



Power spectra of the natural input to the visual system

D. Pamplona^a, J. Triesch^a, C.A. Rothkopf^{a,b,*}

^a Frankfurt Institute for Advanced Studies, Goethe University, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany

^b Institute of Cognitive Science, University Osnabrück, Albrechtstr. 28, 49076 Osnabrück, Germany

ARTICLE INFO

Article history:

Received 29 May 2012

Received in revised form 5 December 2012

Available online 1 March 2013

Keywords:

Natural image statistics

Natural vision

Power spectrum

Orientation

ABSTRACT

The efficient coding hypothesis posits that sensory systems are adapted to the regularities of their signal input so as to reduce redundancy in the resulting representations. It is therefore important to characterize the regularities of natural signals to gain insight into the processing of natural stimuli. While measurements of statistical regularity in vision have focused on photographic images of natural environments it has been much less investigated, how the specific imaging process embodied by the organism's eye induces statistical dependencies on the natural input to the visual system. This has allowed using the convenient assumption that natural image data are homogeneous across the visual field. Here we give up on this assumption and show how the imaging process in a human model eye influences the local statistics of the natural input to the visual system across the entire visual field. Artificial scenes with three-dimensional edge elements were generated and the influence of the imaging projection onto the back of a spherical model eye were quantified. These distributions show a strong radial influence of the imaging process on the resulting edge statistics with increasing eccentricity from the model fovea. This influence is further quantified through computation of the second order intensity statistics as a function of eccentricity from the center of projection using samples from the dead leaves image model. Using data from a naturalistic virtual environment, which allows generation of correctly projected images onto the model eye across the entire field of view, we quantified the second order dependencies as function of the position in the visual field using a new generalized parameterization of the power spectra. Finally, we compared this analysis with a commonly used natural image database, the van Hateren database, and show good agreement within the small field of view available in these photographic images. We conclude by providing a detailed quantitative analysis of the second order statistical dependencies of the natural input to the visual system across the visual field and demonstrating the importance of considering the influence of the sensory system on the statistical regularities of the input to the visual system.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Considerable evidence has been accumulated showing that biological sensory systems reflect the ecological circumstances to which they are exposed (see e.g. Geisler, 2008; Simoncelli & Olshausen, 2001 for reviews). Based on the ideas of Shannon (1948) it was argued by Attneave (1954) and Barlow (1961) that sensory systems of animals compute signal representations that have reduced the redundancy compared to their sensory input. Therefore it is essential to analyze this redundancy by quantifying the statistics of the natural stimuli that organisms encounter in their natural environment. It has been investigated much less, how the observer itself influences these statistics, i.e. there are few investigations

distinguishing between the environmental statistics and the statistics of the input to the sensory system (Burge & Geisler, 2011; Reinagel & Zador, 1999; Rothkopf, Weisswange, & Triesch, 2009).

In vision this redundancy is present in the set of natural images that animals are exposed to, which is only a small subset of all possible images (Daugman, 1989; Field, 1987; Kersten, 1987). One way of quantifying this redundancy is by noting that neighboring image pixels' intensities are not independent but instead their intensities are closely related. A direct perceptual measurement of this redundancy was performed by Kersten (1987). Human subjects were asked to estimate the intensity of individual missing pixels in natural images of size 128×128 pixels with 16 gray levels corresponding to 4 bits. When 1% of the image's pixels were missing, subjects were able to guess the missing gray values correctly with their first guess on 78% of trials, reflecting the spatial regularities in neighboring pixels within the image. Based on these results it was possible to estimate the uncertainty of individual pixels' gray values by computing the upper bound on the entropy

* Corresponding author at: Frankfurt Institute for Advanced Studies, Goethe University, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany.

E-mail addresses: pamplona@fias.uni-frankfurt.de (D. Pamplona), triesch@fias.uni-frankfurt.de (J. Triesch), rothkopf@fias.uni-frankfurt.de (C.A. Rothkopf).

of the gray level distribution, which was found to be 1.4 bits per pixel.

Redundancies can also be quantified with respect to specific statistical models. As no complete generative statistical model for natural images has been developed, researchers have quantified statistical dependencies with different measures. One of the fundamental measures is the intensity autocorrelation function, which quantifies the expected value of the product of pixel intensities as a function of their spatial separation, and it has been used extensively to measure these second order statistical dependencies. A closely related measurement is the power spectrum, which can be expressed as the magnitude of the Fourier transform of the autocorrelation function according to the Wiener–Khinchin theorem (e.g. Hecht, 2001). A fundamental assumption for the applicability of this theorem is that of shift invariance, i.e. that these correlations only depend on the relative separation between image pixels and not on their absolute position. Furthermore, the two-dimensional power spectrum has usually been reduced to a one-dimensional function of spatial frequency by averaging rotationally within the two-dimensional frequency plane under the assumption that the autocorrelation function of natural images is homogeneous in all directions.

The second order statistics of natural images have been investigated empirically by measuring intensity correlations between neighboring image pixels and were first reported by Deriugin (1956) in the context of television signals. Extensive further analysis of the autocorrelation function and the power spectrum of natural images has shown that for large image ensembles the average power spectrum falls off with radial frequency as $1/f_r^\alpha$, where $f_r = \sqrt{f_x^2 + f_y^2}$ and the value for α is empirically estimated to be close to 2 (e.g. Burton & Moorhead, 1987; Carlson, 1978; Field, 1987; Ruderman & Bialek, 1994; van der Schaaf & van Hateren, 1996).

However, when considering the power spectrum of individual images, their second order statistics can depart considerably from this pattern (Langer, 2000; Tolhurst, Tadmor, & Chao, 1992). Further work has shown, that the power spectrum varies significantly not only across orientations (Baddeley, 1997) but also with the considered habitat (Balboa & Grzywacz, 2003) suggesting significant differences for the redundancy reduction processes in animal species inhabiting different environments. More recent research (Oliva & Torralba, 2001) has demonstrated that the power spectrum can vary considerably across images from different scene types, a fact that these authors utilized to distinguish between classes of environments from summary statistics that correspond to the notion of a scene's gist.

The aforementioned studies all investigated the statistical redundancies present in photographic images of natural scenes and not necessarily the statistical redundancies that are present in the natural input to the visual system, because they do not take the imaging process into account. But surely the large variety of vision systems (Land & Nilsson, 2002) influence the statistical regularities present in the input to a visual system. Here we move closer towards a more realistic description of the natural input to the visual system, which is specific to vision of organisms (Ballard, 1991), by explicitly modeling the imaging process and the detector geometry.

Furthermore, previous studies investigated the global second order statistics across the visual field and not the local second order statistical regularities. But these statistics may vary systematically across the visual field and we therefore employ a local analysis thus abandoning the assumption of shift invariance in the natural image input to the visual system. We also abandon the assumption that the power spectrum of the natural input to the visual system is circularly symmetric, thus power is estimated as a function of the vertical and horizontal frequencies and not the

radial frequency. By modeling the imaging process together with the geometry of the image detector the local power spectrum is shown to vary significantly and systematically across the visual field. These statistical dependencies are quantified with a generalized expression for the power spectrum as a function of angular position with respect to the fovea. This function can be related directly to the well known form of $1/f_r^\alpha$, thereby showing the similarities and differences to previous work on quantifying the second order statistics of natural images.

