Natural Image Statistics for Mouse Vision

Authors: Luca Abballe[1] and Hiroki Asari[2*]

Affiliations:

1. Department of Biomedical Engineering, Sapienza University of Rome, Rome RM, 00185, Italy

2. Epigenetics and Neurobiology Unit, EMBL Rome, European Molecular Biology Laboratory, Monterotondo RM, 00015, Italy

∗**Corresponding author:** Hiroki Asari, E-mail: asari@embl.it

Acknowledgments: This work was supported by research grants from EMBL (H.A.). We thank all the members of the Asari lab at EMBL Rome for many useful discussions.

Author contributions: L.A. collected and processed the image data sets. H.A. designed the study, analyzed the results, and wrote the manuscript.

Competing interests: The authors declare no competing financial interests.

Abstract

The mouse has dichromatic colour vision based on two different types of opsins: short (S)- and middle (M)-wavelength-sensitive opsins with peak sensitivity to ultraviolet (UV; 360 nm) and green light (508 nm), respectively. In the mouse retina, the cone photoreceptors that predominantly express the S-opsin are more sensitive to contrasts, and denser towards the ventral retina, preferentially sampling the upper part of the visual field. In contrast, the expression of the M-opsin gradually increases towards the dorsal retina that encodes the lower visual field. Such distinct retinal organizations are assumed to arise from a selective pressure in evolution to efficiently encode the natural scenes. However, natural image statistics of UV light have never been examined beyond the spectral analysis. Here we developed a multi-spectral camera and examined the UV and green image statistics of the same natural scenes. We found that the local contrast and the spatial correlation were higher in UV than in green for images above the horizon, but lower in UV than in green for those below the horizon. This suggests that the mouse retina is not necessarily optimal for maximizing the bandwidth of information transmission. Factors besides the coding efficiency, such as visual behavioural requirements, will thus need to be considered to fully explain the characteristic organization of the mouse retina.

Introduction

Sensory systems have been considered to be adapted to the statistical properties of the environment through evolution (Simoncelli and Olshausen, 2001). Animals encounter different types of sensory signals depending on their natural habitats and lifestyles, and this can serve as an evolutionary driving force for each species to optimize its sensory systems for processing those signals that appear more frequently and are relevant for survival (Baden et al., 2020). The optimality of the sensory processing has been broadly supported from an information theoretic viewpoint of coding efficiency (Attneave, 1954; Barlow, 2004). In particular, various physiological properties of sensory neurons can be successfully derived from learning efficient codes of natural images or natural sounds, such as separation of retinal outputs into ON and OFF channels (Gjorgjieva et al., 2014), Gabor-like receptive fields of visual cortical neurons (Olshausen and Field, 1996), and cochlear filter banks (Smith and Lewicki, 2006). Such computational theories and statistical models are, however, often limited to generic features of the sensory processing, and fail to account for species-specific fine details partly due to a lack of proper data sets of natural sensory signals.

In the past decade, the mouse has become a dominant model for studying the visual system mainly because of the wide availability of experimental tools (Huberman and Niell, 2011). Compared to other mammalian model animals such as cats and primates, however, the mouse vision has certain peculiar properties. For example, mice are dichromats as many other mammals are, but their retina expresses ultraviolet (UV)-sensitive short (S)-wavelength sensitive opsins and green-sensitive middle (M)-wavelength sensitive opsins (Jacobs et al., 1991; Lyubarsky et al., 1999; Nikonov et al., 2006). While UV vision is common in amphibians, birds and insects, it has not been identified in mammals except for a few species including rodents (Cronin and Bok, 2016; Hunt et al., 2001; Peichl, 2005). Moreover, the mouse retina has no fovea but a prominent dorsoventral gradient in the expression pattern of the two opsins (Applebury et al., 2000; Lyubarsky et al., 1999; Röhlich et al., 1994; Szél et al., 1992). A vast majority of the mouse cone photoreceptors ($\sim$95%) co-express the two opsins but with a dominant expression of S- and M-opsins in the ventral and dorsal parts of the retina, respectively (Baden et al., 2013; Jacobs et al., 1991; Lyubarsky et al., 1999; Wang et al., 2011). This makes

the upper visual field more sensitive to UV than green, and vice versa for the lower visual field (Jacobs and Williams, 2007). It is natural to assume that this functional segregation of the mouse vision has evolved due to an adaptation to the natural light distribution as the sunlight is the major source of UV radiation. It remains unclear, though, how optimal the mouse visual system is to natural scene statistics *per se*.

While natural image statistics have been extensively studied thus far (Hyvärinen et al., 2009; Simoncelli and Olshausen, 2001), those outside the spectral domain of human vision remain to be fully explored (Baden et al., 2020, 2013). Here we thus developed a multi-spectral camera system to sample images that spectrally match the mouse photopic vision, and analyzed the first- and second-order statistics of the UV and green image data sets. We identified distinct statistical properties in the UV and green channels between the upper and lower visual-field images; however, these image statistics were not necessarily consistent with what the efficient coding hypothesis would predict from the functional organization of the mouse retina.

Materials and Methods

All data and codes are available upon request.

Multi-spectral camera

Design

We built a multi-spectral camera system based on a beam-splitting strategy (Grahn and Geladi, 2007; Hagen and Kudenov, 2013) to acquire images of the same scenes with ultraviolet (UV)- and green-transmitting channels that match the spectral sensitivity of the mouse photopic vision (Figure 1A; Jacobs et al., 1991; Lyubarsky et al., 1999; Nikonov et al., 2006). The light coming from a commercial camera lens (Nikon, AF Nikkor 50 mm f/1.8D) was collimated with a near-UV achromatic lens (effective focal length, 50 mm; Edmund Optics, 65-976) and split with a dichroic filter (409 nm; Edmund Optics, 68-330). The reflected light, on the one hand, passed through a UV-selective filter set (HOYA U-340 and short-pass filter at 500 nm; Edmund Optics,

84-708) and formed the UV images focused on the first global-shutter camera (Imaging Source, DMK23UX174) with a near-UV achromatic lens (effective focal length, 50 mm; Edmund Optics, 65-976). The transmitted light, on the other hand, passed through a band-pass filter (500±40 nm; Edmund Optics, 65-743) and a lens (Edmund Optics, 65-976), and formed the green images sampled by the second camera (Imaging Source, DMK23UX174). To maximize the dynamic range of the two camera sensors (used with the same settings), we attenuated the light intensity of the green channel using an absorptive neutral density (ND) filter (optical density ranging from 1.0 to 2.0) on a filter wheel (Thorlabs, LTFW6) because the sunlight has much higher power in green than in UV (Figure 1B). The optical components are all mounted with standard light-tight optomechanical components (Thorlabs, 1-inch diameter lens tubes).

Spectral analysis

The spectral sensitivity of the multi-spectral camera system (Figure 1B) was calculated by convolving the relative transmission spectra of the optics for each channel with the spectral sensitivity of the camera sensor (Sony, IMX174 complementary metal–oxide–semiconductor; CMOS). The relative transmission spectra were measured with a spectrometer (Thorlabs, CCS200/M; 200–1000 nm range) by taking the ratio of the spectra of a clear sunny sky (indirect sunlight) with and without passing through the camera optics.

For a comparison, we modelled the spectral sensitivity of the mouse visual system by convolving the transmission spectra of the mouse eye (Henriksson et al., 2010) with the absorption spectra of the mouse cone photoreceptors (Figure 1B). We used a visual pigment template (Govardovskii et al., 2000) with the center frequency at 360 nm and 508 nm to simulate the short (S)- and middle (M)-wavelength-sensitive opsins in the mouse retina, respectively (Jacobs et al., 1991; Lyubarsky et al., 1999; Nikonov et al., 2006).

