Stability and Instability of Perceived Shape from Shading

Kosma P. Stanowski

Faculty of Science

The University of Sydney

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

2024

Declaration of Originality

I certify that the intellectual content of this thesis is, to the best of my knowledge, original and my own except where acknowledged in the text. I wrote the thesis and my supervisor provided feedback on its construction. The work presented in this thesis has not been submitted in fulfillment of any other degree.

Kosma Stanowski

Abstract

Images of an object can vary considerably due to variation in illumination. Despite this, humans are surprisingly adept at recovering stable estimations of shape, suggesting the presence of invariant structure in images upon which our perception relies. Failures of shape constancy also occur where a change in one factor (e.g., illumination) is misperceived as a change in another (e.g., shape). This thesis examines conditions predictive of success and failure of shape constancy and identifies image structure that corresponds with our perception. Our experiments suggest that specific alignment between surface geometry and illumination results in misperception of local and adjacent shape features. Anisotropic shape features, for example a cylindrical ridge, produces minimal shading when the illuminant is aligned with its axis of minimal curvature, causing its apparent deletion. This local deletion of shape was found to produce illusory segmentation of adjacent surface geometry. We also present examples of non-affine changes to perceived shape beyond an affine reduction of perceived relief of anisotropic shape features related to increased contrast of shading gradients produced by slant variation parallel to the illumination direction, and reduced contrast produced by slant variation in the perpendicular direction. A ridge-like feature with a slight depression in its centre, resembling mountains along a mountain range were more likely to be misperceived as a smooth cylindrical ridge when the illumination direction was perpendicular to the feature's axis of elongation. The addition of specular reflections reduced the size of the effects, however the pattern of shape misperception was preserved. These effects were also observed when rendered using a light map containing multiple, distributed strong light sources, suggesting that the perception of Lambertian surfaces is primarily determined by low frequency structure of the illuminant. We conclude that illumination affects shape perception in a complex, piecewise manner and propose directions for future research.

Acknowledgements

I wish to express my deepest gratitude for my supervisor, Prof Bart Anderson, for his patience and support over the many years. Your deep curiosity for challenging problems both within and beyond perception has been inspirational. Thank you for allowing me the freedom to discover my own interests and your encouragement through the numerous false starts at the beginning of my journey. To Phil Marlow, it has been a pleasure working alongside you in the lab. You have been the first port of call guiding my ideas through their infancy. Thank you for always being available to discuss all matters in perception and life, and your continued support – from helping me understand the more esoteric problems in vision, to the nitty gritty of experimental design.

I would like to thank Dr Caleb Owens for what would amount to one of my most fulfilling experiences during my PhD journey, teaching first year psychology. I have learnt a lot from the integrity with which you approach all aspects of education and cannot imagine a better representative for our School for new students. To my fellow tutors, I thank you for your friendship over the many years and for being there during what would otherwise be an unbearably isolating time. In particular, I want to thank Sarah Walker for her role in bringing tutors together for group marking sessions. They were invaluable regardless of how much marking actually got done.

To my parents, thank you for always allowing me to tread my own path. To Beata, for knowing me better than I likely ever will, and to Stefan, for your genuine interest in my work. And to Sophie, I am forever indebted to you for your resilience and support during a challenging period of my life.

Table of Contents

List of Figures

List of Tables

Chapter 1. Image Formation and Estimating Shape From Shading

1.1. Interactions between light and surface material

The human visual system is adept at extracting rich information from the environment. When viewing an object, we simultaneously experience its 3D shape, the material it is composed of, and from what direction the light falls upon it. Given how readily qualities of shape, material, and illumination are perceived, it is surprising that the process by which the visual system manages to disentangle the physical causes of image structure is not fully understood. The complexity of the problem is due to its seemingly ill-posed nature requiring the computation of three independent parameters.

The appearance of surfaces depends on the structure of the illuminating environment, and how the surface reflects or scatters that light to our eyes. This section will introduce basic physics of how light interacts with surfaces to generate the characteristic appearance of different materials. The following section will explain how variations in surface orientation (i.e., shape) influence image structure.

1.1.1. Illumination

Natural illuminating environments are structured by both primary and secondary light sources. Emissive light sources, such as the sun or artificial lamps, are primary sources and are the highest intensity sources of light. Light incident to a surface is scattered or reflected depending on its shape and material properties and therefore act as secondary sources of light. The intensity and chroma of incident light at any given point is determined by light coming from all directions 'visible' from that point in 3D space. The visible portion of the illuminating environment varies due to changes in the projected positions of objects blocking parts of a light source (i.e., casting a shadow), and positions of the light sources themselves. Most natural light sources that are small may be modelled as a point light source from which light rays radiate outward, and the behaviour of larger light sources may be modelled as the summed contribution of many point sources. The intensity of light corresponds to the density of light rays reaching some point in space. The increasing divergence of light rays originating from a point source results in a decrease in light intensity received by a point in space proportional to the square of the distance away from the source. Some examples of light sources that do not adhere to the inverse square law are lasers, and torches that use lenses or parabolic reflectors to concentrate light into a beam of parallel rays. These variations are caused by light sources that are close to a surface relative to distances between points on that surface. However, for very distant sources such as the sun, or sometimes ceiling lights, whose position can reasonably be approximated as infinitely distant relative to the scale of surface features, the intensity and relative position (orientation) of light is identical over all points in a scene. As a point light source emitting diverging light rays becomes more distant, the rays reaching that point in space increasingly converge, and in the limit approach a beam of parallel rays, or collimated light source (see Figure 1). Therefore, a collimated light source illuminates surfaces uniformly.

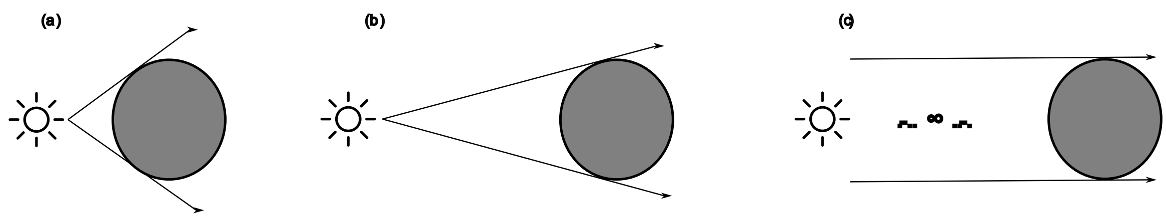


Figure 1. Illumination from a point light source at varying distances. As the distance of the light source increases, the orientation of the incident light rays converge and approach parallel in the limit. Once the light source is very distant relative to surface features, all light rays are parallel and illumination is invariant over that surface.

The 3D structure of a natural illuminating environment is modelled by a light field, which is a function that describes the intensity of light coming from all directions across all points in space (Gershun, 1939, cited by Mury et al., 2007b). Despite the complexities outlined above, simplified descriptions using spherical harmonics have been found to closely predict the interaction of matte materials under several qualitatively different light fields. Spherical harmonics depict three-dimensional periodic functions of various frequencies on the surface of a sphere. An ideal matte surface scatters light evenly in all directions and is therefore mainly influenced by low frequency components of the light field that are captured by the first three orders of spherical harmonics (see Figure 2). These spherical harmonics relate to physically and phenomenologically relevant qualities of the light field (Kartashova et al., 2019). The zeroth order corresponds to the integral of radiance over all directions on a sphere and may also be referred to as the ambient component of illumination. The first order specifies the net direction of light flux through a point in space. The second order term describes either a ring light (middle column), or light clamp (the region between two light sources of oppositely-directed flux, peripheral columns) depending on its polarity, and has been termed the "squash tensor" by Mury et al. (2007b). These low-order spherical harmonics are largely constant across different points in the light field and sufficiently describe the interaction between a wide range of qualitatively different light fields and matte materials (Mury et al., 2007a, 2009; Ramamoorthi & Hanrahan, 2001). Calculating the coefficients of each spherical harmonic order may also be used to describe the qualitative structure of light fields. The zeroth and first order spherical harmonics of the light field respectively describe the diffuse and directional components of illumination. A "diffuseness metric" proposed by Xia et al. (2017) uses the ratio of strengths of the zeroth and first orders to provide a continuous description of the diffuseness and directedness of a light field. The relative intensities of directional and diffuse light sources may be varied to simulate 'harsh' light under a directional-dominant light field (e.g., a clear sunny day) or 'soft' light under more diffuse illumination (e.g., a cloudy day). Psychophysical experiments confirm that human observers can intuitively report on these qualities of illumination (Kartashova et al., 2019; Koenderink et al., 2007; Xia et al., 2014, 2017).

