Vision Research 117 (2015) 117–135

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

A perceptual space of local image statistics

Jonathan D. Victor *, Daniel J. Thengone, Syed M. Rizvi, Mary M. Conte

Brain and Mind Research Institute, Weill Cornell Medical College, 1300 York Avenue, New York, NY 10065, United States

ARTICLE INFO

Article history:
Received 29 October 2014
Received in revised form 28 May 2015
Accepted 30 May 2015
Available online 16 September 2015

Keywords:
Local features
Visual textures
Metamers
Multipoint correlations
Intermediate vision

ABSTRACT

Local image statistics are important for visual analysis of textures, surfaces, and form. There are many kinds of local statistics, including those that capture luminance distributions, spatial contrast, oriented segments, and corners. While sensitivity to each of these kinds of statistics have been well-studied, much less is known about visual processing when multiple kinds of statistics are relevant, in large part because the dimensionality of the problem is high and different kinds of statistics interact. To approach this problem, we focused on binary images on a square lattice – a reduced set of stimuli which nevertheless taps many kinds of local statistics. In this 10-parameter space, we determined psychophysical thresholds to each kind of statistic (16 observers) and all of their pairwise combinations (4 observers). Sensitivities and isodiscrimination contours were consistent across observers. Isodiscrimination contours were elliptical, implying a quadratic interaction rule, which in turn determined ellipsoidal isodiscrimination surfaces in the full 10-dimensional space, and made predictions for sensitivities to complex combinations of statistics. These predictions, including the prediction of a combination of statistics that was metameric to random, were verified experimentally. Finally, check size had only a mild effect on sensitivities over the range from 2.8 to 14 min, but sensitivities to second- and higher-order statistics was substantially lower at 1.4 min. In sum, local image statistics form a perceptual space that is highly stereotyped across observers, in which different kinds of statistics interact according to simple rules.

1. Introduction

The analysis of image statistics underlies many key components of intermediate visual processing, including not only visual texture, but also visual characterization of surfaces and segmentation of images into objects. Although each of these tasks might at first seem deterministic, each is fundamentally statistical. For example, identification of surface materials (such as wood, grass, or hair) is not carried out by matching the image to a stored sample, but rather, by their image statistics, such as the range of contrasts and colors and the distribution of oriented contours at different scales (Karklin & Lewicki, 2009). Segmentation of an image is a statistical task as well, because it is fundamentally ambiguous: multiple scene interpretations are consistent with a single image, and image statistics play a key role in assessing which one is chosen as the most plausible. For example, contours due to a shadow or change in illumination are not typically coincident with a change in material properties, while real object boundaries typically have such changes, and hence, changes in image statistics.

Thus, understanding the processing of image statistics has broad importance as part of a foundation for understanding many aspects of intermediate visual processing. Visual textures, the focus here, present image statistics in their purest form.

While natural textures are characterized by many kinds of statistical features, systematic approaches to studying visual texture (with few exceptions (Motoyoshi & Kingdom, 2007; Saarela & Landy, 2012; Victor, Chubb, & Conte, 2005)) usually explore just one kind of feature, such as luminance distributions (Chubb, Econopouly, & Landy, 1994; Chubb, Landy, & Econopouly, 2004), color (Li & Lennie, 1997), orientation (Landy & Oruc, 2002; Wolfson & Landy, 1995, 1998), or curvature (Ben-Shahar & Zucker, 2004). There are two main reasons for this. One is the high dimensionality of the problem: if all kinds of statistical features were explored, the number of parameters (i.e., the number of different image statistics) would be impractically large. The other is that image statistics exhibit a high degree of interdependency. Edges cannot exist without local changes in luminance, and corners cannot exist without edges at multiple orientations, so these statistics cannot be considered to be independent attributes. Here, we attempt to address both issues, by constructing a texture space of large but manageable dimension (10), whose coordinates take into account the interactions implied by geometry. The data show that once these steps are taken, the perceptual interactions of image statistics obey simple rules that (a) are highly consistent across subjects, (b) accurately predict sensitivity to complex

* Corresponding author. Fax: +212 746 8050.
E-mail address: jdvicto@med.cornell.edu (J.D. Victor).

http://dx.doi.org/10.1016/j.visres.2015.05.018
0042-6989/

combinations of image statistics, and (c) are approximately preserved across a range of spatial scales.

To overcome the problem of high dimensionality (specifically, that an image statistic can be defined from the joint probabilities of any set of gray levels at any configuration of nearby points), we restricted consideration to black-and-white images on a checkerboard. By restricting the analysis to a single scale and only two luminance levels, we can then consider *all* possible local images statistics – i.e., the probabilities of all configurations of black and white checks within a 2×2 neighborhood. This set of image statistics has 10 free parameters (summarized here in Methods; detailed in (Victor & Conte, 2012)). It encompasses not only the intuitively-important features of luminance, contrast, edge, and corner, but also, its four-point correlations are independently informative for natural images (Tkačik et al., 2010). Thus, although it is a reduced space, it has image statistics of many different types and levels of complexity.