Image acquisition

In total, we collected 168 images of natural scenes without any artificial object in the suburbs of Lazio/Abruzzo regions in Italy from July 2020 to November 2020. All the images were acquired using a custom-code in Matlab (Image Acquisition Toolbox) without any automatic

image correction, such as gain or contrast adjustment. The two cameras were set with the same parameter values adjusted to each scene, such as the exposure length. The camera system was placed on the ground to follow the viewpoint of mice, and the following meta-data were also recorded upon image acquisition: date, time, optical density of ND filter in the green channel, weather condition (sunny; cloudy), distance to target object (short, within a few meters; medium, within tens of meters; or long), and camera elevation angle (looking up; horizontal; looking down). We also took a uniform image of a clear sunny sky (indirect sunlight) as a reference image for vignetting correction (see below Eq.(1)).

Image registration

The raw images from the two cameras (16 bit, 1920-by-1200 pixels each) were pre-processed to form a registered image in Matlab (Image Processing Toolbox). First, we corrected the optical vignetting by normalizing the pixel intensity of the raw image $I_{\text{raw}}(x, y)$ for each channel by the ratio of the pixel and the maximum intensities of the reference image $I_{\text{ref}}(x, y)$:

$$I_{\text{corrected}}(x, y) = I_{\text{raw}}(x, y) \cdot \frac{\max I_{\text{ref}}(x, y)}{I_{\text{ref}}(x, y)}. \tag{1}$$

We next applied a two-dimensional median filter (3-by-3 pixel size) to remove salt-and-pepper noise from the corrected images for each channel. Then we applied a projective transformation based on manually selected control points to register the UV image to the green image. Finally, we manually cropped the two images to select only those areas in focus. The cropped images resulted in the pixel size ranging from 341 to 1766 pixels in the horizontal axis and from 341 to 1120 pixels in the vertical axes (see examples in Figure 2).

Image analysis

We analyzed the first- and second-order image statistics of the obtained natural scenes in UV and green channels. Here we excluded a small set of the horizontal images ($N = 9$) from the analysis, and focused on the following two major image groups: 1) looking-up images taken with a positive camera elevation angle ($N = 73$), presumably falling in the ventral retina and

thus perceived in the upper part of an animal's visual-field; and 2) looking-down images with a negative camera elevation angle ($N = 86$) perceived in the lower visual field (i.e., the dorsal retina). For each image data set, we used a sign-test to compare the image statistics parameter values between the UV and green channels (Figures 3–5; significance level, 0.05). All image analysis was done in Matlab (Mathworks).

Normalized light intensity

The visual system adapts its sensitivity to the range of light intensities in each environment (Gollisch and Meister, 2010; Shapley and Enroth-Cugell, 1984). We thus first normalized the pixel intensity of each UV and green image to have the intensity value ranging from zero to one (by subtracting the minimum value of the image, followed by the division by the maximum value), and then calculated the histogram (bin size, 0.01) to compare the normalized intensity distributions of the UV and green images for the upper and lower visual fields (Figure 3A,B).

Local contrast

To calculate the local statistical structure of the normalized intensity images (Figure 3C,D), we used the second-derivative (Laplacian) of a two-dimensional Gaussian filter:

$$\mathrm{LoG}(x, y) = \frac{1}{\pi\sigma^4}\left(1 - \frac{x^2 + y^2}{2\sigma^2}\right)\exp\left[-\frac{x^2 + y^2}{2\sigma^2}\right], \tag{2}$$

with the standard deviation $\sigma = 10$ pixels for the spacial range $x, y \in [-3\sigma, 3\sigma]$. The local contrast distribution was then fitted to the two-parameter Weibull distribution:

$$w(x) = \beta\gamma|x|^{\gamma-1}\exp\left[-\beta|x|^{\gamma}\right], \tag{3}$$

where x is the local contrast value, $\beta > 0$ is the scale parameter (width) of the distribution, and $\gamma > 0$ is the shape parameter (peakedness). In particular, a larger β value indicates higher contrast; and a larger γ value indicates more fine textures in the images. Sign-tests were used to compare these parameter values between UV and green images (Figure 3E–H).

Power spectral density

The power spectral density of the normalized intensity image $I(x, y)$ was computed with the fast Fourier transform (FFT; Figure 4):

$$F(\omega_x, \omega_y) = \text{FFT}\left[I(x, y)\right] \tag{4}$$

$$S(\omega_x, \omega_y) = F(\omega_x, \omega_y)F^*(\omega_x, \omega_y), \tag{5}$$

where the superscript $*$ denotes complex conjugate, and ω_x and ω_y represent the horizontal and vertical spacial frequency (ranging from -0.5 to 0.5 cycles/pixel), respectively. As the average power spectrum of natural images generally falls with a form $1/f^\alpha$ over the spatial frequency f with a slope $\alpha \sim 2$ (Burton and Moorhead, 1987; Simoncelli and Olshausen, 2001; Tolhurst et al., 1992), we fitted the power function b/ω^a to $S(\omega_x, 0)$ and $S(0, \omega_y)$, where a and b indicate the slope and Y-intercept in the log-log space. We used a sign-test to compare these parameter values between UV and green channels (Figure 4I–P).

Spatial autocorrelation

Following the Wiener–Khinchin theorem, the spatial autocorrelation $R(x, y)$ was computed with the inverse FFT of $S(\omega_x, \omega_y)$ in Eq.(5):

$$R(x, y) = \text{IFFT}\left[S(\omega_x, \omega_y)\right], \tag{6}$$

where x and y represent horizontal and vertical distances of the two pixel points in the target image, respectively (Figure 5). Sign-tests were used to compare the $R(d_\text{h}, d_\text{v})$ values at representative data points: $[d_\text{h}, d_\text{v}] = [0, 50], [50, 0]$ (Figure 5I–L).

Results

Multi-spectral camera for the mouse vision

The mouse retina expresses short (S)- and middle (M)-wavelength sensitive opsins that are maximally sensitive to ultraviolet (UV; ∼360 nm) and green (∼508 nm) wavelengths of light, respectively (Jacobs et al., 1991; Lyubarsky et al., 1999; Nikonov et al., 2006). Existing public databases of natural scenes contain a diverse set of images including both natural and artificial objects in both gray and color scales visible to humans, but do not cover any UV images (e.g., Geisler and Perry, 2011; Hateren and Schaaf, 1998; Tkačik et al., 2011; Zhou et al., 2014). To examine the natural image statistics of the mouse vision, we thus set out to build a multi-spectral camera system for acquiring images of the same scenes in both UV and green spectral domains (Figure 1).

We first modelled the spectral sensitivity of the mouse dichromatic vision to determine the center wavelengths of the two channels. Because the lens and cornea absorb shorter wavelength light (e.g., UV rays) more than longer wavelength light, we corrected the absorption spectra of the mouse cone photoreceptors (Govardovskii et al., 2000) with the transmission spectra of the whole eye optics (Henriksson et al., 2010). This resulted in a slight shift of the center wavelengths to a longer wavelength by several nanometers: from ∼360 nm to ∼365 nm for the S-cone and from ∼508 nm to ∼512 nm for the M-cone (Figure 1B). Thus the ocular transmittance had only minor effects on the spectral sensitivity of the mouse vision, reassuring its sensitivity to near-UV light (Jacobs and Williams, 2007; Jacobs et al., 2004).