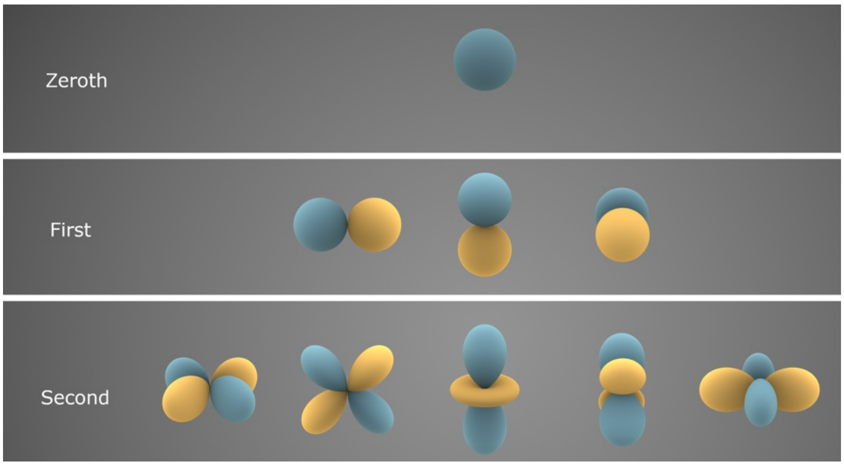


Figure 2. Visualisation of the first three orders of spherical harmonics. The sign of the function is depicted as blue (positive) and yellow (negative), with the distance from the origin indicating the absolute value of the function in that angular direction. Image from Inigo Quilez – Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=32782753.

Stimuli used in perception research are often computer-generated images (CGI) as it allows the easy and direct manipulation of variables of interest. The physics of surface-illumination interaction are well-understood and the primary challenge for CGI is in reducing computational demands. Early approaches used simple light fields that contain a limited, usually singular, range of dominant illumination directions with a uniform diffuse component to ensure that the entirety of the surface is illuminated. The dominant component may be a point light emitting diverging light rays, collimated source emitting light of a single orientation, or an area light emitting light over a limited range of orientations. Simple light fields are limited by their lack of higher-frequency structure present in natural illumination and often produce unrealistic images. Complex light maps are commonly used now to more closely capture natural illumination and promote natural appearance of CGI (Fleming et al., 2003). Complex light maps are images of natural environments captured by a panoramic camera or by photographing a mirror sphere, and often implemented in CGI software as spherical high dynamic range (HDR) images (Debevec & Malik, 1997). Since complex light maps are real-world images, they capture the spatial variation present due to secondary light sources absent in simple light fields. Although complex light maps generally produce convincing renderings of objects, they are limited in that they do not accurately reproduce all variations in local light field across different points on a surface. The important distinction between a light *map* and light *field* is that the light maps used in CGI are 2D, with each point on the light map positioned equidistantly from the origin, as opposed to a light field, where the relative positions and distances of light sources vary across the environment. Since the spherical images are produced by calculating the net luminance at a single point (i.e., the camera sensor), they do not model changes in the local light field caused by variations in the 3D position of each point (i.e., parallax). However, such variations primarily affect the higher-order structure of illumination (Mury et al., 2009) and so are of negligible consequence when rendering matte materials (Ramamoorthi & Hanrahan, 2001).

1.1.2. Surface Material

Images are structured by how interactions between illumination and surfaces redirect light toward our eyes or camera sensor. The material of the surface affects how incident light is reflected, absorbed, and transmitted. For opaque materials, these interactions may be modelled by a bidirectional reflectance distribution function (BRDF) that takes the irradiance from a given incoming direction and calculates the ratio of light reflected along a given outgoing direction. For example, a matte or Lambertian BRDF will scatter incident light uniformly in all directions, whereas with a specular BRDF, the angles between the incident and reflected light ray, and the surface normal, are equal. Most natural materials are modelled by combinations of BRDFs. For example, an apple's red colour would be modelled by a Lambertian BRDF, and its glossy highlights by a specular BRDF. The following section will explain the physics of how light interacts with ideal materials modelled by these BRDFs, and later how reflected light is influenced by 3D shape.

Matte Materials

An idealised Lambertian material reflects incident light uniformly in all directions, and therefore does not vary with changes of viewing angle (i.e., the outgoing direction). This is because rather than being reflected at the interface with the surface, incident light penetrates a matte surface where it is scattered before being re-emitted, a process called subsurface scattering. For matte materials, the depth at which subsurface scattering occurs is very shallow, meaning light exits the surface close to the point at which it entered. Because of this, matte materials are well modelled as a diffuse reflector, with light bouncing off the surface in all directions. Incident light is absorbed by a matte material's particles where it is then redirected stochastically. A light ray may interact multiple times in this fashion before exiting the surface in a random direction, causing the characteristic diffuse reflectance of matte materials.

The process of subsurface scattering is also responsible for the lightness (or albedo) and colour of most materials. Whenever incident light is absorbed by a material particle, it is re-emitted with reduced radiant energy, and the capacity of the material to absorb this radiant energy explains the proportion of incident illumination that is re-emitted from the surface (i.e., how light or dark it is). Diffuse reflectance also gives most surfaces their colour if its material absorbs certain wavelengths of light more than others due to spectral filtering of the incident illumination. For example, if a material selectively absorbs short and medium wavelengths, a higher proportion of long wavelengths will be visible in the re-emitted light, which human observers experience as 'red'. Diffuse reflectance is the most common cause of surface colour but other causes include certain metals like gold that reflect light specularly with a yellow tinge, and "structural colour" that emerges due to diffraction of light and gives some butterfly wings or bird feathers their colour.

The image intensity at a given patch of Lambertian surface is a function of irradiance, or density of light flux, and the surface's albedo (i.e., the proportion of incident light it reflects). For surfaces of uniform albedo, the amount of reflected light is a fixed proportion of incident light (irradiance) which depends on surface orientation relative to the illuminant.

Parts of the surface that are facing the illuminant (i.e., perpendicular to the light rays) will be brightest, with radiance varying smoothly as the surface slants away from the light source (see Figure 3). Since the appearance of matte materials is predominantly influenced by the low frequency structure of the illuminant, which is largely uniform across space (see Section 1.1.1.), variations in radiance over the surface covary with surface orientation and are referred to as 'shading'. Many objects have smooth variations in their local surface orientation and so matte surfaces are characterised by smooth gradients. For matte materials, the mean depth that light penetrates is shallow with light re-emerging close to the point at which it entered, meaning that diffuse reflectance is closely related to local surface orientation. Translucent materials have deeper subsurface scattering meaning light may exit such materials some distance away from where it entered, weakening the degree to which intensity is related to local surface orientation. Intensity gradients generated by translucent surfaces are more directly linked to the sign of curvature (convexities and concavities; Marlow & Anderson, 2021).

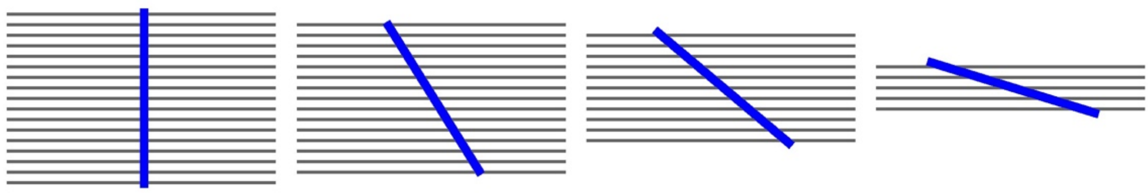


Figure 3. The density of light rays (depicted by thin grey lines) over a given surface patch (depicted by the thick blue line) gradually decreases as the surface slants away from the illumination direction. The radiance of a Lambertian material is proportional to the irradiance at that point.

The relationship between the intensity of light reflected by an ideal matte surface is modelled by Lambert's cosine law (Lambert, 1760). Assuming uniform illumination, the radiance depends only on the surface attitude, varying with the cosine of the angle subtended by the surface normal and incident light ray according to the equation

$$L_r = \frac{\rho}{\pi} \cdot \cos\theta_i \cdot E_0$$

where ρ is the albedo of the surface and E_0 is the irradiance. Variations in intensity of diffuse shading correspond to variations in orientation, or the first derivative, of the surface.

Lambert's law is an idealised model of a diffuse reflecting material, and although it generates realistic-looking images, it neglects other factors that influence image structure considerably in certain instances. The amount of light reaching a surface patch can be reduced when neighbouring surface features cast a shadow, referred to as 'shadowing', or limit the portion of the light field "visible" to that patch, referred to as 'vignetting' or 'masking'. Vignetting can cause concave regions to appear darker than what would be predicted by a simple cosine relationship. Under diffuse illumination, surfaces receive light evenly from all directions i.e., there is no single dominant direction of illumination. Under such conditions, surface luminance varies primarily with surface aperture, the proportion of the light field visible from a point on the surface, rather than the surface normal and follows a "dark-is-deep" rule (Langer & Zucker, 1994). Interreflection refers to the secondary illumination of surface patches by light reflected from elsewhere on the surface. Interreflections reduce the contrast of concave regions where light reflected from one part of the surface is likely to reach another part. This effect is more pronounced for surfaces of higher albedo as a larger proportion of incident light is reflected between points on the surface.

