligned}$$

To compensate truncation while preserving the norms, the standard-deviation parameter of $\bar{v}_u^{\rho, \vartheta}$ is adjusted through (49) and is eventually larger than that of $v_u^{\rho, \vartheta}$, as can be observed in the figure (right).

Figure 10 compares a radial foveation operator $\mathcal{F}_{\rho, \theta}$ (top), $\rho = 4.0$, $\theta = 0$, with the corresponding self-map operator $\mathcal{F}_{\rho, \theta}^{\text{self}}$ (bottom). Both operators are constructed from the same windowing kernel \mathbf{k} visualized in Figure 11. In the center and right of Figure 10 we show the left and right singular functions from the SVD of the operators $\mathcal{F}_{\rho, \theta}$ and $\mathcal{F}_{\rho, \theta}^{\text{self}}$. The

⁵ Note that $\bar{v}_{u_i}^{\rho, \vartheta}(u_j)$ is nothing but the inner product between a Dirac patch at u_i and $\mathcal{B}_{\rho, \theta}$ applied to another Dirac patch at u_j : $\bar{v}_{u_i}^{\rho, \vartheta}(u_j) = \langle \delta_{u_i}, \mathcal{B}_{\rho, \theta}[\delta_{u_j}] \rangle$.

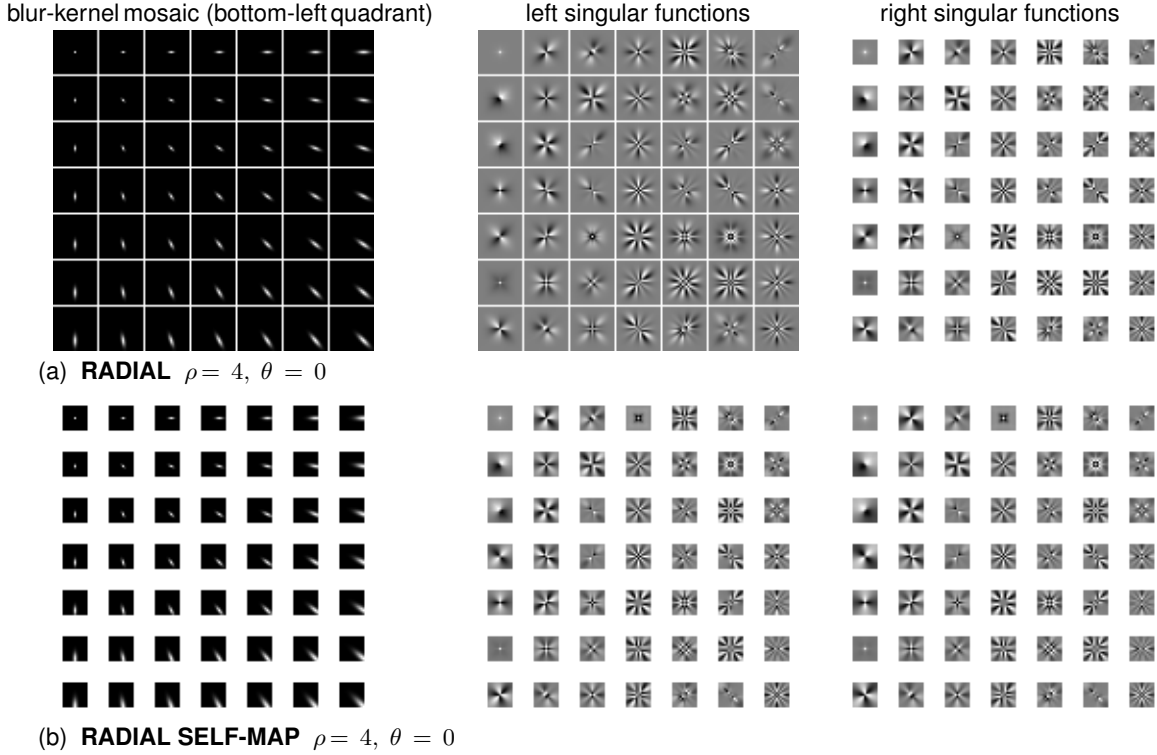


Fig. 10 Radial anisotropic foveation operators $\mathcal{F}_{4,0}$ (a) and the *self-map* $\mathcal{F}_{4,0}^{\text{self}}$ (b) constructed from the 13×13 windowing kernel \mathbf{k} in Figure 11. At the left, we illustrate only the lower-left quadrant of the blur-kernel mosaic: due to symmetry, this effectively represents all blur kernels, since all four quadrants are identical upon rotation of 90 degrees. The central kernel \bar{v}_0 is at the top-left corner of the shown quadrant. Please see also Figure 9, when comparing blur kernels $v_u^{4,0}(\cdot - u)$ from $\mathcal{F}_{4,0}$ with blur kernels $\bar{v}_u^{4,0}(\cdot - u)$ from $\mathcal{F}_{4,0}^{\text{self}}$. The first 49 (out of 169) left and right singular functions of these two foveation operators are shown at the center and right. The functions are sorted columnwise, according to the decreasing amplitude of the respective singular values shown in Figure 11. Observe that there is no essential structural difference between the functions corresponding to $\mathcal{F}_{4,0}$ or $\mathcal{F}_{4,0}^{\text{self}}$. Left and right singular functions of $\mathcal{F}_{4,0}^{\text{self}}$ are visually very similar but not identical, with differences in contrast discernible at the periphery.

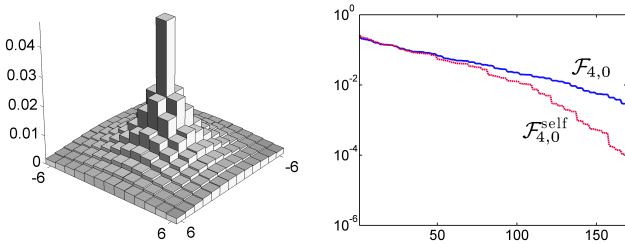


Fig. 11 The windowing kernel \mathbf{k} of size 13×13 used in the construction of the operators $\mathcal{F}_{4,0}$ and $\mathcal{F}_{4,0}^{\text{self}}$ in Figure 10 (left), and the respective singular values of these operators (right).

functions are sorted columnwise according to the decreasing amplitude of the respective singular values, which are plotted in Figure 11. The faster decay of the singular values of the self-map operator is likely due to $\zeta_u^{*,\theta}$ (49) being larger than ζ_u^* (38) (see also Figure 9). In the case of the self-map operator $\mathcal{F}_{\rho,\theta}^{\text{self}}$, we have that \mathbf{k} , the blur kernels $\bar{v}_u^{\rho,\theta}(\cdot - u)$, the left singular functions, and the right singular functions, all share the same support U . This is in contrast with the case of the operator $\mathcal{F}_{\rho,\theta}$, for which only the right singular functions have the same support of the windowing kernel \mathbf{k} , and blur kernels and left singular functions are defined on larger sup-

ports. However, as can be seen in Figure 10, when adopting the self-map construction, the main structural features of blur kernels and singular functions is maintained.

8 Foveated NL-means

We refer to *Foveated NL-means* as the modified NL-means algorithm where the windowed distance d^{WIN} (7) is replaced by a foveated distance d^{FOV} (14) computed by means of a foveation operator. Accordingly, instead of $\{w^{\text{WIN}}(x_1, x_2)\}_{x_2 \in X}$ (13), Foveated NL-means uses adaptive weights determined by the similarity between the foveated patches $\mathbf{z}_x^{\text{FOV}}$:

$$w^{\text{FOV}}(x_1, x_2) = e^{-\frac{d^{\text{FOV}}(x_1, x_2)}{h^2}} / \sum_{x \in X} e^{-\frac{d^{\text{FOV}}(x_1, x)}{h^2}}. \quad (51)$$

This is the only difference between Foveated and standard NL-means, all other parts of these algorithms are the same. The output of the Foveated NL-means, computed as in (12) using weights (51), is denoted by \hat{y}^{FOV} :

$$\hat{y}^{\text{FOV}}(x_1) = \sum_{x_2 \in X} w^{\text{FOV}}(x_1, x_2) z(x_2), \quad \forall x_1 \in X. \quad (52)$$

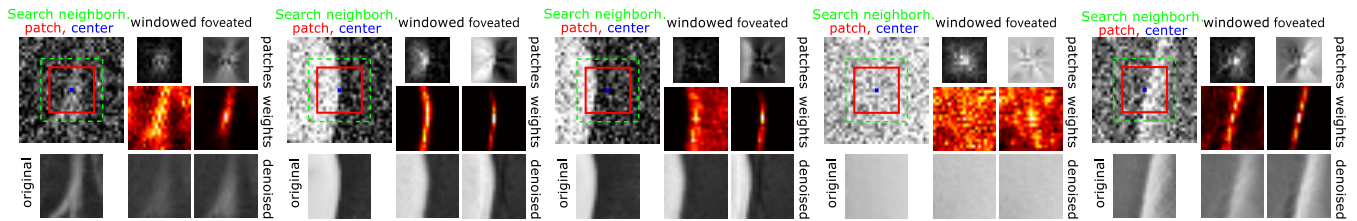


Fig. 12 Details of five 15×15 patches extracted from the noisy *Lena* image ($\sigma = 50$). For each of the five center points x (blue dots) we show the noisy patch \mathbf{z}_x (solid red line), the corresponding 21×21 search neighborhood (dashed green line), the windowed patch $\mathbf{z}_x^{\text{WIN}}$, and the foveated patch $\mathbf{z}_x^{\text{FOV}}$. The last two are scaled by $1/\sqrt{k(0)}$ for visualization, like in Figure 7. These patches are used for computing the displayed NL-means weights w^{WIN} and w^{FOV} over the search neighborhood. These weights give the pixelwise denoised estimates $\hat{y}^{\text{WIN}}(x)$ and $\hat{y}^{\text{FOV}}(x)$; for the sake of visualization, we display also all such estimates $\hat{y}^{\text{WIN}}(x')$ and $\hat{y}^{\text{FOV}}(x')$ for x' over the whole search neighborhood. The weights from foveated distance are better localized around edges, details, and textures and, as a result, foveation preserves better than windowing the original image structures. The foveated and windowed distances provide very different weights in the off-the-edge patch (middle example) since the foveated patch still features the nearby edge that is instead barely detectable in the windowed patch. In contrast, in areas where the difference between distinct patches is unstructured (white noise, as in the fourth example from left), the weights are quite similar and consequently the two denoised estimates are similar too.

We refer to it as the *Anisotropic Foveated NL-means* when the weights (51) are computed using anisotropic foveation operators such as those of Section 6.2.

A Matlab implementation of the Foveated and Anisotropic Foveated NL-means is available online⁶. Our software follows the Matlab codes by Manjon-Herrera and Buades (2008), as a reference of the original NL-means algorithm.

Figure 12 illustrates the behavior of the two algorithms at five selected locations x . The windowed and foveated patches are shown together with the corresponding weights, which have been computed over the whole search neighborhood around x . These weights determine the denoised outputs $\hat{y}^{\text{WIN}}(x)$ and $\hat{y}^{\text{FOV}}(x)$. For the sake of visualization, we show the denoised outputs not only at the center x , but also at other locations within the search neighborhood. It clearly emerges that weights computed from the foveated distance (51) feature a sharper response around image edges, details, and textures than do the weights computed from the windowed distance (13). Thus, foveation is better than windowing at discriminating image structures and, in Section 9, we show that Foveated NL-means indeed yields denoised outputs with better sharpness, contrast, and higher overall visual quality. However, such an improved localization of weights might represent a drawback in some rare cases. For example, in the off-the-edge patch (middle example), the weights of the Foveated and standard NL-means are very different, and the latter provide stronger noise attenuation. This happens because the foveated patch still features the nearby edge, which is instead barely visible in the windowed patch. Therefore, the w^{FOV} weights for the off-the-edge patch are qualitatively similar to those for the over-the-edge patch (second example from left) and, as such, do not take full advantage the dark background in the vicinity of the bright edge.

Whenever the difference between distinct patches is unstructured so that it can be treated as white noise, we can leverage the compatibility requirement (20), and conclude that the windowed and foveated distances should coincide

in expectation. Assuming that the differences in the higher-order moments of the distributions of d^{FOV} and d^{WIN} are small compared to the fluctuations in the unstructured differences between patches, it follows that the weights w^{FOV} are quite similar to w^{WIN} , and therefore the two denoised estimates $\hat{y}^{\text{FOV}}(x)$ and $\hat{y}^{\text{WIN}}(x)$ are similar too. This situation can be observed in the fourth example from left in Figure 12. Thus, the constrained design of foveation operators allowed us to avoid any adjustment of the tuning parameter h , and use instead the same value for both Foveated and standard NL-means.

Hence, we expect the Foveated and the standard NL-means to perform comparably in any region characterized by ubiquity of structurally identical patches. Where such ideal assumption is met to a weaker extent, or not met at all, d^{FOV} diverges from d^{WIN} , typically leading to an effective advantage of foveation, as we shall see in the next section.

9 Experiments

To quantitatively assess the effectiveness of foveated self-similarity as a regularization prior for natural images, we consider the removal of additive white Gaussian noise and the NL-means algorithm (Buades et al., 2005). In fact, the denoising performance summarizes at once the ability to identify similar patches and to distinguish between different ones in a noisy environment.

First, we demonstrate a clear advantage of the isotropic foveated self-similarity over the windowed self-similarity, given the superior denoising performance of Foveated NL-means and its ability to operate effectively over larger search neighborhoods (Section 9.1). Next, we dedicate several experiments to anisotropic foveation, demonstrating that radial operators can further improve the denoising, particularly at edges (Section 9.2). We then show that the same advantage holds when using compactly supported self-map foveation operators (Section 9.3) and that it is not a mere consequence of a noise attenuation operated by foveation (Section 9.4). Foveated self-similarity is shown to yield better results also on

⁶ <http://www.cs.tut.fi/~foi/FoveatedNL>

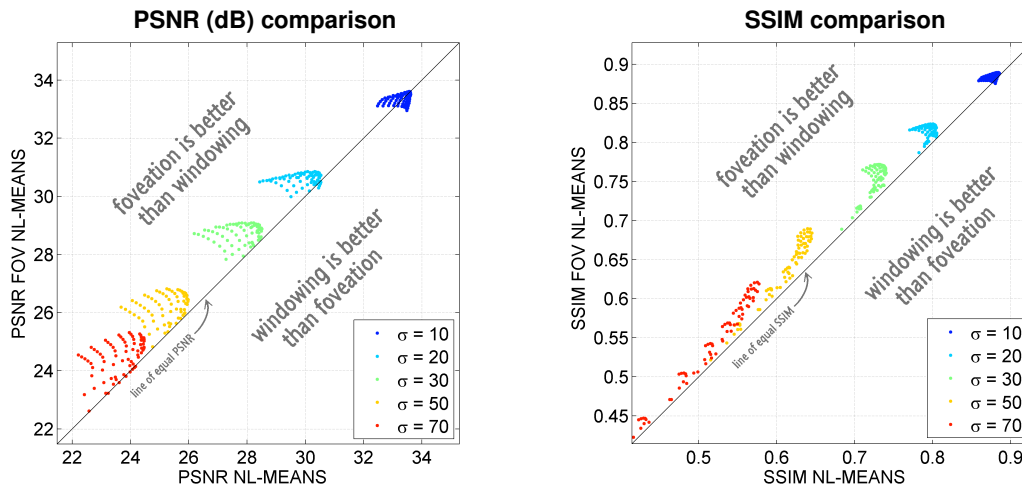


Fig. 13 Scatterplots of PSNR (left) and SSIM (right) of standard NL-means (horizontal axis) vs Foveated NL-means (vertical axis). Each point represents the PSNR or SSIM for a specific combination of patch size and search-neighborhood size. Values are averaged over the *Hill*, *Lena*, *Boats*, *Barbara*, *Man*, and *Peppers* images, each corrupted by 5 independent noise realizations. Each color indicates a separate noise level. When $\sigma > 10$, Foveated NL-means outperforms standard NL-means in all considered configurations, while when $\sigma = 10$, the best configuration of Foveated NL-means and standard NL-means yield the same performance. For the sake of visualization, we do not plot results corresponding to patch size 3×3 and search neighborhood 3×3 and 5×5 , which are very low (see Figure 14).

textures that are especially challenging for nonlocal methods (Section 9.5). Finally, we analyze the complexity and demonstrate that adopting foveation instead of windowing induces only a negligible computational overhead to NL-means (Section 9.6).

In our experiments we consider grayscale test images in the intensity range $[0, 255]$, that have been corrupted by additive white Gaussian noise with standard deviation $\sigma \in \{10, 20, 30, 50, 70\}$, according to the observation model (1).

We measure the objective denoising performance in terms of peak signal-to-noise ratio (PSNR, dB) and SSIM quality index (Wang et al., 2004), while for a subjective visual assessment we show a few image fragments (Fig. 15 and 16). A comparison in terms of method noise (Buades et al., 2005) is provided in Section Suppl.6.

9.1 Windowed vs isotropic foveation

To enable a fair and direct comparison between the standard NL-means (Buades et al., 2005) and the Foveated NL-means (8), we consider several combinations of *patch size* (ranging from 3×3 to 19×19) and *search-neighborhood size* (ranging from 3×3 to 44×44), and set the tuning parameter h equal to σ , as this choice is found to yield the best results for standard NL-means, as well as for the Foveated NL-means, consistently with the compatibility requirement. Therefore, the two algorithms differ only in the employed patch distance. Yet, the foveated distance is computed using the isotropic foveation operator \mathcal{F} (Section 6.1) constructed from the same windowing kernel \mathbf{k} of the windowed distance.

Comparison results are reported in Figure 13: in these scatterplots each point indicates the denoising performance in terms of either PSNR or SSIM index for a specific (*patch*

size, *search-neighborhood size*) combination and for a specific value of σ . The denoising performance is averaged over six standard test images, namely *Hill*, *Lena*, *Boats*, *Barbara*, *Man*, and *Peppers* (see Figure Suppl.1), each corrupted by five independent noise realizations. These values are also reported in Figure 14, arranged according to *patch size*, *search-neighborhood size*, and σ . In particular, each cloud in Figure 13 corresponds to a subfigure in Figure 14.

