HAL open science

Commutativity of spatiochromatic covariance matrices in natural image statistics

Yiye Jang, Jérémie Bigot, Edoardo Provenzi

To cite this version:

Yiye Jang, Jérémie Bigot, Edoardo Provenzi. Commutativity of spatiochromatic covariance matrices in natural image statistics. Mathematics in Engineering, 2020. hal-02141503v2

HAL Id: hal-02141503

https://hal.science/hal-02141503v2

Submitted on 14 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Commutativity of spatiochromatic covariance matrices in natural image statistics

Yiye Jang, Jérémie Bigot and Edoardo Provenzi*[1]

[1]Université de Bordeaux, CNRS, Bordeaux INP, IMB, UMR 5251, F-33400, 351 Cours de la Libération, Talence, France

January 14, 2020

Abstract

Statistic of natural images is a growing field of research both in vision and image processing. On the vision research side, fine statistical details about object distribution in real-world scenes help understanding the human visual system behavior. On the image processing side, by using the information gathered from statistics of natural scenes, we can obtain reliable priors and insights that can be used in many models. In has been rigorously proven in [16] that, if second order stationarity and commutativity of spatiochromatic covariance matrices hold true for natural scenes, then the codification of spatial and chromatic information by the human visual system can be separated through a tensor product. Spatial features are coded via local and oriented Fourier basis elements, while color features are coded via a triad given by an achromatic channel followed by two color opponent channels. In this paper, we will show that, while stationarity is guaranteed, commutativity is not. However, we shall see that commutativity of spatiochromatic covariance matrices can be approached if the database of images used to model visual scenes is modified accordingly to a suitable transformation that describes the response of retinal photoreceptors to light absorption: the Michaelis-Menten formula. A thorough investigation of the effects of a parameter of this formula will be performed and its influence on commutativity of covariance matrices will be detailed.

1 Introduction

In [16] a mathematical result about the separability of spatial and chromatic features of natural images has been established under two hypothesis: second order stationarity and commutativity of the so-called spatiochromatic covariance matrices. Among other consequences, this result guarantees the well-posedness of several image codification models.

The closest representation of physical irradiance of a visual scene can be obtained through multispectral and high dynamic range (HDR) techniques. Unfortunately, nowadays technology permits the acquisition of artifact-free multispectral and HDR images only for still-life scenes, thus constraining too much the available data-set that can be used for statistical purposes. For this reason, knowing that we are considering a rough approximation of the true irradiance, we were forced to build our datasets by using Raw images obtained via RGB cameras.

In [16] it has been shown that spatiochromatic covariance matrices of raw images do not commute perfectly. However, the commutativity properties improve considerably if the information

*yiye.jang,jeremie.bigot,edoardo.provenzi@math.u-bordeaux.fr

carried by raw images is modified according to an important transformation called *Michaelis-Menten formula* that can be written like this: $\mathbf{u}_\mu(x) \mapsto \mathbf{u}_\mu^\gamma(x)/(\mathbf{u}_\mu^\gamma(x) + m_\mu^\gamma)$, where $\mathbf{u}$ is the RGB raw image function, x is the pixel location, μ is one of the RGB chromatic channel, so that $\mathbf{u}_\mu(x)$ is the image intensity in the pixel x and in the chromatic channel μ and m represents the average value. More specifically, we want to investigate the influence of the parameter γ on the commutativity.

The reason why we consider this formula instead of others is that it has been empirically discovered in [20] that it fits the behavior of retinal cones in the transduction processed from light radiance to neuronal electric potential. Thus, the transformed images after the application of the Michaelis-Menten formula can be considered as a good approximation of the first signal input for the visual chain of events that characterize the human visual system.

We stress that the mathematical theory developed in this paper is aimed at formalizing and extending the results of [16], with the hope that they may be useful for vision scientists to refine their models.

The paper is structured as follows. In section 2, we recall the most important information that we need from the state of the art about statistics of natural images in order to introduce our work. In section 3, we discuss an estimation strategy of spatiochromatic covariance matrices of images, followed by a measure of their commutativity. In section 4.1, we present the image database that we used in our research. Most importantly, we introduce a crucial pre-processing filtering tool in our procedure: the *sky classifier*. In section 5, we discuss some result about the commutativity change after the application of the Michaelis-Menten transformation. Finally, in section 6 we resume our contributions and put our work in perspective.

2 Spatiochromatic features of natural images

There is a general agreement about the fact that the Human Visual System (HVS from now on) has evolved in order to optimize the elaboration of visual signals coming from scenes of the natural world (which will be shorten with *natural scenes* from now on, following the traditional nomenclature). Attneave [1], MacKay [11] and Barlow [2] pioneered the idea that the HVS may optimize the processing of natural signals by performing a redundancy reduction, however they did not quantify these ideas with a computational theory that can provide a coding for natural images.

Two kind of redundancies can be distinguished in the interaction between humans and the natural environment: firstly, natural scenes have plenty of homogeneous areas and the light reflected from spatial locations belonging to those areas will be interpreted as the same visual information, this implies a strong *spatial correlation*. Secondly, light signals are absorbed by the three L, M, S-type cones in the retina, whose sensitivity is not independent because they are highly overlapping. This implies a strong *chromatic correlation*. When both effects are taken into account, one speaks about *spatiochromatic correlation*.

The literature about natural image statistics is vast and its exhaustive presentation is far beyond the scope of this paper. Here we will deal only with the statistical information gathered by considering the spatial representation of the image and emphasizing the results from [4] and [19], which are essential to understand the development of our paper.

Before describing the results of [4] and [19], let us recall that when principal component analysis (PCA) is performed on small natural image patches, the basic features that result are Fourier descriptors, see for instance [13]). This fact is a consequence of spatial stationarity.

2.1 Chromatic redundancy in natural images

The first statistical information about chromatic redundancy has been experimentally obtained in [12] in the framework of color segmentation of RGB images. For each picture of a database of 8 RGB images, the authors computed the covariance matrix C of the distribution of the values of R, G and B at each pixel. They found that the eigenvectors of the covariance matrix are

approximately the following ones for each image of the database: $\mathbf{v}_1 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^t$, $\mathbf{v}_2 = \left(\frac{1}{2}, 0, -\frac{1}{2}\right)^t$, $\mathbf{v}_3 = \left(-\frac{1}{4}, \frac{1}{2}, -\frac{1}{4}\right)^t$. These vectors correspond to the three following uncorrelated color features: $X_1 = \frac{R+G+B}{3}$, $X_2 = \frac{R-B}{2}$, $X_3 = \frac{2G-(R+B)}{4}$.

This shows that the feature related to the largest variance is the luminance X_1 (or *achromatic channel*) and the other two features are described by the *opponent channels* X_2 (red-blue) and X_3 (green-violet).

Buchsbaum and Gottshalk approached in [4] the problem of finding uncorrelated color features from a purely theoretical point of view. They considered the abstract ensemble of all possible visual stimuli (radiances) $\mathcal{S} \equiv \{S(\lambda),\ \lambda \in \mathcal{L}\}$, where $\mathcal{L}$ is the spectrum of visible wavelengths. From a given representative $S(\lambda) \in \mathcal{S}$, a weighted integration of $S(\lambda)$ over the visual spectrum, with weights given by the Vos-Walraven spectral sensitivity functions $L(\lambda), M(\lambda), S(\lambda)$, yields the three *cone activation values* $L = \int_{\mathcal{L}} S(\lambda)L(\lambda)\, d\lambda$, $M = \int_{\mathcal{L}} S(\lambda)M(\lambda)\, d\lambda$, $S = \int_{\mathcal{L}} S(\lambda)S(\lambda)\, d\lambda$.

Assuming that the stimulus $S(\lambda)$ (coming from a fixed point $\bar{x}$ of a scene) is a random variable, a covariance matrix can be build from the three random variables L, M, S. This matrix, called the *chromatic covariance matrix* is defined as:

$$C = \begin{bmatrix} C_{LL} & C_{LM} & C_{LS} \\ C_{ML} & C_{MM} & C_{MS} \\ C_{SL} & C_{SM} & C_{SS} \end{bmatrix}, \tag{2.1}$$

where $C_{LL} \equiv \mathbb{E}[L \cdot L] - (\mathbb{E}[L])^2$, $C_{LM} \equiv \mathbb{E}[L \cdot M] - \mathbb{E}[L]\mathbb{E}[M] = C_{ML}$, and so on, $\mathbb{E}$ being the expectation operator.

Let $K(\lambda, \mu) = \mathbb{E}[S(\lambda)S(\mu)] - \mathbb{E}[S(\lambda)] \cdot \mathbb{E}[S(\mu)]$ be the *covariance function*, then the entries of the covariance matrix can be written as $C_{LL} = \iint_{\mathcal{L}^2} K(\lambda, \mu)L(\lambda)L(\mu)\, d\lambda d\mu$, and so on.

To be able to perform explicit calculations, the analytical form of the covariance function $K(\lambda, \mu)$ must be specified. In the absence of a database of multispectral images, Buchsbaum and Gottschalk used abstract non-realistic data, i.e. they chose the easiest covariance function corresponding to visual stimuli maximally uncorrelated with respect to their energy at different wavelengths, i.e. $K(\lambda, \mu) = \delta(\lambda - \mu)$, δ being the Dirac distribution. As the authors themselves observe, this condition is satisfied only if the ensemble $\mathcal{S}$ is made of monochromatic signals.

With this choice, the entries of the covariance matrix C are all positives and they can be written as $C_{LL} = \int_{\mathcal{L}} L^2(\lambda)\, d\lambda$, $C_{LM} = \int_{\mathcal{L}} L(\lambda)M(\lambda)\, d\lambda$, and so on. C is also real and symmetric, so it has three positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ with corresponding eigenvectors $\mathbf{v}_i$, $i = 1, 2, 3$. If W is the matrix whose columns are the eigenvectors of C, i.e. $W = [\mathbf{v}_1|\mathbf{v}_2|\mathbf{v}_3]$, then the diagonalization of C is given by $\Lambda = W^t C W = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$.

The eigenvector transformation of the cone excitation values L, M, S, in the special case of monochromatic stimuli, is then

$$\begin{pmatrix} A(\lambda) \\ P(\lambda) \\ Q(\lambda) \end{pmatrix} = W^t \begin{pmatrix} L(\lambda) \\ M(\lambda) \\ S(\lambda) \end{pmatrix}. \tag{2.2}$$

The transformed values A, P, Q are *uncorrelated* and their covariance matrix is Λ. A is the achromatic channel, while P and Q are associated to the opponent chromatic channels.

The *key point* in Buchsbaum and Gottschalk's theory is the application of the *Perron-Frobenius theorem* (see e.g. [3] for more details), which assures that positive matrices, i.e. matrices whose entries are all strictly greater than zero, have one and only one eigenvector whose entries have all the positive sign, and this eigenvector corresponds to the largest eigenvalue, i.e. λ_1. So, only the transformed A channel will be a linear combination of the cone activation values L, M, S with positive coefficients, while the channels P and Q will show opponency. This is the theoretical reason underlying the evidence of post-retinal chromatic opponent behavior, following Buchsbaum and Gottschalk.

Using the data obtained above, Buchsbaum and Gottschalk could write explicitly the trans-

formation from (L, M, S) to (A, P, Q) as follows:

$$\begin{cases} A \simeq 0.887L + 0.461M \\ P \simeq -0.46L + 0.88M \\ Q = 0.004L - 0.01M + 0.99S. \end{cases} \tag{2.3}$$

More information about the relationship between Buchsbaum and Gottschalk's theory and other well-known color spaces can be found in [10].