We then designed a multi-spectral camera system accordingly using a beam-splitting strategy (Figure 1A; see Methods for specifications; Grahn and Geladi, 2007; Hagen and Kudenov, 2013). By convolving the transmission spectrum of the camera optics with the sensitivity spectrum of the camera sensors, we identified that our imaging device had the sensitivities to ∼368±10 nm and ∼500±30 nm (center wavelength ± half-width at half maximum; HWHM) for the UV and green channels, respectively (Figure 1B). This confirms that the UV and green channels of our device were spectrally well isolated, and that the two

channels largely matched to the spectral sensitivity of the mouse vision (Jacobs et al., 1991; Lyubarsky et al., 1999; Nikonov et al., 2006).

Ultraviolet and green image collection

To collect images that mice would encounter in their natural habitats, we went out to natural fields and wild forests in the countryside and mountain area of Lazio/Abruzzo regions in Italy across different seasons. We placed the multi-spectral camera on the ground at about a height of the mouse eye, and acquired images of natural objects alone at various distances (e.g., clouds, trees, and flowers), excluding any artificial objects. These images were taken with different camera angles in the presence of ample natural light. The images were preprocessed to correct optical vignetting and remove salt-and-pepper noise, and cropped to exclude areas out of focus on the edges (see Methods for details). This led to a set of 168 pairs of UV and green images of various "mouse-view" natural scenes.

Besides well-known facts that UV light is reflected well by open water and some plants (Cronin and Bok, 2016; Hunt et al., 2001), we noticed several distinct features between the UV and green images (see examples in Figure 2). First, clouds often appeared dark and faint in the UV images than in the green ones. In some cases, even negative contrast was formed for the coulds in UV while positive contrast in green. Second, fine textures were more visible in the green images than in the UV ones. In particular, objects in the upper-field UV images were often dark in a nearly uniform manner due to back-light, whereas fine details of the objects were nevertheless visible in the corresponding green images despite a high contrast against the sky. For the lower field images, in contrast, distinct brighter spots stood out in UV due to reflections of shiny leaves and cortices, while more shades and shadows were visible in green. These qualitative observations already suggest that the UV and green images have distinct statistical properties.

Normalized intensity and contrast distributions of UV and green images

To analyze the image statistics more formally, we first calculated the normalized intensity distribution of the UV and green channels for the upper and lower visual-field images

(Figure 3A,B). Because the visual system adapts its sensitivity to the range of light intensities in each environment (Gollisch and Meister, 2010; Shapley and Enroth-Cugell, 1984), we normalized the pixel intensity of each UV and green image to be within the range from zero to unity. We then found that, for the upper visual-field images, the probability distributions of both UV and green intensity values were bimodal (Figure 3A). The two peaks of the UV intensity distribution, however, were higher and more separated than those of the green intensity distribution, suggesting that luminance contrast is higher in UV than in green when animals look up. In contrast, the normalized intensity distributions of the lower field images were unimodel and skewed to the right for both color channels. The distribution was more heavy-tailed for the green than for the UV images (Figure 3B), indicating higher contrast in green than in UV when animals look down.

To better examine the contrast in the two different spectral domains, we calculated the local image contrast using the second derivative (Laplacian) of a two-dimensional Gaussian filter (Eq.(2) in Methods). This filter follows the antagonistic center-surround receptive fields of early visual neurons (e.g., retinal ganglion cells; Enroth-Cugell and Robson, 1966; Rodieck and Stone, 1965) that are sensitive to local contrast, and is commonly used for edge detection in computer vision (Ghosh et al., 2007; Lindeberg, 2013; Marr and Hildreth, 1980). Consistent with what was implicated by the intensity distributions (Figure 3A,B), we found that the probability distribution of local contrast was wider for the upper visual-field UV images than for the corresponding green images (Figure 3C), but narrower for the lower visual-field UV images than for the green counterparts (Figure 3D). To quantify these differences, we fitted a two-parameter Weibull function (Eq.(3) in Methods) to the local contrast distribution of each image in each channel (Geusebroek and Smeulders, 2005; Gijsenij and Gevers, 2011), where the first scale parameter (β) describes the width of the distribution, hence a larger value indicating higher contrast; and the second shape parameter (γ) describes the peakedness, with a larger value indicating more fine textures in the image. For the images above the horizon, the UV channel had significantly larger scale parameter values and smaller shape parameter values than the green channel (Figure 3E,G). In contrast, for the images below the horizon, the UV channel had significantly smaller scale parameter values than the green channel (Figure 3F), with no

difference in the shape parameter values (Figure 3H). Thus the image statistics showed distinct characteristics between the upper and lower visual-field image data sets, with higher contrast and less fine textures in UV than in green for the upper visual-field images, while higher contrast in green for the lower visual-field images.

Importantly, such differences in the local contrast distributions do not agree well with what the efficient coding hypothesis implies from the physiological and anatomical properties of the mouse retina (Attneave, 1954; Barlow, 2004). Purely from an information theoretic viewpoint, a narrower contrast distribution is better encoded with a more sensitive cone type to maximize its bandwidth (Laughlin, 1981). In the mouse retina, the functional S-cones are more sensitive to contrast than the functional M-cones (Baden et al., 2013; Jacobs and Williams, 2007; Jacobs et al., 2004; Röhlich et al., 1994; Wang et al., 2011); and the functional S-cones are denser towards the ventral part of the retina, preferentially sampling the upper part of the visual field, while the functional M-cones towards the dorsal retina, sampling the lower visual field (Applebury et al., 2000; Baden et al., 2013; Szél et al., 1992). Therefore, this particular retinal organization is optimal if the upper visual-field images had lower contrast in UV than in green, and the lower visual-field images had higher contrast in UV than in green. Our image analysis, however, showed the opposite trend in the "mouse-view" visual scenes (Figure 3).

Power spectrum and autocorrelation of UV and green images

We next analyzed the second-order statistics of the acquired images. Specifically, we computed the power spectrum (Figure 4) and spatial autocorrelation (Figure 5) that describes the relationship of the two pixel intensity values as a function of their relative locations in the images (see Methods for details). As expected (Hyvärinen et al., 2009; Simoncelli and Olshausen, 2001), the power spectra generally followed $1/\omega^a$ on the spatial frequency ω for both UV and green channels irrespective of the camera angles (in log-log axes; Figure 4A–H); and were higher for the vertical direction than for the horizontal direction (Figure 4A–H)—i.e., the spatial autocorrelation was elongated in the vertical direction (Figure 5A–D).

There are, however, several distinct properties between the UV and green channels for the upper and lower visual-field images. First, the slope of the power spectra a was larger for the

lower visual-field images than for the upper visual-field images (Figure 4I–L); equivalently, the spatial autocorrelation was narrower for the lower visual-field images (Figure 5E–H), indicating the presence of more fine textures in those images. Second, for the upper visual-field images, the UV power spectra were higher than the green ones for both vertical and horizontal directions (e.g., the Y-intercept b showing the log-power at the spatial frequency of 1 cycle/pixel; Figure 4M,N). In contrast, for the lower visual-field images, the UV power spectra were lower with a larger slope than the green counterparts (Figure 4K,L,O,P). Equivalently, the spatial autocorrelation was wider and narrower in UV than in green for the upper and lower visual-field images, respectively (Figure 5E–H).

Under an efficient coding hypothesis, a higher spatial autocorrelation implies that less cones are needed to faithfully encode the scenes (Attneave, 1954; Barlow, 2004; Laughlin, 1981). One would then expect from the "mouse-view" image statistics that the functional S- and M-cones should be denser on the dorsal and ventral parts of the mouse retina, respectively, to achieve an optimal sampling; however, the opposite is the case with the mouse retina (Applebury et al., 2000; Baden et al., 2013; Szél et al., 1992), suggesting that the cone distribution bias in the mouse retina cannot be simply explained by the optimality principle from an information theoretic viewpoint.