The scatterplots in Figure 13 show that the Foveated NL-means substantially outperforms the standard NL-means, with the only exception of low noise levels, where the two algorithms perform comparably.

When $\sigma > 10$, the PSNR achieved by the Foveated NL-means is significantly higher than the standard NL-means for all the settings, since the point clouds lie clearly above the diagonal. Furthermore, using the foveated distance guarantees improvements over the standard NL-means even with settings optimized for the latter (since the rightmost point in each cloud is well above the diagonal). Similarly, the highest point of each cloud, which corresponds to the best parameter combination for Foveated NL-means, indicates a remarkable improvement over standard NL-means. In particular, under heavy noise ($\sigma \geq 50$), foveation substantially outperforms windowing (about 1 dB). At $\sigma = 10$, the uppermost and rightmost point of the cloud lies on the diagonal, i.e. the best PSNR values from the two algorithms coincide and are obtained for the same combination of patch size and search-neighborhood size, as shown in Figure 14.

In terms of SSIM scores (Figure 13, right), foveation is always beneficial, at low noise levels too. As can be seen from the various fragments shown in Figures 15 and 16, there is a considerable visual difference between the outputs of standard NL-means and those of Foveated NL-means, with those

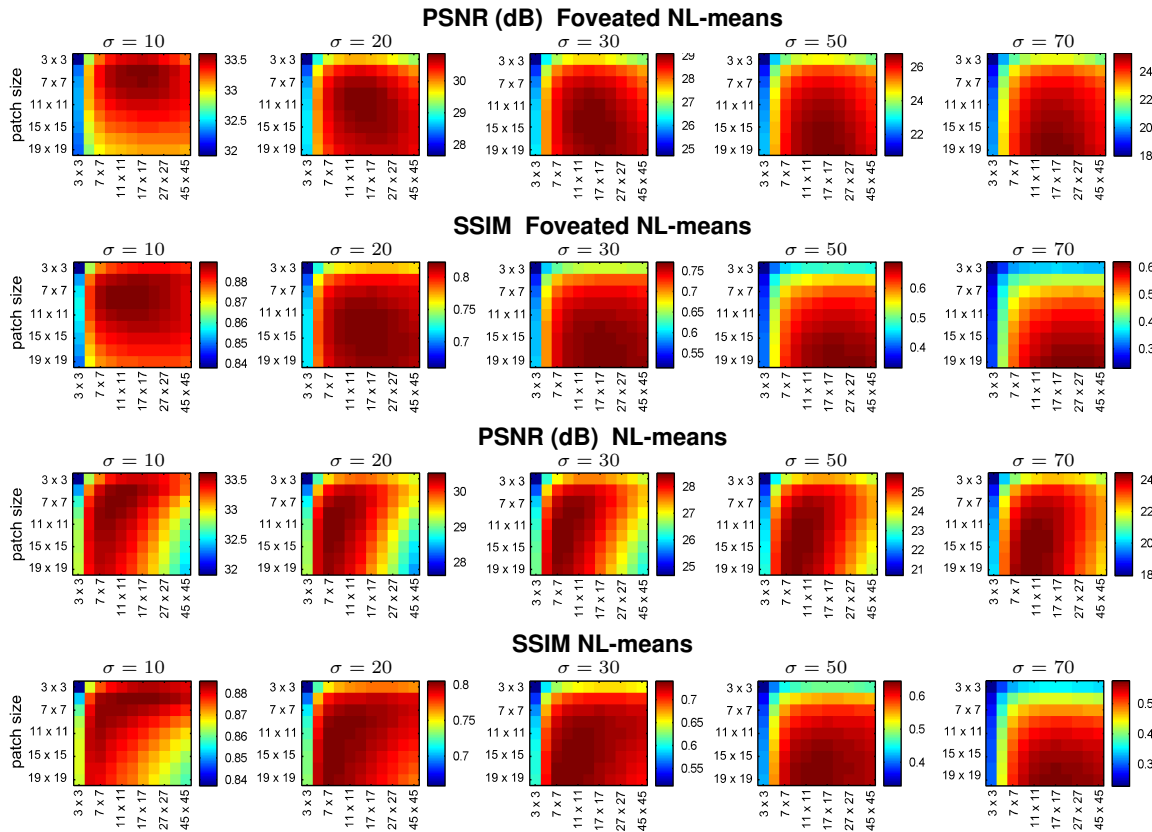


Fig. 14 PSNR (dB) and SSIM score of Foveated and standard NL-means as function of patch size and search-neighborhood size. Average results over the *Hill*, *Lena*, *Boats*, *Barbara*, *Man*, and *Peppers* images, each corrupted by 5 different noise realizations. Throughout the paper we use these results to set patch size and search-neighborhood size so to maximize the respective denoising quality of each algorithm. Specifically, for $\sigma = 10, 20, 30, 50, 70$ respectively: in standard NL-means the patch size is set to $5 \times 5, 11 \times 11, 11 \times 11, 13 \times 13, 15 \times 15$ and the search-neighborhood size is set to $11 \times 11, 9 \times 9, 9 \times 9, 11 \times 11, 13 \times 13$; in Foveated NL-means the patch size is set to $7 \times 7, 11 \times 11, 13 \times 13, 17 \times 17, 19 \times 19$, while the search-neighborhood size is set to 17×17 irrespective of σ .

by the latter characterized by a better contrast, sharper appearance, and increased detail preservation.

In the remainder, the sizes of patch and of search neighborhood are chosen so to maximize the respective denoising quality of each algorithm according to Figure 14 and as specified in its caption.

9.1.1 Local analysis To identify in which image regions d^{FOV} yields the largest improvement over d^{WIN} , we process 100 independent realizations of *Lena* corrupted by noise having $\sigma = 50$ with both Foveated and standard NL-means. We compute the pixelwise RMSE for both algorithms, namely $r^{\text{FOV}}(x)$ and $r^{\text{WIN}}(x)$, by averaging the squared errors of each algorithm over the 100 realizations. The image displayed in Figure 17 is built in the *hue-saturation-value* (HSV) colorspace and each pixel x is drawn red (*hue* $H(x) = 1$) where the Foveated outperforms standard NL-means, i.e. $r^{\text{FOV}}(x) < r^{\text{WIN}}(x)$, or blue ($H(x) = 2/3$) otherwise. To highlight the pixels where the performance gap is larger, the *saturation* $S(x)$ proportional to $|r^{\text{FOV}}(x) - r^{\text{WIN}}(x)|$. The *value* V is fixed equal to 1, yielding white pixels where there is no substantial difference between $r^{\text{FOV}}(x)$ and $r^{\text{WIN}}(x)$. Consequently, the color in each pixel indicates the best performing algorithm,

with the color getting stronger at locations where the performance gap between the two algorithms gets larger.

The color that dominates Figure 17 is red, confirming that Foveated NL-means outperforms the standard NL-means in terms of global quality indexes. Furthermore, red pixels indicate that d^{FOV} is remarkably more effective than d^{WIN} at image edges and at fine details, corroborating the conclusions one can draw from Figures 15 and 16. Flat areas of Figure 17 are instead white, because the difference between distinct patches is unstructured, and the two algorithms perform similarly. Near T-junctions and at off-the-edge regions we can see some blue pixels, indicating that there windowing is better than foveation. These are instances of the so-called rare-patch effect (see Louchet and Moisan, 2011, and particularly Figure 7 therein), an issue common to the NL-means and other nonlocal methods. The increased discrimination provided by the foveated distance (that can be deduced also from the distribution of weights of Foveated NL-means in Figure 12 at patches close to the image edges), as well as the larger patch size used in Foveated NL-means, make this effect slightly more prominent. Several adaptive methods for dealing with this phenomenon exist Sutour et al. (2014, and references



Fig. 15 Comparison between outputs by the NL-means algorithm and by the proposed Isotropic and Anisotropic ($\rho=4$) Foveated NL-means. The numbers between parentheses are the PSNR (db) and SSIM scores computed for the entire image. Results are given under two combinations of patch size and search neighborhood size, one ideal for the standard NL-means, another ideal for the Foveated NL-means (see Figure 14). The shown fragments have size 155×175 pixel.

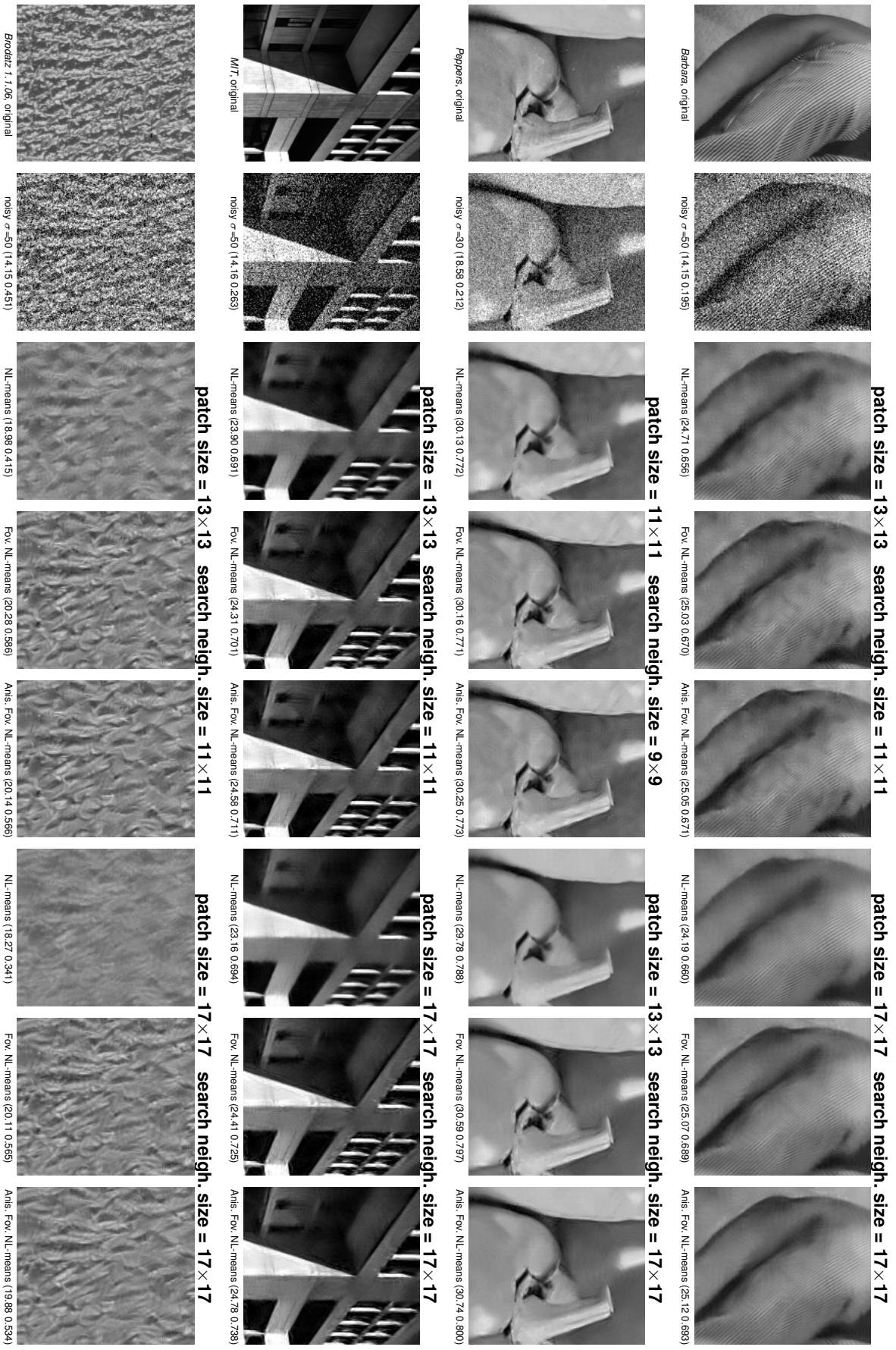


Fig. 16 Comparison between outputs by the NL-means algorithm and by the proposed Isotropic and Anisotropic ($\rho=4$) Foveated NL-means. The numbers between parentheses are the PSNR (db) and SSIM scores computed for the entire image. Results are given under two combinations of patch size and search neighborhood size, one ideal for the standard NL-means, another ideal for the Foveated NL-means (see Figure 14). The shown fragments have size 155×175 pixel.



Fig. 17 Red color indicates pixels where Foveated NL-means outperforms (i.e. has lower RMSE) the standard NL-means, and blue otherwise. The color saturation is proportional to the pixelwise absolute RMSE difference between the two algorithms; at white pixels this difference is negligible. Saturated red pixels indicate that foveation is better at edges and fine details, while blue pixels indicate that windowing is better in few other regions like T-junctions and off-the-edge regions. These results are consistent with the distribution of NL-means weights shown in Figure 12. However, red pixels are more frequent and more saturated than the blue ones. This means that not only foveation is most often more effective than windowing, but also that this advantage is consistently large.

therein) and are arguably valid for the Foveated NL-means also.

More experiments with comparable outcome are reported by us in (Foi and Boracchi, 2012) and by Postec (2012).

Overall, the experiments show that the foveated distance is remarkably better than the windowed distance, and that it can be seamlessly plugged into the NL-means to improve the denoising, particularly at high noise levels. The superior results achieved through patch foveation are not a consequence of some sort of parameter adjustment, but rather indicate towards the foveated distance being a more effective measure for assessing the nonlocal self-similarity in images. In particular, the PSNR and SSIM values reported in Figure 14 show that Foveated NL-means is able to effectively operate on larger search neighborhoods than the standard NL-means. In general, increasing the search-neighborhood size brings more potentially similar patches, but also more potentially irrelevant patches; moreover, the probability of finding repeated noise patterns grows with the size of the search neighborhood. Thus, since typically the most similar patches are found near the neighborhood center, large search neighborhoods are hardly beneficial unless the distance measure can effectively discriminate which patches are truly similar and which not. Therefore, the results in Figure 14 confirm that foveation allows gathering more similar patches farther from the center of the search neighborhood while it enables a more effective rejection of the many dissimilar patches that can be found far from the center of the search neighborhood, thus better coping with the increased probability of false positives.

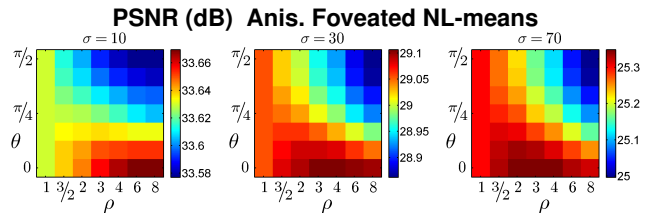


Fig. 18 Denoising performance of Anisotropic Foveated NL-means for various combinations of the θ and ρ parameters, measured as average PSNR (dB) over the *Hill*, *Lena*, *Boats*, *Barbara*, *Man*, and *Peppers* images, each corrupted by 5 independent noise realizations. The best results can be achieved by anisotropic foveation operators characterized by blur kernels elongated along the meridians toward the patch center (i.e. $\theta = 0, \rho > 0$). In contrast, operators with tangential blur kernels (i.e. $\theta = \pi/2, \rho > 0$) are particularly ineffective and even lead to a performance loss with respect to the isotropic ones. SSIM results are similar and are reported in Figure Suppl.5.

9.2 Anisotropic foveation

We here evaluate how the denoising performance of Anisotropic Foveated NL-means depends on the kernels elongation and orientation parameters. Specifically, we consider $\mathcal{F}_{\rho, \theta}$, $\rho \in [1, 8]$ and $\theta \in [0, \pi/2]$, thus including both radial, tangential and isotropic operators.⁷

Figure 18 reports the average PSNR scores as a function of ρ and θ , for separate levels of noise standard deviation σ . For each value of σ , we select the best patch size and search-neighborhood size for the isotropic Foveated NL-means as reported in Section 9.1. These results indicate that patch similarity is more effectively measured by radial anisotropic foveation operators ($\theta = 0, \rho > 1$), which outperform isotropic ($\rho = 1$), chiral ($0 < \theta < \pi/2, \rho > 1$), and tangential ones ($\theta = \pi/2, \rho > 1$). The corresponding SSIM results support the same conclusions and are reported in Figure Suppl.5.

9.2.1 Local analysis To further investigate this point, we generate 50 noisy realizations of *Lena* ($\sigma = 50$) and we denoise them using the Foveated NL-means with patch distance defined by means of 13 foveation operators $\{\mathcal{F}_{\rho, 0}, \rho = 1/8, \dots, 1, \dots, 8\}$, including radial ($\rho > 1$), isotropic ($\rho = 1$), and tangential ($\rho < 1$) ones⁸. For each value of ρ , we compute $r_{\rho}^{\text{FOV}}(x)$, namely the pixelwise RMSE (as average over different noise realizations), and we report in Figure 19 the value of ρ of the operator that yields the lowest $r_{\rho}^{\text{FOV}}(x)$ at an individual pixel x . We adopt a HSV color coding similar to that of Figure 17. Specifically, the *hue* is set to red ($H(x) = 1$) at those pixels where the lowest restoration error is achieved by the most radially elongated foveation operator ($\rho = 8$), while it is blue ($H(x) = 2/3$) where the best operator is tangential with $\rho = 1/8$; the *hue* for $1/8 < \rho < 8$ is defined by logarithmic interpolation, therefore a pixel x is magenta

⁷ The selected parameter range is general enough since on natural images the denoising performance of *left-hand* chiral operators ($\rho \geq 1, -\pi/2 < \theta < 0$) is practically identical to those of *right-hand* chiral operators ($\rho \geq 1, 0 < \theta < \pi/2$), as shown in Section Suppl.4.

⁸ For Gaussian blur kernels, $\mathcal{F}_{\rho, 0} = \mathcal{F}_{1/\rho, \pi/2}$.