2. Material and methods

2.1. Modeling the image acquisition process

When considering how visual signals' statistics are influenced by an imaging process, we need to select a specific model for this imaging process. Here, we define the projective transform based on the thin lens model projecting light onto a spherical surface, corresponding to the reduced eye model proposed by Emsley (1948). First, the thin lens model is a good first approximation of the imaging process under the assumption that the thickness of the lens is negligible when compared to the focal length (e.g. Hecht, 2001). This allows formulating an equivalent thin lens model for the human eye that captures many of its refractive properties (e.g. Coletta, 2010; Emsley, 1948). Consequently light does not suffer any kind of aberration when projected into the plane. Secondly, although aberrations have been measured in the human eye, there are considerable variations across individuals, which would make their inclusion specific to individual observers. A solution that takes into account the general blurring due to the imaging process is to apply the position dependent modulation transfer function (MTF) that has been measured for human observers (Navarro, Artal, & Williams, 1993) to the geometric projections. Third, moving to this basic imaging model in which the geometric distortions are combined with the position dependent blurring in the eye will already reveal a rich set of properties of the statistics of the natural input to the visual system. Finally, the employed thin lens model is general enough to accommodate future work on accommodation, different object distances, and differently shaped retinas.

We give an overview of the involved projective transforms and refer the reader to Appendix A, which spells out the mathematical details. The used projection model, which is illustrated in Fig. 1a starts from a three dimensional point P in space. Usually, when considering the statistics of natural images, the analysis uses planar image projection points p . We call this image *planar*, due to the fact that its coordinates are defined over a plane. Fig. 1b is an example of this image, where for display purposes the image was rotated right-side-up. However, this model assumes that points of the 3d world are projected into a plane, which is true for usual cameras, and therefore for the images in databases commonly used in vision science, but not necessarily for animal eyes. Human photoreceptors lay on an approximately spherical surface, the eye ball. Therefore, in order to model more realistically the human imaging process, we define the thin lens projection onto a spherical surface with parameters matching those of the human visual system.

To model this particular projection we introduce a generalization of the thin lens model presented above in such a way that points of the 3d world are projected into the back of a sphere according to Fig. 1c. The resulting image will be called *spherical*, due to the fact that its coordinates are defined over a spherical surface. Fig. 1d shows the corresponding projection of the visual scene into the spherical surface, which again was rotated right-side-up to facilitate interpretation. Because of the geometry of the projections, the angle between the rays and the plane of projection is the same in the planar and spherical case. Therefore these two maps can be related directly by a transform \mathcal{R} , allowing us to

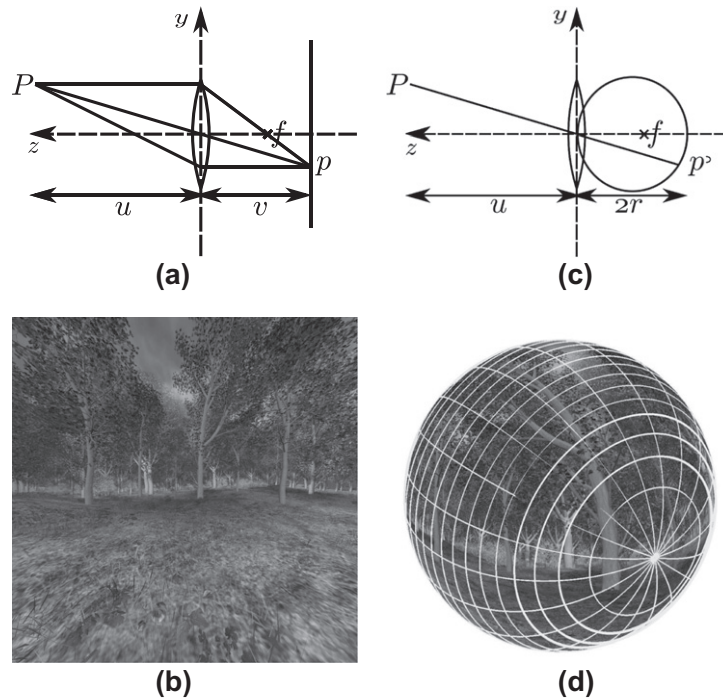


Fig. 1. Comparison of imaging process with planar and spherical projections. (a) Geometry of planar image projection according to the thin lens model. (b) Resulting planar projection of naturalistic VR scene. Note that the image was rotated right-side-up for visualization purposes. (c) Imaging projections into a spherical surface. (d) Resulting spherical projection of naturalistic VR scene. Note that the image was rotated right-side-up for visualization purposes.

obtain the spherical coordinates from the planar ones, and consequently the spherical image.¹

2.2. Analysis of spherical images

The above transformation projects points within the visual scene onto the spherical model eye. One could analyze the autocorrelation of the intensity data through a projection onto spherical harmonics, but this would lead to results not easily comparable with the literature on image statistics, which usually considers image statistics on planar surfaces and not spherical surfaces. Therefore, we will consider two different projections that allow measuring relevant quantities for further analysis in planar images, i.e. the projections result in planar representations of the image data, which allow computing image statistics as in the known power spectra literature.

For the analysis of edge orientations we decided to project the semi-circles into segments in one plane with the *stereographic transform* (e.g. Carathéodory, 1998). This transform is a well known mapping from spherical surfaces onto planes. It does not change angles between inputs, meaning that if the angle between two semi-circles is θ then the angle between the transformed semicircle is also θ . However, this transform is not equal-area nor equidistant, meaning, the length and distance between two segments is not maintained nor scaled proportionally at all eccentricities. Fig. 1 shows the relevant quantities. Points P in the scene are projected into the spherical model eye onto points p' , where for convenience the origin of the coordinate system was positioned at the center of the thin lens. With the stereographic projection the point P is then projected onto the point p'' on a plane at the position $(0, 0, -2\tilde{r})$, where \tilde{r} is the radius of the model eye ball. Note that this projection corresponds to the standard projection when using a thin lens equivalent camera. Appendix A contains the mathematical details of the stereographic transform.

While the stereographic transform can be used for the analysis of edge orientations, as these are maintained by the projection, it introduces other distortions to the input, which preclude it from being used for the analysis of the power spectra. Instead, we define the *generalized stereographic transform* which projects points on the sphere onto a plane that is locally tangential to the sphere. This transform also introduces distortions that increase with the distance from the center of projection, but it is suitable for local analysis in the neighborhood of the center of projection. This means that we consider regions on the spherical surface and project them onto the tangent plane to the sphere at each point. For small distances on the plane to the tangent point considered in the present analysis, these distortions are negligible. All further analyses are based on these data. Again, the mathematical details of this projection and rotation are detailed in Appendix B.

2.3. Position dependent blurring of image data

The optical properties of the eye are only approximated by the ideal thin lens model of image formation, as a wealth of optical aberrations degrade the image quality available at the back of the retina. Empirical studies have quantified the degradation of monochromatic image quality in the human eye across a wide visual field (Navarro, Artal, & Williams, 1993) with natural pupil (4 mm) and accommodation (three diopters). For eccentricities of up to 60° Navarro, Artal, and Williams (1993) report the modulation transfer function (MTF). In the present study we calculated the MTF for each location of the extracted image regions with known position across the visual field and applied the corresponding MTF to the image to obtain the locally degraded image.

