GENERALIZED RECTIFIER WAVELET COVARIANCE MODELS FOR TEXTURE SYNTHESIS

Anonymous authors
Paper under double-blind review

ABSTRACT

State-of-the-art maximum entropy models for texture synthesis are built from statistics relying on image representations defined by convolutional neural networks (CNN). Such representations capture rich structures in texture images, outperforming wavelet-based representations in this regard. However, conversely to neural networks, wavelets offer meaningful representations, as they are known to detect structures at multiple scales (e.g. edges) in images. In this work, we propose a family of statistics built upon non-linear wavelet based representations, that can be viewed as a particular instance of a one-layer CNN, using a generalized rectifier non-linearity. These statistics significantly improve the visual quality of previous classical wavelet-based models, and allow one to produce syntheses of similar quality to state-of-the-art models, on both gray-scale and color textures.

1 INTRODUCTION

Textures ares spatially homogeneous images, consisting of similar patterns forming a coherent ensemble. In texture modeling, one of the standard approaches to synthesize textures relies on defining a maximum entropy model (Jaynes, 1957) using a single observed image (Raad et al., 2018). It consists of computing a set of prescribed statistics from the observed texture image, and then generating synthetic textures producing the same statistics as the observation. If the statistics correctly describe the structures present in the observation, then any new image with the same statistics should appear similar to the observation. A major challenge of such methods resides in finding a suitable set of statistics, that can generate both high-quality and diverse synthetic samples.

A key question along this line of research is to find *what it takes to generate natural textures.* This problem was originally posed in Julesz (1962), in which the author looks for a statistical characterization of textures. In the classical work of Portilla & Simoncelli (2000), the authors presented a model whose statistics are built on the wavelet transform of an input texture image. These statistics were carefully chosen, by showing that each of them captured a specific aspect of the structure of the image. This model produces satisfying results for a wide range of textures, but fails to reproduce complex geometric structures present in some natural texture images. Figure 1 presents a typical example composed of radishes, and synthetic images from three state-of-the-art models developed over the last few decades. To address this problem, the work of Gatys et al. (2015) proposes to use statistics built on the correlations between the feature maps of a deep CNN, pre-trained on the ImageNet classification problem (Deng et al., 2009; Simonyan & Zisserman, 2014). While this model produces visually appealing images, these statistics are hard to interpret. The work of Ustyuzhaninov et al. (2017) made a significant simplification of such statistics, by using the features maps of a one-layer rectifier CNN. A crucial aspect of this simplification relies on using multi-scale filters, which are naturally connected to the wavelet transform. In this work, we propose a wavelet-based model to synthesize complex geometric structures, built on the recent development of the phase harmonics for image representations and non-Gaussian stationary process modeling (Mallat et al., 2020; Zhang & Mallat, 2021).

The phase harmonics are non-linear transformations that adjust the phase of a complex number. In Portilla & Simoncelli (2000); Zhang & Mallat (2021), the authors illustrate that the phase dependencies between wavelet coefficients across scales contain important information about the geometric structures in textures and turbulent flows, and that they can be captured by applying the phase harmonics to complex wavelet coefficients. Remarkably, Mallat et al. (2020) show that the phase

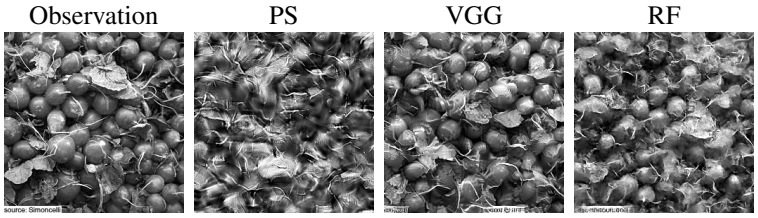

Figure 1: Example of syntheses from three texture models in chronological order. From left to right: the observed texture in gray-scale, synthesis from PS (Portilla & Simoncelli, 2000), from VGG (Gatys et al., 2015), and from RF[1](Ustyuzhaninov et al., 2017).

harmonics admit a dual representation, closely related to the rectifier non-linearity in CNNs. The main contribution of this paper is to develop a texture model based on the wavelet transform and a generalized rectifier non-linearity, that significantly improves the visual quality of the classical wavelet-based model of Portilla & Simoncelli (2000) on a wide range of textures. It also indicates the possibility of reducing significantly the number of statistics in the state-of-the-art work of Ustyuzhaninov et al. (2017), to achieve a similar visual quality.

The rest of the paper is organized as follows: Section 2 reviews the framework of microcanonical maximum-entropy models, build upon a general family of covariance statistics. We then present our model for both gray-scale and color textures in Section 3. Section 4 shows synthesis results of our model, compared with state-of-the-art models. Finally, in section 5, we discuss possible improvements of our model, and general texture models evaluation.