Fig. 19 The colors indicate the foveation operator yielding the lowest RMSE in denoising the current pixel: red corresponds to radial, magenta to isotropic, and blue to tangential foveation. Colors are more saturated where the RMSE difference between different operators is larger; pixels are white where all operators yield the same RMSE. Radial foveation is the best one at capturing edges and contours, which are mainly red. Isotropic foveation is instead better than others in irregular textures, like the feathers attached to the hat, which are mostly magenta. Pale blue pixels can be observed where levels sets are parallel to a flanking edge, as the blur kernels at the periphery align with the edge, thus focusing the weights along the levels sets. The improvement remains however marginal. In flat areas the image is basically white, as there all the operators are practically equivalent.

($H(x)=5/6$) when the best operator is the isotropic one. The *saturation* $S(x)$ reflects the pixelwise standard deviation of $\{r_\rho^{\text{FOV}}(x), \rho = 1/8, \dots, 1, \dots, 8\}$, while the *value* V is fixed equal to 1 ($V(x)=1$). Thus, like in Figure 17, colors are more saturated (hence better visible) where the denoising performance varies more with respect to ρ , and the image is instead white where all operators perform similarly. It emerges that along edges and fine details, patch similarity is more effectively assessed by means of radial foveation operators. This can be explained by the fact that radial foveation preserves the substantial edge structure, since it blurs along rather than across the edges. Similar arguments have driven the development of the “Anisotropic NL-means” (Maleki et al., 2013), where the patch domain U (or equivalently the windowing kernel \mathbf{k}) is elongated along the edge orientation, to emphasize such structure in the patch comparison.

9.2.2 Analysis on images dominated by sharp edges As an additional test to demonstrate that radial foveation is more effective in measuring self-similarity at edges, we consider the *MIT* and *Cameraman* test images (see Fig. Suppl.2), which are characterized by a large amount of sharp edges. Indeed, the corresponding PSNR analysis in Figure 20 shows a more marked advantage of radial foveation than in Figure 18. SSIM results (see Fig. Suppl.6) conform to this trend. This enhanced noise suppression at edges can be observed, upon careful inspection, in Figure 16.

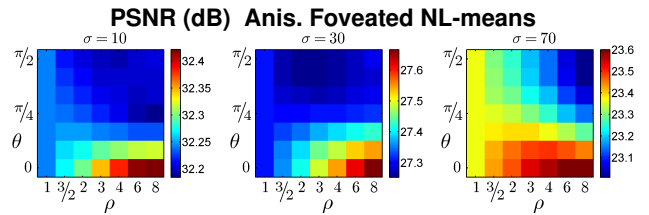


Fig. 20 Denoising performance of Anisotropic Foveated NL-means computed as in Figure 18 but over *Cameraman* and *MIT*. These images, which are characterized by a large amount of sharp edges, show a marked improvement of radial operators over the others. The better ability of radial foveation operators to denoise sharp edges is consistent with the local analysis reported in Figure 19. SSIM results are similar and are reported in Figure Suppl.6.

Although the performance gap between radial and isotropic foveation reported in Figure 18 and 20 is small in comparison with the gap between (isotropic) foveation and windowing shown in Section 9.1, such improvement is consistently found in all our experiments on complex natural images, and indicates that radial foveation yields a stronger prior for measuring the nonlocal similarity with structured content (e.g., objects, edges, contours).

9.3 Self-map foveation operators

Because the support of some blur kernels of the foveation operator (22) exceed the patch neighborhood U , one might legitimately speculate whether this is the reason of the performance gap between the Foveated and standard NL-means. In fact, computing d^{FOV} involves also pixels outside U , and this could bring an advantage to the Foveated over the standard NL-means, since d^{WIN} is instead computed exclusively from pixels within U . To further investigate the reason of such performance gap, we have compared the Foveated NL-means of Section 9.1 against the Foveated NL-means given by the self-map foveation operator $\mathcal{F}_{1,0}^{\text{self}}$ of Section 7. To enable a meaningful comparison, both operators $\mathcal{F} = \mathcal{F}_{1,0}$ and $\mathcal{F}_{1,0}^{\text{self}}$ are isotropic ($\rho = 1, \theta = 0$) and obtained from the same windowing kernel \mathbf{k} . The two mainly differ in the pixels considered for computing the patch distance: \mathcal{F} uses also pixels *outside* U , while $\mathcal{F}_{1,0}^{\text{self}}$ does not. PSNR and SSIM results are reported in Section Suppl.2, and show that $\mathcal{F}_{1,0}^{\text{self}}$ is only marginally less effective than its counterpart \mathcal{F} , with PSNR differences of at most 0.1 dB and typically well below 0.05 dB, and SSIM differences of at most 0.005, over the *Lena* image and $10 \leq \sigma \leq 70$. Such differences appear insignificant especially in comparison with those between the Foveated and standard NL-means. Indeed, the direct numerical comparison between standard NL-means and self-map Foveated NL-means reported in the second and third columns of Table 1 reaffirms for the self-map operators the same appraisal expressed in Section 9.1 on the non-self map operators. Hence, the advantage provided by the foveated distance cannot be explained by the supports of the blur kernels exceeding U . Therefore, the source of such advantage should

PSNR (dB) and SSIM comparison

σ	noisy		w^{WIN} and w^{FOV} computed from z			oracle $w^{\text{WIN,Oracle}}$ and $w^{\text{FOV,Oracle}}$ computed from y		
			NLM	Fov. NLM	Anis. Fov. NLM	NLM	Fov. NLM	Anis. Fov. NLM
10	28.13dB	0.668	33.60dB 0.886	33.52dB 0.887	33.62dB 0.890	34.14dB 0.888	33.75dB 0.884	33.96dB 0.889
20	22.11dB	0.401	30.53dB 0.805	30.79dB 0.820	30.87dB 0.822	30.81dB 0.806	30.99dB 0.818	31.16dB 0.823
30	18.59dB	0.265	28.51dB 0.736	29.00dB 0.766	29.08dB 0.768	28.84dB 0.751	29.16dB 0.768	29.34dB 0.774
50	14.15dB	0.140	25.93dB 0.634	26.71dB 0.682	26.80dB 0.685	26.14dB 0.664	26.77dB 0.696	26.95dB 0.702
70	11.23dB	0.085	24.37dB 0.562	25.25dB 0.616	25.34dB 0.620	24.52dB 0.607	25.29dB 0.645	25.48dB 0.652

Table 1 Denoising performance of standard, Foveated, and Anisotropic Foveated NL-means ($\rho = 4$), reported in terms of mean PSNR(dB) and SSIM over the *Hill*, *Lena*, *Boats*, *Barbara*, *Man*, and *Peppers* images, each corrupted by five independent noise realizations. The three columns at the right refer to the special experiment with oracle weights w^{WIN} and w^{FOV} computed from the noise-free image, as described in Section 9.3. In these experiments, foveation is always produced by self-map foveation operators.

be traced back to an enhanced capability of the foveated distance to assess the nonlocal self-similarity in natural images.

In the third and fourth column of Table 1 we compare self-map isotropic and self-map radial anisotropic operators with $\rho = 4$, and confirm a limited but consistent gain of radial over isotropic foveation for the self-map operators also.

9.4 Oracle weights

Patch foveation can be interpreted as special smoothing applied to noisy patches. This is well described by the decaying spectra of foveation operators shown in Figure 8(right) and Figure 11, with the first few singular functions being smoother than the rest. One might even legitimately doubt whether the advantage of foveation over windowing follows mainly from the attenuation of noise operated by the blur kernels, rather than from a better representation of the patch. Indeed, as we observed in Section 6.1.1, although corresponding pixels of windowed and foveated patches are subject to the same noise variance (41), the windowed patches suffer from an inferior SNR. We address this issue through a simulation where both windowed and foveated patches are not degraded by noise and enjoy ideal infinite SNR. Specifically, in the following experiment, we replace the windowed patch distance d^{WIN} (6) with its *oracle* counterpart $d^{\text{WIN,Oracle}}$,

$$d^{\text{WIN,Oracle}}(x_1, x_2) = \|\mathbf{y}_{x_1}^{\text{WIN}} - \mathbf{y}_{x_2}^{\text{WIN}}\|_2^2,$$

i.e. the Euclidean difference between noise-free windowed patches $\mathbf{y}_x^{\text{WIN}} = \mathbf{y}_x \sqrt{\mathbf{k}}$. This *oracle windowed patch distance* can be used in place of d^{WIN} when computing weights (13), yielding *oracle weights* $w^{\text{WIN,Oracle}}$ for NL-means:

$$\hat{y}^{\text{WIN,Oracle}}(x_1) = \sum_{x_2 \in X} w^{\text{WIN,Oracle}}(x_1, x_2) z(x_2) \quad \forall x_1 \in X.$$

Here the only modification concerns the weights and the estimate is still given as a weighted average of noisy pixels. In the same manner, oracle weights $w^{\text{FOV,Oracle}}$ can be computed for the Foveated NL-means, substituting d^{FOV} in (51) with the *oracle foveated patch distance*

$$d^{\text{FOV,Oracle}}(x_1, x_2) = \|\mathbf{y}_{x_1}^{\text{FOV}} - \mathbf{y}_{x_2}^{\text{FOV}}\|_2^2,$$

where $\mathbf{y}_x^{\text{FOV}} = \mathcal{F}[y, x]$. Neither $w^{\text{WIN,Oracle}}$ nor $w^{\text{FOV,Oracle}}$ should be interpreted as the best weights one can possibly use in (12) or (52) to, e.g., maximize PSNR (e.g., Jin et al., 2011). These weights simply remove the impact of noise from the patch distance, hence allowing a more direct comparison of the effectiveness of the underlying self-similarity models.

Results in Table 1 show that oracle weights (fifth to seventh column) provide a marginal improvement over the usual weights w^{WIN} and w^{FOV} (second to fourth column), and that the performance gaps between foveation and windowing is substantially unaltered. As a matter of fact, and perhaps surprisingly, when $\sigma \geq 20$, the standard NL-means with oracle weights $w^{\text{WIN,Oracle}}$ is even outperformed by the Foveated NL-means with usual weights w^{FOV} computed from noisy foveated patches. We can therefore conclude that foveation works better not simply because foveated patches feature a higher SNR, but essentially because it enables a more effective measure of the nonlocal self-similarity.

9.5 Texture

Although very regular periodic textures possibly represent the ideal conditions for nonlocal algorithms, in some practical cases textures are characterized by sporadic patterns and thin random details that make very difficult to find a close matching replica of a sufficiently large patch. In these cases, textures can prove very challenging for nonlocal algorithms. Therefore, it is important to benchmark the windowed and the foveated distance separately on such textural patterns. To this purpose, we selected six images from the Brodatz dataset (USC-SIPI) as a challenging dataset for nonlocal denoising (Fig. Suppl.3), repeating the experimental study with multiple combinations of patch and neighborhood sizes from Section 9.1. The results (see Section Suppl.5) conform to those on natural images in Figure 13 and demonstrate that foveated self-similarity is effective also on challenging textures. In particular, the quality of the filtered images (see Figure 16(bottom) and Figures Suppl.10 and Suppl.11) confirms that foveation can be substantially more convenient than windowing also without any specific parameter adjustment to match the content of the processed images.

PSNR (dB) vs compute time (sec.) Matlab single-threaded on Intel Mobile T8300 @ 2.6GHz

	patch size = 11×11, search-neighb. size = 9×9						patch size = 13×13, search-neighb. size = 17×17					
	NLM		Fov. NLM		Anis. Fov. NLM		NLM		Fov. NLM		Anis. Fov. NLM	
<i>Lena</i>	29.85dB	2.8s	30.24dB	6.5s	30.26dB	6.0s	29.39dB	3.4s	30.65dB	9.6s	30.68dB	9.0s
<i>Man</i>	27.84dB	139.9s	28.34dB	140.1s	28.40dB	140.0s	27.06dB	288.1s	28.26dB	288.1s	28.34dB	288.0s
<i>Peppers</i>	30.18dB		30.17dB		30.26dB		29.83dB		30.60dB		30.74dB	

Table 2 PSNR vs compute time of NL-means, Foveated NL-means, and Anisotropic Foveated NL-means (noise $\sigma = 30$, image size 512×512). The patch and search-neighborhood sizes of the first three columns are those that yield best overall results for NL-means, while those of the last three columns are optimal for (isotropic) Foveated NL-means (see Fig. 14). For each image and algorithm we report the average PSNR over five noise realizations. For each algorithm we report the average compute times in seconds over the three images and five noise realizations, separating the time needed for either windowing or foveating the patches (top) from that needed for computing weights and averaging (bottom).

9.6 Computational complexity

Table 2 shows that the computation times of Foveated NL-means are marginally higher than those of the standard NL-means. The overhead is due to the computation of the foveated patches (in lieu of windowing), which is nevertheless executed only once on the whole image, before computing d^{FOV} . Once either patch windowing or patch foveation are computed, calculation of d^{WIN} or d^{FOV} have identical complexity. These pairwise patch comparisons in (13) are the most time consuming operation in the algorithm. Thus, the overhead is negligible with respect to the overall computation time.

Foveated patches can be conveniently computed by convolving the whole image z against the distinct blur kernels (modulo translations) v_u of the foveation operator (22). The foveated patches can be then assembled by suitably selecting values from these convolved images, as will be also illustrated in Section 10.3. This is a particularly efficient strategy in case of isotropic foveation operators, where the number of distinct blur kernels can be much smaller than the number of pixels in the patch: in particular, there is one distinct blur kernel for each distinct value that the windowing kernel k takes over U . As an illustrative example, the kernel k reported in Figure 5 takes only 5 distinct values, thus all foveated patches can be obtained by convolving z against the 5 distinct blur kernels shown in Figure 6(second row). However, the number of distinct blur kernels in an anisotropic foveation operator is typically $(|U| + 1)/2$, where $|U|$ is the number of pixels in the patch, as can be observed in Figure 8.

10 Discussion

The superior sharpness and contrast achieved by the Foveated NL-means come perhaps unexpected, because the self-similarity is assessed by means of foveated patches that are indeed blurry (see foveated patches in Figures 7 and 12). This paradox is however only apparent, because what actually matters in the weighted average (12) is that the weight $w^{\text{WIN}}(x_1, x_2)$ or $w^{\text{FOV}}(x_1, x_2)$ is kept small when $|y(x_1) - y(x_2)|$ is large and vice versa. The superior denoising performance achieved through foveation suggests that the distance between $y(x_1)$ and $y(x_2)$ correlates better with

the foveated patch distance $d^{\text{FOV}}(x_1, x_2)$ than with the windowed patch distance $d^{\text{WIN}}(x_1, x_2)$. This should not surprise, since high frequencies enjoy a much shorter-range correlation than the low frequencies. In particular, when making inference about the patch center, the high-frequency content at the patch periphery is not as relevant as the low-frequency content. It follows that foveation, which attenuates the high-frequency content at the periphery of the patch, indeed emphasizes the information useful for the purpose of nonlocal denoising of the patch center.

10.1 Comparison with other nonlocal denoising methods

In spite of the notable improvement achieved when introducing foveation in NL-means, the performances of both Foveated and Anisotropic Foveated NL-means are still inferior to that of more sophisticated nonlocal filters, such as, e.g., BM3D (Dabov et al., 2007b), BM3D-SAPCA (Dabov et al., 2009), NL-Bayes (Lebrun et al., 2013) or SAFIR (Kervrann and Boulanger, 2006). As a figure of merit, when denoising *Lena* corrupted by noise with $\sigma = 30$, Foveated NL-means and Anisotropic Foveated NL-means achieve a PSNR of 30.62dB and 30.65dB, respectively, versus 29.82dB of NL-means (see Figure 15), while BM3D achieves 31.26dB, BM3D-SAPCA 31.42dB, NL-Bayes 31.19dB, and SAFIR 31.16dB. Nevertheless, Foveated NL-means outperforms many other variants of NL-means that leverage alternative patch-similarity measures, such as the NL-means variant presented by Thaipanich et al. (2010) (30.04dB on *Lena* with $\sigma = 30$), the NL-means measuring patch similarity by means of Zernike moments (Ji et al., 2009) (30.35dB on *Lena* with $\sigma = 30$), the NL-means using rotation-invariant block-matching by Grewenig et al. (2011) (PSNR of 31.84dB on *Lena* with $\sigma = 20$, while Foveated NL-means achieves 32.45dB), the ‘‘Anisotropic NL-means’’ by Maleki et al. (2013) (PSNR of 26.45dB and 23.88dB on *Boats* with $\sigma = 38.25 = 0.15 \times 255$ and $\sigma = 63.75 = 0.25 \times 255$, while Foveated NL-means achieves 27.24dB and 24.99dB, respectively), and the patchwise NL-means (Buades et al., 2011) (29.51dB on *Lena* with $\sigma = 30$). However, our contribution is not to be intended as the development of yet another denoising algorithm, but rather as the exploration of a new form of nonlocal self-similarity.