2.2 Spatiochromatic redundancy in natural images

The most influential paper about spatiochromatic feature is [19], where Ruderman, Cronin and Chiao proposed a *patch-based* spatiochromatic coding and tested Buchsbaum-Gottschalk's theory on a database of 12 multispectral natural images of *foliage*. The authors have shown that the scatterplots in the LM and LS planes of the L, M, S cone activations values (corresponding to 1000 pixels randomly selected in the database) show a high degree of correlation but also asymmetry. So, they decided to study these data by first reducing their asymmetry: they modified the LMS values by taking their decimal logarithm and then they subtracted the average image value in the logarithmic domain. They obtained the so-called *Ruderman-Cronin-Chiao coordinates*, i.e. $\tilde{L} = \operatorname{Log} L - \langle \operatorname{Log} L \rangle$, $\tilde{M} = \operatorname{Log} M - \langle \operatorname{Log} M \rangle$ and $\tilde{S} = \operatorname{Log} S - \langle \operatorname{Log} S \rangle$. Following [19], if $\tilde{L}$, $\tilde{M}$, $\tilde{S}$ are the basis vectors in the logarithmically-transformed space, then the application of the PCA gives the following three principal axes:

$$\begin{cases} \ell = \frac{1}{\sqrt{3}}(\tilde{L} + \tilde{M} + \tilde{S}) \\ \alpha = \frac{1}{\sqrt{6}}(\tilde{L} + \tilde{M} - 2\tilde{S}) \\ \beta = \frac{1}{\sqrt{2}}(\tilde{L} - \tilde{M}). \end{cases} \tag{2.4}$$

The color space spanned by these three principal axes is called $\ell\alpha\beta$ *space*.

The key point of the paper is the idea to study spatiochromatic features by considering 3×3 patches, with each pixel containing a 3-vector color information, so that every patch is converted in a vector with 27 components that were analyzed with the PCA. The principal axes of these small patches in the logarithmic space show that the first fluctuations are in the achromatic channel, followed by blue-yellow fluctuations in the α direction and red-green ones in the β direction.

The spatial axes are largely symmetrical and can be represented by Fourier features, in line with the translation-invariance of natural images, as argued in [5]. It is important to stress that no pixel within the patches appear other than the primary gray, blue-yellow or red-green colors, i.e. no mixing of ℓ, α, β has been found in any 3×3 patch. This means that not only the single-pixel principal axes ℓ, α, β, but also the spatially-dependent principal axes $\ell(x), \alpha(x), \beta(x)$, viewed as functions of the spatial coordinate x inside the patches, are decorrelated. These results have been confirmed by [14].

2.3 Second order stationarity

Let us now analyze the consequence of second order stationarity in natural images on their decorrelated spatiochromatic features. For the sake of clarity, we will first start with the simplest case of gray-level images, where stationarity implies that the principal components are Fourier basis functions. We will then extend this result to the color case and show that a supplementary hypothesis on color covariance matrices yields principal components given by the tensor product between Fourier basis functions on one side, and achromatic plus opponent color coordinates on the other.

2.4 The gray-level case

Let I be a gray-level natural image of dimension $W \times H$. We denote the H rows of I as $r^0, \dots, r^{H-1}$ and the position of each pixel of I row-wise as follows[1] :

$$I = \{r_k^j;\ j = 0, \dots, H-1,\ k = 0, \dots, W-1\}. \tag{2.5}$$

Each row $r^j = (r_0^j, \dots, r_{W-1}^j)$ will be interpreted as a W-dimensional random vector and each component r_k^j as a random variable.

Let us define the *spatial covariance of the two random variables* r_k^j, $r_{k'}^{j'}$:

$$\mathrm{cov}(r_k^j, r_{k'}^{j'}) \equiv c_{k,k'}^{j,j'} = \mathbb{E}[r_k^j r_{k'}^{j'}] - \mathbb{E}[r_k^j]\mathbb{E}[r_{k'}^{j'}]. \tag{2.6}$$

Due to the symmetry of covariance we have $c_{k,k'}^{j,j'} = c_{k',k}^{j',j}$. We write the *spatial covariance matrix of the two random vectors* r^j, $r^{j'}$ as $\mathrm{cov}(r^j, r^{j'}) \equiv C^{j,j'}$, where $C^{j,j'}$ is the $W \times W$ matrix:

$$C^{j,j'} = \begin{bmatrix} c_{0,0}^{j,j'} & c_{0,1}^{j,j'} & \cdots & c_{0,W-1}^{j,j'} \\ c_{1,0}^{j,j'} & c_{1,1}^{j,j'} & \cdots & c_{1,W-1}^{j,j'} \\ \vdots & \vdots & \ddots & \vdots \\ c_{W-1,0}^{j,j'} & \cdots & \cdots & c_{W-1,W-1}^{j,j'} \end{bmatrix}. \tag{2.7}$$

Finally, the *spatial covariance matrix* C of the image I can be written as:

$$C = \begin{bmatrix} C^{0,0} & C^{0,1} & \cdots & C^{0,H-1} \\ C^{1,0} & C^{1,1} & \cdots & C^{1,H-1} \\ \vdots & \vdots & \ddots & \vdots \\ C^{H-1,0} & \cdots & \cdots & C^{H-1,H-1} \end{bmatrix}. \tag{2.8}$$

Notice that C is a $HW \times HW$ matrix because each sub-matrix $C^{j,j'}$ is a $W \times W$ matrix.

Hypothesis 1. From now on, the covariance of I is assumed to be invariant under translations of the row and column index, i.e. $c_{k,k'}^{j,j'} = c_{|k-k'|}^{|j-j'|}$.

Hypothesis 1 is weaker than the typical definition of second order stationarity because here we do not assume the translation invariance of the mean. Alongside this hypothesis, we add a technical requirement on the geometry of digital images which is implicitly assumed every time the Fourier transform is considered, i.e. we will consider a *symmetrized spatial domain with a toroidal distance*, which means that we will perform the identification $r_k^j = r_{k'}^{j'}$ when $j \equiv j' \pmod H$ and $k \equiv k' \pmod W$, i.e. every time there exist $a, b \in \mathbb{Z}$ such that $j' - j = aH$ and $k' - k = bW$.

As a covariance matrix, C is real, symmetric and positive-definite. Now, as a consequence of the previous hypotheses, the matrix C is also block-circulant with circulant blocks. Indeed, the $C^{j,j'}$ are *circulant matrices*, i.e. matrices where each row vector is rotated one element to the right relative to the preceding row vector[2], or, equivalently, each column vector is rotated one element down with respect to the preceding column vector. If we use the convenient shorthand notation 'circ( )' to denote a circulant matrix, by specifying only the first row (or, equivalently, the first column, due to symmetry) between the round brackets, then $C^{j,j'}$ can be written as follows:

$$C^{j,j'} = \mathrm{circ}\left(c_{0,0}^{j,j'}, c_{0,1}^{j,j'}, \dots, c_{0,W-1}^{j,j'}\right). \tag{2.9}$$

[1] To avoid cumbersome repetitions of the indexes variability, from now on, we will suppose that $j, j' \in \{0, \dots H-1\}$ and $k, k' \in \{0, \dots W-1\}$, unless otherwise specified.

[2] This can be easily verified by noticing that $c_{k,k'}^{j,j'} = c_{k+1,k'+1}^{j,j'}$.

Now, if we write $C^j \equiv C^{0,j}$, $j = 0, \ldots, H-1$ it is straightforward to see that the covariance matrix C is block-circulant and can be explicitly written as:

$$C = \mathrm{circ}\left(C^0, C^1, \ldots, C^{H-1}\right). \tag{2.10}$$

It is well known that an $n \times n$ circulant matrix has n eigenvalues corresponding to the components of the DFT of the finite sequence given by the first row of the matrix itself, and its eigenvectors are the Fourier basis functions, see e.g. [6] or [8].

Let us apply this general result to the $W \times W$ circulant matrices C^j. The set of eigenvalue equations $C^j \mathbf{e}_m = \lambda^j_m \mathbf{e}_m$, $\lambda^j \in \mathbb{C}$ and $\mathbf{e} \in \mathbb{C}^W$, $m = 0, \ldots, W-1$, can be written as the following matrix equation $C^j E_W = \Lambda^j E_W$, where[3]:

$$\Lambda^j = \sqrt{W}\mathrm{diag}(\hat{c}^j_m;\ m = 0, \ldots, W-1), \quad \hat{c}^j_m = \frac{1}{\sqrt{W}} \sum_{k=0}^{W-1} c^j_k e^{-\frac{2\pi i m k}{W}}, \tag{2.11}$$

and E_W is the Vandermonde matrix which implements the DFT, i.e. the so-called Sylvester matrix :

$$\begin{aligned} E_W &= [\mathbf{e}_0|\mathbf{e}_1|\cdots|\mathbf{e}_{W-1}] \\ &= \left[\mathbf{e}_m = \frac{1}{\sqrt{W}}\left(1, e^{-\frac{2\pi i m}{W}}, \ldots, e^{-\frac{2\pi i m (W-1)}{W}}\right)^t\right]_{m=0,\ldots,W-1} \\ &= \frac{1}{\sqrt{W}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{-\frac{2\pi i}{W}} & \cdots & e^{-\frac{2\pi i (W-1)}{W}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-\frac{2\pi i (W-1)}{W}} & \cdots & e^{-\frac{2\pi i (W-1)^2}{W}} \end{bmatrix}. \end{aligned} \tag{2.12}$$

The following remark will help us understanding how to extend the previous diagonalization procedure to the whole matrix C.

Remark 1. Let $M = \mathrm{circ}(M^0, \ldots, M^{H-1})$ be a block-circulant matrix and let us assume that the blocks M^j can be diagonalized on the *same* basis B. If we write $E_H = [\mathbf{e}_0|\mathbf{e}_1|\cdots|\mathbf{e}_{H-1}]$, with the vectors $\mathbf{e}_j$ defined as in eq. (2.12) for all $j = 0, \ldots, H-1$, then it can be verified by direct computation that $E_H \otimes B$ is a basis of eigenvectors of M, where $\otimes$ denotes the Kronecker product.

In the case of our spatial covariance matrix C, all the submatrices C^j have the same basis of eigenvectors E_W, thus the result stated in the previous remark can be directly applied on the matric C to guarantee that

$$E_H \otimes E_W = \left[\mathbf{e}_{m,l} = \frac{1}{\sqrt{HW}}\left(1, e^{-2\pi i\left(\frac{m}{W}+\frac{l}{H}\right)}, \ldots, e^{-2\pi i\left(\frac{m(W-1)}{W}+\frac{l(H-1)}{H}\right)}\right)^t\right]_{m,l}, \tag{2.13}$$

for $m = 0, \ldots, W-1$, and $l = 0, \ldots, H-1$ provides a basis of eigenvectors for the matrix C.

Actually, due to the symmetry of covariance matrices, the complex parts of the exponentials cancel out (see [8]) and so the 2D cosine Fourier basis also constitutes a basis of eigenvectors of C:

$$\mathbf{e}_{m,l} = \frac{1}{\sqrt{HW}}\left(1, \cos\left(2\pi\left(\frac{m}{W}+\frac{l}{H}\right)\right), \ldots, \cos\left(2\pi\left(\frac{m(W-1)}{W}+\frac{l(H-1)}{H}\right)\right)\right)^t. \tag{2.14}$$

[3]We have used the simplified notation $c^j_m \equiv c^{0,j}_{0,m}$ to denote the matrix element of position m in the first row of the matrix $C^j \equiv C^{0,j}$, $m = 0, \ldots, W-1$.

2.5 The color case

Let us consider now an RGB image function $\mathbf{u} : \Omega \to [0, 255]^3$, where Ω is the spatial domain, and, for all $(j, k) \in \Omega$, $\mathbf{u}(j, k) = (R(j, k), G(j, k), B(j, k))$ is the vector whose components are the red, green and blue intensity values of the pixel defined by the coordinates (j, k).

