SIAM J. Imaging Sciences
Vol. 11, No. 1, pp. 407–441

Dense Scale Selection Over Space, Time, and Space-Time[*]

Tony Lindeberg[†]

Abstract. Scale selection methods based on local extrema over scale of scale-normalized derivatives have been primarily developed to be applied sparsely—at image points where the magnitude of a scale-normalized differential expression additionally assumes local extrema over the domain where the data are defined. This paper presents a methodology for performing dense scale selection, so that hypotheses about local characteristic scales in images, temporal signals, and video can be computed at every image point and every time moment. A critical problem when designing mechanisms for dense scale selection is that the scale at which scale-normalized differential entities assume local extrema over scale can be strongly dependent on the local order of the locally dominant differential structure. To address this problem, we propose a methodology where local extrema over scale are detected of a quasi quadrature measure involving scale-space derivatives up to order two and propose two independent mechanisms to reduce the phase dependency of the local scale estimates by (i) introducing a second layer of postsmoothing prior to the detection of local extrema over scale, and (ii) performing local phase compensation based on a model of the phase dependency of the local scale estimates depending on the relative strengths between first- and second-order differential structures. This general methodology is applied over three types of domains: (i) spatial images, (ii) temporal signals, and (iii) spatio-temporal video. Experiments demonstrate that the proposed methodology leads to intuitively reasonable results with local scale estimates that reflect variations in the characteristic scales of locally dominant structures over space and time.

Key words. scale, scale selection, spatial, temporal, spatio-temporal, scale invariance, scale space, feature detection, differential invariant, video analysis, image analysis, computer vision

AMS subject classifications. 65D18, 65D19, 68U10

DOI. 10.1137/17M114892X

1. Introduction.

The notion of scale is essential when computing features from image data in biological or artificial visual perception. Results from biological research regarding the early visual areas in the lateral geniculate nucleus (LGN) and the primary visual cortex (V1) (Hubel and Wiesel [19, 20, 21], DeAngelis et al. [12, 11]) as well as theoretical results from normative theory of visual operations (Lindeberg [39, 43]) based on scale-space theory (Iijima [22]; Witkin [67]; Koenderink [26, 27]; Koenderink and van Doorn [28, 30]; Lindeberg [33, 38]; Florack [14]; Sporring et al. [59]; Weickert, Ishikawa, and Imiya [66]; ter Haar Romeny et al. [62, 61]) state that local image measurements in terms of receptive fields constitute both a natural and an efficient model for expressing early visual operations.

When applying such spatial or spatio-temporal receptive fields at multiple spatial and

[*] Received by the editors September 22, 2017; accepted for publication (in revised form) November 27, 2017; published electronically February 13, 2018.

http://www.siam.org/journals/siims/11-1/M114892.html

Funding: This work was supported by the Swedish Research Council under contract 2014-4083.

[†] Computational Brain Science Lab, Department of Computational Science and Technology, School of Computer Science and Communication, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden (tony@kth.se, https://www.kth.se/profile/tony).

temporal scales, a basic observation is that the responses that are obtained from the receptive field operators can be strongly dependent on the scale levels at which they are applied. Thus, one may raise the question of whether it is possible from the data itself to generate hypotheses about local appropriate scales in the image data, so as to adapt subsequent processing to the local image structures. Initially, this problem could possibly be seen as intractable. Would it at all be possible to generate hypotheses about interest scale levels before recognizing the objects we are interested in or defining the specific purpose for which the scale estimates are to be used? Research of Lindeberg [33, 35, 34], as well as follow-up work by Bretzner et al. [5, 4]; Chomat et al. [8]; Lowe [49]; Mikolajczyk and Schmid [51]; Lazebnik, Schmid, and Ponce [31]; Rothganger et al. [56]; Bay et al. [2]; Tuytelaars and Mikolajczyk [64]; Negre et al. [52]; and Lindeberg [40, 42], has, however, demonstrated that such an approach is feasible (see [37, 41] for overviews). A general framework for automatic selection of local characteristic scales can be formulated based on the detection of local extrema over scale of scale-normalized feature responses. Specifically, such local scale estimates transform in a scale-covariant way under spatial scaling transformations of the image domain, which is a highly desirable property of a scale selection mechanism, since it implies that the scale estimates will automatically follow local scale variations in the image data. Corresponding local scale selection mechanisms can also be expressed over temporal and spatio-temporal domains (Lindeberg [46, 44, 45]).

A common property of most successful applications of scale selection to computer vision applications, however, is that the scale selection method is applied sparsely over the image domain, most commonly at interest points. If one is attempting to perform dense scale selection based on the two most common rotationally invariant differential invariants for interest point detection, the spatial Laplacian, or the determinant of the spatial Hessian, then the results of scale selection will usually not be stable or useful far away from the interest points.

To address this problem, an initial mechanism for dense scale selection was proposed in Lindeberg [35] based on the detection of local extrema over scale of a spatial quasi quadrature measure that constitutes a rotationally invariant measure of the amount of energy in the first- and second-order differential structures of a spatial image. Modifications of this approach were used by Almansa and Lindeberg [1] for estimating the local scale of fingerprint patterns for fingerprint recognition and were specifically shown to improve the quality of minutiae extraction. Related methods for scale selection have been developed by detecting peaks of weighted entropy measures or Lyapunov functionals over scale (Kadir and Brady [24]; Sporring, Colios, and Trahanias [58]), by minimizing normalized error measures over scale (Lindeberg [36]), by determining local scales for variable bandwidth mean shift from the scale bandwidth that maximizes the norm of the normalized mean shift vector (Comaniciu, Ramesh, and Meer [10]), by detecting maxima of steered energy responses over scales (Ng and Bharath [53]), by comparing reliability measures from statistical classifiers for texture analysis at multiple scales (Kang, Morooka, and Nagahashi [25]), by measuring the size variations of regions to which pixels belong under total variation (TV) flow (Brox and Weickert [6]), by measuring local oscillations in signals (Jones and Le [23]), and by computing image segmentations from the scales at which a supervised classifier delivers class labels with the highest reliability measure (Loog et al. [48, 32]). Specifically, a more algorithmic way of generating scale estimates for image matching away from the locations of interest points was recently proposed by Hassner et al. [18] and Tau and Hassner [60] by considering subspaces generated by local image descriptors computed over multiple scales to improve the performance of stereo matching.

Closely related issues of estimating dominant scales in signals have been studied in wavelet theory and local frequency analysis (Cohen [9]). For example, a local Gaussian-weighted windowed Fourier transform of a 1-D signal corresponds to filtering with Gabor functions [15]. Mallat and Hwang [50] proposed to characterize singularities in terms of Lipshitz exponents and detected maxima in the wavelet transform. Methods based purely on wavelets and/or local frequency analysis have, however, been less developed for 2-D image data or 2+1-D video data.

The purpose of this article is to perform a deeper study into the problem of dense scale selection for images and video. A basic problem that can be observed if performing dense scale selection based on the basic quasi quadrature measure in [35] is that the local scale estimates can be strongly phase dependent. If applied to a sine wave in one or two dimensions, the scale estimates can be biased depending on the relative strength of first-order vs. second-order differential structure. To reduce this phase dependency, we will consider two independent mechanisms in terms of (i) spatial smoothing and (ii) local phase compensation. We will specifically analyze the properties of these mechanisms, determine free parameters in the corresponding methods, and show that these mechanisms may substantially reduce the local phase dependency. We will also generalize this dense scale selection mechanism to the spatio-temporal domain, to perform simultaneous dense scale selection of both local spatial and temporal scales. Compared to wavelet-based approaches for local frequency analysis, the methods that we propose are invariant to rotations over the spatial domain. Additionally, we prove that both the spatial and the temporal scale estimates are provably covariant under independent scaling transformations of the spatial and the temporal domains, implying that the local scale estimates are guaranteed to automatically follow local variations in the spatial extent and the temporal duration of spatial, temporal, or spatio-temporal image structures, which is a highly desirable property of a scale selection mechanism. Experiments on different types of spatial images, temporal signals, and spatio-temporal video demonstrate that the proposed theory leads to dense spatial and temporal scale maps with intuitively reasonable properties.

2. Dense spatial scale selection over a purely spatial domain.

The context we consider is a spatial scale-space representation $L(x, y;\ s)$ defined from any 2-D image $f(x, y)$ by convolution with Gaussian kernels

$$g(x, y;\ s) = \frac{1}{2\pi s} e^{-(x^2+y^2)/2s} \tag{1}$$

at different spatial scales s (see [22, 67, 26, 28, 30, 33, 38, 14, 59, 66, 61]),

$$L(\cdot, \cdot;\ s) = g(\cdot, \cdot;\ s) * f(\cdot, \cdot), \tag{2}$$

and with γ-normalized derivatives defined at any scale s according to (see Lindeberg [35])

$$\partial_\xi = s^{\gamma_s/2}\, \partial_x, \quad \partial_\eta = s^{\gamma_s/2}\, \partial_y. \tag{3}$$

If attempting to perform dense local scale selection in the possibly most straightforward manner, by detecting local extrema of the scale-normalized Laplacian $\nabla^2_{norm} L = s\,(L_{xx} + L_{yy})$

or the scale-normalized determinant of the Hessian $\det \mathcal{H}_{norm} L = s^2\,(L_{xx}L_{yy} - L_{xy}^2)$ at points that are not interest points, one will soon find that the resulting scale estimates will be strongly dependent on the points at which they are computed. The reason for this is that the underlying interest point detectors primarily respond to very specific aspects of the second-order differential structure; see Figure 1 for an illustration. Scale selection by the scale-normalized Laplacian or the scale-normalized determinant of the Hessian operators is therefore primarily intended for image structures that lead to strong responses for these differential operators, such as spatial interest points (see Lindeberg [35, 40, 42]).

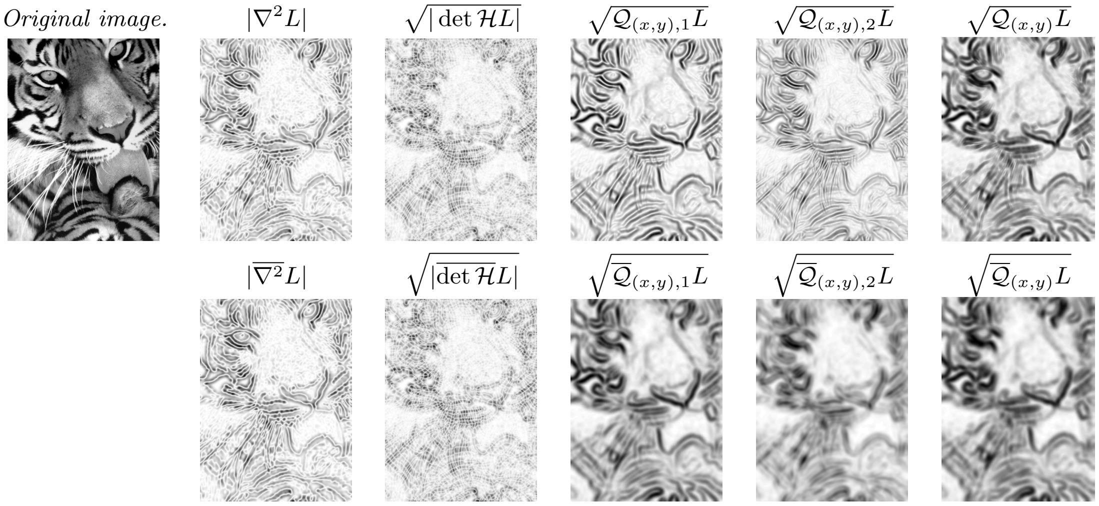

Original image.
$|\nabla^2 L|$
$\sqrt{|\det \mathcal{H}L|}$
$\sqrt{\mathcal{Q}_{(x,y),1}L}$
$\sqrt{\mathcal{Q}_{(x,y),2}L}$
$\sqrt{\mathcal{Q}_{(x,y)}L}$
$|\overline{\nabla^2 L}|$
$\sqrt{|\overline{\det \mathcal{H}L}|}$
$\sqrt{\overline{\mathcal{Q}}_{(x,y),1}L}$
$\sqrt{\overline{\mathcal{Q}}_{(x,y),2}L}$
$\sqrt{\overline{\mathcal{Q}}_{(x,y)}L}$

Figure 1. *The result of computing the unsmoothed quasi quadrature measure $\mathcal{Q}_{(x,y),norm}L$ (top row) and the postsmoothed quasi quadrature entity $\overline{\mathcal{Q}}_{(x,y),norm}L$ (bottom row) with their underlying first- and second-order components at scale $t = 16$ for $C_s = 1/\sqrt{2}$. For comparison, the Laplacian and the determinant of the Hessian responses (top row) as well as the result of applying spatial smoothing to these differential operators for relative postsmoothing scale $c = 1$ (bottom row) are also shown at the same scale to illustrate how the Laplacian and the determinant of the Hessian responses will be limited to specific aspects of local image structures and thereby the limitations of using the Laplacian or the determinant of the Hessian operators only for dense scale selection. (Image size: 480 × 640.)*

To perform dense scale selection, it is therefore more natural to seek a differential expression that responds to wider classes of image structures, comprising both first- and second-order differential structures, and without bias toward primarily specific aspects of the second-order differential image structure.

2.1. A spatial quasi quadrature measure. By combining the rotationally invariant differential invariants (i) the *gradient magnitude*

$$|\nabla L|^2 = L_x^2 + L_y^2 \tag{4}$$

as a measure of the amount of first-order structure, and (ii) the *Frobenius norm of the Hessian matrix*

$$\|\mathcal{H}L\|_F^2 = L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2 \tag{5}$$

as a measure of the amount of second-order structure, we will consider extensions of the following *quasi quadrature* measure (Lindeberg [35, eq. (63)]):

$$\mathcal{Q}_{(x,y),norm} L = s\,(L_x^2 + L_y^2) + C_s\, s^2\,(L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2) \tag{6}$$

based on scale-normalized derivatives for $\gamma_s = 1$. This differential entity can be seen as an approximation of the notion of a quadrature pair of an odd and an even filter (Gabor [15]) as more traditionally formulated based on a Hilbert transform (Bracewell [3, pp. 267–272]) and then extended to 2-D image space, while being confined within the family of differential expressions based on Gaussian derivatives and additionally being rotationally invariant.

If complemented by spatial integration, the components of this quasi quadrature measure are specifically related to the following class of energy measures over the frequency domain (Lindeberg [35, App. A.3]) (here expressed in terms of multi-index notation for the partial derivatives $x^\alpha = x_1^{\alpha_1} \dots x_D^{\alpha_D}$ with $|\alpha| = \alpha_1 + \dots + \alpha_D$):

$$E_{m,\gamma-norm} = \int_{x \in \mathbb{R}^D} \sum_{|\alpha|=m} s^{m\gamma_s}\, L_{x^\alpha}^2\, dx = \frac{s^{m\gamma_s}}{(2\pi)^D} \int_{\omega \in \mathbb{R}^D} |\omega|^{2m}\, \hat{g}^2(\omega;\ s)\, d\omega. \tag{7}$$

For the specific choice of $C_s = 1/2$, the quasi quadrature measure (6) coincides with the proposal by Loog [47] and Griffin [16] to define a metric of the N-jet in scale space.