2.4. Power spectrum estimation

The mean power spectrum was estimated by averaging the squared amplitude of the discrete Fourier transform of the windowed image samples, where K is the number of samples, M , N the image size, w the radially symmetric Hamming window at

¹ This process is implemented using Matlab and is available from the website of the authors.

the respective position and I_k the image of index k . The mean power spectrum PS was obtained as:

$$PS(f_x, f_y) = \frac{1}{K} \sum_{k=0}^K \left| \sum_{x=0}^M \sum_{y=0}^N w(x, y) I_k(x, y) e^{-2\pi i(xf_x + yf_y)} \right|^2. \quad (1)$$

2.5. Power spectra fitting functions

To compare our results to previous literature on the second order statistics of natural images, we fit power spectra with a rotationally symmetric power law:

$$PS(f_x, f_y) = A \frac{1}{(f_x^2 + f_y^2)^\alpha}. \quad (2)$$

This equation is equivalent to those used by Tolhurst, Tadmor, and Chao (1992) and Ruderman and Bialek (1994), where the power spectrum was fit by a power law over radially averaged frequencies $f_r = \sqrt{f_x^2 + f_y^2}$. In these studies the authors verified that the power spectrum is well approximated by the function $PS(f_r) = A \frac{1}{f_r^\alpha}$, with $\alpha \approx 2$, which is equal in our parameterization with $\alpha \approx 1$, because this allows a more direct interpretation of the parameter a in the new parameterization function below.

However, when visually inspecting the resulting power spectra we confirmed previous observations (e.g. Oliva & Torralba, 2001) that the power spectrum can be described as a superposition of an approximately circular component and components with predominant power along the horizontal and vertical directions. Accordingly, to allow for more general dependencies in the power spectra than those described by Tolhurst, Tadmor, and Chao (1992) and Ruderman and Bialek (1994) it was necessary to find a more general expression. Therefore, we fit the power spectrum with the sum of an oriented elliptical power law and a scaled hyperbola:

$$PS(f_x, f_y) = A \left((1 - B) \frac{1}{(f_{xr}^2 + \frac{f_{yr}^2}{a})^\alpha} + B \frac{1}{|f_x f_y|^\beta} \right), \quad (3)$$

where $(f_{xr}, f_{yr}) = (f_x \cos(\theta) + f_y \sin(\theta), -f_x \sin(\theta) + f_y \cos(\theta))$.

Empirically, this function can capture the orientation bias (θ), the shape of the main ellipse (a) and the strong influence of the very low frequencies on the vertical and horizontal direction (B and β). Furthermore, it can fit the observed power spectra well and can be brought in direct relationship to previous functional expressions. Specifically, the parameterization of spectra used in Tolhurst, Tadmor, and Chao (1992) and Ruderman and Bialek (1994) corresponds to setting $\alpha = 1$, $a = 1$, $B = 0$. We fit the empirically determined power spectra with a maximum likelihood criterion on the parameters under Gaussian noise assumption.

2.6. Sensory input

To gain an initial insight into the effects of the projective transform we first considered the distribution of orientations of edge elements projected onto a sphere. For this, artificial scenes consisting of uniformly sampled edge elements were constructed, passed through the thin lens projection, and histograms of the orientation distributions on the planes tangential to the spherical surface were obtained.

We also sampled realizations from the dead-leaves image model, which has second order image statistics closely resembling those of natural scenes (Balboa, Tyler, & Grzywacz, 2001; Lee, Mumford, & Huang, 2001; Matheron, 1975; Ruderman, 1997). To understand how the projection on the sphere affects the power spectra of images depending on the eccentricity and polar angle, we generate a batch of 10,000 dead leaves images of size 2024×2024 pixels and project them into a sphere with radius 8.5 mm. These images are generated by superposition of disks of random radii and gray values, Fig. 2c is one of these images. One of the important properties of these images for this study is that for a given σ they are scale invariant, thus the average power spectrum of these images is approximately $1/f_r^2$ at any eccentricity.

To move closer to the analysis of the power spectra of the natural input to the model visual system we generated a naturalistic data set with highly controlled and reproducible properties. To this end, we constructed parametric naturalistic virtual environments.

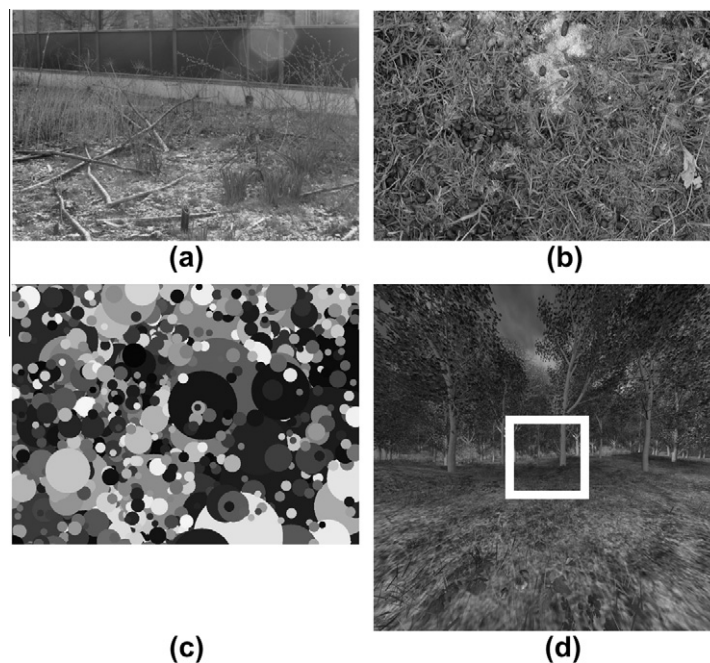


Fig. 2. (a) Image from the van Hateren and van der Schaaf database. (b) Image of the UPenn dataset. (c) Image generated from the dead leaves model. (d) Naturalistic parametric image covering 120° of field of view. The central square marked in yellow corresponds to a region of $26^\circ \times 26^\circ$.

These scenes were constructed as wooded environments using highly detailed naturalistic tree models by Xfrog™. The scenes were rendered using Vizard by Worldviz™. Images were taken at the position of the eye of an agent of height 1.82 m walking through the environment. We considered that the eye was relaxed i.e. looking at infinity with a corresponding focal length of 16.67 mm. The observer's pitch angle of the viewpoint was randomly chosen between -5° and 5° and the field of view covered 120° horizontally and vertically. In order to have the same image resolution in the center of the field of view as reported by van der Schaaf and van Hateren (1996) and to have samples from the peripheral field of view the entire scene covers 5954×5954 pixels. We generated images of size 512×512 pixels at all locations within the field of view necessary for subsequent analysis. The ray tracing method was adapted in a way that mimics the thin lenses model, thus these images have no aberrations related to the shape of lenses. Note that previous studies confirmed that the second order statistics of such rendered image ensembles closely resemble those of natural images (Rothkopf & Ballard, 2009) so that the conclusions drawn in the following are not an artifact of the image generation process.

