Vision Research xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

On the second order spatiochromatic structure of natural images

Edoardo Provenzi [a,*], Julie Delon [a], Yann Gousseau [b], Baptiste Mazin [b]

[a] Laboratoire MAP5 (UMR CNRS 8145), Université Paris Descartes, Sorbonne Paris Cité, 45 rue des Saints Pères, 75006 Paris, France
[b] Télécom ParisTech, LTCI, CNRS, 46 rue Barrault, 75013 Paris, France

ARTICLE INFO

Article history:
Received 7 July 2014
Received in revised form 14 February 2015
Available online xxxx

Keywords:
Natural image statistics
Color images
Spatio-chromatic correlation
Second order stationarity
Fourier basis
Opponent channels

ABSTRACT

We provide a theoretical analysis of some empirical facts about the second order spatiochromatic structure of natural images in color. In particular, we show that two simple assumptions on the covariance matrices of color images yield eigenvectors made by the Kronecker product of Fourier features times the triad given by luminance plus color opponent channels. The first of these assumptions is second order stationarity while the second one is commutativity between color correlation matrices. The validity of these assumptions and the predicted shape of the PCA components of color images are experimentally observed on two large image databases. As a by-product of this experimental study, we also provide novel data to support an exponential decay law of the spatiochromatic covariance between pairs of pixels as a function of their spatial distance.

1. Introduction

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 and transmission of visual signals originating from natural scenes by getting rid of redundant information. In fact, there is a hundred million retinal photoreceptors against one million optic nerve neurons, therefore the retinal output must be re-coded to allow the salient visual information passing to subsequent stages.

The origin of redundancy in the interaction between humans and natural scenes is twofold: on one side, natural scenes contain strong *spatial correlation*, since nearby points are likely to send similar radiance information to the eyes, unless they lie in the proximity of a sharp edge. On the other side, light signals (in photopic conditions) are absorbed by the three L,M,S-type cones, whose sensitivity is not independent, see e.g. the picture of Vos-Walraven's cone sensitivity functions depicted in Buchsbaum and Gottschalk (1983, p. 92). In particular, the $L(\lambda)$ and $M(\lambda)$ have a wide overlap area, thus every broadband visual stimulus will excite both the L and the M-type cones, resulting in a strong *chromatic correlation*. When both effects are taken into account, one speaks about *spatio-chromatic correlation*.

The simplest way to look at spatial redundancy within images is through the second order statistics between pixel values. Two noticeable and well-known facts are the *Fourier-like structure* of Principal Component Analysis (PCA), as a result of *spatial stationarity*, and the *power-law decay* of the power spectrum, as a possible consequence of *scale-invariance*. Higher order statistics have also been largely investigated, for instance through wavelets or sparse coding, as recalled in Section 2. On the other hand, several works have been concerned with chromatic redundancy in images, mostly through second order property and in connection with opponent color spaces. However, the spatio-chromatic structure of color images has been less studied. One of the most striking known empirical observation is that the spatio-chromatic covariance matrices resemble a tensor product between a Fourier basis and color opponent channels, as pointed out in Section 2. In this work, we focus on this statistical characteristic, both from a theoretical and an experimental perspective. In Section 3 we show that second order stationarity,[1] together with another technical assumption, namely that covariance matrices must commute for any distance between pixels, implies that the eigenvectors of the spatio-chromatic covariance matrix are represented by the tensor product of the 2D cosine Fourier basis and the triad given by the achromatic plus color opponent channels. In Section 4, we show experimentally by using two large image databases that these two

* Corresponding author.
E-mail addresses: edoardo.provenzi@parisdescartes.fr (E. Provenzi), julie.delon@parisdescartes.fr (J. Delon), yann.gousseau@telecom-paristech.fr (Y. Gousseau), baptiste.mazin@telecom-paristech.fr (B. Mazin).

[1] Second order stationarity is defined as the invariance under translations of both the mean and the covariance of images. For the results presented in Section 3, *only the hypothesis of stationarity of the covariance is needed*. Nevertheless, we will use the term 'second order stationarity' to describe the corresponding hypothesis.

http://dx.doi.org/10.1016/j.visres.2015.02.025
0042-6989/

assumptions hold true and that the tensor structure of covariance matrices is satisfied. One of these bases is made of images gathered from the internet, for which we have no information about the formation process. The other one is made of RAW images that are free of compression artefacts, white balance and gamma correction. As previously said, the second key assumption that guarantees this result is that the spatiochromatic covariance matrices must commute for any distance between pixels. In Section 4, to test this assumption, we will analyze the decay of spatiochromatic covariance matrix elements, showing that it can be modeled through an exponential law, in contrast to the power law decay commonly thought to characterize natural images. In Section 5 we will discuss these theoretical and empirical results and comment about future perspectives for computer vision purposes.

2. State of the art in natural color image statistics

The literature about natural image statistics is vast and its exhaustive presentation is beyond the scope of this paper. We concentrate here on important achievements that are related to the present paper. In Section 2.1, we recall some classical facts about the covariance structure of gray level images and also quote some related and more profound results on the structure of image patches. In Sections 2.2 and 2.3 we discuss results dealing with the chromatic and spatiochromatic redundancy of natural images, respectively. In these two last sections, a particular emphasis is given to the results from Buchsbaum and Gottschalk (1983) and from Ruderman, Cronin and Chiao (1998), which are both closely related to our results.

2.1. Spatial redundancy in natural images

There is a large body of works dealing with spatial statistics in natural images, as e.g. reviewed in Srivastava et al. (2003). In the present work, we will focus mostly on relatively simple second order property of natural images, and mostly on their covariance. Our motivation is that such simple structures are, to the best of our knowledge, not fully understood in the case of spatiochromatic dependency. In particular, we will not consider in this work the non-gaussianity of natural images, i.e. the appearance of sparse features with high kurtosis, although it is related to the most geometric aspects of image structure, see e.g. Mumford and Gidas (2001).

2.1.1. Image patches decomposition

Attneave (1954), MacKay (1956) and Barlow (1961) pioneered the idea that the HVS, in order to deal with the great amount of information that it constantly receives, should have developed a scheme to get rid of redundant information. However, they did not quantify these ideas with a computational theory that could provide a coding for natural images. The simplest observations in this direction concern principal component analysis (PCA) on small image patches. These are well known (see for instance the experiments in Olshausen & Field (1996)) to yield Fourier basis elements. This fact is a consequence of spatial stationarity, as will be recalled in Section 3.1. More elaborated patch decompositions, relying on the minimization of redundancy, as in Atick and Redlich (1990) and Atick (1992), or on sparse decompositions, as in Olshausen and Field (1996), Olshausen and Field (1997), yield localized, band-pass and oriented filters resembling wavelet basis elements.

Since these early works, sparse dictionary representations have become a standard tool for image restoration, for their ability to economically represent geometric structures, see e.g. Elad and Aharon (2006). Analogous elementary patches have been obtained with the use of Independent Component Analysis (ICA), see e.g. Hyvärinen, Hurri, and Hoyer (2009) or from convolutional neural networks, see the experiments in Krizhevsky, Sutskever, and Hinton (2012).

2.1.2. Power spectrum

One of the most striking fact about image statistics is that most pictures exhibit some form of scale-invariance. Roughly speaking, statistical observations on an image and on a zoomed version of it are qualitatively similar. The most well known among such observations concerns the power spectrum of images. Experiments conducted over different databases of natural images, see e.g. Field (1987) and Ruderman and Bialek (1994), have shown that, while the power spectra of different images change considerably, if we compute the average power spectrum over a sufficiently large number of images and all orientations, then we find a power-law behavior. That is, if we write $S(f)$ for the power spectrum (the square of the amplitude of the Fourier transform) averaged over all directions, we find that $S(f) \approx f^{-\beta}$, as a function of the frequency magnitude f. The value of β varies roughly from 1.5 to 3, with a cluster of values around 2, depending on the image database used, see Tolhurst, Tadmor, and Chao (1992) and Pouli, Cunningham, and Reinhard (2010) for some examples of β values. This decreasing behavior of the power spectrum is usually associated with scale invariance, since the value $\beta = 2$ corresponds to this case, see Mumford and Gidas (2001).

By Wiener-Khinchin's theorem (see Papoulis, 1991), under the hypothesis of second order stationarity, power spectrum and covariance form a Fourier pair. In Ruderman (1996, p. 3397), it is underlined that if the power spectrum of an image follows a power law $S(f) \approx \frac{1}{f^{2-\eta}}, \eta$ being the so-called 'anomalous exponent', then the covariance C as a function of the distance d among pixels has the following expression $C(d) = \frac{a}{d^{\eta}} - b,\ a, b > 0$, i.e. apart from an offset, the covariance also decreases with a power-law. This power-law decay has been proven to fail at large frequencies and distances, both for the power spectrum, e.g. in Langer (2000), and the covariance, see Huang and Mumford (1999) and Huang (2000). In Section 4, we will confirm the failure of the power-law decay of the covariance at large distances. We will show that an exponential model is more accurate, and discuss the relation of such a model with the spatiochromatic covariance properties of natural images.

2.2. Chromatic redundancy in natural images

The first statistical information about chromatic redundancy has been experimentally obtained in Ohta, Kanade, and Sakai (1980) 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:

$$\begin{cases} \mathbf{v}_1 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^t \\ \mathbf{v}_2 = \left(\frac{1}{2}, 0, -\frac{1}{2}\right)^t \\ \mathbf{v}_3 = \left(-\frac{1}{4}, \frac{1}{2}, -\frac{1}{4}\right)^t, \end{cases} \tag{1}$$

These vectors correspond to the three following uncorrelated color features

$$\begin{cases} X_1 = \frac{R+G+B}{3} \\ X_2 = \frac{R-B}{2} \\ X_3 = \frac{2G-(R+B)}{4}. \end{cases} \tag{2}$$

Please cite this article in press as: Provenzi, E., et al. On the second order spatiochromatic structure of natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.02.025

This shows that the feature that corresponds 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 Gottschalk (1983) approached the problem of finding uncorrelated color features from a purely theoretical point of view. Following the already quoted ideas of Attneave, Barlow and MacKay, they analyzed the problem of an efficient post-retinal information transmission by performing a PCA on the LMS *cone activation values*. We shall now give a detailed presentation of this work, to which our contributions are closely related.

Buchsbaum and Gottschalk considered the abstract ensemble of all possible visual stimuli (radiances), i.e. $\mathcal{S} \equiv \{S(\lambda),\ \lambda \in \mathcal{L}\}$, 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:

$$\begin{cases} L = \int_{\mathcal{L}} S(\lambda) L(\lambda)\, d\lambda \\ M = \int_{\mathcal{L}} S(\lambda) M(\lambda)\, d\lambda \\ S = \int_{\mathcal{L}} S(\lambda) S(\lambda)\, d\lambda. \end{cases} \tag{3}$$

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 *chromatic covariance matrix*, is defined as:

$$C = \begin{bmatrix} C_{LL} & C_{LM} & C_{LS} \\ C_{ML} & C_{MM} & C_{MS} \\ C_{SL} & C_{SM} & C_{SS} \end{bmatrix}, \tag{4}$$

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.