2.1.1. Complementary scale normalization.

To allow for richer degrees of freedom regarding the scale selection properties, we allow for complementary scale normalization of the form

$$\mathcal{Q}_{(x,y),\Gamma-norm} L = \frac{s\,(L_x^2 + L_y^2) + C_s\, s^2\,(L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2)}{s^{\Gamma_s}}, \tag{8}$$

which is still within the class of scale-normalized differential expressions as obtained from γ-normalized derivatives

$$\mathcal{Q}_{(x,y),\gamma-norm} L = s^{\gamma_1}\,(L_x^2 + L_y^2) + C_s\, s^{2\gamma_2}\,(L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2) \tag{9}$$

for $\gamma_1 = 1 - \Gamma_s$ and $\gamma_2 = 1 - \frac{\Gamma_s}{2}$. A major motivation for introducing the parameter Γ_s in (8) is that if we were to use $\Gamma_s = 0$, then it can be shown (Lindeberg [34]) that the selected scale would be infinite for any diffuse step edge, which is not a desirable property for a dense scale selection mechanism, while if we use a value of $\Gamma_s > 0$, the selected scale for a diffuse edge will be finite [34, eq. (23)]:

$$\hat{s} = \frac{\gamma_1}{1 - \gamma_1}\, s_0 = \frac{1 - \Gamma_s}{\Gamma_s}\, s_0. \tag{10}$$

Concerning the choice of Γ_s, it can be observed that setting $\Gamma_s = 1/2$ leads to $\gamma_1 = 1/2$ and $\gamma_2 = 3/4$, which are the values derived for edge detection and ridge detection, respectively, to make the scale estimate for a diffuse step edge reflect the diffuseness of the edge and to make the scale estimate for a Gaussian ridge reflect the width of the ridge [34]. For blob detection based on second-order derivatives, $\gamma_s = 1$ corresponding to $\Gamma_s = 0$ is, on the other hand, the preferred choice to ensure that the scale level for a rotationally symmetric or affine deformed Gaussian blob reflects the scale of the blob [35, 40]. From these indications, we could expect to choose the complementary scale normalization parameter Γ_s in the range $\Gamma_s \in]0, \frac{1}{2}]$.

2.1.2. Complementary spatial postsmoothing. While the combination of first- and second-order information in (6) and (8) will decrease the spatial dependency of the differential expression compared to using only either first- or second-order information, the resulting differential expressions will not produce a constant scale estimate for a sine wave, unless the computations are performed at a scale level perfectly adapted to the wavelength of the signal. To reduce the local ripples caused by this *phase dependency*, we introduce complementary smoothing of the quadrature entity $\mathcal{Q}_{(x,y),\Gamma-norm}$ using an integration scale parameter s_{int} proportional to the local scale parameter s used for computing the spatial derivatives

$$\overline{\mathcal{Q}}_{(x,y),\Gamma-norm} L = \mathcal{E}_{s_{int}}(\mathcal{Q}_{(x,y),\Gamma-norm}), \tag{11}$$

where $\mathcal{E}_{s_{int}}$ denotes a Gaussian averaging operation with scale parameter $s_{int} = c^2\,s$. We also define the first- and second-order components of this entity as

$$\overline{\mathcal{Q}}_{(x,y),1,\Gamma-norm} L = s^{-\Gamma_s}\,\mathcal{E}_{s_{int}}(|\nabla L|^2), \tag{12}$$

$$\overline{\mathcal{Q}}_{(x,y),2,\Gamma-norm} L = C_s\,s^{-\Gamma_s}\,\mathcal{E}_{s_{int}}(\|\mathcal{H}L\|_F^2). \tag{13}$$

Figure 1 shows the result of computing these quasi quadrature measures as well as their underlying first- and second-order components for a grey-level image that contains image textures at different scales. As can be seen from the results, (i) the quasi quadrature is less sensitive to the spatial variability in a dense textured image pattern compared to the Laplacian or the determinant of the Hessian operators, and (ii) the postsmoothing operation further decreases the sensitivity.

2.1.3. Basic scale selection method. A basic method for dense scale selection therefore consists of detecting local extrema over scale of either the pointwise scale-normalized quasi quadrature entity (8)

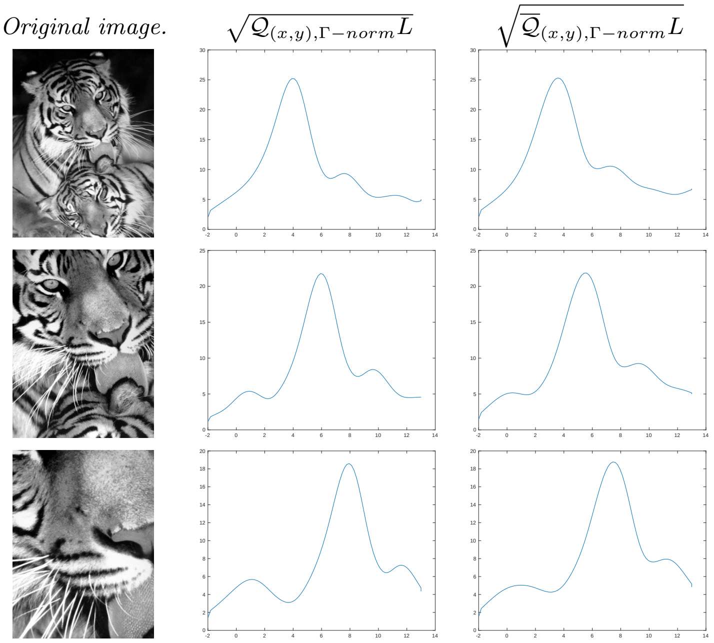
$$\hat{s}_{\mathcal{Q}L} = \operatorname{argmaxlocal}_{s>0} \mathcal{Q}_{(x,y),\Gamma-norm} L \tag{14}$$

or the corresponding postsmoothed entity (11)

$$\hat{s}_{\overline{\mathcal{Q}}L} = \operatorname{argmaxlocal}_{s>0} \overline{\mathcal{Q}}_{(x,y),\Gamma-norm} L. \tag{15}$$

Figure 2 illustrates the effects of these operations for three different images of the same poster taken at different distances between the poster and the camera. As can be seen from the graphs, the scale values at which the local extrema over scale are assumed do adapt to the size variations in the image domain and are assumed at coarser scales relative to the fixed image resolution as the camera approaches the object.

2.2. Scale selection properties. When applying this methodology in practice, there are a number of additional issues that need to be considered, which we will illustrate by closed form theoretical analysis using idealized image models representing a dense texture pattern or sparse image features, respectively.



Original image. $\sqrt{\mathcal{Q}_{(x,y),\Gamma-norm}L}$ $\sqrt{\overline{\mathcal{Q}}_{(x,y),\Gamma-norm}L}$

Figure 2. *Scale-space signatures computed at corresponding central points for three images of a poster taken at different distances. Each scale-space signature shows the variation over scale of the scale-normalized quasi quadrature measure at the center as the function of effective scale approximated by* $s_{eff} = \log_2(s_0 + s)$ *for* $s_0 = 1/8$*. Middle column: the unsmoothed quasi quadrature measure* $\mathcal{Q}_{(x,y),\Gamma-norm}L$ *using* $\Gamma_s = 1/4$*. Right column: the postsmoothed quasi quadrature measure* $\overline{\mathcal{Q}}_{(x,y),\Gamma-norm}L$ *using* $\Gamma_s = 1/4$ *and* $c = 1$*. Note how the scale levels at which the local maxima over scale are assumed to follow the size variations in the image domain caused by varying the distance between the camera and the poster. All results have been computed using* $C_s = 1/\sqrt{(1-\Gamma_s)(2-\Gamma_s)}$*. (Scale range:* $s \in [0.1, 8192]$*. Image size:* 480×640 *pixels.)*

2.2.1. Two-dimensional sine wave.

For a two-dimensional sine wave

$$f(x,y) = \sin(\omega_0 x) + \sin(\omega_0 y), \tag{16}$$

the scale-space representation can be computed in closed form as

$$L(x,y;\ t) = e^{-\omega_0^2 s/2}(\sin(\omega_0 x) + \sin(\omega_0 y)). \tag{17}$$

If we disregard the spatial postsmoothing step by setting the proportionality parameter c between the local scale parameter and the integration scale parameter to $c = 0$, then the quasi quadrature entity assumes the form

$$\begin{aligned}\mathcal{Q}_{(x,y),\Gamma-norm}L =& \omega_0^2 e^{-\omega_0^2 s} s^{1-\Gamma_s} \\ & \times \left(\cos^2(\omega_0 x) + \cos^2(\omega_0 y) + C_s \omega_0^2 s \left(\sin^2(\omega_0 x) + \sin^2(\omega_0 y)\right)\right).\end{aligned} \tag{18}$$

By differentiating this expression with respect to scale s, it follows that for the image points $(x,y) = (m\pi/\omega_0, n\pi/\omega_0)$ at which only the first-order component

$$\mathcal{Q}_{(x,y),1,\Gamma-norm}L = \frac{|\nabla_{norm}L|^2}{s^{\Gamma_s}} = \frac{s\,(L_x^2 + L_y^2)}{s^{\Gamma_s}} \tag{19}$$

responds, the local extrema over scales are assumed at

$$\hat{s}_1 = \frac{1 - \Gamma_s}{\omega_0^2}. \tag{20}$$

Correspondingly, for the spatial positions $(x, y) = ((\pi/2 + m\pi)/\omega_0, (\pi/2 + n\pi)/\omega_0)$ at which only the second-order component

$$\mathcal{Q}_{(x,y),2,\Gamma-norm} L = \frac{C_s \, \|\mathcal{H}_{norm} L\|_F^2}{s^{\Gamma_s}} = \frac{C_s \, s^2 \, (L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2)}{s^{\Gamma_s}} \tag{21}$$

responds, the local extrema are assumed at

$$\hat{s}_2 = \frac{2 - \Gamma_s}{\omega_0^2}. \tag{22}$$

In this respect, the combination of first- and second-order derivatives in the quasi quadrature entity will lead to a strong *phase dependency in the scale estimates*.

At the intermediate points $(x, y) = ((\pi/4 + m\pi/2)/\omega_0, (\pi/4 + n\pi/2)/\omega_0)$, the scale estimate is given by

$$\hat{s}_{intermed} = \frac{\sqrt{C_s^2(\Gamma_s - 2)^2 - 2C_s\Gamma_s + 1} - C_s(\Gamma_s - 2) - 1}{2C_s\omega_0^2}. \tag{23}$$

If we require the scale estimates at the intermediate points to be equal to the geometric average of the extreme cases

$$\hat{s}_{intermed} = \sqrt{\hat{s}_1 \, \hat{s}_2} = \frac{\sqrt{(1 - \Gamma_s)(2 - \Gamma_s)}}{\omega_0^2}, \tag{24}$$

then this implies that C_s should be chosen as

$$C_s = \frac{1}{2 - \Gamma_s}. \tag{25}$$

Alternatively, if we determine the weighting parameter C_s such that the relative strengths of the first- and second-order components become equal at the midpoints $(x, y) = (\pi/4 + m\pi/2)/\omega_0, \pi/4 + m\pi/2)/\omega_0)$ between the extreme points for the scale corresponding to the geometric average $\sqrt{\hat{s}_1 \, \hat{s}_2}$ of the extreme values

$$\mathcal{Q}_{(x,y),1,\Gamma-norm} L\Big|_{x=\frac{\pi}{4\omega_0}, y=\frac{\pi}{4\omega_0}, s=\sqrt{\hat{s}_1 \, \hat{s}_2}} = \mathcal{Q}_{(x,y),2,\Gamma-norm} L\Big|_{x=\frac{\pi}{4\omega_0}, y=\frac{\pi}{4\omega_0}, s=\sqrt{\hat{s}_1 \, \hat{s}_2}}, \tag{26}$$

then this implies that the relative weighting factor C_s between first- and second-order derivative responses should be chosen as

$$C_s = \frac{1}{\sqrt{(1 - \Gamma_s)(2 - \Gamma_s)}}. \tag{27}$$

2.3. Phase-compensated scale estimate. Given this understanding of how the scale estimates depend on the local phase of a sine wave, we can define a *phase-compensated scale estimate* according to either

$$\begin{aligned}\hat{s}_{\mathcal{Q}L,comp} =& \frac{\sqrt{(1-\Gamma_s)(2-\Gamma_s)}}{(\mathcal{Q}_{(x,y),1,\Gamma-norm}L + \mathcal{Q}_{(x,y),2,\Gamma-norm}L)} \\ & \times \left(\frac{\mathcal{Q}_{(x,y),1,\Gamma-norm}L}{1-\Gamma_s} + \frac{\mathcal{Q}_{(x,y),2,\Gamma-norm}L}{2-\Gamma_s} \right) \hat{s}_{\mathcal{Q}L} \end{aligned} \tag{28}$$

or

$$\begin{aligned}&\hat{s}_{\mathcal{Q}L,comp} \\ &= \frac{\sqrt{(1-\Gamma_s)(2-\Gamma_s)}}{(1-\Gamma_s)^{\frac{\mathcal{Q}_{(x,y),1,\Gamma-norm}L}{\mathcal{Q}_{(x,y),1,\Gamma-norm}L+\mathcal{Q}_{(x,y),2,\Gamma-norm}L}} (2-\Gamma_s)^{\frac{\mathcal{Q}_{(x,y),2,\Gamma-norm}L}{\mathcal{Q}_{(x,y),1,\Gamma-norm}L+\mathcal{Q}_{(x,y),2,\Gamma-norm}L}}} \hat{s}_{\mathcal{Q}L}.\end{aligned} \tag{29}$$

These expressions are defined to be equal to the geometric average

$$\hat{s}_{geom} = \sqrt{\hat{s}_1\, \hat{s}_2} = \frac{\sqrt{(1-\Gamma_s)(2-\Gamma_s)}}{\omega_0^2} \tag{30}$$

of the extreme values when only one of the first- and second-order components in $\mathcal{Q}_{\Gamma-norm}L$ responds and using blending of these responses by transfinite interpolation [13] using the relative strengths of the first- and second-order responses, respectively, to achieve a much lower variability of the scale estimates in between (see Figure 3).

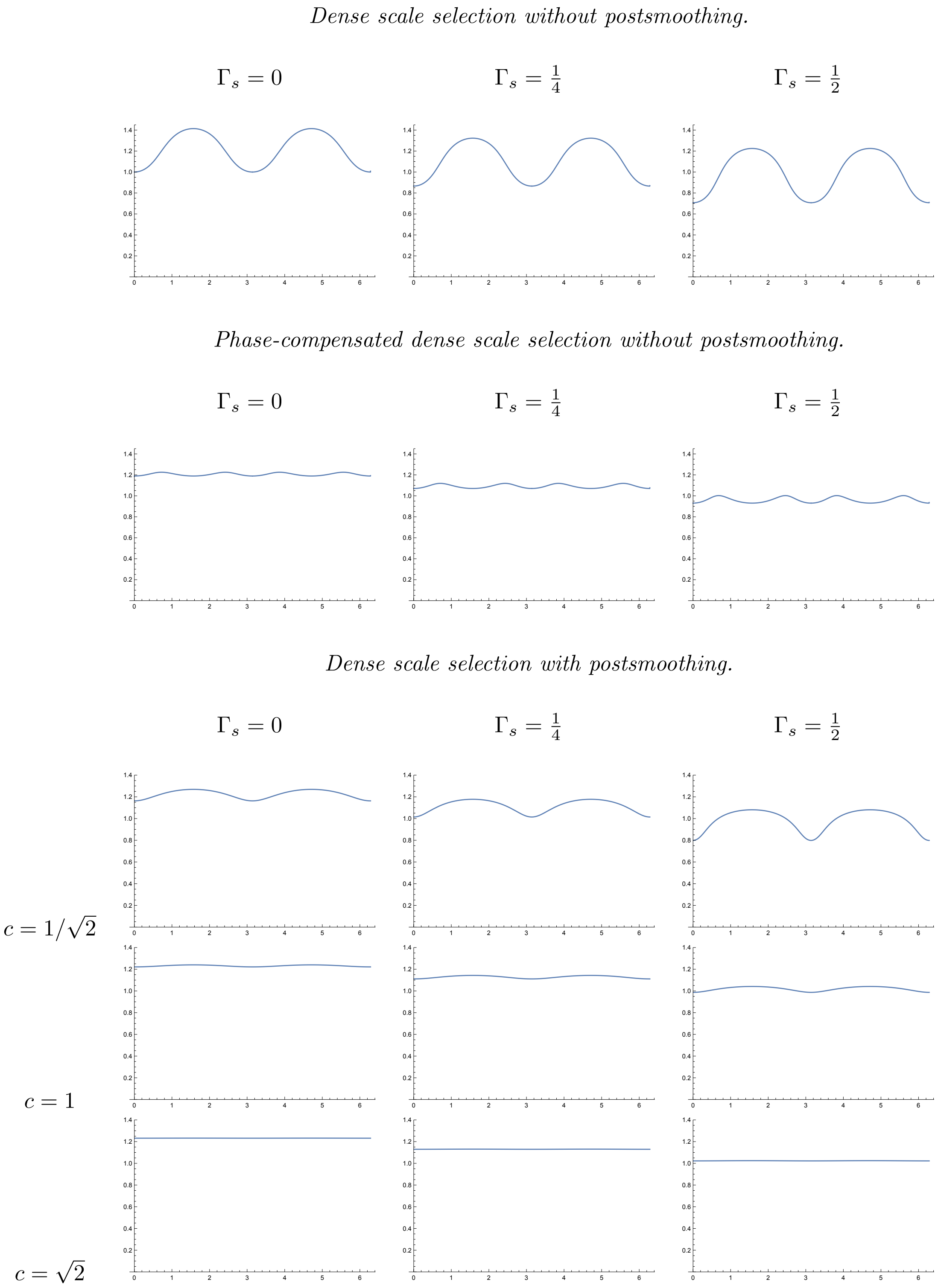
The motivation for the definitions of (28) and (29) is to express interpolation functions on a simple form that compensate for the phase dependency of the scale estimates depending on the relative strengths of the first- and second-order components. When only the first-order component responds and the second-order component is zero, the scale estimate is $\hat{s}_1 = (1-\Gamma_s)/\omega_0^2$. When only the second-order component responds and the first-order component is zero, the scale estimate is $\hat{s}_2 = (2-\Gamma_s)/\omega_0^2$.