Specular Materials

Unlike matte materials, specular materials reflect light at the surface interface. An ideally smooth specular surface resembles a mirror and reflects each ray of light in a single direction such that the angles about the surface normal of the incident and reflected rays are equal (see Figure 4). As there is minimal scattering of light, such surfaces generate sharp reflections of the environment. Since each light ray is reflected at a specific orientation, the visibility of these rays is dependent on the position of the observer and the image will be composed of only the light rays that project to the line of sight, unlike Lambertian materials that do not vary with changes in viewpoint. This change in the reflected light with relative motion between the surface and viewing direction causes specular reflections to slide over the surface.

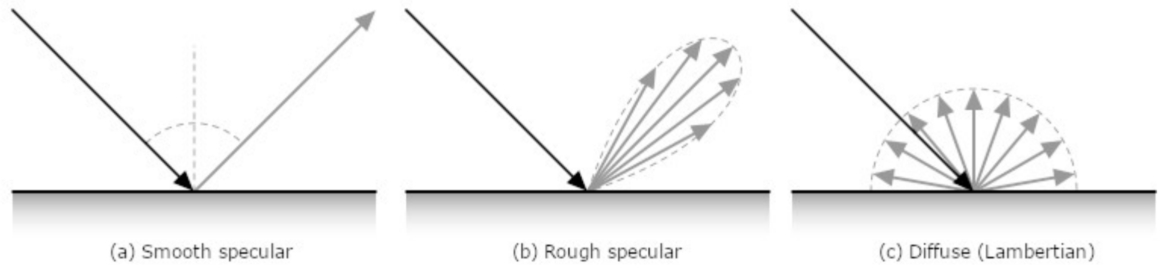


Figure 4. Diagrams depicting reflectance behaviour of three BRDFs. Incident light rays are shown as black arrows with light reflected by the materials shown by grey arrows. A (a) smooth specular surface (e.g., mirror) reflects incident rays in a specific direction equal to the angle of incidence. Introducing roughness (microscale bumps) to a specular surface (b) causes the reflected rays to spread due to small variations in surface orientation. Higher levels of roughness cause the material to behave more similarly to a diffuse (Lambertian) material (c).

Natural specular materials have varying degrees of surface roughness that progressively blur the light they reflect. Materials like unpolished metals are covered in microscale bumps that reflect light at different angles due to small variations in surface orientation. Surface roughness can also be anisotropic, or uniformly rougher in one direction than another as with brushed metal on the underside of a saucepan. Rather than isotropic bumps, a saucepan contains concentric rings of ridges, meaning the surface is roughest along radial directions and blurs reflections more strongly in these directions. At intermediate scales, this blurs reflections due to widening of the reflected ray into a lobe (Figure 4b), and in the limit can behave like diffuse shading.

The images reflected by specular surfaces are systematically distorted by variations in surface geometry. Planar specular surfaces, such as an achromatic (silver) mirror, generate undistorted images of their environment, both in scale and colour. With curved surfaces, variations in surface orientation induce corresponding distortions in the structure of the reflections. Unlike Lambertian shading, specular features are related to the curvature, or second derivative, of a surface. Each point of a specular surface reflects the light sampled from the point on the sphere of the light map corresponding to the view vector reflected about the surface normal. Regions of the surface where the surface normal changes quickly (i.e., regions of high curvature) compress a large portion of the light field within a small area of the surface and the degree of compression is reduced for flatter regions (Figure 5). Since regions of high curvature compress a large portion of the light field onto a small surface area, their reflections change little with small variations in orientation between the surface and light field. Consequently, specular features tend to 'stick' to regions of high curvature, and flow rapidly over flatter regions with relative movement between the light field and surface. Variation in the degree of compression of reflections forms a kind of 'texture' that corresponds to the principal curvatures of the surface.

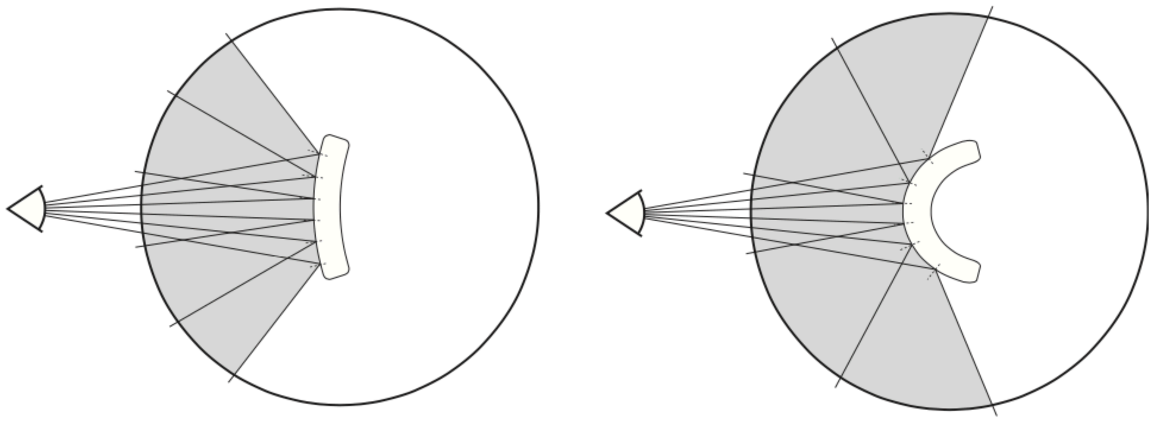


Figure 5. Regions of higher curvature (right image) reflect a larger portion of the light field to the viewer than regions of lower curvature (left image). This causes greater compression of the light field along the direction of higher curvature, as well as a tendency for specular reflections to 'stick' to regions of high curvature as the surface moves slightly relative to the environment. Figure from Fleming et al. (2004).

Principal curvatures describe the directions and magnitudes of minimal and maximal curvatures at a point on the surface. The axes of principal curvature are always perpendicular and describe the directions in which the reflected light field is maximally and minimally compressed. Figure 6 shows purely specular objects illuminated by a noise texture light map along the top row and close-ups of the centres of these images along the bottom row. The rightmost image depicts a sphere that is equally curved in all directions, and the close-up shows that the reflected noise texture is undistorted (or equally compressed in all directions). When the object is stretched into an egg shape, the curvature along the major axis is much lower than along the minor axis (i.e., the surface curves away more slowly along the major axis). A corresponding change occurs in the reflected pattern, which now consists of parallel streaks along the direction of minimum curvature, as well as relative compression of the texture in the orthogonal direction along the axis of maximum curvature. Processing this texture with oriented odd-symmetric filters generates an 'orientation field' that describes variation in the direction of stretching of the reflected light field (see Figure 7). This orientation field provides constraints on the directions of the principal curvatures, and information about their relative magnitudes are provided by the degree of anisotropy. Orientation fields generated by two different shapes under three different illuminations are illustrated in Figure 8. The hues in the right columns indicate the peak orientation, and the saturation the degree of anisotropy. The orientation fields generated by each shape are strikingly stable for each light field.

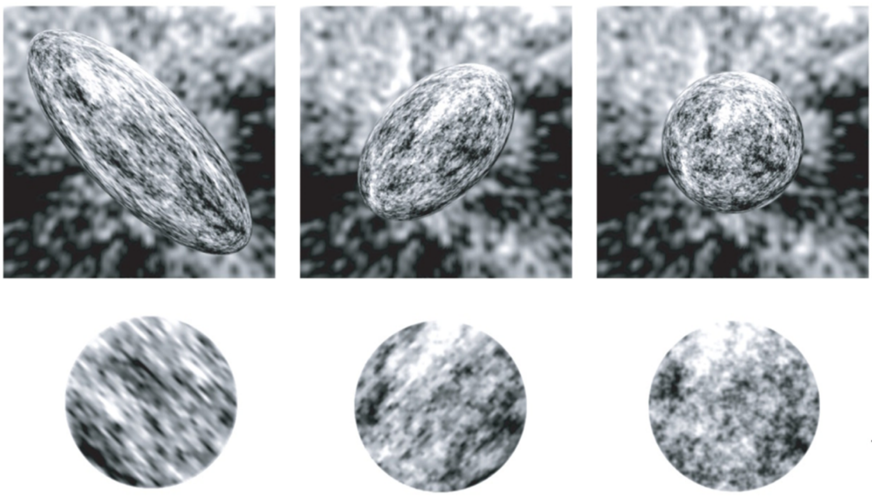


Figure 6. Images of specular egg shapes (top left, middle) and a sphere (top right) rendered in a 1/f noise light map with close-ups of image centre (bottom row). The sphere is equally-curved in all directions and reflects the light map with distortion. The egg shapes curve more gradually along the axis of elongation, causing the reflections to appear stretched in this direction, relative to the stronger compression in the orthogonal direction. This generates a pattern of distortion that specifies the orientation and relative magnitude of the principal curvatures. Figure from Fleming et al. (2004).

