Chapter 9

# Vision as a Fundamentally Statistical Machine

Zhiyong Yang

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/50165

## 1. Introduction

As a vital driving force of systems neuroscience, visual neuroscience had its conceptual framework established more than 40 years ago based on Hubel and Wiesel's groundbreaking work on the receptive-field properties of visual neurons (Hubel & Wiesel, 1977). This framework was subsequently strengthened by David Marr's influential book (Marr, 2010). In this paradigm, visual neurons are conceived to perform bottom-up, image-based processing to build a series of symbolic representations of visual stimuli. This paradigm, however, is deeply misleading since the generative sources in the three-dimensional (3**D**) physical world of any stimulus, to which visual animals must respond successfully, cannot be determined by image-based processing (due to the inverse optics problem). This is perhaps the reason why "Now, thirty years later, the main problems that occupied Marr remain fundamental open problems in the study of perception" (Marr, 2010), as assessed by two prominent vision scientists and Marr's close associates.

During the last 30 years, dramatic progress in computing hardware, digital imaging, statistical modeling, and visual neuroscience has promoted researchers to re-examine the computations and representations (see above) for natural vision examined in Marr's book. A range of new ideas have been proposed, many of which are summarized in books (Knill & Richards, 1996; Rao et al., 2002; Purves & Lotto, 2003; Doya et al., 2007; Trommershauser et al., 2011) and reviews (Simoncelli & Olshausen, 2001; Yuille & Kersten, 2006; Geisler, 2008; Friston, 2010). The unified theme is that vision and visual system structure and function must be understood in statistical terms. How this feat can be achieved, however, is not clear at all.

Since humans and other visual animals must respond successfully to visual stimuli whose generative sources cannot be determined in any direct way, the visual system can only generate percepts according to the probability distributions (**PDs**) of visual variables underlying the stimuli. The information pertinent to the generation of these PDs, namely, the statistics of natural visual environments, must have been incorporated into the visual circuitry by successful behavior in the world over evolutionary and developmental time.

During the last two decades, this statistical concept of vision has been successful in explaining aspects of vision that would be difficult to understand otherwise (see references cited above). In this chapter, I will describe several recent studies that relate the statistics of 2D and 3D natural visual scenes to visual percepts of brightness, saliency, and 3D space.

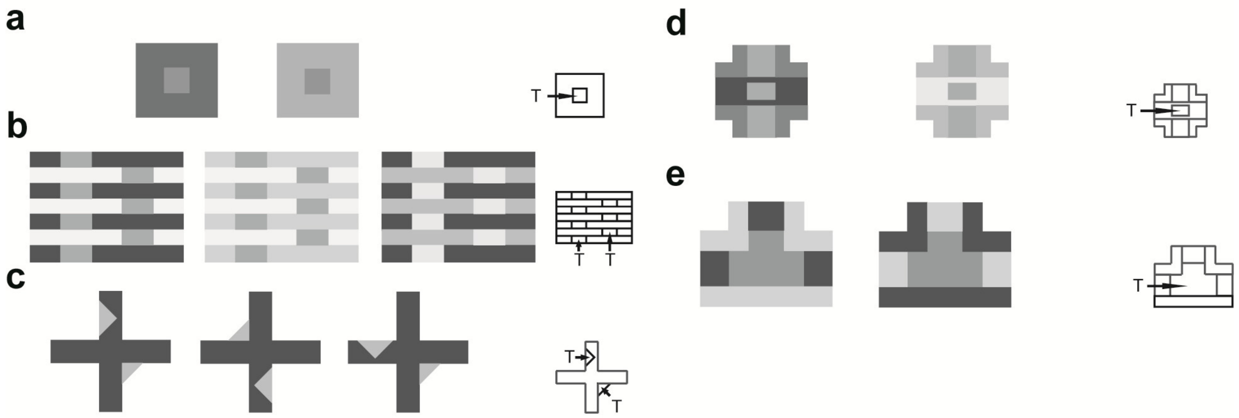
In the second section of this chapter, I will discuss how the PDs of luminance in specific contexts in natural scenes, referred to as the context-mediated PDs in natural scenes, predict brightness, the perception elicited by the luminance of a visual target. Our results show that brightness generated on this statistical basis accounts for a range of observations, whose causes have been debated for a long time without consensus. In the third section, I will present a simple, elegant model of the context- mediated PDs in natural scenes and a measure of visual saliency derived from these PDs. Our results show that this measure of visual saliency is a good predictor of human gaze in free-viewing both static and dynamic natural scenes. In the fourth section, I will present the statistics of 3D natural scenes and their relationship to human visual space. Our results show that human visual space is not a direct mapping of the 3D physical space but rather generated probabilistically. Finally, I will discuss the implications of these and other results for our understanding of the response properties of visual neurons, the intricate visual circuitries, the large-scale cortical organizations, the operational dynamics of the visual system, and natural vision.

## 2. The statistical structure of natural light patterns determines perceived light intensity

### 2.1. Introduction

In this section, I present evidence that the context-mediated PDs of luminance in natural scenes predict brightness, the perception elicited by the luminance of a visual target. A central puzzle in understanding how such percepts are generated by the visual system is that brightness does not correspond in any simple way to luminance. Thus, the same amount of light arising from a given region in a scene can elicit dramatically different brightness percepts when presented in different contexts (Fig. 1) (Kingdom, 2011). For example, in Fig.1 (a), the central square (T) in the left panel appears brighter than the same target in the right panel. This is the standard simultaneous brightness contrast effect.

A variety of explanations have been suggested since the basis for such phenomena was first debated by Helmholtz, Hering, Mach, and others (Gichrist et al., 1999; Purves et al., 2004; Kingdom, 2011). Although lateral inhibition in the early visual processing has often been proposed to account for these “illusions”, this mechanism cannot explain instances in which similar overall contexts produce different brightness effects (compare Fig. 1 (a) with Figs. 1 (b) and (e); see also Fig. 1 (c)). This failure has led to several more recent suggestions, including complex filtering (Blakeslee & McCourt, 2004), the idea that brightness depends on detecting edges and junctions that promote the grouping of various luminances into interpretable spatial arrangements (Adelson, 2000; Anderson & Winawer, 2005), and the proposal that brightness is “re-synthesized” from 3D scene properties “inferred” from the stimulus (Wishart et al., 1997).



(a), Standard simultaneous brightness contrast effect. The central square in the dark surround (left panel) appears brighter than the equiluminant square in the light surround (right panel). (b), White's illusion. Although the gray rectangles in the left panel are all equiluminant, the ones surrounded by the generally lighter context on the left appear brighter than those surrounded by the generally darker context on the right. When, however, the luminance of the target rectangles is the lowest (middle panel) or highest (right panel) value in the stimulus, the targets in the generally lighter context (on the left in the middle and right panels) appear less brighter than ones in the generally darker context (called the "inverted White's effect"). (c), Wertheimer-Benary illusion. The triangle embedded in the arm of the black cross appears brighter than the one that abuts the corner of the cross. The slightly different brightness of the equiluminant triangles is maintained whether the presentation is upside down (middle panel), or reflected along the diagonal (right panel). (d), The intertwined cross illusion. The target on the left appears substantially brighter than the equiluminant target on the right. (e), The inverted T-illusion. The inverted T-shape on the left appears somewhat brighter than the equiluminant target on the right (modified from Yang & Purves, 2004).

**Figure 1.** The influence of spatial patterns of luminance on the apparent brightness of a target (the targets [T] in each stimulus are equiluminant, and are indicated in the insets on the right).

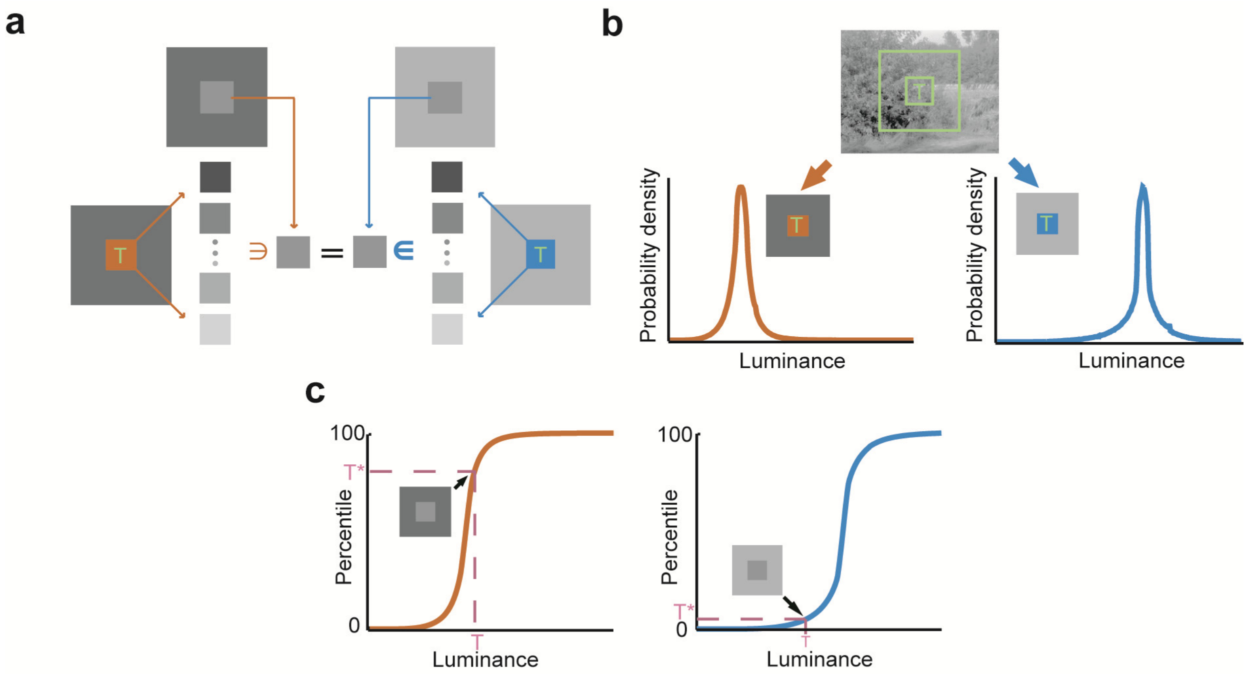
## 2.2. Context-mediated PDs of luminance in natural scenes

To examine whether the statistics of natural light patterns predict the perceptual phenomena shown in Fig. 1, we obtained the relevant PDs of luminance in natural scenes by sampling a database of natural scenes (van Hateren & van der Schaaf, 1998) with target-surround configurations that had the same local geometry as the stimuli in Fig. 1. As a first step, these configurations were superimposed on the images to find light patterns in which the luminance values of both the surround and target regions were approximately homogeneous; for those configurations in which the surround comprised more than one region of the same luminance (see Fig. 1), we also required that the relevant sampled regions meet this criterion. The sampling configurations were moved in steps of one pixel to screen the full image. The mean luminance values of the target and the surrounding regions in the samples were then calculated, and their occurrences tallied.

## 2.3. Brightness signifies context-mediated PDs of luminance in natural scenes

Natural environments comprise objects of different sizes at various distances that are related to each other and the observer in a variety of ways (Yang & Purves, 2003a,b). When the light arising from objects is projected onto an image plane, these complex relationships are transformed into 2D patterns of light intensity with highly structured statistics. Thus, the PD of the luminance of, say, the central target in a standard simultaneous brightness contrast stimulus (Fig. 2 (a)) depends on the surrounding luminance values (Fig. 2 (b)).