We define the *spatiochromatic covariance matrix among two pixels of position* (j, k) *and* (j', k') by extending eq. (2.6) as follows:

$$\mathbf{c}_{k,k'}^{j,j'} = \begin{bmatrix} c_{k,k'}^{j,j'}(R,R) & c_{k,k'}^{j,j'}(R,G) & c_{k,k'}^{j,j'}(R,B) \\ c_{k,k'}^{j,j'}(G,R) & c_{k,k'}^{j,j'}(G,G) & c_{k,k'}^{j,j'}(G,B) \\ c_{k,k'}^{j,j'}(B,R) & c_{k,k'}^{j,j'}(B,G) & c_{k,k'}^{j,j'}(B,B) \end{bmatrix} \tag{2.15}$$

where we defined $c_{k,k'}^{j,j'}(R,R) = \mathbb{E}[R(j,k)R(j',k')] - \mathbb{E}[R(j,k)]\mathbb{E}[R(j',k')]$, $c_{k,k'}^{j,j'}(R,G) = \mathbb{E}[R(j,k)G(j',k')] - \mathbb{E}[R(j,k)]\mathbb{E}[G(j',k')]$, and similarly for the remaining matrix elements. Of course the matrix $\mathbf{c}_{k,k'}^{j,j'}$ is symmetric because $c_{k,k'}^{j,j'}(G,R) = \mathbb{E}[G(j,k)R(j',k')] - \mathbb{E}[G(j,k)]\mathbb{E}[R(j',k')] = c_{k,k'}^{j,j'}(R,G)$, and similarly for all the other off-diagonal elements.

Naturally, we can extend Hypothesis 1 to RGB image case, as follows.

Hypothesis 1 (RGB case). The spatiochromatic covariance of $\mathbf{u}$ of the chromatic channels μ, ν is assumed to be invariant under translations of row and column index, i.e. $c_{k,k'}^{j,j'}(\mu,\nu) = c_{|k-k'|}^{|j-j'|}(\mu,\nu)$, for all $\mu, \nu \in R, G, B$.

With the same technical requirements, we know that $C(\mu, \nu)$ is also block-circulant with circulant blocks.

In the particular case defined by $j' = j$ and $k' = k$, we will call $\mathbf{c}_{k,k'}^{j,j'}$ '*chromatic autocovariance*' and denote it simply as $\mathbf{c}^0$. Notice that the matrix analyzed in [4] is the chromatic autocovariance of the LMS values.

We then define the *spatiochromatic covariance matrix* $\boldsymbol{C}^{j,j'}$ *among the two random vectors* r^j, $r^{j'}$ given by the j-th and j'-the rows of the spatial support of $\mathbf{u}$ by extending eq. (2.7) as follows:

$$\mathbf{C}^{j,j'} = \begin{bmatrix} \mathbf{c}_{0,0}^{j,j'} & \mathbf{c}_{0,1}^{j,j'} & \cdots & \mathbf{c}_{0,W-1}^{j,j'} \\ \mathbf{c}_{1,0}^{j,j'} & \mathbf{c}_{1,1}^{j,j'} & \cdots & \mathbf{c}_{1,W-1}^{j,j'} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{c}_{W-1,0}^{j,j'} & \cdots & \cdots & \mathbf{c}_{W-1,W-1}^{j,j'} \end{bmatrix}. \tag{2.16}$$

Finally, we define the *spatiochromatic covariance matrix* $\boldsymbol{C}$ *of the RGB image* $\boldsymbol{u}$ by extending eq. (2.8) to the $3HW \times 3HW$ matrix defined in this way:

$$C = \begin{bmatrix} \mathbf{C}^{0,0} & \mathbf{C}^{0,1} & \cdots & \mathbf{C}^{0,H-1} \\ \mathbf{C}^{1,0} & \mathbf{C}^{1,1} & \cdots & \mathbf{C}^{1,H-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}^{H-1,0} & \cdots & \cdots & \mathbf{C}^{H-1,H-1} \end{bmatrix}. \tag{2.17}$$

Now, supposing that all the elements of the matrices (2.15) are positive, thanks to the Perron-Frobenius theorem we can assure that each of these $\mathbf{c}_{k,k'}^{j,j'}$ matrices has a basis of eigenvectors that can be written as a triad of achromatic plus opponent chromatic channels. If we further *assume that the matrices (2.15) can be diagonalized on the same basis of eigenvectors* (A, P, Q), then, thanks to Remark 1, we have that the eigenvectors of the spatiochromatic covariance matrix $C(R, G, B)$ can be written as the following Kronecker product:

$$(A, P, Q) \otimes \mathbf{e}_{m,l} \in \mathbb{R}^{3HW}, \tag{2.18}$$

which is precisely the type of eigenvectors that have been exhibited experimentally in [18]. A standard result of linear algebra guarantees that a set of matrices can be diagonalized on the same basis of eigenvectors if and only if they commute[4]. Thanks to the hypothesis of translation invariance of covariance, this is verified if and only if the generic covariance matrix $\mathbf{c}^{j,j'}_{k,k'}$ commutes with the chromatic autocovariance matrix $\mathbf{c}^0$.

The following proposition holds true [16].

Proposition 2.1 *Let $\boldsymbol{u} : \Omega \to [0, 255]^3$ be an RGB image function, with a periodized spatial domain Ω, and suppose that*

1. *All matrices $\boldsymbol{c}^{j,j'}_{k,k'}$ are positive, i.e. their elements are strictly greater than 0;*
2. *The spatiochromatic covariance matrices $\boldsymbol{c}^{j,j'}_{k,k'}$ defined in (2.15) depend only on the distances $|j - j'|$, $|k - k'|$, i.e. the covariance of $\boldsymbol{u}$ is stationary;*
3. *The following commutation property holds:*

$$[\boldsymbol{c}^0, \boldsymbol{c}^{j,j'}_{k,k'}] = 0 \qquad \forall (j, k), (j', k') \in \Omega. \tag{2.19}$$

Then, the eigenvectors of the spatiochromatic covariance matrix $\boldsymbol{C}$ defined in (2.17) can be written as the Kronecker product $(A, P, Q) \otimes \boldsymbol{e}_{m,l}$, where (A, P, Q) is the achromatic plus opponent color channels triad and $\boldsymbol{e}_{m,l}$ is the 2D cosine Fourier basis defined in eq. (2.14).

Proposition 2.1 defines a mathematical framework where the empirical result of Ruderman et al. can be formalized and understood in terms of statistical properties of natural images. In [16], the hypotheses of the proposition above have been checked thanks to simulations performed on databases of natural images: the first two hypotheses have been confirmed, while the third will be discussed in detail in the following part of the paper.

2.6 Commutativity and exponential decay of covariance matrices

The experiments conducted in [16] empirically discovered a linear relationship in the semi-logarithmic scale between spatiochromatic covariance $\mathbf{c}^{j,j'}_{k,k'}(\mu, \nu)$ and pixel distance $d = \sqrt{(j - j')^2 + (k - k')^2}$:

$$\log(\mathbf{c}^d_{\mu\nu}) = \alpha_{\mu\nu} + \beta_{\mu\nu} d. \tag{2.20}$$

This, of course, implies an exponential decay for spatiochromatic covariance that corrected the power-law decay that was commonly supposed to hold true. These analytic expressions will allow us performing a theoretical analysis of the covariance estimators $\hat{\mathbf{c}}^d_{\mu\nu}$, which will play an important role in the discussion of commutativity.

It must be underlined that the exponential decay of $\mathbf{c}^d_{\mu\nu}$ holds true with a great amount of precision only for an intermediate pixel distance range and it slightly deviates from it when d is very small or very large. These deviations are to be expected, because of two different reasons. When d is small, noise and the convolution kernel used by image sensors [15] introduce non linearity between irradiance and pixel intensities; when d is large, the sensor response is altered by optical phenomena as vignetting [7] or an incorrect camera aperture.

This is the reason why, in papers dealing with databases of natural images (see, e.g. [19]), it is common to consider a reduced range of distances to compute statistical features. We will follow the same convention while dealing with experiments. However, for the theoretical part of the paper, we will allow the validity of the exponential decay for any distance $d \geq 0$.

[4]We recall that, given two generic matrices A and B for which the products AB and BA is well defined, $[A, B] \equiv AB - BA$ is called the 'commutator' between them. Of course A and B commute if and only if $[A, B] = 0$.

3 Estimation of spatiochromatic covariances

In this section, we will propose a reliable method to estimate the spatiochromatic covariances and then we will analyze their properties.

3.1 Construction of covariance estimators

Let us start by introducing some notation and nomenclature. $\mathbf{u}_n : \Omega \to [0,1]^3$, $n = 1, 2, \ldots, N$, is the n-th RGB image function with spatial support Ω (common to all images). For each fixed pixel location $x \in \Omega$, $\mathbf{u}_n(x) = (R_n(x), G_n(x), B_n(x))$ is the RGB value of x in the three chromatic channels. The values $\{\mathbf{u}_n(x),\ x \in \Omega\}_{n=1,\ldots,N}$ are i.i.d. image samples with finite population mean and covariance. First of all, we compute the average image $\bar{\mathbf{u}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{u}_n = (\bar{R}, \bar{G}, \bar{B})$ and subtract it from each image to get centered images $\tilde{\mathbf{u}}_n = \mathbf{u}_n - \bar{\mathbf{u}} = (R_n - \bar{R}, G_n - \bar{G}, B_n - \bar{B})$. The μ-channel of the n-th centered image will be written as $\tilde{\mathbf{u}}_{\mu,n}$.

The main task that we must perform is to estimate the coefficients $\beta_{\mu\nu}$ in eq. (2.20). For this, we decided to use the classical ordinary least squares (OLS) estimation [17], which requires the samples in the regression model to be independent. However, in practice, we face the problem that the number of raw images that we have at disposal in our database (and, in general, in the databases publicly available) is not sufficient to provide enough independent samples.

To have a quantitative idea, the final image samples that we have for empirical studies of covariance are 701, suppose that we build a regression model according to the exponential decay and we fit it with the OLS estimation. Suppose also that the distance d ranges from 0 to 100 with step 1 and $\mu, \nu \in \{R, G, B\}$, this implies the need of 909 covariance independent estimators given by $\log(\hat{\mathbf{c}}^d_{\mu\nu})$.

Let us describe how we have computed the estimators accordingly to the OLS prescriptions and overcame this problem. In each centered image, we sample P location pairs given by a center and its neighbor with a fixed step size s.

Centers are fixed in each image and their locations have coordinates

$$\left\{(j_p, k_p),\ j_p = 0, s, 2s, \ldots, \left[\frac{H-1}{s}\right] s,\ k_p = 0, s, 2s, \ldots, \left[\frac{W-1}{s}\right] s,\ p = 1, \ldots, P\right\},$$

where, $P = \left(\left[\frac{H-1}{s}\right] + 1\right)\left(\left[\frac{W-1}{s}\right] + 1\right)$, and [] takes the floor. Neighbours are also fixed throughout all images, while we do not restrict their locations as long as their distance from the centers remains d, for each fixed value of d. We write their locations as (j'_p, k'_p), with the same range variability as (j_p, k_p).

We notice that the set of centers can be identified with a downsampled version of the original image, so that we can consider the stationary hypothesis and its consequences to hold true also for the set of centers.

Finally, we estimate $\mathbf{c}^d_{\mu\nu}$ as follows:

$$\hat{\mathbf{c}}^d_{\mu\nu} = \frac{\sum_{n=1}^{N}\sum_{p=1}^{P} \tilde{\mathbf{u}}_{\mu,n}(j_p, k_p)\tilde{\mathbf{u}}_{\nu,n}(j'_p, k'_p)}{(N-1)P}. \tag{3.1}$$

To simplify the notation, let us write $x^\mu_{np} = \tilde{\mathbf{u}}_{\mu,n}(j_p, k_p)$ for the centers and $y^\nu_{np} = \tilde{\mathbf{u}}_{\nu,n}(j'_p, k'_p)$ for the neighbors.