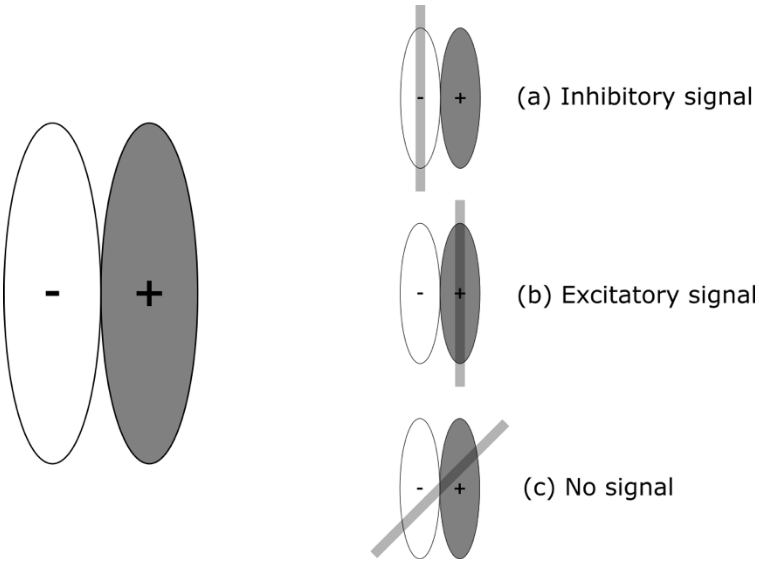


Figure 7. Diagram of odd-symmetric filter as used in image analysis of Fleming et al. (2004). The regions bounded by ellipses (receptive fields) register inhibitory (white) and excitatory (grey) signals when covered by a light bar (transparent grey). Odd-symmetric filters are orientation-selective as they output the net activity registered by both receptive fields such that an oblique stimulus produces zero signal (c). Both inhibition (a) and excitation (b) provides signals by determining the polarity of an edge (i.e., light to dark or dark to light). Populations of such filters at different orientations determine the orientation field of an image.

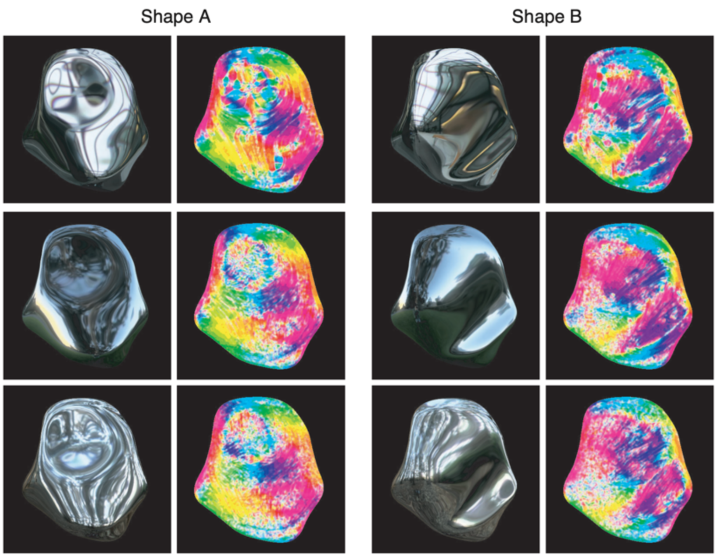


Figure 8. Two shapes rendered with specular BRDF under three different light maps with corresponding visualisations of their orientation fields. Hue denotes peak orientation response and saturation denotes size of peak and provides a cue to the orientation and anisotropy of the principal curvatures, respectively. Both the orientation fields and perceived shape of each object are strikingly stable across variations in illumination. Figure from Fleming et al. (2004).

1.2. Extracting 3D Shape From Shading

The previous section described the generative processes, which are concerned with how the interactions between illumination, material, and 3D shape structure the image received by the viewer. The physics are well understood and the generative problem is mathematically well-posed being a destructive mapping where interactions in a 3D scene collapse onto a 2D image. However, the inverse problem of perception is a constructive mapping, where a visual system or computer algorithm must generate a 3D interpretation based on a 2D input image.

Inverse optics refers to a process by which the visual system reverses the image formation physics to extract the material, shape, and illumination of a scene. This computation involves three variables requiring a priori knowledge of at least two to resolve. The fact that the human visual system operates without any direct knowledge of these variables suggests that inverse optics is an intractable problem, and that the visual system must employ some other cues or strategies. This section will describe the ambiguities and challenges inherent in the inverse problem by reviewing computational work.

It is unclear what the human brain precisely represents about 3D shape. One possibility is that the visual system integrates over local surface orientations to recover shape. This approach is taken by many computer vision algorithms (e.g., Horn, 1970; Ikeuchi & Horn, 1981; Pentland, 1984). However, it is possible that the visual system represents shape differently, and this has inspired other computational approaches that focus on identifying salient and highly informative 'landmarks' within image structure to provide a loose scaffold from which shape estimates may be constructed (e.g., DeCarlo et al., 2003; Judd et al., 2007; Koenderink & van Doorn, 1980; Kunsberg et al., 2018).

1.2.1. Describing shape with local 3D orientation - tilt and slant

Surface orientation may be defined relative to some reference direction, typically the observers' line of sight with two parameters – slant φ and tilt θ. Slant refers to the angle between the line of sight and the surface normal. Described with reference to the image plane, slant is the degree of deviation of the surface tangent plane from frontoparallel. For opaque surfaces, slant is therefore limited to values between 0° and 90°, respectively defining a tangent plane that is frontoparallel and perpendicular to the image plane. Slant is limited at 90° as surface patches greater than this value are necessarily self-occluded by the 'front' of the surface (Figure 9). Tilt describes the orientation of the surface normal's projection onto the image plane measured anti-clockwise from the rightward pointing horizontal vector with values ranging from 0° to 360°. Tilt can also be understood as *slant direction* in that it describes the *direction* that the normal is slanted into the image plane, where slant refers to the *slant* magnitude. In Figure 10, slant increases with the radial distance from the centre and tilt varies with the clockwise rotation from the upward-pointing vector.

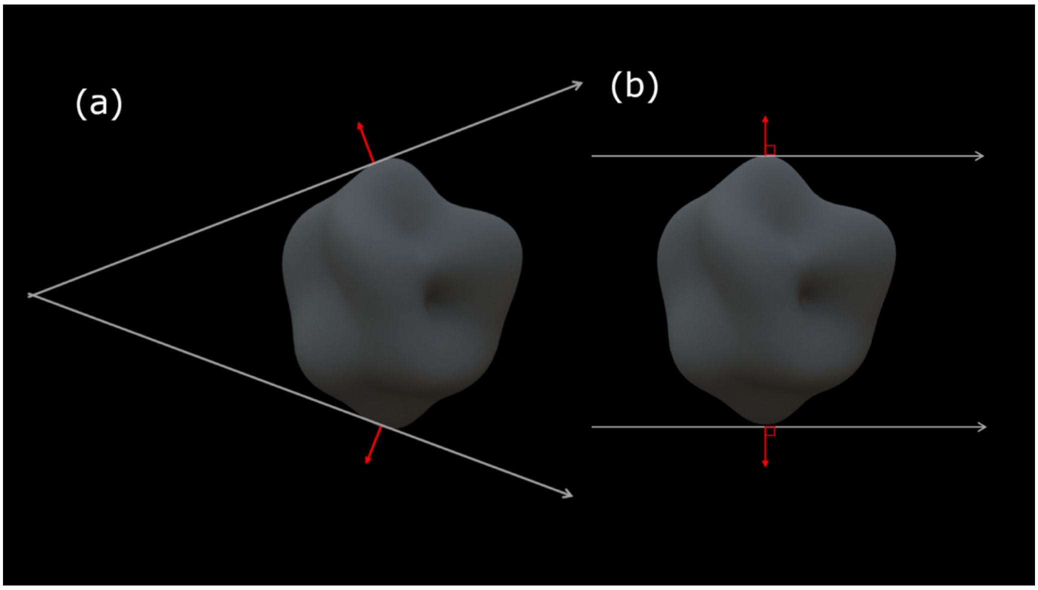


Figure 9. Slant approaches a limit of 90° at the self-occluding rim. The grey arrows denote lines of sight of a view to the left of the shape. The visible boundary of the object is generated by parts of the surface that are tangent to the view vectors. Under (a) perspective projection, the visible boundary is generated by parts of the surface that are in front of the self-occluding rim, and therefore specifies parts of the surface that have a slant angle less than 90°. As viewing distance increases, the image will approach the characteristics of an orthographic projection. Under (b) orthographic projection, the view vectors run perpendicular to the image plane, and the visible boundary therefore occurs at the self-occluding rim where slant is equal to 90°.

