



# Provably Scale-Covariant Networks from Oriented Quasi Quadrature Measures in Cascade

Tony Lindeberg<sup>(✉)</sup>

Computational Brain Science Lab,  
Division of Computational Science and Technology,  
KTH Royal Institute of Technology, Stockholm, Sweden  
[tony@kth.se](mailto:tony@kth.se)

**Abstract.** This article presents a continuous model for hierarchical networks based on a combination of mathematically derived models of receptive fields and biologically inspired computations. Based on a functional model of complex cells in terms of an oriented quasi quadrature combination of first- and second-order directional Gaussian derivatives, we couple such primitive computations in cascade over combinatorial expansions over image orientations. Scale-space properties of the computational primitives are analysed and it is shown that the resulting representation allows for provable scale and rotation covariance. A prototype application to texture analysis is developed and it is demonstrated that a simplified mean-reduced representation of the resulting QuasiQuadNet leads to promising experimental results on three texture datasets.

## 1 Introduction

The recent progress with deep learning architectures has demonstrated that hierarchical feature representations over multiple layers have higher potential compared to approaches based on single layers of receptive fields. A limitation of current deep nets, however, is that they are not truly scale covariant. A deep network constructed by repeated application of compact  $3 \times 3$  or  $5 \times 5$  kernels, such as AlexNet [1], VGG-Net [2] or ResNet [3], implies an implicit assumption of a preferred size in the image domain as induced by the discretization in terms of local  $3 \times 3$  or  $5 \times 5$  kernels of a fixed size. Thereby, due to the non-linearities in the deep net, the output from the network may be qualitatively different depending on the specific size of the object in the image domain, as varying because of *e.g.* different distances between the object and the observer. To handle this lack of scale covariance, approaches have been developed such as spatial transformer networks [4], using sets of subnetworks in a multi-scale fashion [5] or by combining deep nets with image pyramids [6]. Since the size normalization performed

---

The support from the Swedish Research Council (contract 2018-03586) is gratefully acknowledged.

by a spatial transformer network is not guaranteed to be truly scale covariant, and since traditional image pyramids imply a loss of image information that can be interpreted as corresponding to undersampling, it is of interest to develop continuous approaches for deep networks that guarantee true scale covariance or better approximations thereof.

The subject of this article is to develop a continuous model for capturing non-linear hierarchical relations between features over multiple scales in such a way that the resulting feature representation is provably scale covariant. Building upon axiomatic modelling of visual receptive fields in terms of Gaussian derivatives and affine extensions thereof, which can serve as idealized models of simple cells in the primary visual cortex [7–9], we will propose a functional model for complex cells in terms of an oriented quasi quadrature measure. Then, we will combine such oriented quasi quadrature measures in cascade, building upon the early idea of Fukushima [10] of using Hubel and Wiesel’s findings regarding receptive fields in the primary visual cortex [11] to build a hierarchical neural network from repeated application of models of simple and complex cells.

We will show how the scale-space properties of the quasi quadrature primitive in this representation can be theoretically analyzed and how the resulting hand-crafted network becomes provably scale and rotation covariant, in such a way that the multi-scale and multi-orientation network commutes with scaling transformations and rotations over the spatial image domain. Experimentally, we will investigate a prototype application to texture classification based on a substantially mean-reduced representation of the resulting QuasiQuadNet.

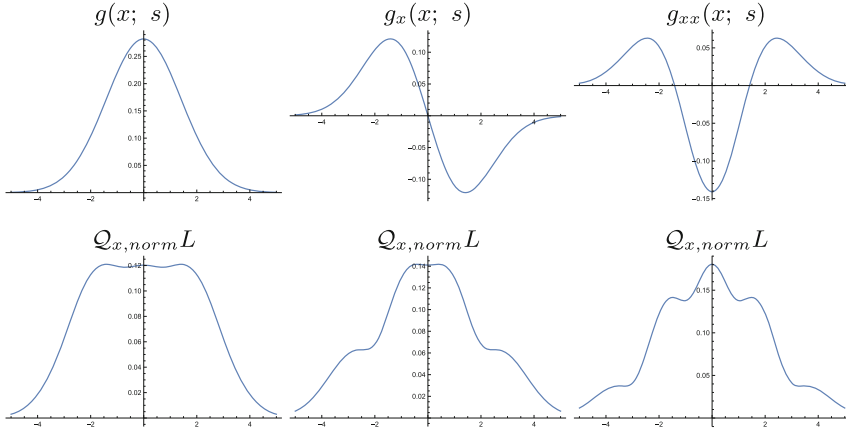
## 2 The Quasi Quadrature Measure over a 1-D Signal

Consider the scale-space representation  $L(x; s)$  of a 1-D signal  $f(x)$  defined by convolution with Gaussian kernels  $g(x; s) = \exp(-x^2/2s)/\sqrt{2\pi s}$  and with scale-normalized derivatives according to  $\partial_{\xi^n} = \partial_{x^n, \gamma-norm} = s^{n\gamma/2} \partial_x^n$  [12].