Notations Throughout the paper, N denotes a positive integer. A gray-scale image x is an element of $\mathbb{R}^{N\times N}$, i.e. $x = x(u)$, $u \in \Omega_N$, with $\Omega_N := \{0, \cdots, N-1\}^2$. A color image $x = \{x^c\}_{c=1,2,3}$ is an element of $\mathbb{R}^{3\times N\times N}$, or equivalently, each $x^c \in \mathbb{R}^{N\times N}$. We shall denote $\bar{x}$ the observed texture (observation), which is assumed to be a realisation a random vector X. For any complex number $z \in \mathcal{C}$, z^* is the complex conjugate of z, Real(z) its real part, $|z|$ its modulus, and $\varphi(z)$ its phase.

2 MICROCANONICAL COVARIANCE MODELS

We briefly review the standard framework of micro-canonical maximum-entropy models for textures. To reliably estimate the statistics in these models, we assume that a texture is a realization of a stationary and ergodic process (restricted to Ω_N). We then review a special family of statistics that are used in the state-of-the-art texture models (mentioned in Figure 1), based on covariance statistics of an image representation.

2.1 FRAMEWORK

Given a observation texture $\bar{x}$, we aim at generating new texture images, similar but different from $\bar{x}$. To that end, a classical method is to define a set of statistics $C\bar{x}$, computed on the observation, and try to sample from the *microcanonical set*

$$\{x : \|Cx - C\bar{x}\| \leq \epsilon\},$$

where $\|\cdot\|$ denotes the L^2 norm. Assume that this set is compact for some $\epsilon > 0$. The maximum-entropy distribution exists, and it is uniform on this set. The major challenges in defining such models are to choose a set of statistics and a sampling algorithm such that the resulting samples achieve a trade-off between the visual quality of the syntheses, and the diversity of the different syntheses (i.e. the model produces syntheses that are similar but different form each other, which can fail if part of the observation texture is memorized and reproduced).

This framework is often used in texture modeling, as under the stationary and ergodic assumption, one can construct Cx as a statistical estimator of $\mathbb{E}(CX)$. It is usually based on the spatial averaging

[1]As the model from Ustyuzhaninov et al. (2017) uses on random filters, we shall use the abbreviation RF.

of a representation Rx of an image x. More formally, we shall regard the complex-valued representation $Rx(\gamma, u) \in \mathbb{C}$ as a function of (γ, u) in an index set $\Gamma \times \Omega_N$. The set of covariance statistics Cx of a model can then be constructed by computing an averaging over the spatial variable u, i.e.

$$Cx(\gamma, \gamma', \tau) := \frac{1}{|\Omega_N|} \sum_{u \in \Omega_N} Rx(\gamma, u) Rx(\gamma', u - \tau)^*, \tag{1}$$

for $(\gamma, \gamma', \tau) \in \Upsilon \subseteq \Gamma \times \Gamma \times \Omega_N$. Given the representation, specifying the model then amounts to selecting the indices in Υ that we choose to correlate[2]. The statistics $Cx(\gamma, \gamma', \tau)$ can be interpreted as estimating the covariance (resp. correlations) between $RX(\gamma, u)$ and $RX(\gamma', u - \tau)$ for zero-mean RX (resp. non-zero mean RX).

The ergodicity assumption ensures that when N is large enough and the covariance set Υ is fixed, the approximation $C\bar{x} \simeq \mathbb{E}(CX)$ should hold with high probability. Intuitively this implies that, with a suitable choice of ϵ of the micro-canonical set, another realization x of X should also belong to this set. Therefore it makes sense to sample this set in order to generate new texture samples. The choices of the representation R and the covariance indices in Υ thus have a crucial impact on the quality and the diversity of the samples drawn from this set. To draw the samples, we follow gradient-based sampling algorithms, suitable in high-dimensions (Bruna & Mallat, 2018), to minimize the objective function $E(x) := \|Cx - C\bar{x}\|^2$, starting from an initial sample from a normal distribution.

This framework encompasses a wide range of state-of-the-art texture models, such as Portilla & Simoncelli (2000); Gatys et al. (2015); Ustyuzhaninov et al. (2017); Zhang & Mallat (2021). The model developed in Portilla & Simoncelli (2000) (that we shall note the PS model in the paper) takes inspiration from the human early visual system to define a multi-scale representation based on the *wavelet transform* of the image (Heeger & Bergen, 1995). We shall next review a family of covariance model which generalizes the statistics in the PS model. We write C^{M} the correlations for a specific model M that uses the representation R^{M}.

2.2 Wavelet phase harmonic covariance models

We review a family of microcanonical covariance models defined by a representation built upon the wavelet transform and phase harmonics. It defines a class of covariance statistics that capture dependencies between wavelet coefficients across scales.

2.2.1 Wavelet transform

The wavelet transform is a powerful tool in image processing to analyze signal structures, by defining a sparse representation (Mallat, 2001). For texture modeling, we consider oriented wavelets to model geometric structures in images at multiple scales. They include the Morlet wavelets and steerable wavelets, proposed in Goupillaud et al. (1984); Simoncelli & Freeman (1995); Unser & Chenouard (2013). In particular, the Simoncelli steerable wavelets have been used to model a diverse variety of textures in Portilla & Simoncelli (2000).