To overcome the second hurdle, the interdependency of different kinds of stimulus features, we used a "maximum-entropy" approach. That is, we specify stimuli by the prevalence of one or more elementary features, and then synthesize an ensemble of images that meet these specifications but are otherwise as random as possible. This limits the interdependence of features to what is implied by geometry, so that observed interactions at the level of neural or perceptual responses can be more readily interpreted.

1.1. Texture space and color space: their geometry and its implications

The above considerations lead to the construction of a "texture space", in which each point corresponds to a specific combination of image statistics that together specify luminance distributions and the prevalence of edges and corners at different orientations (Victor & Conte, 2012). The experiments presented here determine the perceptual distances in this space, focusing on the region near its origin.

The analogy with trichromatic color space provides a helpful geometrical framework. In both color space and texture space, points represent stimuli and the origin represents the neutral point (in color space, a white light; here, the random texture). The present experiments, which consist of measuring thresholds for perceiving that a texture is not random, correspond to measuring thresholds to changes in color and intensity near the white point. In both spaces, a line segment space represents mixtures. In color space, the points on a line segment are the colors that can be created by mixing the lights that correspond to the endpoints. In the space of local image statistics, the points on a line segment are the textures that can be created by mixing the textures that correspond to the endpoint. In color space, mixtures are created by physical mixing of lights; here, mixtures are created at the level of statistics: at the level of the frequency of each way that a 2×2 block can be colored with black and white checks (as described in (Victor & Conte, 2012). In color space and in texture space, a ray emanating from the origin corresponds to a set of stimuli that are progressively more saturated. Thus, determining the point along this ray that is first discriminable from the origin is a way of quantifying sensitivity to the combination of features represented by the direction of the ray. By determining the thresholds for rays that emanate from the origin in many directions, one can map out the "isodiscrimination surface," which summarizes the perceptual sensitivities in the neighborhood of the origin. In the case of color space, the isodiscrimination surfaces are approximately ellipsoids (the "Macadam ellipses" (Macadam, 1942)), and below we find that this holds in texture space as well.

The notion of navigating the space by moving along a straight line trajectory brings up an important mathematical distinction between the geometries of the two spaces. In color space, moving along a line is straightforward: it corresponds to increasing or decreasing the intensity of a light. For textures, this is not the case. For example, increasing the number of edges may also increase the number of intersections, and the proportionality between corners and intersections is typically nonlinear. These nonlinear dependencies underlie the maximum entropy approach (Victor & Conte, 2012) for navigating the space: a direction in the space corresponds to a specified coordinate, and movement along this direction may take a curved path to minimize the introduction of further structure. That is, the maximum-entropy approach yields a locally flattened coordinate system. Here, since we are studying discrimination thresholds, we work in these local coordinates, and ignore the impact of global curvature.

Color space and texture space have other important differences, and these allow us to interpret the sensitivity measurements in a way that has no immediate analogy in color space. The differences go beyond the difference in dimensionality or global curvature, and trace back to a fundamental difference in the way that the coordinate systems are defined. For color space, the origin of the coordinate system – the white point – is defined subjectively. For image statistics, the origin of the texture space has an *a priori* mathematical definition: it is the texture in which each check is randomly and independently assigned to black or white. A similar distinction applies to the axes: for color space, axes are defined empirically based on cone excitations (MacLeod & Boynton, 1979) or combinations motivated by physiological and psychophysical measurements (Derrington, Krauskopf, & Lennie, 1984); for image statistics, axes are defined *a priori* mathematically, in terms of correlations.

The kind of geometry that applies to the two spaces is also different (Zaidi et al., 2013). In color space, any of several coordinate systems (Derrington, Krauskopf, & Lennie, 1984; MacLeod & Boynton, 1979; Wyszecki & Stiles, 1967), each based on its own set of empirical observations, are equally valid descriptions of the space. Changing from one set of axes to another is a general linear transformation, which means that distances and angles computed from the coordinates in one system (via the Pythagorean rule and dot-products) need not match values computed with another. In the space of local image statistics, the coordinates are defined by mathematical considerations. This means that there is a standard definition of distance, and a "sphere" is a well-defined term: it is the locus of points that are at an equal distance from its center.

Because of the mathematics underlying the texture-space coordinates, spheres centered at the origin have another interpretation. Specifically, spheres are the isodiscrimination surfaces for an ideal observer ((Victor & Conte, 2012), Appendix B), i.e., an observer who is able to make full use of all image statistics. Of course we do not anticipate that human performance will resemble this. Rather, we expect that human observers will be selective, and make use of some image statistics more efficiently than others. This will distort the human isodiscrimination surface away from a spherical shape. For example, if sensitivity is reduced along one axis, then the isodiscrimination surface will become elongated in that direction, into an ellipsoid. If sensitivity is different for positive vs. negative changes in a coordinate, the surface will be asymmetrically distorted (i.e., it will become egg-shaped). If cues along different coordinates are not combined, the shape of the isodiscrimination surface will become squared-off. But as the results show, only the first kind of distortion is prominent, and this enables a concise, predictively accurate model for sensitivity to complex combinations of image statistics.

2. Methods

2.1. The stimulus space

The goal of these experiments is to determine visual sensitivity to local image statistics, individually and in combination. To do