Local image statistics have a perceptual metric that is nearly Euclidean

Jonathan D. Victor, Daniel J. Thengone, Mary M. Conte

Department of Neurology and Neuroscience, Weill Cornell Medical College

16.552

VSS 2012

Motivation and Overview

The visual system is sensitive to many kinds of image statistics, including those of low and high order. In natural scenes, these different kinds of image statistics are intertwined in a complex fashion. This motivated the development of a library of synthetic images that enables testing the salience of individual image statistics and how they interact (Victor & Conte, 2012).

To simplify the problem, we consider only local statistics in binary images. Based on the findings of Tkačik et al. (2010), we focus on the statistics of 2x2 neighborhoods. This yields 10 distinct image statistics that range in order from 1 to 4 and capture image features including luminance, contrast, orientation, blobs, edges, and corners.

Here, we fully characterize the sensitivity to each of these 10 image statistics and their pairwise combinations. Sensitivity to individual statistics is highly conserved across observers, and a Euclidean model (a quadratic form) accurately accounts for their interactions. These findings define the geometry of a perceptual space based on the informative statistics of natural images.

Methods and Psychometric Functions

SUBJECTS
13 subjects
VA: 20/20, with correction if needed
Practice: approx 1600 trials

CONDITIONS
8 repeats of 20 on-axis points
16 repeats of 8 off-axis points
288 trials per block, random order
15 blocks = 4320 trials per plane
Feedback during practice only

STIMULI
Pixel Size: 14 min
Display Size: 14.8 deg^2
Binocular viewing at 1m
Contrast: 1.0
Duration: 120 ms (followed by mask)
Target: 16 x 64 pixels on a 64 x 64 array
Trials either have a structured target on a random background, or random target on structured background

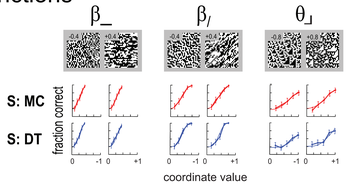
TASK
Find the location of the target stripe (4-AFC, top, right, bottom, left)

In this sample stimulus, the target stripe (defined by $\theta_{\llcorner}$) is on the bottom on a random background.

Here, the target stripe (on the left) is random on a structured ($\theta_{\llcorner}$) background.

Thresholds were not significantly different for these two conditions and did not depend on target location.

Each plot shows the fraction correct for stimuli that vary along a single texture coordinate. Performance is similar for positive and negative excursions of a coordinate, and was highly consistent across subjects (shown: MC and DT). Curves are maximum-likelihood fits to Weibull functions (shape parameter typically 2.2 to 2.6). Error bars are 95% confidence limits.



Coordinates for a Local Binary Texture Space

Textures are defined by the distribution of colorings of 2x2 blocks. There are 16 (=$2^{2\times2}$) possible colorings (right). However, there are constraints among these probabilities because the probabilities of smaller blocks cannot depend on *where* they occur within the 2x2 block.

Since there are 6 independent linear constraints on 16 probabilities, there are 16-6=10 free parameters., each of which corresponds to a coordinate.

One coordinate for luminance: $\gamma = -p[\blacksquare] + p[\square]$

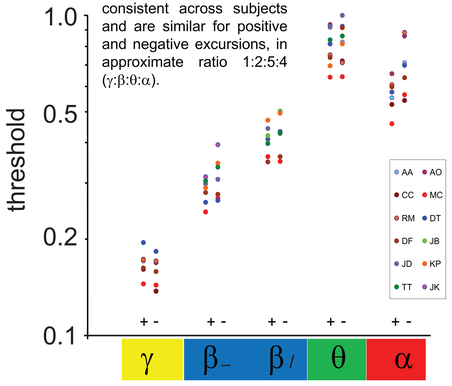
Four coordinates for pairwise correlation: $\beta_{|} = p[\blacksquare] - p[\blacksquare] - p[\blacksquare] + p[\square]$ (similarly for $\beta_{_}$, $\beta_{\backslash}$, $\beta_{/}$)

Four coordinates for the parity of triplets: $\theta_{\ulcorner} = -p[\blacksquare] + p[\blacksquare] + p[\blacksquare] - p[\blacksquare] + p[\blacksquare] - p[\blacksquare] - p[\blacksquare] + p[\square]$ (similarly for $\theta_{\lrcorner}$, $\theta_{\urcorner}$, θ_{L})

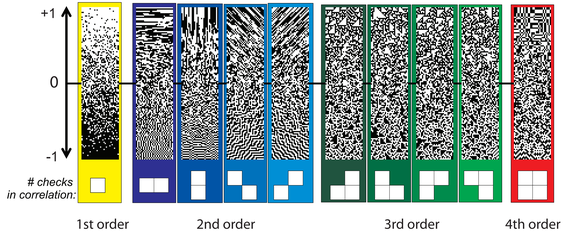
One coordinate for the parity of quadruplets: $\alpha = p[\blacksquare] - p[\blacksquare] - p[\blacksquare] + p[\blacksquare] - p[\blacksquare] + p[\blacksquare] - p[\blacksquare] - p[\blacksquare] + p[\blacksquare] + p[\blacksquare] - p[\blacksquare] - p[\blacksquare] - p[\blacksquare] + p[\blacksquare]$

Sensitivity on Each Coordinate Axis

Thresholds are highly consistent across subjects and are similar for positive and negative excursions, in approximate ratio 1:2:5:4 (γ:β:θ:α).