If we introduce the *covariance function*, $K(\lambda, \mu) = \mathbb{E}[S(\lambda)S(\mu)] - \mathbb{E}[S(\lambda)] \cdot \mathbb{E}[S(\mu)]$, then the entries of the covariance matrix can be written as $C_{LL} = \iint_{\mathcal{L}^2} K(\lambda, \mu) L(\lambda) L(\mu)\, d\lambda d\mu, C_{LM} = \iint_{\mathcal{L}^2} K(\lambda, \mu) L(\lambda) M(\mu)\, d\lambda d\mu$, and similarly for the others. This shows that the correlation among the $L,\ M,\ S$ cone activations does not depend only on the overlap among the sensitivity curves, but also on the prevalence of certain wavelengths in the ensemble of visual stimuli $\mathcal{S}$ with respect to others.

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 to compute $K(\lambda, \mu)$. They chose the easiest covariance function, i.e. $K(\lambda, \mu) = \delta(\lambda - \mu), \delta$ being the Dirac distribution, corresponding to visual stimuli maximally uncorrelated with respect to their energy at different wavelengths. As the authors observe, this condition is satisfied only if the ensemble $\mathcal{S}$ is made of monochromatic signals, thus it is higly non-realistic with respect to natural stimuli.

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 $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3$ 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 $L,\ M,\ S$ is then

$$\begin{pmatrix} A \\ P \\ Q \end{pmatrix} = W^t \begin{pmatrix} L \\ M \\ S \end{pmatrix}.$$

In the special case of monochromatic stimuli, we can also parameterize with the wavelength λ:

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

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 *Perron–Frobenius theorem* (see e.g. Berman & Plemmons (1987) for more details), which assures that positive matrices,[2] 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.

We underline that the positivity of C in Buchsbaum–Gottschalk's theory is a consequence of their non-realistic selection of monochromatic visual stimuli. However, signals coming from real scenes are broad-band, so there is no theoretical guarantee that C has all positive entries. In Section 4, we will see that C is positive also when it is computed through natural RGB images, in which case the HVS sensitivity functions are replaced with the sensitivity functions of cameras.

The monochromatic signal energy of the channels has the following property:

$$\int_{\mathcal{L}} A^2(\lambda)\, d\lambda : \int_{\mathcal{L}} P^2(\lambda)\, d\lambda : \int_{\mathcal{L}} Q^2(\lambda)\, d\lambda = \lambda_1 : \lambda_2 : \lambda_3. \tag{5}$$

The explicit form of the matrices C, W^t and Λ within Buchsbaum–Gottschalk's theory are the following:

$$C = \begin{pmatrix} 77.0622 & 38.6204 & 0.0649 \\ 38.6204 & 22.8099 & 0.0646 \\ 0.0649 & 0.0646 & 0.0151 \end{pmatrix}, \tag{6}$$

$$W^t = \begin{pmatrix} 0.887 & 0.461 & 0.0009 \\ -0.46 & 0.88 & 0.01 \\ 0.004 & -0.01 & 0.99 \end{pmatrix}, \tag{7}$$

$$\Lambda = \mathrm{diag}(97.2,\ 2.78,\ 0.015). \tag{8}$$

The large covariance values between L and M and the very small ones between these two channels and S are a direct consequence of the use of Vos-Walraven's sensitivity functions and the hypothesis $K(\lambda, \mu) = \delta(\lambda - \mu)$. In Section 4 we will see that if we compute C from a database of natural images, then the difference among covariance values is rather small.

Using the data obtained above, Buchsbaum and Gottschalk could write explicitly the transformation from $(L,\ M,\ S)$ to $(A,\ P,\ Q)$ as follows:

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

the energy ratios among $A,\ P$ and Q being $97.2 : 2.78 : 0.015$. Again, we observe that the unrealistic hypothesis of maximally

[2] Notice that the definition of a *positive matrix* A, i.e. a matrix with all positive entries, is different than that of a *positive-definite matrix A*, for which the following property is required: $x^t A x > 0$, for every non zero vector x.

Please cite this article in press as: Provenzi, E., et al. On the second order spatiochromatic structure of natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.02.025

uncorrelated visual signals implies that the achromatic channel accounts for the great majority of the energy transmitted and the blue channel has practically no influence in the computation of the achromatic stimulus.

2.3. Spatio-chromatic redundancy in natural images

Buchsbaum also developed the first computational model of spatio chromatic image coding in early vision in the paper Derrico and Buchsbaum (1991). In that paper, only the L and M signals are taken into account, because the authors claim that they contain almost the whole energy of retinal output and the opponent $L-M$ ganglion cell receptive fields[3] represent 90% of the total ganglion cells on–off receptive fields. Their model consists in a two-stage process: the first (chromatic) step consists in the same PCA operated in Buchsbaum and Gottschalk (1983) to decorrelate the L and M signals, which gives the achromatic part $L+M$ and the opponent chromatic channel $L-M$. The second (spatial) step consists in applying a linear predictive coding (similar to that used by Srinivasan, Laughlin, & Dubs (1982)) which optimizes the transmission of the achromatic part by attenuating the low spatial frequencies.

Also Atick, Li, and Redlich (1992) considered only the L and M signals and, by postulating translation-invariance of natural light stimuli and separability among chromatic and spatial correlation, they built a linear operator able to decorrelate the signals L and M into $L+M$ and $L-M$ (see Atick, 1992, p. 245; Atick, Li, & Redlich, 1992, p. 566).

In Ruderman, Cronin and Chiao (1998), Ruderman, Cronin and Chiao proposed a *patch-based* spatio-chromatic coding and tested Buchsbaum–Gottschalk's theory on a database of 12 multispectral natural images of *foliage*.

Because of the proximity between our contribution and this work, we now give a detailed account of the experiments performed in Ruderman, Cronin and Chiao (1998). Each multispectral image of their database consists of 43 successive images taken at 7–8 nm intervals from 403 to 719 nm, thus they chose $\mathcal{L}=\{403,410,\dots,719\}$ nm. In order to build the cone activation values L, M, S, the authors did not follow the same procedure as Buchsbaum and Gottschalk, but they write $(L,\ M,\ S)=\sum_{\lambda\in\mathcal{L}}\vec{Q}(\lambda)R(\lambda)J(\lambda)$, where $\vec{Q}=(Q_L,\ Q_M,\ Q_S)$ are the cone sensitivity functions detailed in Stockman, MacLeod, and Johnson (1993), $J(\lambda)$ is the standard D65 CIE illuminant used to model daylight spectrum, and $R(\lambda)$ is an estimation of scene's reflectance. $R(\lambda)$ is obtained by placing in each photographed scene a MacBeth chart with known spectral reflectance and re-calibrating the multispectral values in order to match those of the chart. Of course this procedure is approximated, since the illumination of the scene can vary in space and time, for this reason the authors analyzed only the 128×128 central region of each image.

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 (higher in the LM plane than in the LS one due to the overlap of L and M spectral sensitivity functions) but also asymmetry, see Ruderman, Cronin and Chiao (1998, p. 2037).

The authors 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. $\tilde{L}=\mathrm{Log}L-\langle\mathrm{Log}L\rangle$, $\tilde{M}=\mathrm{Log}M-\langle\mathrm{Log}M\rangle$ and $\tilde{S}=\mathrm{Log}S-\langle\mathrm{Log}S\rangle$. This transform is motivated by the fact that, following Weber–Fechner's law, uniform logarithmic changes in stimulus intensity tend to be equally perceptible, see Goldstein (2013). Moreover, second-order statistics of log-transformed data is similar to that of linear images, see Ruderman and Bialek (1994). Instead, the motivation for the average substraction is to assess the data independently on the illumination level, analogously to a von Kries procedure (see von Kries, 1902).

The choice of logarithmic coordinates is nevertheless questionable. Ruderman, Cronin and Chiao claim that the linear relationship among logarithmic data and Weber–Fechner's contrast metric gives a reason to select the principal component analysis among other non-orthogonal analysis because the orthogonal transformations involved in the PCA preserve the space metric. However, other researcher, e.g. Simoncelli and Olshausen (2001) have criticized this observation, claiming that considering high-level perceptual features, as Weber-Fechner's law, in early vision models is misleading.

Following Ruderman, Cronin and Chiao (1998), 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:

$$\begin{cases} l=\frac{1}{\sqrt{3}}\left(\tilde{L}+\tilde{M}+\tilde{S}\right)\\ \alpha=\frac{1}{\sqrt{6}}\left(\tilde{L}+\tilde{M}-2\tilde{S}\right)\\ \beta=\frac{1}{\sqrt{2}}\left(\tilde{L}-\tilde{M}\right). \end{cases} \tag{9}$$

The color space spanned by these three principal axes is called $l\alpha\beta$ *space*. The standard deviations of the l, α, β coordinates are $\sigma_l=0.353$, $\sigma_\alpha=0.0732$ and $\sigma_\beta=0.00745$. Notice that there is an *inversion in the importance of opponent channels* with respect to Buchsbaum and Gottschalk (1983): here the $L-M$ channel has the lowest variance.

To study spatiochromatic decorrelated features, Ruderman, Cronin and Chiao considered 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 they analyzed with the PCA. The principal axes of these small patches in the logarithmic space are depicted in Ruderman, Cronin and Chiao (1998) at page 2041. The *color rendering* that the authors performed is the following: firstly, the R, G, B values are linearly related to the logarithmic $\tilde{L}$, $\tilde{M}$, $\tilde{S}$ values via these formulae $R=128(\tilde{L}+1)$, $G=128(\tilde{M}+1)$, $B=128(\tilde{S}+1)$. Notice that $\tilde{X}=0$ if and only if $\mathrm{Log}X=\langle\mathrm{Log}X\rangle$, for $X=L,\ M,\ S$, so the previous relations set the average image values in the logarithmic domain to 128. Finally, the $(R,\ G,\ B)$ values so obtained are linearly stretched to [0,255]. The first principal axis shows fluctuations 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 Field (1987). It is important to stress that in the results of Ruderman et al. 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. These means that not only the single-pixel principal axes l,α,β, but also the spatially-dependent principal axes $l(x)$, $\alpha(x)$, $\beta(x)$, viewed as functions of the spatial coordinate x inside the patches, are decorrelated.

These results have been confirmed by Párraga, Troscianko, and Tolhurst (2002). A strong, but *not perfect*, spatio-chromatic decorrelation has been confirmed in Hyvärinen, Hurri, and Hoyer (2009, p. 323), where the authors performed experiments on 50000 patches of size 12×12 selected in a basis of 20 RGB (and

[3] The typical representation of a ganglion cell $L-M$ receptive field is given by a center disk surrounded by a ring. The center is excited by the information arriving from the L (resp. M) cones, but its response is inhibited by the information arriving from the M (resp. L) cones on the surrounding ring.

Please cite this article in press as: Provenzi, E., et al. On the second order spatiochromatic structure of natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.02.025

not LMS) images. The imperfection in the decorrelation is put in evidence by the appearance of mixed colors, as e.g. orange. In Section 4, we will perform similar experiments on much larger databases.

Ruderman, Cronin and Chiao proposed the following separable form for the spatio-chromatic principal axes: $p((L,M,S),x) = c(L,M,S)\cdot s(x)$, i.e. the product of two uncorrelated eigenfunctions, namely $c(L,M,S)$, given by the principal axes l,α,β and $s(x)$, given by the Fourier basis. They also suggest that the lack of spatial dependence of the chromatic components can be a consequence of scale invariance in natural images. Such a separable basis has been recently exploited also in Chakrabarti, and Zickler (2011) in the context of hyperspectral images representation and reconstruction.

In Wachtler, Lee, and Sejnowski (2001), the authors applied the ICA to study a set of 8 multispectral images of terrestrial natural scenes containing mainly plants and rocks. The measured values of the ICA basis functions for single pixels are coherent with those of Ruderman, Cronin and Chiao, however, they have proven that if one considers patches of 7×7 pixels, then colors other than the principal ones can appear.