Finally, the estimators of spatiochromatic covariance matrices are:

$$\hat{\mathbf{c}}^d = \begin{bmatrix} \hat{\mathbf{c}}^d_{RR} & \hat{\mathbf{c}}^d_{RG} & \hat{\mathbf{c}}^d_{RB} \\ \hat{\mathbf{c}}^d_{GR} & \hat{\mathbf{c}}^d_{GG} & \hat{\mathbf{c}}^d_{GB} \\ \hat{\mathbf{c}}^d_{BR} & \hat{\mathbf{c}}^d_{BG} & \hat{\mathbf{c}}^d_{BB} \end{bmatrix}. \tag{3.2}$$

We resume this construction in the scheme visualized in Figure 1, 2.

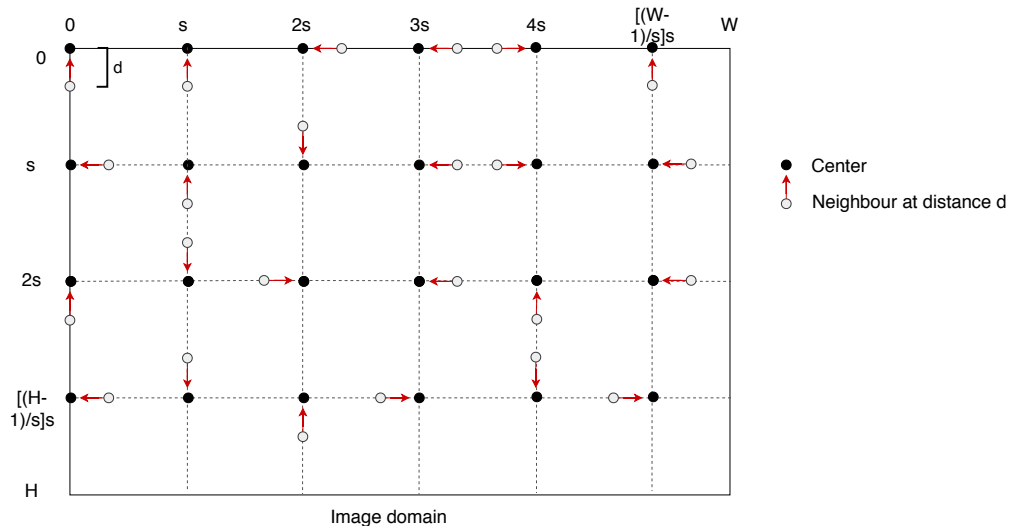

Figure 1: Sampling strategy.

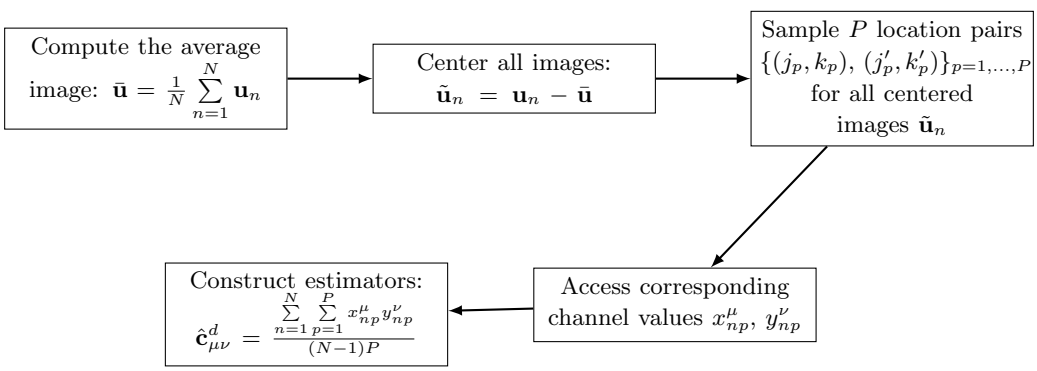

Figure 2: Construction of spatiochromatic covariance estimators.

3.2 Properties of $\hat{\mathbf{c}}^d_{\mu\nu}$

Let us now discuss the properties of the estimators that we have built in the previous section. First of all, $\hat{\mathbf{c}}^d_{\mu\nu}$ are *unbiased*: in fact $\mathbb{E}(\tilde{\mathbf{u}}_n) = 0$, and $\mathrm{cov}(\tilde{\mathbf{u}}_n) = \frac{N-1}{N}\mathrm{cov}(\mathbf{u}_n)$, thus $\mathbb{E}(\hat{\mathbf{c}}^d_{\mu\nu}) = \frac{\sum_{n=1}^{N}\sum_{p=1}^{P} E(x^u_{np} y^v_{np})}{(N-1)P} = \mathbf{c}^d_{\mu\nu}$.

Then, we observe that, during the construction process to compute the estimators, we will keep introducing covariance values. Nevertheless, thanks to the covariance exponential decay, this accumulation will not cause a large variance of the estimators. In fact, we can prove that it exists an upper bound for $\mathrm{cov}(\mathbf{c}^d_{\mu\nu}, \mathbf{c}^{d'}_{\mu'\nu'})$, where μ, ν, d and μ', ν', d' are two different set of parameters, and that this quantity decreases with P, the number of samples.

In order to do that, we first notice that, by direct computation, it can be verified that if C is a block-circulant matrix with circulant blocks, i.e. $C^{j,j'} = \mathrm{circ}(c^{j,j'}_{0,0}, c^{j,j'}_{0,1}, ..., c^{j,j'}_{0,W-1})$ and $C = \mathrm{circ}(C^0, C^1, ..., C^{H-1})$, then $\sum_{j,j'=0}^{H-1}\sum_{k,k'=0}^{W-1} c^{j,j'}_{k,k'} = HW \sum_{j=0}^{H-1}\sum_{k=0}^{W-1} c^{0,j}_{0,k}$.

Since the centers $\{x^{\mu}_{np}\}_p$ constitute a downsampled version of the image $\mathbf{u}_{\mu,n}$, its spatiochromatic covariance is also endowed with properties mentioned in section 2, as well as the one above, thus

$$
\begin{aligned}
\operatorname{cov}(\hat{\mathbf{c}}^{d}_{\mu\nu}, \hat{\mathbf{c}}^{d'}_{\mu'\nu'}) &= \operatorname{cov}\left(\frac{\sum_{n=1}^{N}\sum_{p=1}^{P} x^{\mu}_{np} y^{\nu}_{np}}{(N-1)P}, \frac{\sum_{n'=1}^{N}\sum_{p'=1}^{P} x^{\mu'}_{n'p'} y^{\nu'}_{n'p'}}{(N-1)P}\right) \\
&= \frac{1}{(N-1)^2 P^2}\sum_{n=1}^{N}\sum_{p=1}^{P}\sum_{p'=1}^{P} \mathbb{E}(x^{\mu}_{np} y^{\nu}_{np} x^{\mu'}_{np'} y^{\nu'}_{np'}) - \frac{1}{N}\mathbf{c}^{d}_{\mu\nu}\mathbf{c}^{d'}_{\mu'\nu'},
\end{aligned}
\tag{3.3}
$$

thus

$$
|\operatorname{cov}(\hat{\mathbf{c}}^{d}_{\mu\nu}, \hat{\mathbf{c}}^{d'}_{\mu'\nu'})| \leq \frac{1}{(N-1)^2 P^2}\sum_{n=1}^{N}\sum_{p=1}^{P}\sum_{p'=1}^{P} |\mathbb{E}(x^{\mu}_{np} y^{\nu}_{np} x^{\mu'}_{np'} y^{\nu'}_{np'})| + \frac{1}{N}|\mathbf{c}^{d}_{\mu\nu}||\mathbf{c}^{d'}_{\mu'\nu'}|. \tag{3.4}
$$

Moreover, since $y^{\mu}_{np} \leq 1$ and $y^{\nu'}_{np'} \leq 1$, then $|\mathbb{E}(x^{\mu}_{np} y^{\nu}_{np} x^{\mu'}_{np'} y^{\nu'}_{np'})| \leq |\mathbb{E}(x^{\mu}_{np} x^{\mu'}_{np'})| = \frac{N-1}{N}\mathbf{c}^{\mathrm{dist}[(j_p,k_p),(j_{p'},k_{p'})]}_{\mu\mu'}$.

Therefore $\sum_{p=1}^{P}\sum_{p'=1}^{P} |\mathbb{E}(x^{\mu}_{np} y^{\nu}_{np} x^{\mu'}_{np'} y^{\nu'}_{np'})| \leq \frac{N-1}{N}\sum_{p=1}^{P}\sum_{p'=1}^{P} \mathbf{c}^{\mathrm{dist}[(j_p,k_p),(j_{p'},k_{p'})]}_{\mu\mu'}$. From the remark above about circulant block matrices and from Hypothesis 1 (RGB case), it follows that:

$$
\begin{aligned}
\sum_{p=1}^{P}\sum_{p'=1}^{P} \mathbf{c}^{\mathrm{dist}[(j_p,k_p),(j_{p'},k_{p'})]}_{\mu\mu'} &= P\sum_{p=1}^{P} \mathbf{c}^{\mathrm{dist}[(j_1,k_1),(j_p,k_p)]}_{\mu\mu'} \\
&= P e^{\alpha_{\mu\mu'}} \sum_{l=0}^{\left[\frac{H-1}{s}\right]}\sum_{m=0}^{\left[\frac{W-1}{s}\right]} e^{\beta_{\mu\mu'} s\sqrt{l^2+m^2}}.
\end{aligned}
\tag{3.5}
$$

Since $\beta_{\mu\nu} < 0$, then:

$$
\begin{aligned}
\sum_{l=0}^{\left[\frac{H-1}{s}\right]}\sum_{m=0}^{\left[\frac{W-1}{s}\right]} e^{\beta_{\mu\mu'} s\sqrt{l^2+m^2}} &\leq \sum_{l=0}^{\left[\frac{H-1}{s}\right]}\sum_{m=0}^{\left[\frac{W-1}{s}\right]} e^{\beta_{\mu\mu'} s\sqrt{l^2+0}} \\
&= \sum_{l=0}^{\left[\frac{H-1}{s}\right]} e^{\beta_{\mu\mu'} sl}\left(\left[\frac{W-1}{s}\right]+1\right) \\
&= \frac{1-e^{\beta_{\mu\mu'} s\left(\left[\frac{H-1}{s}\right]+1\right)}}{1-e^{\beta_{\mu\mu'} s}}\left(\left[\frac{W-1}{s}\right]+1\right).
\end{aligned}
\tag{3.6}
$$

Moreover, since $P = \left(\left[\frac{H-1}{s}\right]+1\right)\left(\left[\frac{W-1}{s}\right]+1\right)$, we can get:

$$
\frac{1}{P}\sum_{l=0}^{\left[\frac{H-1}{s}\right]}\sum_{m=0}^{\left[\frac{W-1}{s}\right]} e^{\beta_{\mu\mu'} s\sqrt{l^2+m^2}} \leq \frac{1-e^{\beta_{\mu\mu'} s\left(\left[\frac{H-1}{s}\right]+1\right)}}{\left(1-e^{\beta_{\mu\mu'} s}\right)\left(\left[\frac{H-1}{s}\right]+1\right)}, \tag{3.7}
$$

therefore:

$$
|\operatorname{cov}(\hat{\mathbf{c}}^{d}_{\mu\nu}, \hat{\mathbf{c}}^{d'}_{\mu'\nu'})| \leq \frac{e^{\alpha_{\mu\mu'}}}{(N-1)}\frac{1-e^{\beta_{\mu\mu'} s\left(\left[\frac{H-1}{s}\right]+1\right)}}{\left(1-e^{\beta_{\mu\mu'} s}\right)\left(\left[\frac{H-1}{s}\right]+1\right)} + \frac{1}{N}|\mathbf{c}^{d}_{\mu\nu}\mathbf{c}^{d'}_{\mu'\nu'}|, \tag{3.8}
$$

but $\frac{H-1}{s} - 1 < \left[\frac{H-1}{s}\right] \leq \frac{H-1}{s}$, thus, if we plug the previous inequality into the upper bound in eq. (3.8), then we get:

$$
|\operatorname{cov}(\hat{\mathbf{c}}^{d}_{\mu\nu}, \hat{\mathbf{c}}^{d'}_{\mu'\nu'})| < \frac{e^{\alpha_{\mu\mu'}}}{(N-1)}\frac{1-e^{\beta_{\mu\mu'}(H-1+s)}}{\left(1-e^{\beta_{\mu\mu'} s}\right)\left(\frac{H-1}{s}\right)} + \frac{1}{N}|\mathbf{c}^{d}_{\mu\nu}\mathbf{c}^{d'}_{\mu'\nu'}|. \tag{3.9}
$$

So we get an upper bound for the covariances and, furthermore, when the set of parameters μ, ν, d is equal to the set μ', ν', d', then we automatically have upper bound for the variances of the estimators.