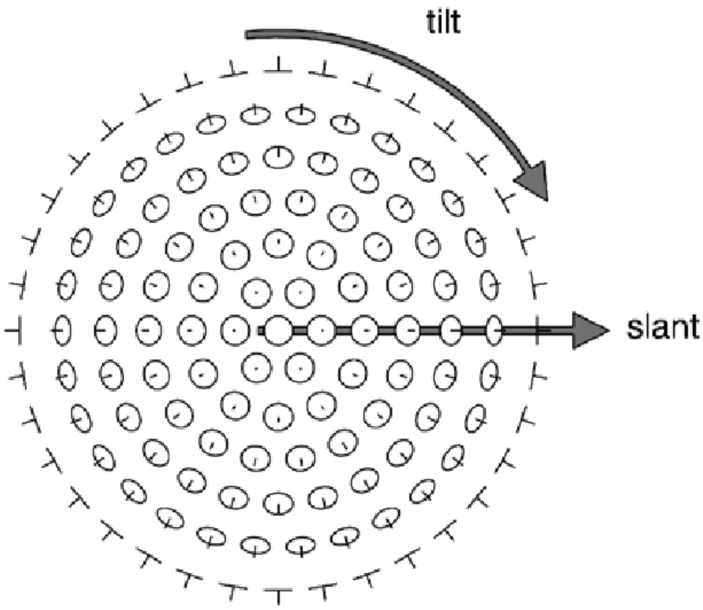


Figure 10. Diagram of surface normals of varying tilt and slant over the full range of the visible hemisphere. Tilt, or *slant direction*, specifies the orientation of the surface normal's projection on the image plane, conventionally measured as the clockwise rotation from the upward-pointing vector. Slant, or *slant magnitude*, is 0° in the centre of the diagram where the normal is frontoparallel to the image plane and increases to a maximum value of 90° (see Figure 9. Slant approaches a limit of 90° at the self-occluding rim. The grey arrows denote lines of sight of a view to the left of the shape. The visible boundary of the object is generated by parts of the surface that are tangent to the view vectors. Under (a) perspective projection, the visible boundary is generated by parts of the surface that are in front of the self-occluding rim, and therefore specifies parts of the surface that have a slant angle less than 90°. As viewing distance increases, the image will approach the characteristics of an orthographic projection. Under (b) orthographic projection, the view vectors run perpendicular to the image plane, and the visible boundary therefore occurs at the self-occluding rim where slant is equal to 90°.) with the radial distance from the centre. Figure from Cornelis et al. (2009).

1.2.2. Ambiguity in the Estimation of Ground-Truth Shape From Shading

Surface orientation cannot be uniquely determined from image intensities, presenting challenges for computer vision algorithms attempting to recover local shape across all points on a visible surface. The intensity of light reflected by an idealised Lambertian surface is proportional to the cosine of the angle between the incident light ray and surface normal. Variations in the angle between illumination and surface can occur due to changes in the surface orientation (i.e., shape) or the illumination direction. Therefore, shape cannot be directly computed from intensity values in the image, as there are infinite pairs of illumination and surface orientations that could generate a given radiance. Even under uniform illumination, radiance does not specify a unique surface orientation. If slant and tilt are specified in relation to the illumination direction (i.e., taking the light vector as the line-of-sight), the luminance of a Lambertian surface depends only on the slant angle between the light ray and surface and is independent of the tilt angle. If we consider a sphere illuminated from a single direction, connecting points of equal luminance will form circles corresponding to a single slant value and each point along these circles corresponds to a tilt value between 0° and 360° (Figure 11). This means that even in situations where the illumination direction is known, only surface slant can be computed from radiance; tilt remaining ambiguous. The anchoring of image intensity to slant values is also complicated by the fact that the intensity of reflected light depends not only on surface attitude, but also the strength of illumination and albedo. Since surface orientation cannot be recovered from image intensity alone, the second order image structure, or shading flow, is analysed. Some strategies include propagation, local, and minimisation approaches (see Zhang et al., 1999 for a review). We consider these approaches next.

Figure 11. Illustration of a shaded sphere with illumination direction parallel to its pole denoted by the black arrow. If surface tilt θ and slant φ is measured relative to illumination direction (i.e., taking the light vector as the line-of-sight), we can see that radiance varies directly with surface slant. The isoluminant contours (isophotes) form concentric circles around the sphere's pole, demonstrating that image intensity is independent of surface tilt. Figure from Marlow et al. (2019).

Propagation Methods

Propagation methods involve an integration over shading flow. Such methods require high-fidelity images, as errors accumulate with increasing distance from points where the computation was initiated. Horn's (1970) early work used singular points where image intensity is maximal. For a Lambertian BRDF, these are points of the surface where the normal is directly facing the illumination and can therefore be computed if the illumination direction is known if we assume that the full range of visible surface normals (i.e., tilt values ranging from 0° to 360°, and slant values ranging from 0° to 90°) are present in the image, and that one of these normal directions are indeed facing the illuminant. These assumptions hold in principle for globally smooth convex surfaces that are not illuminated from behind since globally convex surfaces necessarily contain all possible surface orientations in the visible hemisphere. In practice however, these assumptions may not be adequately satisfied, for example in situations where the surface is largely flat and dramatically slants out of view near the bounding contour. In such instances, a substantial range of slant values may not be salient in the image and thus effectively absent.

If the above assumptions are reasonably satisfied, characteristic strips are computed from these singular points, which are paths of steepest ascent relative to the illumination direction (i.e., paths that follow the direction of maximal shading gradient). Slant values are recovered from image intensity and tilt is equal to the direction the characteristic strip is travelling. Regions in between the characteristic strips are recovered by interpolating between neighbouring strips. Although Horn's method provides a solution under idealised conditions, a limitation of propagation approaches is their high sensitivity to noise and quantisation errors in the image.

Local Methods

Pentland's (1984) approach manages to recover estimates of local surface orientation at each point, however it assumes each point is locally spherical and therefore cannot determine the type of shape (i.e., elliptical, cylindrical, or saddle, see Section 1.4.1.). By assuming a surface is isotropically curved at each point, six independent measures that are directly available from the shading information suffice to constrain recovery of the surface normal, which itself has six degrees of freedom – the tilt and slant of both the surface normal and illumination, the radius of curvature, and the product of illumination intensity and albedo. Pentland (1984) relies on the image intensity I, the first partial derivatives I_x and I_y, and the second partial derivatives I_{xx}, I_{yy}, and I_{xy}. Apart from the inability to recover the type of shape using this approach, there also remains a concave/convex ambiguity since the direction of shading flow may be reversed by either inverting shape convexity or rotating the direction of illumination by 180°.

Minimisation Methods

Minimisation methods work by minimising an energy function over the entire image through a series of iterative computations. They therefore involve a lengthy process that is ill-suited to applications requiring near-instantaneous performance (i.e., perceptual systems). Constraining assumptions are implemented through the energy function to penalise surface reconstructions that violate these assumptions. Ikeuchi and Horn (1981) proposed the first minimisation method and implemented brightness and smoothness constraints in their energy function. The brightness constraint penalises surface reconstructions that produce image intensities that deviate from the input, while the smoothness constraint forces computed surface normals to vary gradually across neighbouring points. A further constraint was implemented by holding the surface normals around the bounding contour constant between iterations where shape could be unambiguously derived. Brooks and Horn (1985) implemented a similar approach that does not require knowledge of the reflectance map by estimating both shape and illumination direction by alternating between shape and light estimation across iterations, moving the light source to minimise the error of the current shape reconstruction. Although the smoothness constraint is necessary to achieve convergence of minimisation methods, it causes the generated shape to deviate from ground truth in many cases. Indiscriminate smoothing causes flattening of the generated surface, introducing significant errors at discontinuities of shading flow. Forcing surface normals to vary gradually over the surface means regions of high curvature, where image intensity is varying quickly, are inaccurately recovered. Furthermore, although Ikeuchi and Horn's (1981) model can exploit information provided by the bounding contour, it is unable to accurately process images containing interior smooth occlusions (e.g., the contour produced by the nose of an obliquely viewed face).

1.3. Computational Approaches to Recovery of Qualitative Shape From Shading

The previous section outlines some of the challenges inherent in computing ground truth shape from shading. Approaches that make some progress on this problem necessitate several restrictive assumptions which limit their applicability in situations where precise knowledge of the rendering function is not known or where shape is to be recovered effectively in real-time. Other approaches constrain the problem by computing families of solutions that are related by some form of geometric equivalence rather than defining shape as a series of discrete surface normals, or by identifying 'landmarks' of image structure which are robust to variations in the rendering function and provide reliable cues to shape. We consider those next.

1.3.1. Shape Equivalence in Euclidean and Affine Geometry

Descriptions of the generative process of image formation are typically based in Euclidean geometry, or in terms of the scene's metric structure. The locations of points in a scene may be described using three coordinates that define their position in space along three dimensional axes (x, y, z). Although some shapes can be explicitly defined in a Euclidean framework, such as a sphere being defined as a locus of points equidistant from its centre, more complex forms are defined in terms of their equivalence with other shapes. For a given geometric framework, shape A is equivalent to shape B if A can be perfectly superimposed on B using a group of permitted transformations. What emerges is a hierarchy of progressively loosened definitions of shape equivalence that permit a larger set of transformations. Of relevance to us is how this describes relationships in 3D, however the same principles may be applied when considering higher-dimensional coordinate systems.