Oriented wavelets are defined by the dilation and rotation of a complex function $\psi : \mathbb{R}^2 \mapsto \mathbb{C}$ on a plane. Let r_θ denote the rotation by angle θ in $\mathbb{R}^2$. They are derived from ψ with dilations by factors 2^j, for $j \in \{0, 1, \cdots, J-1\}$, and rotations r_θ over angles $\theta = \ell\pi/L$ for $0 \leq \ell < L$, where L is the number of angles in $[0, \pi)$. The wavelet at scale j and angle θ is defined by

$$\psi_{j,\theta}(u) = 2^{-2j}\psi(2^{-j} r_\theta u), \quad u \in \mathbb{R}^2$$

Scales equal or larger than J are carried by a low-pass filter ϕ_J. These continuous wavelets are then discretized (with periodic boundary condition) to define a discrete wavelet transform in the spatial domain Ω_N.

The wavelet transform of an image $x \in \mathbb{R}^{N \times N}$ is a family of functions obtained by the convolution of x with the discrete wavelets. Let $\Lambda := \{0, \cdots, J-1\} \times \frac{\pi}{L}\{0, \cdots, L-1\}$ be the index set of the wavelet coefficients. For notation convenience, we denote the resulting coefficients by

$$x \star \psi_{j,\theta}(u) = \sum_{v \in \Omega_N} x(u - v)\psi_{j,\theta}(v), \quad u \in \Omega_N, \quad (j, \theta) \in \Lambda. \tag{2}$$

[2]More precisely, the spatial shifts are defined periodically, for $u - \tau \notin \Omega_N$. See Appendix D.3 for details about how we treat border effects for non-periodic images.

The low-pass coefficients $x \star \phi_J$ are defined similarly.

2.2.2 WAVELET PHASE HARMONICS AND THE PS MODEL

To model natural textures, it has been shown (Portilla & Simoncelli, 2000; Zhang & Mallat, 2021) that it is crucial to capture statistical dependencies between wavelet coefficients across scales. This can be achieved by using a wavelet phase harmonic representation, which is defined by the composition of a linear wavelet transform of x, and a non-linear phase harmonic transform.

In Mallat et al. (2020), the authors introduce the phase harmonics to adjust the phase of a complex number $z \in \mathbb{C}$. More precisely, the phase harmonics $\{[z]^k\}_{k\in\mathbb{Z}}$ of a complex number $z \in \mathbb{C}$ are defined by multiplying its phase $\varphi(z)$ of z by integers k, while keeping the modulus constant, i.e.

$$\forall\, k \in \mathbb{Z},\ [z]^k := |z| e^{ik\varphi(z)}. \tag{3}$$

The wavelet phase harmonic representation (WPH) is then defined by

$$R^{\mathrm{WPH}} x(\gamma, u) = [x \star \psi_{j,\theta}(u)]^k - \mu_\gamma, \quad \gamma = (j, \theta, k) \in \Gamma = \Lambda \times \mathbb{Z}, \tag{4}$$

where μ_γ is defined as the spatial average of $R^{\mathrm{WPH}}\bar{x}(\gamma, \cdot)$. For $k = 0$, μ_γ measures the sparsity of the wavelet coefficients, which (loosely speaking) counts the number of edges in the observation $\bar{x}$, of length 2^j and orientation θ.

The PS model can be interpreted as a particular case of the WPH covariance model, with a special choice of Υ. It is composed of the following three categories: 1) Raw coefficient correlations ($k = k' = 1$): they capture 2nd order statistics of a stationary process X, i.e. the correlations between $X(u)$ and $X(u')$ for $(u, u') \in \Omega_N^2$. 2) Coefficient magnitude statistics ($k = k' = 0$): they capture information of X beyond the 2nd order statistics. Very often, nearby scales and angles are considered in the model, such as $j' \approx j$ and $\theta' \approx \theta$. 3) Cross-scale phase statistics ($k = 1, k' = 2$): they capture local phase alignments of the wavelet coefficients at nearby scales $j' = j + 1$, which are complementary to the magnitude correlations. One major issue of the PS model is that it uses a restricted set of statistics: first, it is defined with low-order phase harmonics k and k', and second, the spatial shift parameter τ is restricted to be positive only for equal wavelet coefficients, $\gamma = \gamma'$. This results in a small index set Υ compared to the number of pixels N^2. In the next section, we shall use a dual representation of the phase harmonic operator to define a covariance model, closely related to state-of-the-art models based on CNN representations.

3 GENERALIZED RECTIFIER WAVELET COVARIANCE MODEL

In the previous section, we presented a class of models, built from the wavelet phase harmonic representation. A dual representation of the phase harmonic operator $[\cdot]^k$ can be defined via a generalized rectified linear unit, that we review in Section 3.1. We then discuss in Section 3.2 how to define an appropriate index set of Γ for gray-scale textures. Section 3.3 extends the model to color textures.