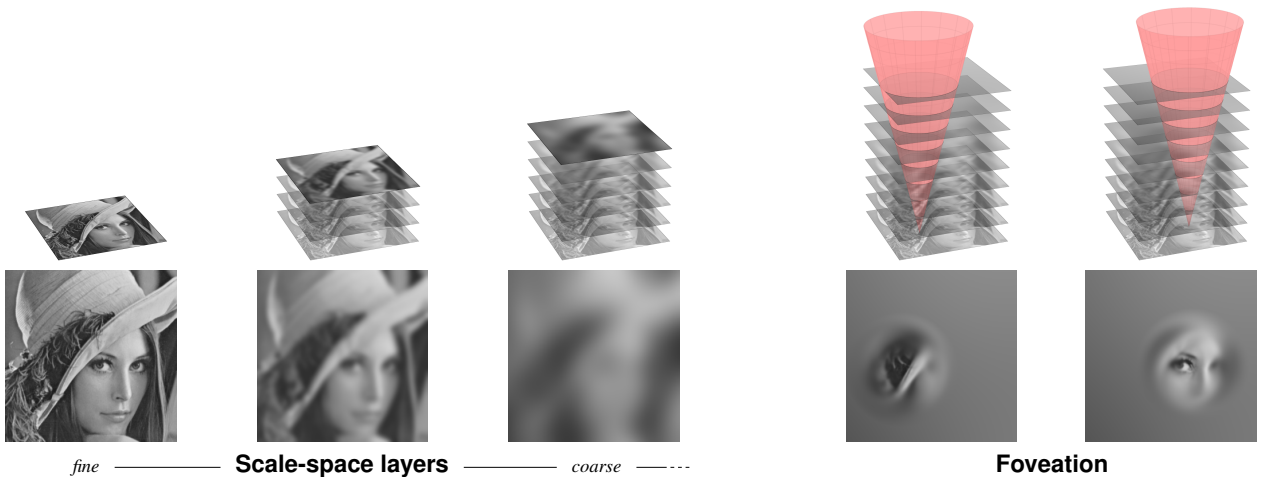


Fig. 21 Isotropic foveation can be interpreted as conical sectioning of a scale-space representation of the image, i.e. the foveated image is composed by pixels that are extracted from concentric subset in different layers of the scale-space, each layer portraying a different degree of coarseness. Each of these concentric subset is found as the intersection of the layer with a cone stemming from the fixation point in the finest layer. In the illustration, the bottom and finest layer is the fully sharp original image; as we move to the higher and coarser layers the image is progressively blurred, with the top and coarsest layer being the limiting case of the average of the whole image.

10.2 Other applications

Since our first work on the subject (Foi and Boracchi, 2012), the foveated patch self-similarity has been employed for a diverse range of imaging applications, which go well beyond our basic denoising examples. We wish to mention the works by Chierchia et al., where the foveated self-similarity defines the weights in nonlocal total-variation for recovering images from blurred, noisy, and incomplete observations (most samples missing due to random decimation) (Chierchia et al., 2014a; Chierchia et al., 2014), and later also combined with a nonlocal structure tensor for restoring multi-component (color and multi/hyper-spectral) images (Chierchia et al., 2014b). The use of patch foveation is not even limited to image restoration, but can be relevant also for various image analysis problems. Ciresan et al. (2012) successfully addressed the segmentation of electron microscopy image stacks in neuroanatomy by a deep neural network that takes as input foveated patches, winning a challenge at the IEEE ISBI conference. Similarly, patch foveation is used by Haloi (2015) for preprocessing the input of a deep neural network trained for the automatic detection of retinopathic microaneurysm in fundus images.

We provide an open-source implementation of our foveation operators⁶, and encourage researchers to experiment adopting patch foveation within their patch-based algorithms.

10.3 Multiscale patch descriptors

In contrast with the conventional windowing, which is only spatially selective and attenuates in an equal way sharp details and smooth areas, patch foveation provides selectivity in both space and frequency domain. In fact, foveated patches embed pixels from fine-scale (the fixation point) to coarse-scale (pixels at the patch periphery) representations of patches, and, re-

markably, such a space/frequency selectivity is achieved by means of a single, non-adaptive operator. Foveation operators therefore provide a compact and non redundant multi-scale representation of each image patch. In particular, foveation can be interpreted as a conical sectioning of the scale-space representation of an image, as illustrated in Figure 21. The central-acuity and the pixelwise-compatibility requirements can be directly interpreted in terms of such scale-space sectioning. In particular, the central acuity implies that the central pixel is extracted from the lowest layer z , while the pixelwise compatibility implies that upper and coarser layers are progressively selected as the distance from the center increases. The actual shape of the cone depends directly on the windowing kernel k , and in particular on its decay rate and on whether it is circularly symmetric or not.

To investigate the importance of such a regularity in the scale-space sectioning, Section Suppl.7 presents an experiment using modified “foveation” operators that violate the central-acuity and pixelwise-compatibility requirements.

10.4 Extension to patchwise multipoint estimators

The foveated self-similarity framework and the Foveated NL-means algorithm developed in the previous sections are focused on a nonlocal pointwise estimation (see, e.g., Katkovnik et al., 2010) where a single pixel is estimated at a time. Specifically, in the Foveated NL-means algorithm, the particular adaptive weighted average (52) produces the estimate $\hat{y}^{\text{FOV}}(x_1)$ only; any other estimate at a position $x_3 \neq x_1$ has to be computed separately from a different set of weights $w^{\text{FOV}}(x_3, x_2)$, $x_2 \in X$. Further, the blurred pixels at the periphery of the patches $z_{x_2}^{\text{FOV}}$, $x_2 \in X$, although essential for determining the weights $w^{\text{FOV}}(x_1, x_2)$ (51), do not enter directly into the summation (52), which instead combines noisy samples $z(x_2)$. What matters here is that $z(x_2)$ are, apart from a

trivial scaling by $\sqrt{\mathbf{k}(0)}$, the central pixels of the patches $\mathbf{z}_{x_2}^{\text{FOV}}$ because of the central acuity requirement. This implies that the foveated patch model that underpins the self-similarity is completely transparent with respect to the data samples that enter the weighted average. When this is violated the estimation may degenerate, as exemplified in Section [Suppl.7](#).

Different from the pointwise estimators, many of the abovementioned competitive denoising methods are in fact nonlocal multipoint estimators (see, e.g., [Katkovnik et al., 2010](#)), which typically produce simultaneously estimates for each pixel in a patch. For example, the BM3D achieves its efficiency by pushing the concept to the extremes: at once, it produces an individual estimate for each and every pixel belonging to any of the mutually similar patches that are involved in the estimation (this is the so-called collaborative filtering, [Dabov et al., 2007b](#)).

Extending the foveated self-similarity to the class of nonlocal multipoint estimators requires special care. In fact, patch foveation not only affects the selection of patches or the calculation of relative weights used in filtering, but bears also structural changes to the patch regularity model leveraged in multipoint estimation. Specifically, substituting in a multipoint filter its native patch distance with the foveated patch distance, installs a similarity model between the patches $\mathbf{z}_x^{\text{FOV}}$ that can substantially deviate from the canonical similarity between the patches \mathbf{z}_x . It means that a modification of the distance must be accompanied by a modification of the way mutual regularity is exploited for noise attenuation and signal recovery. These are additional technical challenges specific to multipoint methods which were not met in the straightforward transition from NL-means to Foveated NL-means.

We recently made some preliminary steps toward the use of foveation in multipoint estimators. One such instance are the Foveated Nonlocal Principal Components ([Foi and Boracchi, 2014](#)), where we applied the foveated distance to identify suitable patch clusters for adaptive dictionary learning and low-rank patch approximation in noisy conditions. While we refer the reader to ([Foi and Boracchi, 2014](#)) for additional details, we mention here some of the key elements of this method, so to provide an intuition how far the change of the similarity metric impacts on the baseline method. In this case, the baseline method is a similarity-weighted principal component analysis (PCA) where weights are given by the Euclidean distance with respect to a given ‘‘centroid’’ or ‘‘reference’’ patch; patches close to the centroid can thus be projected on a low-rank subspace given by the first few PCs. Similar archetypal approximations are at the core of state-of-the-art methods such as BM3D-SAPCA ([Dabov et al., 2009](#)). In contrast with NL-means, where we are able to replace the patch distance with a suitably constrained foveated distance without need of further changes to the algorithm, with the nonlocal principal components foveation must spread also to many other elements of the algorithm. Firstly, the weights (51) that are used for weighted PCA are relevant for the foveated patches $\mathbf{z}_x^{\text{FOV}}$ and not the patches \mathbf{z}_x , therefore the PCA has to be computed for a cluster of foveated patches. While the found eigenvectors are in principle able to approximate

any patch, their construction guarantees good approximation for foveated patches only. Being orthonormal, the PCA corresponds to a rotation of the coordinates and alone it cannot incorporate foveation, which is a more general non-symmetric linear transformation. Hence, we endow the analysis and synthesis operators with respect to this eigendecomposition with, respectively, the foveation operator and its de-foveating inverse⁹. In spite of the obvious burden due to the additional foveation and de-foveation steps, the overall procedure is successful, in the sense that if we perform low-rank approximation using clusters of mutually similar foveated patches, the approximation of the patches \mathbf{z}_x is closer to the noise-free patches \mathbf{y}_x than what one would obtain using an approximation of same low rank originated from the Euclidean or windowed distance ([Foi and Boracchi, 2014](#)).

10.5 Connection with HVS orientation preference

The improvement attained by radial foveation surprisingly recalls the *radial orientation bias* of the HVS (see Section 2.1.1), which is the enhanced sensitivity to radial orientations in the visual field, i.e. orientations collinear with the meridians intersecting the center of the gaze. The layout of blur kernels in radial foveation operators (see the examples in Figure 8) agrees with such orientation preference. Furthermore, we observe that patches blurred by radial foveation operators actually preserve edges and sharp details that are oriented towards the patch center. In this sense, radial foveation induces a patch similarity measure that mimics the HVS sensitivity. To further inspect the connection between radial foveation and the radial orientation bias of the HVS, we designed the following experiment.

The orientation preference of the HVS has been also investigated by means of functional Magnetic Resonance Imaging (fMRI). In a recent experiment ([Freeman et al., 2011](#)), an annular region in the human primary visual cortex (V1) was monitored, and the fMRI responses to a visual stimulus consisting in a rotating sinusoidal grating were recorded. This experiment revealed a strong response to stimuli that are oriented along the angular direction of the retinally mapped V1 voxel, and the resulting topographic map of orientation preference was characterized by radial orientations, congruent with previous studies on the radial bias.

To investigate whether a similar topographic map can be also found in anisotropic foveation, we run an experiment where a specific foveated distance mimics the sensitivity of a single directional *receptor*. Let us consider anisotropic foveation operators $\rho > 1$ and define

$$\begin{aligned} d_{\rho, \theta, \bar{u}}(x_1, x_2) &= \\ &= (\mathcal{F}_{\rho, \theta}[z, x_1](\bar{u}) - \mathcal{F}_{\rho, \theta}[z, x_2](\bar{u}))^2 \frac{\|\mathbf{k}\|_1}{\mathbf{k}(\bar{u})}, \quad (53) \end{aligned}$$

⁹ Invertibility is feasible with self-map operators $\mathcal{F}_{\rho, \theta}^{\text{self}}$ provided that the inequality (50) is met with a strict lower bound. As a matter of fact, the greatest advantage of self-map operators is that they can be represented through $\mathcal{B}_{\rho, \theta}$ (46) as square matrices of size $|U| \times |U|$.

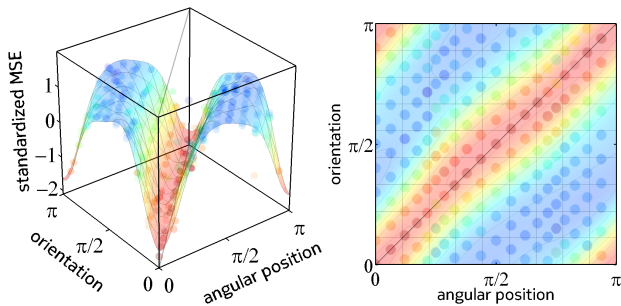


Fig. 22 Human visual system and Anisotropic Foveated NL-means share exactly the same orientation preference. Patch similarity here is measured by (53) which mimics the sensitivity of a single receptor placed at a pixel \bar{u} within the patch. The average standardized MSE of the denoised images is reported as a function of the angular position ($\angle\bar{u}$) and orientation ($\angle\bar{u} + \theta$), which in fact characterize the blur kernel $v_{\bar{u}}$. The best denoising performance is achieved when the angular position and orientation coincide, i.e. when the blur kernels are oriented radially towards the patch center. This plot is strikingly similar to Figure 3A in (Freeman et al., 2011), which investigated the orientation preference of the HVS by means of fMRI recordings.

which corresponds to measuring the patch distance as the difference between two foveated patches in the specific pixel \bar{u} , where the blur kernel yielding foveation was oriented towards $\angle\bar{u} + \theta$ ¹⁰. The scaling factor $\|\mathbf{k}\|_1/k(\bar{u})$ guarantees that (53) satisfies the compatibility requirement (20).

Of course, a lot of information is ignored when comparing patches using (53), and we expect a substantial loss in the quality of the denoised image if we replace d^{FOV} or d^{WIN} (6) with $d_{\rho,\theta,\bar{u}}$. Nevertheless, the denoising performance achieved when using (53) indicates how a single receptor with *angular position* $\angle\bar{u}$ and *orientation* $\angle\bar{u} + \theta$ is effective at assessing the nonlocal self-similarity. Hence, through (53), we can determine, for any given angular position, whether there exist any particular receptor orientation that is better than others at measuring patch similarity.

To this purpose, we take the *Hill, Lena, Boats, Barbara, Man, Peppers, Cameraman*, and *MIT* images, generate six noisy realizations of each image with $\sigma = 30$, and process each such realization by Foveated NL-means using the special distance (53) with $\rho = 6$ and varying θ and \bar{u} in the following sets:

$$\begin{aligned} \theta &\in \{k\pi/10, k = -4, -3, -2, -1, 0, 1, 2, 3, 4, 5\}, \\ \bar{u} &\in \{u \in U, \text{ s.t. } \|u\|_{\infty} = 5\}. \end{aligned} \quad (54)$$

The set (54) corresponds to the peripheral annular region of interest from (Freeman et al., 2011). We then standardize the MSE separately for each individual angular position $\angle\bar{u}$, and report this standardized MSE values as function of angular position and orientation in Figure 22, where a fitting surface is also displayed for the sake of visualization. The standardization is essential in order for the orientation preference to emerge from the data: without standardization, the angular

¹⁰ In fact, $\mathcal{F}_{\rho,\theta}[z, x_1](\bar{u})$ is proportional to the value in \bar{u} of the convolution between z and the blur kernel $v_{\bar{u}}$, and this latter has a main axis oriented as $\angle\bar{u} + \theta$

positions corresponding to the most pronounced directions in the image content (typically vertical and horizontal, see, e.g., van der Schaaf and van Hateren, 1996; Baddeley, 1997; Torralba and Oliva, 2003) would otherwise dominate, making data from different angular positions inconsistent one with each other.

Figure 22 indicates a clear analogy with the orientation preference in the HVS, since the best performance is achieved when angular position and orientation of the receptor coincide, i.e. $\theta = 0$. Thus the best receptor in assessing the nonlocal self-similarity is often the one following a radial direction, and radial foveation operators are built upon such blur kernels. Remarkably, Figure 22 is actually very close to Figure 3A in (Freeman et al., 2011), which displays the preferred orientations against the preferred angular positions determined from the fMRI experiment, with a measure of the SNR of the cortical responses to oriented stimuli. This is a striking relation between anisotropic foveation in digital images and features of the HVS, which well substantiates the use of radial foveation operators. Most importantly, it uncovers an apparent functional justification for the orientation bias, exclusively leveraging the nonlocal statistics of natural images.

11 Conclusions

Exploiting foveation, an essential principle of the HVS, we have proposed a direct modification of the way self-similarity is assessed within nonlocal imaging filters. Instead of measuring the patch similarity through a windowed Euclidean distance, we employ foveation operators, and measure the Euclidean distance between foveated patches. Thanks to a general, constrained, design of foveation operators, we have installed a simple – yet fully explicit – connection between traditional windowed self-similarity and the foveated self-similarity. We have also constructed families of isotropic and anisotropic operators characterized by possibly elongated and oriented blur kernels. Through extensive experiments, mainly within a nonlocal image denoising framework, we have shown that foveated self-similarity is a far more effective regularization prior for natural images than the conventional windowed self-similarity. We have also pointed out that radial anisotropic foveation operators exhibit an enhanced ability at recovering edges and structures. Strikingly, such radial design is consistent with HVS features, and this study might provide a functional justification in terms of natural image statistics for the radial bias of HVS.

Ongoing research considers foveated self-similarity within multiscale nonlocal transform-domain algorithms. Furthermore, we find particularly relevant examining the connections between foveation and transsaccadic integration (Demeyer et al., 2009; Herwig and Schneider, 2014; Krishna et al., 2014) and exploring their joint applications in image filtering.