Fig. 2 (c) illustrates the supposition that, for any context, the visual system generates the brightness of a target according to the value of its luminance in the probability distribution function (PDF, the integral of PD) of the possible target luminance experienced in that context (Yang & Purves, 2004). This value is referred to subsequently as the percentile of the target luminance among all possible luminance values that co-occur with the contextual luminance pattern in the natural environment. In formal terms, this supposition means that the visual system generates brightness percepts according to the relationship Brightness=$A\Phi(P)+A_0$, where A and $A_0$ are constants, and $\Phi(P)$ is a monotonically increasing function of the PDF, P.



(a), The brightness elicited by a given target luminance in any context depends on the frequency of occurrence of that luminance relative to all the possible target luminance values experienced in that context in natural environments. This concept is illustrated here using the standard simultaneous brightness contrast stimulus in Fig. 1 (a). The series of squares with different luminance values indicate all the possible occurrences of luminance in the target (T) in the two different contexts; the symbol ∊ indicates the relationship of a particular occurrence of luminance to the all possible occurrences of target luminance values experienced in the two contexts in natural environments. (b), This statistical relationship can be derived from the PD of target luminance values co-occurring with the luminance values and pattern of the two contexts of interest. The red and blue curves indicate the PDs of the luminance of the targets in (a), obtained by sampling the natural image database. The size of the sampling configuration was 1°x1°. (c), The brightness elicited by the luminance of the targets in (a) is based on the percentile of that luminance in the PDFs for the two different contexts, which are indicated by the icons (modified from Yang & Purves, 2004).

**Figure 2.** Brightness percepts signify context-mediated PDs of luminance in natural scenes.

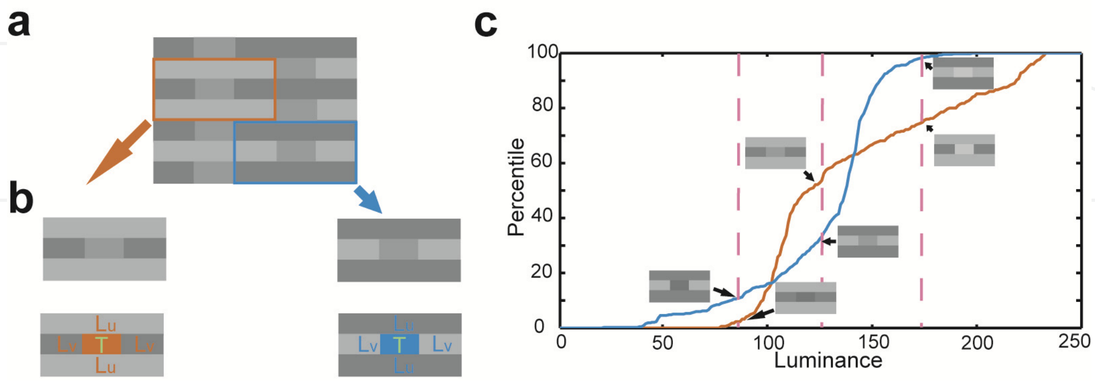
By definition, then, the percentile of target luminance for the lowest luminance value within any contextual light pattern is 0% and corresponds to the perception of maximum darkness; the percentile for the highest luminance within any contextual pattern is 100% and corresponds to the maximum perceivable brightness. In any given context, a higher luminance will always have a higher percentile, and will always elicit a perception of greater brightness compared to any luminance that has a lower percentile. Since the relation Brightness=AΦ(P)+$A_0$ is not based on a particular luminance within the context in question, but rather on the entire PD of possible luminance values experienced in that context, the context-dependent relationship between brightness and luminance is highly nonlinear (see Fig. 2 (c)). In consequence, the same physical difference between two luminance values will often signify different percentile differences, and thus perceived differences in brightness. Furthermore, because the percentiles change more rapidly as the target luminance approaches the luminance of the surround one would expect greater changes of brightness, an expectation that corresponds to the well known "crispening" effect in perception.

Finally, because the same value of target luminance will often correspond to different percentiles in the PDFs of target luminance in different contexts, two targets having the same luminance can elicit different brightness percepts, the higher percentile always corresponding to a brighter percept. Thus, in the standard simultaneous brightness contrast stimulus in Fig. 1 (a), the target (T) in the left panel in Fig. 2 (a) appears brighter than the equiluminant target in the right panel.

## 2.4. White's illusion

White's illusion (Fig. 1 (b)) presents a particular challenge for any explanation of brightness (White, 1979). The equiluminant rectangular areas surrounded by predominantly more luminant regions in the stimulus (on the left in the left panel of Fig. 1 (b); see also the area in the red frame in Fig. 3 (a) appear brighter than areas of identical luminance surrounded by less luminant regions (on the right in the left panel of Fig. 1 (b); see also the area in the blue frame in Fig. 3 (a). The especially perplexing characteristic of this percept is that the effect is opposite that elicited by standard simultaneous brightness contrast stimuli (Fig. 1 (a)). Even more puzzling, the effect reverses when the luminance of the rectangular targets is either the lowest or highest value in the stimulus (middle and right panels in Fig. 1 (b).

The explanation for White's illusion provided by the statistical framework outlined above is shown in Fig. 3. When presented separately, as in Fig. 3 (a), the components of White's stimulus elicit much the same effect as in the usual presentation. By sampling the images of natural visual environments using configurations based on these components (Fig. 3 (b)), we obtained the PDFs of the luminance of a rectangular target (T) embedded in the two different configurations of surrounding luminance in White's stimulus. As shown in Fig. 3 (c), when the target in the intermediate range of luminance values (i.e., in between the luminance values at the two crossover points) abuts two dark rectangles laterally (left panel in Fig. 3 (b)), the percentile of the target luminance (red line) is higher than the percentile when the target abuts the two light rectangles (right panel in Fig. 3(b); blue line in Fig. 3 (c)). If, as we suppose, the percentile in the PDF of target luminance within any specific context determines the brightness perceived, the target with an intermediate luminance on the left in Fig. 3 (b) should appear brighter than the equiluminant target on the right. Finally, when all the luminance values in the stimulus are limited to a very narrow range (e.g., from 0 to 100 cd/m$^2$ or from 1000 to 1100 cd/m$^2$), when the sampling configurations are orientated vertically, or when the aspect ratio of the sampling configurations is changed (e.g., from 1:2 to 1:5), the PDFs derived from the database are not much different. These further results are consistent with the observations that White's stimulus elicits the similar effect when presented at a wide range of overall luminance levels, in a vertical orientation, or with different aspect ratios.



(a), The usual presentation of White's illusion; boxed areas indicate the basic components of the stimulus, which when separated elicit about the same effect as the usual presentation. (b), The sampling configurations used to obtain the PDFs of target luminance (the red and blue rectangles), given a pattern of surrounds with luminance values $L_u$ and $L_v$ (size of the sampling configuration in this example was 0.6°[H]×0.3°[V] and the unit of luminance was $cd/m^2$). (c), The PDFs of the luminance of the targets in these contexts (red curve: $L_u$=145, $L_v$=105; blue curve: $L_u$=105, $L_v$=145). For the middle luminance values lying within the two crossover points (at ~105 and 145), the red curve is above the blue curve; as a result, the luminance configurations in (b) generate White's illusion (as indicated by the insets). For other luminance values of the target, the blue curve is above the red curve; as a result, the luminance configurations in (b*)* generate the inverted White's effect (modified from Yang & Purves, 2004).

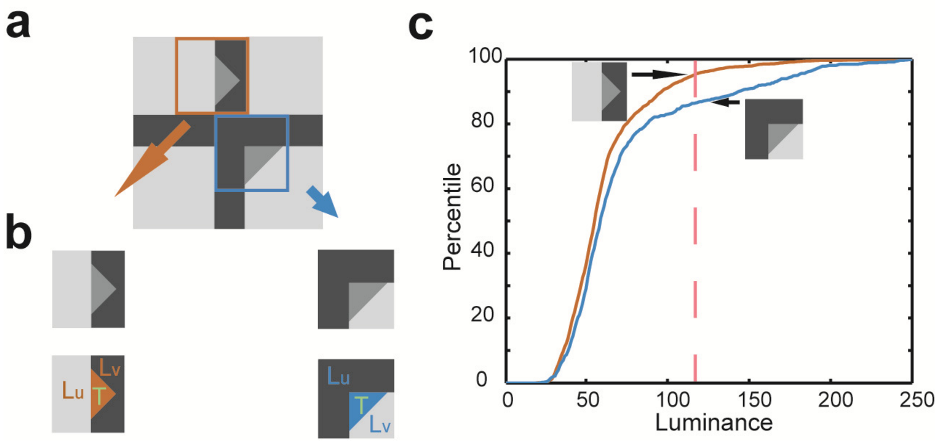
**Figure 3.** Statistical explanation of White's illusion.

An aspect of White's illusion that has been particularly difficult to explain is the so-called "inverted White's effect": when the target luminance is either the lowest or the highest value in the stimulus, the effect is actually opposite the usual percept (see the middle and right panels in Fig. 1 (b)). The explanation for this further anomaly is also evident in Fig. 3 (c). When the target luminance is the lowest value in the presentation (see insets), the blue curve is above the red curve. As a result, a relatively dark target surrounded by more light area should now appear darker, as it does (see also the middle panel of Fig. 1 (b)). By the same token, when the target luminance is the highest value in the stimulus (see insets), the blue curve is also above the red curve. Accordingly, the relatively light target surrounded by more dark area should appear brighter, as it does (see also the right panel of Fig. 1 (b)). Thus the statistical structure of natural light patterns predicts not only White's illusion, but the inverted White's effect as well. Notice further that the two crossover points of the blue and red curves shift to the right when the contextual luminances increase, and to the left when they decrease; thus the inverted effect will be apparent, although altered in magnitude, for any luminance values of the surrounding areas.

## 2.5. The Wertheimer-Benary illusion

In the Wertheimer-Benary illusion (Fig. 1(c)), the equiluminant gray triangles appear differently bright, the triangle embedded in the arm of the cross looking slightly brighter than the triangle in the corner of the cross.

The explanation of the Wertheimer-Benary illusion provided by the statistical framework outlined above is shown in Fig. 4. By sampling the images of natural environments using configurations based on the components of the stimulus (Fig. 4 (b)), we obtained the PDFs of target luminance in these contexts. As shown in Fig. 4 (c), when the triangular patch is embedded in a dark bar with its base facing a lighter area, the percentile of the luminance of the triangular patch (red line) is higher than the percentile when the triangular patch abuts a dark corner with its base facing a similar light background (blue line). Accordingly, the same gray patch should appear brighter in the former context than in the latter, as is the case. The PDFs obtained after changing the triangles to rectangles, rotating the configurations in Fig. 4 (b) by 180°, or reflecting the configurations along the diagonal of the cross (cf. middle and right panels in Fig. 1 (c)) were much the same as those shown in Fig. 4 (c). These several observations accord with the fact that the Wertheimer-Benary effect is little changed by such manipulations.