Finally, we applied our analysis to images from well known databases of natural images. The study of natural image statistics has led to a number of high quality calibrated image databases (see e.g. Doi et al., 2003; Tkačik et al., 2011; van Hateren & van der Schaaf, 1998). Nevertheless, these databases have been obtained with specific purposes in mind that have different consequences for the image data and, consequently, the statistical regularities present therein. For example, the van Hateren and van der Schaaf (1998) database was obtained for subsequent analyses that were based on the image isotropy assumption. Accordingly, the authors selected images that cover a horizontal field of view of only 26° (13° in each direction), while human's field of view is 60° in the nasal direction, 105° in the temporal, 75° up and 60° down (see Fig. 2). We revisit the image material from some of these databases and analyze their second order statistical regularities and compare them to those obtained on the dataset generated with the naturalistic rendered scenes, which have a much larger field of view.

2.7. Coordinates and parameters

For the estimation of the orientations of projected edges onto a spherical surface and for the estimation of the local power spectrum we consider 17 positions distributed across 3 eccentricities (ϵ) and eight polar angles (χ) according to Table 1.

The dead leaves images were projected at 11 eccentricities between 0° and 60° of visual angle in steps of 5° . At each position, statistics were computed on spherical patches of 128×128 pixels corresponding to 4.2° degrees of field of view. Finally, the three-dimensional scenes and VR images are available in color, but we convert them to grayscale. Every image was rescaled to values between 0 and 1.

Table 1
Angular positions across the visual field at which statistics were computed.

(ϵ, χ)	0	45	90	135	180	225	270	315
0	×							
30	×	×	×	×	×	×	×	×
50	×	×	×	×	×	×	×	×

3. Results and discussion

3.1. Distribution of edges on the sphere

In order to understand the effect of the imaging process on the input to the visual system we first consider its influence of the distribution of edge orientations. The artificially generated 3d edge elements of random length and uniformly distributed orientations were projected onto the sphere using the stereographic transform and their distribution sampled at 17 different positions given by Table 1 at a radial distance of 2 m. By starting with edge elements that are oriented uniformly we can quantify the influence of the imaging process by measuring the distribution of edge element orientations after projection.

The polar histograms of the distribution of projected edge elements is shown in Fig. 3a. First, the uniform distribution of orientations in the artificial input data is maintained for the projection at the model fovea. This result is important in that it confirms that the distribution of edge elements present in the visual scenes is not altered at the fovea by the considered imaging process. Thus, foveally the distribution of edge elements corresponds to the orientations present in the natural image input. Accordingly, previous work related to the distribution of edges in natural images (e.g. Coppola et al., 1998; Geisler & Perry, 2009) is not influenced by the imaging process, at least foveally. By contrast, for peripheral regions of the visual field the distributions are biased towards radial orientations. This means, that edge elements present in the natural environment will be projected on the eye ball in such a way that their radial components will be emphasized and that this effect increases with eccentricity.

We quantified the strength of the radial bias across the visual field by fitting separate von Mises distributions (e.g. Evans, Hastings, & Peacock, 2000) to the obtained distributions of edge orientations. The von Mises distribution is appropriate in that it is the analog of the Gaussian distribution but on a circular quantity, such as orientation, i.e. it is a continuous probability distribution defined for values between 0° and 360° . The corresponding polar plot in Fig. 3b confirms and quantifies the observed radial orientation preference.

3.2. Power spectrum of dead leaves on the sphere

Given that the distributions of edge elements projected on the eye ball show an increase of their radial components with increasing eccentricity as described above we now turn to the influence of the imaging process on the second order intensity statistics as quantified by the power spectra. To separate the influence of the projection onto a spherical surface and the position dependent blurring of the eye, we compare the power spectra for four different imaging cases.

We first use images sampled from the dead-leaves model at 11 positions across the visual field with increasing eccentricities $0^\circ, 5^\circ, \dots, 60^\circ$ and estimate the power spectrum at each position. Separately at each position within these planar images, we fit the resulting power spectrum with Eq. (2) and, as expected, the exponent α is constant across the field of view (Fig. 4, planar normal case). After that, we blur these images with the corresponding MTF function at each position within the field of view (Navarro, Artal, & Williams, 1993). From a signal processing point of view, the modulation transfer functions of the human eye are low pass filters where the cut off frequency decreases with eccentricity. Thus, the high frequencies are more and more attenuated towards the periphery. Accordingly, when fitting the power spectra with Eq. (2) the parameter α now increases with the eccentricity (Fig. 4 planar MTF case).

The procedure was repeated with spherical images leading to two more imaging cases. First, we projected the above images sampled

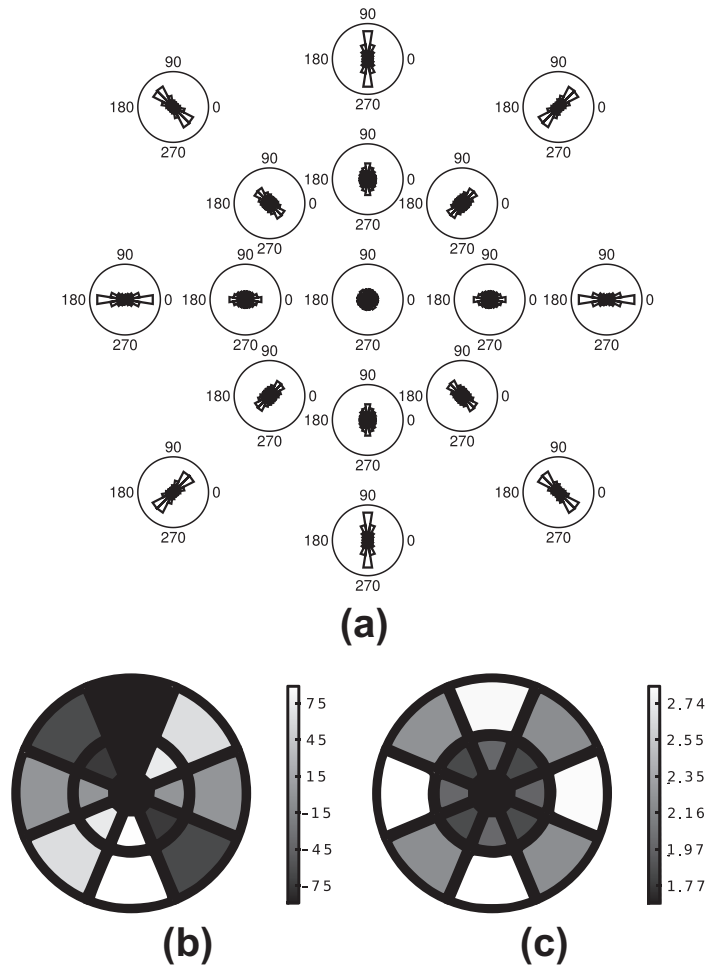


Fig. 3. (a) Histograms of edge orientations at different positions across the visual field as given in Table 1. (b) Mean of von Mises distribution fitted to the distribution of edges across the visual field. (c) Inverse-variance of von Mises distribution fitted to the distribution of edges across the visual field. The histograms on the upper part of the figure correspond to image data in the upper part of the visual field.

from the dead-leaves model onto a spherical model retina an fitted the power spectra separately across the visual field. While the exponent of the power spectra of the spherical images decreases with the eccentricity (Fig. 4 spherical normal case), additional blurring with the position dependent MTF results in an increase of the parameter α with the eccentricity. However, this increase is much smaller than in the planar case (Fig. 4 spherical MTF case). This shows that the strong low pass filtering properties of the human eye, which increase with eccentricity across the visual field, are attenuated by the spherical projection, thus both factors shape the statistics of the input signals to the retina. One should also notice that only foveally α is approximately 1, thus previous results only approximate the statistics of the natural input at the exact center of the retina.