The Kronecker product relationship between spatial and chromatic correlations is also assumed (without proof) in the very recent and interesting treatise Li (2014).

3. Relationship between second order stationarity and the decorrelated spatiochromatic features of natural images

In this section we will 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.

3.1. 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[4]:

$$I=\left\{r_k^j;\ j=0,\dots,H-1,\ k=0,\dots,W-1\right\}. \tag{10}$$

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* $r_k^j, r_{k'}^{j'}$:

$$cov(r_k^j,r_{k'}^{j'})\equiv c_{k,k'}^{j,j'}=\mathbb{E}\left[r_k^j r_{k'}^{j'}\right]-\mathbb{E}[r_k^j]\mathbb{E}[r_{k'}^{j'}]. \tag{11}$$

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 $cov(r^j,r^{j'})\equiv C^{j,j'}$, where $C^{j,j'}$ is the $W\times W$ matrix:

$$C^{j,j'}=\begin{bmatrix} c_{0,0}^{j,j'} & c_{0,1}^{j,j'} & \cdots & c_{0,W-1}^{j,j'}\\ c_{1,0}^{j,j'} & c_{1,1}^{j,j'} & \cdots & c_{1,W-1}^{j,j'}\\ \vdots & \vdots & \ddots & \vdots\\ c_{W-1,0}^{j,j'} & \cdots & \cdots & c_{W-1,W-1}^{j,j'}\end{bmatrix}. \tag{12}$$

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

$$C=\begin{bmatrix} C^{0,0} & C^{0,1} & \cdots & C^{0,H-1}\\ C^{1,0} & C^{1,1} & \cdots & C^{1,H-1}\\ \vdots & \vdots & \ddots & \vdots\\ C^{H-1,0} & \cdots & \cdots & C^{H-1,H-1}\end{bmatrix}. \tag{13}$$

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 will be tested in Section 4.3 and, as said before, it 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 $j\equiv j'$ (mod H) and $k\equiv k'$ (mod 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,[5] 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:

$$C^{j,j'}=\mathrm{circ}\left(c_{0,0}^{j,j'},c_{0,1}^{j,j'},\dots,c_{0,W-1}^{j,j'}\right). \tag{14}$$

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

$$C=\mathrm{circ}\left(C^0,C^1,\dots,C^{H-1}\right). \tag{15}$$

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. Frazier (2001) or Gray (2006).

Let us apply this general result to the $W\times W$ circulant matrices C^j. The set of eigenvalue equations $C^j\mathbf{e}_m=\lambda_m^j\mathbf{e}_m, \lambda^j\in\mathbb{C}$ and $\mathbf{e}\in\mathbb{C}^W, m=0,\dots,W-1$, can be written as the following matrix equation $C^jE_W=\Lambda^jE_W$, where[6]:

$$\Lambda^j=\sqrt{W}\,\mathrm{diag}(\hat{c}_m^j;\ m=0,\dots,W-1),\quad \hat{c}_m^j=\frac{1}{\sqrt{W}}\sum_{k=0}^{W-1}c_k^j e^{-\frac{2\pi i m k}{W}}, \tag{16}$$

and E_W is the Vandermonde matrix which implements the DFT, i.e.

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

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

[6] We have used the simplified notation $c_m^j\equiv c_{0,m}^{0,j}$ to denote the matrix element of position m in the first row of the matrix $C^j\equiv C^{0,j}, m=0,\dots,W-1$.

Please cite this article in press as: Provenzi, E., et al. On the second order spatiochromatic structure of natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.02.025

$$\begin{aligned} E_W &= [\mathbf{e}_0|\mathbf{e}_1|\cdots|\mathbf{e}_{W-1}] \\ &= \left[\mathbf{e}_m = \tfrac{1}{\sqrt{W}}\left(1, e^{-\frac{2\pi i m}{W}}, \ldots, e^{-\frac{2\pi i m(W-1)}{W}}\right)^t\right]_{m=0,\ldots,W-1} \\ &= \frac{1}{\sqrt{W}}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{-\frac{2\pi i}{W}} & \cdots & e^{-\frac{2\pi i(W-1)}{W}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-\frac{2\pi i(W-1)}{W}} & \cdots & e^{-\frac{2\pi i(W-1)^2}{W}} \end{bmatrix}. \end{aligned} \tag{17}$$

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. (17) 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

$$E_H \otimes E_W = \left[\mathbf{e}_{m,l} = \frac{1}{\sqrt{HW}}\left(1, e^{-2\pi i\left(\frac{m}{W}+\frac{l}{H}\right)}, \ldots, e^{-2\pi i\left(\frac{m(W-1)}{W}+\frac{l(H-1)}{H}\right)}\right)^t\right]_{m,l}, \tag{18}$$

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 Gray, 2006) and so the 2D cosine Fourier basis also constitutes a basis of eigenvectors of C:

$$\begin{aligned} \mathbf{e}_{m,l} = \frac{1}{\sqrt{HW}}\Bigg(1, \cos\left(2\pi\left(\frac{m}{W}+\frac{l}{H}\right)\right), \ldots, \\ \cos\left(2\pi\left(\frac{m(W-1)}{W}+\frac{l(H-1)}{H}\right)\right)\Bigg)^t. \end{aligned} \tag{19}$$

3.2. The color case

Let $\mathbf{u} : \Omega \to [0, 255]^3$ be an RGB image function, 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. (11) as follows:

$$\boldsymbol{c}^{j,j'}_{k,k'} = \begin{bmatrix} c^{j,j'}_{k,k'}(R,R) & c^{j,j'}_{k,k'}(R,G) & c^{j,j'}_{k,k'}(R,B) \\ c^{j,j'}_{k,k'}(G,R) & c^{j,j'}_{k,k'}(G,G) & c^{j,j'}_{k,k'}(G,B) \\ c^{j,j'}_{k,k'}(B,R) & c^{j,j'}_{k,k'}(B,G) & c^{j,j'}_{k,k'}(B,B) \end{bmatrix} \tag{20}$$

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

In the particular case defined by $j' = j$ and $k' = k$, we will call $\boldsymbol{c}^{j,j'}_{k,k'}$ '*chromatic autocovariance*' and denote it simply as $\boldsymbol{c}^0$. Notice that the matrix analyzed in Buchsbaum and Gottschalk (1983) is the chromatic autocovariance of 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. (12) as follows:

$$\boldsymbol{C}^{j,j'} = \begin{bmatrix} \boldsymbol{c}^{j,j'}_{0,0} & \boldsymbol{c}^{j,j'}_{0,1} & \cdots & \boldsymbol{c}^{j,j'}_{0,W-1} \\ \boldsymbol{c}^{j,j'}_{1,0} & \boldsymbol{c}^{j,j'}_{1,1} & \cdots & \boldsymbol{c}^{j,j'}_{1,W-1} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{c}^{j,j'}_{W-1,0} & \cdots & \cdots & \boldsymbol{c}^{j,j'}_{W-1,W-1} \end{bmatrix}. \tag{21}$$

Finally, we define the *spatiochromatic covariance matrix* $\boldsymbol{C}$ *of the RGB image* $\mathbf{u}$ by extending Eq. (13) to the $3HW \times 3HW$ matrix defined in this way:

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}^{0,0} & \boldsymbol{C}^{0,1} & \cdots & \boldsymbol{C}^{0,H-1} \\ \boldsymbol{C}^{1,0} & \boldsymbol{C}^{1,1} & \cdots & \boldsymbol{C}^{1,H-1} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{C}^{H-1,0} & \cdots & \cdots & \boldsymbol{C}^{H-1,H-1} \end{bmatrix}. \tag{22}$$

Now, supposing that all the elements of the matrices (20) are positive, thanks to the Perron-Frobenius theorem we can assure that each of these $\boldsymbol{c}^{j,j'}_{k,k'}$ 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 (20) 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 $\boldsymbol{C}$ can be written as the following Kronecker product:

$$(A,P,Q) \otimes \mathbf{e}_{m,l} \in \mathbb{R}^{3HW}, \tag{23}$$

which is precisely the type of eigenvectors that have been exhibited experimentally in Ruderman (1996). 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.[7] Thanks to the hypothesis of translation invariance of covariance, this is verified if and only if the generic covariance matrix $\boldsymbol{c}^{j,j'}_{k,k'}$ commutes with the chromatic autocovariance matrix $\boldsymbol{c}^0$.

It is convenient to resume all the hypotheses made and results obtained so far in the following proposition.

Proposition 1. *Let* $\mathbf{u} : \Omega \to [0,255]^3$ *be an RGB image function, with a periodized spatial domain* Ω*, and suppose that*

1. *The spatiochromatic covariance matrices matrices* $\boldsymbol{c}^{j,j'}_{k,k'}$ *defined in (20) depend only on the distances* $|j-j'|, |k-k'|$*, i.e. the covariance of* $\mathbf{u}$ *is stationary;*
2. *All matrices* $\boldsymbol{c}^{j,j'}_{k,k'}$ *are positive, i.e. their elements are strictly greater than 0;*
3. *The following commutation property holds:*

$$\left[\boldsymbol{c}^0, \boldsymbol{c}^{j,j'}_{k,k'}\right] = 0 \qquad \forall (j,k),(j',k') \in \Omega. \tag{24}$$

Then, the eigenvectors of the spatiochromatic covariance matrix $\boldsymbol{C}$ *defined in (22) can be written as the Kronecker product* $(A,P,Q) \otimes \mathbf{e}_{m,l}$*, where* (A,P,Q) *is the achromatic plus opponent*

[7] 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 $[A,B] = 0$.

color channels triad and $\mathbf{e}_{m,l}$ *is the 2D cosine Fourier basis defined in Eq. (19).*

Proposition 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 the following section we will test this framework with the help of two large databases of RGB images.

A direct corollary of Proposition 1 is the following.

Corollary 1. *If the hypotheses of Proposition 1 are valid, then the following decomposition formula holds:*

$$\mathbf{u} = \sum_{m=0}^{W-1}\sum_{l=0}^{H-1}\langle \mathbf{u}, (A,P,Q)\otimes \mathbf{e}_{m,l}\rangle (A,P,Q)\otimes \mathbf{e}_{m,l}, \tag{25}$$

where $\langle\,,\rangle$ *is the scalar product in* $\mathbb{R}^{3HW}$.

4. Validations on natural image databases

As stated in the previous section, the validity of Proposition 1, which expresses the spatiochromatic basis as the Kronecker product of the 2D cosine Fourier basis by the triad given by one achromatic plus two opponent color channels, is constrained by some hypotheses. In this section we present the tests that we have performed to check their validity.

To perform our numerical experiences we have selected two databases that are best suited for different scopes. The first one is an excerpt from the database described in Hays and Efros (2007), which consists of 2.3 millions of 1024×768 copyright-free RGB images taken from the popular website Flickr. The presence of semantically similar images in the database can induce biases in the statistical results, for this reason, in our experiments, we have used packets of images randomly selected from the database. As it will be detailed later on, we have also performed experiments with several subsets to check the coherence of the results. The advantage of this first database is its large number of images, which enabled us to check the stability of our results. An excerpt of this database is provided in Fig. 1.

The second database is made of personal RAW photographs of more than 2000 natural scenes[8]. The content of the natural images that constitute the RAW database is given by landscapes, people, woods, parks and rivers. The database has been built by means of two cameras, a Canon 400D and a Nikon D300.

