Displays 46 (2017) 16–24

Contents lists available at ScienceDirect

Displays

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

Generating an image that affords slant perception from stereo, without pictorial cues [☆]

J. Galeotti [a,b,*], K. Macdonald [b], J. Wang [b], S. Horvath [a], A. Zhang [a], R. Klatzky [c]

[a] *Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[b] *Dept. of Biomedical Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA*
[c] *Dept. of Psychology and Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

ARTICLE INFO

Article history:
Received 19 August 2015
Received in revised form 12 October 2016
Accepted 7 November 2016
Available online 9 November 2016

Keywords:
Texture gradients
Depth perception
Pictorial cues
Stereopsis
Fractals
Slant perception

ABSTRACT

This paper describes an algorithm for generating a planar image that when tilted provides stereo cues to slant, without contamination from pictorial gradients. As the stimuli derived from this image are ultimately intended for use in studies of slant perception under magnification, a further requirement is that the generated image be suitable for high-definition printing or display on a monitor. A first stage generates an image consisting of overlapping edges with sufficient density that when zoomed, edges that nearly span the original scale are replaced with newly emergent content that leaves the visible edge statistics unchanged. A second stage reduces intensity clumping while preserving edges by enforcing a broad dynamic range across the image. Spectral analyses demonstrate that the low-frequency content of the resulting image, which would correspond to the pictorial cue of texture gradient changes under slant, (a) has a power fall-off deviating from 1/f noise (to which the visual system is particularly sensitive), and (b) does not offer systematic cues under changes in scale or slant. Two behavioral experiments tested whether the algorithm generates stimuli that offer cues to slant under stereo viewing only, and not when disparities are eliminated. With a particular adjustment of dynamic range (and nearly so with the other version that was tested), participants viewing without stereo cues were essentially unable to discriminate slanted from flat (frontal) stimuli, and when slant was reported, they failed to discriminate its direction. In contrast, non-stereo viewing of a control stimulus with pictorial cues, as well as stereoscopic observation, consistently allowed participants to perceive slant correctly. Experiment 2 further showed that these results generalized across a population of different stimuli from the same generation process and demonstrated that the process did not substitute biased slant cues.

1. Introduction

When people view a surface that deviates from a perpendicular along the line of sight, they have available three classes of cues to judge its slant[1]: the so-called pictorial cues, which would be visible with monocular viewing; stereoscopic disparities, which are available only with two eyes; and oculomotor cues of convergence and accommodation. (Depending on viewing conditions, movement-based cues may also be available.) Considerable research on this topic has concentrated on the pictorial cues; more specifically, on how slant is signaled by gradient changes in spatial elements across a pictorial display. Gibson [1] famously used the term *texture gradient* to capture the progressive change in image statistics with distance from the viewer. Purdy, Gibson's student, attempted in his thesis work [2] to characterize gradient cues more systematically and found that several candidate properties of regular texture elements, such as principal axes and their ratios, were powerful cues to slant. In more recent research, Todd and associates [3,4] have argued that local analysis is insufficient to compute slant from objectively 2D textures; region-based metrics are needed.

Research on pictorial gradients seeks to understand how surface orientation is computed in the absence of stereo disparities, which would directly convey that points on a slanted surface vary in depth from the viewer. In circumstances where judgments are made from a 2D display, this constitutes a cue-conflict situation,

---

[☆] This paper was recommended for publication by Richard H.Y. So.

[*] Corresponding author at: Carnegie Mellon University, Robotics Institute, NSH A525, Pittsburgh, PA 15213, USA.

E-mail addresses: jgaleotti@cmu.edu (J. Galeotti), kam233@pitt.edu (K. Macdonald), jiw86@pitt.edu (J. Wang), shorvath@cmu.edu (S. Horvath), ajzhang@cs.cmu.edu (A. Zhang), klatzky@cmu.edu (R. Klatzky).

[1] The term *slant* is used here for rotations of an image relative to the fronto-parallel plane, with the rotation axis centered on the image and oriented either horizontally (pitch rotation) or vertically (yaw rotation).

http://dx.doi.org/10.1016/j.displa.2016.11.003
0141-9382/

because pictorial cues in a 2D picture will necessarily be in conflict with stereoscopic and oculomotor cues, both of which would indicate the picture lies in the frontal plane. Even if one eye is closed, there is still a conflict between accommodation and pictorial cues in a 2D image.

Relatively little research has addressed the reverse problem, namely, how stereo disparities can cue slant in the absence of pictorial cues. The goal of the present research was to develop an algorithm for generating an image that would approximate statistical invariance over a large range of scales, and hence not give rise to gradient cues when slanted, while still affording stereo perception of local depth—and hence slant. In contrast to classical scale space [5,6], in which statistics change across scale (typically monotonically due to Gaussian blurring), we are trying to maintain invariant statistics across scale.

Zabulis and Backus [7] offered one approach to this problem, "Starry Night" textures, composed of points varying in flux. Pictorial cues of size and perspective were avoided by adjusting each point in the texture to a circular dot of constant radius. While this dynamic approach is possible when the displays are computer-generated, one goal of our research is to eventually produce stimuli that could be printed and presented under a microscope at different slants, in order to investigate how well slant can be perceived under conditions of magnified viewing from stereo cues alone.