In Euclidean geometry, congruency and similarity provide two definitions of shape equivalence. If a pair of shapes is congruent, one may be perfectly superimposed on the other using only a combination of translation and rotation. Two equally sized cubes may therefore be described as having equivalent shapes, even if their projections on the image plane are different due to their positions in space or due to rotation. However, since the visual system receives 2D images as input, translations of objects in depth are ambiguous. When the visual system encounters an object at a further distance, it needs to be capable of recognising that the diminution of its projected image is a consequence of an increased viewing distance and not a change in its physical properties (i.e., its size). Similarity provides a loosened definition of shape equivalence by further permitting uniform scaling of an object's size. Similarity relates objects in which the distances between respective points are in a fixed ratio.

However, variation in illumination direction often affects perceived depth of the visible surface non-uniformly. The intensity of light reflected by a Lambertian surface is dependent on the slant angle between the surface normal and incident illumination, and therefore does not specify variation in surface slant relative to the observer, with the exception of frontal illumination where the view vector and illumination direction are equivalent. Some changes to perceived shape due to variation in illumination direction may be described by an affine transformation. An example of this is the exaggerated depth perceived in shallow wall carvings illuminated at a grazing angle referred to as the bas relief effect (Belhumeur et al., 1997). In affine geometry, shape equivalence is maintained under the transformations described above for Euclidean geometry, as well as stretching (scaling) or uniform twisting (shearing) in depth. The only characteristics preserved under affine transformation are (1) parallelism, where segments that are parallel in the untransformed geometry remain parallel in the transformed geometry, and (2) the ratios of distances between collinear points.

Although it is impossible to uniquely determine surface shape and illumination direction without prior knowledge of either parameter, Belhumeur et al. (1997) show that the set of shape and illumination pairs that may produce a given image are related by an affine "generalised bas-relief" transformation. Their model assumes an idealised Lambertian BRDF and a single infinitely distant light source (i.e., a scene illuminated by parallel rays). A transformed shape that has been scaled and/or sheared in depth will generate an identical image if accompanied by a corresponding transformation of the illumination direction and surface albedo (see Figure 12). The classic example from which the transform gets its name is the exaggerated depth perceived in shallow carvings (i.e., a simple scaling in depth) illuminated at grazing angles. Furthermore, the positions of cast and attached shadows are preserved under this transformation and is robust to small arbitrary rotations of the surface (Figure 12, second column).

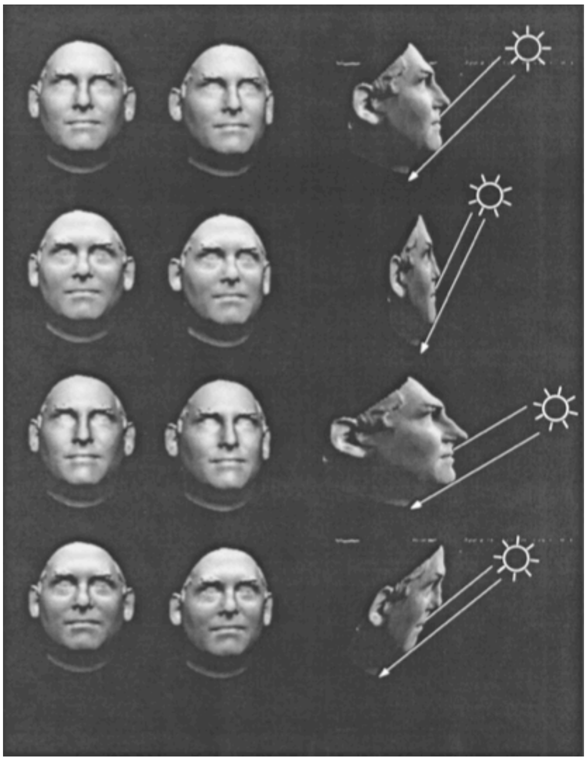


Figure 12. Illustration of the bas-relief ambiguity. The images in the rightmost column depict a sideview showing the scaling and shearing in depth of the surface and corresponding transformation of the illumination. The remaining columns show the images produced from a frontal view (leftmost) and under a small rotation of the surface (middle). These images were generated without transforming surface albedo however they all appear very similar despite the considerably different shapes apparent when viewed side-on. Figure from Belhumeur et al. (1997).

1.3.2. Contours of Image Structure

The bounding contour, or silhouette, of any smooth (i.e., differentiable), globally convex shape is a curve of points where the viewed object smoothly self-occludes such that surface orientation can be directly computed from the image with no or very limited image analysis. Slant along a smooth self-occlusion approaches a limit of 90° depending on the nature of the perspective projection (i.e., the slant at the bounding contour will be less than 90° the smaller the viewing distance) and tilt can be unambiguously determined from the normal direction along the contour at that point.

Zucker (1985) proposes that processing the image globally as an "orientation field" forms the basic input for a visual system. The orientation field is derived from the tangent orientations of shading gradients, which specify the isoluminant contours, or isophotes, in the image. This generates a visualisation akin to a topographic map with the lines indicating level curves of image intensity. This differential representation of the image is related to the differential geometry of the surface. Such a representation describes how a visual system may achieve curve detection (David & Zucker, 1990; Zucker et al., 1989) and account for change in local illumination due to cast shadows (Breton & Zucker, 1996).

There are certain regularities in the derivative of shading that are relatively robust to illumination variation which are reasonable candidates for features that a visual system should be sensitive to. Parabolic lines are a geometric characteristic of surfaces that exhibit several regularities in the shading patterns they generate (Koenderink & van Doorn, 1980). The geometric relevance of parabolic lines is that they connect points along the surface for which one of the principal curvatures is zero (i.e., they curve in only one direction). Parabolic lines segment smooth surfaces into elliptic (convex or concave) and hyperbolic (saddle shaped) patches. Parabolic lines are photometrically specified in the image in several ways. Firstly, local extrema in intensity exist on these curves and will travel along these curves, or be created or destroyed on them, with variations in illumination direction. Parabolic lines demarcate an axis of symmetry of surface geometry by segmenting regions of changing curvature. This symmetry in surface geometry is reflected in symmetry of shading flow (i.e., converging or diverging shading flow) under a wide range of illumination conditions. Secondly, the isoluminant contours, or isophotes, intersect parabolic lines at a constant angle (see Figure 13). Although parabolic lines are a valuable cue in that they relate invariant characteristics of surface geometry with invariant characteristics of shading, it is difficult to determine whether this is a cue exploited by the human visual system. Although there is evidence suggesting humans are capable of implicitly locating parabolic lines in identifying the extent of a convex bump or concave dimple, both parabolic lines and saddle regions are not perceptually salient and seems to be something only experienced psychophysical observers are capable of reliably reporting (Koenderink et al., 2015).

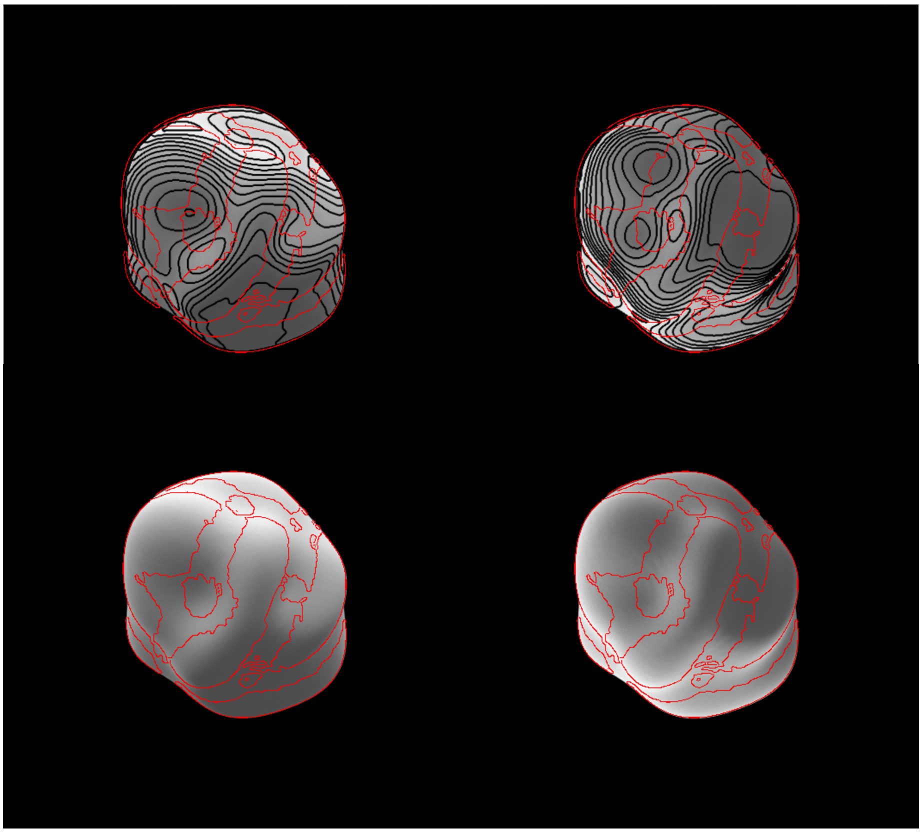


Figure 13. Pair of identical shapes rendered with diffuse reflectance and different illumination directions. The parabolic lines (in red) are a feature of the surface geometry that connect points for which one of the principal curvatures is zero. The shading, which varies with changes in illumination direction, is visualised with isophotes (in black). Sets of isophotes intersect with the parabolic lines at a constant angle, demonstrating a direct relationship between image structure and surface geometry. The same images are presented in the bottom row with parabolic lines overlaid only on the shading pattern without isophotes to demonstrate that parabolic lines are not perceptually salient.