In the first expression (28), the ratios $w_1 = \mathcal{Q}_{(x,y),1,\Gamma-norm}L/(\mathcal{Q}_{(x,y),1,\Gamma-norm}L + \mathcal{Q}_{(x,y),2,\Gamma-norm})$ and $w_2 = \mathcal{Q}_{(x,y),2,\Gamma-norm}L/(\mathcal{Q}_{(x,y),1,\Gamma-norm}L + \mathcal{Q}_{(x,y),2,\Gamma-norm})$, which are positive and sum to one, $w_1 + w_2 = 1$, are used as relative weights in a linear convex combination

$$\hat{s}_{comp} = s_{geom} \left(w_1 \frac{\hat{s}}{\hat{s}_1} + w_2 \frac{\hat{s}}{\hat{s}_2} \right) \tag{31}$$

defined such that $\hat{s}_{comp} = s_{geom}$ if either $(w_1 = 1, w_2 = 0, \hat{s} = \hat{s}_1)$ or $(w_1 = 0, w_2 = 1, \hat{s} = \hat{s}_2)$. In the second expression (29), the same ratios w_1 and w_2 are used as relative weights in a geometric convex combination

$$\hat{s}_{comp} = s_{geom} \left(\frac{\hat{s}}{\hat{s}_1} \right)^{w_1} \left(\frac{\hat{s}}{\hat{s}_2} \right)^{w_2}, \tag{32}$$



Dense scale selection without postsmoothing.

$\Gamma_s = 0$ $\Gamma_s = \frac{1}{4}$ $\Gamma_s = \frac{1}{2}$

Phase-compensated dense scale selection without postsmoothing.

$\Gamma_s = 0$ $\Gamma_s = \frac{1}{4}$ $\Gamma_s = \frac{1}{2}$

Dense scale selection with postsmoothing.

$\Gamma_s = 0$ $\Gamma_s = \frac{1}{4}$ $\Gamma_s = \frac{1}{2}$

$c = 1/\sqrt{2}$

$c = 1$

$c = \sqrt{2}$

Figure 3. *Spatial variability of the local scale estimates in units of $\sigma_s = \sqrt{s}$ for a 1-D sine wave $f(x) = \sin \omega_0 x$ with angular frequency $\omega_0 = 1$. Top row: regular scale estimates without postsmoothing according to (14). Second row: phase-compensated scale estimates on a logarithmic scale and without postsmoothing according to (29). Third, fourth, and bottom rows: scale estimates computed from local extrema over scale of the postsmoothed quasi quadrature entity $\overline{\mathcal{Q}}_{(x,y),\Gamma-norm} L$ according to (15) for relative postsmoothing scales $c = 1/\sqrt{2}$, $c = 1$, and $c = \sqrt{2}$, respectively. (All results have been computed using $C_s = 1/\sqrt{(1-\Gamma_s)(2-\Gamma_s)}$.)*

again defined such that $\hat{s}_{comp} = s_{geom}$ if either $(w_1 = 1, w_2 = 0, \hat{s} = \hat{s}_1)$ or $(w_1 = 0, w_2 = 1, \hat{s} = \hat{s}_2)$.

For the first expression (28), the blending is thus performed on a linear scale with respect to the spatial scale parameter, whereas in the second expression (29) the blending is performed on a more natural logarithmic scale.

From these spatial scale estimates, we can in turn estimate the temporal wavelength of the sine wave according to

$$\hat{\lambda} = \frac{2\pi\sqrt{\hat{s}_{\mathcal{Q}_L,comp}}}{\sqrt[4]{(1-\Gamma_s)(2-\Gamma_s)}}. \tag{33}$$

Although the interpolation functions (28) and (29) in combination with (33) will not lead to exact wavelength estimates for all phases of a sine wave, these functions compensate for the gross behavior of the phase dependency, which substantially decreases the otherwise much higher spatial variability in the spatial scale estimates.

2.4. Scale calibration. In sections 2.2–2.3, as well as in a more detailed analysis in the supplementary material, it is shown how the scale estimates $\hat{s}_{\mathcal{Q}L}$ and $\hat{s}_{\overline{\mathcal{Q}}L}$ according to (14) and (15) are influenced by the parameters Γ_s, c, and C_s in the quasi quadrature measure. To decouple this dependency from the later stage visual modules for which the dense scale selection methodology is intended to be used as an initial preprocessing stage, we introduce the notion of *scale calibration*, which implies that the scale estimates are to be multiplied by uniform scaling factors such that they are either

(i) equal to the scale estimate $\hat{s} = s_0$ obtained by applying the regular scale-normalized Laplacian $\nabla^2_{norm} L = s\,(L_{xx} + L_{yy})$ or the scale-normalized determinant of the Hessian $\det \mathcal{H}_{norm} L = s^2\,(L_{xx}L_{yy} - L_{xy}^2)$ at the center of a Gaussian blob of any spatial extent s_0, or

(ii) equal to the scale estimate $\hat{s} = \sqrt{2}/\omega_0^2$ corresponding to the geometric average of the scale estimates obtained for a sine wave of any angular frequency ω_0 when $\Gamma_s = 0$.

The first method, which aims at similarity with previous scale selection methods at sparse image features, will be referred to as *Gaussian scale calibration*, whereas the second method, which aims at similarity for dense texture patterns, will be referred to as *sine wave scale calibration*.

For the case of either (i) no phase compensation and no postsmoothing, or (ii) phase compensation without postsmoothing, the necessary calibration factors can be obtained from the theoretical results in section 2.2. Ways of deriving the scale calibration factors when using spatial postsmoothing are described in the supplementary material.

2.5. Composed dense scale selection algorithms. Given the above treatment, we can define four types of dense scale selection algorithms as follows:

Algorithm I: without postsmoothing or phase compensation, with local scale estimates at every image point computed according to (14).

Algorithm II: with phase compensation and without postsmoothing, with local scale estimates at every image point computed according to (28) or (29) based on uncompensated local scale estimates according to (14).

Algorithm III: with postsmoothing and without phase compensation, with local scale estimates at every image point computed according to (15).

Algorithm IV: with both postsmoothing and phase compensation, with local phase-compensated scale estimates at every image point computed in an analogous way as (28) or (29), although based on postsmoothing according to (15) and with the factors $(1 - \Gamma_s)$ and $(2 - \Gamma_s)$ in the expressions for phase compensation, which originate from the local extrema over scale when $c = 0$, replaced by $S_{sine,1}(\Gamma_s, c, C_s)$ and $S_{sine,2}(\Gamma_s, c, C_s)$ according to (SM5) in the supplementary material.

The scale estimates from each algorithm can in turn be calibrated using either Gaussian scale calibration or sine wave calibration according to section 2.4.

2.6. Scale covariance of the spatial scale estimates under spatial scaling transformations.

Consider a scaling transformation of the spatial image domain

$$f'(x_1', x_2') = f(x_1, x_2) \qquad \text{for} \qquad (x_1', x_2') = (S_s\, x_1, S_s\, x_2), \tag{34}$$

where S_s denotes the spatial scaling factor. Define the spatial scale-space representations L and L' of f and f', respectively, according to

$$L(x_1, x_2;\ s) = (T(\cdot, \cdot;\ s) * f(\cdot, \cdot,))\,(x_1, x_2;\ s), \tag{35}$$

$$L'(x_1', x_2';\ s') = \big(T(\cdot, \cdot;\ s') * f'(\cdot, \cdot)\big)\,(x_1', x_2';\ s'). \tag{36}$$

Consider a spatial differential expression of the form

$$\mathcal{D}L = \sum_{i=1}^{I} \prod_{j=1}^{J} c_i\, L_{x^{\alpha_{ij}}} = \sum_{i=1}^{I} \prod_{j=1}^{J} c_i\, L_{x_1^{\alpha_{1ij}} x_2^{\alpha_{2ij}}}, \tag{37}$$

required to be *homogeneous* in the sense that the sum of the orders of differentiation in each term does not depend on the index of that term,

$$\sum_{j=1}^{J} |\alpha_{ij}| = \sum_{j=1}^{J} \alpha_{1ij} + \alpha_{2ij} = M. \tag{38}$$

Then, the corresponding homogeneous differential expression $\mathcal{D}_{norm}L$ with the spatial derivatives $L_{x_1^{m_1} x_2^{m_2}}$ replaced by scale-normalized derivatives according to

$$L_{\xi_1^{m_1} \xi_2^{m_2}} = s^{(m_1+m_2)\gamma_s/2}\, L_{x_1^{m_1} x_2^{m_2}} \tag{39}$$

transforms, according to (Lindeberg [35, eq. (25)]), into

$$\mathcal{D}'_{norm}L' = S_s^{M(\gamma_s - 1)}\, \mathcal{D}_{norm}L. \tag{40}$$

Regarding the spatial quasi quadrature measure $\mathcal{Q}_{(x,y),\Gamma-norm}L$, according to (8), which we use for dense spatial scale selection, this differential invariant is not of the homogeneous form

(37). If we split this differential expression into two components based on the orders of spatial differentiation

$$\mathcal{Q}_{(x,y),\Gamma-norm} L = \mathcal{Q}_{(x,y),1,\Gamma-norm} L + \mathcal{Q}_{(x,y),2,\Gamma-norm} L, \tag{41}$$

where

$$\mathcal{Q}_{(x,y),1,\Gamma-norm} L = \frac{s\,(L_x^2 + L_y^2)}{s^{\Gamma_s}}, \tag{42}$$

$$\mathcal{Q}_{(x,y),2,\Gamma-norm} L = \frac{C_s\, s^2\, \left(L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2\right)}{s^{\Gamma_s}}, \tag{43}$$

we can note that each of these expressions is of the homogeneous form (37) and corresponds to γ-normalized scale-space derivatives in the respective cases:

$$\gamma_1 = 1 - \Gamma_s, \qquad \gamma_2 = 1 - \frac{\Gamma_s}{2}. \tag{44}$$

Applying the transformation property (40) to each of the two components of the spatial quasi quadrature measure then gives that they transform according to

$$\mathcal{Q}_{(x',y'),1,\Gamma-norm} L' = S_s^{2(\gamma_1-1)}\, \mathcal{Q}_{(x,y),1,\Gamma-norm} L = S_s^{-2\Gamma_s}\, \mathcal{Q}_{(x,y),1,\Gamma-norm} L, \tag{45}$$

$$\mathcal{Q}_{(x',y'),2,\Gamma-norm} L' = S_s^{2\times 2(\gamma_2-2)}\, \mathcal{Q}_{(x,y),2,\Gamma-norm} L = S_s^{-2\Gamma_s}\, \mathcal{Q}_{(x,y),2,\Gamma-norm} L. \tag{46}$$

In other words, because of the deliberate adding of differential expressions corresponding to different orders of spatial differentiation for the maximally scale-invariant case of $\gamma_s = 1$ prior to postnormalization by the postnormalization power Γ_s, it follows that the two components transform in the same way under spatial scaling transformations, implying that the composed quasi quadrature measure transforms as

$$(\mathcal{Q}_{(x',y'),\Gamma-norm} L')(x', y';\ s') = S_s^{-2\Gamma_s}\, (\mathcal{Q}_{(x,y),\Gamma-norm} L)(x, y;\ s) \tag{47}$$

under uniform scaling transformations of the spatial image domain.

This covariance property under spatial scaling transformations specifically implies that local extrema over spatial scales are preserved under uniform scaling transformations of the spatial image domain and are transformed in a scale-covariant way according to

$$\hat{s}' = S_s^2\, \hat{s} \tag{48}$$

or, in units of the standard deviation $\sigma_s = \sqrt{s}$ of the spatial scale-space kernel, according to

$$\hat{\sigma}'_s = S_s\, \hat{\sigma}_s. \tag{49}$$

This property constitutes the theoretical foundation for dense spatial scale selection and implies that the local spatial scale estimates will automatically adapt to local variations in the dominant spatial scales in the image data.

This scale covariance of the spatial scale estimates also extends to phase-compensated scale estimates according to (28) and (29). This property is straightforward to prove, since the underlying uncompensated spatial scale estimates $\hat{s}_{QL}$ in (28) and (29) are provably scale covariant, and additionally the ratio that determines the scale compensation factor is invariant under independent scaling transformations of the spatial domain provided that the spatial scale levels are appropriately matched, which they are if the phase compensation factors are computed at scale levels corresponding to the spatial scale estimates.

Correspondingly, the scale covariance of the spatial scale estimates also extends to spatial postsmoothing prior to the detection of local extrema over spatial scales. This property follows from the fact that the amount of spatial postsmoothing is proportional to the spatial scale level at which the nonlinear quasi quadrature measure is computed.

Under affine intensity transformations

$$f'(x, y) = a\, f(x, y) + b, \tag{50}$$

the Gaussian derivatives are multiplied by a uniform scaling factor $L'_{x^\alpha y^\beta}(x, y;\, t) = a\, L_{x^\alpha y^\beta}(x, y;\, t)$ and the quasi quadrature measure transforms according to

$$\mathcal{Q}'_{(x,y),\Gamma-norm}(x, y;\ s) = a^2\, \mathcal{Q}_{(x,y),\Gamma-norm}(x, y;\ s). \tag{51}$$

The scale estimates are therefore unaffected by illumination variations, whose effects can be well approximated by local affine transformations over the intensity domain.

The spatial quasi quadrature entities used for scale selection are based on the rotationally invariant differential invariants $|\nabla L|^2$ and $\|\mathcal{H}L\|^2$ and are therefore rotationally invariant. This implies that the resulting scale spatial estimates are covariant under rotations of the spatial image domain.