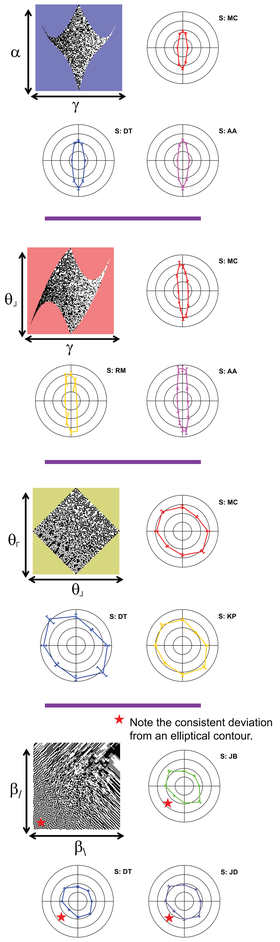


Coordinate Axes



Each strip shows the textures generated by varying one coordinate across its entire range, from -1 to +1. A coordinate value of 0 corresponds to a random texture.

Isodiscrimination Contours in Selected Coordinate Planes



★ Note the consistent deviation from an elliptical contour.

Thresholds and Model Fits in All 15 Unique Coordinate Planes

Ellipsoid model for discrimination in the entire 10-D space

$$d^2 = \sum_{i,j} Q_{i,j} c_i c_j$$

d is the perceptual distance from the random texture (d=1 corresponds to threshold)

$Q_{i,j}$ are the elements of a symmetric, positive-definite matrix

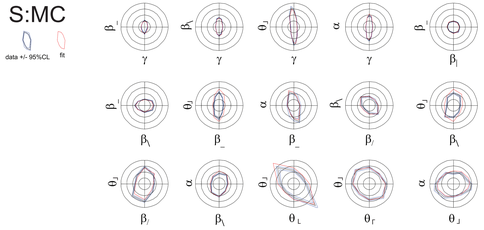
c_i are the texture coordinates

Under the hypothesis of a Euclidean perceptual metric, the isodiscrimination surface is an ellipsoid in the 10-parameter texture space, described by a symmetric matrix Q (see equation). We constrained the elements Q_{ii} to be equal for image statistics that differed only by a 90 deg. rotation. The resulting model is highly over-determined: it has 5 unique diagonal parameters and 15 unique off-diagonal parameters, and there are 128 measured threshold values to account for.

		Goodness of Fit (rmse) Raw	Corrected
Subject	MC	0.065	0.031
	DT	0.086	0.031
	DF	0.091	0.047
	JD	0.099	0.043
Mean		0.085	0.038

The raw rmse is the root-mean-squared deviation between the predictions of the ellipsoid model and the measured thresholds. The corrected rmse compares the predictions of the ellipsoid model with the best possible predictions of any opponent model, i.e., any model that predicts equal thresholds for positive and negative deviations of the image statistics.

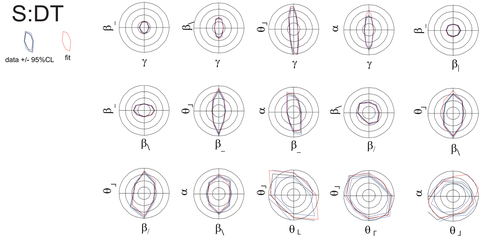
S:MC



The planar plots show all of the experimentally determined thresholds in each of the coordinate planes tested, along with 95% confidence limits (via bootstrap), and fits of the Euclidean model.

Some predicted contours deviate slightly from ellipses because they take into account the values of out-of-plane parameters.

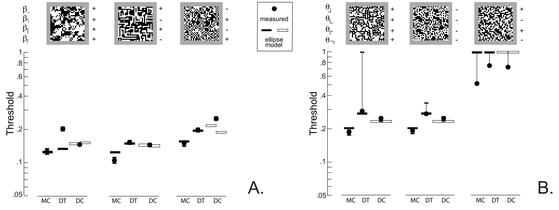
S:DT



Summary and Conclusions

- Local image statistics of low and high order can be described by a 10-parameter stimulus space.
- Both the salience of individual statistics and the pattern of interactions are highly conserved across observers.
- A Euclidean perceptual metric provides a good account for all pairwise interactions (rmse 0.05 to 0.1).
- The Euclidean metric also accounts for thresholds in directions defined by mixtures of four components.
- Deviations from the Euclidean model, though small, are largely consistent across observers. The main deviation is that there is an interaction unique to a specific combination of second-order statistics ($\beta_{\backslash}$ and $\beta_{/}$), suggesting a privileged role for corners.

Measured and predicted thresholds for four-component mixtures



The Euclidean model predicts thresholds for textures defined by four-component mixtures of image statistics.

A. Measured and predicted thresholds for four-component mixtures of second-order statistics.

B. Measured and predicted thresholds for four-component mixtures of third-order statistics.

For subjects MC and DT, model parameters were determined from their isodiscrimination contours in each of the relevant coordinate planes. For subject DC, an observer who only participated in this experiment, model parameters were determined from the subject's sensitivities to the individual image statistics, and the shapes of the isodiscrimination contours for subjects MC (left open symbol) and DT (right open symbol). The images over each set of points show an example of the texture with all parameters set to ±0.225.

Supported by EY7977