3.3. Power spectrum of naturalistic input on the sphere

Having full control over the imaging process by utilizing the rendered naturalistic images, we first computed the power spectrum foveally over 10,000 planar images and verified that their rotationally averaged second order statistics match those that have been described for natural images. We start by estimating the radially averaged statistics according to Eq. (2) at random positions chosen uniformly across the entire visual field. The resulting power spectrum is shown in a log-log plot in Fig. 5 together with the corresponding rotationally averaged power spectrum from the van der Schaaf and van Hateren (1996) database. This confirms that indeed the rendered scenes have second order statistics that

match those described previously (Ruderman & Bialek, 1994; Ruderman, 1994; Tolhurst, Tadmor, & Chao, 1992).

In order to estimate the statistics of the natural input on the sphere, the process of the previous section was repeated with the images from virtual environments. Thus patches of planar images from virtual wooded environments were extracted and projected into the model eye ball according to their position in the visual field and blurred with the corresponding position dependent MTF. Fig. 1b shows an example planar patch and 1d shows the

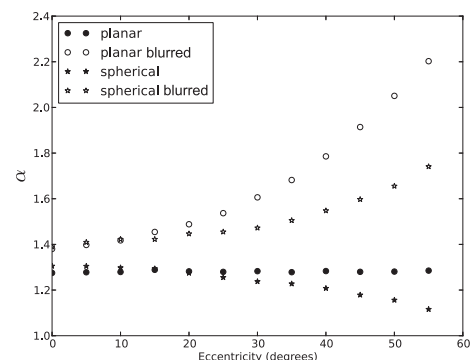


Fig. 4. Dependence of the parameter α describing the power spectra in Eq. (2) across the visual field as a function of eccentricity based on dead-leaves model images. Note the different dependence of the exponent α as a function of eccentricity for the different imaging cases.

corresponding projection onto the sphere for the model eye at the same position and with the same looking direction as the planar camera. The spherical patches were then reprojected into the corresponding tangent planes for further analysis.

For the initial analysis of the spherical images we estimated the average power spectrum irrespective of position in the visual field, i.e. the mean over the entire visual field, to compare our results to previous investigations that also applied such averaging. This average power spectrum is approximately circularly symmetric with additional activations on the vertical and horizontal axes and can be well fitted with Eq. (2) as has been done in previous studies. Thus the global averaged statistics of the spherical blurred images are radially symmetric. However, this does not guarantee that this is also the case for the local power spectra.

If it were the case that the statistics of spherical images were shift-invariant, i.e. their statistics were position independent as in the case of planar dead-leaves model images, the average power spectrum would again be similar to this average across the visual field. To test this hypothesis, we estimated the average power spectrum at each of the positions in the visual field given in Table 1. Fig. 6 shows the average power spectra at each of these positions and shows an elliptical instead of a circular shape, where the major axis of the ellipse points tangentially to radii emanating from the fovea. This verifies that the power spectra differ in their properties across the visual field, i.e. their shape, orientation, and magnitude depend on the eccentricity and polar angle in systematic ways. Thus, the power spectrum of these images is not position invariant.

To quantify this variability, we fitted the power spectra with the linear combination of an elliptical and a hyperbolic component, as given in Eq. (3). This description utilizes six parameters: A is a multiplying factor of both conics corresponding to an overall amplitude of the power spectrum, a corresponds to the quotient between the main axes xx and yy of the elliptical component, θ represents its rotation, and α is the exponent of the ellipse's power law. B is a mixing factor between the elliptical and the hyperbolic components, and β the exponent of the power law associated with the hyperbolic component. Table 2 and Fig. 7 present the resulting parameters at the chosen locations across the visual field.

There are several properties of the distributions of these parameters that should be highlighted. First, the exponent α is approximately 1 in the center of the field of view. Particularly, this is important in that it is in good agreement with previous results. But, over the range of the visual field considered here, both A and α increase with the eccentricity. This property reflects the blurred spherical transform, as demonstrated by the results using the dead-leaves image samples above.

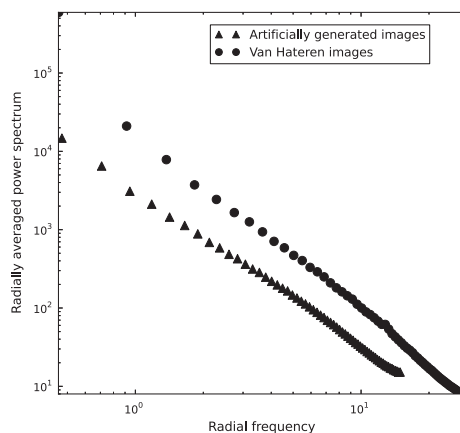


Fig. 5. Estimated power spectrum of planar rendered naturalistic images of wooded environment together with those from the van Hateren database. These power spectra were obtained by averaging radial power spectra from random positions across the entire image planes.

Furthermore, the elliptical component of the power spectrum shows a distinct pattern of orientation changes. The rotation of the ellipse θ reflects a tangential pattern, meaning $\theta \approx \chi + 90^\circ \bmod 180^\circ$. This is particularly evident in the depiction of the value of this parameter across the visual field in Fig. 7. The elliptical component tends to be more circularly symmetric at smaller eccentricities, as expected, and gets more elongated with increasing eccentricities. Note that this is an effect of the projective transformation of the visual input and that a related effect is present in perspective distortions in planar projections of visual scenes (Bruce & Tsotsos, 2006). The same procedure was applied to the planar images and the results were similar in terms of orientation and strong activation on the axis, but the values A and α diverge especially in the periphery. This again shows the importance of considering the full imaging process and the projection of images onto the spherical model eye. For more details please see the [Supplementary Material](#).

To confirm that these results are not artifacts of the naturalistic rendered scenes we applied the same analyses to natural images from the Van Hateren database (van Hateren & van der Schaaf, 1998). Because of their limited field of view, the comparison cannot be carried out for eccentricities larger than 13° and polar angles of $\pm 34^\circ$. Nevertheless, at this eccentricity the comparison in Fig. 8 shows already a small but clear difference between radial and tangential directions and also shows good agreement between the rendered scenes and the Van Hateren database. Note that the magnitude of this effect, i.e. the difference between the radial and tangential power, differs between planar photographic images and blurred spherical images, as expected from the above analysis.

4. Conclusions

The study of sensory systems is tied to the analysis of the statistics of the natural environment through the efficient coding hypothesis. Here we extended the characterization of the second order statistics of natural images by taking into account how the imaging process of a model eye additionally shapes the statistics of the input to the visual system. For that, the thin lense model was adapted in such way that 3D points are projected into a spherical surface instead of the commonly used plane. Under this model, we derived a transform that maps planar images into spherical ones by geometric projections and included the blur introduced by the optical properties of the visual system through the previously empirically measured modulation transfer function across the visual field.