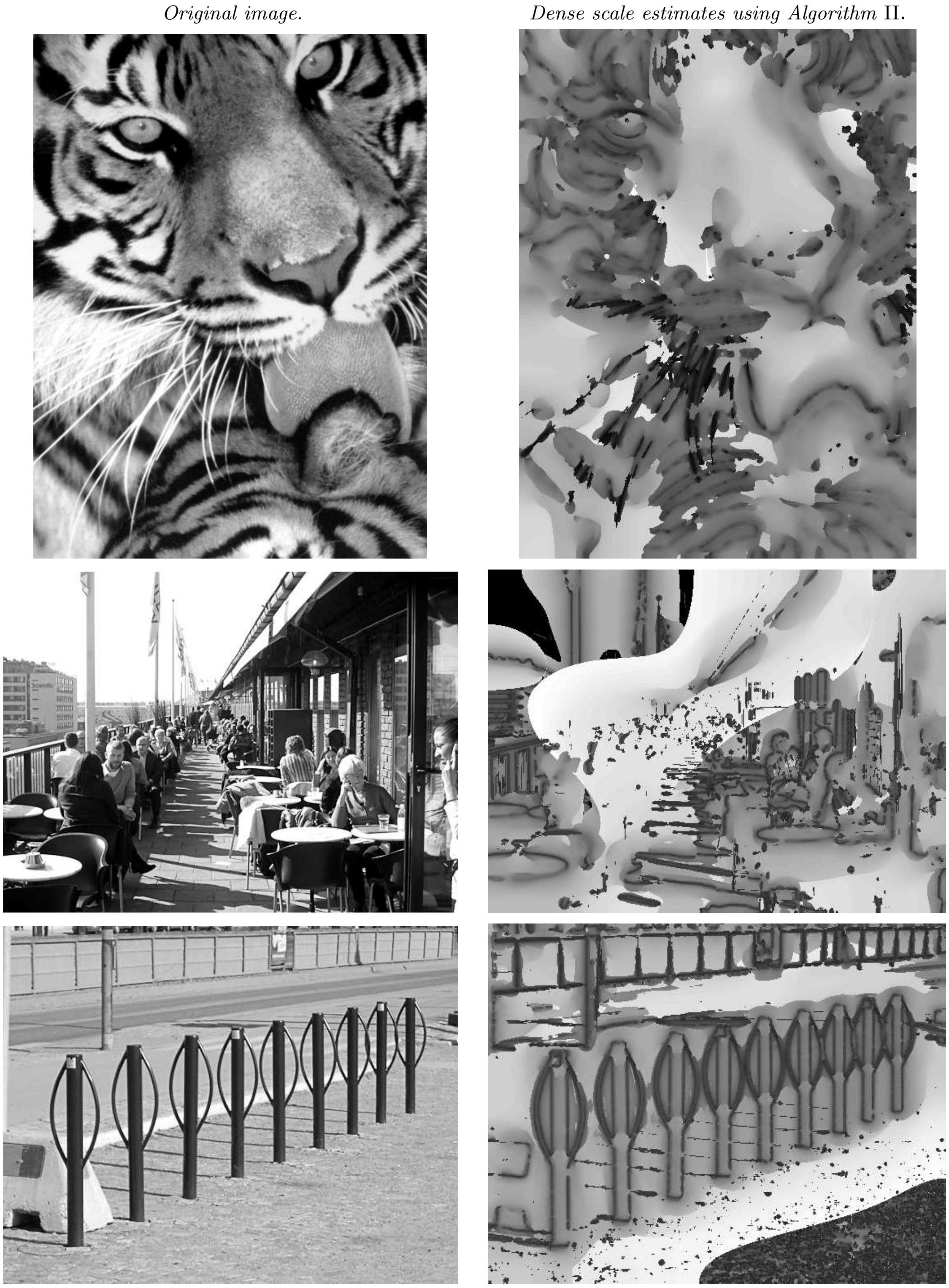
2.7. Experimental results. Figure 4 shows spatial scale maps computed using Algorithm II for three images. All the scale maps have been computed using Gaussian scale calibration for $\Gamma_s = 1/4$. For all images in this illustration, we can note how finer scale estimates are selected near edges, leading to a sketch-like representation of prominent edges. This behavior is in good agreement with the theory and also implies that image features or image descriptors that are computed with this type of scale selection methodology will be well localized near edges.

In general, the detection of local extrema over scale of the quasi quadrature entity $\mathcal{Q}_{(x,y),\Gamma-norm}L$ will sweep out *scale selection surfaces* defined by

$$\begin{cases} \partial_s(\mathcal{Q}_{(x,y),\Gamma-norm}L) = 0, \\ \partial_{ss}(\mathcal{Q}_{(x,y),\Gamma-norm}L) < 0 \end{cases} \tag{52}$$

in the 3-D scale space spanned by the spatial dimensions (x, y) and the scale parameter s. Specifically, when the local image structures contain different types of structures at different scales, multiple local extrema over scale may be detected, corresponding to multiple patches of the scale selection surfaces at different scales, which may represent qualitatively different types of image structures in the image domain while at different scales.

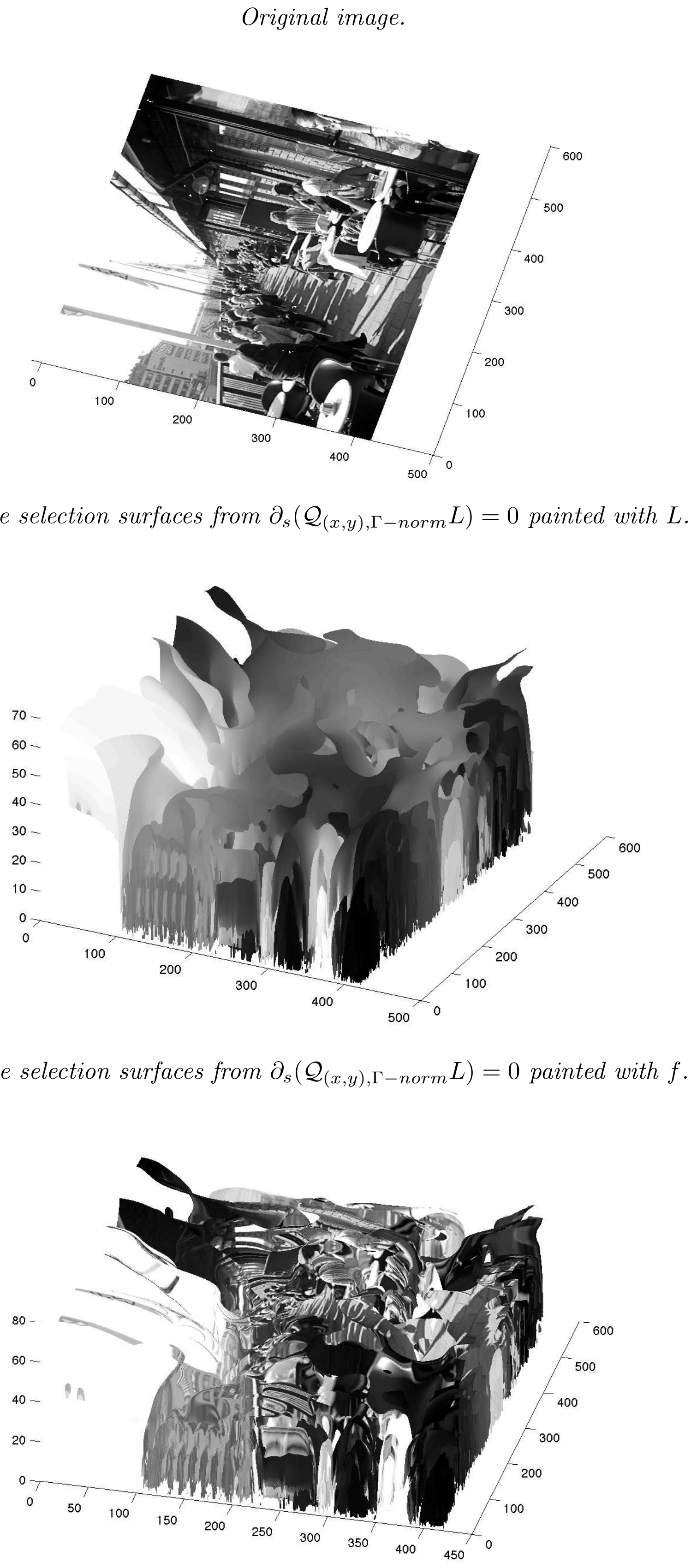
In Figure 4, a much simplified form of visualization has been used, by only showing the scale value of the local maximum over scale that has the maximum response among the possibly multiple local maxima over scale. When moving between different points over the



Original image.

Dense scale estimates using Algorithm II.

Figure 4. *Dense spatial scale maps computed using Algorithm* II *(dense scale selection with phase compensation) for three different images using* $\Gamma_s = 1/4$ *and* $C_s = 1/\sqrt{(1-\Gamma_s)(2-\Gamma_s)}$*. The grey-levels code for effective scale approximated by* $s_{eff} = \log_2(s_0 + s)$ *for* $s_0 = 1/8$ *in such a way that darker means finer spatial scales and brighter indicates coarser spatial scales. See section* 2.7 *for more detailed explanations. Observe, however, that for the quite common phenomenon of when there are multiple local extrema over scale corresponding to different types of dominant structures at different scales, this visualization only shows the single extremum of the scale estimates that has the strongest maximum response. Thereby, situations where the maximum value over scales switches between two scale selection surfaces at difference scales appear as discontinuities in this simplified form of visualization. Such layer discontinuities are therefore artefacts of the visualization method, not the scale selection method. A more appropriate form of visualization is in terms of a 3-D visualization of the scale selection surfaces as shown in Figure* 5*, where multiple scale estimates may be displayed at every image point.*


Original image.

Scale selection surfaces from $\partial_s(\mathcal{Q}_{(x,y),\Gamma-norm}L) = 0$ painted with L.

Scale selection surfaces from $\partial_s(\mathcal{Q}_{(x,y),\Gamma-norm}L) = 0$ painted with f.

Figure 5. *3-D visualization of local scale estimates underlying the dense scale maps shown in Figure 4(right) and shown here as scale selection surfaces in 3-D scale space over space (x, y) and scale s, here also visualizing multiple local extrema over scales at every image point. (Scale values in units of effective scale approximated by $s_{eff} = \log_2(s_0 + s)$ for $s_0 = 1/8$.)*

spatial domain, this global maximum may in some places switch between different patches of the scale selection surfaces at different scales. The discontinuities in the scale maps that can be seen in the right column correspond to such switching between multiple scale selection surfaces and are artefacts of the visualization method, not the scale selection method.

More generally, one should of course treat multiple local extrema over scale as multiple scale hypotheses as done in the more appropriate 3-D visualization of such multiple scale selection surfaces in Figure 5. From a more detailed inspection of the scale selection surface patch corresponding to the edge of the roof in the upper part of the image, one can also find that the selected scale levels decrease with increasing distance from the camera as caused by perspective effects. The similarities between the scale maps for the repetitive structures in the image in the lower part of the figure demonstrate the stability of the dense scale estimates under natural imaging conditions, whereas the relative differences in scale estimates between corresponding parts of the different yet similar pillars reveal the size gradient caused by perspective scaling effects.

To quantify the numerical stability of the scale estimates for a stimulus for which the scale estimates should be approximately constant, we computed dense scale estimates for a set of 2-D sine waves of the form

$$f(x,y) = \sin\omega x + \sin\omega y, \tag{53}$$

with wavelengths $\lambda = 8, 16, 32$, and 64 for each of the four types of algorithms and compared the results with corresponding theoretical predictions based on the scale selection properties of the 1-D sine wave model (24)

$$\hat{\omega} = \frac{\sqrt[4]{(1-\Gamma_s)(2-\Gamma_s)}}{\sqrt{\hat{s}}}. \tag{54}$$

The mean and the standard deviation around the mean were computed for the scale estimates in terms of effective scale s_{eff}, and these measures were transformed into relative measures in units of the scale parameter $\sigma_s = \sqrt{s}$ of dimension [length]. As can be seen from the results in Table 1, the use of phase compensation and complementary postsmoothing substantially decreases the spatial variability in the scale estimates by an order of magnitude in units of σ_s. Specifically, pure phase compensation achieves a reduction in the variability of the same order as pure postsmoothing for $c = 1$, while also allowing for a higher resolution in the scale estimates near edge-like structures.

Table 1

Measures of the accuracy of the scale estimates computed for a 2-D sine wave $f(x,y) = \sin\omega x + \sin\omega y$ and compared to corresponding theoretical predictions based on a 1-D sine wave model according to (33) for Algorithms I–IV using $\Gamma_s = 1/4$, $c = 1$, and $C_s = 1/\sqrt{(1-\Gamma_s)(2-\Gamma_s)}$.

Accuracy of scale estimates for a 2-D sine wave pattern.				
Error measure	Alg. I	Alg. II	Alg. III	Alg. IV
Offset of mean	+ 5.0 %	– 0.6 %	+ 1.6 %	+ 1.5 %
Relative spread	± 11.8 %	± 1.3 %	± 0.6 %	± 0.1 %

3. Dense temporal scale selection over a purely temporal domain. In this section, we develop a corresponding approach for dense scale selection over a purely temporal domain.

3.1. A temporal quasi quadrature measure. Motivated by the fact that first-order derivatives respond primarily to the locally odd component of a signal, whereas second-order derivatives respond primarily to the locally even component of a signal, for dense applications it is natural to aim at a feature detector that combines such first- and second-order temporal derivative responses. By specifically combining the squares of the first- and second-order temporal derivative responses in an additive way, we obtain a temporal quasi quadrature measure of the form

$$\mathcal{Q}_{t,\Gamma-norm} L = \frac{\tau L_t^2 + C_\tau \tau^2 L_{tt}^2}{\tau^{\Gamma_\tau}}, \tag{55}$$

which is a reduction of the 2-D spatial quasi quadrature measure (8) to a 1-D purely temporal domain, where t denotes time and τ temporal scale.

This construction is closely related to a proposal by Koenderink and van Doorn [29] of summing up the squares of the first- and second-order derivative responses of receptive fields and an observation by De Valois et al. [65] that first- and second-order biological receptive fields typically occur in pairs that can be modeled as approximate Hilbert pairs, while here they are instead formalized in terms of scale-normalized temporal scale-space derivatives.

3.2. Scale covariance of temporal scale estimates under temporal scaling transformations. Consider a temporal scaling transformation of the form

$$f'(t') = f(t) \qquad \text{for} \qquad t' = S_\tau \, t. \tag{56}$$

From the temporal scaling transformation of scale-normalized temporal derivatives defined from either the noncausal Gaussian temporal scale space or the time-causal temporal scale space obtained by convolution with the time-causal limit kernel [43], it follows that scale-normalized temporal derivatives of order n are transformed, according to [46, eqs. (10) and (104)], as

$$\partial_{\zeta'^n,norm} L'(t';\ \tau') = S_\tau^{n(\gamma_\tau - 1)} \, \partial_{\zeta^n,norm} L(t;\ \tau) \tag{57}$$

provided that the temporal scale levels are correspondingly matched according to

$$\tau' = S_\tau^2 \, \tau. \tag{58}$$

Applied to the temporal quasi quadrature measure

$$\mathcal{Q}_{t,\Gamma-norm} L = \frac{\tau L_t^2 + C_\tau \tau^2 L_{tt}^2}{\tau^{\Gamma_\tau}} = \mathcal{Q}_{t,1,\Gamma-norm} L + \mathcal{Q}_{t,2,\Gamma-norm} L, \tag{59}$$

with its first- and second-order components

$$\mathcal{Q}_{t,1,\Gamma-norm} L = \frac{\tau L_t^2}{\tau^{\Gamma_\tau}}, \qquad \mathcal{Q}_{t,2,\Gamma-norm} L = \frac{C_\tau \tau^2 L_{tt}^2}{\tau^{\Gamma_\tau}}, \tag{60}$$

and which correspond to γ-normalized temporal derivatives with $\gamma_1 = 1 - \Gamma_\tau$ and $\gamma_2 = 1 - \Gamma_\tau/2$ for the first- and second-order components, respectively, the first- and second-order components transform according to

$$\begin{aligned}(\mathcal{Q}_{t,1,\Gamma-norm} L')(t';\ \tau') &= S_\tau^{2(\gamma_1 - 1)}\,(\mathcal{Q}_{t,1,\Gamma-norm} L)(t;\ \tau)\\ &= S_\tau^{-2\Gamma_\tau}\,(\mathcal{Q}_{t,1,\Gamma-norm} L)(t;\ \tau),\end{aligned} \tag{61}$$

$$\begin{aligned}(\mathcal{Q}_{t,2,\Gamma-norm} L')(t';\ \tau') &= S_\tau^{2\times 2(\gamma_2 - 1)}\,(\mathcal{Q}_{t,2,\Gamma-norm} L)(t;\ \tau)\\ &= S_\tau^{-2\Gamma_\tau}\,(\mathcal{Q}_{t,2,\Gamma-norm} L)(t;\ \tau).\end{aligned} \tag{62}$$

Since the first- and second-order components transform in a similar way because of the deliberate adding of entities depending on temporal derivatives of different orders for the maximally scale-invariant choice of $\gamma_\tau = 1$, it follows that the temporal quasi quadrature measure, despite its inhomogeneity, still transforms according to a power law

$$(\mathcal{Q}_{t,\Gamma-norm} L')(t';\ \tau') = S_\tau^{-2\Gamma_\tau}\,(\mathcal{Q}_{t,\Gamma-norm} L)(t;\ \tau). \tag{63}$$

Specifically, this implies that temporal scale estimates computed from local extrema over temporal scales are preserved and are transformed in a scale-covariant way according to

$$\hat{\tau}' = S_\tau^2\,\hat{\tau} \tag{64}$$

or, in units of the standard deviation $\sigma_\tau = \sqrt{\tau}$ of the temporal scale-space kernel,

$$\hat{\sigma}_\tau' = S_\tau\,\hat{\sigma}_\tau. \tag{65}$$

This does in turn imply that the temporal scale estimates will adapt to local temporal scaling transformations of the input signal, and constitutes the theoretical basis for the dense temporal scale selection methodology.