References

- A. Alahi, R. Ortiz, and P. Vanderghenst. FREAK: Fast retina keypoint. In *Proc. 2012 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 510–517. IEEE, 2012. <http://dx.doi.org/10.1109/CVPR.2012.6247715>.
- S. K. Alexander, E. R. Vrscay, and S. Tsurumi. A simple, general model for the affine self-similarity of images. In A. Campilho and M. Kamel, editors, *Image Analysis and Recognition*, volume 5112 of *Lecture Notes in Computer Science*, pages 192–203. Springer Berlin Heidelberg, 2008. http://dx.doi.org/10.1007/978-3-540-69812-8_19.
- P. Arias, G. Facciolo, V. Caselles, and G. Sapiro. A variational framework for exemplar-based image inpainting. *Int. J. Computer Vision*, 93(3):319–347, 2011. <http://dx.doi.org/10.1007/s11263-010-0418-7>.
- S. Awate and R. Whitaker. Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2005)*, volume 2, pages 44–51, 2005. <http://dx.doi.org/10.1109/CVPR.2005.176>.
- R. Baddeley. The correlational structure of natural images and the calibration of spatial representations. *Cognitive Science*, 21(3): 351–372, 1997. http://dx.doi.org/10.1207/s15516709cog2103_4.
- B. Balas, L. Nakano, and R. Rosenholtz. A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, 9(12), 2009. <http://dx.doi.org/10.1167/9.12.13>.
- M. F. Barnsley. *Fractals everywhere*. Academic Press, second edition, 1993. <http://www.sciencedirect.com/science/book/9780120790616>.
- A. Basu and K. Wiebe. Enhancing videoconferencing using spatially varying sensing. *IEEE Trans. Systems, Man and Cybernetics, Part A: Systems and Humans*, 28(2):137–148, 1998. <http://dx.doi.org/10.1109/3468.661143>.
- J. Bigun and G. H. Granlund. Central symmetry modelling. In *Proc. EUSIPCO-86, Third European Signal Processing Conference*, pages 883–886, Sept. 1986. <http://www2.hh.se/staff/josef/public/publications/bigun86haag.pdf>.
- J. R. Brockmole and D. E. Irwin. Eye movements and the integration of visual memory and visual perception. *Perception & Psychophysics*, 67(3):495–512, 2005. <http://dx.doi.org/10.3758/BF03193327>.
- A. Buades, B. Coll, and J.-M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490, 2005. <http://dx.doi.org/10.1137/040616024>.
- A. Buades, B. Coll, and J.-M. Morel. Nonlocal image and movie denoising. *Int. J. Comput. Vision*, 76(2):123–139, Feb. 2008. <http://dx.doi.org/10.1007/s11263-007-0052-1>.
- A. Buades, B. Coll, J.-M. Morel, and C. Sbert. Self-similarity driven color demosaicking. *IEEE Trans. Image Process.*, 18(6):1192–1202, 2009. <http://dx.doi.org/10.1109/TIP.2009.2017171>.
- A. Buades, B. Coll, and J.-M. Morel. Image denoising methods. a new nonlocal principle. *SIAM Rev.*, 52(1):113–147, Feb. 2010. <http://dx.doi.org/10.1137/090773908>.
- A. Buades, B. Coll, and J.-M. Morel. Non-Local Means Denoising. *Image Processing On Line*, 2011, 2011. http://dx.doi.org/10.5201/ipol.2011.bcm_nlm.
- E.-C. Chang, S. Mallat, and C. Yap. Wavelet foveation. *Appl. Comput. Harmon. Anal.*, 9(3):312–335, 2000. <http://dx.doi.org/10.1006/acha.2000.0324>.
- P. Chatterjee and P. Milanfar. A generalization of non-local means via kernel regression. In *Proc. SPIE Computational Imaging VI*, volume 6814, pages 68140P–68140P–9, 2008. <http://dx.doi.org/10.1117/12.778615>.
- P. Chatterjee and P. Milanfar. Is denoising dead? *IEEE Trans. Image Process.*, 19(4):895–911, Apr. 2010. <http://dx.doi.org/10.1109/TIP.2009.2037087>.
- G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu. Epigraphical splitting for solving constrained convex formulations of inverse problems with proximal tools. *ArXiv e-prints*, Mar. 2014. <http://arxiv.org/abs/1210.5844v3>.
- G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu. Epigraphical projection and proximal tools for solving constrained convex optimization problems. *Signal, Image and Video Processing*, pages 1–13, 2014a. <http://dx.doi.org/10.1007/s11760-014-0664-1>.
- G. Chierchia, N. Pustelnik, B. Pesquet-Popescu, and J.-C. Pesquet. A nonlocal structure tensor-based approach for multicomponent image recovery problems. *IEEE Trans. Image Process.*, 23(12):5531–5544, Dec. 2014b. <http://dx.doi.org/10.1109/TIP.2014.2364141>.
- D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 2843–2851. Curran Associates, Inc., 2012. <http://papers.nips.cc/paper/4741-deep-neural-networks-segment-neuronal-membranes-in-electron-microscopy-images.pdf>.
- A. R. Conn, K. Scheinberg, and L. Vicente. *Introduction to Derivative-Free Optimization*. MPS-SIAM Series on Optimization, 2009. <http://dx.doi.org/10.1137/1.9780898718768>.
- C. Curcio, K. Sloan, R. Kalina, and A. Hendrickson. Human photoreceptor topography. *J. Comparative Neurology*, 292:497–523, 1990. <http://dx.doi.org/10.1002/cne.902920402>.
- K. Dabov, A. Foi, and K. Egiazarian. Video denoising by sparse 3D transform-domain collaborative filtering. In *Proc. 15th European Signal Process. Conf., EUSIPCO 2007*, 2007a. http://www.cs.tut.fi/~foi/GCF-BM3D/VBM3D-EUSIPCO_2007.pdf.
- K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):2080–2095, Aug. 2007b. <http://dx.doi.org/10.1109/TIP.2007.901238>.
- K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image restoration by sparse 3D transform-domain collaborative filtering. In *Proc. SPIE Electronic Imaging*, volume 6812, pages 681207–1 – 681207–12, 2008. <http://dx.doi.org/10.1117/12.766355>.
- K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. BM3D image denoising with shape-adaptive principal component analysis. In R. Gribonval, editor, *SPARS'09 - Signal Processing with Adaptive Sparse Structured Representations*, Saint Malo, France, 2009. INRIA Rennes - Bretagne Atlantique. <http://hal.inria.fr/inria-00369582>.
- S. Daly, J. Ribas-Corbera, and K. Matthews. As plain as the noise on your face: Adaptive video compression using face detection and visual eccentricity models. *Journal of Electronic Imaging*, 10(1):30–46, 2001. <http://dx.doi.org/10.1117/1.1333679>.
- A. Danielyan, A. Foi, V. Katkovnik, and K. Egiazarian. Image and video super-resolution via spatially adaptive block-matching filtering. In J. Astola, K. Egiazarian, and V. Katkovnik, editors, *Proc. Int. Workshop Local and Non-Local Approximation in Image Processing (LNLA)*, 2008. <http://www.eurasip.org/Proceedings/Ext/LNLA2008/papers/cr1026.pdf>.

- A. Danielyan, V. Katkovnik, and K. Egiazarian. BM3D frames and variational image deblurring. *IEEE Trans. Image Process.*, 21(4): 1715–1728, 2012. <http://dx.doi.org/10.1109/TIP.2011.2176954>.
- J. Darbon, A. Cunha, T. Chan, S. Osher, and G. Jensen. Fast non-local filtering applied to electron cryomicroscopy. In *IEEE Int. Symposium on Biomedical Imaging (ISBI): From Nano to Macro*, pages 1331–1334, May 2008. <http://dx.doi.org/10.1109/ISBI.2008.4541250>.
- J. S. De Bonet. Noise reduction through detection of signal redundancy. Rethinking artificial intelligence, MIT AI Lab, 1997. http://www.ai.mit.edu/projects/lv/Abstracts/zz-jsd_3.ps.
- C.-A. Deledalle, L. Denis, and F. Tupin. How to compare noisy patches? Patch similarity beyond Gaussian noise. *Int. J. Computer Vision*, 99(1):86–102, 2012. <http://dx.doi.org/10.1007/s11263-012-0519-6>.
- M. Demeyer, P. De Graef, J. Wagemans, and K. Verfaillie. Transsaccadic identification of highly similar artificial shapes. *J. Vision*, 9(4), 2009. <http://dx.doi.org/10.1167/9.4.28>.
- G. Desbordes. *Vision in the Presence of Fixational Eye Movements: Insights from Psychophysics and Neural Modeling*. PhD thesis, Boston University, 2007. http://cns-web.bu.edu/~gdesbord/Desbordes_PhD_dissertation.pdf.
- B. Dong, J. Ye, S. Osher, and I. Dinov. Level set based nonlocal surface restoration. *Multiscale Modeling & Simulation*, 7(2):589–598, 2008. <http://dx.doi.org/10.1137/070705556>.
- K. Donner and S. Hemilä. Modelling the effect of microsaccades on retinal responses to stationary contrast patterns. *Vision Res.*, 47(9):1166–1177, 2007. <http://dx.doi.org/10.1016/j.visres.2006.11.024>.
- R. O. Duncan and G. M. Boynton. Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, 38(4):659–671, 2003. [http://dx.doi.org/10.1016/S0896-6273\(03\)00265-4](http://dx.doi.org/10.1016/S0896-6273(03)00265-4).
- M. Ebrahimi and E. R. Vrscay. Solving the inverse problem of image zooming using “self-examples”. In M. Kamel and A. Campilho, editors, *Image Analysis and Recognition*, volume 4633 of *Lecture Notes in Computer Science*, pages 117–130. Springer Berlin Heidelberg, 2007. http://dx.doi.org/10.1007/978-3-540-74260-9_11.
- M. Ebrahimi and E. R. Vrscay. Examining the role of scale in the context of the non-local-means filter. In *Proc. 5th Int. Conf. Image Analysis and Recognition (ICIAR '08)*, pages 170–181, Berlin, Heidelberg, 2008a. Springer-Verlag. http://dx.doi.org/10.1007/978-3-540-69812-8_17.
- M. Ebrahimi and E. R. Vrscay. Multi-frame super-resolution with no explicit motion estimation. In *Proc. 2008 Int. Conf. Image Processing, Computer Vision, and Pattern Recognition (ICCV2008)*, 2008b. http://individual.utoronto.ca/ebrahimi/papers/ebvr_iccv08.pdf.
- M. P. Eckstein. Visual search: A retrospective. *J. Vision*, 11(5), 2011. <http://dx.doi.org/10.1167/11.5.14>.
- A. Efros and T. Leung. Texture synthesis by non-parametric sampling. In *Proc. 7th IEEE Int. Conf. Computer Vision (ICCV)*, volume 2, pages 1033–1038, 1999. <http://dx.doi.org/10.1109/ICCV.1999.790383>.
- M. Elad and D. Datsenko. Example-based regularization deployed to super-resolution reconstruction of a single image. *Comput. J.*, 52(1):15–30, Jan. 2009. <http://dx.doi.org/10.1093/comjnl/bxm008>.
- R. Engbert and R. Kliegl. Microsaccades keep the eyes’ balance during fixation. *Psychological Science*, 15(6):431–431, 2004. <http://dx.doi.org/10.1111/j.0956-7976.2004.00697.x>.
- R. Etienne-Cummings, J. Van der Spiegel, P. Mueller, and M.-Z. Zhang. A foveated silicon retina for two-dimensional tracking. *IEEE Trans. Circuits and Systems II: Analog and Digital Signal Processing*, 47(6):504–517, 2000. <http://dx.doi.org/10.1109/82.847066>.
- A. Foi and G. Boracchi. Foveated self-similarity in nonlocal image filtering. In *SPIE Human Vision and Electronic Imaging XVII*, volume 8291, Burlingame, California, USA, 2012. SPIE. <http://dx.doi.org/10.1117/12.912217>.
- A. Foi and G. Boracchi. Anisotropically foveated nonlocal image denoising. In *Proc. IEEE Int. Conf. Image Process. (ICIP 2013)*, pages 464–468, September 15–18 2013a. <http://dx.doi.org/10.1109/ICIP.2013.6738096>.
- A. Foi and G. Boracchi. Anisotropic foveated self-similarity. In *Proc. of SPARS 2013, Signal Processing with Adaptive Sparse Structured Representations*, page 1, July 8–11 2013b. <http://spars2013.epfl.ch/data/uploaded/flash/Paper%20158.pdf>.
- A. Foi and G. Boracchi. Nonlocal foveated principal components. In *Proc. 2014 IEEE Workshop on Statistical Signal Processing (SSP 2014)*, pages 145–148, 2014. <http://dx.doi.org/10.1109/SSP.2014.6884596>.
- J. Freeman and E. P. Simoncelli. Metamers of the ventral stream. *Nature Neuroscience*, 14(9):1195–1201, 2011. <http://dx.doi.org/10.1038/nn.2889>.
- J. Freeman, G. J. J. Brouwer, D. J. Heeger, and E. P. Merriam. Orientation decoding depends on maps, not columns. *J. Neurosci.*, 31(13):4792–4804, 2011. <http://dx.doi.org/10.1523/JNEUROSCI.5160-10.2011>.
- W. Freeman, T. Jones, and E. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2): 56–65, 2002. <http://dx.doi.org/10.1109/38.988747>.
- J. Froment. Parameter-Free Fast Pixelwise Non-Local Means Denoising. *Image Processing On Line*, 4:300–326, 2014. <http://dx.doi.org/10.5201/ipol.2014.120>.
- D. F. Garway-Heath, J. Caprioli, F. W. Fitzke, and R. A. Hitchings. Scaling the hill of vision: the physiological relationship between light sensitivity and ganglion cell numbers. *Investigative Ophthalmology and Visual Science*, 41(7):1774–1782, 2000. <http://iovs.arvojournals.org/article.aspx?articleid=2123396>.
- W. Geisler, J. Perry, B. Super, and D. Gallogly. Edge co-occurrence in natural images predicts contour grouping performance. *Vision research*, 41(6):711–724, 2001. [http://dx.doi.org/10.1016/S0042-6989\(00\)00277-7](http://dx.doi.org/10.1016/S0042-6989(00)00277-7).
- W. S. Geisler and J. S. Perry. Real-time foveated multiresolution system for low-bandwidth video communication. In *Human Vision and Electronic Imaging, SPIE Proceedings*, pages 294–305, 1998. <http://dx.doi.org/10.1117/12.320120>.
- M. Ghazel, G. H. Freeman, and E. R. Vrscay. Fractal image denoising. *IEEE Trans. Image Process.*, 12(12):1560–1578, 2003. <http://dx.doi.org/10.1109/TIP.2003.818038>.
- G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *Multiscale Modeling & Simulation*, 7(3):1005–1028, 2009. <http://dx.doi.org/10.1137/070698592>.
- S. Grewenig, S. Zimmer, and J. Weickert. Rotationally invariant similarity measures for nonlocal image denoising. *J. Visual Communication and Image Representation*, 22(2):117–130, Feb. 2011. <http://dx.doi.org/10.1016/j.jvcir.2010.11.001>.
- M. Haloi. Improved microaneurysm detection using deep neural networks. *ArXiv e-prints*, 2015. <http://arxiv.org/abs/1505.04424>.
- A. Herwig and W. X. Schneider. Predicting object features across saccades: Evidence from object recognition and visual search. *Journal of Experimental Psychology: General*, 2014. <http://dx.doi.org/10.1037/xap0000001>.

- doi.org/10.1037/a0036781.
- A. Jacquin. Image coding based on a fractal theory of iterated contractive image transformations. *IEEE Trans. Image Process.*, 1(1):18–30, 1992. <http://dx.doi.org/10.1109/83.128028>.
- J. A. M. Jennings and W. Charman. Analytic approximation of the off-axis modulation transfer function of the eye. *Vision Res.*, 37(6):697–704, 1997. [http://dx.doi.org/10.1016/S0042-6989\(96\)00238-6](http://dx.doi.org/10.1016/S0042-6989(96)00238-6).
- Z. Ji, Q. Chen, Q.-S. Sun, and D.-S. Xia. A moment-based nonlocal-means algorithm for image denoising. *Information Processing Letters*, 109:1238–1244, 2009. <http://dx.doi.org/10.1016/j.ipl.2009.09.007>.
- Q. Jin, I. Grama, and Q. Liu. Removing Gaussian noise by optimization of weights in non-local means. *arXiv preprint*, 2011. <http://arxiv.org/abs/1109.5640>.
- C. Joselevitch. Human retinal circuitry and physiology. *Psychology & Neuroscience*, 1:141–165, 12 2008. <http://dx.doi.org/10.3922/j.psns.2008.2.008>.
- S. Karlsson and J. Bigun. Synthesis and detection of log-spiral codes. In *Swedish Symposium on Image Analysis (SSBA2011)*, page 4, Mar. 2011. <http://www2.hh.se/staff/josef/public/publications/karlsson11linkoping.pdf>.
- V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola. From local kernel to nonlocal multiple-model image denoising. *Int. J. Comput. Vision*, 86(1):1–32, 2010. <http://dx.doi.org/10.1007/s11263-009-0272-7>.
- C. Kervrann and J. Boulanger. Optimal spatial adaptation for patch-based image denoising. *IEEE Trans. Image Process.*, 15(10):2866–2878, Oct. 2006. <http://dx.doi.org/10.1109/TIP.2006.877529>.
- C. Kervrann, J. Boulanger, and P. Coupé. Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In *Proc. 1st Int. Conf. Scale Space and Variational Methods in Computer Vision*, SSMV'07, pages 520–532, Berlin, Heidelberg, 2007. Springer-Verlag. <http://dl.acm.org/citation.cfm?id=1767926.1767978>.
- S. Kindermann, S. Osher, and P. Jones. Deblurring and denoising of images by nonlocal functionals. *Multiscale Modeling & Simulation*, 4(4):1091–1115, 2005. <http://dx.doi.org/10.1137/05062249>.
- P. Kortum and W. Geisler. Implementation of a foveated image coding system for image bandwidth reduction. In *Human Vision and Electronic Imaging, SPIE Proceedings*, pages 350–360, 1996. <http://dx.doi.org/10.1117/12.238732>.
- E. Kowler. Eye movements: The past 25 years. *Vision Res.*, 51(13):1457–1483, 2011. <http://dx.doi.org/10.1016/j.visres.2010.12.014>.
- B. S. Krishna, A. E. Ipata, J. W. Bisley, J. Gottlieb, and M. E. Goldberg. Extrafoveal preview benefit during free-viewing visual search in the monkey. *Journal of Vision*, 14(1):6, 2014. <http://dx.doi.org/10.1167/13.6.5>.
- D. La Torre, E. R. Vrscay, M. Ebrahimi, and M. F. Barnsley. Measure-valued images, associated fractal transforms, and the affine self-similarity of images. *SIAM J. Img. Sci.*, 2(2):470–507, Apr. 2009. <http://dx.doi.org/10.1137/070696763>.
- M. Lebrun, A. Buades, and J.-M. Morel. A nonlocal bayesian image denoising algorithm. *SIAM Journal on Imaging Sciences*, 6(3):1665–1688, 2013. <http://dx.doi.org/10.1137/120874989>.
- S. Lee, M. Pattichis, and A. Bovik. Foveated video compression with optimal rate control. *IEEE Trans. Image Process.*, 10(7):977–992, 2001. <http://dx.doi.org/10.1109/83.931092>.
- W. Levick and L. Thibos. Analysis of orientation bias in cat retina. *J. Physiol.*, 329:243, 1982. <http://dx.doi.org/10.1113/jphysiol.1982.sp014301>.
- A. Levin, B. Nadler, F. Durand, and W. T. Freeman. Patch complexity, finite pixel correlations and optimal denoising. In *Proc. 12th European Conf. Computer Vision (ECCV)*, volume 5, pages 73–86, 2012. http://dx.doi.org/10.1007/978-3-642-33715-4_6.
- Y. Li and D. Huttenlocher. Sparse long-range random field and its application to image denoising. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision - ECCV 2008*, volume 5304 of *Lecture Notes in Computer Science*, pages 344–357. Springer Berlin Heidelberg, 2008. http://dx.doi.org/10.1007/978-3-540-88690-7_26.
- Y. Lou, P. Favaro, S. Soatto, and A. Bertozzi. Nonlocal similarity image filtering. In *Proc. 15th Int. Conf. Image Analysis and Processing, ICIAP '09*, pages 62–71, Berlin, Heidelberg, 2009. http://dx.doi.org/10.1007/978-3-642-04146-4_9.
- C. Louchet and L. Moisan. Total variation as a local filter. *SIAM Journal on Imaging Sciences*, 4(2):651–694, 2011. <http://dx.doi.org/10.1137/100785855>.
- D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60:91–110, Nov. 2004. <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>.
- S. Lyu and E. Simoncelli. Modeling multiscale subbands of photographic images with fields of gaussian scale mixtures. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 31(4):693–706, 2009. <http://dx.doi.org/10.1109/TPAMI.2008.107>.
- M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian. Video denoising, blocking and enhancement through separable 4-D nonlocal spatiotemporal transforms. *IEEE Trans. Image Process.*, 21(9):3952–3966, Sept. 2012. <http://dx.doi.org/10.1109/TIP.2012.2199324>.
- M. Mahmoudi and G. Sapiro. Fast image and video denoising via nonlocal means of similar neighborhoods. *IEEE Signal Processing Letters*, 12(12):839–842, 2005. <http://dx.doi.org/10.1109/LSP.2005.859509>.
- A. Maleki, M. Narayan, and R. G. Baraniuk. Anisotropic nonlocal means. *Appl. Comput. Harmon. Anal.*, 35(3):452–482, Nov. 2013. <http://arxiv.org/abs/1112.0311>.
- J. V. Manjon-Herrera and A. Buades. Non-local means filter, Matlab code. Matlab Central File Exchange, 2008. <http://www.mathworks.com/matlabcentral/fileexchange/13176-non-local-means-filter>.
- S. Martinez-Conde, J. Otero-Millan, and S. L. Macknik. The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nature Reviews Neuroscience*, 14(2):83–96, 2013. <http://dx.doi.org/10.1038/nrn3405>.
- M. B. McCamy, J. Otero-Millan, S. L. Macknik, Y. Yang, X. G. Troncoso, S. M. Baer, S. M. Crook, and S. Martinez-Conde. Microsaccadic efficacy and contribution to foveal and peripheral vision. *J. Neurosci.*, 32(27):9194–9204, 2012. <http://dx.doi.org/10.1523/JNEUROSCI.0515-12.2012>.
- P. Milanfar. A tour of modern image filtering: New insights and methods, both practical and theoretical. *IEEE Signal Processing Magazine*, 30(1):106–128, 2013. <http://dx.doi.org/10.1109/MSP.2011.2179329>.
- J. P. Monaco, A. C. Bovik, and L. K. Cormack. Active, foveated, uncalibrated stereovision. *Int. J. Comput. Vision*, 85(2):192–207, Nov. 2009. <http://dx.doi.org/10.1007/s11263-009-0230-4>.
- I. Mosseri, M. Zontak, and M. Irani. Combining the power of internal and external denoising. In *2013 IEEE Int. Conf. Computational Photography (ICCP)*, pages 1–9. IEEE, 2013. <http://>