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. We demosaicked each RAW image to build a subsampled RGB image simply by keeping unaltered the R and B information and averaging the G channel.

The advantage of this second database is that RAW images 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 the images of the first database. Using RAW images from two different cameras will permit us to test the influence of camera sensitivity curves on the results of our analysis. An excerpt of this database is provided in Fig. 2.

Before describing our tests, we show in Fig. 3 the 64 first principal components obtained by applying a PCA to a packet of 10^5 images from the Flickr database. These components indeed have the shape predicted by Proposition 1, i.e. they show, in separated component, oscillations of increasing frequency of the achromatic and color opponent channels. Notice that all the principal components that we have obtained respect the previsions of Proposition 1, we show just the first 64 components because the variance of features represented by the subsequent components decreases rapidly.

4.1. Computation of the covariance matrices

Since the test that we have performed are related to covariance matrices, it is logical to start by describing how these matrices are obtained from the images of the databases. First of all, in order to simplify their computation, the expectation of the empirical average image of the ensemble has been subtracted to all images.

We first concentrated on the evaluation of the chromatic autocovariance $\boldsymbol{c}^0$. Even if the simplest way to compute the covariance is via the Fourier transform, which implicitly assumes periodicity and may lead to biases, we chose the alternative strategy that we are going to describe in detail hereafter. First of all, the expectation operator $\mathbb{E}$ involved in its computation has been approximated by randomly selecting a pixel in N different images of the database, storing its RGB values in three N-dimensional row vectors $v_\mu, \mu\in\{R,G,B\}$, and then estimating the elements of the chromatic autocovariance matrices as follows: $\boldsymbol{C}_{RR} = v_R v_R^t/N$, $\boldsymbol{C}_{RG} = v_R v_G^t/N$, and so on. Notice that it is possible to compute the covariances in this way because of the initial subtraction of the average image.

We then turned our attention to the spatiochromatic covariance matrices $\boldsymbol{c}^{j,j'}_{k,k'}$ with $j\neq j'$ and $k\neq k'$. To simplify the notation, from now on we will write $\boldsymbol{c}^{j,j'}_{k,k'}\equiv \boldsymbol{c}^d$, where $d=\sqrt{(j-j')^2+(k-k')^2}$. We compute $\boldsymbol{c}^d$ by randomly selecting a different point $(j,k)\in\Omega$ in each image of the packet and by considering its four neighbors $(j-d,k),(j+d,k),(j,k-d),(j,k+d)$, we then count how many of these neighbors actually fall in the image domain Ω and we create the vectors v_μ and $w_\mu, \mu\in\{R,G,B\}$: in w_μ we store the R,G,B values of the neighbors of (j,k) that fall in Ω, while in v_μ we store the R,G,B values of the central pixel (j,k) repeated as many times as the length of w_μ. We iterate the procedure for all the images of the packet and we concatenate the values of the random pixels and their neighbors in the vectors v_μ and w_μ, respectively. The estimation of the matrix elements of $\boldsymbol{c}^d$ can be done as follows: $\boldsymbol{C}^d_{RR} = v_R w_R^t/L$, $\boldsymbol{C}^d_{RG} = v_R w_G^t/L$, and so on, where here L denotes the common length of v_μ and w_μ.

4.2. Stability of the covariance computation with respect to the number of images and the image content

As previously stated, the very large Flickr database allows us checking the stability of the covariance matrices computation. We will now introduce the details of the stability tests. If we fix a threshold $\varepsilon = 10^{-D}, D\in\mathbb{N}$, then we consider the estimation of $\boldsymbol{c}^0$ *ε-stable* when the relative error defined by $\|\boldsymbol{c}^0_{N+1}-\boldsymbol{c}^0_N\|/\|\boldsymbol{c}^0_{N+1}\|$ is smaller than ε, where $\boldsymbol{c}^0_N$ is the estimation of $\boldsymbol{c}^0$ obtained with N images and $\|\ \|$ is the 2-norm. Due to the law of large numbers, we expect the relative error to decrease proportionally to $1/N$. This is confirmed by our experiments, as can be seen in Fig. 4. A good trade-off between precision and computational time required to perform the experiments is given by $N=10^5$, which implies a 10^{-4}-stable estimation of the covariance.

This also explains why small databases of multispectral or calibrated images, see e.g. Foster et al. (2006); Tkacik et al., 2011; Olmos and Kingdom, 2004 and Vazquez-Corral et al. (2009),

[8] The database of RAW images is available at the following website: http://download.tsi.telecom-paristech.fr/RawDatabase/.

Fig. 1. Excerpt of the Flickr database.

Fig. 2. Excerpt of the RAW database.

though very interesting, cannot be used for the scopes of the present paper.

With this value of N fixed, we tested the stability of the computation with respect to the image content by selecting 10 different packets of N pictures and comparing the estimation of $\boldsymbol{c}^0$. Our tests have reported differences in the estimation of $\boldsymbol{c}^0$ of magnitude 10^{-4}, which is the same order as the stability error, this confirms that the coherence of our results with respect to different subsets of images.

4.3. Testing the spatial invariance of covariance

Here we discuss the tests that we have performed in order to check the translation invariance of the covariance. In Fig. 5 we show the spatial distributions of the chromatic autocovariances and their linearly stretched version obtained by setting the minimum to 0 and the maximum to 255 in order to enhance the visibility of the inhomogeneities. Without stretching the images appear constant. Analogous results have been obtained for the spatiochromatic covariances with distance $d > 0$.

Please cite this article in press as: Provenzi, E., et al. On the second order spatiochromatic structure of natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.02.025

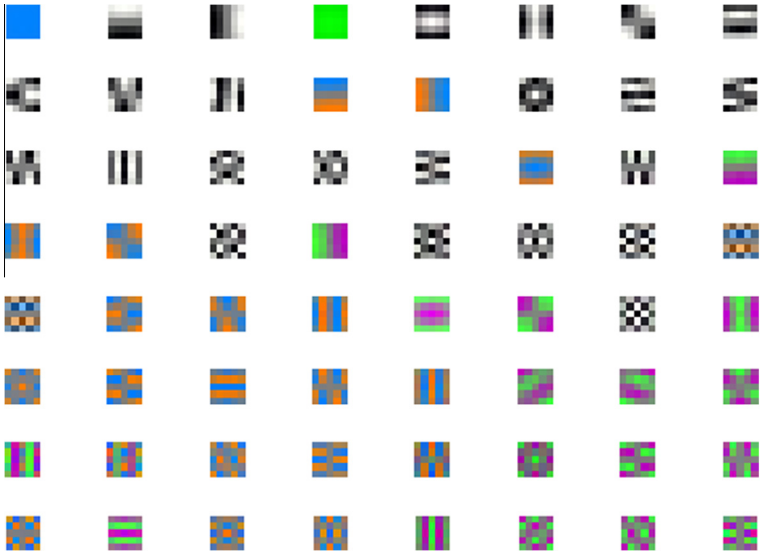

Fig. 3. Given a set of 4,000 96 × 128 images taken from the Flickr database, we have extracted from them 1,900,000 non overlapping patches of dimension 5 × 5. We have then performed the centered PCA on the set of these patches and normalized the principal components between 0 and 1 for visualization purposes. Coherently with the results of the paper, the principal components are represented by 2D Fourier patterns in the achromatic and color opponent channels. A similar PCA has been obtained in Hyvarinen (2009), chapter 15.

All covariances are slightly larger in the upper part of the images and smaller in lower parts, this is due to the presence of the sky in many images of the database. We stress again the fact that the pictorial representation of the second column of Fig. 5 is *exaggerated* by the stretching and that the constant pattern shown by the images of the first column confirms that the hypothesis of translational invariance of covariance can be considered verified with a very good degree of approximation, the variance of the the covariance values along the images being of order 10^{-4}.

4.4. The chromatic autocovariance matrix $\boldsymbol{c}^0$ and its eigenvalues and eigenvectors

The expressions of the chromatic autocovariance matrices relative to the Flickr and RAW databases, $\boldsymbol{c}^0_{\text{Flickr}}$ and $\boldsymbol{c}^0_{\text{RAW}}$ respectively, that we have obtained are the following:

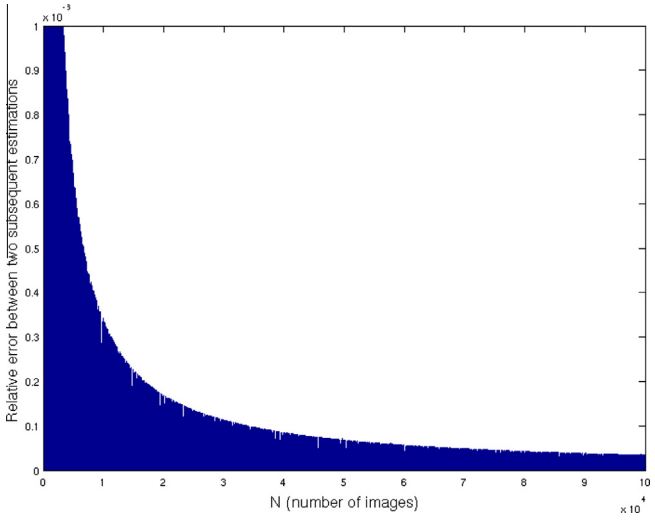

Fig. 4. Relative error decay in the estimation of the chromatic autocovariance matrix elements.

$$\boldsymbol{c}^0_{\text{Flickr}} = \begin{bmatrix} 0.0719 & 0.0651 & 0.0612 \\ 0.0651 & 0.0713 & 0.0710 \\ 0.0612 & 0.0710 & 0.0851 \end{bmatrix} \tag{26}$$

$$\boldsymbol{c}^0_{\text{RAW}} = \begin{bmatrix} 0.0022 & 0.0021 & 0.0021 \\ 0.0021 & 0.0021 & 0.0022 \\ 0.0021 & 0.0022 & 0.0024 \end{bmatrix}, \tag{27}$$

which confirm the positivity assumption on $\boldsymbol{c}^0$. Notice that the covariances observed on the RAW database are much smaller than those observed on the Flickr database. We believe that this is mostly due to the fact that the contrast of images posted on Flickr is often much higher than that of unprocessed RAW images. Despite this difference, the eigenvectors of the previous matrices are very similar:

$$\begin{cases} A_{\text{Flickr}} = (0.5483, 0.5761, 0.6061) \leftrightarrow \lambda_1 = 0.2080, \\ P_{\text{Flickr}} = (0.7179, 0.0474, -0.6945) \leftrightarrow \lambda_2 = 0.0170, \\ Q_{\text{Flickr}} = (0.4289, -0.8160, 0.3876) \leftrightarrow \lambda_3 = 0.0034. \end{cases} \tag{28}$$

and

$$\begin{cases} A_{\text{RAW}} = (0.5679, 0.5683, 0.5954) \leftrightarrow \lambda_1 = 0.0065, \\ P_{\text{RAW}} = (0.7210, 0.0055, -0.6930) \leftrightarrow \lambda_2 = 0.0002, \\ Q_{\text{RAW}} = (0.3971, -0.8228, 0.4066) \leftrightarrow \lambda_3 = 7.8 \cdot 10^{-7}. \end{cases} \tag{29}$$

We can see that, using a database of real RGB images and not the idealized visual stimuli of Buchsbaum and Gottschalk, the blue channel not only appears in the achromatic direction A, but it is even its dominant component. We also notice the similarity between the eigenvectors obtained with the Flickr and RAW database and those reported in Ohta, Kanade, and Sakai (1980), which where also obtained from RGB camera images.