3.1 FROM PHASE HARMONICS TO THE GENERALIZED RECTIFIER

The generalized rectified linear unit of a complex number z, with a phase shifted by $\alpha \in [0, 2\pi]$, is defined by

$$\rho_\alpha(z) = \rho(\mathrm{Real}(e^{i\alpha} z)), \tag{5}$$

where ρ is a rectified linear unit, i.e. for any $t \in \mathbb{R}$, $\rho(t) := \max(0, t)$. In Mallat et al. (2020), it is shown that applying a Fourier transform on $\rho_\alpha(z)$ along the variable α results in the phase harmonics of z (up to some normalization constant). This suggests an alternative model, defined by coefficients of the form

$$R^{\mathrm{ALPHA}} x(\gamma, u) = \rho_\alpha(x \star \psi_{j,\theta}(u)) - \mu_\gamma, \quad \gamma = (j, \theta, \alpha), \tag{6}$$

for $\gamma \in \Gamma = \Lambda \times [0, 2\pi]$, and, as before, μ_γ is defined as the spatial average of $R^{\mathrm{ALPHA}}\bar{x}(\gamma, u)$ over $u \in \Omega_N$.

Relation with high-order phase harmonics Based on the duality between the phase harmonics $k \in \mathbb{Z}$ and the phase shift variable $\alpha \in [0, 2\pi]$, we now present the relation between C^{ALPHA} and the high-order phase harmonics in C^{WPH}. The following result was first proved in Mallat et al. (2020),

Proposition 1 *There exists a complex-valued sequence $\{c_k\}_{k\in\mathbb{Z}}$ such that for all $(j, \theta, \alpha) \in \Gamma$, $(j', \theta', \alpha') \in \Gamma$, and all $\tau \in \Omega_N$,*

$$C^{\mathrm{ALPHA}}x((j,\theta,\alpha),(j',\theta',\alpha'),\tau) = \sum_{(k,k')\in\mathbb{Z}^2} c_k c_k^* C^{\mathrm{WPH}}x((j,\theta,k),(j',\theta',k'),\tau)e^{i(k\alpha-k'\alpha')}.$$

The proof is given in Appendix A. We remark that the sequence $\{c_k\}_{k\in\mathbb{Z}}$ is uniquely determined by the rectifier non-linearity ρ, and they are non-zero if k is even (Mallat et al., 2020). This result shows that for a suitable choice of (α, α'). the covariance statistics $C^{\mathrm{ALPHA}}x$ can implicitly capture $C^{\mathrm{WPH}}x$ with a wide range of k and k'.

Relation with second order statistics Using a simple decomposition of wavelet coefficients into their positive, negative, real and imaginary parts, we can further show that the covariance statistics $C^{\mathrm{ALPHA}}x$ capture the classical second order statistics of wavelet coefficients, also used in the PS model (with phase harmonic coefficients $k = k' = 1$).

Proposition 2 *Let $I = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. Their exists a finite complex-valued sequence $\{w_{\alpha,\alpha'}\}_{(\alpha,\alpha')\in I^2}$ such that for all $(j, \theta) \in \Lambda$, and all $\tau \in \Omega_N$,*

$$\sum_{(\alpha,\alpha')\in I^2} w_{\alpha,\alpha'} C^{\mathrm{ALPHA}}x((j,\theta,\alpha),(j',\theta',\alpha'),\tau) = \frac{1}{\Omega_N}\sum_{u\in\Omega_N}\big(x\star\psi_{j,\theta}(u)\big)\big(x\star\psi_{j',\theta'}(u-\tau)\big)^*. \tag{7}$$

The proof is given in Appendix B. This shows that using only four α uniformly chosen between $[0, 2\pi]$ is sufficient to capture second order statistics. Because the wavelet transform is an invertible linear operator (on its range space), computing the r.h.s of eq. (7) for all (j, θ, τ), as well as the low-pass coefficients carried out by Φ_J, is equivalent to computing the correlation matrix of x.

Relation with the RF model Setting aside the subtraction by the spatial mean μ_γ, the RF model can be viewed as a particular case of models defined by eq. (6). Indeed, the statistics of the RF model take the form of eq. (1), with

$$R^{\mathrm{RF}}x(f,u) = \rho(x\star\psi_f(u)),$$

where $\{\psi_f\}$ being a family of multi-scale random filters. By writing $\rho_\alpha(x\star\psi_{j,\theta}(u-\tau)) = \rho(x\star \mathrm{Real}(\psi^\tau_{j,\theta}e^{i\alpha})(u))$, with $\psi^\tau_{j,\theta}$ denoting the translation of $\psi_{j,\theta}$ by τ, we see that the models are similar, the difference being that our models use wavelet-based filters instead of random ones.

3.2 Defining an appropriate Υ

The choice of the covariance set Υ is of central importance in the definition of the model. A too small set of indices will induce a model that could miss important structural information about the texture that we want to synthesize. Conversely, if Υ contains too many indices, the syntheses can have good visual quality, but the statistics of the model may have a large variance, leading to the memorization of some patterns of the observation. There is a trade-off between these two aspects: one must capture enough information to get syntheses of good visual quality, but not much, so as not to reproduce parts of the original image. To illustrate this point, we shall study the model ALPHA defined with three different sets Υ : A smaller model $\mathrm{ALPHA_S}$ with a limited amount of elements in Υ, an intermediate model $\mathrm{ALPHA_I}$, and a larger model $\mathrm{ALPHA_L}$.