Following an analogous procedure, we can get another upper bound w.r.t to the dimension W, i.e.

$$|\mathrm{cov}(\hat{\mathbf{c}}^d_{\mu\nu}, \hat{\mathbf{c}}^{d'}_{\mu'\nu'})| < \frac{e^{\alpha_{\mu\mu'}}}{(N-1)} \frac{1 - e^{\beta_{\mu\mu'}(W-1+s)}}{(1 - e^{\beta_{\mu\mu'}s})\left(\frac{W-1}{s}\right)} + \frac{1}{N}|\mathbf{c}^d_{\mu\nu}\mathbf{c}^{d'}_{\mu'\nu'}|. \tag{3.10}$$

This upper bound implies that the estimators are under control. Since $\beta_{\mu\mu'} < 0$, this upper bound increases monotonically w.r.t s and thus it decreases w.r.t P. So, it makes sense for us to sample more pairs even in one image.

Let us compute the limit of the upper bound when P tends to infinity. The only part of it that depends on P is the kernel $\frac{1-e^{\beta_{\mu\mu'}(H-1+s)}}{(1-e^{\beta_{\mu\mu'}s})\left(\frac{H-1}{s}\right)}$, thus we confine the computation on it:

$$\begin{aligned} \frac{1 - e^{\beta_{\mu\mu'}(H-1+s)}}{(1 - e^{\beta_{\mu\mu'}s})\left(\frac{H-1}{s}\right)} &= \frac{s}{H-1} + \frac{1 - e^{\beta_{\mu\mu'}(H-1)}}{\frac{e^{-\beta_{\mu\mu'}s}-1}{s}(H-1)} \\ &\xrightarrow{s\to 0} 0 + \frac{1 - e^{\beta_{\mu\mu'}(H-1)}}{-\beta_{\mu\mu'}(H-1)} > 0. \end{aligned} \tag{3.11}$$

In figure 3, we show the behavior of the kernel w.r.t. P for the image dimensions of our database and for some β values of the same magnitude of those reported in [16].

We can notice an asymptotic behavior of the upper bound w.r.t. P, which implies that the error introduced in the computation by limiting ourselves to a value of P that can be managed by an ordinary computer is negligible.

We conclude this section with two remarks. The first one is that, if we set $(\mu, \nu, d) = (\mu', \nu', d')$, then, by equations (3.9) and (3.10), then the covariance $\mathrm{cov}(\hat{\mathbf{c}}^d_{\mu\nu}, \hat{\mathbf{c}}^{d'}_{\mu'\nu'})$ becomes the variance $\mathrm{var}(\hat{\mathbf{c}}^d_{\mu\nu})$ and so, when $N \to +\infty$, the upper bounds tend to 0 and thus the variance will tend to 0 as well.

$$\mathrm{var}(\hat{\mathbf{c}}^d_{\mu\nu}) < \frac{e^{\alpha_{\mu\mu}}}{(N-1)} \frac{1 - e^{\beta_{\mu\mu}(H-1+s)}}{(1 - e^{\beta_{\mu\mu}s})\left(\frac{H-1}{s}\right)} + \frac{1}{N}(\mathbf{c}^d_{\mu\nu})^2, \tag{3.12}$$

$$\mathrm{var}(\hat{\mathbf{c}}^d_{\mu\nu}) < \frac{e^{\alpha_{\mu\mu}}}{(N-1)} \frac{1 - e^{\beta_{\mu\mu}(W-1+s)}}{(1 - e^{\beta_{\mu\mu}s})\left(\frac{W-1}{s}\right)} + \frac{1}{N}(\mathbf{c}^d_{\mu\nu})^2, \tag{3.13}$$

The second remark is that, the previous information plus the unbiasedness of the estimators imply that the estimators $\hat{\mathbf{c}}^d_{\mu\nu}$ converge to $\mathbf{c}^d_{\mu\nu}$ in L^2 sense, so that they are $\sqrt{N}$-consistent. Furthermore, because of the dedicated sampling strategy, $\hat{\mathbf{c}}^d_{\mu\nu}$ are asymptotically normal estimators. By the delta method, we get that $\log(\hat{\mathbf{c}}^d_{\mu\nu})$ is also a $\sqrt{N}$-consistent estimator of $\log(\mathbf{c}^d_{\mu\nu})$, for every strictly positive $\mathbf{c}^d_{\mu\nu}$.

3.3 Regression

We now pass to the analysis of the regression step in order to estimate the slopes $\beta_{\mu\nu}$ with the OLS technique, as previously mentioned, which can be written as follows:

$$\log(\hat{\mathbf{c}}^d_{\mu\nu}) = \alpha_{\mu\nu} + \beta_{\mu\nu} d + \epsilon^d_{\mu\nu}, \quad \mu, \nu \in \{R, G, B\}, \tag{3.14}$$

where d ranges in the intermediate pixel range mentioned in section 2.6. We will denote the OLS estimators of the slopes $\beta_{\mu\nu}$ as $\hat{\beta}^{OLS}_{\mu\nu}$.

We start by pointing out two problems related with the use of $\log(\hat{\mathbf{c}}^d_{\mu\nu})$: the first one is that they are correlated, so that the noise terms $\epsilon^d_{\mu\nu}$ are correlated too. The second one is that $\log(\hat{\mathbf{c}}^d_{\mu\nu})$ is likely to be biased because of the non-linearity of the logarithmic function.

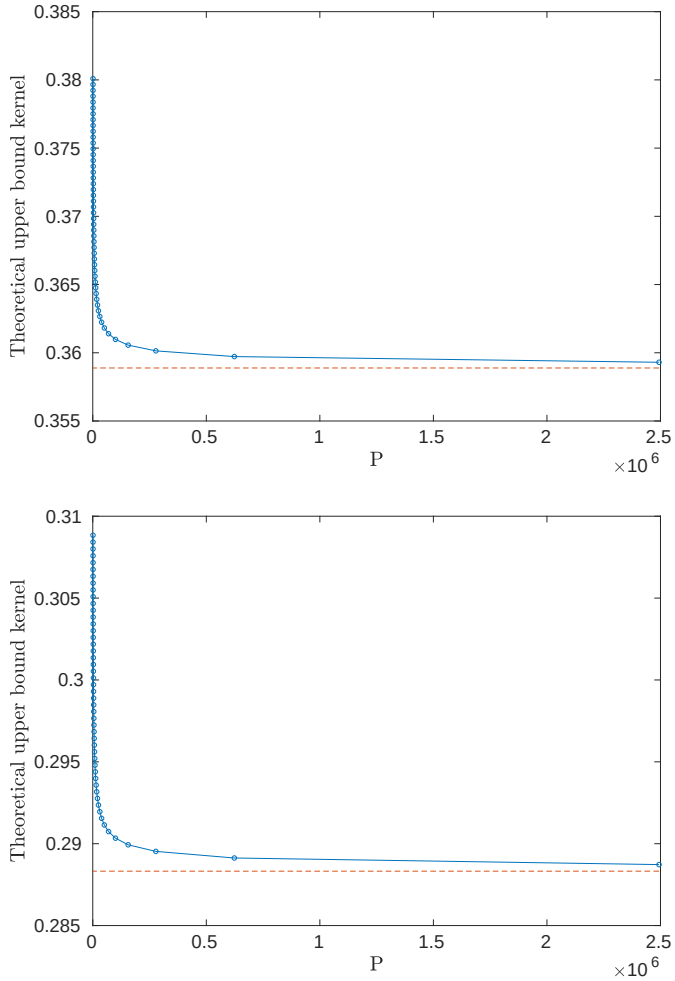

Figure 3: *Theoretical upper bound kernel vs P.* $H = 1288$, $W = 1936$, $\beta = -0.0026$ (top), -0.0020 (bottom). Right dashed line is the limit of the upper bound kernel.

As we will now underline, these two problems will have a limited impact on our computation.

Actually, in spite of the fact that the noise terms are correlated, the OLS estimators are still unbiased, the only adverse effect of correlation is that the variance of $\hat{\beta}_{\mu\nu}^{OLS}$ will become larger. Formally speaking, the $\hat{\beta}_{\mu\nu}^{OLS}$ will not be the so-called BLUE, which stands for Best Linear Unbiased Estimators.

Passing to the second problem, even if $\log(\hat{\mathbf{c}}_{\mu\nu}^{d})$ is biased w.r.t $\log(\mathbf{c}_{\mu\nu}^{d})$, as we previously mentioned, it remains $\sqrt{N}$-consistent. By definition of consistency, if we observe a tiny variance of $\log(\hat{\mathbf{c}}_{\mu\nu}^{d})$, then its biasedness can be ignored.

As we will show in more detail in section 5, the variance of $\log(\hat{\mathbf{c}}_{\mu\nu}^{d})$ that we measured in practice is almost null and so is the variance of $\hat{\beta}_{\mu\nu}^{OLS}$, thus biasedness is not a problem for our computations.

4 Analysis of spatiochromatic covariance matrices commutativity

To study the commutativity of spatiochromatic covariance matrices quantitatively, we need to select a measure. Let us report here some standard definition that will be useful in this section.

Quoting [9], we call a set $F \subset M(n, \mathbb{R})$ of matrices a *commuting family of matrices* if every pair of matrices in F commutes. F is said to be *simultaneously diagonalizable* if there is a single non-singular matrix $V \in M(n, \mathbb{R})$ such that $V^{-1}AV$ is diagonal for every $A \in F$.

From classical linear algebra, it is well known that $F \subset M(n, \mathbb{R})$ is a commuting family if and only if it is a simultaneously diagonalizable family. Moreover, for any given $A \in F$ and for any given ordering $\lambda_1, \ldots, \lambda_n$ of the eigenvalues of A, there is a non-singular matrix $V \in M(n, \mathbb{R})$ such that $V^{-1}AV = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Finally, if A is symmetric, then V is orthogonal, i.e. $V^{-1} = V^T$, the transposed of V.

These considerations allow us the possibility to measure the commutativity properties of the set of estimates $\{\hat{\mathbf{c}}^d\}_d$ without having to compute all the commutators. Instead, we will compute the matrix V which best simultaneously diagonalizes the family of matrix estimates and we will measure the lack of commutativity by this value:

$$\text{JD-obj} = \sum_d \sum_{i \neq j} \left[(V^T \hat{\mathbf{c}}^d V)_{ij} \right]^2, \tag{4.1}$$

where JD-obj stands for joint diagonalisability objective and it is the sum of the square off-diagonal elements of $V^T \hat{\mathbf{c}}^d V$, where d runs from 0 to some maximal distance value used compute the covariances. Of course, in the case of perfect commutativity, JD-obj would be zero while, for an almost-commuting family, the value of JD-obj will be small but not perfectly null.