*Quasi Quadrature in 1-D.* Motivated by the fact that the first-order derivatives primarily respond to the locally odd component of the signal, whereas the second-order derivatives primarily respond to the locally even component of a signal, it is natural to aim at a differential feature detector that combines locally odd and even components in a complementary manner. By specifically combining the first- and second-order scale-normalized derivative responses in a Euclidean way, we obtain a quasi quadrature measure of the form

$$Q_{x,norm}L = \sqrt{\frac{s L_x^2 + C s^2 L_{xx}^2}{s^\Gamma}} \quad (1)$$

as a modification of the quasi quadrature measures previously proposed and studied in [12, 13], with the scale normalization parameters  $\gamma_1$  and  $\gamma_2$  of the first- and second-order derivatives coupled according to  $\gamma_1 = 1 - \Gamma$  and  $\gamma_2 = 1 - \Gamma/2$  to enable scale covariance by adding derivative expressions of different orders only for the scale-invariant choice of  $\gamma = 1$ . This differential entity can be seen as an approximation of the notion of a quadrature pair of an odd and even filter



**Fig. 1.** 1-D Gaussian derivatives up to orders 0, 1 and 2 for  $s_0 = 1$  with the corresponding 1-D quasi quadrature measures computed from them at scale  $s = 1$  for  $C = 8/11$ . (Horizontal axis:  $x \in [-5, 5]$ .)

as more traditionally formulated based on a Hilbert transform, while confined within the family of differential expressions based on Gaussian derivatives.

Figure 1 shows the result of computing this quasi quadrature measure for a Gaussian peak as well as its first- and second-order derivatives. As can be seen, the quasi quadrature measure is much less sensitive to the position of the peak compared to *e.g.* the first- or second-order derivatives. Additionally, the quasi quadrature measure also has some degree of spatial insensitivity for a first-order derivative (a local edge model) and a second-order derivative.

*Determination of C.* To determine the weighting parameter  $C$  between local second-order and first-order information, let us consider a Gaussian blob  $f(x) = g(x; s_0)$  with spatial extent given by  $s_0$  as input model signal. By using the semi-group property of the Gaussian kernel  $g(\cdot; s_1) * g(\cdot; s_2) = g(\cdot; s_1 + s_2)$ , the quasi quadrature measure can be computed in closed form

$$Q_{x, norm L} = \frac{s^{\frac{1-\Gamma}{2}} e^{-\frac{x^2}{2(s+s_0)}} \sqrt{x^2(s+s_0)^2 + Cs(s+s_0-x^2)^2 + 2}}{\sqrt{2\pi}(s+s_0)^{5/2}}. \tag{2}$$

By determining the weighting parameter  $C$  such that it minimizes the overall ripple in the squared quasi quadrature measure for a Gaussian input

$$\hat{C} = \operatorname{argmin}_{C \geq 0} \int_{x=-\infty}^{\infty} (\partial_x(Q_{x, norm L}^2))^2 dx, \tag{3}$$

we obtain

$$\hat{C} = \frac{4(s+s_0)}{11s}, \tag{4}$$

which in the special case of choosing  $s = s_0$  corresponds to  $C = 8/11 \approx 0.727$ . This value is very close to the value  $C = 1/\sqrt{2} \approx 0.707$  derived from an equal contribution condition in [13, Eq. (27)] for the special case of choosing  $\Gamma = 0$ .

*Scale Selection Properties.* To analyze the scale selection properties of the quasi quadrature measure, let us consider the result of using Gaussian derivatives of orders 0, 1 and 2 as input signals, *i.e.*,  $f(x) = g_{x^n}(x; s_0)$  for  $n \in \{0, 1, 2\}$ .