Discussion

To study the natural image statistics for the mouse vision, here we collected a set of 168 "mouse-view" two-color images of various natural scenes across different seasons using a custom-made multi-spectral camera (Figures 1 and 2). We identified distinct properties in the first- and second-order image statistics for the two channels between the images above and below the horizon (Figures 3–5). Specifically, both the local contrast and the spatial autocorrelation were higher in UV than in green for the upper visual-field images, while they were both lower in UV than in green for the lower visual-field images. This is in disagreement with what the efficient coding hypothesis implies (Attneave, 1954; Barlow, 2004) from the functional division of the mouse retina along the dorsoventral axis (Applebury et al., 2000; Baden et al., 2013; Szél et al., 1992).

We thus suggest that the given retinal organization in mice should have evolved not only to efficiently encode natural scenes from an information theoretic perspective, but likely to meet some other ethological demands in their specific visual environments.

"Mouse-view" natural image database

We employed a beam-splitting strategy to simultaneously acquire UV and green images of the same scenes (Figure 1) because it has certain advantages over other hyper- or multi-spectral imaging techniques (Grahn and Geladi, 2007; Hagen and Kudenov, 2013). First, a previous study used a hyperspectral scanning technique where a full spectrum of each point in space was measured by a spectrometer (Baden et al., 2013). While the photoreceptor response could be better estimated by using its absorption spectra, the scanned images through a pinhole aperture inevitably had lower spatial and temporal resolutions than the snapshot images acquired with our device. Second, a camera array can be used for multi-spectral imaging with each camera equipped with appropriate filters and lenses (Genser et al., 2020). This is easy to implement and will perform well for distant objects; however, because angular disparity becomes larger for objects at a shorter distance, one would have a difficulty in taking close-up images that small animals such as mice would normally encounter in their everyday lives. Finally, our single-lens-two-camera design is simple and cost-effective compared to other snapshot spectral imaging methods (Hagen and Kudenov, 2013). In particular, commercially available devices are often expensive and inflexible, hence not suitable for our application to collect images that spectrally match the mouse vision.

There are several conceivable directions to expand the "mouse-view" natural image database. First, we could take high dynamic range images using a series of different exposure times. This works only for static objects, but can be useful to collect images at night during which nocturnal animals such as mice are most active. Second, we could take a movie to analyze the space-time statistics of natural scenes. It would be interesting to miniaturize the device and mount it on an animal's head to collect time-lapse images with more natural self-motion dynamics (Meyer et al., 2018; Sattler and Wehr, 2021). Expanding our "mouse-view" natural image datasets will be critical to better understand the visual environment of mice and

develop a theoretical explanation on species-specific and non-specific properties of the mouse visual system.

Optimality of the mouse retina

What selective pressures have driven the mouse retina to favour UV sensitivity over blue and evolve the dorsovental gradient in the opsin expression? Our image analysis suggests that the coding efficiency alone with respect to the natural image statistics cannot fully explain the peculiar organization of the mouse retina (Figures 3–5). Thus we will likely need to consider more ethological demands that are directly relevant for survival and reproduction. For example, the mouse urine reflects UV very well, and this has been suggested to serve as a con-specific visual cue for their territories and trails besides an olfactory cue (Chàvez et al., 2003). The UV sensitivity can also be advantageous for the hunting behaviour of mice because many nocturnal insects are attracted to UV light. Furthermore, increased UV sensitivity in the ventral retina may improve the detection of tiny dark spots in the sky, such as aerial predators (Yilmaz and Meister, 2013). Indeed, the S-opsin-dominant cones in mice have higher sensitivity to dark contrasts than the M-opsin-dominant ones (Baden et al., 2013), and turning the anatomical M-cones into the functional S-cone by co-expressing the S-opsin will dramatically increase the spatial resolution in the UV channel because the mouse retina has only a small fraction of the uniformly distributed genuine S-cones ($\sim$5 %) compared to the co-expressing cones ($\sim$95 %; Applebury et al., 2000; Haverkamp et al., 2005; Nikonov et al., 2006; Röhlich et al., 1994).

These arguments, however, are difficult to generalize because each species has presumably taken its own strategy to increase the fitness in its natural habitat, leading to convergent and divergent evolution. On the one hand, UV sensitivity was identified in some mammals that live in a different visual environment than mice, including diurnal small animals such as the degu and gerbil (Chàvez et al., 2003; Govardovskii et al., 1992; Jacobs et al., 2003) and even large animals such as the Arctic reindeer (Hogg et al., 2003). On the other hand, some species showing a similar behavioural pattern as mice do not have the dorsoventral division of the retinal function (Cronin and Bok, 2016; Hunt et al., 2001; Peichl, 2005). For example, even within the genus *Mus*, some species do not have the dorsoventral gradient of the S-opsin expression,

and others completely lack the S-cones (Szél et al., 1996). It is even possible that the cone distribution bias may have nothing to do with the perception of the color vision, but may arise just because of the developmental processes. Indeed, the center of the human fovea is generally devoid of S-cones (Curcio et al., 1991; Wald, 1967), and there is a huge diversity in the ratio of M- and L-cones in the human retina across subjects with normal color vision (Hofer et al., 2005; Roorda and Williams, 1999). Behavioural tests across species will then be critical for validating the ethological arguments to better understand the structure and function of the visual system (Baden et al., 2020). We expect that the "mouse-view" natural image datasets will contribute to designing such studies.

Figure legends

Figure 1: Multi-spectral camera system for the mouse vision

(A) Schematic diagram of the camera optics. Incoming light was split into UV and Green channels by a dichroic mirror and further filtered to match the spectral sensitivity of the mouse visual system (see panel B). A neutral density filter with the optical density value from 1.0 to 2.0 was used for the Green channel to maximize the dynamic range of the camera sensor to be used with the same parameter settings as the UV channel.

(B) Relative spectral sensitivity of the camera system (UV channel, violet area; Green channel, green area). For comparison, the spectral sensitivity of the mouse S- and M-cone photoreceptors (Govardovskii et al., 2000) corrected with the transmission spectrum of the mouse eye optics (Henriksson et al., 2010) was shown in violet and green, respectively, as well as the sunlight spectrum in gray.

Figure 2: Example images of the natural scenes in UV and green channels

(A) Upper visual field images taken with positive camera elevation angles (UV, Green, and merged images from left to right). These images typically contain trees and branches with sky backgrounds.

(B) Lower visual field images taken with negative camera elevation angles, often containing a closer look of grasses and flowers.

Figure 3: Light intensity and local contrast distributions of the "mouse-view" natural images

(A,B) Normalized light intensity distributions of the upper (A) and lower (B) visual0field images for UV (violet) and Green (green) channels (median and interquartile range).

(C,D) Local contrast distributions computed with the Laplacian-of-Gaussian filter (σ=10 in Eq.(2); see Methods for details). The distribution of the UV channel is more strongly heavy-tailed than that of the Green channel for the upper visual field images (C) but the Green channel's distribution is wider than the UV channel's for the lower visual-field images (D).

(E–H) Scale (β; E, F) and shape (γ; G, H) parameters from the Weibull distribution fitted to each image (Eq.(3); see Methods for details). For the upper field images (E,G), the UV channel has significantly larger β (E) and smaller γ (G) values than the green channel. In contrast, for the lower field images (F,H), the Green channel has significantly larger β (F) but comparable γ (H) values than the UV channel. P-values are obtained from sign-tests.

