Vision Research 115 (2015) 271–279

Contents lists available at ScienceDirect

Vision Research

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

Invariant texture perception is harder with synthetic textures: Implications for models of texture processing

Benjamin Balas [a,*], Catherine Conlin [b]

[a] *Department of Psychology, Center for Visual and Cognitive Neuroscience, North Dakota State University, Fargo, ND, USA*
[b] *Department of Psychology, North Dakota State University, Fargo, ND, USA*

ARTICLE INFO

Article history:
Received 16 June 2014
Received in revised form 25 January 2015
Available online 7 February 2015

Keywords:
Texture perception
Summary statistics
Invariance

ABSTRACT

Texture synthesis models have become a popular tool for studying the representations supporting texture processing in human vision. In particular, the summary statistics implemented in the Portilla–Simoncelli (P–S) model support high-quality synthesis of natural textures, account for performance in crowding and search tasks, and may account for the response properties of V2 neurons. We chose to investigate whether or not these summary statistics are also sufficient to support texture discrimination in a task that required illumination invariance. Our observers performed a match-to-sample task using natural textures photographed with either diffuse overhead lighting or lighting from the side. Following a briefly presented sample texture, participants identified which of two test images depicted the same texture. In the *illumination change* condition, illumination differed between the sample and the matching test image. In the *no change* condition, sample textures and matching test images were identical. Critically, we generated synthetic versions of these images using the P–S model and also tested participants with these. If the statistics in the P–S model are sufficient for invariant texture perception, performance with synthetic images should not differ from performance in the original task. Instead, we found a significant cost of applying texture synthesis in both lighting conditions. We also observed this effect when power-spectra were matched across images (Experiment 2) and when sample and test images were drawn from unique locations in the parent textures to minimize the contribution of image-based processing (Experiment 3). Invariant texture processing thus depends upon measurements not implemented in the P–S algorithm.

1. Introduction

Natural visual stimuli can vary substantially in appearance as a function of illumination conditions, the observer's distance to the stimulus, viewpoint or pose relative to the observer, and planar rotation. Indeed, a single real stimulus out in the world (e.g. an object or texture) may present infinite variations in 2D appearance depending on the specific viewing conditions. Nonetheless, observers are typically able to cope with appearance variation reasonably well, achieving useful (if limited) levels of perceptual constancy with complex stimuli like familiar faces (Burton et al., 1999), real and nonce objects (Bulthoff & Edelman, 1992), and scenes (Xiao et al., 2010).

Texture and material perception both also exhibit invariance to ecologically-relevant changes in appearance to some extent. Observers can typically recognize or match textures across changes in planar rotation, changes in scale, and changes in illumination and can rapidly categorize images of materials taken in unconstrained settings (Sharan, Rosenholtz, & Adelson, 2009; Wiebel, Valsecchi, & Gegenfurter, 2013). Also, as Fleming (2014) observes for material perception, the sheer range of diverse images that can reliably be labeled as "plastic" or "metal", for example, suggests that the human visual system has some impressive means of compensating for variation in a range of parameters and extracting robust estimates of properties that are relevant to material categorization. To some extent, observers' ability to be both selective about what images they assign a material category (e.g. "glossy" to and generalize the same category to a diverse set of appearances suggests that some simple features that are useful tools for material perception in some settings (e.g., skewness of the pixel intensity histogram; Motoyoshi et al., 2007; Sharan et al., 2008) are likely not the sole basis for material inference (Anderson & Kim, 2009). Also, texture and material constancy is not perfect – Ho, Landy, and Maloney (2006) demonstrated for example, that roughness judgments regarding artificial textures were affected by illumination, a result that suggests that the measurements used to characterize illumination and roughness may to

---

* Corresponding author at: 1210 Albrecht Blvd., Fargo, ND 58108-6050, USA.
E-mail address: Benjamin.balas@ndsu.edu (B. Balas).

http://dx.doi.org/10.1016/j.visres.2015.01.022

some extent be confounded by the visual system. Variation in viewpoint also appear to affect roughness judgments in a similar fashion (Ho, Maloney, & Landy, 2007), which may suggest that whatever degree of perceptual constancy the visual system is able to achieve for textures may be constrained by some set of features (what Ho et al. refer to as *pseudocues*) that do not provide perfect information for invariant recognition.