We have seen that the process of projecting 3D locations from a scene into a spherical surface introduces orientation and metric biases to edge elements and that these effects depend on the

Table 2

Numerical values of the parameters of the generalized power spectra of spherical images in function of angular position.

(ϵ, χ)	A	a	α	θ	B	β	$\chi + 90 \bmod 180$
(0,0)	3376	1.0	1.1	81	4.8e-02	1.3	90
(30,0)	4112	1.0	1.3	79	2.2e-02	1.6	90
(30,45)	6422	0.9	1.3	141	5.0e-04	1.5	135
(30,90)	6657	0.9	1.3	2	7.8e-04	1.5	0
(30,135)	6251	0.9	1.2	33	2.0e-03	1.5	45
(30,180)	3955	1.0	1.3	78	2.4e-02	1.6	90
(30,225)	5769	0.8	1.4	123	3.0e-02	1.7	135
(30,270)	5338	1.0	1.4	8	2.3e-02	1.6	0
(30,315)	5812	0.7	1.3	56	3.4e-02	1.7	45
(50,0)	3374	0.9	1.7	78	1.3e-02	2.1	90
(50,45)	6138	0.7	1.6	130	1.0e-08	2.0	135
(50,90)	5809	0.7	1.7	0	3.2e-04	2.0	0
(50,135)	5943	0.7	1.6	45	1.0e-08	2.0	45
(50,180)	3340	0.9	1.7	78	1.2e-02	2.0	90
(50,225)	7076	0.6	1.8	123	6.3e-03	2.3	135
(50,270)	5955	0.8	1.8	0	2.1e-03	2.1	0
(50,315)	7106	0.6	1.8	56	6.4e-03	2.3	45

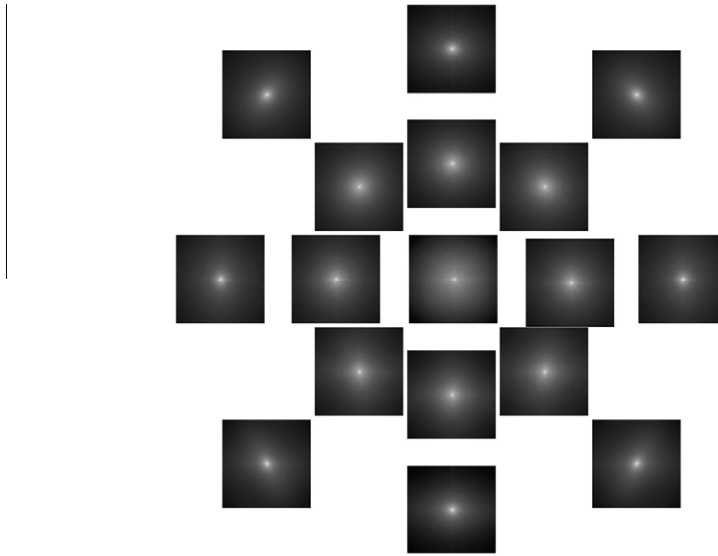


Fig. 6. Local power spectra across the visual field sampled at locations given in Table 1. Note that the positions were rotated with respect to their position on the retina so that the upper spectra correspond to the upper half of the visual field (towards the sky) whereas the spectra on the bottom half correspond to spectra on the lower half of the visual field (towards the ground plane).

position within the visual field. We verified that the average power spectra of the natural input systematically deviate from the assumption of circular symmetry and instead depend both on the frequencies f_x and f_y and the angular position within the visual field. The power spectra of the so projected images are well described by the proposed generalized power law. We verified that overall the component along the cardinal directions contributes more power towards the center of the field of view and that the orientation of the elliptical power component is approximately orthogonal to the radial direction with respect to the center of projection. These signatures of the systematic deviations in the power spectra were also found in the widely used van Hateren natural image database, although these effects are small because of the small field of view of these images. In future work we will investigate the consequences of the properties of local power spectra across the visual field for model retinal Ganglion cells under different

formulations of optimal sensory coding. Furthermore, we will include the influence of natural gaze behavior on these statistics (Rothkopf, Weisswange, & Triesch, 2009).

It should also be mentioned that we did not find any image database that fulfills the requirements of large field of view and calibration in order to obtain the correct projections onto planes tangential to a model eye ball. We therefore rendered naturalistic scenes in a virtual environment. This way, we could control the parameters of the image acquisition process, as well as the position and orientation of the imaging system corresponding to a human walking through a forest. The results obtained on the basis of this artificial data set confirm the usefulness of constructing naturalistic virtual environments with a full three dimensional scene layout as it allows accessing the relevant scene points.

We conclude that the properties of the natural input to the visual system not only depend on the statistics of visual scenes in

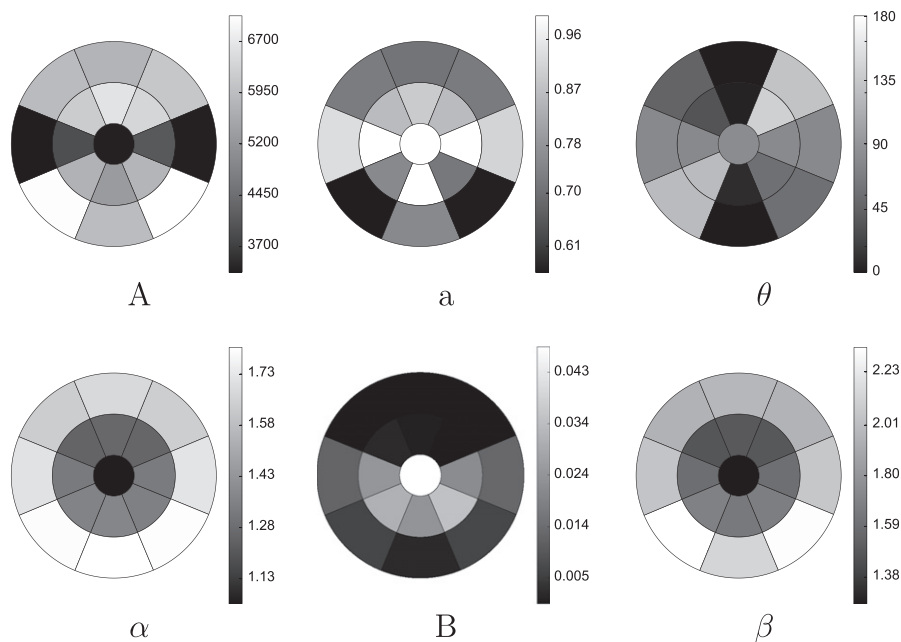


Fig. 7. Spatial distribution of the parameters used to fit the power spectra across the visual field.