Figure 4: Power spectrum of the "mouse-view" natural images

(A–D) The average power spectra of the upper (A,B) and lower (C,D) visual-field images for the UV (A,C) and Green (B,D) channels.

(E–H) The power spectra in the vertical (E,G) and horizontal (F,H) directions (median and interquartile range) for the upper (E,F) and lower (G,H) visual-field images.

(I–M) The slope (a; I–L) and Y-intercept (b; M–P) parameters of the power function b/ω^a in the log-log space fitted to the power spectra of each image in the vertical (I,K,M,O) and horizontal (J,L,N,P) directions. For the upper visual-field images (I,J,M,N), the UV channel has significantly larger b (M,N) but comparable a (I,j) values than the Green channel. For the lower field images (K,L,O,P), in contrast, the Green channel has significant larger b (O,P) and smaller a (K,L) values than the UV channel. P-values are obtained from sign-tests.

Figure 5: Spatial autocorrelation of the "mouse-view" natural images

(A–D) The average spatial autocorrelation of the upper (A,B) and lower (C,D) visual-field images for the UV (A,C) and Green (B,D) channels, respectively.

(E–H) The spatial autocorrelation in the vertical (E,G) and horizontal (F,H) directions (median and interquartile range). The UV channel has a higher and wider spatial correlation for the upper visual-field images (E,F), while the Green channel has a higher and wider spatial correlation for the lower visual-field images (G,H).

(I–L) Representative spatial correlation values of the pixels horizontally (I,K) or vertically (J,L) separated by 50 pixels for the upper (I,J) and lower (K,L) visual-field images. P-values were obtained from sign-tests.

References

Applebury, M. L., Antoch, M. P., Baxter, L. C., Chun, L. L., Falk, J. D., Farhangfar, F., Kage, K., Krzystolik, M. G., Lyass, L. A., and Robbins, J. T. (2000). Cone photoreceptor: a single cone type expresses both S and M opsins with retinal spatial patterning. Neuron, 27, 513–523.

Attneave, F. (1954). Some informational aspects of visual perception. Psychol Rev, 61, 183–193.

Baden, T., Euler, T., and Berens, P. (2020). Understanding the retinal basis of vision across species. Nat Rev Neurosci, 21, 5–20.

Baden, T., Schubert, T., Chang, L., Wei, T., Zaichuk, M., Wissinger, B., and Euler., T. (2013). A tale of two retinal domains: Near-optimal sampling of achromatic contrasts in natural scenes through asymmetric photoreceptor distribution. Neuron, 80, 1206–1217.

Barlow, H. B. (2004). Possible principles underlying the transformation of sensory messages. In Sensory Communication, W. A. Rosenblith, ed. (MIT Press), pp. 217–234.

Burton, G. J. and Moorhead, I. R. (1987). Color and spatial structure in natural scenes. Appl Opt, 26, 157–170.

Chàvez, A. E., Bozinovic, F., Peichl, L., and Palacios, A. G. (2003). Retinal spectral sensitivity, fur coloration and urine reflectance in the genus *Octodon* (rodentia): implications for visual ecology. Invest Ophthalmol Vis Sci, 44, 2290–2296.

Cronin, T. W. and Bok, M. J. (2016). Photoreception and vision in the ultraviolet. J Exp Biol, 219, 2790–2801.

Curcio, C. A., Allen, K. A., Sloan, K. R., Lerea, C. L., Hurley, J. B., Klock, I. B., and Milam, A. H. (1991). Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. J Comp Neurol, 312, 610–624.

Enroth-Cugell, C. and Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. J Physiol, 187, 517–552.

Geisler, W. S. and Perry, J. S. (2011). Statistics for optimal point prediction in natural images. J Vis, 11, 14.

Genser, N., Seiler, J., and Kaup, A. (2020). Camera array for multi-spectral imaging. IEEE Trans Image Process, 29, 9234–9249.

Geusebroek, J. M. and Smeulders, A. W. M. (2005). A six-stimulus theory for stochastic texture. Int J Comput Vis, 62, 7–16.

Ghosh, K., Sarkar, S., and Bhaumik, K. (2007). Understanding image structure from a new multi-scale representation of higher order derivative filters. Image Vis Comput, 25, 1228–1238.

Gijsenij, A. and Gevers, T. (2011). Color constancy using natural image statistics and scene semantics. IEEE Trans Pattern Anal Mach Intell, 33, 687–698.

Gjorgjieva, J., Sompolinsky, H., and Meister, M. (2014). Benefits of pathway splitting in sensory coding. J. Neurosci, 34, 12127–12144.

Gollisch, T. and Meister, M. (2010). Eye smarter than scientists believed: Neural computations in circuits of the retina. Neuron, 65, 150–164.

Govardovskii, V. I., Fyhrquist, N., Reuter, T., Kuzmin, D. G., and Donner, K. (2000). In search of the visual pigment template. Vis Neurosci, 17, 509–528.

Govardovskii, V. I., Röhlich, P., Szél, A., and Khokhlova, T. V. (1992). Cones in the retina of the mongolian gerbil, *Meriones unguiculatus*: an immunocytochemical and electrophysiological study. Vision Research, 32, 19–27.

Grahn, H. and Geladi, P. (2007). Techniques and Applications of Hyperspectral Image Analysis. (John Wiley and Sons).

Hagen, N. and Kudenov, M. W. (2013). Review of snapshot spectral imaging technologies. Opt Eng, 52, 090901.

Hateren, J. H. v. and Schaaf, A. v. d. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. Proc Biol Sci, 265, 359–366.

Haverkamp, S., Wässle, H., Duebel, J., Kuner, T., Augustine, G. J., Feng, G., and Euler, T. (2005). The primordial, blue-cone color system of the mouse retina. J Neurosci, 25, 5438–5445.

Henriksson, J. T., Bergmanson, J. P. G., and Walsh, J. E. (2010). Ultraviolet radiation transmittance of the mouse eye and its individual media components. Exp Eye Res, 90, 382–387.

Hofer, H., Carroll, J., Neitz, J., Neitz, M., and Williams, D. R. (2005). Organization of the human trichromatic cone mosaic. J Neurosci, 25, 9669–9679.

Hogg, C., Neveu, M., Stokkan, K.-A., Folkow, L., Cottrill, P., Douglas, R., Hunt, D. M., and Jeffery, G. (2003). Arctic reindeer extend their visual range into the ultraviolet. J Exp Biol, 214, 2014–2019.

Huberman, A. D. and Niell, C. M. (2011). What can mice tell us about how vision works? Trends Neurosci, 34, 464–473.

Hunt, D. M., Wilkie, S. E., Bowmaker, J. K., and Poopalasundaram, S. (2001). Vision in the ultraviolet. Cell Mol Life Sci, 58, 1583–1598.

Hyvärinen, A., Hurri, J., and Hoyer, P. O. (2009). Natural Image Statistics: A Probabilistic Approach to Early Computational Vision. (Springer-Verlag London).

Jacobs, G. H., Calderone, J. B., Fenwick, J. A., Krogh, K., and Williams, G. A. (2003). Visual adaptations in a diurnal rodent, *Octodon degus*. J Comp Physiol A, 189, 347–361.

Jacobs, G. H., Neitz, J., and Deegan II, J. F. (1991). Retinal receptors in rodents maximally sensitive to ultraviolet light. Nature, 353, 655–656.

Jacobs, G. H. and Williams, G. A. (2007). Contributions of the mouse UV photopigment to the ERG and to vision. Doc Ophthalmol, 115, 137–144.