For the zero-order Gaussian kernel, the scale-normalized quasi quadrature measure at the origin is given by

$$\mathcal{Q}_{x,norm}L|_{x=0,n=0} = \frac{\sqrt{C}s^{1-\Gamma/2}}{2\pi(s+s_0)^2}. \quad (5)$$

For the first-order Gaussian derivative kernel, the scale-normalized quasi quadrature measure at the origin is

$$\mathcal{Q}_{x,norm}L|_{x=0,n=1} = \frac{s_0^{1/2}s^{(1-\Gamma)/2}}{2\pi(s+s_0)^2}, \quad (6)$$

whereas for the second-order Gaussian derivative kernel, the scale-normalized quasi quadrature measure at the origin is

$$\mathcal{Q}_{x,norm}L|_{x=0,n=2} = \frac{3\sqrt{C}s_0s^{1-\Gamma/2}}{2\pi(s+s_0)^3}. \quad (7)$$

By differentiating these expressions with respect to scale, we find that for a zero-order Gaussian kernel the maximum response over scale is assumed at

$$\hat{s}|_{n=0} = \frac{s_0(2-\Gamma)}{2+\Gamma}, \quad (8)$$

whereas for first- and second-order derivatives, respectively, the maximum response over scale is assumed at

$$\hat{s}|_{n=1} = \frac{s_0(1-\Gamma)}{3+\Gamma}, \quad \hat{s}|_{n=2} = \frac{s_0(2-\Gamma)}{4+\Gamma}. \quad (9)$$

In the special case of choosing  $\Gamma = 0$ , these scale estimates correspond to

$$\hat{s}|_{n=0} = s_0, \quad \hat{s}|_{n=1} = \frac{s_0}{3}, \quad \hat{s}|_{n=2} = \frac{s_0}{2}. \quad (10a-c)$$

Thus, for a Gaussian input signal, the selected scale level will for the most scale-invariant choice of using  $\Gamma = 0$  reflect the spatial extent  $\hat{s} = s_0$  of the blob, whereas if we would like the scale estimate to reflect the scale parameter of first- and second-order derivatives, we would have to choose  $\Gamma = -1$ . An alternative motivation for using finer scale levels for the Gaussian derivative kernels is to regard the positive and negative lobes of the Gaussian derivative kernels as substructures of a more complex signal, which would then warrant the use of finer scale levels to reflect the substructures of the signal ((10b) and (10c)).

### 3 Oriented Quasi Quadrature Modelling of Complex Cells

In this section, we will consider an extension of the 1-D quasi quadrature measure (1) into an oriented quasi quadrature measure of the form

$$Q_{\varphi, norm} L = \sqrt{\frac{\lambda_{\varphi} L_{\varphi}^2 + C \lambda_{\varphi}^2 L_{\varphi\varphi}^2}{s^T \Sigma s}}, \tag{11}$$

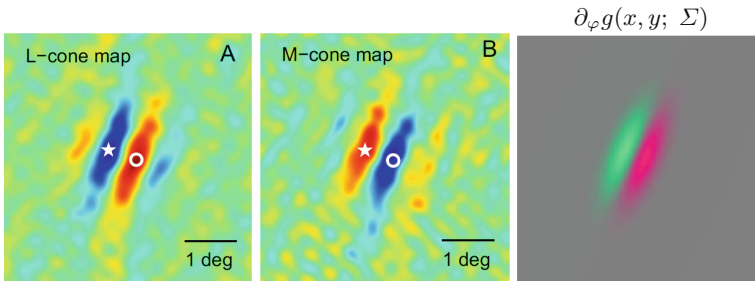
where  $L_{\varphi}$  and  $L_{\varphi\varphi}$  denote directional derivatives of an affine Gaussian scale-space representation [14, ch. 15] of the form  $L_{\varphi} = \cos \varphi L_{x_1} + \sin \varphi L_{x_2}$  and  $L_{\varphi\varphi} = \cos^2 \varphi L_{x_1 x_1} + 2 \cos \varphi \sin \varphi L_{x_1 x_2} + \sin^2 \varphi L_{x_2 x_2}$ , and with  $\lambda_{\varphi}$  denoting the variance of the affine Gaussian kernel (with  $x = (x_1, x_2)^T$ )

$$g(x; s, \Sigma) = \frac{1}{2\pi s \sqrt{\det \Sigma}} e^{-x^T \Sigma^{-1} x / 2s} \tag{12}$$

in direction  $\varphi$ , preferably with the orientation  $\varphi$  aligned with the direction  $\alpha$  of either of the eigenvectors of the composed spatial covariance matrix  $s \Sigma$ , with

$$\Sigma = \frac{1}{\max(\lambda_1, \lambda_2)} \begin{pmatrix} \lambda_1 \cos^2 \alpha + \lambda_2 \sin^2 \alpha & (\lambda_1 - \lambda_2) \cos \alpha \sin \alpha \\ (\lambda_1 - \lambda_2) \cos \alpha \sin \alpha & \lambda_1 \sin^2 \alpha + \lambda_2 \cos^2 \alpha \end{pmatrix} \tag{13}$$

normalized such that the main eigenvalue is equal to one.