(a), The usual presentation of the Wertheimer-Benary stimulus. The separated components of the stimulus (boxed areas) elicit about the same effect as the usual presentation. (b), Configurations used to sample the database (size=0.4°×0.4°). (c), The PDFs of target luminance, given the surrounding luminances in (b) (The unit of luminance was cd/m$^2$). The red curve corresponds to the condition shown in the left panel in (b) ($L_u$=205, $L_v$=45), and the blue curve to the condition shown in the right panel in (b) ($L_u$=45, $L_v$=205) (modified from Yang & Purves, 2004).

**Figure 4.** Statistical explanation of the Wertheimer-Benary illusion.

The context-mediated PDs of luminance in natural scenes predict equally well other brightness phenomena shown in Fig. 1 (Yang & Purves, 2004).

## 2.6. Discussion

### *2.6.1. The statistical nature of perception*

I showed that brightness percepts do not encode luminance as such, but rather the statistical relationship between the luminance in an area within a particular contextual light pattern and all possible occurrences of luminance in the context that have been experienced by humans in natural environments during evolution. The statistical basis for this aspect of visual perception is quite different from traditional approaches to rationalizing brightness. In the "relational approach" (Gichrist et al., 1999), an idea that evolved from the late 19th C.

debate between Helmholtz, Hering, and others, brightness percepts are "recovered" by the visual system from explicitly coded luminance contrasts and gradients. Another idea is that brightness depends on intermediate-level visual processes that detect edges, gradients and junctions, which are then grouped into specific spatial layouts (Adelson, 2000; Anderson & Winawer, 2005). Finally, the brightness elicited by a given luminance has been also considered as being "re-synthesized" by processing at several levels of the visual system that is based on inferences about the possible arrangements of surfaces in 3D, their material properties and their illumination (Wishart et al., 1997).

The common deficiency of these several ways of thinking about brightness is their failure to relate the statistics of light patterns experienced in the course of evolution to what the corresponding brightness percepts need to signify (namely, the relationship of a particular occurrence of luminance to all possible occurrences of luminance in a given context). Since light patterns on the retina are the only information the visual system receives, basing brightness percepts on the statistics of natural light patterns allows visual animals to deal optimally with all possible natural occurrences of luminance, employing the full range of perceivable brightness to represent the physical world.

### *2.6.2. Neural instantiation of context-mediated PDs of luminance in natural scenes*

What sort of neural mechanisms, then, could incorporate these statistics of natural light patterns and relate them to brightness percepts? Although the answer is not known, the present results suggest that the circuitry at all levels of the visual system instantiates the statistical structures of light patterns in natural environments. In this conception, the center-surround organization of the receptive fields of retinal ganglion cells provides the initial basis for representing the necessary statistics. A further speculation would be that neural circuitry at the level of visual cortex is organized to instantiate the statistics of luminance patterns with arbitrary target and context shapes and sizes. As a result, the neuronal response at each location would signify the percentile of the target luminance in the PDF pertinent to a given context.

## 3. Visual saliency emerging from context-mediated PDs in natural scenes

### 3.1. Introduction

In this section, I present a simple model of the context-mediated PDs in natural scenes and derive a measure of visual saliency from these PDs. Visual saliency is the perceptual quality that makes some items in visual scenes stand out from their immediate contexts (Itti & Koch, 2001). Visual saliency plays important roles in natural vision in that saliency can direct eye movements and facilitate object detection and scene understanding. We developed a model of the context-mediated PDs in natural scenes using a modified algorithm for independent component analysis (**ICA**) (Hyvarinen, 1999) and demonstrated that visual saliency based on the context-mediated PDs in natural scenes is a good predictor of human gaze in free-viewing both static and dynamic natural scenes (Xu et al., 2010).

### 3.2. Context-mediated PDs in natural scenes and visual saliency

A visual feature is a random variable and co-occurs at certain probabilities with other visual features in natural scenes. We call these the context-mediated PDs in natural scenes. Here, a context refers to the natural scene patch that co-occurs with a visual target in question in space and/or time domains. We proposed to represent the context-mediated PDs in natural scenes using independent components (**ICs**) of natural scenes. There are two reasons for this. First, it has been argued extensively that the early visual cortex represents incoming stimuli in an efficient manner (Simoncelli & Olshausen, 2001). Second, the filters of the ICs of natural scenes are very much like the receptive fields of simple cells in V1 (van Hateren & van der Schaaf, 1998).

To model the context-mediated PDs in static natural scenes, we used a center-surround configuration in which the scene patch within the circular center serves as the target and the scene patch in the annular surround as the context (Xu et al., 2010). We sampled a large number of scene patches from the McGill calibrated color image database of natural scenes (Olmos & Kingdom, 2004). Thus, each sample is a pair of a patch in center ( $X_c$ ) and a patch in the surrounding area ( $X_s$ ) (Fig. 5 (a)). We developed a model of natural scenes in this configuration (Eq. (1)). In Eq. (1), $A_s$ , $A_c$ , and $A_{sc}$ are ICs. This model allows us to calculate the ICs for the context first and then the other ICs of natural scenes.

$$\begin{bmatrix} X_s \\ X_c \end{bmatrix} = \begin{bmatrix} A_s & 0 \\ A_{sc} & A_c \end{bmatrix} \begin{bmatrix} U_s \\ U_{sc} \end{bmatrix} \quad (1)$$

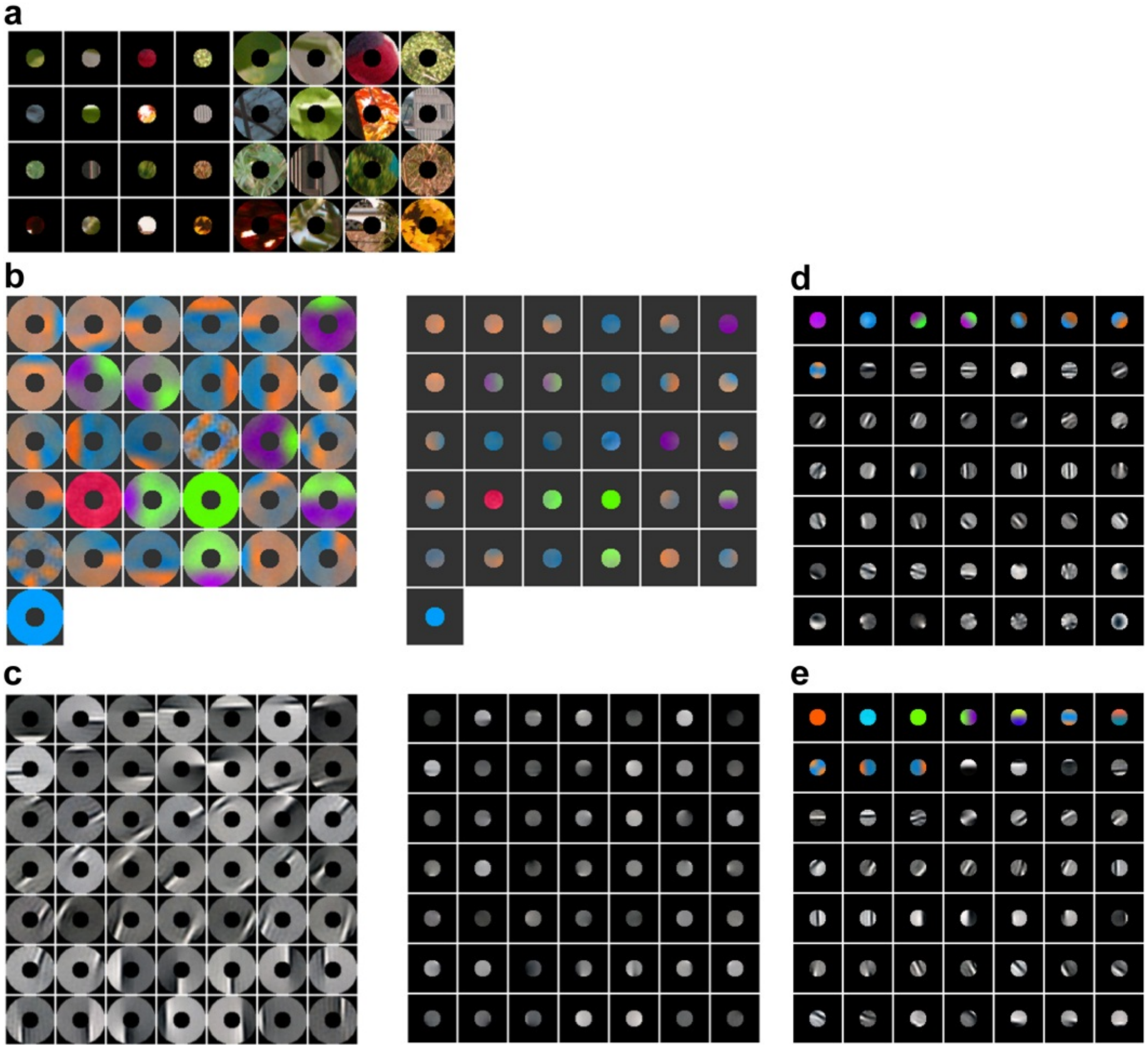
ICA filters (i.e., $W_s, W_{sc}, W_c$ ) can be obtained as follows

$$\begin{bmatrix} U_s \\ U_{sc} \end{bmatrix} = \begin{bmatrix} W_s & 0 \\ W_{sc} & W_c \end{bmatrix} \begin{bmatrix} X_s \\ X_c \end{bmatrix} \quad (2)$$

Therefore, we obtained three sets of ICs. First, the columns of $A_s$ are the ICs for $X_s$ . Second, the columns of $A_{sc}$ are the ICs for $X_c$ that are paired with the ICs for $X_s$ . Finally, the columns of $A_c$ are the ICs for $X_c$ that are not paired with any ICs for $X_s$ .

Fig. 5 (b) shows paired chromatic ICs for $X_c$ and $X_s$ . Fig. 5 (c) shows paired achromatic ICs for $X_c$ and $X_s$ . The chromatic ICs for the surround have red-green (L-M) or blue-yellow [S-(LM)] opponency. The chromatic paired ICs for the center are extensions of the ICs for the surround. Fig. 5 (d) shows the ICs for $X_c$ , including chromatic and achromatic ICs, that are not paired with any ICs for $X_s$ . Fig. 5 (e) shows examples of the ICs for the center computed alone.

To obtain the context-mediated PDs in dynamic natural scenes, we used sequences of image patches in which the current frame severed as the target and the three preceding frames as the context. We sampled a large number of sequences of image patches (~ 490,000) from a video database (Itti & Baldi, 2009) and performed the ICA according to Eq. (1). Fig. 6 (a) shows the paired chromatic spatiotemporal ICs. Fig. 6 (b) shows the paired achromatic spatiotemporal ICs. Fig. 6 (c) shows the unpaired ICs for the current frame, which are oriented bars and have red-green or blue-yellow opponency.