4.5. The exponential decay of spatiochromatic covariance matrix elements

All the spatiochromatic matrices $\boldsymbol{c}^d$ that we have estimated turned out to be *positive*. The most prominent cause of positive correlation values is probably the fact that the spectral sensitivity functions of cameras are also highly overlapping, see

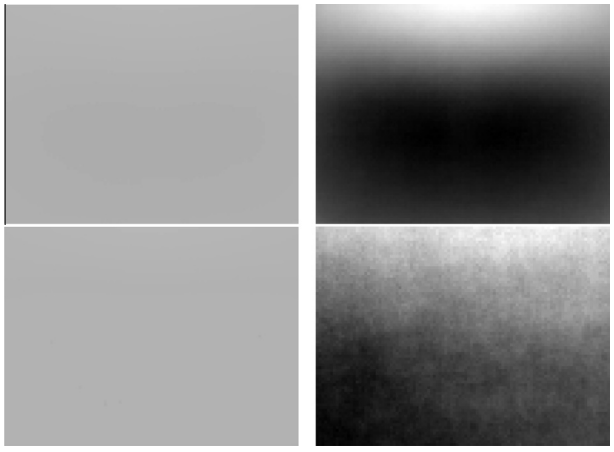

Fig. 5. First column: distribution of chromatic autocovariances computed through the Flickr (first row) and RAW (second row) databases between the red channels. Second column: the intensity of the images of the first column has been stretched between 0 and 255 to enhance the visibility of the spatial inhomogeneity, that is not possible to perceive with the non-stretched values. We stress that we do not introduce the other color channel combinations because they are very similar to those shown here.

for instance Jiang et al. (2013). Their decay with respect to increasing values of d is reported in the linear, bi-logarithmic and semi-logarithmic scale in Fig. 6 for the Flickr database and in Fig. 7 for the RAW database.

Let us write the generic element of the matrix $\boldsymbol{c}^d$ as $\boldsymbol{c}^d_{\mu\nu}, \mu, \nu \in \{R, G, B\}$. Notice that a power-law behavior for $c^d_{\mu\nu}$ would be represented by a linear relationship in the bi-logarithmic scale, i.e. $\log(\boldsymbol{c}^d_{\mu\nu}) = \alpha_{\mu\nu} + \beta_{\mu\nu}\log(d), \beta_{\mu\nu} < 0$, which, in fact, is equivalent to $\boldsymbol{c}^d_{\mu\nu} = e^{\alpha_{\mu\nu}} d^{\beta_{\mu\nu}}$. However, as can be seen in Figs. 6 and 7, the graphs in the bi-logarithmic scale show a significant deviation from a linear behavior from $d = 100$, these distance being expressed in pixels. This confirm in the color case the fact that *the power-law decay of the covariance is not valid for large pixels distances*, a fact already noticed for gray level images in Huang and Mumford (1999).

Moreover, notice that the graphs of Figs. 6 and 7 in the semi-logarithmic scale show a linear decay for all the distances that we have tested (from 1 to 300 pixels). To quantify this behavior we have performed a linear fit. The graphical and analytical expressions of the straight lines approximating the covariance decay in the semi-logarithmic scale are reported in Figs. 8 and 9. Note that the largest discrepancy with respect to the linear behavior is found at very small distances. Nonetheless, we stress that the linear approximation is very precise, as shown by the value of the coefficient of determination R^2 (which expresses the percentage of empirical data variance that is described by the linear approximation) which is greater than 0.98 for all curves.

A linear behavior in the semilogarithmic domain corresponds to an exponential decay: $\log(\boldsymbol{c}^d_{\mu\nu}) = \alpha_{\mu\nu} + \beta_{\mu\nu} d, \beta_{\mu\nu} < 0$, is equivalent to $\boldsymbol{c}^d_{\mu\nu} = e^{\alpha_{\mu\nu}} e^{\beta_{\mu\nu} d}$. Since $\boldsymbol{c}^0_{\mu\nu} = e^{\alpha_{\mu\nu}}$, we can write

$$\boldsymbol{c}^d_{\mu\nu} = \boldsymbol{c}^0_{\mu\nu} e^{\beta_{\mu\nu} d}, \ \mu, \nu \in \{R, G, B\}, \tag{30}$$

where $\boldsymbol{c}^0_{\mu\nu}$ is the generic element of the chromatic autocovariance matrix.

From the point of view of differential equations, this means that the spatiochromatic covariance $\boldsymbol{c}_{\mu\nu}(d)$, interpreted as a function of the pixel distance d, satisfies the following initial value problem:

$$\begin{cases} \dot{\boldsymbol{c}}_{\mu\nu}(d) = \beta_{\mu\nu} \boldsymbol{c}_{\mu\nu}(d) \\ \boldsymbol{c}_{\mu\nu}(0) = \boldsymbol{c}^0_{\mu\nu} \end{cases} \tag{31}$$

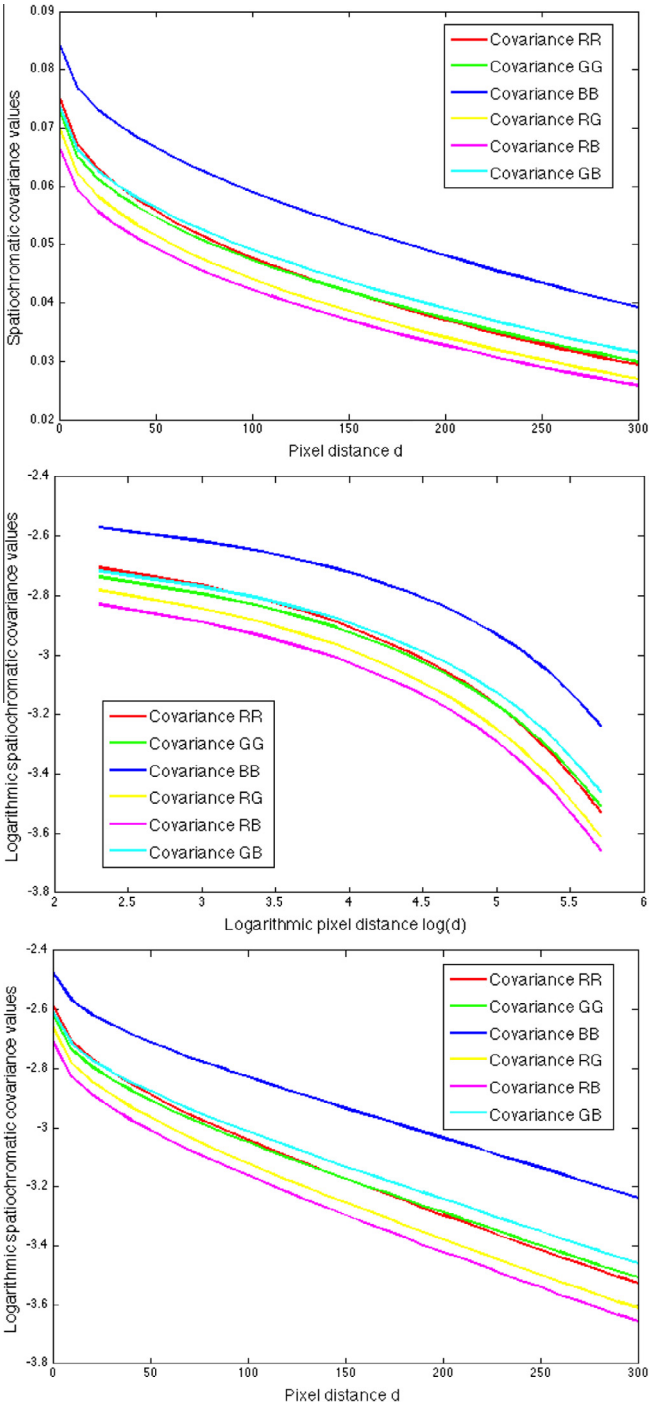

Fig. 6. Graphs of the six distinct spatiochromatic covariance matrix elements in the linear, (natural) bi-logarithmic and semi-logarithmic scale, respectively, as a function of the pixel distance d. The values were obtained using the Flickr database.

i.e. the speed of decay of $\boldsymbol{c}_{\mu\nu}(d)$ is directly proportional to $\boldsymbol{c}_{\mu\nu}(d)$ via the coefficient $\beta_{\mu\nu}$. The value of these coefficients are listened in Table 1.

It can be seen that the spatiochromatic covariance relative to the blue channel decreases less rapidly than that of the red and green channels. This may be explained by the fact that pictures in which the sky is present are characterized by large homogeneous areas dominated by the blue channel.

The explicit analytical expressions of $\boldsymbol{c}^d$ that we have managed to obtain are interesting for two reasons: firstly, they provide an

Please cite this article in press as: Provenzi, E., et al. On the second order spatiochromatic structure of natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.02.025

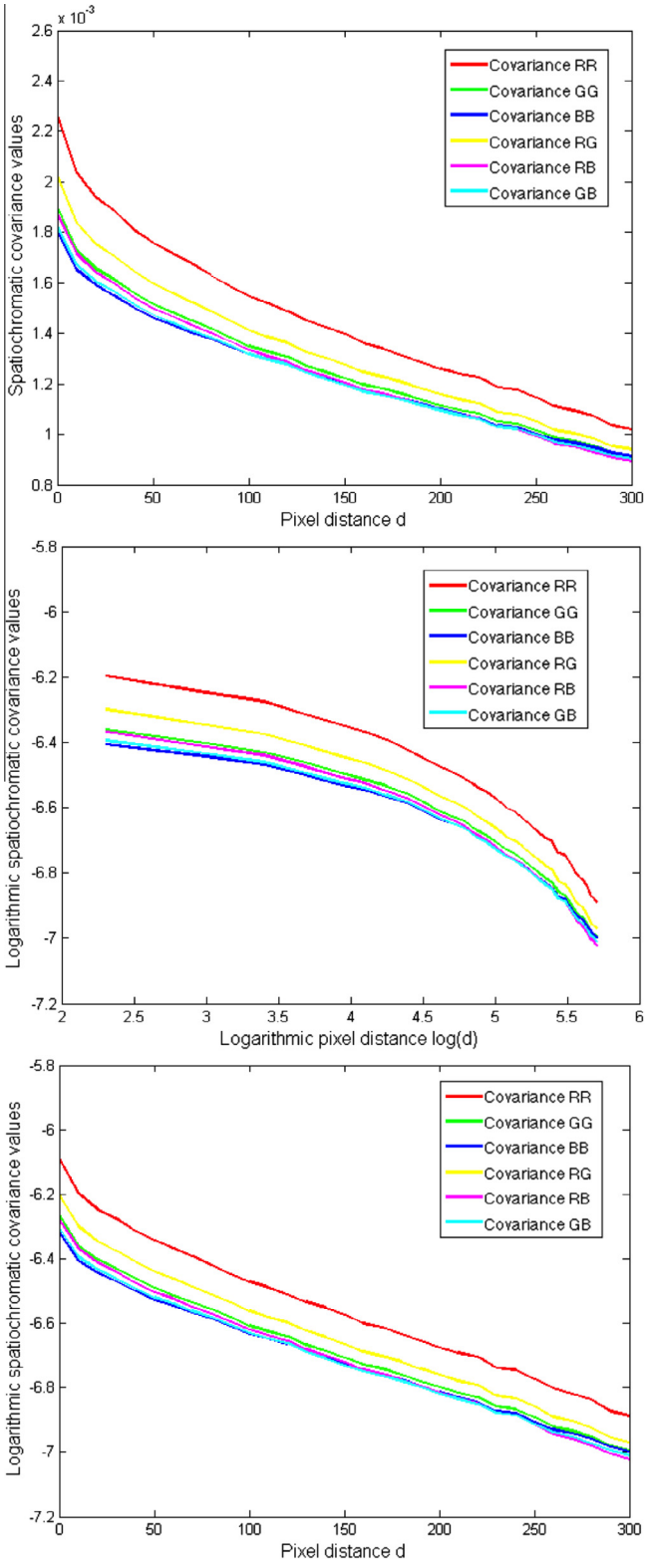

Fig. 7. Graphs of the six distinct spatiochromatic covariance matrix elements in the linear, (natural) bi-logarithmic and semi-logarithmic scale, respectively, as a function of the pixel distance d. The values were obtained using the RAW database.

accurate model for the covariance that corrects the power-law decay; secondly, they allow computing the commutators $[\boldsymbol{c}^0, \boldsymbol{c}^d]$ for every distance $d > 0$. If the coefficients $\beta_{\mu\nu}$ were all equal, then these commutators would be null matrices, however, the differences in the coefficients of the exponentials make the matrix elements of the commutators different than zero. Fig. 10 shows the absolute and normalized 2-norms of the commutators as a function of d. It can be seen that, for small values of d, the commutators can be considered approximately null, however, as d increases, they show deviations from the zero matrix, but they are still small.