This temporal scale covariance property also extends to phase-compensated temporal scale estimates according to (29)

$$\hat{\tau}_{\mathcal{Q}_t,comp} = \frac{\sqrt{(1-\Gamma_\tau)(2-\Gamma_\tau)}\,\hat{\tau}_{\mathcal{Q}_t}}{(1-\Gamma_\tau)^{\frac{\mathcal{Q}_{t,1,\Gamma-norm}L}{\mathcal{Q}_{t,1,\Gamma-norm}L+\mathcal{Q}_{t,2,\Gamma-norm}L}}\,(2-\Gamma_\tau)^{\frac{\mathcal{Q}_{t,2,\Gamma-norm}L}{\mathcal{Q}_{t,1,\Gamma-norm}L+\mathcal{Q}_{t,2,\Gamma-norm}L}}}, \tag{66}$$

since the underlying uncompensated temporal scale estimates $\hat{\tau}_{\mathcal{Q}_t}$ transform in a scale-covariant way, and the ratios $w_1 = \mathcal{Q}_{t,1,\Gamma-norm}L/(\mathcal{Q}_{t,1,\Gamma-norm}L + \mathcal{Q}_{t,2,\Gamma-norm}L)$ and $w_2 = \mathcal{Q}_{t,2,\Gamma-norm}L/(\mathcal{Q}_{t,1,\Gamma-norm}L + \mathcal{Q}_{t,2,\Gamma-norm}L)$ that determine the scale compensation factors are invariant under temporal scaling transformations, provided that the temporal scale levels are appropriately matched.

The temporal scale covariance of the temporal scale estimates is also preserved under temporal postsmoothing, since the amount of temporal postsmoothing is proportional to the local temporal scale for computing the temporal derivatives.

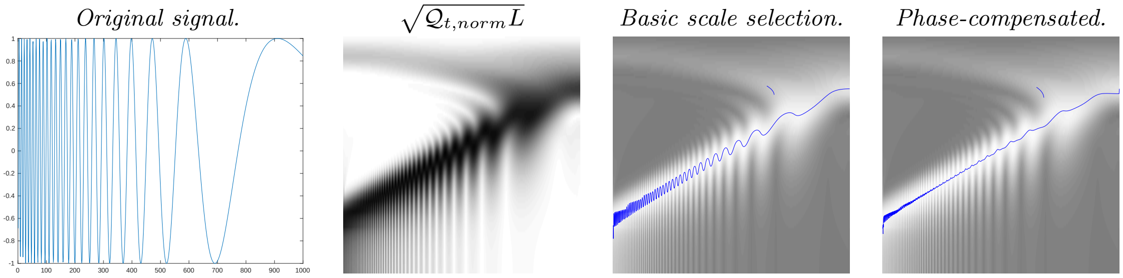

Original signal. $\sqrt{\mathcal{Q}_{t,norm} L}$ *Basic scale selection.* *Phase-compensated.*

Figure 6. *Dense temporal scale selection from local extrema over scale of the temporal quasi quadrature measure $\mathcal{Q}_{t,\Gamma-norm} L$ applied to a synthetic sine wave signal $f(t) = \sin(\exp((b-t)/a))$ for $a = 200$ and $b = 1000$ with temporally varying frequency so that the wavelength increases with time t. Left: Original temporal signal. Middle left: The square root of the temporal quasi quadrature measure $\mathcal{Q}_{t,\Gamma-norm} L$ computed for the most scale-invariant choice of the complementary scale normalization parameter $\Gamma_\tau = 0$. Middle right: Basic scale estimates $\hat{\tau}_{\mathcal{Q}_{t,\Gamma-norm}}$ from local extrema over scale of the quasi quadrature measure according to (14) reduced to one dimension for $\Gamma_\tau = 0$ and shown as blue curves overlaid on the magnitude map $\mathcal{Q}_{t,\Gamma-norm} L$ with reversed contrast. Right: Phase-compensated scale estimates $\hat{\tau}_{\mathcal{Q}_{t,comp}}$ according to (29) reduced to one dimension for $\Gamma_\tau = 0$ and shown as blue curves overlaid on the magnitude map $\mathcal{Q}_{t,\Gamma-norm} L$ with reversed contrast. All results have been computed using $C_\tau = 1/\sqrt{(1-\Gamma_\tau)(2-\Gamma_\tau)}$. (Horizontal axis: Time $t \in [0, 1000]$; vertical axis in columns 2–4: Effective temporal scale $\tau_{eff} = \log \tau$ over the range from $\sigma_{\tau,min} = 0.25$ to $\sigma_{\tau,max} = 1000$ for $\sigma_\tau = \sqrt{\tau}$.)*

3.3. Experimental results.

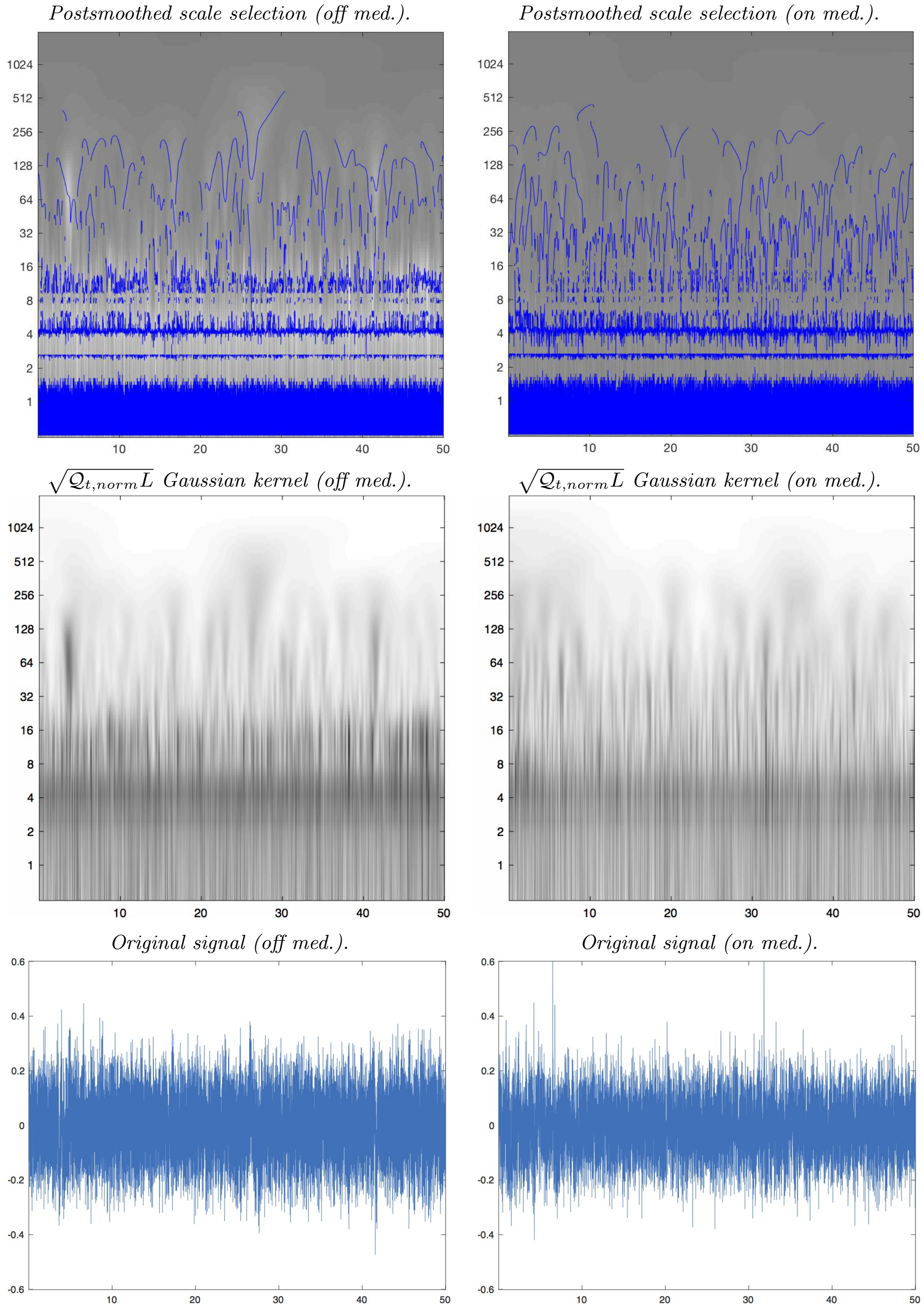
3.3.1. Sine wave with exponentially varying frequency. Figure 6 shows the result of applying this basic form of dense temporal scale selection to a sine wave with exponentially varying frequency of the form

$$f(t) = \sin\left(\exp\left(\frac{(b-t)}{a}\right)\right) \tag{67}$$

for $a = 200$ and $b = 1000$. The left figure shows the raw temporal signal. The middle left figure shows the magnitude map over time and temporal scales of the temporal quasi quadrature measure $\mathcal{Q}_{t,norm} L$ computed using a noncausal Gaussian temporal scale-space representation. The middle right figure shows temporal scale estimates computed as the zero-crossings of $\partial_\tau(\mathcal{Q}_{t,\Gamma-norm} L) = 0$ that satisfy the sign condition $\partial_{\tau\tau}(\mathcal{Q}_{t,\Gamma-norm} L) < 0$. These zero-crossings have been interpolated to higher accuracy along the temporal scale dimension than the sampling density over the temporal scales using parabolic interpolation [46, eq. (115)]. In the rightmost figure, the basic temporal estimates from the middle right have been additionally phase-compensated according to (66).

Note how (i) the temporal scale selection method is able to capture the rapid variations in the temporal scales in the signal, and (ii) the phase compensation method substantially suppresses the phase dependency of the temporal scale estimates.

3.3.2. Real measurement signals. Figure 7 shows an example of performing this type of dense temporal scale selection analysis on two real measurements signals using the noncausal Gaussian temporal scale-space concept. The figures in the bottom row show measurements of the local field potential recorded from the subthalamic nucleus of a conscious human subject


Postsmoothed scale selection (off med.).

Postsmoothed scale selection (on med.).

$\sqrt{\mathcal{Q}_{t,norm} L}$ *Gaussian kernel (off med.).*

$\sqrt{\mathcal{Q}_{t,norm} L}$ *Gaussian kernel (on med.).*

Original signal (off med.).

Original signal (on med.).

Figure 7. *Dense local scale analysis of a neurophysiological signal from Cagnan, Duff, and Brown* [7] *showing the local field potential sampled at* $\nu = 512$ Hz *during* 50 *seconds for an electrode inserted in the subthalamic nucleus of a conscious human subject with Parkinson's disease without medication (left column) and with the medication levodopa (L-dopa) (right column). Note how this local scale analysis method reveals that the uppermost stronger band of dense local scale estimates in the top left figure is spread out over a larger scale range in the top right figure as a result of the medication. Top row: Basic scale estimates* $\hat{\tau}_{\mathcal{Q}_{t,\Gamma-norm}}$ *from local extrema over scale of the quasi quadrature measure according to* (15) *reduced to one dimension for* $c = 3$ *and* $\Gamma_\tau = 0$. *Middle row: The square root of the temporal quasi quadrature measure* $\mathcal{Q}_{t,\Gamma-norm} L$ *computed for the most scale-invariant choice of the complementary scale normalization parameter* $\Gamma_\tau = 0$. *Bottom row: Original temporal signal. All results have been computed using* $C_\tau = 1/\sqrt{(1-\Gamma_\tau)(2-\Gamma_\tau)}$. *(Horizontal axis: Time* $t \in [0, 50]$ *in seconds. Vertical axis in top and middle rows: Temporal scale in milliseconds.)*

with Parkinson's disease, with the patient either off or on the medication L-dopa (labeled "off med." or "on med." and shown in the left and the right columns, respectively).

As can be seen from the dense local scale estimates in the top row, the dense scale analysis (beyond a wide band of responses at finer scales up to temporal scale $\sigma_{\tau,0} = 1.4$ ms) returns three rather strong bands of coarser temporal scale estimates when the patient is off medication. These bands are assumed around temporal scales $\sigma_{\tau,1} = 2.6$ ms, $\sigma_{\tau,2} = 4.5$ ms, and $\sigma_{\tau,4} = 11$ ms with a weaker additional band at temporal scale $\sigma_{\tau,3} = 7.9$ ms. According to the approximate expression (33) for the local scale estimate of a sine wave, these scale estimates correspond to frequencies $\nu = 1/\lambda$ around $\nu_0 = 136$ Hz, $\nu_1 = 73$ Hz, $\nu_2 = 42$ Hz, $\nu_3 = 24$ Hz, and $\nu_4 = 18$ Hz.

When the patient is on medication, the uppermost band of coarser scale estimates around $\sigma_{\tau,4} = 11$ ms is replaced by a sparser set of dense local scale estimates over a wider scale range $[\sigma_{\tau,5}, \sigma_{\tau,6}] = [10, 42]$ ms and corresponding to frequencies in the range $[\nu_6, \nu_5] = [5, 19]$ Hz.

Comparing the results for the patient off vs. on medication, we see that there is also a weaker band of responses over the scale range between $\sigma_{\tau,7} = 44$ ms and $\sigma_{\tau,8} = 260$ ms corresponding to a frequency range between $\nu_7 = 5$ Hz and $\nu_8 = 0.75$ Hz when the patient is off medication, and responses are not as strong in this band when the patient is on medication.

The biological background for this signal analysis problem is that in Parkinson's disease (PD), several prominent rhythms appear in the local field potentials. Among these, the low-frequency ($\sim$5 Hz) rhythms are associated with tremors observed in PD patients. Next, the so-called beta band (15–30 Hz) rhythms are causally related to many motor deficits associated with PD (Hammond, Bergman, and Brown [17]). Recent analysis of local field potentials in the subthalamic nucleus using Fourier analysis revealed that beta band oscillations are not persistent and instead occur in bursts (Tinkhauser et al. [63]). Moreover, administration of L-dopa medication was shown to reduce the frequency of beta bursts; in particular, the long beta bursts are significantly reduced. Indeed, quenching of the beta band oscillations is one of the goals of PD treatment. Specifically, modulation of beta band oscillations can form a basis for an event-triggered deep-brain-stimulation system (Rosin et al. [55]). To that end, however, it is important to correctly isolate the occurrences of beta band activity. Conventional methods based on Fourier transforms, however, have not been found to be very precise for this purpose.

In relation to this biological background, our temporal scale analysis thus reveals how medication by L-dopa affects the temporal dynamics of neurons in bands at multiple scales that are related to the tremors observed in PD patients. Specifically, it shows how the responses in the band around $\nu_4 = 18$ Hz related to pathology are reduced and spread out by the L-dopa medication and that the responses in the band with frequencies below $\nu_7 = 5$ Hz related to tremors are weaker.

4. Dense spatio-temporal scale selection over the joint spatio-temporal domain.

In this section, we shall combine the mechanisms for dense spatial scale selection and dense temporal scale selection developed in sections 2 and 3 to design a mechanism for dense simultaneous selection of spatial and temporal scales over the joint spatio-temporal domain.