Jacobs, G. H., Williams, G. A., and Fenwick, J. A. (2004). Influence of cone pigment coexpression on spectral sensitivity and color vision in the mouse. Vision Res, 44, 1615–1622.

Laughlin, S. (1981). A simple coding procedure enhances a neuron's information capacity. Z Naturforsch C Biosci, 36, 910–912.

Lindeberg, T. (2013). A computational theory of visual receptive fields. Biol Cybern, 107, 589–635.

Lyubarsky, A., Falsini, B., Pennesi, M., Valentini, P., and Pugh Jr., E. (1999). UV- and midwave-sensitive cone-driven retinal responses of the mouse: a possible phenotype for coexpression of cone photopigments. J Neurosci, 19, 442–455.

Marr, D. and Hildreth, E. (1980). Theory of edge detection. Proc R Soc Lond Ser B, 207, 187–217.

Meyer, A. F., Poort, J., O'Keefe, J., Sahani, M., and F.Linden, J. (2018). A head-mounted camera system integrates detailed behavioral monitoring with multichannel electrophysiology in freely moving mice. Neuron, 100, 46–60.e7.

Nikonov, S., Kholodenko, R., Lem, J., and Pugh Jr., E. (2006). Physiological features of the S-and M-cone photoreceptors of wild-type mice from single-cell recordings. J Gen Physiol, 127, 359–374.

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381, 607–609.

Peichl, L. (2005). Diversity of mammalian photoreceptor properties: adaptations to habitat and lifestyle? Anat Rec A Discov Mol Cell Evol Biol, 287, 1001–1012.

Rodieck, R. W. and Stone, J. (1965). Analysis of receptive fields of cat retinal ganlion cells. J Neurophysiol, 28, 833–849.

Röhlich, P., Van Veen, T., and Szél, A. (1994). Two different visual pigments in one retinal cone cell. Neuron, 13, 1159–1166.

Roorda, A. and Williams, D. R. (1999). The arrangement of the three cone classes in the living human eye. Nature, 397, 520–522.

Sattler, N. J. and Wehr, M. (2021). A head-mounted multi-camera system for electrophysiology and behavior in freely-moving mice. Front Neurosci, 24, 592417.

Shapley, R. and Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. Prog Retinal Res, 3, 263–346.

Simoncelli, E. P. and Olshausen, B. A. (2001). Natural image statistics and neural representation. Annu Rev Neurosci, 24, 1193–1216.

Smith, E. and Lewicki, M. (2006). Efficient auditory coding. Nature, 439, 978–982.

Szél, A., Röhlich, P., Caffé, A. R., Juliusson, B., Aguirre, G., and Van Veen, T. (1992). Unique topographic separation of two spectral classes of cones in the mouse retina. J Comp Neurol, 325, 327–342.

Szél, A., Röhlich, P., Caffé, A. R., and Van Veen, T. (1996). Distribution of cone photoreceptors in the mammalian retina. Microsc Res Techn, 35, 445–462.

Tkačik, G., Garrigan, P., Ratliff, C., Milčinski, G., Klein, J. M., Seyfarth, L. H., Sterling, P., Brainard, D. H., and Balasubramanian, V. (2011). Natural images from the birthplace of the human eye. PLOS ONE, 6, 1–12.

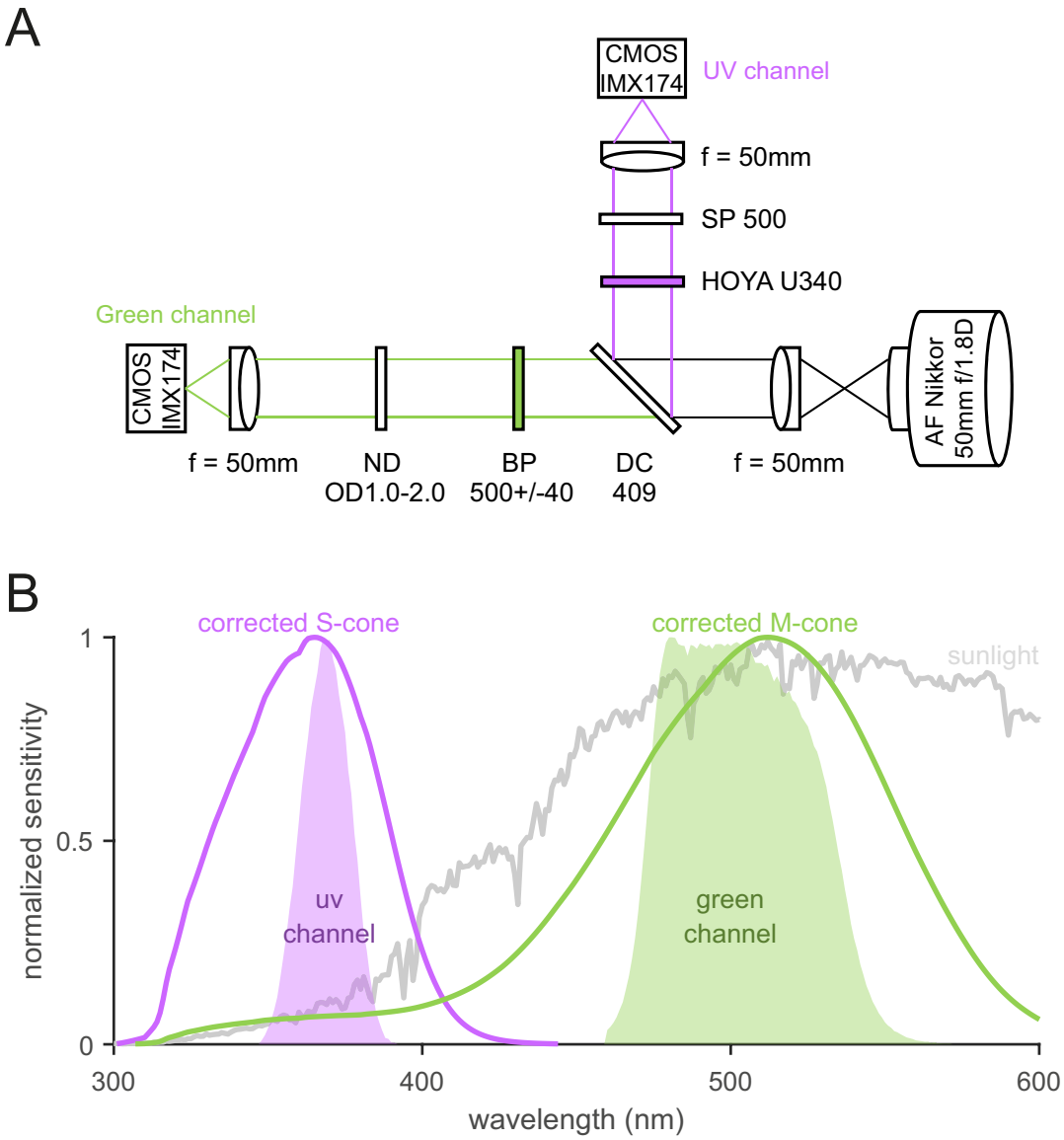
Tolhurst, D. J., Tadmor, Y., and Tang, C. (1992). The amplitude spectra of natural images. Ophthalmic Physiol Opt, 12, 229–232.