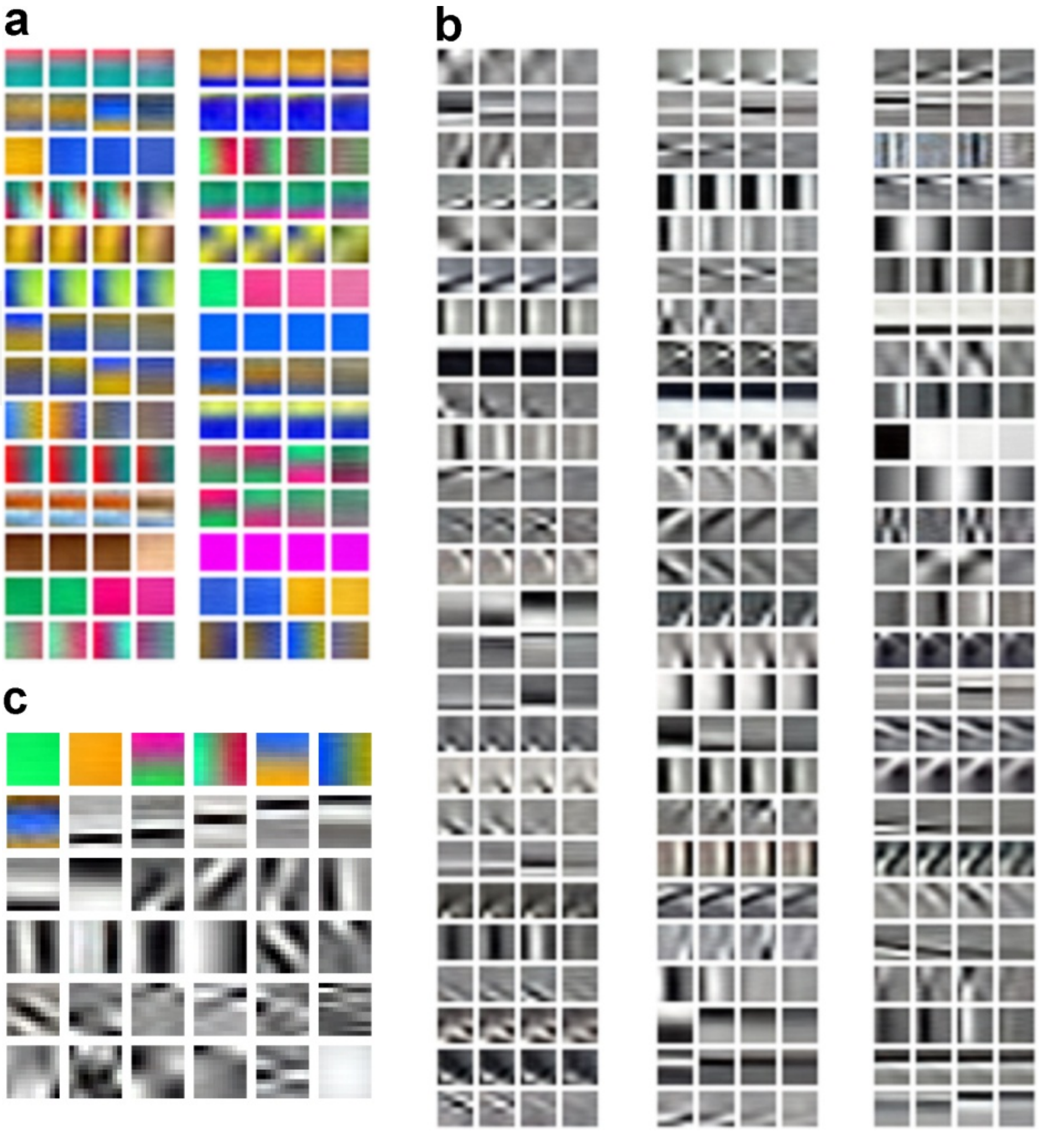
(a), Samples of patches of natural scenes. Each central patch in the left panel is paired with a surrounding patch in the right panel in the same raster order. (b), Examples of paired chromatic center and surround ICs. ICs in the panels are paired in the same raster order. (c), Examples of paired achromatic center and surround ICs. (d), Examples of unpaired center ICs. (e), Examples of the ICs for the center computed alone (modified from Xu et al., 2010).

**Figure 5.** ICs of color images of natural scenes.

The context-mediated PDs of natural scenes, i.e., the conditional PDs, $P(X_c \mid X_s)$, can be derived using the Bayesian formula as follows

$$P(X_c \mid X_s) = \frac{P(X_c, X_s)}{P(X_s)} \propto \frac{P(U_s)P(U_{sc})}{P(U_s)} = \prod_i P(u_{sc}^i) \tag{3}$$

where $u_{sc}^i$ is the amplitude of the ith unpaired IC for $X_c$. Therefore, the context-mediated PDs depend only on the unpaired ICs for $X_c$. We modeled $P(u_{sc}^i)$ as generalized Gaussian PDs.



(a), Selected 28 red/green or blue/yellow paired ICs. (b), Selected 78 paired bright/dark ICs. (c), Examples of unpaired target ICs (modified from Xu et al., 2010).

**Figure 6.** ICs of natural moving scenes.

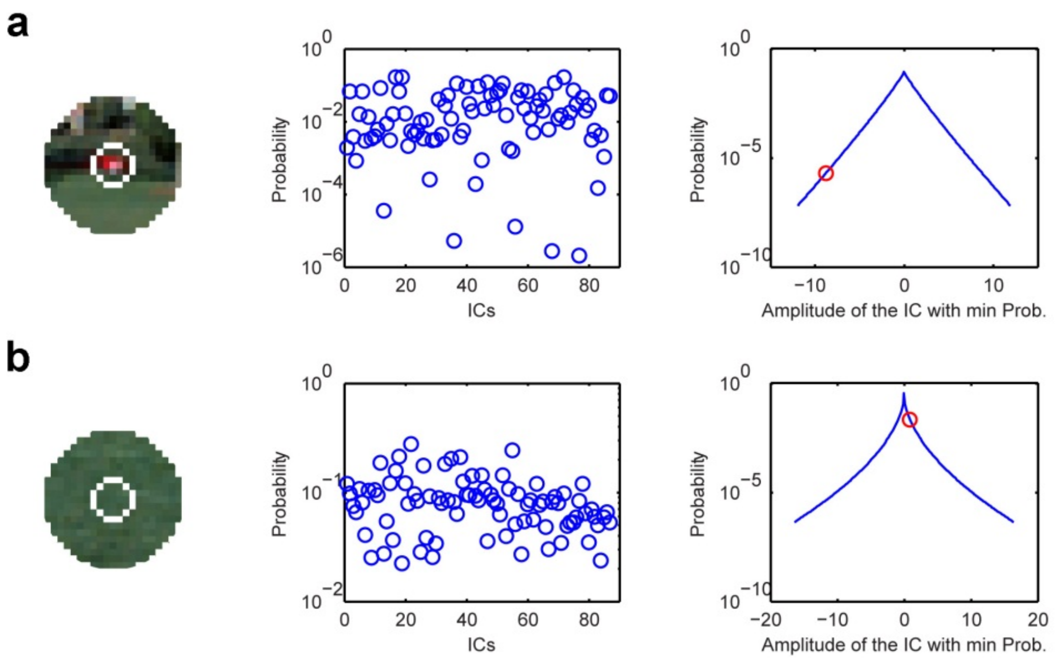
We proposed a measure of visual saliency as

$$S = \ln P_{\max}(X_c \mid X_s) - \ln P(X_c \mid X_s) \tag{4}$$

Substituting Eq. (3) into Eq. (4), we have

$$S = \sum_i \ln P_{\max}(u_{sc}^i) - \sum_i \ln P(u_{sc}^i) \tag{5}$$

where $P_{\max}(X_c \mid X_s)$ is the maximum probability of a target, $X_c$, that co-occurs with a context, $X_s$, in natural scenes. Thus, if the probability of the occurrence of a target is low relative to that of the most likely occurrence in the context in natural scenes, the target is salient within the context (Fig. 7).



(a), An image patch with an salient feature at the center (left), the probabilities of the ICs (middle), and the PD of the IC that has the smallest probability (right). (b), An image patch with an non-salient feature at the center (left), the probabilities of the ICs (middle), and the PD of the IC that has the smallest probability (right). In (a) and (b), the red circles indicate the probability of the amplitude of the IC of the features in the centers (modified from Xu et al., 2010).

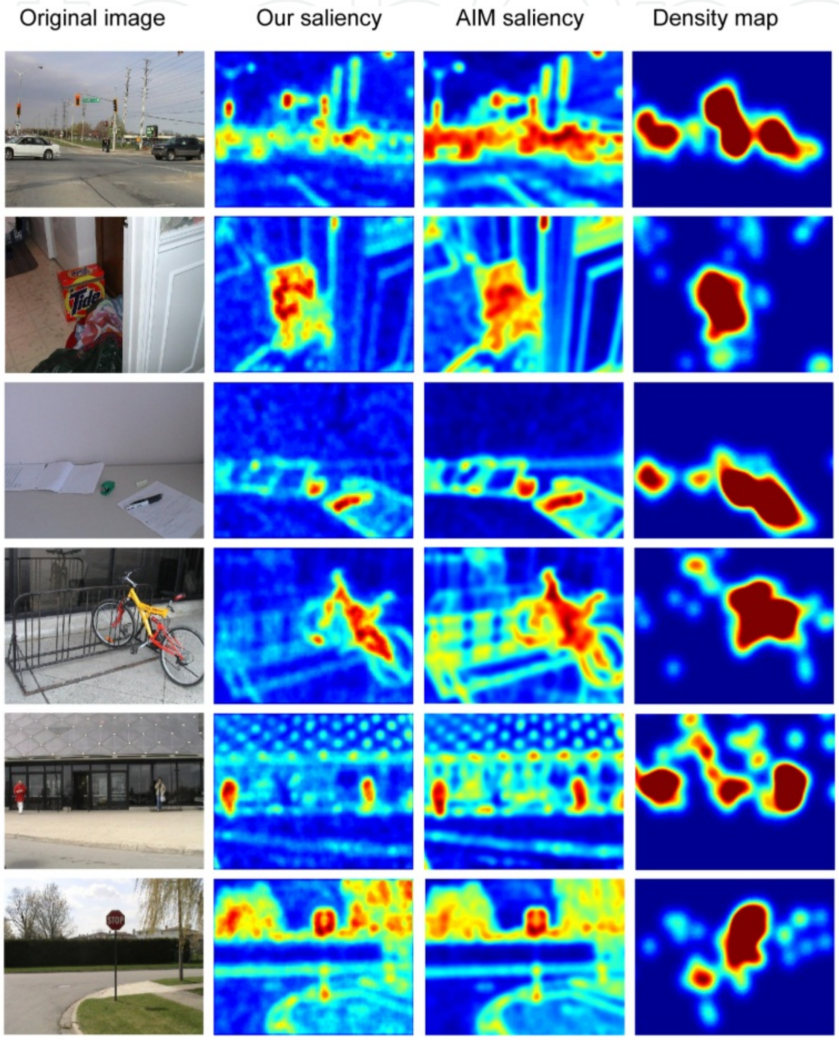
**Figure 7.** Visual saliency based on the context-mediated PDs in natural scenes.

### 3.3. Visual saliency and human gaze in free-viewing static natural scenes

Human gaze in free-viewing natural scenes is probably driven by visual saliency in natural scenes. To test this hypothesis, we used a dataset of human gaze collected from 20 human subjects in free-viewing 120 images (Bruce & Tsotsos, 2009). Fig. 8 shows the saliency maps based on the context-mediated PDs in natural scenes and the density maps of human gaze for six scenes. The saliency maps based on the information maximization (**AIM**) model are also shown (Bruce & Tsotsos, 2009). Evidently, the salient features and objects in these scenes predicted by the saliency maps accord with human observations and the saliency maps predicted by our model qualitatively matched the density maps of human gaze.