4.1. Spatio-temporal scale-space representation.

The context that we initially consider for dense spatio-temporal scale selection is a space-time separable spatio-temporal scale-space representation $L(x, y, t;\ s, \tau)$ defined from any 2+1-D video sequence $f(x, y, t)$ by convolution with space-time separable spatio-temporal Gaussian kernels

$$T(x,y,t;\ s,\tau) = g(x,y;\ s)\, g(t;\ \tau) = \frac{1}{2\pi s} e^{-(x^2+y^2)/2s}\, \frac{1}{\sqrt{2\pi\tau}} e^{-t^2/2\tau} \tag{68}$$

at different spatio-temporal scales (s,τ) (Lindeberg [38]),

$$L(\cdot,\cdot,\cdot;\ s,\tau) = T(\cdot,\cdot,\cdot;\ s,\tau) * f(\cdot,\cdot,\cdot), \tag{69}$$

and with γ-normalized spatio-temporal derivatives defined according to [35, 43],

$$\partial_\xi = s^{\gamma_s/2}\,\partial_x, \quad \partial_\eta = s^{\gamma_s/2}\,\partial_y, \quad \partial_\zeta = \tau^{\gamma_\tau/2}\,\partial_t. \tag{70}$$

Initially, we will develop the basic theory based on a noncausal Gaussian temporal scale-space model, and then in the experiments, for the purpose of also being able to handle real-time image streams, complement this construction with a truly time-causal spatio-temporal scale-space representation (Lindeberg [43]) defined based on temporal smoothing with the time-causal limit kernel $\Psi(t;\ \tau, c)$ having a Fourier transform of the form

$$\hat{\Psi}(\omega;\ \tau, c) = \prod_{k=1}^{\infty} \frac{1}{1 + i\, c^{-k}\sqrt{c^2-1}\sqrt{\tau}\,\omega}, \tag{71}$$

and for which the discrete implementation of the temporal smoothing operation is in turn approximated by a finite number of discrete recursive filters coupled in a cascade.

4.2. A spatio-temporal quasi quadrature measure. In Lindeberg [43], the following spatio-temporal quadrature was considered:

$$\begin{aligned}
&\mathcal{Q}_{3,(x,y,t),norm} L \\
&= \tau\, \mathcal{Q}_{(x,y),norm} L_t + C\,\tau^2\, \mathcal{Q}_{(x,y),norm} L_{tt} \\
&= \tau \left(s\,(L_{xt}^2 + L_{yt}^2) + C\,s^2 \left(L_{xxt}^2 + 2L_{xyt}^2 + L_{yyt}^2 \right) \right) \\
&\quad + C\,\tau^2 \left(s\,(L_{xtt}^2 + L_{ytt}^2)\ + C s^2 (L_{xxtt}^2 + 2L_{xytt}^2 + L_{yytt}^2) \right).
\end{aligned} \tag{72}$$

This differential entity has been constructed to constitute a simultaneous quasi quadrature measure over both the spatial dimensions (x, y) and the temporal dimension t, implying that, instead of combining a pair of first- and second-order derivatives over a single dimension, here we use an octuple of first- and second-order derivatives over the three spatio-temporal dimensions and with additional terms added to make the resulting differential expression rotationally invariant over the spatial domain.

Specifically, this differential entity mimics some of the known properties of complex cells in the primary visual cortex as discovered by Hubel and Wiesel [19, 20, 21] in the sense of (i) being independent of the polarity of the stimuli, (ii) not obeying the superposition principle, and (iii) being rather insensitive to the phase of the visual stimuli. The primitive components of the quasi quadrature measure (the partial derivatives) do in turn mimic some of the known properties of simple cells in the primary visual cortex in terms of precisely localized "on" and "off" subregions (i) with spatial summation within each subregion, (ii) with spatial antagonism between on- and off-subregions, and (iv) whose visual responses to stationary or moving spots can be predicted from the spatial subregions. This model, however, is also simplified in the sense that the variability over different orientations, and eccentricities over the spatial domain as well as over motion directions over joint space-time, have been replaced by primitive components in terms of partial derivatives based on an isotropic scaling parameter over all spatial orientations and space-time separable receptive fields over the joint space-time domain.

This spatio-temporal quasi quadrature measure is intended to measure the local energy of the local spatio-temporal derivatives obtained by combining first- and second-order derivative operators over both the spatial dimensions and the temporal dimension. Specifically, it can be seen as a combination of the previously considered spatial quasi quadrature measure of the form (6) for $\Gamma_s = 0$ and the previously derived temporal quasi quadrature measure (55) for $\Gamma_\tau = 0$. By adding more general Γ-normalization with independent scale normalization parameters Γ_s and Γ_τ over space and time, respectively, we here extend the definition of the differential expression (72) to the following more general form:

$$\begin{aligned}
&\mathcal{Q}_{(x,y,t),\Gamma-norm} L \\
&= \frac{\tau\, \mathcal{Q}_{(x,y),\Gamma-norm} L_t + C_\tau \tau^2\, \mathcal{Q}_{(x,y),\Gamma-norm} L_{tt}}{\tau^{\Gamma_\tau}} \\
&= \frac{1}{s^{\Gamma_s} \tau^{\Gamma_\tau}} \left(\tau \left(s\,(L_{xt}^2 + L_{yt}^2) + C_s\, s^2 \left(L_{xxt}^2 + 2L_{xyt}^2 + L_{yyt}^2\right)\right)\right. \\
&\qquad \left. + C_\tau\, \tau^2 \left(s\,(L_{xtt}^2 + L_{ytt}^2) + C_s\, s^2 (L_{xxtt}^2 + 2L_{xytt}^2 + L_{yytt}^2)\right)\right). \qquad (73)
\end{aligned}$$

By the tight integration of the spatial quasi quadrature $\mathcal{Q}_{(x,y),\Gamma-norm} L$ with the temporal quasi quadrature measure $\mathcal{Q}_{t,\Gamma-norm} L$, the intention of this combined spatio-temporal quasi quadrature is to simultaneously allow for combined scale selective properties over joint space-time, to allow for joint spatio-temporal scale selection. Specifically, the fact that all individual components of this differential invariant (all the partial derivatives $L_{x^{m_1} y^{m_2} t^n}$) are expressed in terms of nonzero orders of spatial differentiation $m_1 + m_2 > 0$ and temporal differentiation $n > 0$ ensures that the resulting expression is localized over both space-time and spatio-temporal scales.

4.3. Scale selection properties for a spatio-temporal sine wave. In the following, we investigate the scale selection properties that this quasi quadrature measure gives rise to for a multidimensional sine wave of the form

$$f(x, y, t) = (\sin(\omega_s x) + \sin(\omega_s y)) \sin(\omega_\tau t) \qquad (74)$$

taken as an idealized model of a dense spatio-temporal structure over both space and time and with the spatio-temporal image structures having spatial extent of size $\lambda_s = 2\pi/\omega_s$ and temporal duration $\lambda_\tau = 2\pi/\omega_\tau$. The spatio-temporal scale-space representation of (74) obtained by Gaussian smoothing will then be of the form

$$L(x, y, t;\ s, \tau) = e^{-\omega_s^2 s/2} e^{-\omega_\tau^2 \tau/2} (\sin(\omega_s x) + \sin(\omega_s y)) \sin(\omega_\tau t). \qquad (75)$$

Let us decompose this quasi quadrature measure into the following four components based on spatial and temporal derivatives of either first or second order:

$$\begin{aligned}&\mathcal{Q}_{(x,y,t),\Gamma-norm}L\\&=\mathcal{Q}_{(x,y),1,\Gamma-norm}L_t+\mathcal{Q}_{(x,y),2,\Gamma-norm}L_t\\&\quad+C_\tau\left(\mathcal{Q}_{(x,y),1,\Gamma-norm}L_{tt}+\mathcal{Q}_{(x,y),2,\Gamma-norm}L_{tt}\right),\end{aligned}\tag{76}$$

where

$$\mathcal{Q}_{(x,y),1,\Gamma-norm}L_t=\frac{s\,\tau\,(L_{xt}^2+L_{yt}^2)}{s^{\Gamma_s}\tau^{\Gamma_\tau}},\tag{77}$$

$$\mathcal{Q}_{(x,y),2,\Gamma-norm}L_t=\frac{C_s\,s^2\,\tau\,\left(L_{xxt}^2+2L_{xyt}^2+L_{yyt}^2\right)}{s^{\Gamma_s}\tau^{\Gamma_\tau}},\tag{78}$$

$$\mathcal{Q}_{(x,y),1,\Gamma-norm}L_{tt}=\frac{s\,\tau^2\,(L_{xtt}^2+L_{ytt}^2)}{s^{\Gamma_s}\tau^{\Gamma_\tau}},\tag{79}$$

$$\mathcal{Q}_{(x,y),2,\Gamma-norm}L_{tt}=\frac{C_s\,s^2\,\tau^2\,\left(L_{xxtt}^2+2L_{xytt}^2+L_{yytt}^2\right)}{s^{\Gamma_s}\tau^{\Gamma_\tau}}.\tag{80}$$

By selecting both spatial and temporal scales from local extrema of the quasi quadrature measure over both spatial and temporal scales

$$(\hat{s}_{\mathcal{Q}_{(x,y,t),\Gamma-norm}},\hat{\tau}_{\mathcal{Q}_{(x,y,t),\Gamma-norm}})=\operatorname{argmaxlocal}_{s,\tau}\,\mathcal{Q}_{(x,y,t),\Gamma-norm}L,\tag{81}$$

it follows that

- at the spatial points $(x=n\pi/\omega_s, y=n\pi/\omega_s)$ at which only the first-order spatial derivatives respond, the selected spatial scale will be

$$\hat{s}_{11}=\frac{1-\Gamma_s}{\omega_s^2};\tag{82}$$

- at the spatial points $(x=(\pi/2+n\pi)/\omega_s, y=(\pi/2+n\pi)/\omega_s)$ at which only the second-order spatial derivatives respond, the selected spatial scale will be

$$\hat{s}_{22}=\frac{2-\Gamma_s}{\omega_s^2};\tag{83}$$

- at the temporal moments $t=n\pi/\omega_\tau$ at which only the first-order temporal derivative responds, the selected temporal scale will be

$$\hat{\tau}_1=\frac{1-\Gamma_\tau}{\omega_\tau^2};\tag{84}$$

- and at the temporal moments $t=(\pi/2+n\pi)/\omega_\tau$ at which only the second-order temporal derivative responds, the selected temporal scale will be

$$\hat{\tau}_2=\frac{2-\Gamma_\tau}{\omega_\tau^2}.\tag{85}$$

Determining the weighting parameters C_s and C_τ, such that the relative strengths of the first- and second-order components become equal at the spatial and temporal midpoints ($x = (\pi/4 + n\pi/2)/\omega_s$, $y = (\pi/4 + n\pi/2)/\omega_s$), and $t = (\pi/4 + n\pi/2)/\omega_\tau$ between the extreme points and at the spatial and temporal scales corresponding to the geometric averages $\sqrt{\hat{s}_1 \hat{s}_2}$ and $\sqrt{\hat{\tau}_1 \hat{\tau}_2}$ of the extreme values, then implies that the relative weighting factors C_s and C_τ between the first- and second-order derivative responses should be chosen as

$$C_s = \frac{1}{\sqrt{(1-\Gamma_s)(2-\Gamma_s)}}, \tag{86}$$

$$C_\tau = \frac{1}{\sqrt{(1-\Gamma_\tau)(2-\Gamma_\tau)}}. \tag{87}$$

Note the structural similarities between these results and the corresponding analysis for the purely spatial quasi quadrature measure $\mathcal{Q}_{(x,y),\Gamma-norm} L$ studied in section 2.

4.4. Spatio-temporal scale covariance of the joint spatio-temporal scale estimates under independent scaling transformations of the spatial and the temporal domains. Consider an independent scaling transformation of the spatial and the temporal domains of a video sequence

$$f'(x_1', x_2', t') = f(x_1, x_2, t) \qquad \text{for} \qquad (x_1', x_2', t') = (S_s\, x_1, S_s\, x_2, S_\tau\, t), \tag{88}$$

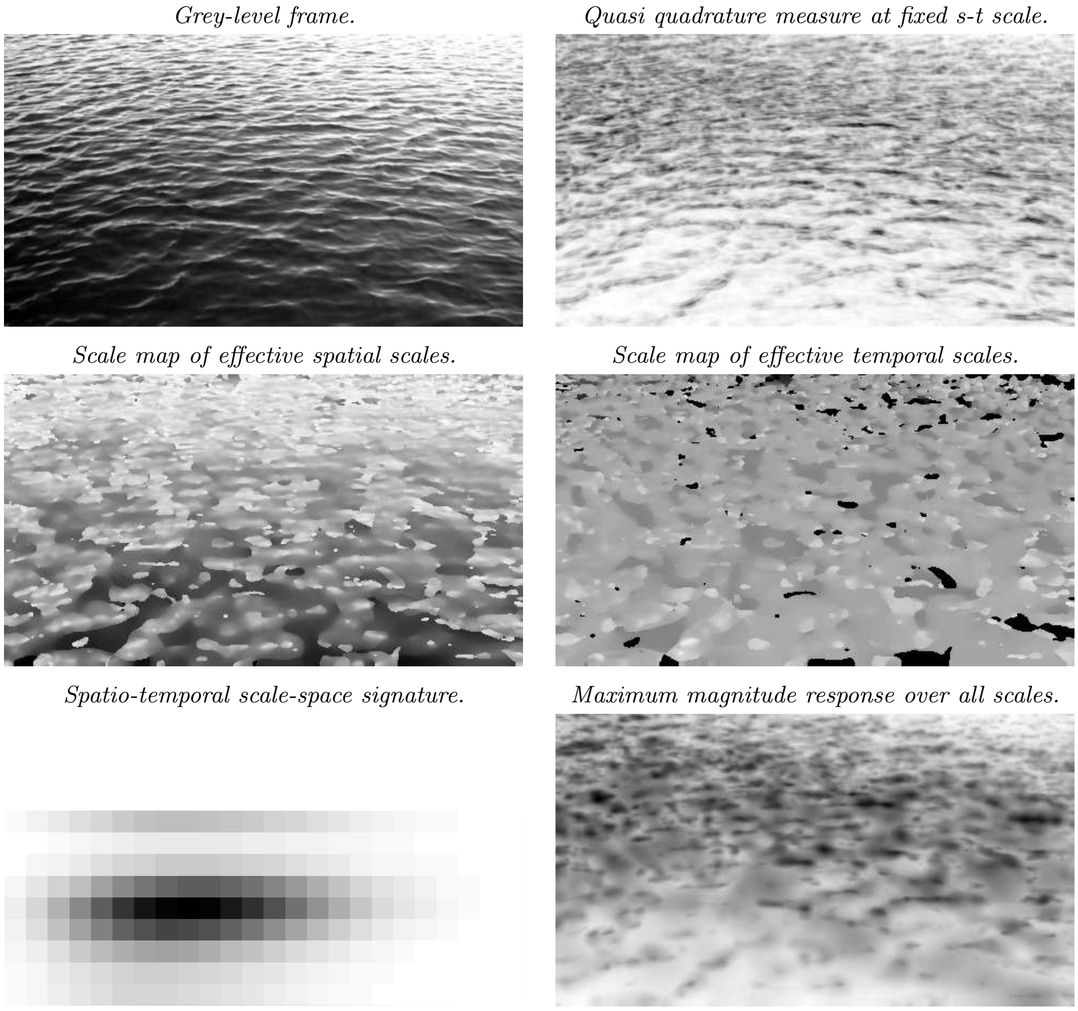
where S_s and S_τ denote the spatial and temporal scaling factors, respectively. Then, the corresponding spatio-temporal scale covariance of the spatio-temporal scale estimates

$$(\hat{s}', \hat{\tau}') = (S_s^2\, \hat{s}, S_\tau^2\, \hat{\tau}) \tag{89}$$

can be proven—provided that the spatial positions (x, y) and the temporal moments t are appropriately matched according to $(x_1', x_2', t') = (S_s\, x_1, S_s\, x_2, S_\tau\, t)$—by combining the ideas in the proof of spatial scale covariance from section 2.6 with the ideas in the proof of temporal scale covariance from section 3.2.

4.5. Experimental results. Figure 8 shows an example of applying dense spatio-temporal scale selection to a real video sequence. For reasons of computational efficiency, we only show results obtained using a time-causal and time-recursive spatio-temporal scale-space representation obtained by convolution with Gaussian kernels over the spatial domain and convolution with the time-causal limit kernel over the temporal domain. Because of the time-recursive implementation of this scale-space concept, it is not necessary to explicitly compute and build the five-dimensional spatio-temporal scale-space representation over space-time (x, y, t) and spatio-temporal scales (s, τ). Instead, the time-recursive implementation builds a four-dimensional representation over the spatial domain (x, y) and the spatio-temporal scale parameters (s, τ) at every temporal image frame t. Then, this representation is recursively updated to the next frame, using only the temporal scale-space representation at the previous frame as a sufficient temporal buffer of past information, using the methodology of time-causal and time-recursive spatio-temporal receptive fields developed in [43].