Bounding contours are highly informative and perceptually salient image features. This has prompted several theories to consider features within the bounding contour that approximate the behaviour of bounding contours as an additional source of reliable shape information. DeCarlo et al. (2003) demonstrated that although line drawings that are generated based on the regions where the surface slants out of sight, or self-occludes, can effectively convey coarser shape features, information about finer features that would be captured by a human illustrator are absent. The addition of ‘suggestive contours’, which are self-occlusions that would emerge with a small rotation to the object, produces a compelling impression of 3D shape using only line drawing (see Figure 14). Apparent ridges, developed by Judd et al. (2007), consider lines of maximal view-dependent curvature. This differs from ordinary surface curvature in that it specifically considers the rate of change in the surface normals relative to the image plane. Both suggestive contours and apparent ridges typically occur where surface slant is high and can therefore generally be identified in the image itself from the steep shading gradients they produce without relying on prior knowledge of surface geometry. The capacity for simple line drawings generated using suggestive contours and apparent ridges to compellingly convey 3D shape indicates the importance of these features in recovering 3D shape from 2D images. These techniques suffice to convey familiar 3D forms such as faces, however unfamiliar forms may appear ambiguous or cluttered.

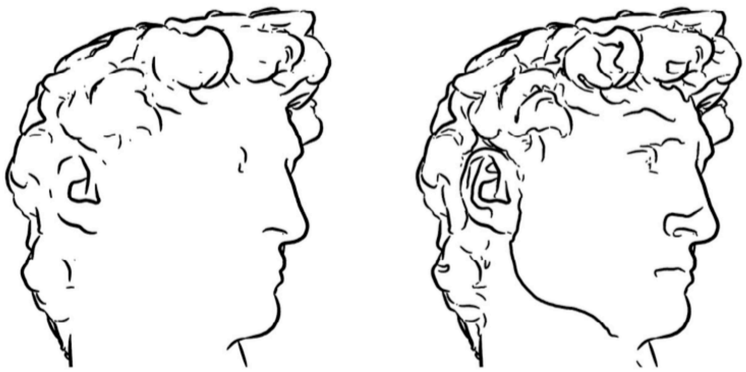


Figure 14. Line drawings tracing smooth self-occlusions only (left), and with the inclusion of suggestive contours (right). Suggestive contours convey information about shape features that would become a smooth self-occlusion following a small rotation. Smooth self-occlusions alone convey only vague shape information, and the image on the left is only recognisable as a face due to our particular atunement to identifying faces. The depiction of features *approaching* a self-occlusion makes shape far more apparent without cluttering the image with irrelevant detail. Image from DeCarlo et al. (2003).

The highly-informative image structure contained within the bounding contour demonstrated by DeCarlo et al.'s (2003) suggestive contours inspired Kunsberg and Zucker to consider how certain invariant features of shading flow organise qualitative shape globally. They approach the problem of visual perception with the assumption that certain regions of an image are more informative than others. Kunsberg and Zucker (2018) have proposed critical contours as an important invariant, related to bounding contours in that they trace curves of maximal surface slant. Their approach involves a topological representation of surface geometry using the Morse-Smale (MS) complex (Figure 15). The lines (1-cells) of the MS complex are integral curves (i.e., they trace paths that monotonically decrease in depth) connecting local maxima of the surface slant function to saddle points, and saddle points to local minima. The 1-cells form quadrilateral regions (2-cells) that define the descending manifold between the local maximum and minimum. The 1-cells can be understood as a ridge along a mountain range, and the significance of the 2-cells can be appreciated by imagining the paths that water dropped on a local maximum peak would follow. Most of the water would follow an integral path along the descending manifold directly to the nearest local minimum, but a small proportion would follow the ridge line to reach a saddle point. These 1-cells connect singular points along the surface gradient function, forming the boundary between adjacent descending manifolds. A slight disturbance to the water's path along a 1-cell changes its course from the connected saddle point to one of two different local minima. Critical contours are the subset of 1-cells that (a) connect local maxima of the slant function to its saddle points and (b) the shading gradient orthogonal to the 1-cell is sufficiently steep. Critical contours are perceptually salient, forming steep, symmetric shading gradients which vary little under different illuminations. Critical contours correspond to regions of high surface slant, and in this sense approximate the behaviour of bounding contours, providing quasi-self-occlusions within the object. Kunsberg and Zucker's (2021) qualitative approach to shape from shading also predicts aspects of human vision such as bistable perception of convex or concave regions.

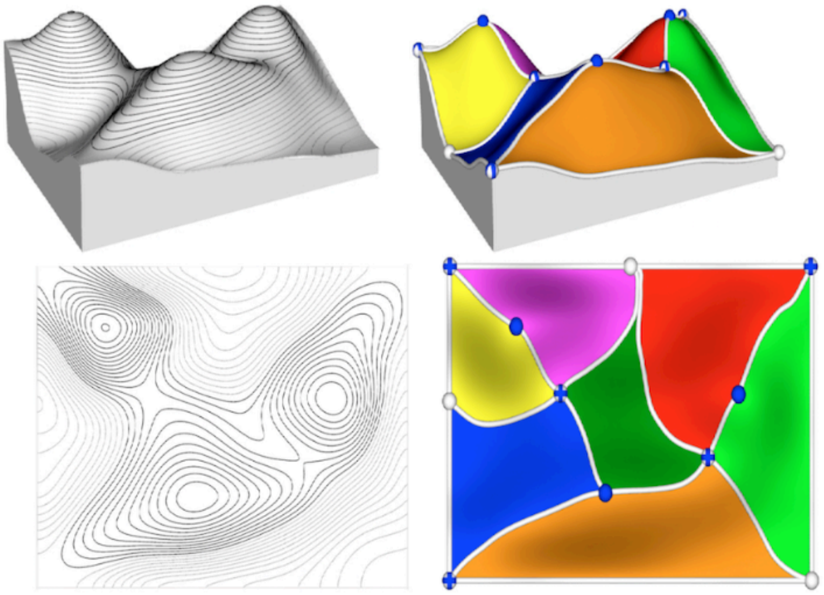


Figure 15. A bumpy surface with contours denoting level sets in depth (left column) and its MS complex (right column). The maxima and minima (solid blue and white points, respectively) are connected to saddle points (blue-white crosses) by 1-cells (white contours), which segment the image into 2-cells of the MS complex (coloured regions). The 2-cells define the ascending/descending manifold composed of integral curves connecting the 2-cell's local maximum and minimum. A ball dropped on any part of the 2-cell will follow a particular integral curve toward the local minimum. Figure from Kunsberg and Zucker (2018).

1.4. Human Perception of Shape

Computer vision research first formalised the shape from shading problem and has made considerable advances in recovering 3D shape from 2D images. Although many computer vision algorithms can recover shape with impressive fidelity, in the most restrictive cases they require strong assumptions about the generative conditions, and in cases where these assumptions are weaker, they require significant computation time, particularly for iterative solutions like minimisation techniques. This is at odds with the human visual system that routinely resolves this problem with limited a priori knowledge and effectively in real-time. This section will review evidence for the capacity of the human visual system and the cues that it exploits to derive shape from shading. First, we introduce Koenderink and van Doorn's (1992) 'shape index', which provides a succinct description of the qualitative nature and metric curvature of a surface patch in terms relevant to human observers. Second, we review psychophysical research of human shape perception. Third, we review more recent work applying Bayesian methods, suggesting that sufficient information is present in images themselves without necessitating a priori assumptions about the generative function.

1.4.1. Curvature and Shape Index

Tilt and slant are useful parameters as they describe surface orientation in image-based coordinates relative to some reference vector, typically the line of sight of the camera or observer. The values of these parameters describe the image as seen by the observer, such as where the self-occluding boundaries exist. However, tilt and slant do not describe what people typically mean by 'shapes' like round bumps or elongated furrows. Curvature refers to the second order behaviour of a surface, or the rate of change of orientation over some region. It is therefore defined with reference to local surface geometry and does not vary with changes in viewpoint.