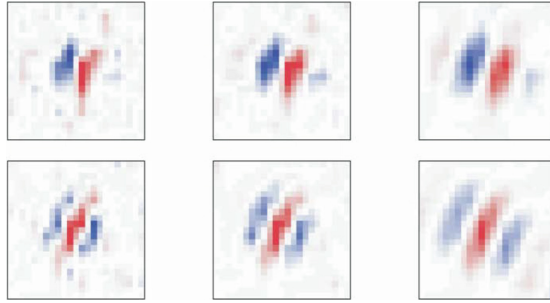


**Fig. 2.** Example of a colour-opponent receptive field profile for a double-opponent simple cell in the primary visual cortex (V1) as measured by Johnson *et al.* [15] (Fig. 1(a–b) Copyright Society for Neuroscience with permission): (left) Responses to L-cones corresponding to long wavelength red cones, with positive weights represented by red and negative weights by blue. (middle) Responses to M-cones corresponding to medium wavelength green cones, with positive weights represented by red and negative weights by blue. (right) Idealized model of the receptive field from a first-order directional derivative of an affine Gaussian kernel  $\partial_{\varphi} g(x, y; \Sigma)$  according to (14) for  $\sigma_1 = \sqrt{\lambda_1} = 0.6$ ,  $\sigma_2 = \sqrt{\lambda_2} = 0.2$  in units of degrees of visual angle,  $\alpha = 157^{\circ}$  and with positive weights for the red-green colour-opponent channel  $U = R - G$  with positive values represented by red and negative values by green. (Color figure online)

*Affine Gaussian derivative model for linear receptive fields.* According to the normative theory for visual receptive fields in Lindeberg [8,9], directional derivatives of affine Gaussian kernels constitute a canonical model for visual receptive fields over a 2-D spatial domain. Specifically, it was proposed that simple cells in the primary visual cortex (V1) can be modelled by directional derivatives of affine Gaussian kernels, termed *affine Gaussian derivatives*, of the form

$$T_{\varphi^m}(x_1, x_2; s, \Sigma) = \partial_{\varphi}^m (g(x_1, x_2; s, \Sigma)). \quad (14)$$

Figure 2 shows an example of the spatial dependency of a colour-opponent simple cell that can be well modelled by a first-order affine Gaussian derivative over an R-G colour-opponent channel over image intensities. Corresponding modelling results for non-chromatic receptive fields can be found in [8,9].



**Fig. 3.** Significant eigenvectors of a complex cell in the cat primary visual cortex, as determined by Touryan *et al.* [16] (Fig. 5(b) Copyright Elsevier with permission) from the response properties of the cell to a set of natural image stimuli, using a spike-triggered covariance method (STC) that computes the eigenvalues and the eigenvectors of a second-order Wiener kernel using three different parameter settings (cutoff frequencies) in the system identification method (from left to right). Qualitatively, these kernel shapes agree well with the shapes of first- and second-order affine Gaussian derivatives.

*Affine Quasi Quadrature Modelling of Complex Cells.* Figure 3 shows functional properties of a complex cell as determined from its response properties to natural images, using a spike-triggered covariance method (STC), which computes the eigenvalues and the eigenvectors of a second-order Wiener kernel (Touryan *et al.* [16]). As can be seen from this figure, the shapes of the eigenvectors determined from the non-linear Wiener kernel model of the complex cell do qualitatively agree very well with the shapes of corresponding affine Gaussian derivative kernels of orders 1 and 2. Motivated by this property and theoretical and experimental motivations for modelling receptive field profiles of simple cells by affine Gaussian derivatives, we propose to model complex cells by a possibly post-smoothed (spatially pooled) oriented quasi quadrature measure of the form (11)

$$(\overline{Q}_{\varphi, norm} L)(\cdot; s_{loc}, s_{int}, \Sigma_{\varphi}) = \sqrt{g(\cdot; s_{int}, \Sigma_{\varphi}) * (Q_{\varphi, norm}^2 L)(\cdot; s_{loc}, \Sigma_{\varphi})} \quad (15)$$

where  $s_{loc} \Sigma_\varphi$  represents an affine covariance matrix in direction  $\varphi$  for computing directional derivatives and  $s_{int} \Sigma_\varphi$  represents an affine covariance matrix in the same direction for integrating pointwise affine quasi quadrature measures over a region in image space.