4.6. Transformations of the RAW images and their effects on the statistics

In order to test the effect of the photographic development pipeline, and of other non-linear transformations, on the observed statistics, we have transformed the RAW images of the databases with the logarithmic and an average gamma function of 0.5. We have found no noticeable difference in the decay. The only change is in the coefficient β_{BB} which becomes -0.0019 for the Nikon RAW database and -0.0018 for the Canon RAW database, but, apart from this slight change, all the other coefficients remain the same.

We have also performed the transformation from RGB to LMS by using the chromatic adaptation matrix M_{CAT02} from the CIECAM02 model, see e.g. Fairchild (2005), to have an indication of what we would get if we repeated our experiments on cone activation data. Qualitatively, the behavior of spatiochromatic covariances remains the same, the only small quantitative change being represented by the fact that the exponential decay is slightly faster, as can be seen by the coefficients reported in Table 2.

A further and, in our opinion, more significative test that we have performed on RAW images is the following: we have transformed the initial data with the so-called Michaelis–Menten (also known as Naka-Rushton) equation, see Shapley and Enroth-Cugell (1984, Chap. 9), i.e. $I_\mu(x) \mapsto I_\mu^\gamma(x)/(I_\mu^\gamma(x) + m_\mu^\gamma)$, where m_μ is the average intensity value in the chromatic channel μ and $\gamma = 0.74$ (the value corresponding to the rhesus monkey retinal photoreceptors). This equation models the photochemical transduction from radiance to action potential performed by retinal photoreceptors and plays a major role in the adaptation mechanisms of human vision.

We then repeated the spatiochromatic covariance computations on this new transformed database and we have found again an exponential decay, but this time with exactly the same coefficients for all the chromatic combinations, i.e. $\beta_{\mu\nu} = -0.0033$, for all $\mu, \nu \in \{R, G, B\}$. This can be explained by the fact that the photochemical transduction process normalizes the dynamic range of the data to the interval $[0, 1]$ and sets the average value of each chromatic channel to $1/2$.

The by-product of photoreceptors' photochemical transduction is the rearranging of radiance values so that all the spatiochromatic covariance matrices commute perfectly. Up to our knowledge, this test has never been performed before.

We notice that the presence of the three average radiance values in the photoreceptors' photochemical transduction formula and the fact that they are normalized to the same value $1/2$ is crucial for the rearrangement of the slopes $\beta_{\mu\nu}$.

As already remarked, RAW data provide a good approximation of real physical irradiance, thus we believe that this test gives an interesting hint about the consequence of photochemical transduction on covariance of natural visual stimuli. However, a better choice would be to perform this test on multispectral radiance values, in fact the irradiance acquired in the RAW data is integrated with respect to the RGB camera sensitivity functions, which are very different than the cone sensitivity curves. This is not yet possible due to the lack of a large database of natural multispectral images. We will turn back to this issue in the discussion section.

Please cite this article in press as: Provenzi, E., et al. On the second order spatiochromatic structure of natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.02.025

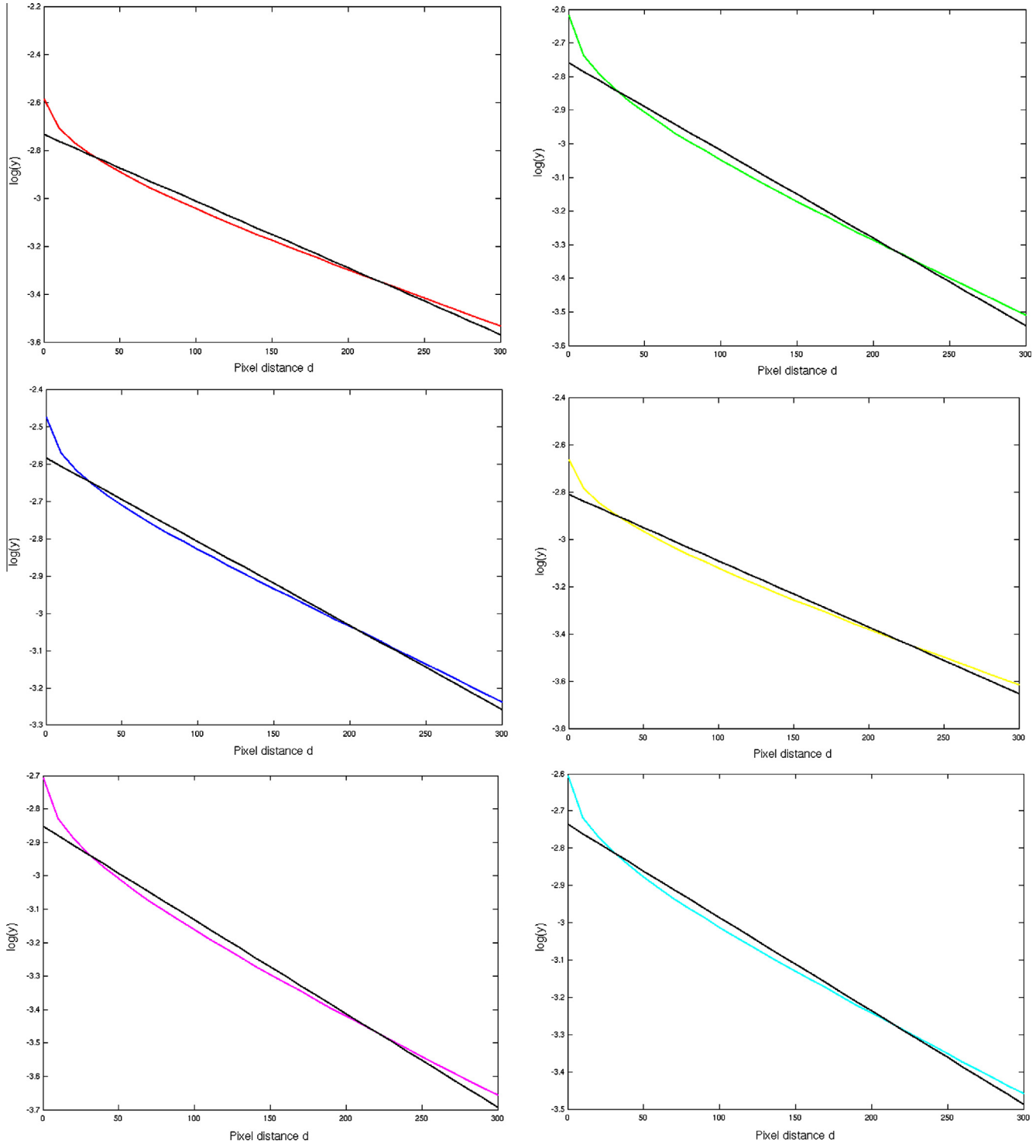

Fig. 8. Graphics of the linear approximations of the spatiochromatic covariance matrix elements in the semi-logarithmic scale (data obtained from the Flickr database). The coefficient of determination R^2 is greater than 0.98 for all the linear approximations.

4.7. The power spectrum

We are now going to analyze the consequence of an exponential model for the covariance decay on the power spectrum. As recalled in Section 2.1, we can theoretically relate the autocovariance decay with the power spectrum $S(f)$ via the Fourier transform. Using the notation of Section 2.1.2 and applying the Fourier transform to the analytical expression of the spatiochromatic autocovariance given by Eq. (30) we get that

$$S_\mu(f) \propto \frac{1}{1+\left(\frac{f}{\beta_{\mu\mu}}\right)^2}, \quad \mu \in \{R, G, B\}. \tag{32}$$

Please cite this article in press as: Provenzi, E., et al. On the second order spatiochromatic structure of natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.02.025

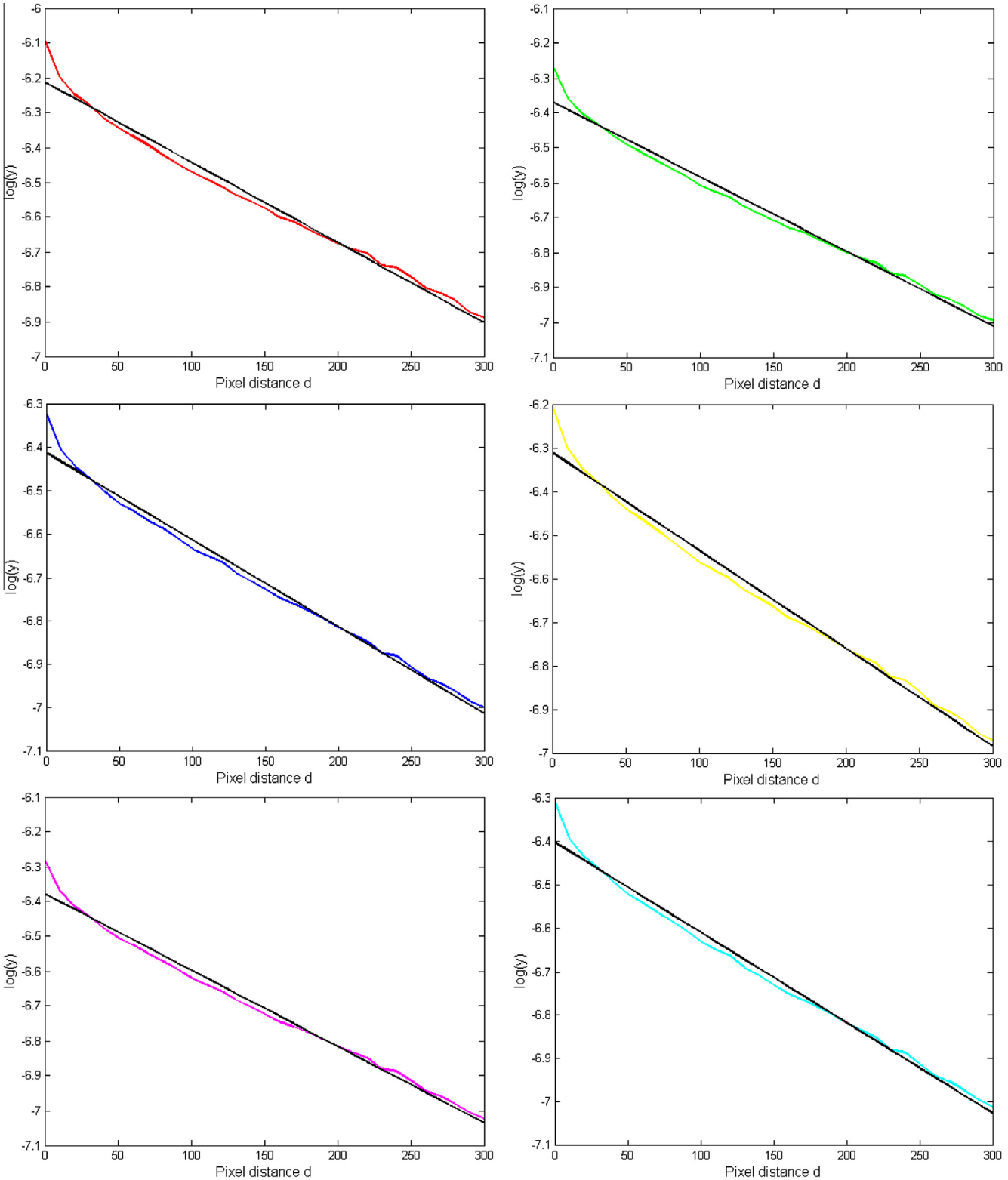

Fig. 9. Graphics of the linear approximations of the spatiochromatic covariance matrix elements in the semi-logarithmic scale (data obtained from the RAW database). The coefficient of determination R^2 is greater than 0.98 for all the linear approximations.