Because the notion of phase compensation is not yet fully developed for the time-causal limit kernel, we did not use local phase compensation in this experiment. Instead, we restricted



Grey-level frame.

Quasi quadrature measure at fixed s-t scale.

Scale map of effective spatial scales.

Scale map of effective temporal scales.

Spatio-temporal scale-space signature.

Maximum magnitude response over all scales.

Figure 8. *Results of dense spatio-temporal scale selection applied to a video sequence with water waves. The results have been computed with a time-causal and time-recursive spatio-temporal scale-space representation obtained by convolution with a Gaussian kernel over the spatial domain and the time-causal limit kernel over the temporal domain. Top left: Grey-level image. Top right: The spatio-temporal quasi quadrature measure computed at a fixed spatio-temporal scale. Middle left: Selected spatial scale levels in units of effective scale* $s_{eff} = \log_2 \sigma_s$*. Middle right: Selected temporal scale levels in units of effective scale* $\tau_{eff} = \log_2 \sigma_\tau$*. Bottom left: Spatio-temporal scale-space signature showing the magnitude variations of the spatio-temporal quasi quadrature measure over both spatial and temporal scales, with effective spatial scale increasing linearly from left to right and effective temporal scale increasing linearly from bottom to top. While this illustration shows the average over all the image points at the given image frame for the purpose of suppressing the influence of local spatial variations, the general dense scale selection method is otherwise local, based on individual scale-space signatures at every image point and for every time moment. Bottom right: The maximum magnitude response over all spatial and temporal scales at every image point. Note from the maps of the selected spatial and temporal scales that there is a clear vertical size gradient for the selected spatial scales, whereas there is no size gradient for the selected temporal scales. The reason for this is that there are size variations over the spatial domain because of perspective depth effects affecting the spatial scales, whereas the temporal scales are stationary over the image domain, since temporal scale levels are not affected by the perspective mapping. Note that the contrast of the maps showing magnitude information has been set so that dark shading corresponds to larger values and light shading to lower values. In addition, the magnitude maps have been stretched by a square root function. (Results computed using* 24 *logarithmically distributed spatial scale levels between* $\sigma_{s,min} = 0.25$ *pixels and* $\sigma_{s,max} = 24$ *pixels and using* 9 *logarithmically distributed temporal scale levels between* $\sigma_{\tau,min} = 10$ *ms and* $\sigma_{\tau,max} = 2.56$ *sec for* $C_s = 1/\sqrt{(1-\Gamma_s)(2-\Gamma_s)}$ *and* $C_\tau = 1/\sqrt{(1-\Gamma_\tau)(2-\Gamma_\tau)}$ *using complementary scale normalization parameters* $\Gamma_s = 0$ *and* $\Gamma_\tau = 0$*.) (Image size:* 480×270 *pixels. Frame* 100 *of* 250 *frames at* 25 *frames per second.)*

ourselves to spatial postsmoothing, noting that the approach can be extended to temporal postsmoothing in a straightforward manner by adding a second stage of recursive temporal smoothing to the quasi quadrature measures computed at every image frame. To make the magnitude maps maximally scale invariant for purposes of visualization, we used $\Gamma_s = \Gamma_\tau = 0$.

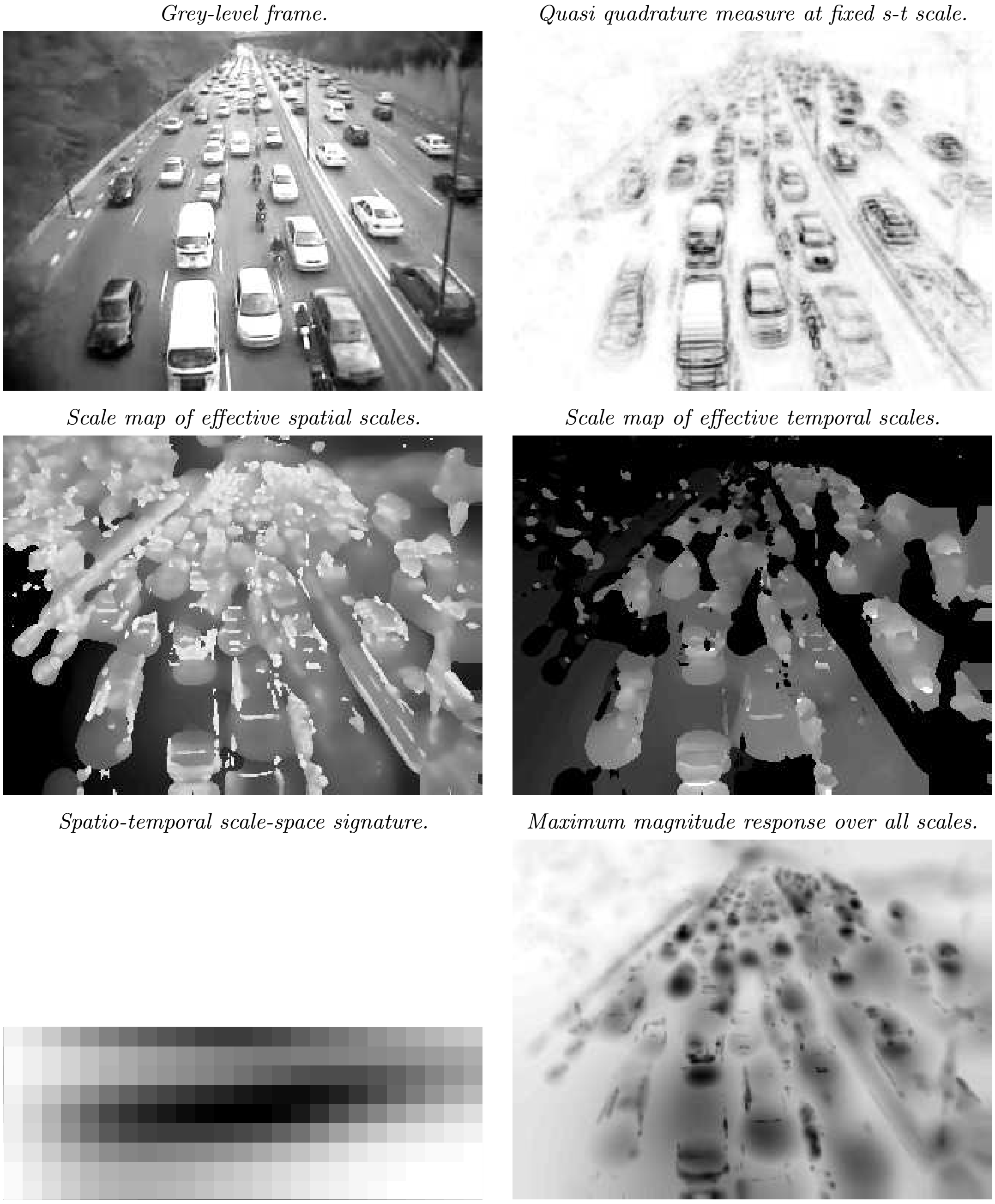
At every image frame, we computed a discrete approximation of the spatio-temporal quasi quadrature measure at all spatial and temporal scales and detected two-dimensional local extrema over spatial and temporal scales as candidates for local spatio-temporal scale levels. These local extrema were then interpolated to higher resolution over spatial and temporal scales using parabolic interpolation according to [46, eq. (115)]. For simplicity, the results shown in the second row display only the global extremum over spatio-temporal scales at every image point. However, when applying the scale selection methodology in practice, multiple local extrema over spatio-temporal scales should instead be considered to make it possible to handle multiple characteristic spatio-temporal scale levels at any image point.

From the scale maps in the middle row, we note that the selected spatial scale levels well reflect the perspective size gradient over the vertical direction in the image domain, if we assume that the water waves have a stationary distribution of wavelengths over the water surface, while these spatial lengths are shortened because of the perspective scaling and foreshortening effects. For the selected temporal scale levels, the distribution is more stationary over the image domain, which can be understood from the assumption that the temporal wavelengths of the waves should be stationary over the water surface, while at the same time the temporal scales are not affected by the perspective transformation (the temporal periodicity of a wave remains the same under imaging transformations).

In the spatio-temporal scale-space signature, showing the average over all the image points of the scale-normalized spatio-temporal quasi quadrature measure as a function of the spatial and temporal scales, we can see that for this video sequence there is a narrow range of dominant spatial and temporal scales. The spread over the spatial scale levels is, however, wider than the spread over the temporal scales, caused by the additional variabilities induced by the perspective scaling and foreshortening effects.

When comparing the maximum magnitude response over all spatio-temporal scales to the quasi quadrature measure at a fixed spatio-temporal scale, we can observe that the variability in the maximum over all spatio-temporal scales is lower than the variability in the response at a fixed scale.

Figures 9–10 show results of applying corresponding dense spatio-temporal scale selection to videos of other dynamic scenes. In the traffic scene in Figure 9, we note that distinct responses in the spatial and temporal scale maps are obtained for the different moving cars, again with a vertical size gradient in the spatial scale estimates reflecting the perspective scaling and foreshortening effects, whereas the temporal scale estimates are essentially unaffected by the perspective transformation. Additionally, we can observe that large spatial scales and long temporal scales are selected in the smooth stationary regions on the road and in some parts of the background. For the video of breaking waves in Figure 10, there are two dominant spatio-temporal scales in the scene: one for the larger scale of the overall waves and one for fine-scale spatio-temporal structures where the waves break. These two spatio-temporal scale levels are in turn reflected as horizontal stripes in the maps of the selected spatial and temporal scales.



Grey-level frame.

Quasi quadrature measure at fixed s-t scale.

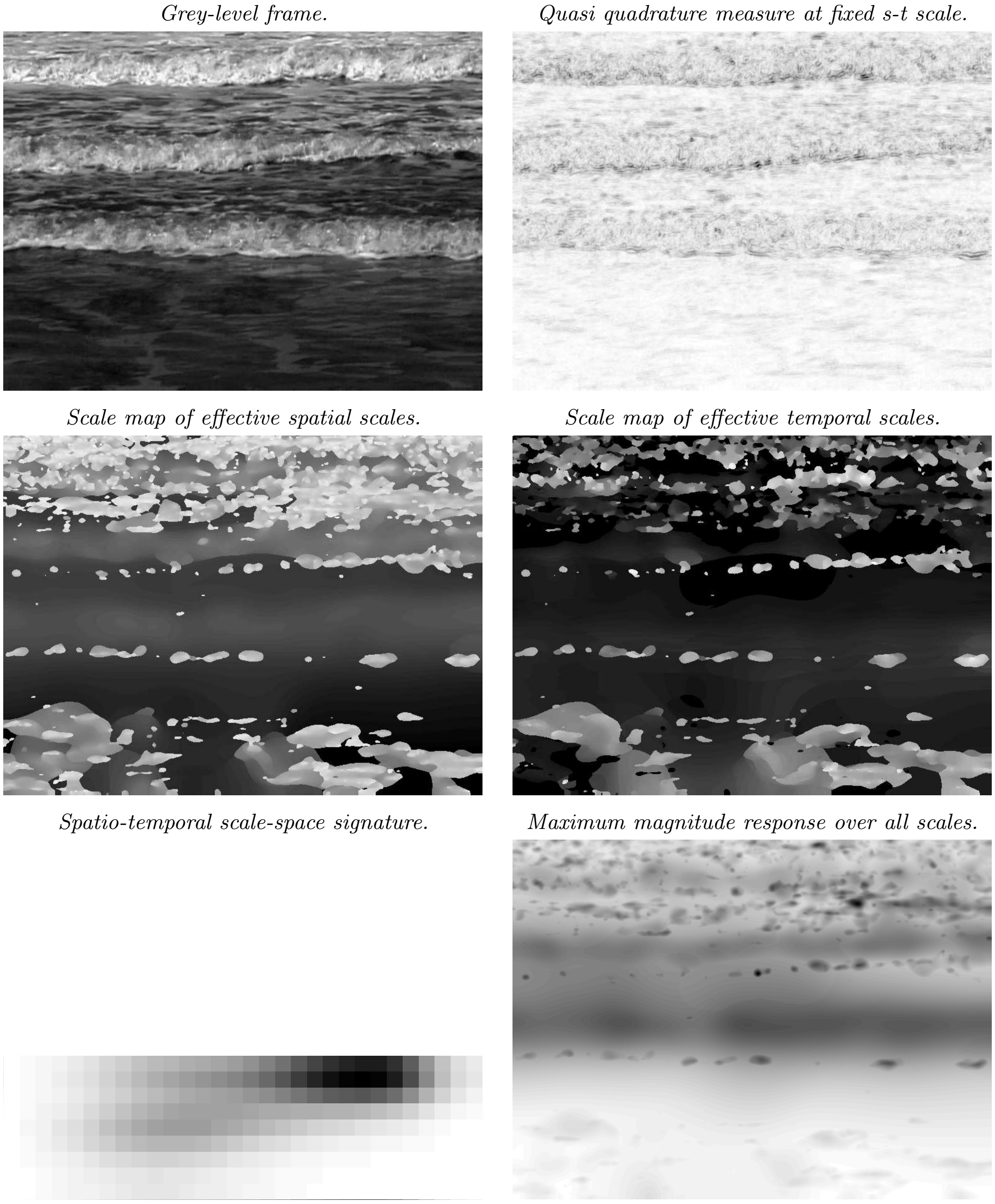
Scale map of effective spatial scales.

Scale map of effective temporal scales.

Spatio-temporal scale-space signature.

Maximum magnitude response over all scales.

Figure 9. *Results of dense spatio-temporal scale selection applied to a traffic scene (video "smooth_traffic*05*" from the Maryland dynamic scene dataset* [57]*). Note that distinct responses in the spatial and temporal scale maps are obtained for the different moving cars, again with a size gradient in the spatial scale estimates reflecting the perspective scaling effects, whereas the temporal scale estimates are essentially unaffected by the perspective transformation. Additionally, we can observe that large spatial scales and long temporal scales are selected in the smooth stationary regions on the road and in some parts of the background (Image size:* 320×240 *pixels. Frame* 80 *of* 1217 *frames at* 30 *frames per second.)*



Grey-level frame.

Quasi quadrature measure at fixed s-t scale.

Scale map of effective spatial scales.

Scale map of effective temporal scales.

Spatio-temporal scale-space signature.

Maximum magnitude response over all scales.

Figure 10. *Results of dense spatio-temporal scale selection applied to a scene with breaking waves (from the DynTex dataset* [54]*). Notice how horizontal stripes of finer spatial and temporal scales are selected at the breaking waves. Notice additionally that there are two dominant spatio-temporal scales in the scene: one for the larger scale of the overall waves and one for fine-scale spatio-temporal structures where the waves break. (Image size:* 768×576 *pixels. Frame* 50 *of* 250 *frames at* 30 *frames per second.)*

5. Summary and discussion. We have presented a general methodology for performing dense scale selection by detecting local extrema over scale of scale-normalized quasi quadrature entities, which constitute local energy measures of the combined strength of first- and second-order scale-space derivatives. Specifically, we have (i) analyzed how local scale estimates may in general be strongly dependent on the relative strengths of first- vs. second-order image information at every image point, and (ii) proposed two mechanisms to substantially reduce this phase dependency using postsmoothing and pointwise phase compensation.

Based on the presented theoretical analysis of scale selection properties over a purely spatial image domain, in section 2 we presented four types of algorithms for dense scale selection, depending on whether the mechanisms of phase compensation and postsmoothing are included or excluded. For Algorithms II–IV that involve such mechanisms, we have shown that these mechanisms substantially reduce the spatial variability of the local scale estimates compared to the baseline Algorithm I that does not make use of either phase compensation or postsmoothing. These four methods all lead to provable scale invariance in the sense that the local scale estimates perfectly follow scaling transformations over image space, and so do image features and image descriptors that are computed at scales proportional to these local scale estimates.