To quantitatively examine how well this model of visual saliency predicts human fixation, we used the receiver operating characteristic (**ROC**) metric and the Kullback–Leibler (**KL**) divergence. The ROC metric measures the area under the ROC curve. To calculate this metric, we used visual saliency as a feature to classify the locations where the saliency measures are greater than a threshold as fixations and the rest as nonfixated locations. By varying the threshold, we obtained an ROC curve and calculated the area under the curve, which indicates how well the saliency maps predict human gaze.

To avoid a central tendency in human gaze, we used the ROC measure described in (Tatler et al., 2005). We compared the saliency measures at the attended locations to the saliency measures in that scene at the locations that are attended in different scenes in the dataset, called shuffled fixations. The average area under the ROC curve is 0.6803, which means the saliency measures at fixations are significantly higher than the saliency measures at shuffled fixations. Similarly, we measured the KL divergence between two histograms of saliency measures: the histogram of saliency measures at the fixated locations in a test scene and the histogram of saliency measures at the same locations in a different scene randomly selected from the dataset (Zhang et al., 2008).



First column: input scenes. Second column, saliency maps produced by our model. Third column: saliency maps given by the AIM model. Fourth column: density maps of human fixation. Saliency is coded in color-scale (red/blue: high/low saliency) (modified from Xu et al., 2010).

**Figure 8.** Examples of saliency maps of natural scenes.

Our model of visual saliency is a good predictor of human gaze in free-viewing static natural scenes, outperforming all other models that we tested. As shown in Table 1 (Xu et al., 2010), our model has an average KL divergence of 0.3016 and the average ROC measure is 0.6803. The average KL divergence and ROC measure for the AIM model are 0.2879 and 0.6799 respectively, which were calculated using the code provided by the authors. The results for other models in Table 1 were given in (Zhang et al., 2008).

| Model | KL (SE) | ROC (SE) |
|---|---|---|
| Bruce et al. (2009) | 0.2879(0.0048) | 0.6799(0.0024) |
| Itti et al. (1998) | 0.1130(0.0011) | 0.6146(0.0008) |
| Gao et al. (2009) | 0.1535(0.0016) | 0.6395(0.0007) |
| Zhang et al.: DOG (2008) | 0.1723(0.0012) | 0.6570(0.0007) |
| Zhang et al.: ICA (2008) | 0.2097(0.0016) | 0.6682(0.0008) |
| Our model | 0.3016(0.0051) | 0.6803(0.0027) |

**Table 1.** ROC metric and KL-divergence for saliency maps of static natural scenes (SE: standard error).

## 3.4. Visual saliency and human gaze in free-viewing natural movies

We used a database of human gaze collected from 8 subjects in free-viewing 50 videos, including indoor scenes, outdoor scenes, television clips, and video games (Itti & Baldi, 2009). Fig. 9 shows the saliency maps for selected frames in 6 videos. The 3 contextual video frames and the target frame are shown to the left and the saliency maps to the right. As predicted by the saliency maps, the moving objects in these videos appear to be salient (e.g., the character in the game video, the falling water drop, the soccer player and the ball, the moving car and the walking policeman, and the jogger and the football player). These predictions accord well with human observations.

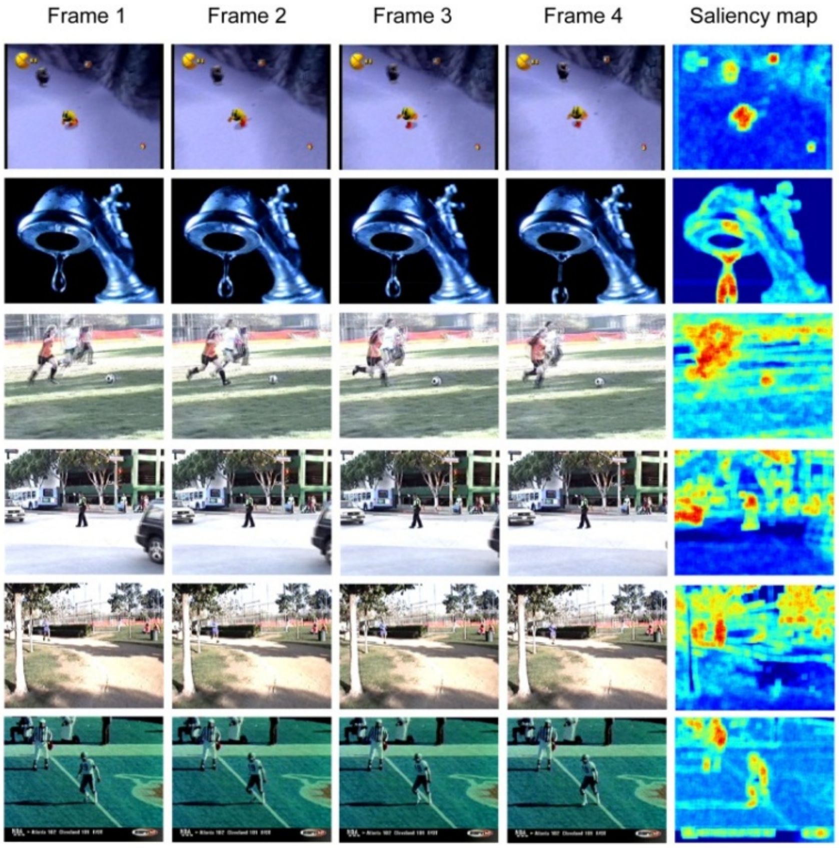


**Figure 9.** Saliency maps of dynamic natural scenes. Examples of contextual frames (3 left columns) and target frames (4th column) in 6 video clips and saliency maps (rightmost column) (modified from Xu et al., 2010).

We calculated the KL-divergence for this dataset as described above. Humans tend to gaze at visual features that have high saliency, as shown by the KL divergence measures in Table 2 (Xu et al., 2010). The KL-divergence measure for our model is 0.3153, which is higher than the saliency metric (0.205) (Itti et al., 1998) and the surprise metric 0.241 (Itti & Baldi, 2009), but slightly lower than the AIM model (0.328) (Bruce & Tsotsos, 2009).

## 3.5. Discussion

### *3.5.1. Distinctions from other models of visual saliency*

Our model of visual saliency is different from all other models. There are four classes of models of visual saliency. The first class of models do not use PDs in natural scenes but involve complex image-based computation that includes feature extraction, feature pooling, and normalization (Itti et al., 1998). The second class of models make use of PDs computed from the current scene the subject is seeing (Bruce & Tsotsos, 2009). The third class of models are based on PDs in natural scenes that are not dependent on specific contexts (Zhang et al., 2008). Finally, there is a biologically inspired neural network model (Zhaoping & May, 2007). Our model is unique in that: 1) the PDs are computed from an ensemble of natural scenes that presumably approximate the statistics human experienced during evolution and development; and 2) the PDs are dependent on specific contexts in natural scenes.

| **Model** | **KL (SE)** |
|---|---|
| Bruce et al. (2009) | 0.328(0.009) |
| Itti et al. (2009) | 0.241(0.006) |
| Zhang et al. (2009) | 0.181 |
| Itti et al. (1998) | 0.205(0.006) |
| Our model | 0.315(0.003) |

**Table 2.** KL-divergence for saliency maps of dynamic natural scenes (SE: standard error).

### *3.5.2. Neurons as estimators of context-mediated PDs in natural scenes*

These results support the notion that neurons in the early visual cortex act as estimators of the context-mediated PDs in natural scenes. This way, any single neuron relates an occurrence of any visual variable to the underlying PD in natural scenes. These PDs are related to all possible stimuli in natural scenes experienced by the visual animals over evolutionary and developmental time.

This hypothesis is distinct from the conventional view of neurons as feature detectors, the efficient coding hypothesis (Simoncelli & Olshausen, 2001), predictive coding (Rao & Ballard, 1999), the proposal that neurons encode logarithmic likelihood functions (Rao, 2004), and several recent V1 neuronal models that involve complex spatial-tempo structures but don't function as estimators of PDs in natural scenes (Rust et al., 2005; Chen at al., 2007).

Since the response of any single neuron encodes and decodes the PD of the visual variable in natural scenes, this concept is also different from probabilistic population codes where populations of neurons automatically encode PDs due to varying tuning among neurons and noise (Ma et al., 2005).

## 4. Statistics of 3D natural scenes and visual space

### 4.1. Introduction

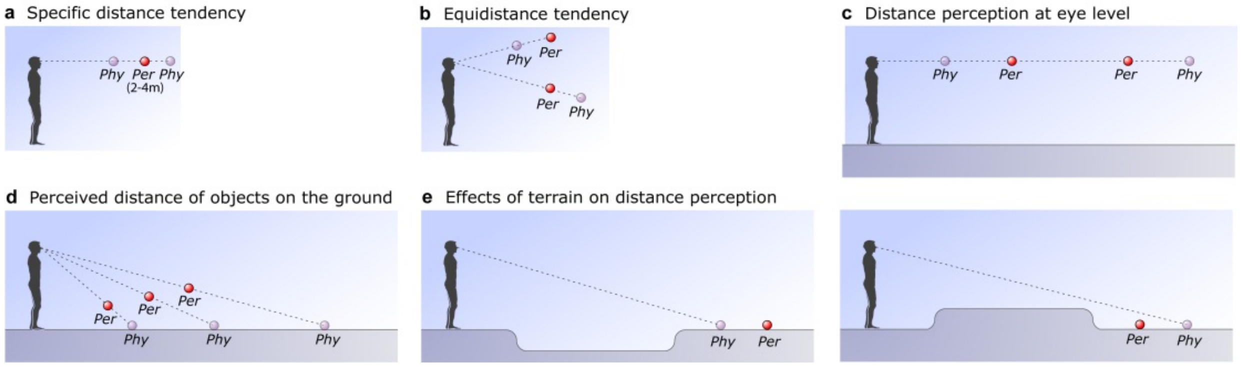
In the last two sections, I presented evidence that aspects of human natural vision are generated on the basis of the PDs of visual variables in 2D natural scenes. However, the most fundamental task of vision is to generate visual percepts and visually guided behaviors in the 3D physical world. In this section, I present PDs in 3D natural scenes and relate them to the characteristics of human visual space.

Visual space is characterized by perceived geometrical properties such as distance, linearity, and parallelism. An appealing intuition is that these properties are the result of a direct transformation of the Euclidean characteristics of physical space (Hershenson, 1998; Loomis et al., 1996; Gillam, 1996). This assumption is, however, inconsistent with a variety of puzzling and often subtle discrepancies between the predicted consequences of any direct mapping of physical space and what people actually see. A number of examples in perceived distance, the simplest aspect of visual space, show that the apparent distance of objects bears no simple relation to their physical distance from the observer (Loomis et al., 1996; Gillam, 1996) (Fig. 10). Although a variety of explanations have been proposed, there has been little or no agreement about the basis of this phenomenology.

We tested the hypothesis that these anomalies of perceived distance are all manifestations of a probabilistic strategy for generating visual percepts in response to inevitably ambiguous visual stimuli (Knill & Richards, 1996; Purves & Lotto, 2003; Trommershauser et al., 2011). A straightforward way of examining this idea in the case of visual space is to analyze the statistical relationship between geometrical features (e.g., points, lines and surfaces) in the image plane and the corresponding physical geometry in representative visual scenes. Accordingly, we used a database of natural scene geometry acquired with a laser range scanner to test whether the otherwise puzzling phenomenology of perceived distance can be explained in statistical terms (Fig. 11).