This goal led us to develop an algorithm for generating images with three critical properties: First, as the detection of stereo correspondence is facilitated by the uniqueness of shape features and the continuity of edges (e.g., Marr and Poggio [8]), there should be dense edge content. Second, in order to isolate the contribution of stereo cues, perspective changes under slant should be avoided (although foreshortening alone may not be sufficient to induce slant perception [9]). Third, given the known importance of gradient changes for pictorial slant perception, perceivable edge density should remain constant as the image is rotated relative to the viewer's line of sight. At a local level, rotation can be treated as a change in scale. Thus a critical feature of the generation algorithm, especially given our interest in slant under magnified viewing, is that the statistics of the resulting images are not indicative of viewing scale. In particular, as will be discussed later, the algorithm is constructed with the goal that the edge statistics (i.e., statistics of high-frequency content) are scale-invariant, while texture gradient information is suppressed by perturbation of low frequency content.

The goal of scale invariance suggests the use of a fractal image, which is scale invariant by definition. However, fractals are constrained only to have equivalent non-regularity across scales, and as is exhibited by well-known fractal sets [10,11], sub-regions can have relatively sparse, object-like shapes, the density of which changes with rescaling. The approach taken here was to determine a means of generating images with the desired properties *de novo* and then to test the extent to which they cued slant. Under the goal of scale invariance, the generated content consisted of overlapping edges with sufficient density that when zoomed in, edges that nearly spanned the original scale disappeared, but were replaced with newly emergent edge content that left the statistics unchanged. (These newly emergent edges are always in the generated image, but given the resolution limits of either a human eye or a computer monitor, they are too small to see when zoomed out.)

2. Image generation

2.1. Generation algorithm

In general, if a source image is viewed within a rectangular aperture, zooming in by a factor of k will reduce the viewed source area to $1/k^2$ of its previous value, while the source data that are currently in view will now be assigned k^2 its previous number of pixels. Maintaining invariance under zoom thus requires k^2 times as many emergent edges in the entire image as there are vanishing edges.

To achieve this goal, the algorithm additively superimposes a large number of randomly shaped and oriented, overlapping triangles, where the scales of the triangles are distributed so as to maintain the inverse-square law. Shapes are used as the route to edges, because a directly generated edge is a 2D shape (albeit very small on one of the dimensions) that will change in size as it is rescaled. Triangles, in particular, were chosen as edge generators because the literature on slant from texture indicates certain shapes that are precluded *a priori*, particularly circles and rectangles, which will create undesirable perspective cues in the image, as well as shapes with constant parameters such as ellipses with a common axis ratio, and even families of blobs with underlying statistical regularities.

The process of generating a single triangle is illustrated by Fig. 1: A circle (not rendered) of radius r was placed at a randomly selected point, three points were randomly selected around the circle, and tangent lines were drawn to these points, with the intersections constituting the three vertices of the triangle. Additional constraints were placed on the configuration so that the tangents intersected within a reasonable distance, avoiding near-parallel edges: Relative to the first point of tangency (p_1), p_2 was restricted to angular distances in the range ⟨(15–165°), (195–345°)⟩, and p_3 must fall on the larger arc formed between p_1 and p_2 such that (1) the resulting triangle will contain the inscribed circle and (2) p_3 will not fall within 20° of p_1 or p_2.

The scale of a triangle is defined as the radius of the largest circle that can be inscribed within it. To produce a distribution of scales, radii of circles used to generate triangles were drawn from a distribution of sizes according to

$$r = \frac{a}{\left(\frac{a}{r_{min}} - \gamma\right)} \tag{1}$$

where a is defined as:

$$a = \frac{1}{\frac{1}{r_{min}} - \frac{1}{r_{max}}} \tag{2}$$

In (1) r represents the radius of the circle to be plotted, and r_{min} and r_{max} are the limits of the radii allowed. The variable γ is drawn from a uniform distribution between 0 and 1. Eq. (1) achieves the goal of biasing the triangle distribution toward small scales.

To localize the triangles within the image, the centers of generating circles were placed randomly over its area. An extension on the boundaries was included in the computation as needed to generate triangles centered beyond them, so as to create uniform content, even at the extremities. The brightness values were then determined by additively superimposing the triangles, according to the rule that each pixel's final intensity equals the number of triangles covering it. Fifteen million triangles were sufficient to generate a "source image" at a resolution of 45,600 pixels wide by 31,200 pixels high with a broad range of intensities.

The source size was selected to allow printing at a large (12+ in.) size. The large source image also allows us to cut smaller patches that can be used to create experimental stimuli, either by printing or presenting on a monitor. In order for the displayed stimulus to exhibit the properties intended by the generation algorithm, its pixilation must be imperceptible to the observer under all viewing conditions across our range of slant. For our experiments, the monitor pixel size and viewing distance must be implemented so as to ensure that this goal is met. Another important constraint that applies to printed and computer-generated stimuli is that the pixel size of the source image must be as small as or