To precisely define these models, let us note $\mathfrak{J} := \{0, \cdots, J-1\}$, $\Theta := \frac{\pi}{L}\{0, \cdots, L-1\}$, and $\mathcal{A}_A = \frac{2\pi}{A}\{0, \cdots, A-1\}$. Let us also define the set of spatial shifts $\mathfrak{T} := \{0\} \cup \{2^j(\cos(\theta), \sin(\theta))\}_{0\le j<J,\, \theta\in\frac{\pi}{L}\{0,\cdots,2L-1\}}$. Table 1 summarizes the conditions that all parameters

Model	Scales	Angles	Phase shift	Spatial shift
$\mathrm{ALPHA_S}$	$(j,j')\in\mathfrak{J}^2$ $\lvert j'-j\rvert\leq 1$	$(\theta,\theta')\in\Theta^2$	$(\alpha,\alpha')\in\mathcal{A}_4\times\mathcal{A}_1$	$\tau\in\mathfrak{T}$ if $(j,\theta)=(j',\theta')$ $\tau=0$ otherwise.
$\mathrm{ALPHA_I}$	$(j,j')\in\mathfrak{J}^2$	$(\theta,\theta')\in\Theta^2$	$(\alpha,\alpha')\in\mathcal{A}_4\times\mathcal{A}_1$	$\tau\in\mathfrak{T}$
$\mathrm{ALPHA_L}$	$(j,j')\in\mathfrak{J}^2$	$(\theta,\theta')\in\Theta^2$	$(\alpha,\alpha')\in\mathcal{A}_4^2$	$\tau\in\mathfrak{T}$

Table 1: List of indices in Υ for different ALPHA models.

have to satisfy to be contained in these sets. Additionally, these models include large scale information through the covariance of a low-pass filter, i.e. the spatial average of $x\star\phi_J(\cdot)x\star\phi_J(\cdot-\tau)$, for $\tau\in\mathfrak{T}$.

The small model $\mathrm{ALPHA_S}$ is inspired from the PS model, as it only takes into account of the interactions between nearby scales (i.e. $|j'-j|\leq 1$), and the spatial shift correlations are only considered for $(j,\theta)=(j',\theta')$. On the contrary, the intermediate model uses a much wider range of scale and spatial shift interactions. The larger model further takes into account of a wider range of phase shift interactions. The size (number of statistics) of the $\mathrm{ALPHA_S}$, $\mathrm{ALPHA_I}$ and $\mathrm{ALPHA_L}$ models are in the order of $J|\Theta|^2|\mathcal{A}_A|+J|\Theta||\mathcal{A}_A||\mathfrak{T}|$, $J^2|\Theta|^2|\mathcal{A}_A||\mathfrak{T}|$, and $J^2|\Theta_4|^2|\mathcal{A}_A|^2|\mathfrak{T}|$ respectively. The size of these models thus depends linearly on the number of spatial shifts in $\mathfrak{T}$. To keep this number from being too large, instead of taking all shifts in a square box, such as in Portilla & Simoncelli (2000), we choose to select only shift of dyadic moduli, and with the same orientations as the wavelets.

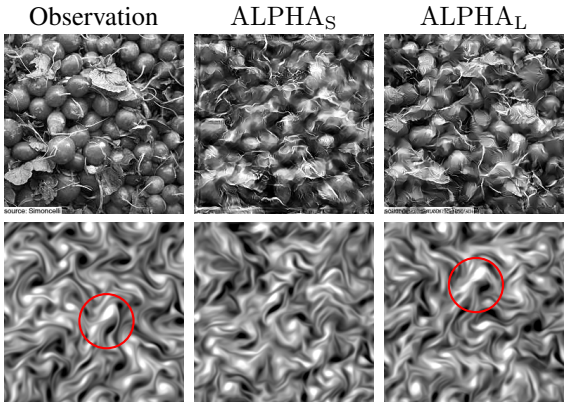

Figure 2: Examples of syntheses from models defined in Table 1.

To illustrate the trade-off on the visual quality and the diversity of the model, we present syntheses from models $\mathrm{ALPHA_S}$ and $\mathrm{ALPHA_L}$ on different texture examples. In Figure 2, we observe that the larger model $\mathrm{ALPHA_L}$ is able to capture and reproduce more geometric structures than the smaller model, resulting in more visually appealing syntheses, particularly for the first example. However, the last row show an example where the synthesis from the $\mathrm{ALPHA_L}$ model is too similar to the observation, as it produces patterns too similar to some we observe in $\bar{x}$. In order to avoid this memorization effect, we shall, in the rest of the paper, consider only the intermediate model.

3.3 Modelling color interactions

In order to generate color images, the covariance model $\mathrm{ALPHA_I}$ defined in Section 3.2 could be directly applied to each R, G and B color channel independently. However, it would not take into account the color coherence in the structures of the observation.