4.1 Dataset description

The database we used consists of 732 raw images, of size 1288×1936, taken by a Canon 400D. In order to explore the largest possible variety of visual content of natural scenes, we have diversified as much as possible the pictures that we have taken. In Each 4-neighborhood of pixels in a raw image contains two pixels corresponding to the R and B channels and two pixels corresponding to the G channel. Each RAW image was demosaicked to build a subsampled sRGB image simply by keeping unaltered the R and B information and averaging the G channel.

The advantage of raw images is that they are free from post-processing operations such as gamma correction, white balance or compression, thus, modulo camera noise, they provide a much better approximation of irradiance than other images, as e.g. jpeg ones. An excerpt of this database is provided in Figure 4.

However, we found out that the proportion of images containing large areas of the sky (called *sky images* hereafter) dominates the semantic content of the database. Too many sky images will cluster a subset, which will have a different covariance structure than the rest. Therefore, prior to the numeric studies, we need to filter part of sky images out, to balance the database.

For this purpose, we have developed a sky classifier, that we will describe in detail in the appendix.

After filtering, there are 701 images left.

4.2 Implementation

The only hyperparameter that we need to assign beforehand is the step size s. In practice, we have approximately $\min(H, W)$ options for s. Since we need to decrease the variance as much as possible, considering the execution time, we chose $s = 2$ ($P = 623392$).

The computational complexity of eq. (3.1) is $\mathcal{O}(NP)$. We use Matlab 9.4 [5] to implement the method. In the case of our database, it takes around 14 hours to compute 281 (pixel distances)

[5] Codes to reproduce our experiments are available at `https://github.com/yiyej/spatiochromatic_cov`.

Figure 4: Excerpt of the raw image database.

$\times$ 9 (channel combinations) estimates for raw images, while it needs 23 to 24 hours to perform a Michaelis-Menten transformation and the same estimation.

5 Numeric results

In this section we present and discuss the numerical results that we have obtained through our simulations.

5.1 Validations of estimators' properties

To validate the properties of estimator $\hat{\mathbf{c}}^d_{\mu\nu}$, we group n_0 images to mimic once realization. So we have $\frac{N}{n_0}$ realizations in total. Figure 5 shows the unbiasedness of $\hat{\mathbf{c}}^d_{\mu\nu}$. Figure 6 shows the empirical upper bounds of var($\hat{\mathbf{c}}^d_{\mu\nu}$).

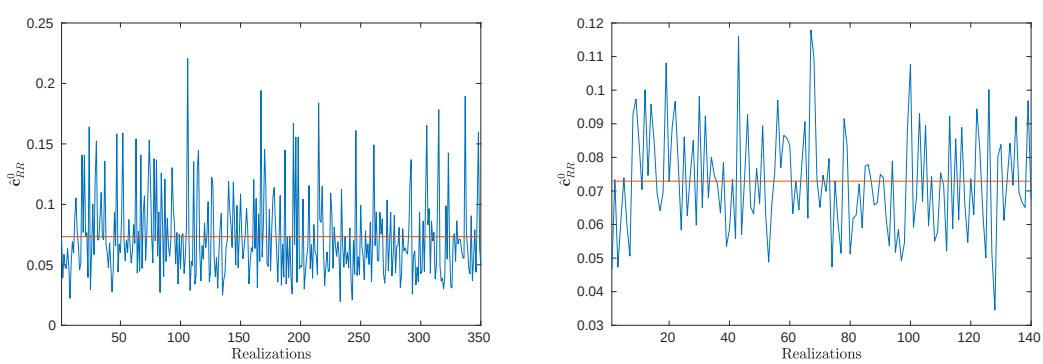

Figure 5: *Unbiasedness.* Estimates of $\hat{\mathbf{c}}^0_{RR}$ from different realizations. Left: $n_0 = 2$ (350 realizations); right: $n_0 = 5$ (140 realizations)

. We can find that the estimator is unbaised.

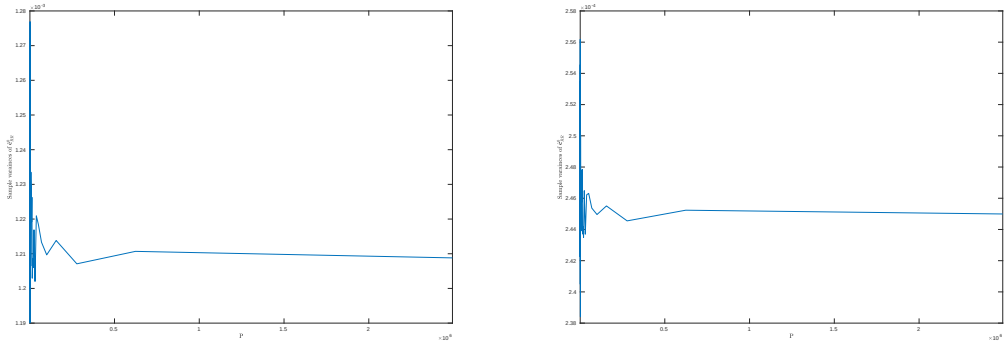

Figure 6: *Empirical upper bounds.* Sample variances of $\hat{\mathbf{c}}^0_{RR}$ vs P. Left: $n_0 = 2$ (350 realizations); right: $n_0 = 5$ (140 realizations). First, the empirical upper bound decreases dramatically as P goes up, sharing the shape with theoretical upper bound kernel. Second, the effect from P on covariances is independent of N. Third, with 5 images, the magnitude of sample variance has already reached 10^{-4} with sufficient large P.

β_{RR}	β_{RG} (β_{GR})	β_{RB} (β_{BR})
-0.00228	-0.00228(-0.00228)	-0.00229(-0.00229)

β_{GG}	β_{GB} (β_{BG})
-0.00222	-0.00218(-0.00218)

β_{BB}
-0.00210

Table 1: Estimates of the slopes from our database of raw images after the application of the sky classifier.

5.2 Exponential decay of spatio-chromatic covariance of raw images

In Figure 7 we show the decay of $\log(\hat{\mathbf{c}}^d_{\mu\nu})$ computed for the raw images of our database, after the sky classifier, with respect to d. Since we find a linear relationship, the exponential decay holds. We built the regression model and fit it with OLS. Estimates of the slopes are provided in Table 1. Notice that, up to the accuracy 10^{-4}, the straight lines relative to the combinations of chromatic channels RR, RG (GR), RB (BR) are parallel and the same is true for those relatives to the combinations GG and GB (BG). All the straight lines are parallel up to the accuracy 10^{-3}.

5.3 Effects of the Michaelis-Menten transformation on commutativity

In this subsection, we will be focusing on the effects of the Michaelis-Menten transformation, $\mathbf{u}_\mu(x) \mapsto \mathbf{u}^\gamma_\mu(x)/(\mathbf{u}^\gamma_\mu(x) + m^\gamma_\mu)$ on the commutativity properties of spatio-chromatic covariance matrices.

Firstly, we compute the JD-obj measure, eq. (4.1) for the original raw images of our database after the action of the sky classifier. Then, we transform the raw images applying the Michaelis-Menten formula with 9 different γ values, ranging from 0.2 to 1 with step 0.1 and we compute again the JD-obj measure.

It is clear that, if $\gamma \to 0$, then the Michaelis-Menten formula will turn all the image pixels to 1/2, thus leaving with constant images and all spatiochromatic covariance matrices would be identical and perfectly commuting. Since we want to avoid this trivial situation, we remain far

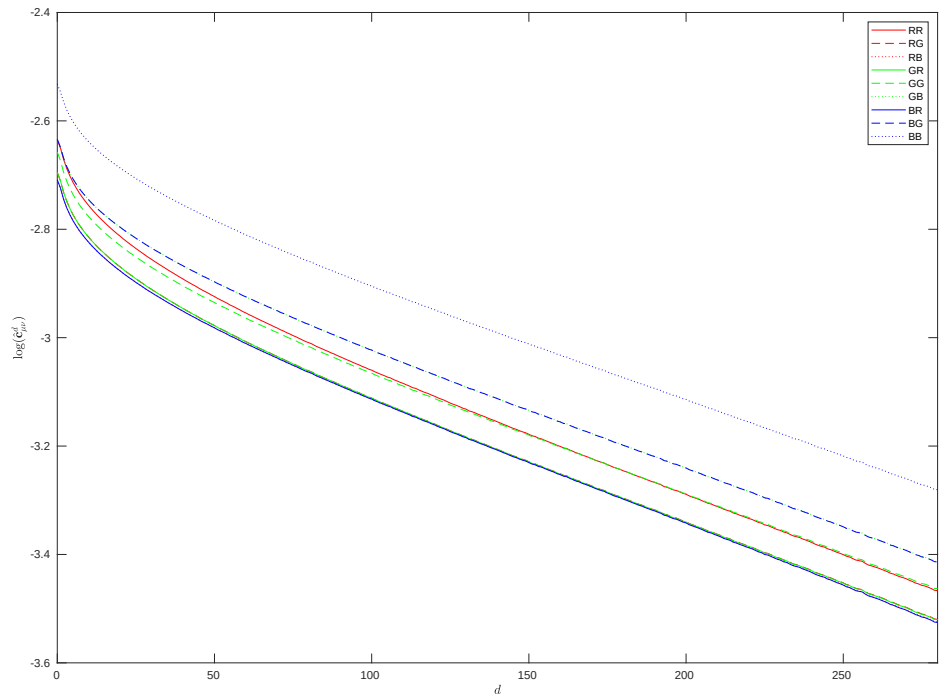

Figure 7: *Exponential decay of raw images.* Parameter setting: $s = 2(P = 623392)$, $d = 0, 1, ...280$. We built the regression model and fit it with OLS on the range $[70, 280]$, where the covariances exhibit exponential decay precisely.

from the value $\gamma = 0$ by starting with $\gamma = 0.2$.

In Figure 8 it can be seen that also after the Michaelis-Menten transformation the exponential decay of covariance holds true[6].

Let us now discuss the quantitative results about the commutativity measure, i.e. the JD-obj values. For the sake of clarity, let us write JD-obj(raw) and JD-obj(MM) for the JD-obj values obtained with the original raw images and the transformed ones, with the Michaelis-Menten transformation, respectively.

We have performed experiments to compute these values on the family of matrices $\{\hat{\mathbf{c}}^d\}_{d=0,\ldots,d_M}$ by changing the value of γ.

One interesting result is that for all d_M and for all γ greater than 0.5 JD-obj(MM) $<$ JD-obj(raw). This remains true also for some values of γ smaller than 0.5 but, in this case, the relationship between JD-obj(MM) and JD-obj(raw) depends on d_M.

In Figure 9 we show the JD-obj(MM) with different γ parameters and with d_M ranging from 10 to 90. The horizontal line is JD-obj(raw). For all these distance values, we empirically verified that when $\gamma \geq 0.6$, then the Michaelis-Menten transformation improves the commutativity of the family $\{\hat{\mathbf{c}}^d\}_{d=0,\ldots,75}$, whilst, for $\gamma \leq 0.5$, the Michaelis-Menten transformation may improve or not the commutativity.

In Figure 10 we show the JD-obj(MM) with different γ parameters and with d_M ranging from 100 to 280. Here there is no horizontal line showing JD-obj(raw) because, for all these distance values and for all γ the Michaelis-Menten transformation improves the commutativity of the family $\{\hat{\mathbf{c}}^d\}_{d=0,\ldots,75}$. However, we can notice that the optimal value of γ, corresponding the absolute minimum of the curve representing JD-obj(MM), gradually shifts towards 0. One

[6]It is worth mentioning that the Michaelis-Menten transformation interchanges the position of some straight lines. For example, when $\gamma = 0.7$, the lines $RB(BR)$ and RR are shifted up w.r.t $RG(GR)$ and GG, respectively.