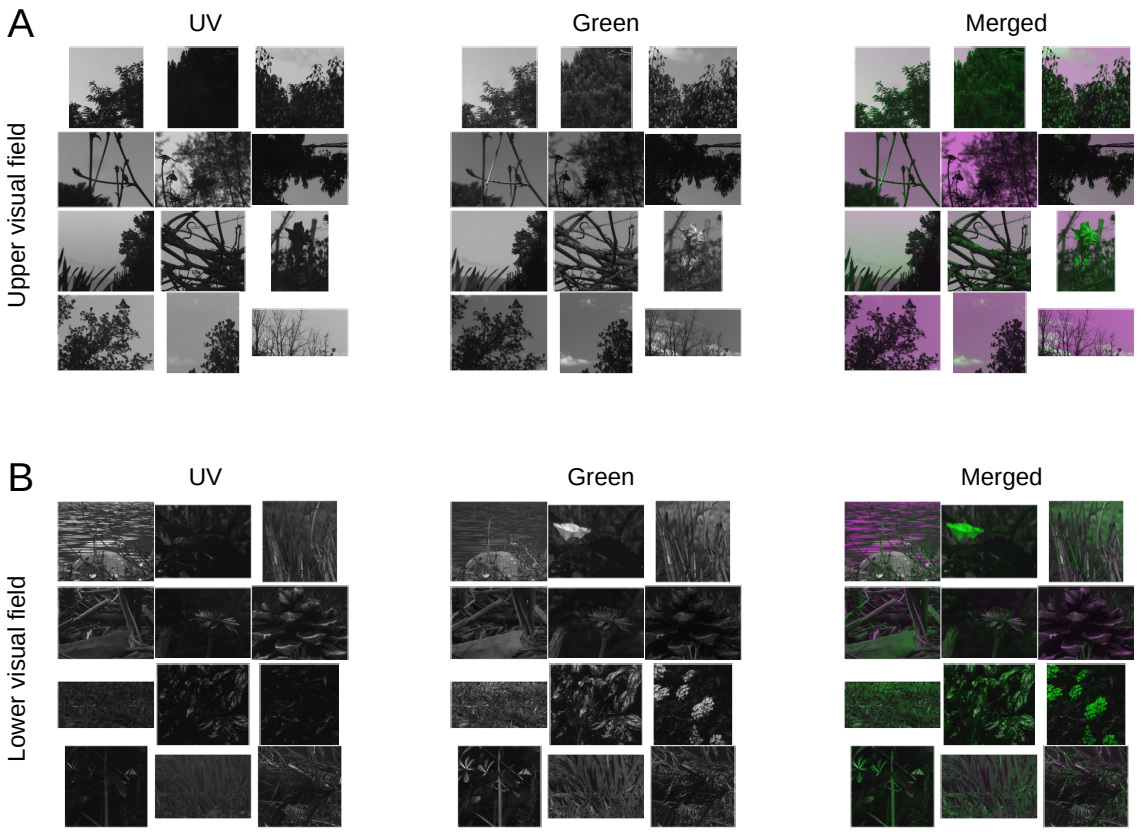
Wald, G. (1967). Blue-blindness in the normal fovea. J Opt Soc Am, 57, 1289–1301.

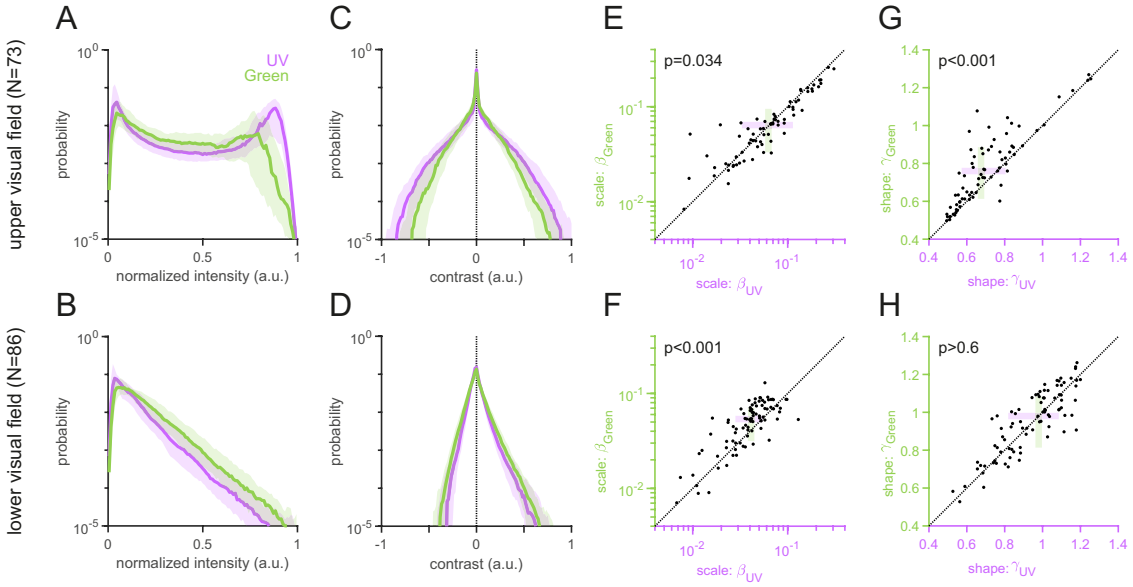
Wang, Y. V., Weick, M., and Demb, J. B. (2011). Spectral and temporal sensitivity of cone-mediated responses in mouse retinal ganglion cells. J Neurosci, 31, 7670–7681.

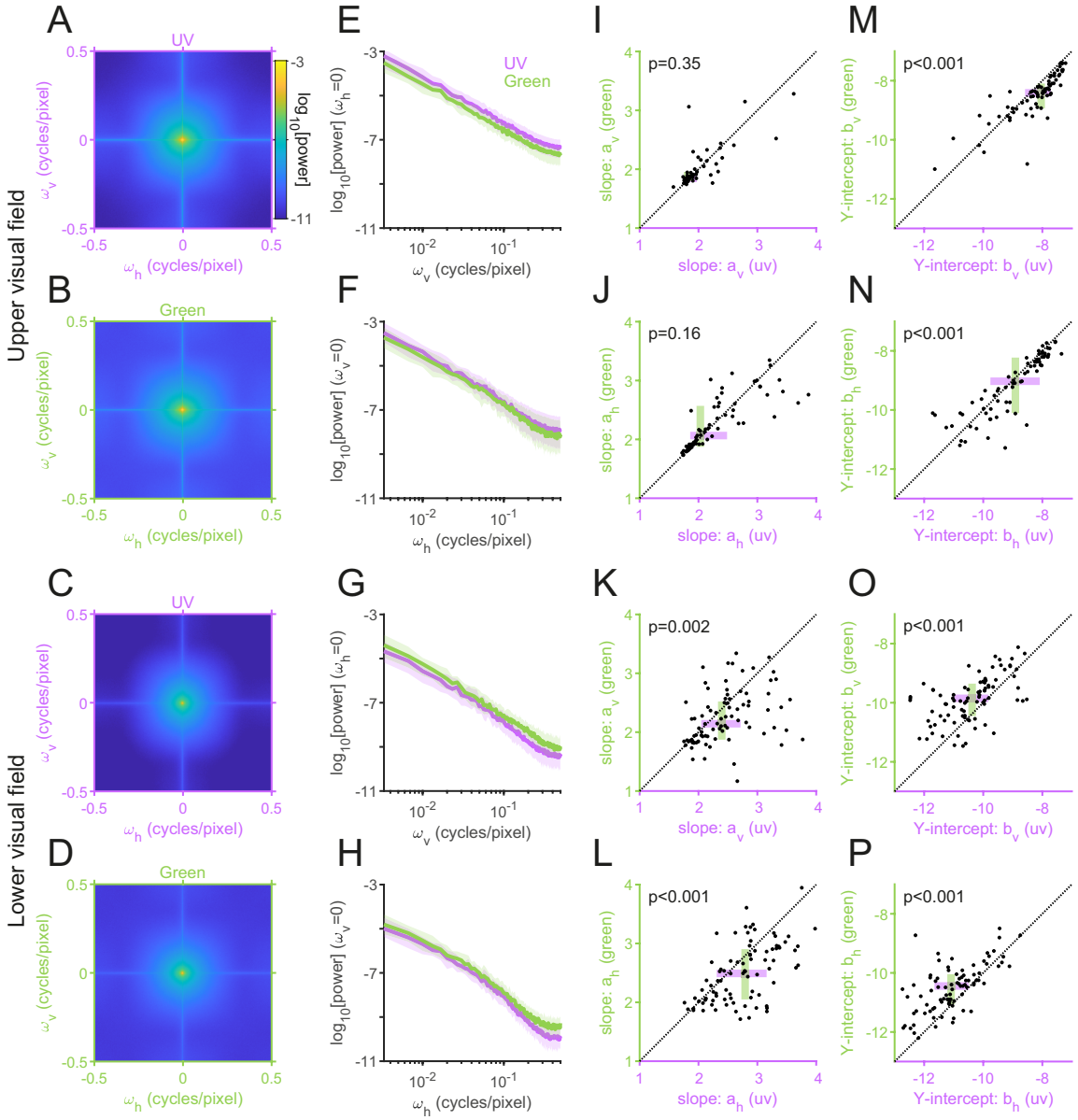
Yilmaz, M. and Meister, M. (2013). Rapid innate defensive responses of mice to looming visual stimuli. Curr Biol, 23, 2011–2015.