If $\left(f/\beta_{\mu\mu}\right)^2 \gg 1$, i.e. if $f \gg \beta_{\mu\mu}$, then we can neglect the term 1 at the denominator of $S_\mu(f)$ and approximate it with a power law with respect to f. Since the order of magnitude of $\beta_{\mu\mu}$ is 10^{-3}, this means that $S_\mu(f)$ can be approximated by a power law for values of $f \gg 10^{-3}$.

The plots of $S_\mu(f)$ in the bi-logarithmic scale reported in Fig. 11 confirm this fact: for small values of f (which correspond to large values of d for the covariance) the power spectrum deviates from the linear fit, as f increases (corresponding to smaller values of d for the covariance), the linear fit becomes more and more precise.

Please cite this article in press as: Provenzi, E., et al. On the second order spatiochromatic structure of natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.02.025

Table 1
Slopes of the straight lines which approximate the spatiochromatic covariance graphs in the semilogarithmic scale for the Flickr and the RAW databases.

Flickr Database	RAW Database (Nikon)	RAW Database (Canon)
$\beta_{RR} = -0.0028$	$\beta_{RR} = -0.0023$	$\beta_{RR} = -0.0022$
$\beta_{GG} = -0.0026$	$\beta_{GG} = -0.0021$	$\beta_{RR} = -0.0020$
$\beta_{BB} = -0.0022$	$\beta_{BB} = -0.0020$	$\beta_{RR} = -0.0019$
$\beta_{RG} = \beta_{GR} = -0.0028$	$\beta_{RG} = \beta_{GR} = -0.0022$	$\beta_{RR} = -0.0021$
$\beta_{RB} = \beta_{BR} = -0.0028$	$\beta_{RB} = \beta_{BR} = -0.0022$	$\beta_{RR} = -0.0021$
$\beta_{GB} = \beta_{BG} = -0.0025$	$\beta_{GB} = \beta_{BG} = -0.0021$	$\beta_{RR} = -0.0020$

This is coherent with the graph of the covariance in the bi-logarithmic scale reported in Fig. 6: for small values of d the graph is linear, but as d increases the curve deviates from linearity.

This explains why the covariance can be thought to have a power-law decay when it is analyzed only by means of the power spectrum.

5. Discussion and perspectives

We have provided a theoretical analysis of the relationship between translation invariance of the covariance and the decorrelated spatiochromatic features of digital RGB images, supported by several numerical tests.

Table 2
Slopes of the straight lines which approximate the spatiochromatic covariance graphs in the semilogarithmic scale for the RAW databases *in the LMS coordinate system*. These slopes are slightly higher (in absolute value) with respect to those reported in the second and third column of Table 1.

(LMS) RAW Database (Nikon)	(LMS) RAW Database (Canon)
$\beta_{RR} = -0.0026$	$\beta_{RR} = -0.0025$
$\beta_{GG} = -0.0025$	$\beta_{GG} = -0.0023$
$\beta_{BB} = -0.0022$	$\beta_{BB} = -0.0021$
$\beta_{RG} = \beta_{GR} = -0.0026$	$\beta_{RG} = \beta_{GR} = -0.0024$
$\beta_{RB} = \beta_{BR} = -0.0024$	$\beta_{RB} = \beta_{BR} = -0.0024$
$\beta_{GB} = \beta_{BG} = -0.0022$	$\beta_{GB} = \beta_{BG} = -0.0022$

Our analysis has been motivated by the will to understand the basic mathematical reasons underlying the appearance of a separable spatiochromatic basis of uncorrelated features when the PCA is performed over patches or whole natural images.

In order to investigate this property, we have built the spatiochromatic covariance matrix of an abstract three-chromatic image and we have shown that, under the assumption of spatial invariance and commutativity, their eigenvectors can be written as the Kronecker product of the cosine Fourier basis times an achromatic plus color opponent triad.

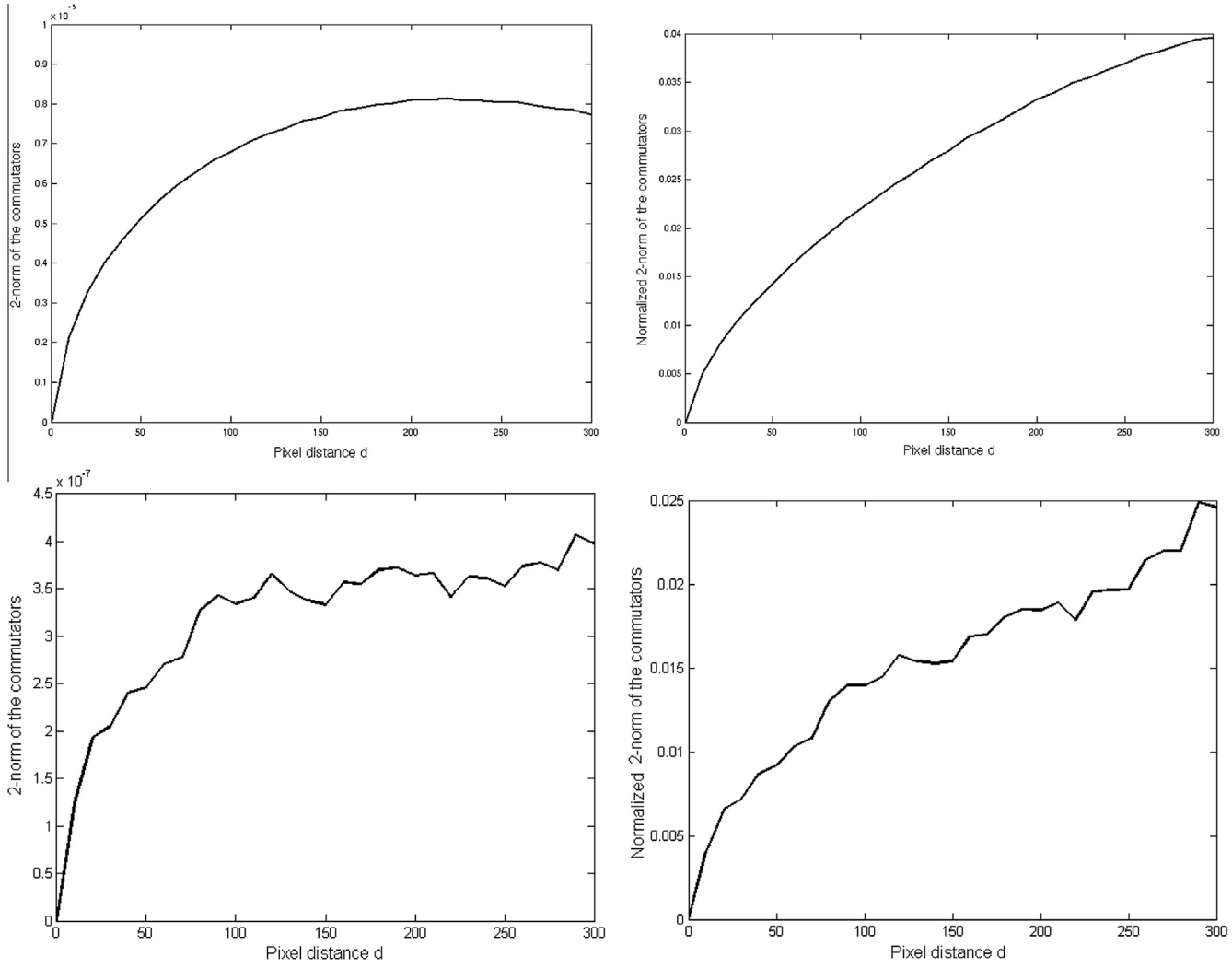

Fig. 10. *Left*: Graph of the 2-norm of the commutators between the spatiochromatic covariance matrices as a function of pixel distance d. *Right*: Graph of the normalized 2-norm of the commutators, the normalization is done over the mean value of 2-norm of the product matrix performed from left to right and from right to left. *First row*: data obtained with the Flickr database. *Second row*: data obtained with the RAW database.

Please cite this article in press as: Provenzi, E., et al. On the second order spatiochromatic structure of natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.02.025