- [//dx.doi.org/10.1109/ICCPHOT.2013.6528298](http://dx.doi.org/10.1109/ICCPHOT.2013.6528298).
- A. Olmedo-Payá, A. Martínez-Álvarez, S. Cuenca-Asensi, J. Ferrández-Vicente, and E. Fernández. Modeling the effect of fixational eye movements in natural scenes. In J. Ferrández Vicente, J. Álvarez Sánchez, F. Paz López, and F. Toledo Moreo, editors, *Natural and Artificial Models in Computation and Biology*, volume 7930 of *Lecture Notes in Computer Science*, pages 332–341. Springer Berlin Heidelberg, 2013. http://dx.doi.org/10.1007/978-3-642-38637-4_34.
- J. Orchard, M. Ebrahimi, and A. Wong. Efficient nonlocal-means denoising using the SVD. In *Proc. 15th IEEE Int. Conf. Image Process. (ICIP 2008)*, pages 1732–1735, Oct. 2008. <http://dx.doi.org/10.1109/ICIP.2008.4712109>.
- S. Osindero, M. Welling, and G. E. Hinton. Topographic product models applied to natural scene statistics. *Neural Computation*, 18(2):381–414, 2006. <http://dx.doi.org/10.1162/089976606775093936>.
- G. Peyré. Sparse modeling of textures. *J. Mathematical Imaging and Vision*, 34(1):17–31, 2009a. <http://dx.doi.org/10.1007/s10851-008-0120-3>.
- G. Peyré. Manifold models for signals and images. *Computer Vision and Image Understanding*, 113(2):249–260, 2009b. <http://dx.doi.org/10.1016/j.cviu.2008.09.003>.
- M. Poletti, C. Listorti, and M. Rucci. Microscopic eye movements compensate for nonhomogeneous vision within the fovea. *Current Biology*, 23(17):1691–1695, 2013. <http://dx.doi.org/10.1016/j.cub.2013.07.007>.
- S. Postec. *Quelques remarques en débruitage des images liées à des propriétés de similarité, de régularité et de parcimonie*. PhD thesis, Université de Bretagne-Sud, 2012. <http://web.univ-ubs.fr/lmam/postec/these.pdf>.
- M. Protter, M. Elad, H. Takeda, and P. Milanfar. Generalizing the nonlocal-means to super-resolution reconstruction. *IEEE Trans. Image Process.*, 18(1):36–51, 2009. <http://dx.doi.org/10.1109/TIP.2008.2008067>.
- M. Ranzato, V. Mnih, J. Susskind, and G. Hinton. Modeling natural images using gated MRFs. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 35(9):2206–2222, 2013. <http://dx.doi.org/10.1109/TPAMI.2013.29>.
- R. Rosenholtz, J. Huang, A. Raj, B. J. Balas, and L. Ilie. A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, 12(4), 2012. <http://dx.doi.org/10.1167/12.4.14>.
- G. Rosman, A. Dubrovina, and R. Kimmel. Patch-collaborative spectral point-cloud denoising. *Computer Graphics Forum*, 2013. <http://dx.doi.org/10.1111/cgf.12139>.
- S. Roth and M. Black. Fields of experts: a framework for learning image priors. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2005)*, volume 2, pages 860–867 vol. 2, 2005. <http://dx.doi.org/10.1109/CVPR.2005.160>.
- M. Rucci, R. Iovin, M. Poletti, and F. Santini. Miniature eye movements enhance fine spatial detail. *Nature*, 447(7146):852–855, 2007. <http://dx.doi.org/10.1038/nature05866>.
- J. Salmon. On two parameters for denoising with non-local means. *IEEE Signal Process. Letters*, 17(3):269–272, Mar. 2010. <http://dx.doi.org/10.1109/LSP.2009.2038954>.
- Y. Sasaki, R. Rajimehr, B. Kim, L. Ekstrom, W. Vanduffel, and R. Tootell. The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. *Neuron*, 51(5):661–670, Sept. 2006. <http://dx.doi.org/10.1016/j.neuron.2006.07.021>.
- M. Sigman, G. A. Cecchi, C. D. Gilbert, and M. O. Magnasco. On a common circle: natural scenes and gestalt rules. *Proceedings of the National Academy of Sciences*, 98(4):1935–1940, 2001. <http://dx.doi.org/10.1073/pnas.98.4.1935>.
- H. Strasburger, I. Rentschler, and M. Jüttner. Peripheral vision and pattern recognition: A review. *J. Vision*, 11(5), 2011. <http://dx.doi.org/10.1167/11.5.13>.
- C. Sutour, C.-A. Deledalle, and J.-F. Aujol. Adaptive regularization of the NL-means: Application to image and video denoising. *IEEE Trans. Image Process.*, 23(8):3506–3521, Aug. 2014. <http://dx.doi.org/10.1109/TIP.2014.2329448>.
- T. Tasdizen. Principal neighborhood dictionaries for nonlocal means image denoising. *IEEE Trans. Image Process.*, 18(12):2649–2660, Dec. 2009. <http://dx.doi.org/10.1109/TIP.2009.2028259>.
- T. Thaipanich, B. Oh, P.-H. Wu, D. Xu, and C.-C. Kuo. Improved image denoising with adaptive nonlocal means (ANL-means) algorithm. *IEEE Trans. Consum. Electr.*, 56(4):2623–2630, Nov. 2010. <http://dx.doi.org/10.1109/TCE.2010.5681149>.
- A. Toet and D. M. Levi. The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Res.*, 32(7):1349–1357, 1992. [http://dx.doi.org/10.1016/0042-6989\(92\)90227-A](http://dx.doi.org/10.1016/0042-6989(92)90227-A).
- C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. 6th Int. Conf. Computer Vision*, pages 839–846, 1998. <http://dx.doi.org/10.1109/ICCV.1998.710815>.
- F. Tong and Z.-N. Li. Reciprocal-wedge transform for space-variant sensing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(5):500–511, May 1995. <http://dx.doi.org/10.1109/34.391393>.
- A. Torralba and A. Oliva. Statistics of natural image categories. *Network: Comput. Neural Syst*, 14:391–412, 2003.
- USC-SIPI. University of Southern California Signal and Image Processing Institute image database. <http://sipi.usc.edu/database/>.
- A. van der Schaaf and J. van Hateren. Modelling the power spectra of natural images: statistics and information. *Vision Res.*, 36(17):2759–2770, 1996. [http://dx.doi.org/10.1016/0042-6989\(96\)00002-8](http://dx.doi.org/10.1016/0042-6989(96)00002-8).
- R. S. Wallace, P.-W. Ong, B. B. Bederson, and E. L. Schwartz. Space variant image processing. *Int. J. Computer Vision*, 13(1):71–90, 1994. <http://dx.doi.org/10.1007/BF01420796>.
- B. A. Wandell. *Foundations of Vision*. Sinauer Associates, Inc., 1995.
- J. Wang, Y. Guo, Y. Ying, Y. Liu, and Q. Peng. Fast non-local algorithm for image denoising. In *Proc. IEEE Int. Conf. Image Process. (ICIP 2006)*, pages 1429–1432, Oct. 2006. <http://dx.doi.org/10.1109/ICIP.2006.312698>.
- Z. Wang and A. Bovik. Embedded foveation image coding. *IEEE Trans. Image Process.*, 10(10):1397–1410, 2001. <http://dx.doi.org/10.1109/83.951527>.
- Z. Wang and A. Bovik. Foveated image and video coding. In H. R. Wu and K. R. Rao, editors, *Digital Video, Image Quality and Perceptual Coding*, pages 431–457. CRC Press, 2005. <http://dx.doi.org/10.1201/9781420027822.ch14>.
- Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error measurement to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, Apr. 2004. <http://dx.doi.org/10.1109/TIP.2003.819861>.
- J. Wei and Z.-N. Li. Efficient disparity-based gaze control with foveate wavelet transform. In *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, volume 2, pages 866–871 vol.2, 1998. <http://dx.doi.org/10.1109/IROS.1998.727309>.
- L.-Y. Wei and M. Levoy. Fast texture synthesis using tree-structured vector quantization. In *Proc. 27th Conf. Computer Graphics and Interactive Techniques, SIGGRAPH '00*, pages 479–488, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co. <http://dx.doi.org/10.1145/344779.345009>.

- C. F. Weiman and G. Chaikin. Logarithmic spiral grids for image processing and display. *Computer Graphics and Image Processing*, 11(3):197–226, 1979. [http://dx.doi.org/10.1016/0146-664X\(79\)90089-3](http://dx.doi.org/10.1016/0146-664X(79)90089-3).
- Y. Weiss and W. Freeman. What makes a good model of natural images? In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2007)*, pages 1–8, 2007. <http://dx.doi.org/10.1109/CVPR.2007.383092>.
- T. Wertheim. Über die indirekte sehschärfe. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 7:172–187, 1894. <http://echo.mpiwg-berlin.mpg.de/MPIWG:YSKVNH9>.
- D. R. Williams, P. Artal, R. Navarro, M. J. McMahon, and D. H. Brainard. Off-axis optical quality and retinal sampling in the human eye. *Vision Res.*, 36(8):1103–1114, 1996. [http://dx.doi.org/10.1016/0042-6989\(95\)00182-4](http://dx.doi.org/10.1016/0042-6989(95)00182-4).
- R. Wodnicki, G. Roberts, and M. Levine. A foveated image sensor in standard CMOS technology. In *Proc. IEEE Conf. Custom Integrated Circuits*, pages 357–360, 1995. <http://dx.doi.org/10.1109/CICC.1995.518202>.
- L. Yaroslavsky. *Digital Picture Processing, An Introduction*. Springer Verlag, Berlin, Heidelberg, 1985. <http://dx.doi.org/10.1007/978-3-642-81929-2>.
- D. Zhang and Z. Wang. Image information restoration based on long-range correlation. *IEEE Trans. Circuits and Systems for Video Technology*, 12(5):331–341, 2002. <http://dx.doi.org/10.1109/TCSVT.2002.1003472>.

page left intentionally blank

supplementary material begins on next page

Supplementary Material

Foveated Nonlocal Self-Similarity

Alessandro Foi, Giacomo Boracchi

Contents:

This supplementary material is organized as follows:

- Section [Suppl.1](#) presents the considered image datasets;
- Section [Suppl.2](#) concerns self-map operators and compares the denoising performance of Foveated NL-means based on either self-maps or non self-map foveation operators;
- Section [Suppl.3](#) provides additional results on anisotropic foveation operators;
- Section [Suppl.4](#) investigates chiral foveation operator;
- Section [Suppl.5](#) provides additional experiments on texture denoising by means of Foveated NL-means;
- Section [Suppl.6](#) compares the standard NL-means against the anisotropic Foveated NL-means in terms of method noise;
- Section [Suppl.7](#) presents an experiment to stress the importance of the constrained design of foveation operators. In particular, we show that operators that violate the design constraints are less effective at estimating the nonlocal self-similarity.

Suppl.1 Test images

In our experiments we use three different sets of grayscale images. The first set consists of six standard test images of size 512×512 , namely *Hill*, *Lena*, *Boats*, *Barbara*, *Man*, *Peppers* (see [Figure Suppl.1](#)), which were selected to assess the foveated self-similarity on natural images featuring complex structures. The second set is composed of two images of size 256×256 dominated by sharp edges, namely *Cameraman* and *MIT* (see [Figure Suppl.2](#)), through which we can better expose the link between denoising performance at edges and anisotropy of the foveation operator. Finally, the third set comprises six texture image of size 512×512 from the Brodatz database ([USC-SIPI](#)) (see [Figure Suppl.3](#)); these images, because of their sporadic patterns and random details, are particularly challenging for nonlocal algorithms and have thus been separately considered to compare the foveated and the windowed distance in this difficult scenario.



Fig. Suppl.1 The six 512×512 grayscale test images used in the denoising experiments reported in [Figures 13, 14, 15, 16, 18, 22](#), [Figures Suppl.4, Suppl.5](#), and [Suppl.14](#), and [Table 1](#).

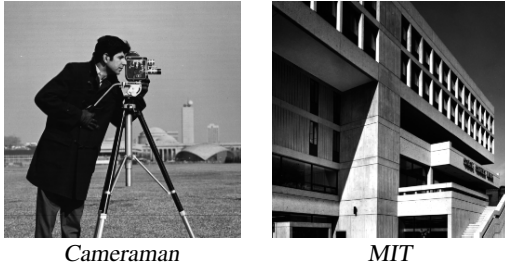


Fig. Suppl.2 *Cameraman* and *MIT* are two 256×256 grayscale images characterized by a large amount of sharp edges. These test images are used for the experiments reported in Figures 16, 20, 22 and Figure Suppl.6.

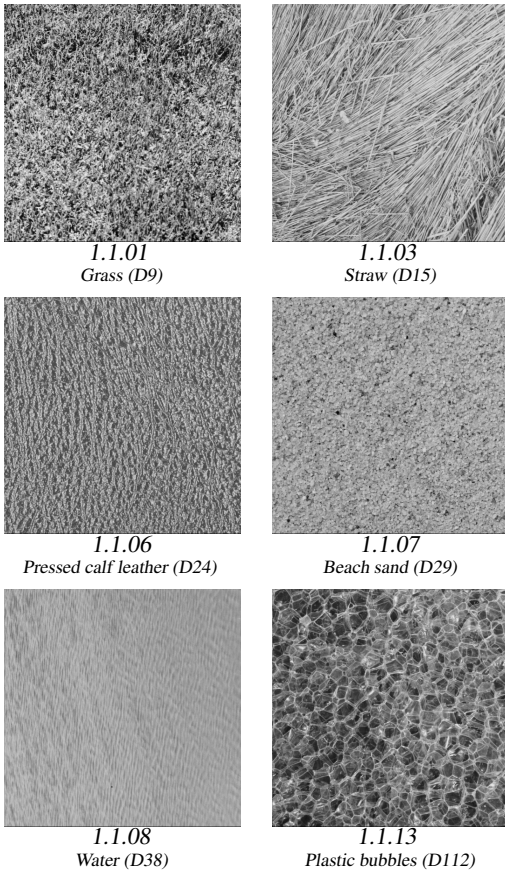


Fig. Suppl.3 The six 512×512 grayscale textures from the Brodatz dataset (USC-SIPI), used for the experiments reported in Figures Suppl.9, Suppl.10, and Suppl.11.