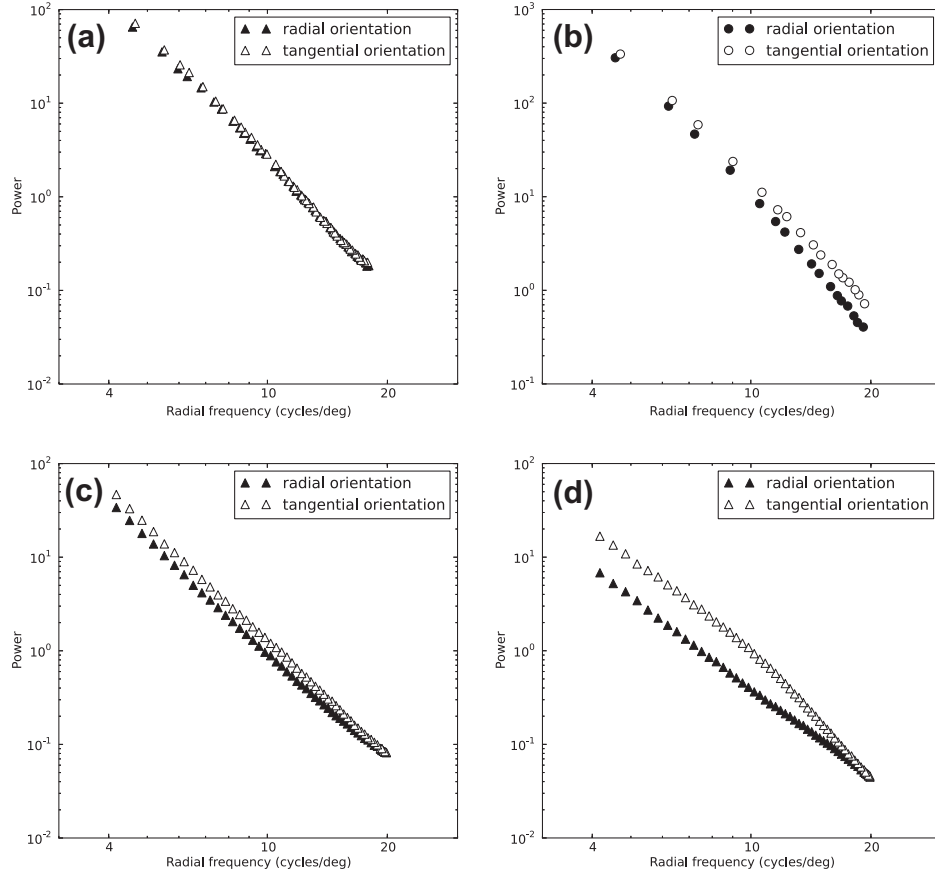


Fig. 8. Comparison of estimated power spectra of spherical Van Hateren images and spherical naturalistic rendered images (a) Radial and tangential power spectra of spherical blurred artificially generated images at eccentricity 13° and polar angle -34°. (b) Profile of the power spectrum of spherical blurred Van Hateren images at eccentricity 13° and polar angle -34°. (c) Same as in (a) but at eccentricity 30°, (d) Same as in (a) but at eccentricity 50°.

the environment, but also on the statistics imposed on the stimulus by the imaging process as embodied by the animal's eye. These effects depend on the parameters of the imaging system such as focal length, field of view, and eye's position and orientation. Accordingly, to understand the properties of sensory coding it is important to not only consider the statistics of natural images by using natural image databases but also the statistics imposed by the respective animal eyes.

Acknowledgments

This research was supported by the BMBF Project Bernstein Fokus: Neurotechnologie Frankfurt, FKZ 01GQ0840. The authors would like to thank Andrew Worzella for the help in generating the artificial images and the comments by one of the anonymous reviewers.

Appendix A. Projective geometry

Here we detail the projective transform from world coordinates to points on the spherical model eye ball as shown schematically in Fig. 1c. This model extends the usual projective transform and allows the position of the projecting plane ($z = v$) to depend not only on the focal length f but also on the distance to the focused point u , according to the law $\frac{1}{f} = \frac{1}{v} + \frac{1}{u}$. Thus $v = -\frac{fu}{f-u}$. A point in the 3d world $P = (X, Y, Z, 1)$ is projected to the point $(x, y, v, 1)$, by the transform:

$$(x, y, z, 1)^T = \frac{v}{Z} I_4(X, Y, Z, 1)^T, \quad (4)$$

where I_4 is the identity matrix of size 4×4 .

This projection, generalizes the thin lenses model presented above in such a way that points of the 3d world are projected into the back of a sphere whose radius depends on the focal length and the distance to the object by $\tilde{r} = \frac{1}{2} \frac{fu}{u-f}$, and centered at $(0, 0, -\tilde{r})$ (note: $\tilde{r} = |v/2|$). The projective transform on the sphere \mathcal{S} is then defined by:

$$(x', y', z', 1)^T = -2\tilde{r} \frac{Z}{R^2} I_4(X, Y, Z, 1)^T, \quad (5)$$

where $R = \sqrt{X^2 + Y^2 + Z^2}$. The resulting image will be called *spherical*, since its coordinates are defined over a spherical surface. As an example, Fig. 1d shows the projection of the visual scene onto the spherical surface.

Because of the geometry of the projections, the angle between the rays and the plane $y = 0$ is the same in the planar and spherical case. Therefore these two maps are related by the transform \mathcal{R} :

$$(x', y', z', 1)^T = \frac{v^2}{r^2} I_4(x, y, v, 1)^T, \quad (6)$$

where $r = \sqrt{x^2 + y^2 + v^2}$ allowing us to obtain the spherical coordinates from the planar ones, and consequently the spherical image. The spherical pixel value $I_s(x', y', z')$ is equal to the value of the projective pixel value $I(x, y, v)$ if $(x', y', v', 1) = \mathcal{R}(x, y, v, 1)$. Meaning $I_s(x', y', z') = I(\mathcal{R}^{-1}(x', y', z'))$. In case $\mathcal{R}^{-1}(x', y', z')$ is not integer, we used a bicubic interpolation.

Appendix B. Analysis of spherical images

We first consider the *stereographic transform*, which projects points P onto points P' on the plane at the back of the model eye ball according to Fig. 1c. The sphere of radius \tilde{r} is centered at

$(0, 0, -\tilde{r})$, the center of projection is $(0, 0, -2\tilde{r})$, and the North Pole, the opposite point to the center of projection on the sphere, is $(0, 0, 0)$. The projecting plane, which is tangential to the center of projection, is given by the equation: $z = -2\tilde{r}$. Accordingly, the projection is such that each point on the sphere $P = (x', y', z')$ is projected on the point P'' of intersection between the plane and the segment $\overline{P'NP}$. This can be expressed by:

$$(x'', y'', z'')^T = (NP, 1)^T + \mu I_4((x', y', z', 1) - (NP, 1))^T. \quad (7)$$

Substituting $\mu = -\frac{(2\tilde{r})^2}{z}$ and $NP = (0, 0, 0)$ this simplifies to:

$$(x'', y'', z'')^T = \mu I_4(x', y', z', 1)^T. \quad (8)$$

The second transform used in the analysis of the power spectra is the *generalized stereographic transform*. The general idea is the same as for the stereographic transform. Given any point on the spherical surface as the center of projection (x_c, y_c, z_c) its opposite point is given by: $NP = (-x_c, -y_c, -2\tilde{r} - z_c)$, and the tangential plane is $\tau: x_c x + y_c y + (z_c + \tilde{r})z + \tilde{r}z_c = 0$. Each point on the sphere $p' = (x', y', z')$ is then projected onto the intersection point between the plane τ and the segment $\overline{p'NP}$, as in the previous case. We define the generalized stereographic transform, updating only the values of NP to $(-x_c, -y_c, -2\tilde{r} - z_c)$, and μ to $\frac{2\tilde{r}^2}{xx_c + yy_c + z(z_c + \tilde{r}) + \tilde{r}(2\tilde{r} + z_c)}$ in the expression (7), resulting in the following expression:

$$(x'', y'', z'')^T = (-x_c, -y_c, -2\tilde{r} - z_c, 1)^T + \frac{2\tilde{r}^2}{xx_c + yy_c + z(z_c + \tilde{r}) + \tilde{r}(2\tilde{r} + z_c)} I_4((x', y', z', 1) - (NP, 1))^T. \quad (9)$$