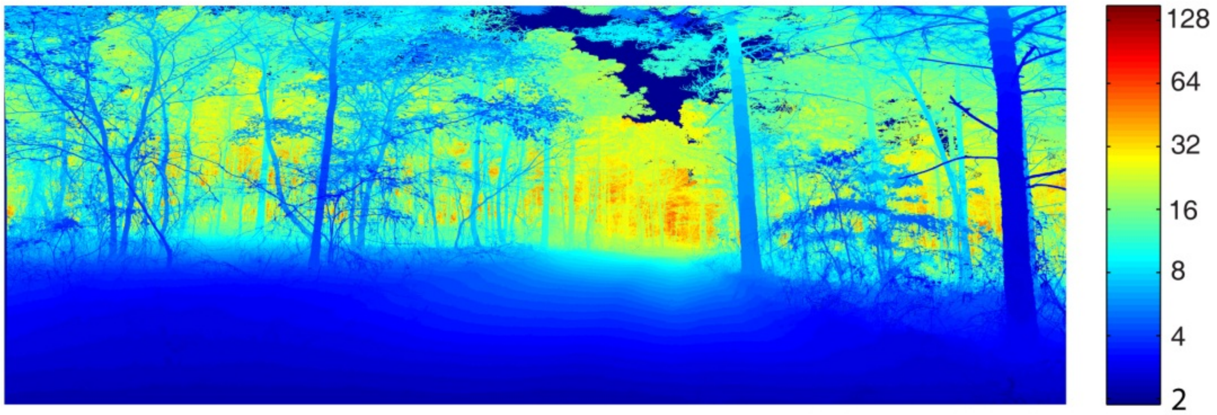
### 4.2. A probabilistic concept of visual space

The challenge in generating perceptions of distance (and spatial relationships more generally) is the inevitable ambiguity of visual stimuli. When any point in space is projected onto the retina, the corresponding point in projection could have been generated by an infinite number of different locations in the physical world. In consequence, the relationship between any projected image and its source is inherently ambiguous. Nevertheless, the PD of the distances of un-occluded object surfaces from the observer must have a potentially informative statistical structure. Given this inevitable ambiguity, it seems likely that highly evolved visual systems would have taken advantage of this probabilistic information in generating perceptions of physical space.


a Specific distance tendency
Phy Per Phy
(2-4m)
b Equidistance tendency
Phy Per
Per Phy
c Distance perception at eye level
Phy Per Per Phy
d Perceived distance of objects on the ground
Per Per Per
Phy Phy Phy
e Effects of terrain on distance perception
Phy Per
Per Phy

(a), Specific distance tendency. When a simple object is presented in an otherwise dark environment, observers usually judge it to be at a distance of 2-4m, regardless of its actual distance. In these diagrams, which are not to scale, 'Phy' indicates the physical position of the object and 'Per' the perceived position. (b), Equidistance tendency. Under these same conditions, an object is usually judged to be at about the same distance from the observer as neighboring objects, even when their physical distances differ. (c), Perceived distance of objects at eye-level. The distances of nearby objects presented at eye-level tend to be overestimated, whereas the distances of further objects tend to be underestimated. (d), Perceived distance of objects on the ground. An object on the ground a few meters away tends to appear closer and slightly elevated with respect to its physical position. Moreover, the perceived location becomes increasingly elevated and relatively closer to the observer as the angle of the line of sight approaches the horizontal plane at eye-level. (e), Effects of terrain on distance perception. Under more realistic conditions, the distance of an object on a uniform ground-plane a few meters away from the observer is usually accurately perceived. When, however, the terrain is disrupted by a dip, the same object appears to be further away; conversely, when the ground-plane is disrupted by a hump, the object tends to appear closer than it is (modified from Yang & Purves, 2003a).

**Figure 10.** Anomalies in perceived distance.



The distance (in meters) of each pixel is indicated by color coding. Black areas are regions where the laser beam did not return a value.

**Figure 11.** A range image acquired by laser range scanning.

This probabilistic strategy can be formalized in terms of Bayesian inference (Knill & Richards, 1996; Trommershauser et al., 2011). In this framework, the PD of the physical sources underlying a visual stimulus, P(S|I) can be expressed as

$$P(S|I)=P(I|S)P(S)/P(I) \qquad (6)$$

where S represents the parameters of physical scene geometry and I the visual image. P(S) is the PD of scene geometry in typical visual environments (the prior), P(I|S) the PD of stimulus I generated by the scene geometry S (the likelihood function), and P(I) a normalization constant.

If visual space is indeed determined by the PD of 3D scene geometry underlying visual stimuli, then, under reduced-cue conditions, the prior PD of distances to the observer in typical viewing environments should bias perceived distances. By the same token, the PD of the distances between locations in a scene should bias the apparent relative distances among them. Finally, when additional information pertinent to distance is present, these biases will be reduced.

### 4.3. PDs of distances in natural scenes

The information at each pixel in the range image database is the distance, elevation, and azimuth of the corresponding location in the physical scene relative to the laser scanner (Fig. 11). These data were used to compute the PD of distances from the center of the scanner to locations in the physical scenes in the database.

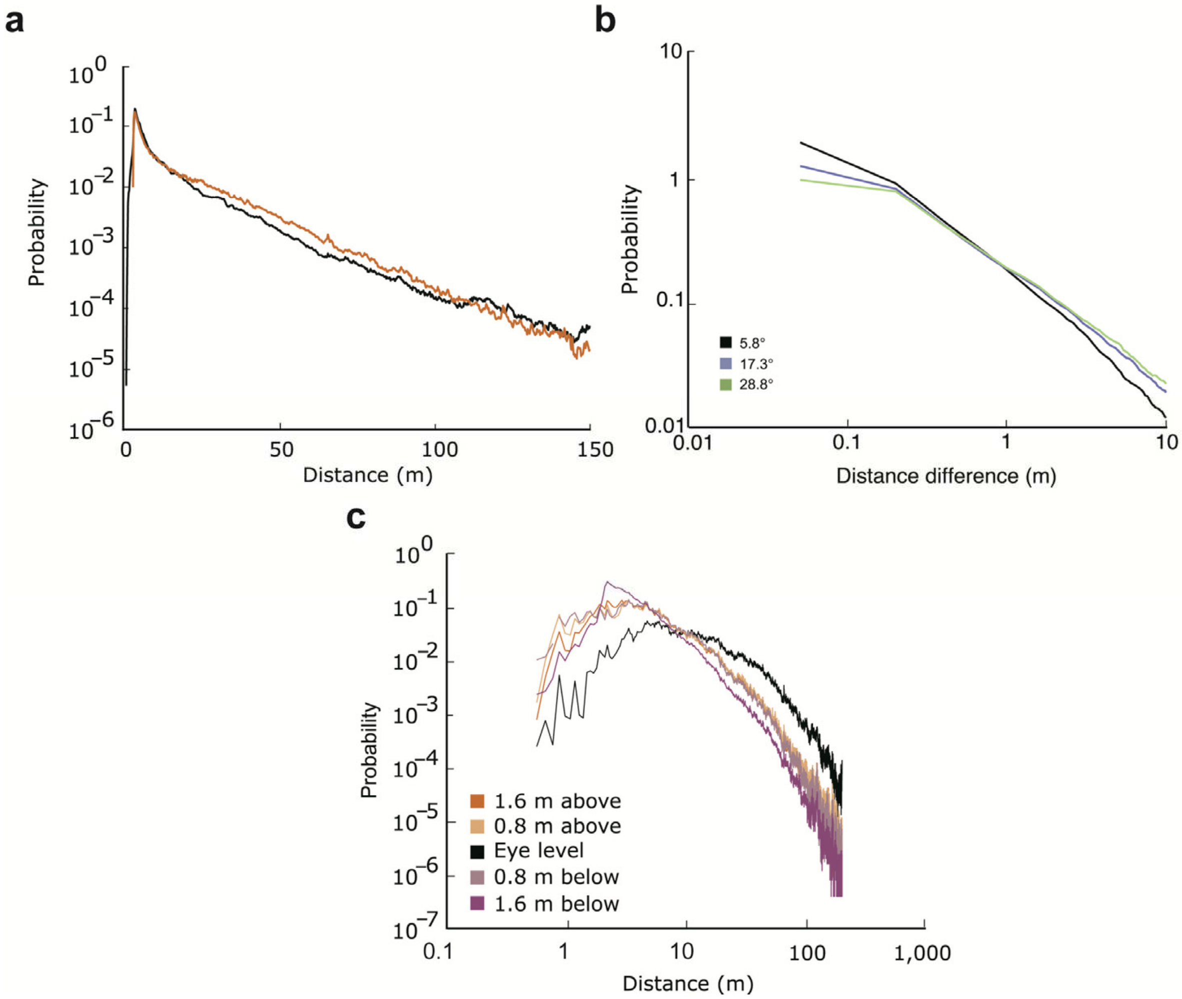
The first of several statistical features apparent in the analysis is that the PD of the radial distances from the scanner to physical locations in the scenes has a maximum at about 3 m, declining approximately exponentially over greater distances (Fig. 12 (a)). This PD is scale-invariant, meaning that any scaled version of the geometry of a set of natural scenes will, in statistical terms, be much the same (Lee et al., 2001). A simple model of natural 3D scenes generates a scaling-invariant PD of object distances nearly identical to that obtained from natural scenes (see legend of Fig. 12).

A second statistical feature of the analysis concerns how different physical locations in natural scenes are typically related to each other with respect to distance from the observer. The PD of the differences in the distance from the observer to any two physical locations is highly skewed, having a maximum near zero and a long tail (Fig. 12 (b)). Even for physical separations as large as 30°, the most probable difference between the distances from the image plane of two locations is minimal.

A third statistical feature is that the PD of horizontal distances from the scanner to physical locations changes relatively little with height in the scene (the height of the center of scanner was always 1.65m above the ground, thus approximating eye-level of an average adult) (Fig. 12 (c)). The PD of physical distances at eye-level has a maximum at about 4.7 m and decays gradually as the distances increase. The PDs of the horizontal distances of physical locations at different heights above and below eye-level also tend to have a maximum at about 3m, and are similar in shape.

### 4.4. Perceived distances in impoverished settings

How, then, do these scale-invariant PDs of distances from the image plane in natural scenes account for the anomalies of visual space summarized in Fig. 10?



(a), The scale-invariant PD of the distances from the center of the laser scanner to all the physical locations in the database (black line). The red line represents the PD of distances derived from a simple model in which 1000 planar rectangular surfaces were uniformly placed at distances from 2.5-300 m, from 150 m left to 150 m right, and from the ground to 25 m above the ground (which was 1.65m below the image center). The sizes of these uniformly distributed surfaces ranged from 0.2-18 m. Five hundred 512×512 images of this model made by a pinhole camera method showed statistical behavior similar to that derived from the range image database for a wide variety of specific values, although with different slopes and modes. The model also generated statistical behavior similar to that shown in (b) and (c) (not shown). (b), PDs of the differences in the physical distances of two locations separated by three different angles in the horizontal plane. (c), PDs of the horizontal distances of physical locations at different heights with respect to eye-level (modified from Yang & Purves, 2003a).