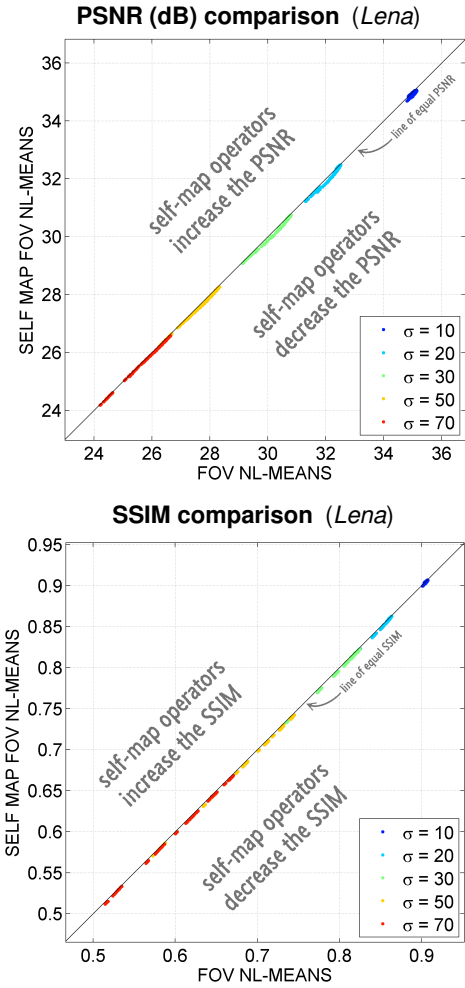


Fig. Suppl.4 PSNR and SSIM scatterplots of isotropic Foveated NL-means $\mathcal{F} = \mathcal{F}_{1,0}$ (horizontal axis) vs isotropic Foveated NL-means using self-map operator $\mathcal{F}_{1,0}^{\text{self}}$ (vertical axis). Each point represents the PSNR or SSIM value for a specific parameter pair (patch size, search neighborhood size) for the *Lena* image. Each color corresponds to a separate noise level. The two algorithms yield nearly the same performance, with only negligible differences in favor of the Foveated NL-means.

Suppl.2 Relative performance of self-map operators

We compare the Foveated NL-means of Section 9.1 against the Foveated NL-means given by the corresponding self-map foveation operator $\mathcal{F}_{1,0}^{\text{self}}$ of Section 7. To enable a meaningful comparison, both operators $\mathcal{F} = \mathcal{F}_{1,0}$ and $\mathcal{F}_{1,0}^{\text{self}}$ are isotropic ($\rho = 1, \theta = 0$) and obtained from the same windowing kernel \mathbf{k} . The scatterplots in Figure Suppl.4 report the PSNR and SSIM values obtained for the *Lena* image over the same parameters pairs and noise levels as in the plots in Figure 13. These results demonstrate that there is no significant loss of performance from using of self-map foveation operators.

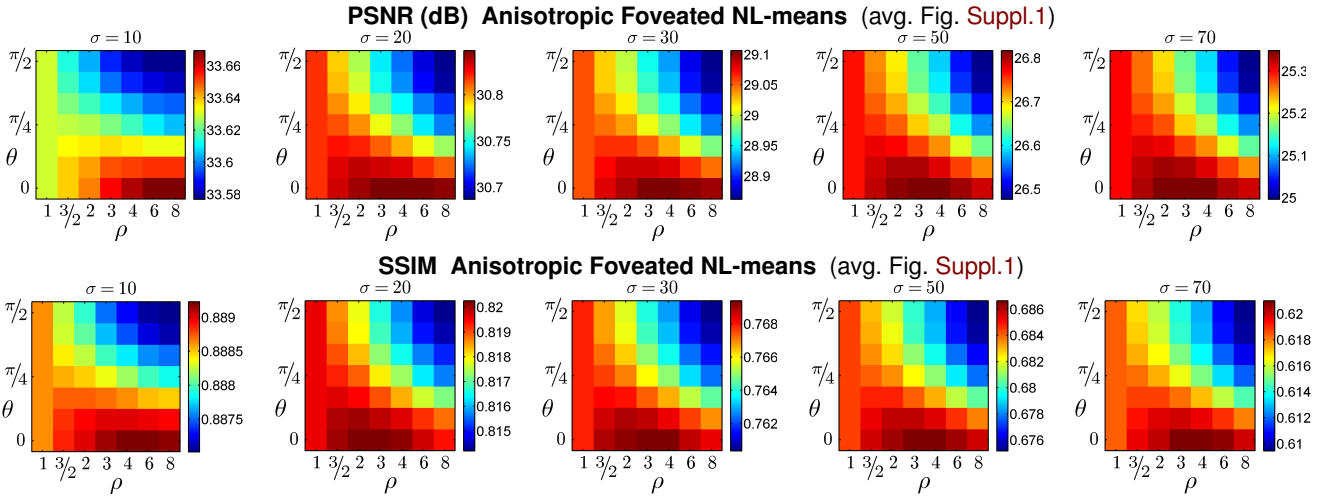


Fig. Suppl.5 Denoising performance of Anisotropic Foveated NL-means for various combinations of the θ and ρ parameters, measured as the average PSNR (dB) (top) and SSIM score (bottom) over the six images in Figure Suppl.1, each corrupted by 5 independent noise realizations. The best restoration performance can be achieved by anisotropic foveation operators characterized by blur kernels elongated along the meridians toward the patch center (i.e. $\theta = 0, \rho > 0$). In contrast, operators with tangential blur kernels (i.e. $\theta = \pi/2, \rho > 0$) are particularly ineffective and even lead to a performance loss with respect to the isotropic ones.

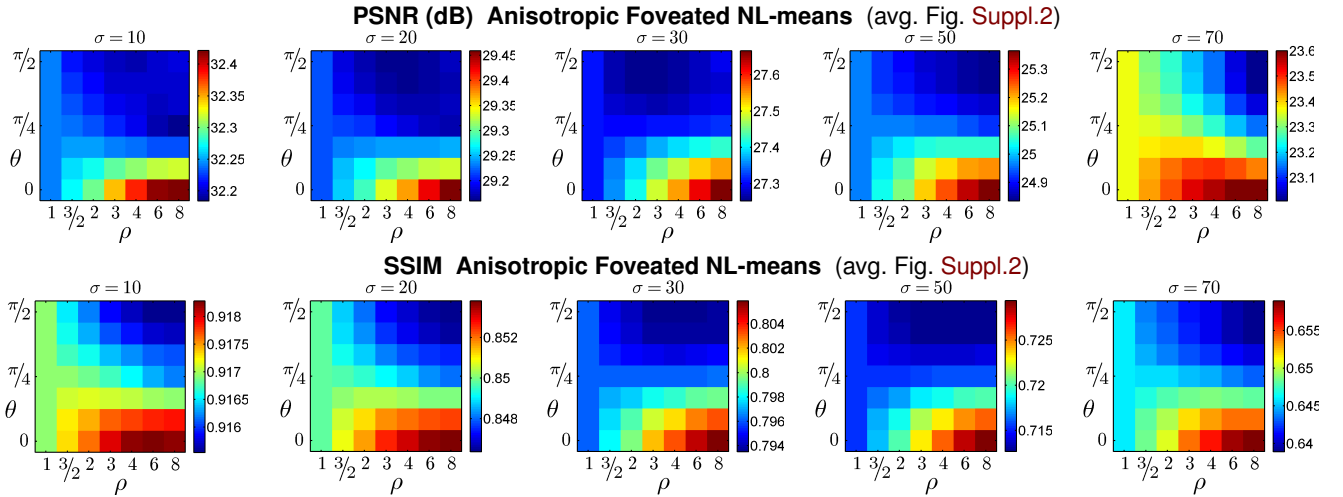


Fig. Suppl.6 Denoising performance of Anisotropic Foveated NL-means computed as in Figure Suppl.5 but over the two images in Figure Suppl.2, which are characterized by plenty of sharp edges. Note the more marked improvement from radial operators compared to Figure Suppl.5. The better ability of radial foveation operators to denoise sharp edges is consistent with the local analysis reported in Figure 19.

Suppl.3 Advantage of radial anisotropic foveation

We here provide the complete results from the comparative analysis of denoising with anisotropic foveation operators $\mathcal{F}_{\rho, \theta}$ for different combinations of elongation and orientation parameters, $\rho \in [1, 8]$, $\theta \in [0, \pi/2]$.

First, in Figure Suppl.5, we report PSNR and SSIM results averaged over the six test images in Figure Suppl.1 and five noise realizations. These results complete those included in Section 9.2 (Figure 18).

Next, in Figure Suppl.6, we report PSNR and SSIM results averaged over the two images with plenty of sharp edges (Figure Suppl.2) and five noise realizations. These results complete those included in Section 9.2.2 (Figure 20). A comparison with Figure Suppl.5 exposes the more marked improvement from using well elongated radially oriented ker-

nels when denoising images characterized by edges, in agreement with the analysis in Section 9.2.1.

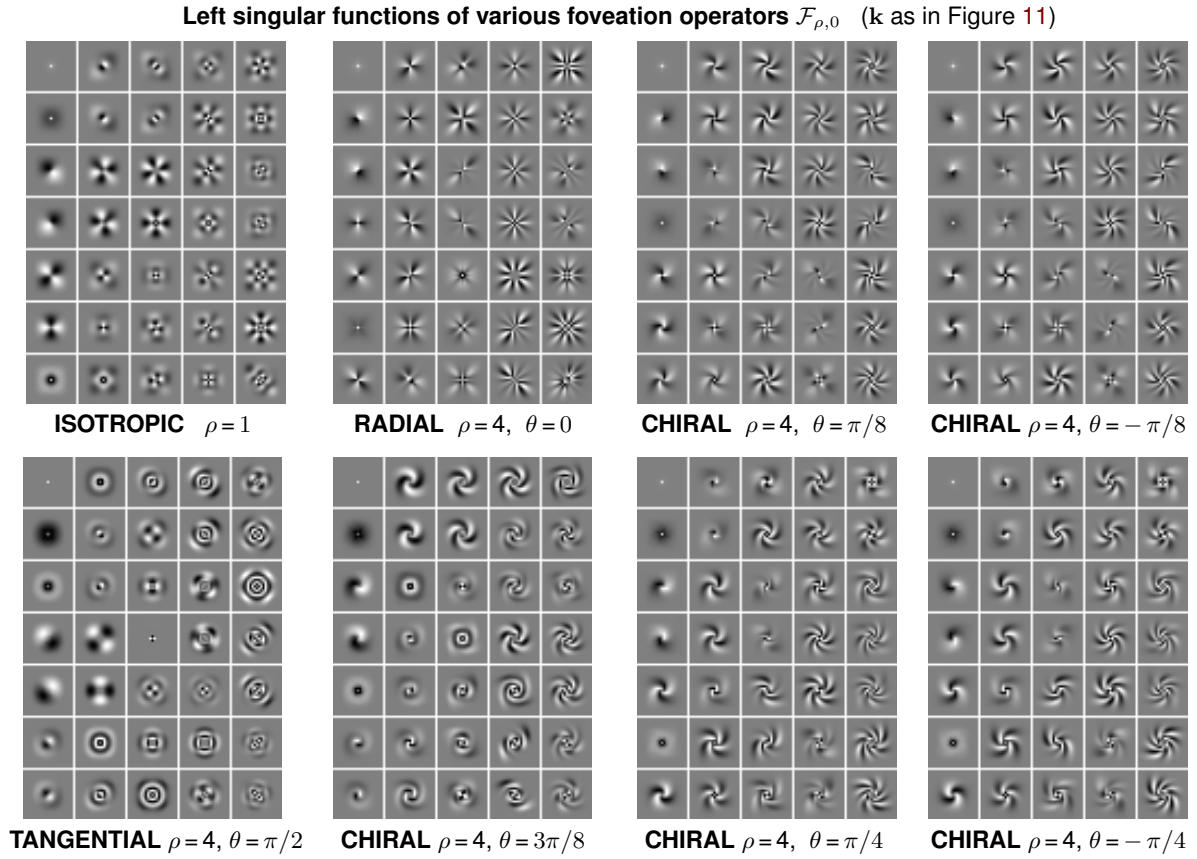


Fig. Suppl.7 The first 35 left singular functions of foveation operators $\mathcal{F}_{\rho,\theta}$ constructed from the 13×13 windowing kernel k in Figure 11, for various combinations of ρ and θ , encompassing isotropic, radial, right-hand chiral, left-hand chiral, as well as tangential operators. The singular functions are sorted columnwise, according to the decreasing magnitude of the respective singular value. Chiral operators capture spiral patterns that coil progressively faster as we transit from $\theta = 0$ (radial) to $\theta = \pi/2$ (tangential). Note the reversal of orientation that results from a change of sign of θ .

Suppl.4 Chirality

Throughout the experiments of Section 9 and Section Suppl.3 we have considered $\theta \in [0, \pi/2]$, which when $0 < \theta < \pi/2$ yields *right-hand* chiral operators, such as that in Figure 8(d). Right-hand chiral operators capture spiral pinwheel patterns along inward-twisting clockwise trajectories. Such patterns may be noticed among the singular functions as well as in the overall blur-kernel mosaic in Figure 8(d). These spirals are logarithmic, since the angle θ between the meridian and the orientation ϑ of the blur kernel is constant, as by the definition of anisotropic foveation operators (44). Figure Suppl.7 illustrates the left singular functions associated to various foveation operators; the different handedness of right-hand chiral operators ($\rho > 0$, $0 < \theta < \pi/2$) and left-hand chiral operators ($\rho > 0$, $-\pi/2 < \theta < 0$) can be appreciated, together with the faster coil rate when approaching the tangential operator. The coarsest of these spiral patterns bear a resemblance to the circular harmonic functions by Bigun and Granlund (1986) and Karlsson and Bigun (2011).

Natural images feature very low overall chirality, with pinwheel patterns existing as transitional singularities between iso-oriented regions and with clockwise and counter-clockwise twists occurring in similar proportions (Geisler

et al., 2001; Sigman et al., 2001; Osindero et al., 2006). It is therefore not surprising that for any given ρ and θ , the left-hand operator $\mathcal{F}_{\rho,-\theta}$, $\theta \in (0, \pi/2)$, yields a denoising performance on natural images that is practically identical to that of its right-hand twin $\mathcal{F}_{\rho,\theta}$. Figure Suppl.8 illustrates this fact for the particular case of the *Peppers* image corrupted by noise with $\sigma = 30$.

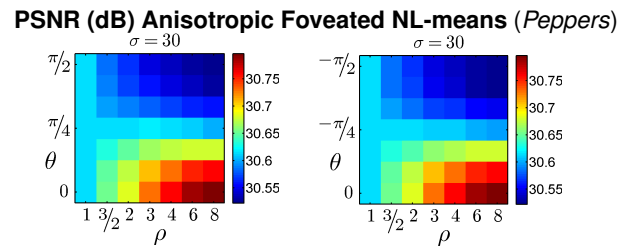


Fig. Suppl.8 Natural images exhibit negligible overall chirality: denoising performance (PSNR, dB) of the Anisotropic Foveated NL-means on the *Peppers* image corrupted by noise with $\sigma = 30$ for various combinations of θ and ρ . The results by the right-hand operator $\mathcal{F}_{\rho,\theta}$ and its left-hand twin $\mathcal{F}_{\rho,-\theta}$ are practically identical.

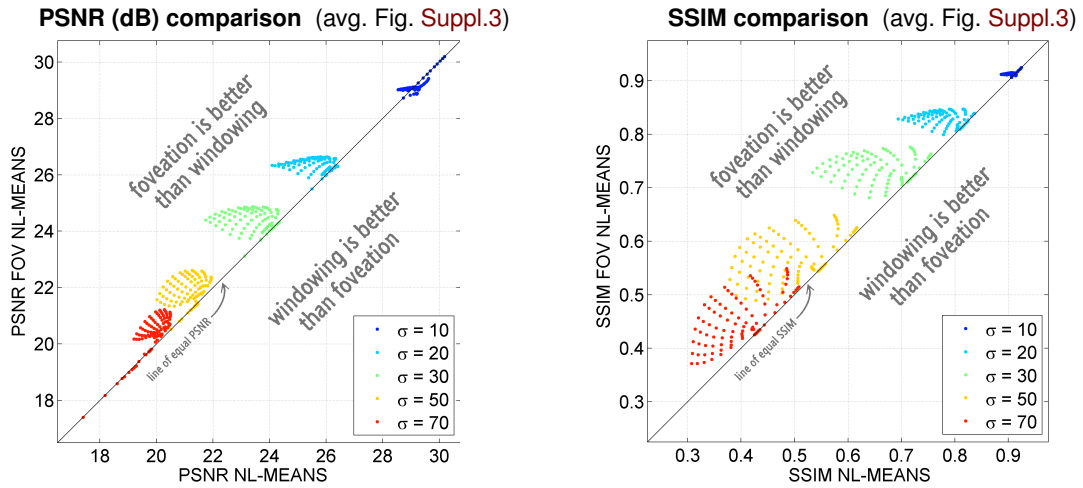


Fig. Suppl.9 Scatterplots of PSNR (left) and SSIM (right) of the standard NL-means (horizontal axis) vs Foveated NL-means (vertical axis) for the texture images in Figure Suppl.3. As in Figure 13, each point represents the average PSNR value (or SSIM score) achieved for a specific parameter pair (patch size, search radius) over the six textures corrupted by 5 independent noise realizations. Even though the values are inferior to those of Figure 13, qualitatively the scatterplots are comparable, and confirm the effectiveness of foveation also on challenging textures.

