Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 45 (2005) 3145–3168
Vision Research
www.elsevier.com/locate/visres

The effects of amplitude-spectrum statistics on foveal and peripheral discrimination of changes in natural images, and a multi-resolution model

C.A. Párraga [a,*], T. Troscianko [a], D.J. Tolhurst [b]

[a] Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Bristol, BS8 1TN, UK
[b] Department of Physiology, University of Cambridge, Downing Street, Cambridge, CB2 3EG, UK

Received 18 November 2004; received in revised form 29 July 2005

Abstract

Psychophysical thresholds were measured for discriminating small changes in spatial features of naturalistic scenes (morph sequences), for foveal and peripheral vision, and under M-scaling. Sensitivity was greatest for scenes with near natural Fourier amplitude slope, perhaps implying that human vision is optimised for natural scene statistics. A low-level model calculated differences in local contrast between pairs of images within a few spatial frequency channels with bandwidth like neurons in V1. The model was "customised" to each observer's contrast sensitivity function for sinusoidal gratings, and it could replicate the "U-shaped" relationships between discrimination threshold and spectral slope, and many differences between picture sets and observers. A single-channel model and an ideal-observer analysis both failed to capture the U-shape.

Keywords: Natural scenes; Contrast; Optimisation; Cortex model; M-scaling; Ideal observer

1. Introduction

Classical psychophysical and electrophysiological studies with simple (usually sinusoidal grating) stimuli have resulted in an impressive understanding of channel characteristics of early vision. Yet, it is undoubtedly true that the stimuli viewed in everyday life are different and hugely more complex than gratings. Furthermore, the tasks carried out in everyday life are often more complex than simple grating detection and discrimination tasks. The present paper is motivated by a desire to learn about the relationship between the perception of "natural scenes" and the well-known properties of channels in early vision, as shown with sinusoidal gratings.

It is now generally hypothesised that the organisation of the visual system and the tuning characteristics of individual channels or neurons are optimisations for dealing with the salient information in the natural visual world (Barlow, 1961; Laughlin, 1983; Marr, 1982). The function of red–green colour opponency has been interpreted in these terms (Osorio & Vorobyev, 1996; Párraga, Troscianko, & Tolhurst, 2002; Regan et al., 2001), as has the contrast coding of single neurons or populations of neurons (Clatworthy, Chirimuuta, Lauritzen, & Tolhurst, 2003; Laughlin, 1981; Tadmor & Tolhurst, 2000). The spatial organisation of primary visual cortex (V1) neuron receptive fields seems to match the "statistics" of spatial features in the visual image (Hancock, Baddeley, & Smith, 1992; Olshausen & Field, 1997; Srinivasan, Laughlin, & Dubs, 1982; Van Hateren & Van Der Schaaf, 1998). However, such "visual ecology" generally looks at how the properties of single visual neurons rather than overall visual performance may be matched to the natural

* Corresponding author. Tel.: +44 117 928 8581; fax: +44 117 928 8588.
E-mail address: alej.parraga@bristol.ac.uk (C.A. Párraga).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.08.006

environment. We propose that, if the visual system really is optimised for the information in the natural environment, then visual detection and discrimination might be best when we use natural scenes as stimuli or, at least, stimuli with certain statistical characteristics of natural scenes (Geisler, Perry, Super, & Gallogly, 2001; Knill, Field, & Kersten, 1990).

To test this proposal, we need to compare detection or discrimination performance with natural and unnatural stimuli. We use digitised monochrome photographs of natural scenes to represent "natural scene stimuli". But, what is an unnatural stimulus? Natural scenes exhibit many statistical regularities (Geisler et al., 2001; Kersten, 1987), and the Fourier amplitude spectra of natural scenes show a remarkably stable relationship between the spatial frequency (f) and the amplitude of that spatial frequency component (their "second-order" statistics):