To capture color interactions in the observation image, we shall impose the covariance between the coefficients of equation 6 for all indices in Υ and all color channels. More precisely, let $x=\{x^c\}_{c=1,2,3}$ be a color image, with the parameter c representing the color channel. The $\mathrm{ALPHA_C}$

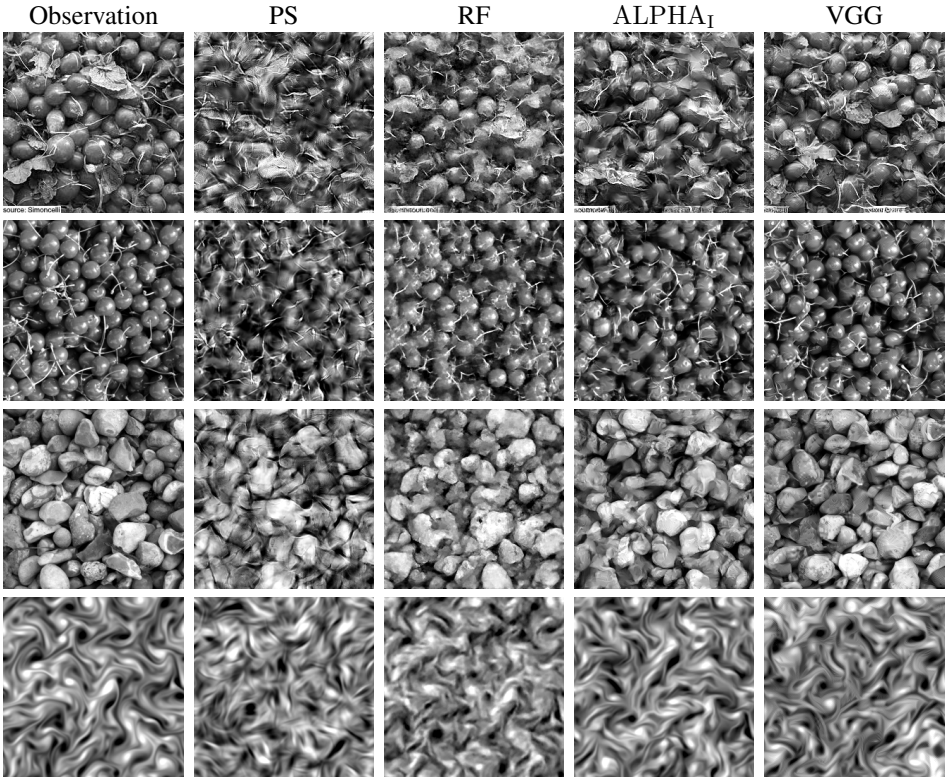

Figure 3: Visual comparison between the gray-scale texture models.

color model is defined by correlations between coefficients of the form:

$$R^{\mathrm{ALPHA_C}} x(\gamma, u) = \rho_\alpha (x^c \star \psi_{j,\theta})(u) - \mu_\gamma, \quad \gamma = (j, \theta, \alpha, c). \tag{8}$$

The set of indices is defined as $\Upsilon^{\mathrm{ALPHA_C}} := \{(\gamma, \gamma', \tau) : ((j, \theta, \alpha), (j', \theta', \alpha'), \tau) \in \Upsilon^{\mathrm{ALPHA_I}}, (c, c') \in \{1, 2, 3\}^2\}$.

4 Numerical results

In this section, we compare our intermediate model to the state-of-art models (PS, RF and VGG) on both gray-scale and color textures. We first specify the experimental setup and detail the size of each model. We then present the synthesis results of various examples, and discuss their quality through visual inspection. A quantitative evaluation of the quality of the syntheses, based on the synthesis loss of the VGG model, and proposed in Ustyuzhaninov et al. (2017), is discussed in Appendix E.

4.1 Experimental Setup

For our experiments, we choose gray-scale and color textures with various visual structures[3]. In the gray-scale examples, we also include a turbulent vorticity flow, which is simulated from Navier-Stokes equations in two dimensions (Schneider et al., 2006). These examples all contain complex geometric structures that are hard to model by the classical PS model.

The texture images presented in this work have a size of $N = 256$, giving a number of pixels of $\sim$65k. For all the ALPHA models, we use Morlet wavelets with a maximal scale $J = 5$ and number of orientations $L = 4$. This choice differs from the PS model, which uses Simoncelli steerable wavelets. In Appendix C, we discuss the impact of these wavelets on our model. Similarly to Gatys et al. (2015); Ustyuzhaninov et al. (2017), we use the L-BFGS algorithm (Nocedal, 1980)

[3]The source for the presented textures are given in Appendix D.