Patches of the spherical image can now be projected into a plane. However this plane is oblique, thus the representation of the image still utilizes three coordinates (x'', y'', z'') for all centers of projection but the four cardinal points. To solve this problem we just need to rotate the plane. The axis $[u, v]$ and the angle of rotation λ are given directly by the center of projection in the following way: $\lambda = \arccos\left(\frac{-(z_c + \tilde{r})}{\sqrt{(x_c^2 + y_c^2 + (z_c + \tilde{r})^2)}}\right)$, $u = -y_c / \text{norm}([x_c, y_c])$; $v = x_c / \text{norm}([x_c, y_c])$. The rotation is done multiplying the coordinates on the tangential plane the by a matrix:

$$\begin{pmatrix} \cos \lambda + u^2(1 - \cos \lambda) & uv(1 - \cos \lambda) & v \sin \lambda & 0 \\ uv(1 - \cos \lambda) & \cos \lambda + v^2(1 - \cos \lambda) & -u \sin \lambda & 0 \\ -v \sin \lambda & u \sin \lambda & \cos \lambda & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (10)$$

After this rotation, the patch is ready to be analyzed with the usual planar methods. Note that this rotation does not introduce any distortion on the image and the distance between points on the plane is maintained by the transform. Throughout the subsequent analysis, we always use a homogeneous coordinate system, where the measurement unit is 1 mm. To map an image in pixels (i, j) to homogeneous coordinates (x, y, z) , we need only three parameters, i.e. the total number of pixels (M, N) , the size of the projecting plane s_N, s_M in a camera sensor size or the total field of view, and finally the distance between the sensor and the nodal point v : $(x, y, z) = (\frac{s_N}{N}(N/2) - j, \frac{s_M}{M}(i - M/2), v)$.

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.visres.2013.01.011>.

References

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61, 183–193.

- Baddeley, R. (1997). The correlational structure of natural images and the calibration of spatial representations. *Cognitive Science*, 21(3), 351–371.
- Balboa, R. M., & Grzywacz, N. M. (2003). Power spectra and distribution of contrasts of natural images power spectra and distribution of contrasts of natural images from different habitats. *Vision Research*, 43, 2527–2537.
- Balboa, R., Tyler, C., & Grzywacz, N. (2001). Occlusions contribute to scaling in natural images. *Vision Research*, 41(7), 955–964.
- Ballard, D. (1991). Animate vision. *Artificial Intelligence Journal*, 48, 57–86.
- Barlow, H. (1961). Sensory communication. In *Ch. Possible principles underlying the transformation of sensory messages* (pp. 217–234). Cambridge, MA, USA: MIT Press.
- Bruce, N. D., & Tsotsos, J. K. (2006). A statistical basis for visual field anisotropies. *Neurocomputing*, 69(10–12), 1301–1304 [Computational Neuroscience: Trends in Research 2006].
- Burge, J., & Geisler, W. (2011). Optimal defocus estimation in individual natural images. *Proceedings of the National Academy of Sciences*, 108(40), 16849–16854.
- Burton, G., & Moorhead, I. (1987). Color and spatial structure in natural scenes. *Applied Optics*, 26(1), 157–170.
- Carathéodory, C. (1998). *Conformal representation*. Dover Publications.
- Carlson, C. (1978). Thresholds for perceived image sharpness. *Photographic Science and Engineering*, 22, 69–71.
- Coletta, V. P. (2010). *Physics fundamentals* (2nd ed.). Physics Curriculum and Instruction.
- Coppola, D., Purves, H., McCoy, A., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences*, 95(7), 4002.
- Daugman, J. (1989). Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Transactions on Biomedical Engineering*, 36(1), 107–114.
- Deriugin, N. (1956). The power spectrum and the correlation function of the television signal. *Telecommunications*, 1(7), 1–12.
- Doi, E., Inui, T., Lee, T., Wachtler, T., & Sejnowski, T. (2003). Spatiochromatic receptive field properties derived from information-theoretic analyses of cone mosaic responses to natural scenes. *Neural Computation*, 15(2), 397–417.
- Emsley, H. (1948). *Visual optics* (5th ed., Vol. 1). Hutton Press.
- Evans, M., Hastings, N., & Peacock, B. (2000). *Statistical distributions* (3rd ed.). Wiley, Ch. von Mises Distribution.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12), 2379–2394.
- Geisler, W. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, 59, 167–192.
- Geisler, W., & Perry, J. (2009). Contour statistics in natural images: Grouping across occlusions. *Visual Neuroscience*, 26(01), 109–121.
- Hecht, E. (2001). *Optics* (4th ed.). Addison Wesley.
- Kersten, D. (1987). Predictability and redundancy of natural images. *Journal of the Optical Society of America A*, 4(12), 2395–2400.
- Land, M., & Nilsson, D. (2002). *Animal eyes*. Oxford University Press.
- Langer, M. (2000). Large-scale failures of $f^{-\alpha}$ scaling in natural image spectra. *Journal of the Optical Society of America A*, 17(1), 28–33.
- Lee, A., Mumford, D., & Huang, J. (2001). Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *International Journal of Computer Vision*, 41(1), 35–59.
- Matheron, G. (1975). *Random sets and integral geometry*. Vol. 261. New York: Wiley.
- Navarro, R., Artal, P., & Williams, D. (1993). Modulation transfer of the human eye as a function of retinal eccentricity. *Journal of the Optical Society of America A*, 10(2), 201–212.
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3), 145–175.
- Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computations in Neural Systems*, 10, 341–350.
- Rothkopf, C., & Ballard, D. (2009). Image statistics at the point of gaze during human navigation. *Visual Neuroscience*, 26(01), 81–92.
- Rothkopf, C., Weisswange, T., Triesch, J., 2009. Learning independent causes in natural images explains the spacevariant oblique effect. In *IEEE 8th International Conference on Development and Learning. IEEE*.
- Ruderman, D. (1994). The statistics of natural images. *Network: Computation in Neural Systems*, 5(4), 517–548.
- Ruderman, D. (1997). Origins of scaling in natural images. *Vision Research*, 37(23), 3385–3398.
- Ruderman, D., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical Review Letters*, 73(6), 814–817.
- Shannon, C. (1948). The mathematical theory of communication: 1963. *MD Computing Computers in Medical Practice*, 14(4), 306–317.
- Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24, 1193–1216.
- Tkačik, G., Garrigan, P., Ratliff, C., Milčinski, G., Klein, J., Seyfarth, L., et al. (2011). Natural images from the birthplace of the human eye. *PLoS One*, 6(6), e20409.
- Tolhurst, D., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic and Physiological Optics*, 12(2), 229–232.
- van der Schaaf, A., & van Hateren, J. (1996). Modelling the power spectra of natural images: Statistics and information. *Vision Research*, 36(17), 2759–2770.
- van Hateren, J., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society London B*, 265, 359–366.