Are there specific computational features that may be reasonable for achieving invariant texture recognition (and explain the various failures of perceptual constancy that have been observed)? The fact that textures (and materials) do not have consistent shape rules out some large classes of visual features that are useful in other domains. For example, hierarchical models of invariant object recognition (e.g. HMAX, Riesenhuber & Poggio, 1999) are likely to be most useful for recognizing objects with well-defined shape and consistent relationships between local and global contours, conditions that textures and materials typically do not meet. For texture recognition and discrimination, representations of visual structure that are less position-dependent are likely to be better. A useful shorthand for such representations is *summary statistics*, by which we refer to a broad class of image measurements that describe appearance using visual features considered in the aggregate: histograms of filter outputs, correlation functions between wavelet coefficients, or moments of intensity histograms. Summary statistics are in general useful for texture encoding because they inherently reflect some basic properties of texture perception. For example, they are naturally invariant to simple transformations of texture appearance like translation. Nonetheless, it remains far from clear what specific summary statistics the visual system may use for texture and material perception in general. A range of different feature vocabularies have been proposed to account for human performance with different kinds of textures and different tasks, including the "needle" statistics proposed by Julesz (1981), center-surround filter outputs (Bergen & Adelson, 1988), and the "back-pocket" model of texture perception (Landy & Graham, 2004). To our knowledge, the specific problem of how invariant texture recognition is achieved has received comparatively little attention – while a number of computational models that purport to achieve some level of invariant texture recognition have been developed (Varma & Zisserman, 2002, 2009; Xu, Ji, & Fermuller, 2009), we are not aware of any work with human observers exploring candidate features that the human visual system may use to recognize and discriminant natural textures in an invariant fashion.

In the present study, we examine the extent to which a specific set of summary statistics, those implemented by the Portilla–Simoncelli (P–S) texture synthesis algorithm, support matching of texture samples and texture properties given changes in illumination. This is a useful candidate model to consider for a number of reasons: first, compared to other parametric models of texture appearance, the P–S algorithm reliably generates high quality synthetic images for a wide range of natural textures. Second, the P–S model has been used in prior behavioral work (Balas, 2006, 2012) to demonstrate that the features used as the basis of the model have some perceptual validity. Finally, the model has also been used in recent years both as a model of peripheral vision in general (Rosenholtz, 2011) and to account for the properties of cells in the ventral visual stream that may process summary statistics of appearance (Freeman & Simoncelli, 2011). Taken together, these various lines of research suggest that the P–S algorithm is a particularly good target model for investigating the extent to which invariant texture perception may be supported by summary statistics. Here, we do this by implementing a texture discrimination task designed to reveal the extent to which the P–S representation of texture appearance is sufficient for texture recognition given naturalistic appearance variation.

We asked observers to perform a match-to-sample task that required them to match sample and test textures across changes in illumination using both original and synthetic versions of the stimuli. Similar to (though see below) the logic employed in recent studies that used "mongrels' of peripherally-viewed stimuli (Balas, Nakano, & Rosenholtz, 2009; Rosenholtz et al., 2011), we assumed that if the P–S algorithm is indeed a sufficient appearance code for invariant texture perception, performing our task with synthetic images should be no more challenging than with original images. Should this be the case, it would support the claim that the summary statistics in the model allow the visual system to cope with changes in illumination by constraining the appearance of images of the same texture under different lighting conditions to be more similar to one another than images of different textures. If, however, we observe a significant cost when observers complete our task using synthetic images, it would suggest that the summary statistics used in the P–S model do not offer a sufficient code for invariant texture processing. This result would suggest that invariant texture recognition may depend on higher-order statistics than those contained in the P–S model, or possibly even that summary statistics in general may not be an adequate tool for invariant texture perception.

An important caveat to our use of synthetic images here is that this study is not an examination of the role of "mongrels" and summary statistics in peripheral vision. While it would certainly be interesting to examine the capabilities of peripheral vision with regard to invariant texture recognition, in the current manuscript we are not characterizing the properties of peripheral vision nor using the P–S model as a proxy for computations that may occur in the periphery. Instead, we are examining the extent to which this particular model of texture appearance carries sufficient information about texture appearance for observers to achieve some level of invariant texture processing. Thus, while this study and prior work with "mongrels" share some qualities (the use of P–S textures) the goals of the current study are distinct and we do not comment here on how these tasks might play out in the visual periphery. Instead, in three complementary experiments, we offer insights into what features do and do not appear to make contributions to observers' ability to achieve invariant texture matching. In Experiment 1, we examine observers' ability to match original and synthetic texture samples subject to either changing illumination or stable illumination. In Experiment 2, we examine the contribution of luminance and contrast to this problem domain by imposing matched power spectra on our test images. Finally, in Experiment 3, we examine invariant texture processing by asking observers to match texture properties (rather than specific samples as in Experiments 1 and 2) of real and synthetic textures subject to changing vs. stable illumination. In all three experiments, we observe a significant cost of synthetic appearance, suggesting that the summary statistics included in the P–S model do not carry sufficient information to account for observers' abilities to match textures in an invariant way.

2. Experiment 1

In our first task, we used the Portilla–Simoncelli model as a means of determining the extent to which a rich appearance code based on summary-statistics was sufficient to support observers' ability to match texture samples under illumination change.

3. Method

3.1. Subjects

We recruited 13 participants (5 female) from the NDSU Introductory Psychology study pool. All participants reported normal