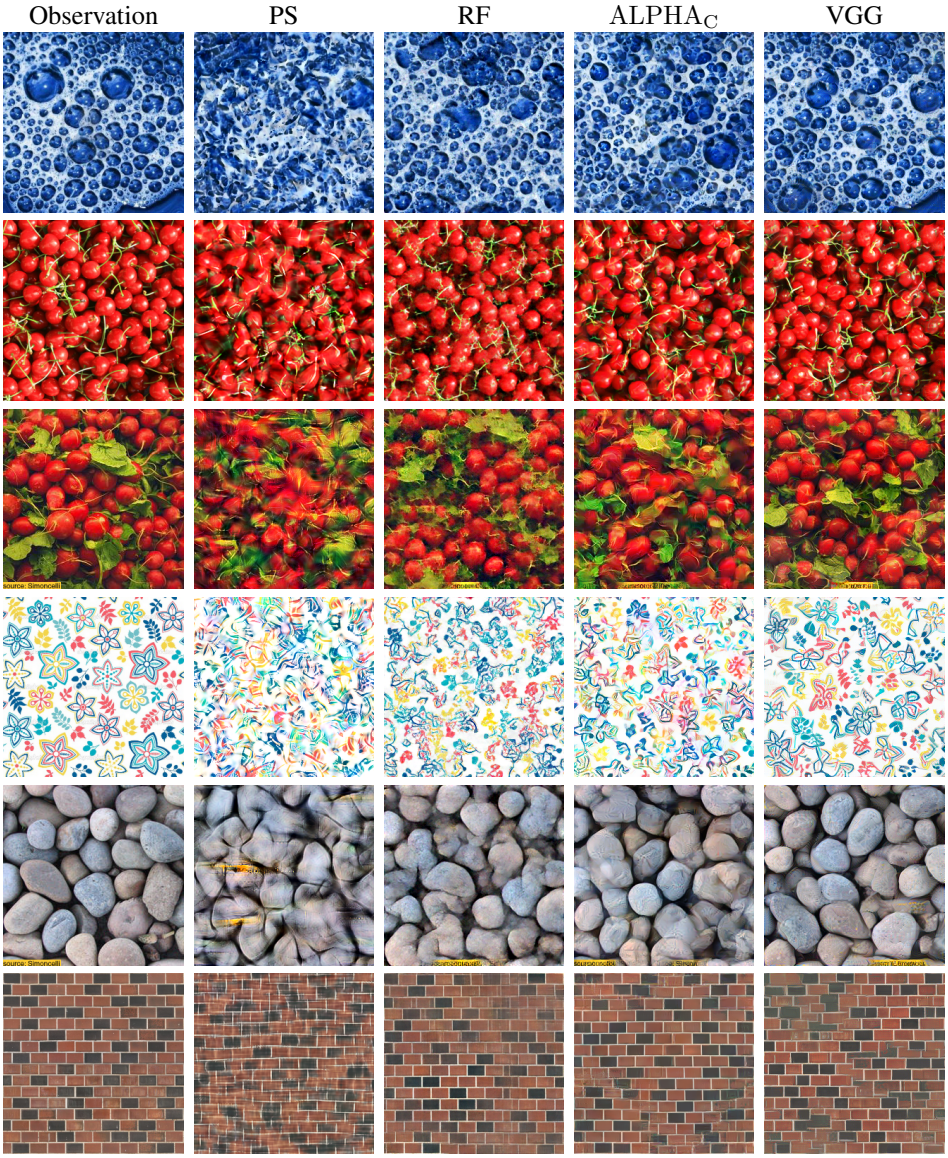

Figure 4: Visual comparison between the color texture models.

for the optimization of the objective. As in the VGG model (Gatys et al., 2015), we further apply a histogram matching procedure as post-processing after optimization. The details of the PS, RF and VGG models, as well as detailed specifications of our models, are given in Appendix D.

For gray-scale textures, the number of the statistics used in the $\mathrm{ALPHA_I}$ model is $\sim$35k. The small model $\mathrm{ALPHA_S}$ has $\sim$3.5k statistics, while the large model $\mathrm{ALPHA_L}$ has $\sim$142k statistics. The PS, RF and VGG models have $\sim$7k, $\sim$1048k , $\sim$177k statistics respectively. For color textures, the number of statistics of the $\mathrm{ALPHA_C}$ model is $\sim$320k, while for the PS, RF and VGG models, it is $\sim$36k, $\sim$1048k, $\sim$177k respectively.

4.2 Results

In Figure 3, we present examples of syntheses from the $\mathrm{ALPHA_I}$, PS, RF and VGG models. We observe that our model produces syntheses of similar visual quality to the RF and VGG models, significantly outperforming the PS model. In Figure 4, we present examples of syntheses for the color textures. The syntheses from our model for the examples of the second and third rows, which

were also presented in Figure 3, appear more visually convincing than their gray-scale versions. This is probably due to the fact that color images already have decomposed the structure information in the different color channels. We observe that our model reproduces the color coherence of textures, and reaches similar visual quality to the RF and VGG models. Note that for the last row example in Figure 4, the VGG model produces less convincing textures, because the long-range correlations present in the image (aligned bricks) are not reproduced. To remedy this issue, it has been proposed in Berger & Memisevic (2017) to add spatial shifts to the correlations of the network feature maps. These shift statistics are similar to the parameter τ in our model.

5 Discussion

In this work, we presented a new generative model for texture images built on a wavelet-based representation of geometric structures. Inspired from the classical work of Portilla & Simoncelli (2000), this model incorporates a wider range of statistics, by computing covariances between rectified wavelet coefficients, at different scales, phases and positions. We showed that this model is able to capture and reproduce complex geometric structures, present in natural textures or physical processes, producing syntheses of similar quality to state-of-the-art models that use CNN-based representations.