Suppl.5 Texture

Although very regular periodic textures possibly represent the ideal conditions for nonlocal algorithms, in some practical cases textures are characterized by sporadic patterns and thin random details that make very difficult to find a close matching replica of a sufficiently large patch. In these cases, textures can prove very challenging for nonlocal algorithms. Therefore, it is important to compare the windowed and the foveated distance separately on a dataset of images containing this sort of textural patterns. To this purpose, we selected the six images in Figure Suppl.3 from the Brodatz dataset (USC-SIPI) as a challenging dataset for denoising. We consider the same multiple combinations of patch and neighborhood sizes used in Section 9.1, and we compare the standard NL-means against Foveated NL-means based on isotropic and anisotropic self-map operators ($\rho = 4$). The results, summarized in Figure Suppl.9, confirm the difficulty in denoising these images, with a substantial drop in PSNR and SSIM scores with respect to the analogous experiment on complex natural images in Figure 13. Nevertheless, the scatterplots of Figure Suppl.9 and Figure 13 are qualitatively similar: when $\sigma > 10$, also on textures foveation leads to a denoising performance that is better, or in the worst case comparable, to that of windowing for any combination of patch size and neighborhood size. Likewise, upon choosing the respective combinations of patch and neighborhood size yielding the best performance for each algorithm, Foveated NL-means outperforms again the standard NL-means for $\sigma > 10$, while for $\sigma = 10$ they achieve the same performance. Figures Suppl.10 and Suppl.11 report a few fragments of the considered textures for visual comparison and attest that foveation leads to the best denoising performance also in terms of visual quality. In particular, isotropic Foveated NL-means provides reconstructed images that are sharper and better contrasted than those by the standard NL-means, and that, at least

numerically, are also marginally better than the images produced by the Anisotropic Foveated NL-means. This agrees with our analysis of Figure 19, where we already noted that on irregular textures isotropic foveation is more effective than the anisotropic one. The latter instead excels at regular and well structured edges and details, as confirmed by the experiments on *MIT* and *Cameraman* reported in Figure Suppl.6 and 16; indeed, on the relatively more structured *1.1.13* texture in Figure Suppl.11, isotropic and anisotropic foveation are on par. Images in Figures Suppl.10 and Suppl.11 are computed using the same general-purpose parameters appropriate for natural images with mixed content (such as those in Figure 13). The quality of the filtered images confirms that, in practice, foveation can be substantially more convenient than windowing also without any specific parameter adjustment to match the content of the processed images. These results are consistent with the experiment on natural images in Section 9.1, and demonstrate that foveated self-similarity is an effective prior also on challenging textures.

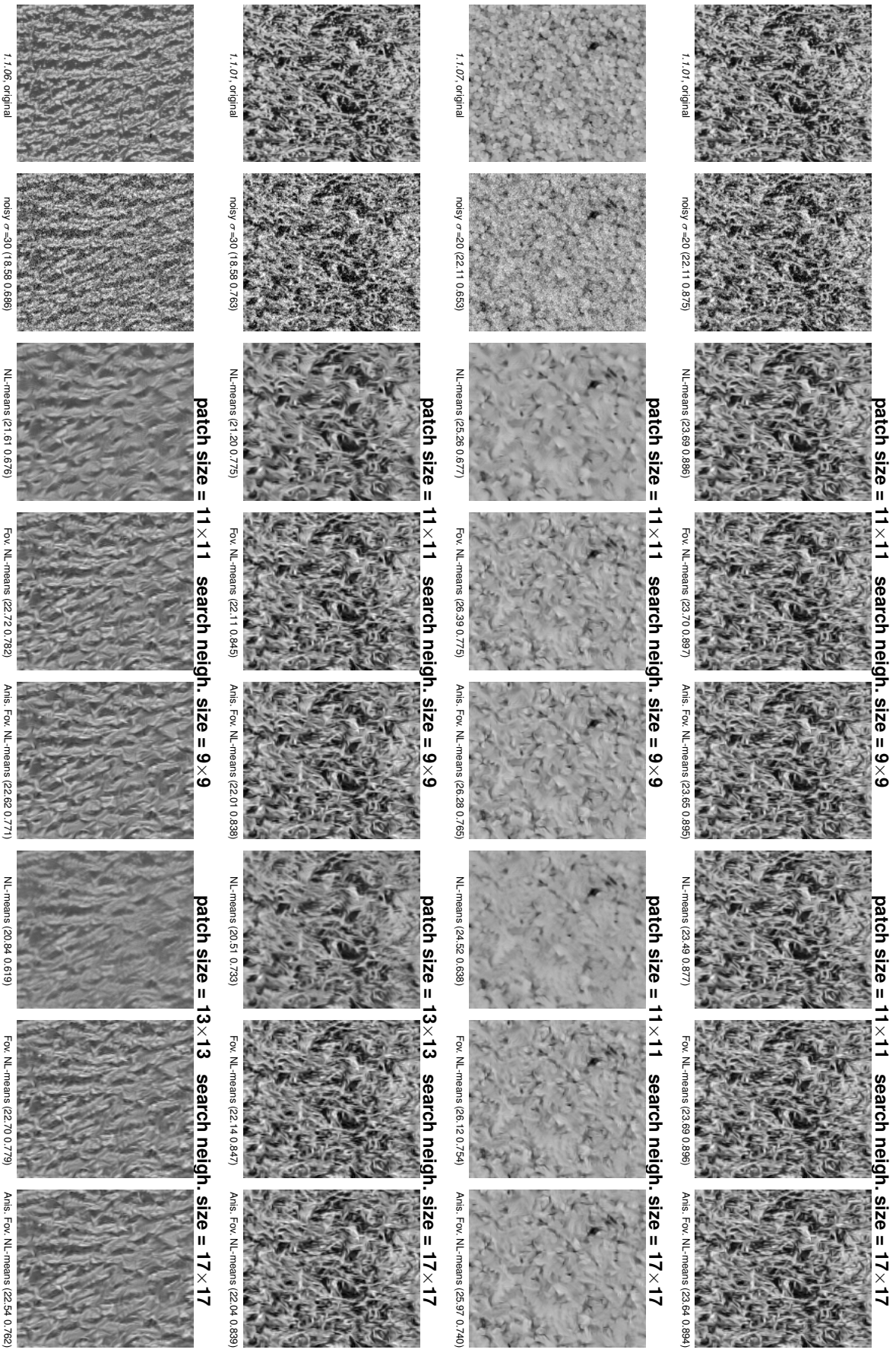


Fig. Suppl.10 Comparison between standard NL-means and Foveated NL-means on textures. PSNR (db) and SSIM values in parentheses are for the entire 512×512 image. The shown fragments have size 155×175 pixel. Results are given under two combinations of patch and search-neighborhood size, as detailed in Figure 14.

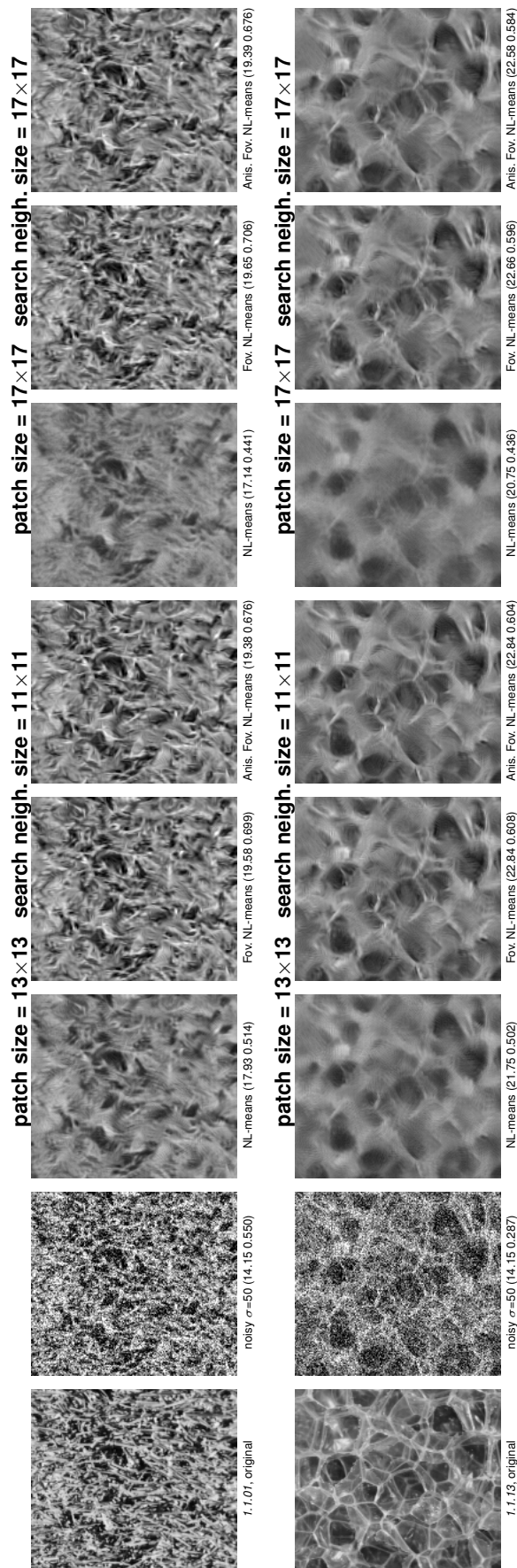


Fig. Suppl.11 Continued from Fig. Suppl.10.

Suppl.6 Method noise

For any given denoising method, the so-called *method noise* (Buades et al., 2005) can be computed as the difference between an image and its denoised version. Typically, this is done without addition of noise to the image, so the method noise corresponds to what the denoising filter has removed from the image as if it were noise. In our notation, it corresponds to applying the denoising filter directly to the noise-free image y while assuming this were corrupted by noise having a given standard deviation σ . As the practical goal of any denoising filter is to remove noise without damaging the underlying unknown noise-free image, the method noise should ideally appear as white noise, without featuring structures or any recognizable details of the image. Visual inspection of the method noise can thus be a reasonable qualitative criterion for judging the properties of a denoising filter, and we specifically use the method noise for comparing NL-means and Foveated NL-means.

In Figure Suppl.12, we report the method noise for the standard (first and third column) and Foveated (second and fourth column) NL-means, where these methods have been tuned to operate on images corrupted by noise having a given standard deviation σ . A larger σ makes the filtering more aggressive. To enable a direct as well as fair comparison, we again apply each filter according to two combinations of patch size and neighborhood size, one combination considered optimal for the standard NL-means (first and second column), and another combination optimal for the Foveated NL-means (third and fourth column). These optimal parameters have been determined based on the analysis of Figure 14, and are reported at the end of Section 9.1.

The visual inspection of Figure Suppl.12 suggests that the method noise of Foveated NL-means contains fewer and fainter structured details than the method noise of the standard NL-means. Expectedly, the differences between the two algorithms increase when these are tuned to suppress stronger noise or operate with larger patches (see Section 6.1.2). For instance, in the bottom row of the figure ($\sigma = 30$), the method noise of standard NL-means contains prominent edges and details. In contrast, for small values of σ , and hence small patch size, the differences become less noticeable. Thus, the method noise analysis attests favorably to foveation, and agrees with the results from the traditional denoising experiments reported in the earlier sections.

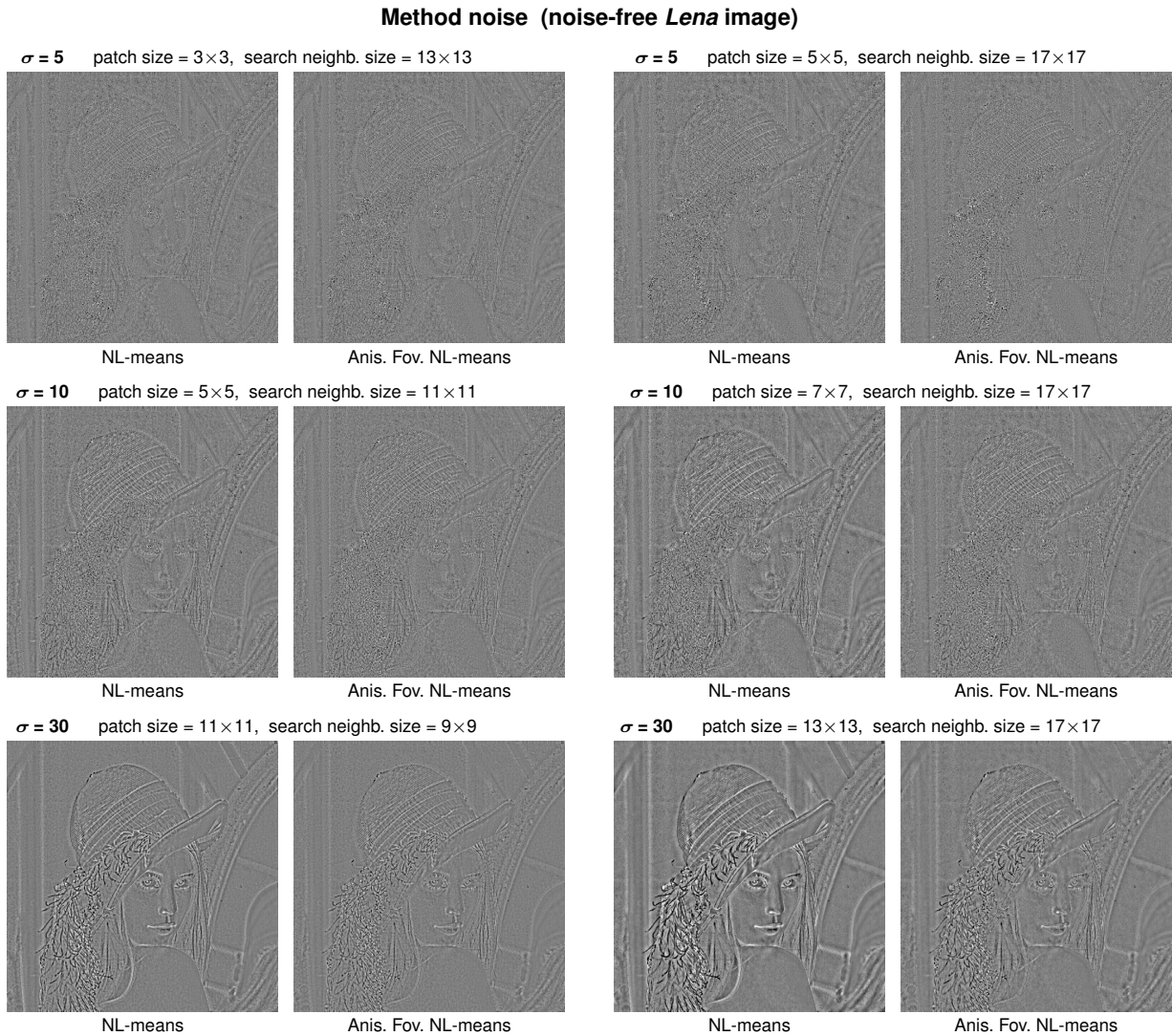


Fig. Suppl.12 The method noise of the standard (first and third columns) and Foveated (second and fourth columns) NL-means when processing the noise-free *Lena* at different configurations of patch size, neighborhood size, and assumed noise standard deviation σ . A larger σ makes the filtering more aggressive and thus more features of the image emerge in the method noise. The differences between the two algorithms also increase with the value of σ , and the method noise for standard NL-means contains always more edges and image structures than the method noise of the Foveated NL-means. Images are scaled to a common range of intensities, separately for each value of σ .

Suppl.7 Irregular design

As a further confirmation of the practical importance of the central-acuity and pixelwise-compatibility requirements, thus of a certain regularity in the above scale-space sectioning, we consider modified “foveation” operators that disobey the decreasing acuity towards the periphery of the patch. Specifically, we construct operators from either randomly permuted windowing kernels (i.e. *randomized* windowing kernels) or from windowing kernels that attain their minimum at the center of the patch and increase towards the periphery (i.e. *inside-out* windowing kernels), as illustrated in Figure Suppl.13. In the randomized case, any blur kernel v_u , $u \in U$, composing \mathcal{F} can be narrow or wide regardless of the position of u with respect to the center of the patch. In the inside-out case, the wider blur kernels correspond to smaller $|u|$ and the most narrow ones are found near the boundary of the patch,

i.e. in completely reversed positions compared to the regular design of a foveation operator. Figure Suppl.14 reports the denoising performance obtained from such irregular “foveation” operators over the natural images in Figure Suppl.1: the performance is notably inferior to that obtained from the regular design. The reversed inside-out case is particularly unfortunate. In the same figure, we report also the corresponding results obtained by the standard NL-means, which is found to be impaired even further by the use of irregular windowing kernels. Note however that as the noise gets heavier ($\sigma \geq 50$), the gap between regular and irregular design gets narrower, possibly highlighting a limit in the resolvability of finer details, which hinders all these operators.

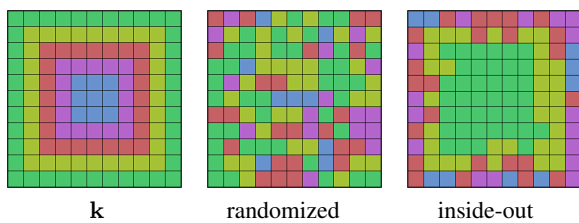


Fig. Suppl.13 A windowing kernel \mathbf{k} of size 11×11 (left) and two windowing kernels obtained by permutation of \mathbf{k} : a completely *randomized* windowing kernel (center) and a reversed *inside-out* windowing kernel (right), which is increasing from the center towards the periphery. These two kernels were obtained by shuffling \mathbf{k} , thus they attain the same values with same proportions (i.e., their histograms coincide), hence also their norms are the same.

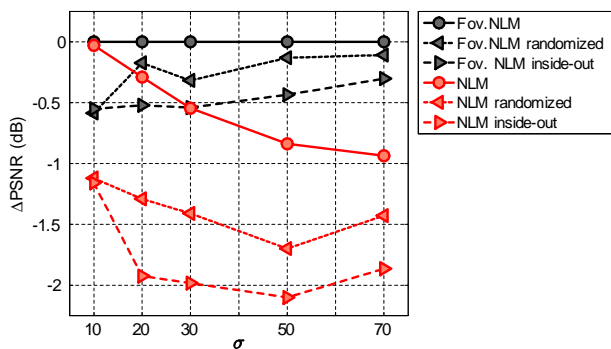


Fig. Suppl.14 Comparison between the Foveated NL-means (based on $\mathcal{F}_{1,0}$) and the Foveated and standard NL-means based on randomized and inside-out versions of the windowing kernel \mathbf{k} , as shown in Figure Suppl.13. The reported value (ΔPSNR) is the difference between the PSNR achieved by Foveated NL-means and the PSNR achieved by other denoising algorithms. This substantial performance gap emphasizes the importance of designing foveator operators according to the constrained construction presented in Section 5, and in particular the relevance of central acuity and pixelwise compatibility requirements.