In section 3, we developed corresponding dense scale selection mechanisms over a purely temporal domain and with corresponding mechanisms of postsmoothing and local phase compensation to reduce the phase sensitivity of the local scale estimates. By experiments on a synthetic sine wave with an exponentially varying wavelength as a function of time, we demonstrated that the local scale estimates adapt well to the variabilities of the time-dependent characteristic temporal scales in the signal. By experiments on a neurophysiological signal with approximate stationarity properties, we demonstrated how the proposed dense scale selection methodology is able to reflect multiple levels of characteristic scales in the signal that are not as visible in a spectral analysis based on Fourier transforms.

In section 4, we combined the above dense scale selection mechanisms over spatial and temporal domains into joint dense spatio-temporal scale selection in video data and demonstrated how the resulting approach is able to generate hypotheses about joint characteristic spatio-temporal scales for different types of dynamic scenes.

A common property of these spatial, temporal, and spatio-temporal scale selection methods is that the scale estimates are computed in a bottom-up fashion from the data in such a way that the scale estimates will be covariant under independent scaling transformations of the spatial and the temporal domains. We propose these forms of dense scale selection as a general mechanisms for estimating local spatial and temporal scales in spatial images, temporal signals, and spatio-temporal video.

As a complement to previous scale selection methodologies, which have been primarily applied sparsely at spatial or spatio-temporal interest points, the proposed dense scale selection methodology is intended for applications where spatial, temporal, or spatio-temporal receptive field responses are to be computed densely at every image point and for every time moment. Potential applications of such dense receptive field responses include texture analysis over a static spatial domain and dynamic texture analysis over a spatio-temporal domain. For example, if the application of dense spatio-temporal scale selection presented in Figure 8 is applied to videos of water waves taken under different wind conditions, then the spatial scale estimates will reflect the spatial extent of the water waves, whereas the temporal scale estimates will reflect their temporal duration. In this way, dynamic parameters of the water waves can be estimated directly, without using a generative physical model of the wave patterns.

More generally, the proposed framework provides a theory for modeling and measuring how dense receptive field measurements respond selectively at different spatial and temporal scales. This theory should be relevant for a large sets of computer vision problems, where receptive field–based image measurements in terms of spatial or spatio-temporal N-jets are used as the basis for image analysis or video analysis applications. The presented theory could also be relevant for computational modeling of biological vision. If we regard the spatio-temporal quasi quadrature measure (73) as modeling important properties of complex cells as detailed in section 4.2, then the proposed dense spatio-temporal scale selection theory can explain how complex cells having receptive fields over different ranges of spatial and temporal scales respond selectively to stimuli of different spatial extent and temporal duration.

The only free parameters are the complementary spatial and temporal scale normalization parameters Γ_s and Γ_τ, the relative integration scales c for optional postsmoothing, and the scale calibration factor by which the generated scale estimates are proportional to the scales at which the local extrema over scales are assumed. These parameters should be optimized to the specific application domain, where the dense scale selection methodology is to be combined with higher-level visual modules.

If suitable values of these parameters can be determined for a specific application domain, then by the general scale covariance property of the scale estimates, the proposed dense scale selection theory guarantees that the resulting spatial, temporal, or spatio-temporal scale estimates will automatically adapt to and follow variabilities in the characteristic scales in the input images, signals, videos, or image streams. In this way, the resulting chain of computer vision/image analysis/signal analysis/video analysis operations can be made provably scale invariant.

Acknowledgments. I would like to thank Prof. Peter Brown at Oxford University for providing the data for the experiments in Figure 7 and Arvid Kumar at KTH Royal Institute of Technology for serving as a link and discussion partner regarding this dataset, specifically regarding the biological background.

REFERENCES

[1] A. Almansa and T. Lindeberg, *Fingerprint enhancement by shape adaptation of scale-space operators with automatic scale-selection*, IEEE Trans. Image Process., 9 (2000), pp. 2027–2042.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool, *Speeded up robust features (SURF)*, Comput. Vision Image Understanding, 110 (2008), pp. 346–359.

[3] R. N. Bracewell, *The Fourier Transform and Its Applications*, 3rd ed., McGraw-Hill, New York, 1999.

[4] L. Bretzner, I. Laptev, and T. Lindeberg, *Hand-gesture recognition using multi-scale colour features, hierarchical features and particle filtering*, in Proc. Face and Gesture, Washington D.C., 2002, pp. 63–74.

[5] L. Bretzner and T. Lindeberg, *Feature tracking with automatic selection of spatial scales*, Comput. Vision Image Understanding, 71 (1998), pp. 385–392.

[6] T. Brox and J. Weickert, *A TV flow based local scale estimate and its application to texture discrimination*, J. Visual Commun. Image Representation, 17 (2006), pp. 1053–1073.

[7] H. Cagnan, E. P. Duff, and P. Brown, *The relative phases of basal ganglia activities dynamically shape effective connectivity in Parkinson's disease*, Brain, 138 (2016), pp. 1667–1678.

[8] O. Chomat, V. de Verdiere, D. Hall, and J. Crowley, *Local scale selection for Gaussian based description techniques*, in Proc. European Conf. on Computer Vision (ECCV 2000), Lecture Notes in Comput. Sci. 1842, Springer-Verlag, Berlin, 2000, pp. 117–133.

[9] L. Cohen, *Time-Frequency Analysis*, Signal Process. Ser. 778, Prentice Hall PTR, Englewood Cliffs, NJ, 1995.

[10] D. Comaniciu, V. Ramesh, and P. Meer, *The variable bandwidth mean shift and data-driven scale selection*, in Proc. International Conference on Computer Vision (ICCV 2001), Vancouver, Canada, 2001, pp. 438–445.

[11] G. C. DeAngelis and A. Anzai, *A modern view of the classical receptive field: Linear and non-linear spatio-temporal processing by V1 neurons*, in The Visual Neurosciences, L. M. Chalupa and J. S. Werner, eds., Vol. 1, MIT Press, Cambridge, MA, 2004, pp. 704–719.

[12] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman, *Receptive field dynamics in the central visual pathways*, Trends Neurosci., 18 (1995), pp. 451–457.

[13] C. Dyken and M. S. Floater, *Transfinite mean value interpolation*, Comput. Aided Geom. Design, 26 (2009), pp. 117–134.

[14] L. M. J. Florack, *Image Structure*, Comput. Imaging Vision, Springer Science+Business Media, Dordrect, 1997.

[15] D. Gabor, *Theory of communication*, J. IEE, 93 (1946), pp. 429–457.

[16] L. D. Griffin, *The second order local-image-structure solid*, IEEE Trans. Pattern Anal. Mach. Intell., 29 (2007), pp. 1355–1366.

[17] C. Hammond, H. Bergman, and P. Brown, *Pathological synchronization in Parkinsons disease: Networks, models and treatment*, Trends Neurosci., 30 (2007), pp. 357–364.

[18] T. Hassner, S. Filosof, V. Mayzels, and L. Zelnik-Manor, *Sifting through scales*, IEEE Trans. Pattern Anal. Mach. Intell., 39 (2017), pp. 1431–1443.

[19] D. H. Hubel and T. N. Wiesel, *Receptive fields of single neurones in the cat's striate cortex*, J Physiol., 147 (1959), pp. 226–238.

[20] D. H. Hubel and T. N. Wiesel, *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex*, J Physiol., 160 (1962), pp. 106–154.

[21] D. H. Hubel and T. N. Wiesel, *Brain and Visual Perception: The Story of a 25-Year Collaboration*, Oxford University Press, Oxford, UK, 2005.

[22] T. Iijima, *Basic theory on normalization of pattern (in case of typical one-dimensional pattern)*, Bull. Electrotech. Lab., 26 (1962), pp. 368–388 (in Japanese).

[23] P. W. Jones and T. M. Le, *Local scales and multiscale image decompositions*, Appl. Comput. Harmon. Anal., 26 (2009), pp. 371–394.

[24] T. Kadir and M. Brady, *Saliency, scale and image description*, Int. J. Comput. Vision, 45 (2001), pp. 83–105.

[25] Y. Kang, K. Morooka, and H. Nagahashi, *Scale invariant texture analysis using multi-scale local autocorrelation features*, in Proc. Scale Space and PDE Methods in Computer Vision (Scale-Space'05), Lecture Notes in Comput. Sci. 3459, Springer, Berlin, 2005, pp. 363–373.

[26] J. J. Koenderink, *The structure of images*, Biol. Cybernet., 50 (1984), pp. 363–370.

[27] J. J. Koenderink, *Scale-time*, Biol. Cybernet., 58 (1988), pp. 159–162.

[28] J. J. Koenderink and A. J. van Doorn, *Representation of local geometry in the visual system*, Biol. Cybernet., 55 (1987), pp. 367–375.

[29] J. J. Koenderink and A. J. van Doorn, *Receptive field families*, Biol. Cybernet., 63 (1990), pp. 291–298.

[30] J. J. Koenderink and A. J. van Doorn, *Generic neighborhood operators*, IEEE Trans. Pattern Anal. Mach. Intell., 14 (1992), pp. 597–605.

[31] S. Lazebnik, C. Schmid, and J. Ponce, *A sparse texture representation using local affine regions*, IEEE Trans. Pattern Anal. Mach. Intell., 27 (2005), pp. 1265–1278.

[32] Y. Li, D. M. J. Tax, and M. Loog, *Scale selection for supervised image segmentation*, Image Vision Comput., 30 (2012), pp. 991–1003.

[33] T. Lindeberg, *Scale-Space Theory in Computer Vision*, Springer, Berlin, 1993.

[34] T. Lindeberg, *Edge detection and ridge detection with automatic scale selection*, Int. J. Comput. Vision, 30 (1998), pp. 117–154.

[35] T. Lindeberg, *Feature detection with automatic scale selection*, Int. J. Comput. Vision, 30 (1998), pp. 77–116.

[36] T. Lindeberg, *A scale selection principle for estimating image deformations*, Image Vision Comput., 16 (1998), pp. 961–977.

[37] T. Lindeberg, *Principles for automatic scale selection*, in Handbook on Computer Vision and Applications, Academic Press, Boston, 1999, pp. 239–274, http://www.csc.kth.se/cvap/abstracts/cvap222.html.

[38] T. Lindeberg, *Generalized Gaussian scale-space axiomatics comprising linear scale-space, affine scale-space and spatio-temporal scale-space*, J. Math. Imaging Vision, 40 (2011), pp. 36–81.

[39] T. Lindeberg, *A computational theory of visual receptive fields*, Biol. Cybernet., 107 (2013), pp. 589–635.

[40] T. Lindeberg, *Scale selection properties of generalized scale-space interest point detectors*, J. Math. Imaging Vision, 46 (2013), pp. 177–210.

[41] T. Lindeberg, *Scale selection*, in Computer Vision: A Reference Guide, K. Ikeuchi, ed., Springer, Berlin, 2014, pp. 701–713.

[42] T. Lindeberg, *Image matching using generalized scale-space interest points*, J. Math. Imaging Vision, 52 (2015), pp. 3–36.

[43] T. Lindeberg, *Time-causal and time-recursive spatio-temporal receptive fields*, J. Math. Imaging Vision, 55 (2016), pp. 50–88.

[44] T. Lindeberg, *Spatio-temporal scale selection in video data*, in Proc. Scale Space and Variational Methods in Computer Vision (SSVM 2017), Lecture Notes in Comput. Sci. 10302, Springer, Berlin, 2017, pp. 3–15.

[45] T. Lindeberg, *Spatio-temporal scale selection in video data*, J. Math. Imaging Vision, 2017, pp. 1–38, https://doi.org/10.1007/s10851-017-0766-9.

[46] T. Lindeberg, *Temporal scale selection in time-causal scale space*, J. Math. Imaging Vision, 58 (2017), pp. 57–101.

[47] M. Loog, *The jet metric*, in Proc. International Conference on Scale Space and Variational Methods in Computer Vision (SSVM 2007), Lecture Notes in Comput. Sci. 4485, Springer, Berlin, 2007, pp. 25–31.

[48] M. Loog, Y. Li, and D. Tax, *Maximum membership scale selection*, in Multiple Classifier Systems, Lecture Notes in Comput. Sci. 5519, Springer, Berlin, 2009, pp. 468–477.

[49] D. G. Lowe, *Distinctive image features from scale-invariant keypoints*, Int. J. Comput. Vision, 60 (2004), pp. 91–110.

[50] S. G. Mallat and W. L. Hwang, *Singularity detection and processing with wavelets*, IEEE Trans. Inform. Theory, 38 (1992), pp. 617–643.

[51] K. Mikolajczyk and C. Schmid, *Scale and affine invariant interest point detectors*, Int. J. Comput. Vision, 60 (2004), pp. 63–86.

[52] A. Negre, C. Braillon, J. L. Crowley, and C. Laugier, *Real-time time-to-collision from variation of intrinsic scale*, Experimental Robotics, 39 (2008), pp. 75–84.

[53] J. Ng and A. A. Bharath, *Steering in scale space to optimally detect image structures*, in Proc. European Conference on Computer Vision (ECCV 2004), Lecture Notes in Comput. Sci. 3021, Springer, Berlin, 2004, pp. 482–494.

[54] R. Péteri, S. Fazekas, and M. J. Huiskes, *DynTex: A comprehensive database of dynamic textures*, Pattern Recognition Lett., 31 (2010), pp. 1627–1632, https://doi.org/10.1016/j.patrec.2010.05.009.

[55] B. Rosin, M. Slovik, R. Mitelman, M. Rivlin-Etzion, S. N. Haber, Z. Israel, E. Vaadia, and H. Bergman, *Closed-loop deep brain stimulation is superior in ameliorating Parkinsonism*, Neuron, 72 (2011), pp. 370–384.

[56] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce, 3*D object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints*, Int. J. Comput. Vision, 66 (2006), pp. 231–259.

[57] N. Shroff, P. Turaga, and R. R. Chellappa, *Moving vistas: Exploiting motion for describing scenes*, in Proc. Computer Vision and Pattern Recognition (CVPR 2010), IEEE, Piscataway, NJ, 2010, pp. 1911–1918, http://www.umiacs.umd.edu/users/nshroff/DynamicScene.html.

[58] J. Sporring, C. J. Colios, and P. E. Trahanias, *Generalized scale selection*, in Proc. Int. Conf. on Image Processing (ICIP'00), Vancouver, Canada, 2000, pp. 920–923.

[59] J. Sporring, M. Nielsen, L. Florack, and P. Johansen, eds., *Gaussian Scale-Space Theory,* Comput. Imaging Vision 8, Springer Science+Business Media, Dordrecht, 1997.

[60] M. Tau and T. Hassner, *Dense correspondences across scenes and scales*, IEEE Trans. Pattern Anal. Mach. Intell., 38 (2016), pp. 875–888.

[61] B. ter Haar Romeny, *Front-End Vision and Multi-Scale Image Analysis*, Springer, Berlin, 2003.

[62] B. ter Haar Romeny, L. Florack, and M. Nielsen, *Scale-time kernels and models*, in Proc. International Conference on Scale-Space and Morphology in Computer Vision (Scale-Space'01), Lecture Notes in Comput. Sci. 2106, Springer, Berlin, 2001.

[63] G. Tinkhauser, A. Pogosyan, H. Tan, D. Herz, A. Kühn, and P. Brown, *Beta burst dynamics in Parkinson's disease OFF and ON dopaminergic medication*, Brain, 140 (2017), pp. 2968–2981.

[64] T. Tuytelaars and K. Mikolajczyk, *A Survey on Local Invariant Features*, Found. Trends Comput. Graphics Vision 3, Now Publishers, Hanover, MA, 2008.

[65] R. L. De Valois, N. P. Cottaris, L. E. Mahon, S. D. Elfer, and J. A. Wilson, *Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity*, Vision Res., 40 (2000), pp. 3685–3702.

[66] J. Weickert, S. Ishikawa, and A. Imiya, *Linear scale-space has first been proposed in Japan*, J. Math. Imaging Vision, 10 (1999), pp. 237–252.

[67] A. P. Witkin, *Scale-space filtering*, in Proc. 8th Int. Joint Conf. Art. Intell., Karlsruhe, Germany, 1983, pp. 1019–1022.