The pointwise affine quasi quadrature measure  $(\mathcal{Q}_{\varphi, norm} L)(\cdot; s_{loc}, \Sigma_\varphi)$  can be seen as a Gaussian derivative based analogue of the energy model for complex cells as proposed by Adelson and Bergen [17] and Heeger [18]. It is closely related to a proposal by Koenderink and van Doorn [19] of summing up the squares of first- and second-order derivative responses and nicely compatible with results by De Valois *et al.* [20], who showed that first- and second-order receptive fields typically occur in pairs that can be modelled as approximate Hilbert pairs.

The addition of a complementary post-smoothing stage as determined by the affine Gaussian weighting function  $g(\cdot; s_{int}, \Sigma_\varphi)$  is closely related to recent results by Westö and May [21], who have shown that complex cells are better modelled as a combination of two spatial integration steps.

By choosing these spatial smoothing and weighting functions as affine Gaussian kernels, we ensure an affine covariant model of the complex cells, to enable the computation of affine invariants at higher levels in the visual hierarchy.

The use of multiple affine receptive fields over different shapes of the affine covariance matrices  $\Sigma_{\varphi, loc}$  and  $\Sigma_{\varphi, int}$  can be motivated by results by Goris *et al.* [22], who show that there is a large variability in the orientation selectivity of simple and complex cells. With respect to this model, this means that we can think of affine covariance matrices of different eccentricity as being present from isotropic to highly eccentric. By considering the full family of positive definite affine covariance matrices, we obtain a fully affine covariant image representation able to handle local linearizations of the perspective mapping for all possible views of any smooth local surface patch.

## 4 Hierarchies of Oriented Quasi Quadrature Measures

Let us in this first study disregard the variability due to different shapes of the affine receptive fields for different eccentricities and assume that  $\Sigma = I$ . This restriction enables covariance to scaling transformations and rotations, whereas a full treatment of affine quasi quadrature measures over all positive definite covariance matrices would have the potential to enable full affine covariance.

An approach that we shall pursue is to build feature hierarchies by coupling oriented quasi quadrature measures (11) or (15) in cascade

$$F_1(x, \varphi_1) = (\mathcal{Q}_{\varphi_1, norm} L)(x) \quad (16)$$

$$F_k(x, \varphi_1, \dots, \varphi_{k-1}, \varphi_k) = (\mathcal{Q}_{\varphi_k, norm} F_{k-1})(x, \varphi_1, \dots, \varphi_{k-1}), \quad (17)$$

where we have suppressed the notation for the scale levels assumed to be distributed such that the scale parameter at level  $k$  is  $s_k = s_0 r^{2(k-1)}$  for some  $r > 1$ , *e.g.*,  $r = 2$ . Assuming that the initial scale-space representation  $L$  is computed at scale  $s_0$ , such a network can in turn be initiated for different values of  $s_0$ , also distributed according to a geometric distribution.

This construction builds upon an early proposal by Fukushima [10] of building a hierarchical neural network from repeated application of models of simple and complex cells [11], which has later been explored in a hand-crafted network based on Gabor functions by Serre *et al.* [23] and in the scattering convolution networks by Bruno and Mallat [24]. This idea is also consistent with a proposal by Yamins and DiCarlo [25] of using repeated application of a single hierarchical convolution layer for explaining the computations in the mammalian cortex. With this construction, we obtain a way to define continuous networks that express a corresponding hierarchical architecture based on Gaussian derivative based models of simple and complex cells within the scale-space framework.

Each new layer in this model implies an expansion of combinations of angles over the different layers in the hierarchy. For example, if we in a discrete implementation discretize the angles  $\varphi \in [0, \pi[$  into  $M$  discrete spatial orientations, we will then obtain  $M^k$  different features at level  $k$  in the hierarchy. To keep the complexity down at higher levels, we will for  $k \geq K$  in a corresponding way as done by Hadji and Wildes [26] introduce a pooling stage over orientations

$$(\mathcal{P}_k F_k)(x, \varphi_1, \dots, \varphi_{K-1}) = \sum_{\varphi_k} F_k(x, \varphi_1, \dots, \varphi_{K-1}, \varphi_k), \quad (18)$$

and instead define the next successive layer as

$$F_k(x, \varphi_1, \dots, \varphi_{k-2}, \varphi_{K-1}, \varphi_k) = (\mathcal{Q}_{\varphi_k, norm} \mathcal{P}_{k-1} F_{k-1})(x, \varphi_1, \dots, \varphi_{K-1}) \quad (19)$$

to limit the number of features at any level to maximally  $M^{K-1}$ . The proposed hierarchical feature representation is termed QuasiQuadNet.