**Figure 12.** PDs of the physical distances from the image plane of points in the range images of natural scenes.

When little or no other information is available in a scene, observers tend to perceive objects at a distance of 2-4m (Owens et al., 1976). In the absence of any distance cues, the likelihood function in Eq. (6) is flat; the apparent distance of a point in physical space should therefore accord with the PD of the distances of all points in typical visual scenes (see Eq. (6)). As indicated in Fig. 12 (a), this distribution has a maximum probability at about 3 m. The agreement between this PD of distances in natural scenes and the relevant psychophysical evidence is thus consistent with a probabilistic explanation of the 'specific distance tendency'.

The similar apparent distance of an object to the apparent distances of its near neighbors in the retinal image (the 'equidistance tendency' (Owens et al., 1976)) also accords with the PD of the distances of locations in the natural scenes. In the absence of additional information about differences in the distances of two nearby locations, the likelihood function is again more or less flat. As a result, the PD of the differences of the physical distances from the image plane to any two locations in natural scenes should strongly bias the perceived difference in their distances. Since this distribution between two locations with relatively small angular separations (the black line in Fig. 12 (b)) has a maximum near zero, any two neighboring objects should be perceived to be at about the same distance from the observer. However, at larger angular separations (the green line in Fig. 12 (b)) the probability associated with small absolute differences in the distance to the two points is lower than the corresponding probabilities for smaller separations, and the distribution relatively flatter. Accordingly, the tendency to see neighboring points at the same distance from the observer would be expected to decrease somewhat as a function of increasing angular separation. Finally, when more specific information about the distance difference is present, this tendency should decrease. Each of these several tendencies has been observed in psychophysical studies of the 'equidistance tendency'.

### 4.5. Perceived distances in more complex circumstances

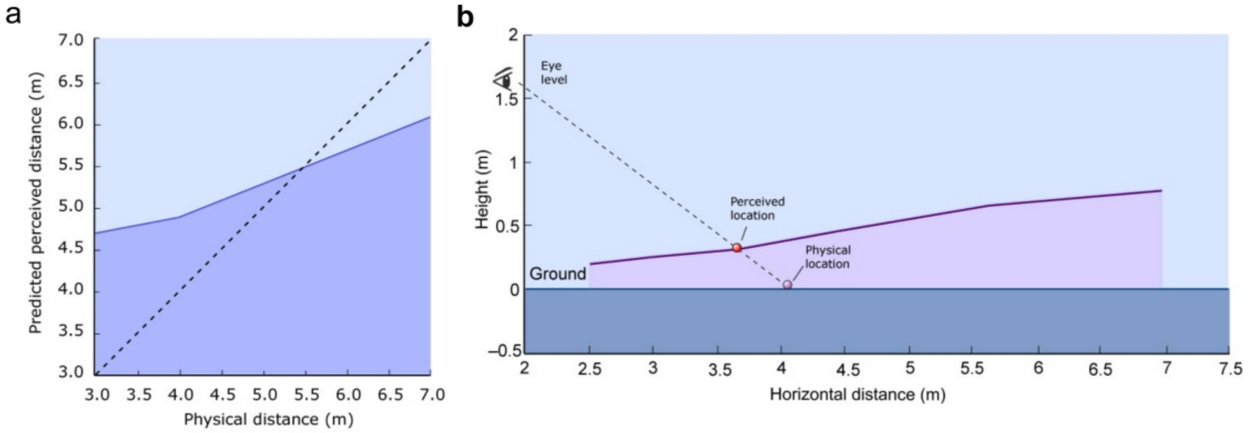
The following explanations for the phenomena illustrated in Figs. 10 (c) and (d) are somewhat more complex since, in contrast to the 'specific distance' and 'equidistance' tendencies, the relevant psychophysical observations were made under conditions that entailed some degree of contextual visual information. Thus, the relevant likelihood functions are no longer flat. Since their form is not known, we used a Gaussian to approximate the likelihood function in the following analyses.

The PD of physical distances at eye-level (the black line in Fig. 12 (c)) accounts for the perceptual anomalies in response to stimuli generated by near and far objects presented at this height (see Fig. 10 (c)). As shown in Fig. 13 (a), the distance that should be perceived on this basis is approximately a linear function of physical distance, with near distances being overestimated and far distances underestimated; the physical distance at which overestimation changes to underestimation is about 5-6 m. The effect of these statistics accords both qualitatively and quantitatively with the distances reported under these experimental conditions (Philbeck & Loomis, 1997).

To examine whether the perceptual observations summarized in Fig. 10 (d) can also be explained in these terms, we computed the PD of physical distances of points at different elevation angles of the laser beam relative to the horizontal plane at eye-level (Fig. 14). As shown in Fig. 14 (a), the PD of distances is more dispersed when the line of sight is directed above rather than below eye-level. The distribution shifts toward nearer distances with increasing absolute elevation angle, a tendency that is more pronounced below than above eye-level. A more detailed examination of the distribution within 30 m shows a single salient ridge below eye-level (indicated in red), extending from ~3 m near the ground to ~10 m at an elevation of -10°(Fig. 14 (b)). The distances of the average physical locations at different elevation angles of the scanning beam form a gentle curve. Below eye-level, the height of this curve is relatively near the ground for closer distances, but increases slowly as the horizontal distance from the observer increases. If the portion of the curve at heights below eye-level in Fig. 14 (c) is taken as an index of the average ground, it is apparent that the average ground is neither a horizontal plane nor a plane with constant slant, but a curved surface that is increasingly inclined toward the observer as a function of horizontal distance.

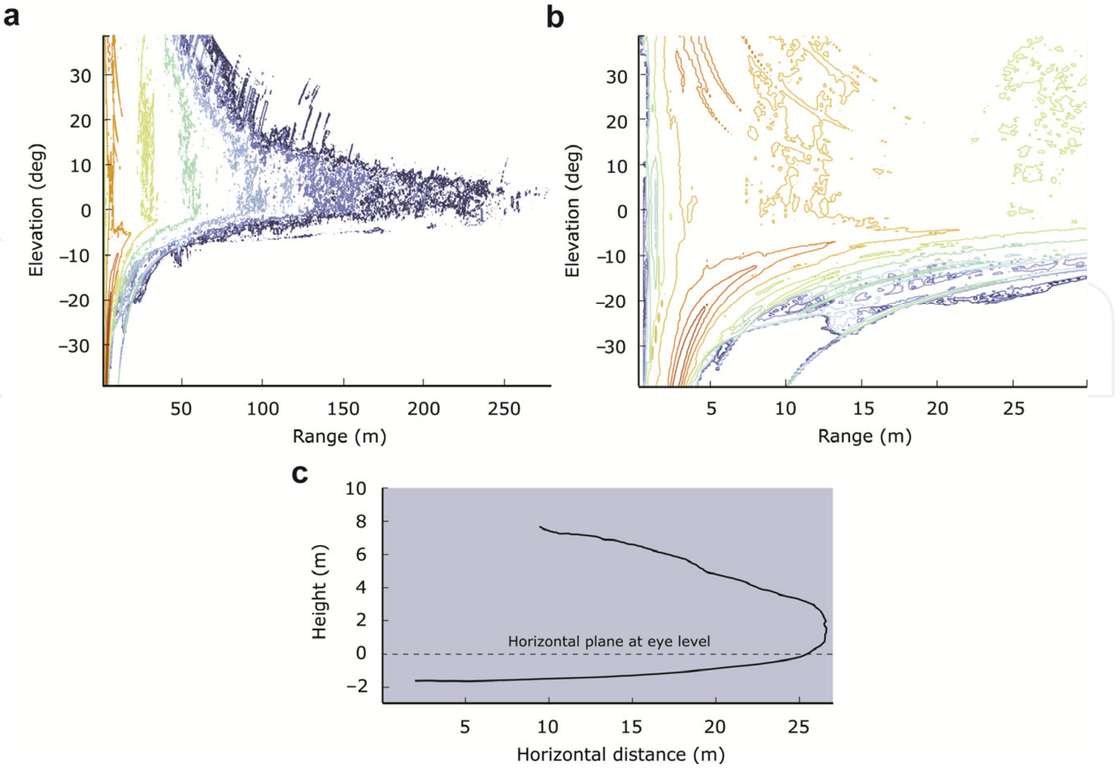
These characteristics of distance as a function of the elevation of the line of sight can thus account for the otherwise puzzling perceptual effects shown in Fig. 10 (d). The perceived location of an object on the ground without much additional information varies according to the declination of the line of sight, the object appearing closer and higher than it really is as a function of this angle. The apparent location of an object predicted by the PDs in Fig. 14 is increasingly higher and closer to the observer as the declination of the line of sight decreases, in agreement with the relevant psychophysical data (Ooi et al., 2001) (Fig. 13 (b)).

Other characteristics of visual space shown in Fig. 10 can be explained in the same way (Yang & Purves, 2003a).



(a), The perceived distances predicted from the PD of physical distances measured at eye-level. The solid line represents the local mass mean of the PD obtained by multiplying the PD in Fig. 12 (c) (black line) by a Gaussian likelihood function of distances with a standard deviation of 1.4 m. The dashed line represents the equivalence of perceived and physical distances for comparison. (b), The perceived distances of objects on the ground in the absence of other information predicted from the PD in Fig. 14 (a). The likelihood function at an angular declination α was a Gaussian function, i.e., $\sim \exp(-(\alpha-\alpha_0)^2/2\Sigma^2)$ ($\alpha_0=\sin^{-1}(H/R)$, $\Sigma=8°$, R is radial distance, and H=1.65m). The prior was the distribution of distance at angular declinations within [α-8°, α+8°]. The ground in the diagram is a horizontal plane 1.65m below eye-level (icon). The predicted perceptual locations of objects on the ground are indicated by the solid line (modified from Yang & Purves, 2003a).

**Figure 13.** The perceived distances predicted for objects located at eye-level and objects on the ground.



(a), Contour plot of the logarithm of the PD of distances at elevation angles indicated by color coding. Red indicates a probability value of ~$10^{-2}$ and blue ~$10^{-5.5}$. (b), Blowup of (a) showing the PD of distances within 30 m in greater detail. In this case, red indicates a probability value of ~$10^{-1.5}$. (c), The average distance as a function of elevation angle, based on the data in (a). The vertical axis is the height relative to eye level; the horizontal axis is the horizontal distance from the image plane. The curve below eye-level, if modeled as a piece-wise plane, would have a slant of about 1.5° from 3-15 m, and about 5° from 15-24 m (modified from Yang & Purves, 2003a).

**Figure 14.** PD of physical distances at different elevation angles.

## 4.6. Discussion

When projected onto the retina, 3D spatial relationships in the physical world are necessarily transformed into 2D relationships in the image plane. As a result, the physical sources underlying any geometrical configuration in the retinal image are uncertain: a multitude of different scene geometries could underlie any particular configuration in the image. This uncertain link between retinal stimuli and physical sources presents a biological dilemma, since an observer's fate depends on visually guided behavior that accords with real-world physical sources.