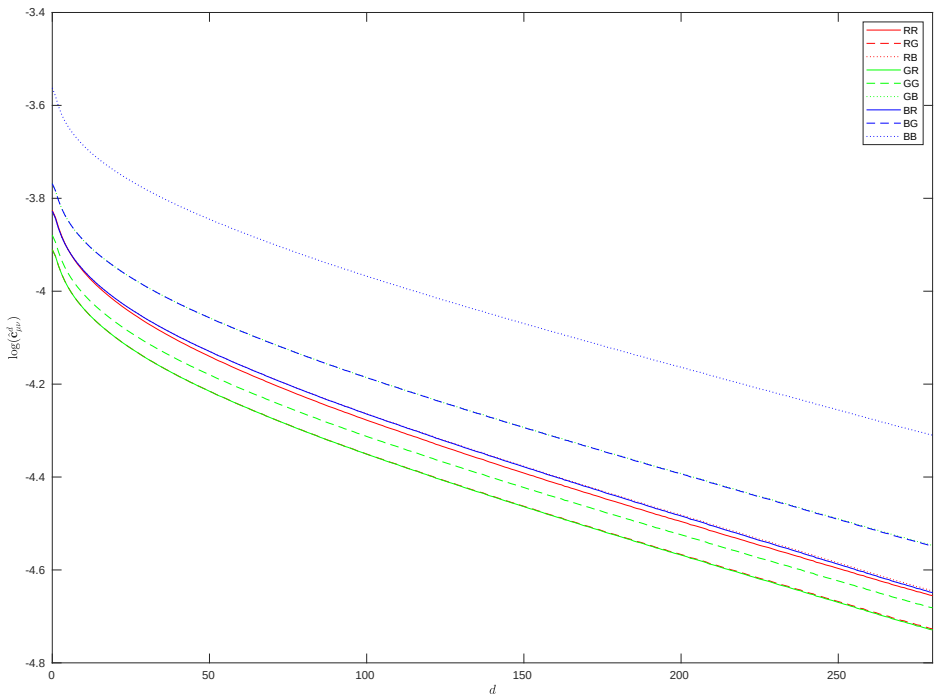

Figure 8: Exponential decay of covariance after the application of the Michaelis-Menten formula with $\gamma = 0.7$. Parameter setting: $s = 2(P = 623392)$, $d = 0, 1, ...280$. The choice of $\gamma = 0.7$ is arbitrary, but we stress that the exponential decay remains true also for all the other γ values that we have considered.

explanation for this behavior is the following: when a large number of matrices is considered, much more noise will be introduced in the computation, in this case, very small values of γ tend to make the family of matrices more homogeneous, thus improving commutativity. Plus, if these matrices do not really commute, then the best way of forcing them to commute is to make them having more homogeneous values.

We believe that, focusing on small d_M will help revealing the true information about the action of the Michaelis-Menten transformation. Interestingly, 0.9 is the first value of γ to be best, before γ decreases back to small values.

In Figure 11 we indicate the best γ value with respect to commutativity for all the distances from 1 to 280.

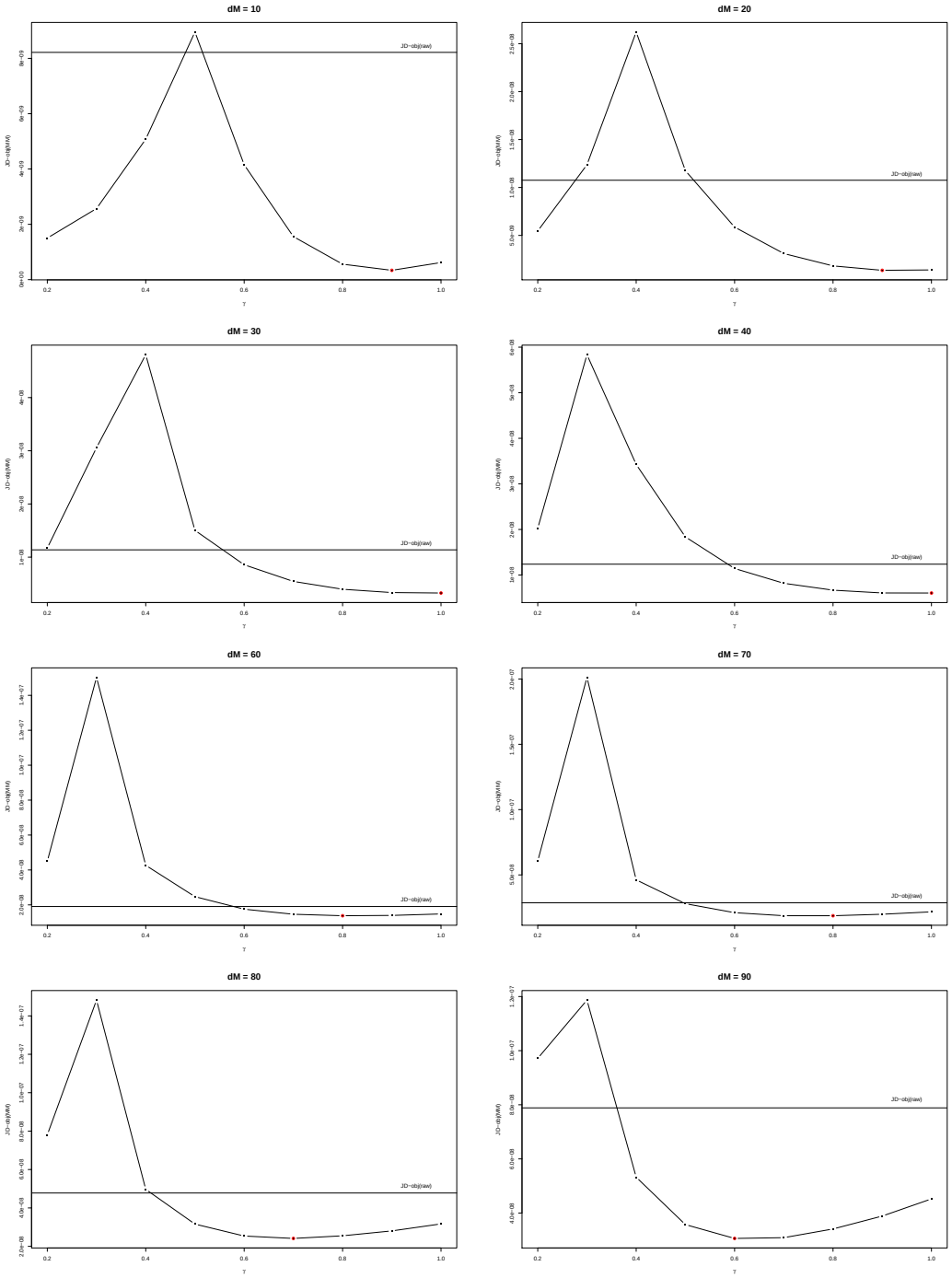

Figure 9: *JD-obj values.* JD-obj(MM) with different γ parameters and with d_M ranging from 10 to 90.

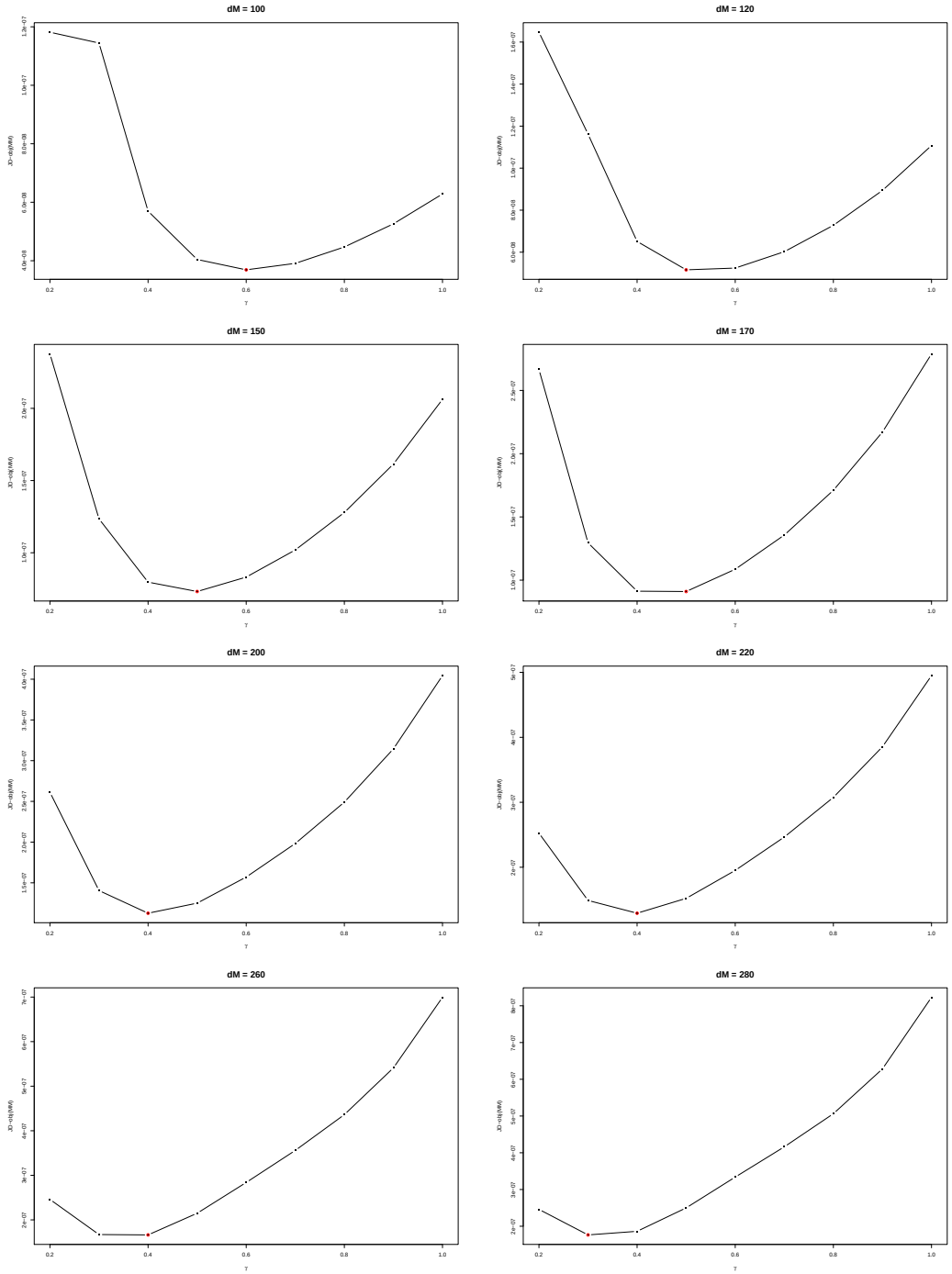

Figure 10: *JD-obj values.* JD-obj(MM) with different γ parameters and with d_M ranging from 100 to 280.