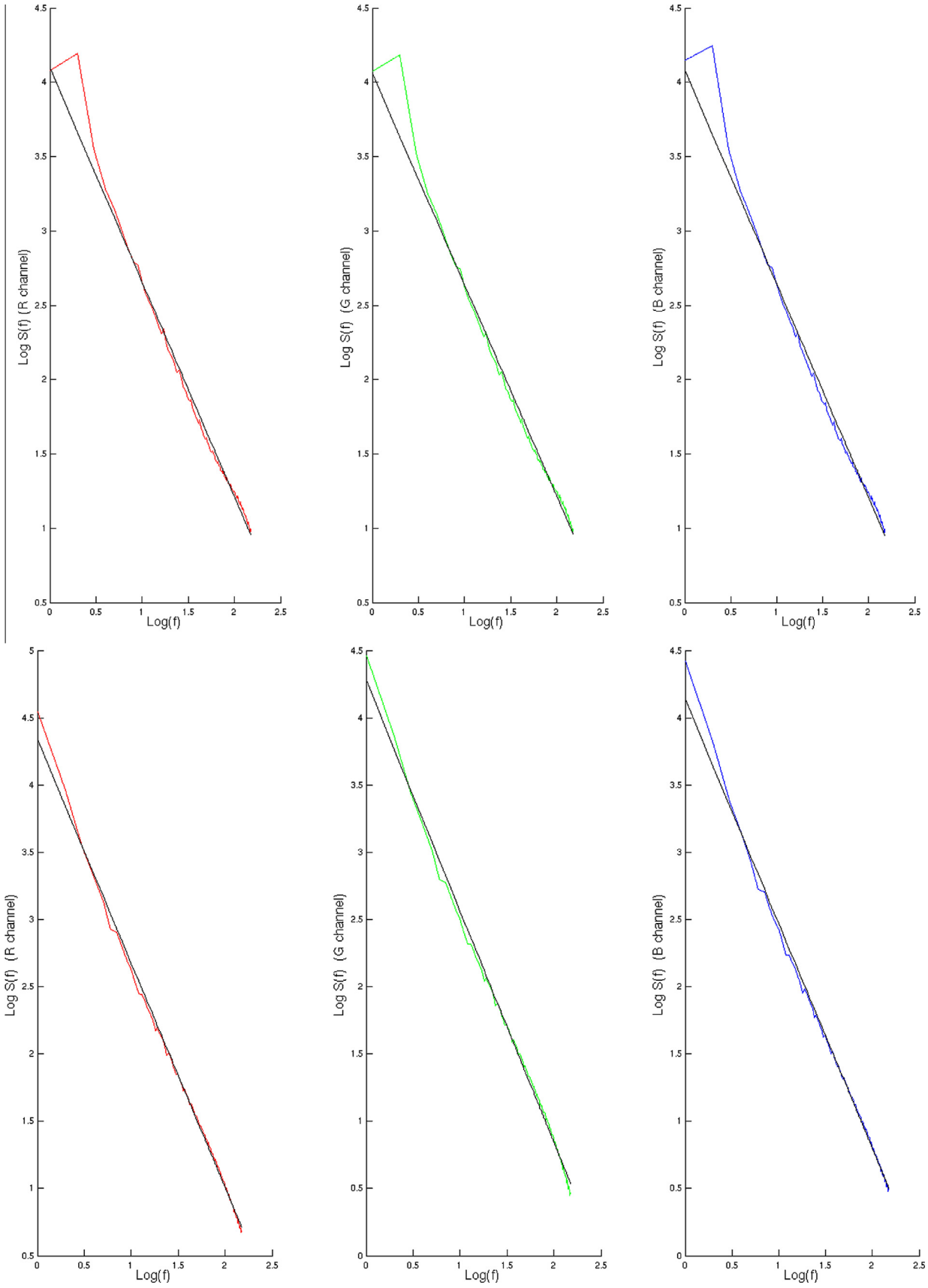

Fig. 11. Average power spectra. *First row*: Flickr database, the slopes of the linear approximations are −1.4429, −1.4229, −1.4384, from left to right. *Second row*: Raw database, the slopes of the linear approximations are −1.6666, −1.7182, −1.6698, from left to right. Notice that the initial bump showed by the power spectra of the images taken by the Flickr database is shared with the graph reported by Pouli, Cunningham, and Reinhard (2010, p. 68), Fig. 10.

Please cite this article in press as: Provenzi, E., et al. On the second order spatiochromatic structure of natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.02.025

The numerical tests that we have conducted have shown that the assumptions are verified with a good degree of approximation, both when we consider the pictures of a large database of a million natural RGB images taken from the popular website Flickr and also those of a database of RAW images that we have taken. In order to explicitly link the correlation radius to the geometrical properties of the images in the database, one would need a database with known parameters: camera sensitivity functions, focal length, distance to the objects in the scene, and so on. The constitution of such a database constitutes an interesting research perspective.

In particular, the analysis of the commutativity of spatiochromatic covariance matrices have led to a lateral result that it is worth underlying: our tests have shown that the spatial covariance decays exponentially and not following a power law. As recalled in the introduction, the failure of the power law decay has already been reported in the literature of natural image statistics, but our result on the exponential decay is novel. Moreover, we have shown that the decay speed is not the same for all the combinations of chromatic channels: the autocovariance decay of the blue channel being the slowest and the R–B covariance decay being the fastest.

The slower decay of the blue autocovariance can be explained by the fact that many pictures of natural environments contain large homogeneous areas of blue. The faster decay of the red-blue covariance instead can be explained by the fact that the sensitivity curves of red and blue have the smallest overlapping.

With the two databases of RAW images that we have built by using two different cameras, we have performed experiments that we deem interesting for research about human vision. Firstly, we have confirmed that the covariance decay of RGB RAW images is qualitatively coherent with that of the Flickr database.

Secondly, since RAW images give a representation of the physical radiance (integrated with respect to the camera sensitivity functions) we operated a transformation of these data to LMS values and repeated the tests about covariance decay. We have found that the transformation does not change the qualitative behavior of previous results, the only qualitative change being a slight increment in the decay speed.

Another experiment that, up to authors' knowledge, provides new information that can be useful for human vision studies is the transformation of the RAW images according to the photochemical transduction formula. Differently from the RGB to LMS conversion, we have found that the photochemical transduction transformation changes significantly the results of our experiments. In fact, the negative coefficients in the exponentials which represent the spatiochromatic covariance decay turned out to be the same for every combination of chromatic channels for the transformed data. This can be explained by the normalization of the dynamic range of the data to the interval $[0, 1]$ and of their average value to $1/2$.

From a mathematical point of view, we stress that this property implies the perfect commutativity of all spatiochromatic covariance matrices. This means that, if we take into account also the first adaptation stage provided the photochemical transduction in the analysis of our visual data, then the mathematical results about tensor product separability of spatiochromatic features are satisfied with even better numerical accuracy.

We stress that, at this stage of the research, the numerical results that we have obtained cannot be considered wholly exhaustive to infer properties of the HVS. To do that, a large unbiased database of multispectral images should be carefully built in such a way that the camera sensitivity functions have no influence on the data of the database.

The visual content of multispectral images can be integrated with the retinal sensitivity functions to generate the LMS cone activation values that will become the new input for our mathematical framework.

Large multispectral databases of natural environments are not yet available because of the difficulty of taking a multispectral image in non-controlled conditions without producing artefacts, a problem similar to the well known 'ghosting effect' in high dynamic range imaging. However, recent advances in camera sensors and post-processing algorithms may allow in the near future the creation of such databases and the extension of the results of this paper.

References

Atick, J. (1992). Could information theory provide an ecological theory of sensory processing? *Network Computation in Neural Systems, 3*, 213–251.
Atick, J., Li, Z., & Redlich, A. (1992). Understanding retinal coding from first principles. *Neural Computation, 4*, 559–572.
Atick, J., & Redlich, A. (1990). Towards a theory of early visual processing. *Neural Computation, 2*, 308–320.
Attneave, F. (1954). Some informational aspects of visual perception. *Physchological Review, 61*, 183–193.
Barlow, H. (1961). Possible principles underlying the transformations of sensory messages. *Sensory Communications*, 217–234.
Berman, A., & Plemmons, R. (1987). *Nonnegative matrices in the mathematical sciences*. SIAM.
Buchsbaum, G., & Gottschalk, A. (1983). Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of Royal Society of London B, 220*, 89–113.
Chakrabarti, A., & Zickler, T. (2011). Statistics of real-world hyperspectral images. *Proceeding of the CVPR*, 193–200.
Derrico, J., & Buchsbaum, G. (1991). A computational model of spatio-chromatic image coding in early vision. *Journal of Visual Communication and Image Representation, 2*, 31–38.
Elad, M., & Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing, 15*, 3736–3745.
Fairchild, M. (2005). *Color appearance models*. Wiley.
Field, D. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America, 4*, 2379–2394.
Foster, D., Amano, K., Nascimiento, S., & Foster, M. (2006). Frequency of metamerism in natural scenes. *Journal of the Optical Society of America, 23*, 2359–2372.
Frazier, M. (2001). *An introduction to wavelets through linear algebra*. Springer.
Goldstein, B. (2013). *Sensation and perception* (9th Edition). Cengage Learning.
Gray, R. (2006). *Toeplitz and circulant matrices: A review*. Now Publishers Inc..
Hays, J. & Efros, A. (2007). Scene completion using millions of photographs, in: ACM Transactions on Graphics (SIGGRAPH).
Huang, J. (2000). *Statistics of natural images and models*. Ph.D. thesis. Brown University <http://www.dam.brown.edu/ptg/MDbook/Huangthesis.pdf>.
Huang, J., & Mumford, D. (1999). Statistics of natural images and models. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.
Hyvärinen, A., Hurri, J., & Hoyer, P. (2009). *Natural image statistics: A probabilistic approach to early computational vision*. Springer.
Jiang, J., Liu, D., Gu, J., Süsstrunk, S. (2013). What is the space of spectral sensitivity functions for digital color cameras? In *WACV, IEEE workshop on the applications of computer vision* (pp. 168–179).
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25*, 1097–1105.
Langer, M. (2000). Large-scale failures of $1/f^{\alpha}$ in natural image spectra. *Journal of the Optical Society of America A, 17*, 17–28.
Li, Z. (2014). *Understanding vision: Theory, models, and data*. Oxford University Press.
MacKay, D. (1956). Towards an information-flow model of human behaviour. *British Journal of Psychology, 47*, 30–43.
Mumford, D., & Gidas, B. (2001). Stochastic models for generic images. *Quarterly of Applied Mathematics, 59*, 85–112.
Ohta, Y., Kanade, T., & Sakai, T. (1980). Color information for region segmentation. *Computer graphics and image processing, 13*, 222–241.
Olmos, A., & Kingdom, F. (2004). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception, 33*, 14631473.
Olshausen, B., & Field, D. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Letters to Nature, 381*, 607–609.
Olshausen, B., & Field, D. (1997). Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research, 37*, 607–609.
Papoulis, A. (1991). *Probability, random variables, and stochastic processes*. McGraw Hill.
Párraga, C., Troscianko, T., & Tolhurst, D. (2002). Spatiochromatic properties of natural images and human vision. *Current Biology, 6*, 483–487.
Pouli, T., Cunningham, D., & Reinhard, E. (2010). Image statistics and their applications in computer graphics. *Eurographics State of the Art Report*, 83–112.
Ruderman, D. (1996). Origin of scaling in natural images. *Vision Research, 37*, 3385–3398.
Ruderman, D., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical Review Letters, 73*, 814–817.

Please cite this article in press as: Provenzi, E., et al. On the second order spatiochromatic structure of natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.02.025

Ruderman, D., Cronin, T., & Chiao, C. (1998). Statistics of cone responses to natural images: Implications for visual coding. *Journal of the Optical Society of America A, 15*, 2036–2045.

Shapley, R., & Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. *Progress in Retinal Research, 3*, 263–346.

Simoncelli, E., & Olshausen, B. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24*, 1193–1216.

Srinivasan, M., Laughlin, S., & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proceedings of Royal Society of London B, 216*, 427–459.

Srivastava, A., Lee, A. B., Simoncelli, E. P., & Zhu, S. C. (2003). On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision, 18*, 17–33.

Stockman, A., MacLeod, D., & Johnson, N. (1993). Spectral sensitivities of the human cones. *Journal of the Optical Society of America, 10*, 2491–2521.

Tkacik, G., Garrigan, P., Ratliff, C., Milcinski, G., Klein, J., Seyfarth, L., et al. (2011). Natural images from the birthplace of the human eye. *PLoS One, 6*, e20409.

Tolhurst, D., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic and Physiological Optics, 12*, 229–232.

Vazquez-Corral, J., Párraga, C., Vanrell, M., & Baldrich, R. (2009). Psychophysical evaluation on a new dataset. *Journal of Imaging Science and Technology, 53*, 031105–031109.

von Kries, J. (1902). Chromatic adaptation. *Festschrift der Albrecht-Ludwigs-Universität*, 145–158.

Wachtler, T., Lee, T. W., & Sejnowski, T. (2001). Chromatic structure of natural scenes. *Journal of Optical Society of America A, 18*, 65–77.

Please cite this article in press as: Provenzi, E., et al. On the second order spatiochromatic structure of natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.02.025