*Scale Covariance.* A theoretically attractive property of this family of networks is that the networks are provably scale covariant. Given two images  $f$  and  $f'$  that are related by a uniform scaling transformation  $f(x) = f'(Sx)$  for some  $S > 0$ , their corresponding scale-space representations  $L$  and  $L'$  will be equal  $L'(x'; s') = L(x; s)$  and so will the scale-normalized derivatives  $s'^{n/2} L'_{x'_i n}(x'; s') = s^{n/2} L_{x_i n}(x; s)$  based on  $\gamma = 1$  if the spatial positions are related according to  $x' = Sx$  and the scale levels according to  $s' = S^2 s$  [12, Eqns. (16) and (20)]. This implies that if the initial scale levels  $s_0$  and  $s'_0$  underlying the construction in (16) and (17) are related according to  $s'_0 = S^2 s_0$ , then the first layers of the feature hierarchy will be related according to  $F'_1(x', \varphi_1) = S^{-\Gamma} F_1(x, \varphi_1)$  [13, Eqns. (55) and (63)]. Higher layers in the feature hierarchy are in turn related according to

$$F'_k(x', \varphi_1, \dots, \varphi_{k-1}, \varphi_k) = S^{-k\Gamma} F_k(x, \varphi_1, \dots, \varphi_{k-1}, \varphi_k) \quad (20)$$

and are specifically equal if  $\Gamma = 0$ . This means that it will be possible to perfectly match such hierarchical representations under uniform scaling transformations.

*Rotation Covariance.* Under a rotation of image space by an angle  $\alpha$ ,  $f'(x') = f(x)$  for  $x' = R_\alpha x$ , the corresponding feature hierarchies are in turn equal if the orientation angles are related according to  $\varphi'_i = \varphi_i + \alpha$  ( $i = 1..k$ )

$$F'_k(x', \varphi'_1, \dots, \varphi'_{k-1}, \varphi'_k) = F_k(x, \varphi_1, \dots, \varphi_{k-1}, \varphi_k). \quad (21)$$



## 5 Application to Texture Analysis

In the following, we will use a substantially reduced version of the proposed quasi quadrature network for building an application to texture analysis.

If we make the assumption that a spatial texture should obey certain stationarity properties over image space, we may regard it as reasonable to construct texture descriptors by accumulating statistics of feature responses over the image domain, in terms of *e.g.* mean values or histograms. Inspired by the way the SURF descriptor [27] accumulates mean values and mean absolute values of derivative responses and the way Bruno and Mallat [24] and Hadji and Wildes [26] compute mean values of their hierarchical feature representations, we will initially explore reducing the QuasiQuadNet to just the mean values over the image domain of the following 5 features

$$\{\partial_\varphi F_k, |\partial_\varphi F_k|, \partial_{\varphi\varphi} F_k, |\partial_{\varphi\varphi} F_k|, \mathcal{Q}_\varphi F_k\}. \quad (22)$$

These types of features are computed for all layers in the feature hierarchy (with  $F_0 = L$ ), which leads to a 4000-D descriptor based on  $M = 8$  uniformly distributed orientations in  $[0, \pi[$ , 4 layers in the hierarchy delimited in complexity by directional pooling for  $K = 3$  with 4 initial scale levels  $\sigma_0 = \sqrt{s_0} \in \{1, 2, 4, 8\}$ .

**Table 1.** Performance results of the mean-reduced QuasiQuadNet in comparison with a selection of among the better methods in the extensive performance evaluation by Liu *et al.* [34] (our results in slanted font).

	KTH-TIPS2b	CUReT	UMD
FV-VGGVD [28] (SVM)	88.2	99.0	99.9
FV-VGGM [28] (SVM)	79.9	98.7	99.9
MRELBP [29] (SVM)	77.9	99.0	99.4
FV-AlexNet [28] (SVM)	77.9	98.4	99.7
<i>mean-reduced QuasiQuadNet LUV (SVM)</i>	78.3	98.6	
<i>mean-reduced QuasiQuadNet grey (SVM)</i>	75.3	98.3	97.1
ScatNet [24] (PCA)	68.9	99.7	98.4
MRELBP [29]	69.0	97.1	98.7
BRINT [30]	66.7	97.0	97.4
MDLBP [31]	66.5	96.9	97.3
<i>mean-reduced QuasiQuadNet LUV (NNC)</i>	72.1	94.9	
<i>mean-reduced QuasiQuadNet grey (NNC)</i>	70.2	93.0	93.3
LBP [32]	62.7	97.0	96.2
ScatNet [24] (NNC)	63.7	95.5	93.4
PCANet [33] (NNC)	59.4	92.0	90.5
RandNet [33] (NNC)	56.9	90.9	90.9