Being defined with a wavelet family instead of multi-scale random filters, the proposed model uses about three times less statistics than the RF model for color textures. For the gray-scale textures, our model has about 30 times less statistics, as it focuses on capturing the geometric structures present in images. However, further reduction of the number of statistics remains an open problem. For instance, the size of the model could be reduced with a more sophisticated selection of the spatial shift parameter τ, depending on the scales of the coefficients being correlated. Additionally, our color model is defined by computing covariances between coefficients of all different color channels, and is therefore nine times larger than the gray-scale model. In model of Portilla & Simoncelli (2000) extended to color images performs a PCA on the color channels (Vacher & Briand, 2021). A similar idea could also be applied to our model for color images. Another line of research to reduce the size of the model could be to incorporate a second layer of wavelet transform (Leonarduzzi et al., 2019; Sifre & Mallat, 2013). Such statistics could also capture additional information to improve the quality of the synthesized textures.

Finding a minimal set of statistics to define a texture model is important because a large number of statistics can result in a high variance of the estimators, and the associated model can suffer from memorization effects. That is, parts of the observation image can be reproduced in the syntheses. This is a problem because the aim of the model is to approximate the underlying distribution of the observation, and therefore produce diverse textures. In this regard, the mere visual evaluation of the synthetic textures can fail to take this aspect of the model into account. A more general evaluation of texture models is an interesting line of research, left for future works.

Reproducibility Statement In order to reproduce the experiments presented in this work, the main parameters to define the model can be found in Section 4.1. More details are given in Appendix D. Additionally, the code shall be made publicly available.

Ethics Statement The authors acknowledge that no potential conflicts of interest, discrimination, bias, fairness concerns or research integrity issues have been raised during the completion of this work.

References

Guillaume Berger and Roland Memisevic. Incorporating long-range consistency in cnn-based texture generation. In *International Conference on Learning Representations, ICLR 2017*. OpenReview.net, 2017.

Joan Bruna and Stephane Mallat. Multiscale Sparse Microcanonical Models. *Mathematical Statistics and Learning*, 1(5):257–315, 2018.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Leon Gatys, Alexander S Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks. *Advances in neural information processing systems*, 28:262–270, 2015.

P. Goupillaud, A. Grossmann, and J. Morlet. Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, 23(1):85–102, 1984. ISSN 0016-7142. Seismic Signal Analysis and Discrimination III.

David J Heeger and James R Bergen. Pyramid-based texture analysis/synthesis. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 229–238, 1995.

Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

Bela Julesz. Visual pattern discrimination. *IRE transactions on Information Theory*, 8(2):84–92, 1962.

Roberto Leonarduzzi, Gaspar Rochette, Jean-Phillipe Bouchaud, and Stéphane Mallat. Maximum-entropy scattering models for financial time series. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5496–5500. IEEE, 2019.

Stéphane Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way, 3rd Edition*. Academic Press, 2001.

Stéphane Mallat, Sixin Zhang, and Gaspar Rochette. Phase harmonic correlations and convolutional neural networks. *Information and Inference: A Journal of the IMA*, 9(3):721–747, 2020.

Jorge Nocedal. Updating quasi-newton matrices with limited storage. *Mathematics of computation*, 35(151):773–782, 1980.

Javier Portilla and Eero P Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision*, 40(1):49–70, 2000.

Lara Raad, Axel Davy, Agnès Desolneux, and Jean-Michel Morel. A survey of exemplar-based texture synthesis. *Annals of Mathematical Sciences and Applications*, 3(1):89–148, 2018.

Kai Schneider, Jörg Ziuber, Marie Farge, and Alexandre Azzalini. Coherent vortex extraction and simulation of 2D isotropic turbulence. *Journal of Turbulence*, 7:N44, 2006.

Laurent Sifre and Stéphane Mallat. Rotation, scaling and deformation invariant scattering for texture discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1233–1240, 2013.

E. P. Simoncelli and W. T. Freeman. The steerable pyramid: a flexible architecture for multi-scale derivative computation. In *Proceedings., International Conference on Image Processing*, volume 3, pp. 444–447 vol.3, Oct 1995. doi: 10.1109/ICIP.1995.537667.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

M. Unser and N. Chenouard. A unifying parametric framework for 2d steerable wavelet transforms. *SIAM Journal on Imaging Sciences*, 6(1):102–135, 2013.

Ivan Ustyuzhaninov, Wieland Brendel, Leon A. Gatys, and Matthias Bethge. What does it take to generate natural textures? In *International Conference on Learning Representations, ICLR 2017*. OpenReview.net, 2017.

Jonathan Vacher and Thibaud Briand. The portilla-simoncelli texture model: towards understanding the early visual cortex. *Image Processing On Line*, 11:170–211, 2021.

Sixin Zhang and Stéphane Mallat. Maximum entropy models from phase harmonic covariances. *Applied and Computational Harmonic Analysis*, 53:199–230, 2021.