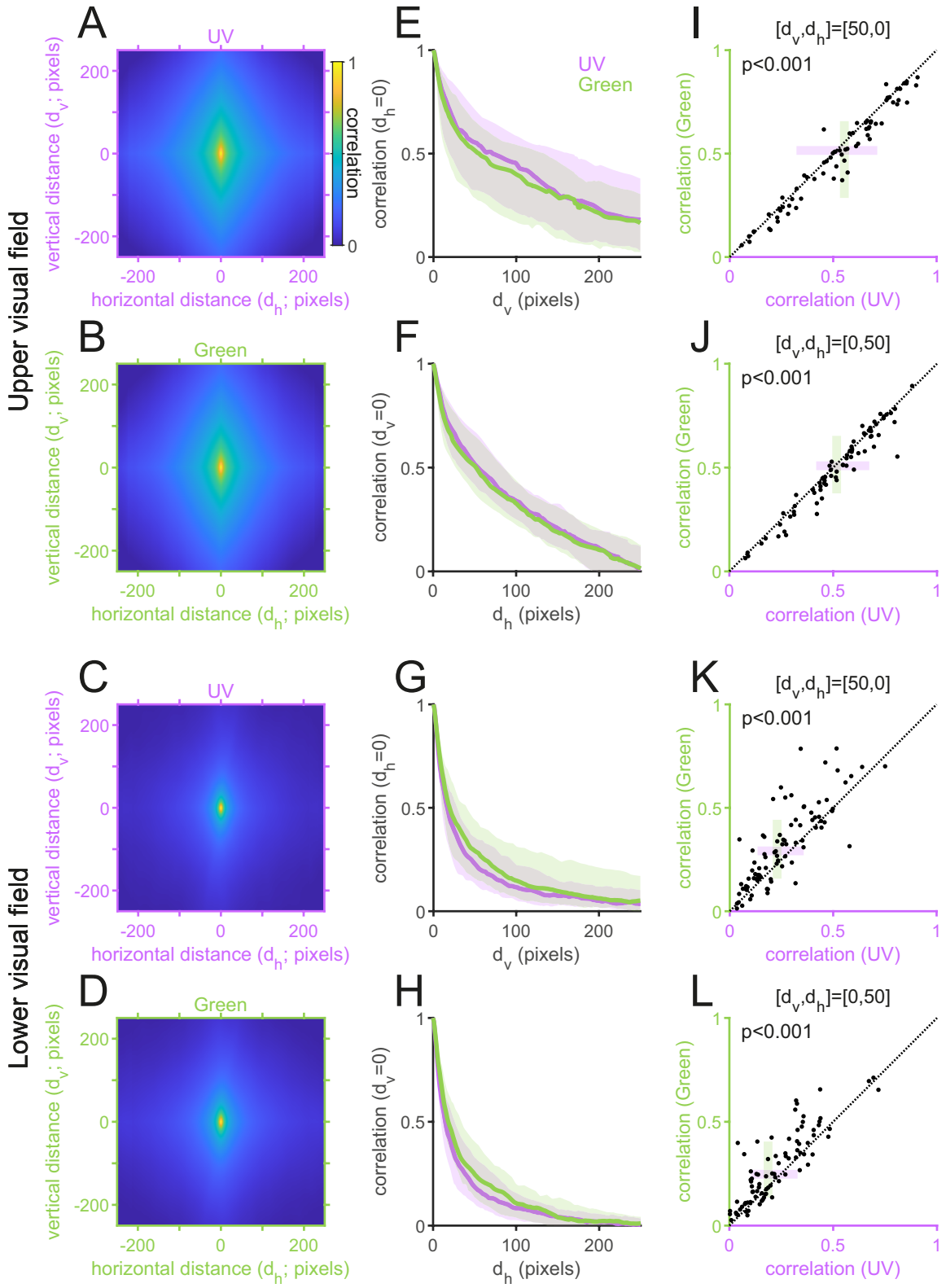
Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., and Oliva, A. (2014). Learning deep features for scene recognition using places database. In Advances in Neural Information Processing Systems, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, eds., vol. 27. (Curran Associates, Inc.).


A
CMOS
IMX174
UV channel
f = 50mm
SP 500
HOYA U340
Green channel
CMOS
IMX174
f = 50mm
ND
OD1.0-2.0
BP
500+/-40
DC
409
f = 50mm
AF Nikkor
50mm f/1.8D
B
corrected S-cone
corrected M-cone
sunlight
uv
channel
green
channel
normalized sensitivity
1
0.5
0
300
400
500
600
wavelength (nm)


A
UV
Green
Merged
Upper visual field
B
UV
Green
Merged
Lower visual field


A
B
C
D
E
F
G
H
upper visual field (N=73)
lower visual field (N=86)
UV
Green
probability
normalized intensity (a.u.)
contrast (a.u.)
p=0.034
p<0.001
p<0.001
p>0.6
scale: β_{Green}
scale: β_{UV}
shape: γ_{Green}
shape: γ_{UV}


A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Upper visual field
Lower visual field
UV
Green
log10[power]
ωv (cycles/pixel)
ωh (cycles/pixel)
log10[power] (ωh=0)
log10[power] (ωv=0)
p=0.35
p=0.16
p=0.002
p<0.001
slope: av (green)
slope: av (uv)
slope: ah (green)
slope: ah (uv)
Y-intercept: bv (green)
Y-intercept: bv (uv)
Y-intercept: bh (green)
Y-intercept: bh (uv)


Upper visual field
Lower visual field
A
UV
B
Green
C
UV
D
Green
vertical distance (d_v; pixels)
horizontal distance (d_h; pixels)
correlation
E
F
G
H
UV
Green
correlation (d_h=0)
correlation (d_v=0)
d_v (pixels)
d_h (pixels)
I
J
K
L
$[d_v,d_h]$=[50,0]
$[d_v,d_h]$=[0,50]
p<0.001
correlation (Green)
correlation (UV)