Given this quandary, we set out to explore the hypothesis that the uncertain relationship of images and sources is addressed by a probabilistic strategy, using the phenomenology of visual space to test this idea. If physical and perceptual space are indeed related in this way, then the characteristics of human visual space should accord with the PDs of 3D natural scene geometry. Observers would be expected to perceive objects in positions substantially and systematically different from their physical locations when countervailing empirical information is not available, or at locations predicted by the altered PDs of the possible sources of the stimulus in question when other contextual information is available. Using a database of range images, we showed that the phenomena illustrated in Fig. 10 can all be rationalized in this framework.

If visual space is indeed generated by a probabilistic strategy, then explaining the relevant perceptual phenomenology will inevitably require knowledge of the statistical properties of natural visual environments with respect to observers. Visual space generated probabilistically will necessarily be a space in which perceived distances are not a simple mapping of physical distances; on the contrary, apparent distance will always be determined by the way all the available information at that moment affects the PD of the gamut of the possible sources of any physical point in the scene.

## 5. Conclusion

These and many other studies present a strong case supporting the concept that vision works as a fundamentally statistical machine. In this concept, even the simplest visual percept has a statistical basis, i.e., it is related to a certain statistics in the natural environments that supports routinely successful visually guided behavior. The statistics of natural visual environments must have been incorporated into the visual circuitry by successful behavior in the world over evolutionary and developmental time.

There are a range of statistics in the natural environments. These include the statistics of 2D and 3D natural scenes in both space and time domains. As discussed here and elsewhere (Geisler, 2008), these statistics are related to a range of aspects of human natural vision. Since natural environments consist of objects of various physical properties that are arranged in 3D space and move in a variety of ways, the statistics of natural objects, activities, and events, though not discussed here, are critical for our understanding of human object recognition and activity and event understanding (Yuille & Kersten, 2006; Doya et al., 2007; Friston, 2010).

What could be the neural mechanisms underlying this fundamentally statistical machine? A broad hypothesis is that the response properties of visual neurons and their connections, the organization of visual cortex, the patterns of activity elicited by visual stimuli, and visual perception are all determined by the PDs of visual stimuli. In this conception, neurons do not detect or encode features, but by virtue of their activity levels, act as estimators of the PDs of the variables underlying any given stimulus. From this perspective, the function of visual cortical circuitry is to propagate, combine, and transform these PDs. The iterated structure of the primary visual cortex in primates may thus be organized in the way it is in order to generate PDs pertinent to simpler aspects of visual stimuli. By the same token, the extrastriate visual cortical areas may serve to generate PDs pertinent to more complex aspects of visual stimuli by propagating, combining, and transforming the PDs elaborated in the V1 area. The activity patterns elicited by any visual stimulus would, in this conception, be determined by the joint PDs of the variables underlying visual stimuli, which, in turn, determine what people actually see.

This statistical concept of vision and visual system structure and function is radically different from the conventional view, where visual neurons are conceived to perform bottom-up, image-based processing (e.g., computing zero-crossing, luminance and texture gradients, stereoscopic and motion correspondence, and grouping) to build a series of symbolic representations of visual stimuli (e.g., primal sketch, 2½) sketch, and 3D representation) (Marr, 2010). Since the statistics of natural scenes, which are, as argued above, fundamental to the generation of natural vision and visually guided behaviors, are not contained in any current stimulus the visual animal is seeing, any image-based feature extraction/representation construction in the current stimulus *per se* will not generate percepts that allow routinely successful behaviors. The results presented here and many others support this statistical concept of vision and visual system structure and function and several recent reviews also point to this new concept (Knill & Richards, 1996; Rao et al., 2002; Purves & Lotto, 2003; Doya et al., 2007; Trommershauser et al., 2011; Simoncelli & Olshausen, 2001; Yuille & Kersten, 2006; Geisler, 2008; Friston, 2010), but much is left to the next generation of neuroscientists.

## Author details

Zhiyong Yang
*Brain and Behavior Discovery Institute and Department of Ophthalmology,*
*Georgia Health Sciences University, USA*

## 6. References

Adelson, E. H. (2000). Lightness and perception and lightness illusion. In: *The New Cognitive Neuroscience (2nd Ed)*, M. Gazzaiga, (Ed.), pp. 339-351, MIT Press, ISBN 0262071959, Cambridge, MA, USA.

Anderson, B. L. & Winawer, J. (2005). Image segmentation and lightness perception. *Nature*, Vol. 434, pp. 79-83.

Blakeslee, B. & McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision Research,* Vol. 44, No. 21, pp. 2483-2503.

Bruce, N. D. & Tsotsos, J. K. (2009). Saliency, attention, and visual search: an information theoretic approach. *Journal of Vision*, Vol. 9, No. 5, pp. 1-24.

Chen, X.; Han, F.; Poo, M. M. & Dan, Y. (2007). Excitatory and suppressive receptive field subunits in awake monkey primary visual cortex (V1). *Proc. Natl. Acad. Sci. USA,* Vol. 104, No. 48, pp. 19120-19125.

Doya, K.; Ishii, S.; Pouget, A. & Rao, R. P. N. (Eds.). (2007). *Bayesian Brain, Probabilistic Approaches to Neural Coding.* MIT Press, ISBN 0-262-04238-X, Cambridge, MA, USA.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, Vol. 11, p. 127-138..

Gao, D.; Han, S. & Vasconcelos, N. (2009). Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 6, pp. 989-1005.

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol*. Vol. 59, pp. 167-192.

Gichrist, A. et al. (1999). An anchoring theory of lightness perception. *Psychological Review,* Vol. 106, No. 4, pp. 795-834.

Gillam, B. (1995). The perception of spatial layout from static optical information. In: *Perception of space and motion, W.* Epstein & S. Rogers, (Eds.), pp. 23-67, Academic Press, Inc., ISBN 978-0122405303, San Diego, CA, USA.

Hershenson, M. (1998). *Visual space perception*: *A Primer.* MIT Press, ISBN 978-0262581677, Cambridge, MA, USA.

Hubel, D. H. & Wiesel, T. N. (1977). Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. Lond. B*, Vol. 198, No. 1130, pp 1-59.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw,* Vol. 10, No. 3, pp. 626-634

Itti, L. & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, Vol. 49, pp. 1295-1306.

Itti, L. & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience,* Vol. 2, No. 3, pp. 194-203.

Itti, L.; Koch, C. & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 11, pp. 1254-1259.

Kingdom, F. A. A. (2011). Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research*, Vol. 51, No. 7, pp. 652-673.

Knill, D. C. & Richards, W. (Eds.). (1996). *Perception as Bayesian Inference.* Cambridge Univ. Press, ISNB 052146109X, New York, USA.

Lee, A. B.; 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,* Vol. 41, No. 1-2, pp. 35-59.

Loomis, J. M.; Da Silva, J. A.; Philbeck, J. W. & Fukusima, S. S. (1996). Visual perception of location and distance. *Current Directions in Psychological Science,* Vol. 5, No. 3, pp. 72-77.

Ma, W. J.; Beck, J. M.; Latham, P. E. & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience,* Vol. 9, pp. 1432-1438.

Marr, D. (2010). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (reprinted from 1982 edition). MIT Press, ISBN 978-0-262-51462-0, Cambridge, MA, USA.

Olmos, A. & Kingdom, F. A. A. (2004). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, Vol. 33, No. 12, pp. 1463 - 1473.

Ooi, T. L.; Wu, B. & He, Z. J. (2001). Distance determined by the angular declination below the horizon. *Nature,* Vol. 414, pp. 197-200.

Owens, D. A. & Leibowitz, H. W. (1976). Oculomotor adjustments in darkness and the specific distance tendency. *Attention, Perception, & Psychophysics,* Vol. 20, No. 1, pp. 2-9.

Philbeck, J. W. & Loomis, J. M. (1997). Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *J. Exp. Psychol. Hum. Percept. Perform.,* Vol. 23, No. 1, pp. 72-85.

Purves, D.; Williams, S. M.; Nundy, S. & Lotto, R. B. (2004). Perceiving the Intensity of Light. *Psychological Review,* Vol. 111, No. 1, pp. 142-158.

Purves, D. & Lotto, R. B. (2003). *Why We See What We Do, AN EMPIRICAL THEORY OF VISION*. Sinauer Associates, Inc., ISBN 0-87893-752-8, Sunderland, MA, USA.

Rao, R. P. (2004). Bayesian computation in recurrent neural circuits. *Neural Computation*, Vol. 16, No. 1, pp. 1-38.

Rao, R. P. N. & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, Vol. 2, pp. 79-87.

Rao, R. P. N.; Olshausen, B. A. & Lewicki, M. S. (Eds.). (2002). *PROBABILISTIC MODELS OF THE BRAIN, Perception and Neural Function.* MIT Press, ISBN 0-262-18224-6, Cambridge, MA, USA.

Rust, N. C.; Schwartz, O.; Movshon, J. A. & Simoncelli, E. P. (2005). Spatiotemporal elements of macaque v1 receptive fields. *Neuron*, Vol. 46, No. 6, pp. 945-956.

Simoncelli, E. P. & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Ann. Rev. Neurosci.* Vol. 24, pp. 1193-1216.

Tatler, B. W.; Baddeley, R. J. & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research*, Vol. 45, No. 5, pp. 643-659.

Trommershauser, J.; Kording, K. & Landy, M. L. (Eds.). (2011). *Sensory Cue Integration*. Oxford Univ. Press, ISBN 978-0-19-538724-7, New York, USA.

van Hateren, J. H. & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Biol. Sci.*, Vol. 265, No. 1394, pp. 359–366.

White, M. (1979). A new effect of pattern on perceived lightness. *Perception*, Vol. 8, No. 4, pp. 413-416.

Wishart, K. A.; Frisby, J. P. & Buckley, D. (1997). The role of 3-D surface slope in a lightness/brightness effect. *Vision Research,* Vol. 37, No. 4, pp. 467-473.

Xu, J.; Yang, Z. & Tsien, J. (2010). Emergence of Visual Saliency from Natural Scenes via Context-mediated Probability Distributions Coding. *PLoS ONE,* 5(12):e15796. doi:10.1371/journal.pone.0015796.

Yang, Z. & Purves, D. (2003a). A statistical explanation of visual space. *Nat. Neurosci.,* Vol. 6, No. 6, pp. 632-640.

Yang, Z. & Purves, D. (2003b). Image/source statistics of surfaces in natural scenes. *Network: Computation in Neural Systems,* Vol. 14, No. 3, pp. 371-390.

Yang, Z. & Purves, D. (2004). The statistical structure of natural light patterns determines perceived light intensity. *Proc. Natl. Acad. Sci. USA*, Vol. 101, No. 23, pp. 8745-8750.

Yuille, A. & Kersten, D. (2006). A Vision as Bayesian inference: analysis by synthesis? *Trends Cogn. Sci.,* Vol. 10, No. 7, pp. 301-308.

Zhang, L.; Tong, M. H. & Cottrell, G. W. (2009). SUNDAy: Saliency using natural statistics for dynamic analysis of scenes. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Amsterdam, Netherlands, August, 2009.

Zhang, L.; Tong, M. H.; Marks, T. K.; Shan, H. & Cottrell, G. W. (2008). SUN: A Bayesian framework for saliency using natural statistics. *Journal of Vision*, Vol. 0, No. 1, pp. 1-20.

Zhaoping, L. & May, K. A. (2007). Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. *PLoS Comput. Biol. 3(4)*: e62. doi:10.1371/journal.pcbi.0030062.