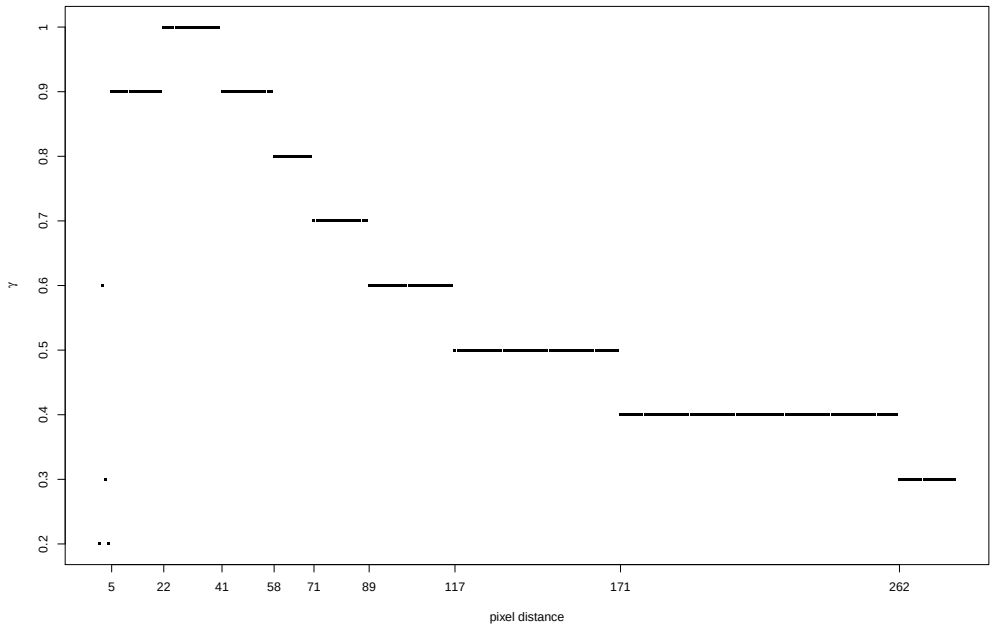

Figure 11: The best γ values for the families $\{\hat{\mathbf{c}}^d\}_{d=0,\ldots,d_M}$. The x-axis records the moving of d_M.

6 Conclusions and perspectives

The work of this paper is inspired by the analysis of spatiochromatic features of natural images provided in [16]. Our contributions to the improvements of this paper are the following.

First of all, we constructed a collection of estimators for spatiochromatic covariance matrices that is reliable from the perspectives of unbiasedness and consistency. This construction is based on a method that permits to exploit as much as possible the information of each image, thus allowing the use of relatively small databases of images, as those typically available for raw or multispectral natural images. Our proposal is general and may be applied to reduces the amount of sample needed by any image statistics model.

Moreover, we devised a sky classifier which allowed us to remove from our database images with statistically redundant information about the sky.

We also verified with great accuracy the exponential decay of spatiochromatic covariance for raw images, showing that, up to a 10^{-3} accuracy, the exponential decay coefficient is the same for each combination of chromatic channels, if we avoid the very first distances which are affected by noise and errors introduced by the convolution kernel of cameras. The consequence is that, up to this accuracy, spatio-chromatic covariance matrices commute and the results of paper [16] about the possibility to separate the codification of spatial and chromatic visual signals into a tensor product hold true.

Finally, we have analyzed the consequences of the application of the Michaelis-Menten transformation to our raw data. If raw data can be associated to the radiance of a visual scene, their Michaelis-Menten transformed can be associated with the output of retinal cones after the absorption of light.

So, it is natural to test if the Michaelis-Menten transformation has an effect on the commutativity of spatio-chromatic covariance matrices, allowing a more precise and efficient tensor product codification of spatial and chromatic visual signals by the optical neurons.

The Michaelis-Menten transformation depends on a parameter γ, which has been measured as 0.74 for the retina of a rhesus monkey. Our tests have confirmed that the exponential decay

is retained after the application of the Michaelis-Menten transformation and, remarkably, that, when γ ranges between 0.6 and 1, the commutativity of spatio-chromatic covariance matrices is actually improved with respect to the original raw image values.

The results that we have obtained are very promising and confirm the conjectures of paper [16] about the importance of the Michaelis-Menten transformation. However, for a full proof of these assumptions a database of natural mulispectral images should be built and analyzed with the techniques described in this paper. Technological limitations of multispectral cameras do not allow this for the moment when movement (e.g. leaves moved by the wind or people walking) is considered.

Appendix

Sky classifier

The key point of this classifier is to control the distribution mass of blue channel and red channel. After a statistical analysis of our database, we found out that, in general, the objects appearing in the pictures are characterized by high values of red, while, of course, the sky is always characterized by high values of blue.

We only consider sky in the daylight. Figure 12 shows one common distribution of daylight sky. We can see that the mass of blue channel distribution of the sky area is located in the high-valued range, so that there is an enough number of pixels capable of exhibiting 'bright blue'. Also, the mass of red channel is located out of the high-valued range. Thus, most of pixels in sky images in our database are characterized by a simultaneous presence of large amount of high-valued blue pixels and a relatively small amount of high-valued red pixels.

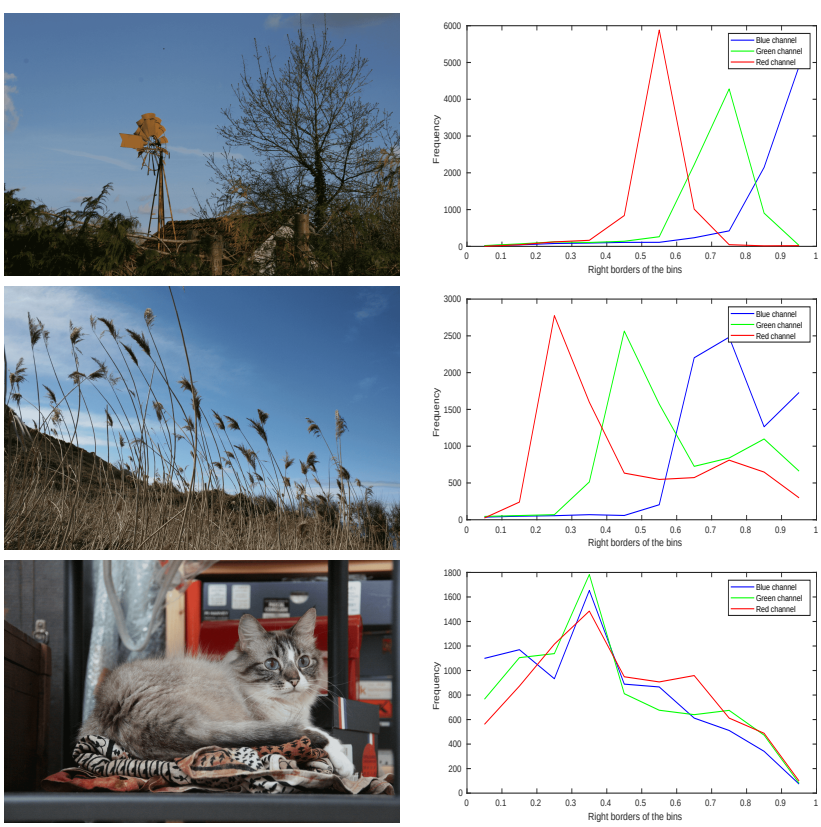

Figure 12: *One typical distribution of sky images.* The histogram is plotted with 10 bins and pixels are taken from the upper 1/3 part of the corresponding image.

Inspired by the analysis above, we defined the Boolean variable that labels sky in our classifier as follows:

$$\text{label} = (\text{Quantile}_B(p_B) > \theta_B) \text{ AND } (\text{Quantile}_R(p_R) < \theta_R),$$

where, $\text{Quantile}_\mu(p_\mu)$ is the quantile of the probability p_μ for the pixel value distribution of the chromatic channels $\mu = B$ or R and θ_B should be located in the high-valued range, θ_R should be located out of high-valued range.

If our database contained only horizontal images, we could limit our sky classifier only to the upper 1/3 part of an image. However, as can be seen from Figure 4, we also have to deal with rotated vertical images, thus, in order to take into account both image geometries, we apply our algorithm only on the top right part of the images. More specifically, we considered the pixels belonging to the top-right part of the image, i.e. the sub-image with coordinates (x, y), $x \geq (1 - \lambda)W$ and $0 \leq y \leq \lambda H$.

We then compute $\text{Quantile}_\mu(p_\mu)$ in each sub-image and the label variable. If the label turns out to be 1, then we take the image out of our database.

Notice that pictures with a significant amount of clouds in the sky will, correctly, not be removed, because in this case there will be a significant amount of bright red pixels.

Moreover, if there is a sufficient amount of detail in the sky, as shown by Figure 13, the classifier will not remove the image from the database.

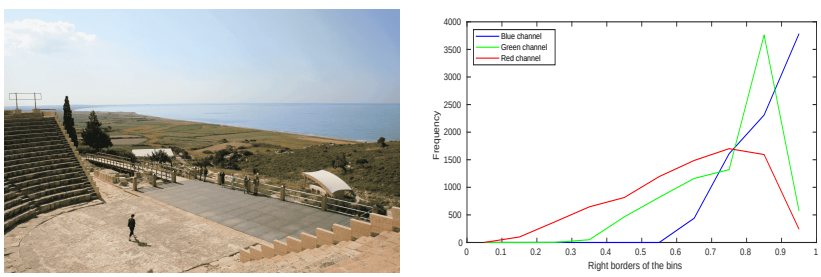

Figure 13: The sky classifier did not remove this image from the database thanks to the presence of the visible gradient which is responsible for the distributions shown in the graph at the right.

By correctly tuning the parameters of the classifier, we can control the proportion of sky images to be taken out. The tuning for our database gave this parameter selection: $\lambda = 0.4$, $p_B = 0.4$, $p_R = 0.6$, $\theta_B = 0.6$, $\theta_R = 0.6$ which classified 31 pictures as sky images over 732, reported in Figure 14.

Figure 14: Pictures taken out of our database because labeled as sky images by our classifier.

Acknowledgments

Jérémie Bigot is a member of Institut Universitaire de France (IUF), and this work has been carried out with financial support from the IUF.

Conflict of interest

The authors declare no conflict of interest.

References

1 F. Attneave. Some informational aspects of visual perception. *Physchol. Rev.*, 61:183–193, 1954.

2 H.B. Barlow. Possible principles underlying the transformations of sensory messages. *Sensory Communications*, pages 217–234, 1961.

3 A. Berman and R.J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. SIAM, 1987.

4 G. Buchsbaum and A. Gottschalk. Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proc. Royal Society of London B*, 220:89–113, 1983.

5 D.J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am.*, 4(12):2379–2394, December 1987.

6 M. W. Frazier. *Introduction to wavelets through linear algebra*. Springer, 2001.

7 R.C. Gonzales and R.E. Woods. *Digital image processing*. Prentice Hall, 2002.

8 R.M. Gray. *Toeplitz and Circulant Matrices: A review*. Now Publishers Inc., 2006.

9 C.R. Johnson and R.A. Horn. *Matrix Analysis*. Cambridge: Cambridge University Press, 1985.

10 Garrett M Johnson, Xioyan Song, Ethan D Montag, and Mark D Fairchild. Derivation of a color space for image color difference measurement. *Color Research & Application*, 35(6):387–400, 2010.

11 D.M. MacKay. Towards an information-flow model of human behaviour. *British Journal of Psychology*, 47:30–43, 1956.

12 Y. Ohta, T. Kanade, and T. Sakai. Color information for region segmentation. *Computer graphics and image processing*, 13:222–241, 1980.

13 B. Olshausen and D.J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research*, 37:607–609, 1997.

14 C. Párraga, T. Troscianko, and D. Tolhurst. Spatiochromatic properties of natural images and human vision. *Current biology*, 6(12):483–487, 2002.

15 W.K. Pratt. *Digital Image Processing*. J. Wiley & Sons, 2007.

16 E. Provenzi, J. Delon, Y. Gousseau, and B. Mazin. On the second order spatiochromatic structure of natural images. *Vision research*, 120:22–38, 2016.

17 C.R. Rao. *Linear statistical inference and its applications*. John Wiley and Sons, 1973.

18 D.L. Ruderman. Origin of scaling in natural images. *Vision Research*, 37:3385–3398, 1996.

19 D.L. Ruderman, T.W. Cronin, and C. Chiao. Statistics of cone responses to natural images: implications for visual coding. *J. Opt. Soc. Am. A*, 15(8):2036–2045, August 1998.

20 R. Shapley and C. Enroth-Cugell. *Visual adaptation and retinal gain controls*, volume 3, chapter 9, pages 263–346. Progress in Retinal Research, 1984.