The second column in Table 1 shows the result of applying this approach to the KTH-TIPS2b dataset [35] for texture classification, consisting of 11 classes (“aluminum foil”, “cork”, “wool”, “lettuce leaf”, “corduroy”, “linen”, “cotton”, “brown bread”, “white bread”, “wood” and “cracker”) with 4 physical samples from each class and photos of each sample taken from 9 distances leading to 9 relative scales labelled “2”, . . . , “10” over a factor of 4 in scaling transformations and additionally 12 different pose and illumination conditions for each scale, leading to a total number of  $11 \times 4 \times 9 \times 12 = 4752$  images. The regular benchmark setup implies that the images from 3 samples in each class are used for training and the remaining sample in each class is used for testing over 4 permutations. Since several of the samples from the same class are quite different from each other in appearance, this implies a non-trivial benchmark which has not yet been saturated.

When using nearest-neighbour classification on the mean-reduced grey-level descriptor, we get 70.2% accuracy, and 72.1% accuracy when computing corresponding features from the LUV channels of a colour-opponent representation. When using SVM classification, the accuracy becomes 75.3% and 78.3%, respectively. Comparing with the results of an extensive set of other methods in Liu *et al.* [34], out of which a selection of the better results are listed in Table 1, the results of the mean-reduced QuasiQuadNet are better than classical texture classification methods such as locally binary patterns (LBP) [32], binary rotation invariant noise tolerant texture descriptors [30] and multi-dimensional local binary patterns (MDLBP) [31] and also better than other handcrafted networks, such as ScatNet [24], PCANet [33] and RandNet [33]. The performance of the mean-reduced QuasiQuadNet descriptor does, however, not reach the performance of applying SVM classification to Fischer vectors of the filter output in learned convolutional networks (FV-VGGVD, FV-VGGM [28]).

By instead performing the training on every second scale in the dataset (scales 2, 4, 6, 8, 10) and the testing on the other scales (3, 5, 7, 9), such that the benchmark does not primarily test the generalization properties between the different very few samples in each class, the classification performance is 98.8% for the grey-level descriptor and 99.6% for the LUV descriptor.

The third and fourth columns in Table 1 show corresponding results of texture classification on the CURET [36] and UMD [37] texture datasets, with random equally sized partitionings of the images into training and testing data. Also for these datasets, the performance of the mean-reduced descriptor is reasonable compared to other methods.

## 6 Summary and Discussion

We have presented a theory for defining hand-crafted hierarchical networks by applying quasi quadrature responses of first- and second-order directional Gaussian derivatives in cascade. The purpose behind this study has been to investigate if we could start building a bridge between the well-founded theory of scale-space representation and the recent empirical developments in deep learning, while at

the same time being inspired by biological vision. The present work is intended as an initial work in this direction, where we propose the family of quasi quadrature networks as a new baseline for hand-crafted networks with associated provable covariance properties under scaling and rotation transformations.

By early experiments with a substantially mean-reduced representation of the resulting QuasiQuadNet, we have demonstrated that it is possible to get quite promising performance on texture classification, and comparable or better than other hand-crafted networks, although not reaching the performance of learned CNNs. By inspection of the full non-reduced feature maps, which could not be shown here because of the space limitations, we have also observed that some representations in higher layers may respond to irregularities in regular textures (defect detection) or corners or end-stoppings in regular scenes.

Concerning extensions of the approach, we propose to: (i) complement the computation of quasi quadrature responses by divisive normalization [38] to enforce a competition between multiple feature responses, (ii) explore the spatial relationships in the full feature maps that are suppressed in the mean-reduced representation and (iii) incorporate learning mechanisms.

## References

1. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105 (2012)
2. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR (2015). [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of Computer Vision and Pattern Recognition (CVPR 2016), pp. 770–778 (2016)
4. Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K.: Spatial transformer networks. In: NIPS, pp. 2017–2025 (2015)
5. Cai, Z., Fan, Q., Feris, R.S., Vasconcelos, N.: A unified multi-scale deep convolutional neural network for fast object detection. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 354–370. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46493-0\\_22](https://doi.org/10.1007/978-3-319-46493-0_22)
6. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: CVPR (2017)
7. Koenderink, J.J., van Doorn, A.J.: Generic neighborhood operators. *IEEE-TPAMI* **14**, 597–605 (1992)
8. Lindeberg, T.: Generalized Gaussian scale-space axiomatics comprising linear scale-space, affine scale-space and spatio-temporal scale-space. *J. Math. Imaging Vis.* **40**, 36–81 (2011)
9. Lindeberg, T.: A computational theory of visual receptive fields. *Biol. Cybern.* **107**, 589–635 (2013)
10. Fukushima, K.: Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **36**, 193–202 (1980)
11. Hubel, D.H., Wiesel, T.N.: *Brain and Visual Perception*. Oxford University Press, New York (2005)