$$\text{Amplitude}(f) \propto f^{-\alpha}, \tag{1}$$

where α is the *spectral slope* of the scene and has values close to 1.2 on average (Burton & Moorhead, 1987; Carlson, 1978; Field, 1987; Párraga, Brelstaff, Troscianko, & Moorhead, 1998; Párraga et al., 2002; Tolhurst, Tadmor, & Chao, 1992). Given that natural scenes generally have this property, it is possible to define the degree of naturalness of related stimuli according to how close the α of an image is to its natural, unperturbed value. It is possible to produce versions of images which have α modified by an amount $\Delta\alpha$; such images may be regarded as increasingly "unnatural" as $|\Delta\alpha|$ increases. There have been several psychophysical investigations of visual discriminations using random-dot and digitised photograph stimuli whose amplitude spectra have been manipulated in such a way (Knill et al., 1990; Párraga & Tolhurst, 2000; Tadmor & Tolhurst, 1994; Thomson & Foster, 1997; Tolhurst & Tadmor, 1997).

In this paper, we investigate what is intended to be a more naturalistic discrimination task than has been used before: detection of small spatial changes in stimuli generated by morphing between two natural-scene images (Benson, 1994). Such a task might, for instance, be the basis of identifying facial identity or expressions, or of distinguishing between two slightly different objects. We measure thresholds for discriminating morphed image sequences for stimuli having natural and unnatural second-order statistics, to ask whether human vision *is* optimised for natural statistics. Primarily, we wish to know whether performance in such tasks and the effects of changes in amplitude spectral slope or viewing eccentricity are explicable in terms of the low-level channel structure of the visual system, so well characterised with grating stimuli.

There are a number of image-difference models designed to predict the visibility, e.g., of targets in natural scenes (Daly, 1993; Doll, McWorter, Wasilewski, & Schmieder, 1998; Lubin, 1993; Menendez & Peli, 1995; Rohaly, Ahumada, & Watson, 1997; Watson, 1987; Watson, 2000). The basis of these models is to split the two images to be compared into several spatial-frequency bands (compare Campbell & Robson, 1968; Peli, 1990), weighted by the contrast sensitivity function (CSF) of the observer. For each equivalent pair of points in the images, one must find whether, at each spatial frequency, the difference in contrast between the image patches is at or above the contrast discrimination threshold. This knowledge is provided by knowing the contrast discrimination function (Legge & Foley, 1980). This information needs to be spatially pooled over the whole of the two images, and some process must exist that allows information from different spatial-frequency bands to be combined (Rohaly et al., 1997). Such models have been used to look at applied issues such as image quality (e.g., to evaluate image compression algorithms) or the visibility of small military targets; it is less clear whether they will be able to account for shape discrimination data in experiments such as those proposed here. In this paper, we investigate whether such a multiple frequency-band model can account for the magnitudes of thresholds for the naturalistic morph-discrimination task for stimuli with natural and unnatural second-order statistics, and whether it can account for differences of thresholds between different observers and different viewing eccentricities.

A preliminary account of some aspects of this project has been published (Párraga, Troscianko, & Tolhurst, 2000). Further work has appeared in Abstract form (Párraga, Tolhurst, & Troscianko, 1999; Párraga et al., 2002; Párraga, Troscianko, Tolhurst, & Gilchrist, 2000).

2. Methods

2.1. The natural-image stimuli

The experiments described here are similar in design to those described in our previous work (Párraga et al., 2000). The stimuli were produced from four achromatic images (128×128 pixels, 8 bits of grey level) containing the face of a man, the face of a woman matched for size; and a bull and a car on grey backgrounds. Two different morph sequences were created, one by "morphing" the two faces (called here *man-to-woman*, courtesy of P.J. Benson) into a sequence of 41 slightly different faces, and another by "morphing" the bull into the car (called *bull-to-car*). Both morph sequences then consisted of a series of pictures varying in shape, contrast and texture in small incremental steps of 2.5% steps in the case of the *man-to-woman* series and of 0.5% steps in the case of the *bull-to-car* series. The difference in step size followed preliminary experiments