12. Lindeberg, T.: Feature detection with automatic scale selection. *Int. J. Comput. Vis.* **30**, 77–116 (1998)
13. Lindeberg, T.: Dense scale selection over space, time and space-time. *SIAM J. Imaging Sci.* **11**, 407–441 (2018)
14. Lindeberg, T.: *Scale-Space Theory in Computer Vision*. Springer, Dordrecht (1993). <https://doi.org/10.1007/978-1-4757-6465-9>
15. Johnson, E.N., Hawken, M.J., Shapley, R.: The orientation selectivity of color-responsive neurons in Macaque V1. *J. Neurosci.* **28**, 8096–8106 (2008)
16. Touryan, J., Felsen, G., Dan, Y.: Spatial structure of complex cell receptive fields measured with natural images. *Neuron* **45**, 781–791 (2005)
17. Adelson, E., Bergen, J.: Spatiotemporal energy models for the perception of motion. *JOSA A* **2**, 284–299 (1985)
18. Heeger, D.J.: Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* **9**, 181–197 (1992)
19. Koenderink, J.J., van Doorn, A.J.: Receptive field families. *Biol. Cybern.* **63**, 291–298 (1990)
20. De Valois, R.L., Cottaris, N.P., Mahon, L.E., Elfer, S.D., Wilson, J.A.: Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. *Vis. Res.* **40**, 3685–3702 (2000)
21. Westö, J., May, P.J.C.: Describing complex cells in primary visual cortex: a comparison of context and multi-filter LN models. *J. Neurophys.* **120**, 703–719 (2018)
22. Goris, R.L.T., Simoncelli, E.P., Movshon, J.A.: Origin and function of tuning diversity in Macaque visual cortex. *Neuron* **88**, 819–831 (2015)
23. Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., Poggio, T.: Robust object recognition with cortex-like mechanisms. *IEEE-TPAMI* **29**, 411–426 (2007)
24. Bruna, J., Mallat, S.: Invariant scattering convolution networks. *IEEE-TPAMI* **35**, 1872–1886 (2013)
25. Yamins, D.L.K., DiCarlo, J.J.: Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* **19**, 356–365 (2016)
26. Hadji, I., Wildes, R.P.: A spatiotemporal oriented energy network for dynamic texture recognition. In: *ICCV*, pp. 3066–3074 (2017)
27. Bay, H., Ess, A., Tuytelaars, T., van Gool, L.: Speeded Up Robust Features (SURF). *CVIU* **110**, 346–359 (2008)
28. Cimpoi, M., Maji, S., Vedaldi, A.: Deep filter banks for texture recognition and segmentation. In: *CVPR*, pp. 3828–3836 (2015)
29. Liu, L., Lao, S., Fieguth, P.W., Guo, Y., Wang, X., Pietikäinen, M.: Median robust extended local binary pattern for texture classification. *IEEE-TIP* **25**, 1368–1381 (2016)
30. Liu, L., Long, Y., Fieguth, P.W., Lao, S., Zhao, G.: BRINT: binary rotation invariant and noise tolerant texture classification. *IEEE-TIP* **23**, 3071–3084 (2014)
31. Schaefer, G., Doshi, N.P.: Multi-dimensional local binary pattern descriptors for improved texture analysis. In: *ICPR*, pp. 2500–2503 (2012)
32. Ojala, T., Pietikäinen, M., Maenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE-TPAMI* **24**, 971–987 (2002)
33. Chan, T.H., Jia, K., Gao, S., Lu, J., Zeng, Z., Ma, Y.: PCANet: A simple deep learning baseline for image classification? *IEEE-TIP* **24**, 5017–5032 (2015)
34. Liu, L., Fieguth, P., Guo, Y., Wang, Z., Pietikäinen, M.: Local binary features for texture classification: taxonomy and experimental study. *Pattern Recogn.* **62**, 135–160 (2017)

35. Mallikarjuna, P., Targhi, A.T., Fritz, M., Hayman, E., Caputo, B., Eklundh, J.O.: The KTH-TIPS2 database. KTH Royal Institute of Technology (2006)
36. Varma, M., Zisserman, A.: A statistical approach to material classification using image patch exemplars. *IEEE-TPAMI* **31**, 2032–2047 (2009)
37. Xu, Y., Yang, X., Ling, H., Ji, H.: A new texture descriptor using multifractal analysis in multi-orientation wavelet pyramid. In: *CVPR*, pp. 161–168 (2010)
38. Carandini, M., Heeger, D.J.: Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**, 51–62 (2012)