The curvature of a 2D curve is typically defined as the reciprocal of the radius of the circle that best fits the curve at that point (an 'osculating' circle). Directional curvatures can be similarly calculated on a 3D surface by intersecting the surface with a plane rotating about the surface normal at that point. The principal curvatures k_1 and k_2 specify the directions and magnitudes of maximum and minimum curvature passing through that point and suffice to describe a 3D surface patch up to the second order. The product of the principal curvatures gives the Gaussian curvature K, which is an intrinsic measure of curvature that depends only on the surface geometry (see Figure 16). When the principal curvatures have the same sign, the Gaussian curvature is positive, and describes a concave 'cup' (k_1 and $k_2 < 0$; green) or convex 'cap' (k_1 and $k_2 > 0$; red). Cups and caps are broadly categorised as elliptic points and the singular case of a concave or convex sphere where the surface is equally curved in all directions are umbilic points. Negative Gaussian curvature describes a saddle or hyperbolic point (white), where one principal curvature is concave and the other is convex. A zero

Gaussian curvature can describe a cylindrical rut (blue) or ridge (yellow) if $k_1 = 0$ and $k_2 < 0$ or $k_2 > 0$, respectively, or a plane where $k_1 = k_2 = 0$.

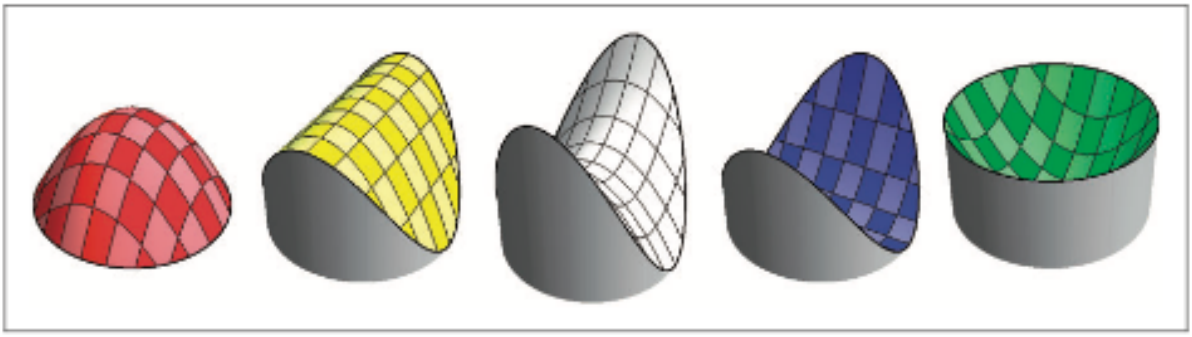


Figure 16. Illustrations of shape categories described by Gaussian curvature *K*. From the left, images show a convex umbilic or elliptic point ($K > 0$), convex cylinder ($K = 0$), saddle ($K < 0$), concave cylinder ($K = 0$), and concave umbilic or elliptic point ($K > 0$). Figure from Koenderink et al. (2015).

To describe shape in a manner relevant to human observers, a scale must be designed that specifies the continuous transformation from one canonical shape to another, for example from a spherical cup to a curved trough to a flat cylindrical rut. By the above descriptions of curvature, a larger sphere has a lower curvature than a smaller sphere since its surface normals vary more gradually across a given distance over its surface. Such a descriptor should therefore be scale-invariant to capture the commonality between shapes of different sizes and amplitudes. Koenderink and van Doorn (1992) address these issues by describing local surface geometry in terms of 'shape index' s and 'curvedness' c. The shape index is derived from the ratio of the principal curvatures and so is independent of curvature magnitude, or curvedness. These measures can be visualised as a polar coordinate system in the $k_1 - k_2$ plane with the shape index given by ray direction and curvedness given by radius (Figure 17). Specifically, the shape index is calculated from the angle between the ray and the $k_1 + k_2 = 0$ axis according to the equation:

$$s = \frac{2}{\pi}\tan^{-1}\frac{k_2 + k_1}{k_2 - k_1} \quad (k_1 \geq k_2)$$

The shape index ranges between -1 and 1 with positive values denoting convex shapes and negative values denoting concave shapes. Shapes of equal index magnitude that vary only in sign form complementary pairs that fit together as a convex ‘stamp’ and concave ‘mould’. The umbilic points are equally curved in all directions and are represented at the extremes of the scale. Since any arbitrary deformation of an umbilic point necessarily moves it toward a cylinder, these are true endpoints of the shape index scale. The cylinders are represented by $|s| = 0.5$, and the transitional saddle, which is neither concave nor convex, is represented at the midpoint of the scale by $s = 0$. Saddle points are unique in that they are congruent to their own moulds, requiring only a 90° rotation about the surface normal of the ‘stamp’ to fit its inverted ‘mould’. The plane has a curvedness of zero and an indeterminate shape index, as it is represented by the origin of the $k_1 - k_2$ plane. Psychophysical experiments have demonstrated that shape index and curvedness scales are intuitive for human observers (Koenderink et al., 2014, 2015).

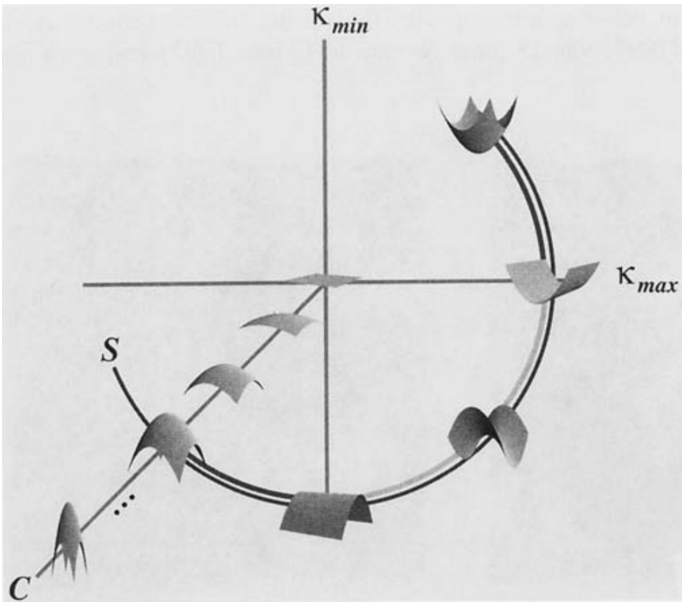


Figure 17. 'Shape index' and 'curvedness' scales depicted as a polar coordinate system. The angle denotes the smooth transition between qualitative ‘types’ of curvature independent of scale (shape index), and radial displacement from the centre denotes curvature magnitude (curvedness). Image from Phillips and Todd (1996).

1.4.2. Role of Illumination in Shape From Shading

Early work on human vision used highly simplified stimuli, often produced under a single distant point light source and a uniform diffuse component to ensure all parts of the surface are illuminated. The conventional stimulus is the simplest example of a shape from shading stimulus that can be generated without need for physics-based rendering. It consists of a smooth linear gradient from either end of the stimulus, typically bounded by a circular contour. As discussed above, this stimulus can be accurately interpreted as any class of curved shape (sphere, saddle, and cylinder), though it is typically experienced as either a convex or concave spherical cap or cup. Ramachandran (1988) used this stimulus to demonstrate the light-from-above prior in that a conventional stimulus that is light to dark from top to bottom is typically perceived as convex, and the concave interpretation is more common when the polarity of the gradient is reversed. These experiences are consistent with illumination directed from overhead. However, this pattern of perception is far from universal and many observers perceive the image as consistently convex regardless of the direction of the intensity gradient. This is supported by a convexity prior demonstrated by Langer and Bulthoff (2001), and it may be that the relative strengths of these priors vary between observers. Idiosyncratic effects like these may inordinately contribute to the perception of overly-simplified stimuli and not describe visual perception generically. Indeed, the light-from-above prior is far from ubiquitous and priors on illumination direction are rapidly updated based on recent experience (Adams et al., 2004).

The use of stimuli generated under constrained and simplified conditions raises the question of the generalisability of these findings. In some sense, these simplified stimuli complicate the recovery of 3D shape for the visual system. For most illumination orientations, illumination from a single light source means that portions of the surface will be shadowed resulting in the absence of image structure that the visual system could exploit. For example, including a diffuse component of illumination may assist disambiguation of convex and concave features by applying a “dark-is-deep” rule (Langer & Zucker, 1994). Since the image intensity of a surface depends on the aperture (the proportion of the light field visible at each point) rather than surface normal under diffuse lighting, concavities are biased to appear darker than convexities. The inverse of this is also true that maxima of image intensity tends to be perceived as the points of highest surface relief, explaining the shearing of shape under oblique illumination described earlier (Christou & Koenderink, 1997, see Figure 12). With more complex stimuli like human torsos, variation in illumination direction were found to affect perceived shape in a piecewise manner, presumably due to the appearance of cast shadows under more oblique illumination directions (Todd et al., 1996). There is evidence to suggest that humans are sensitive to the directional and diffuse components of illumination (Morgenstern et al., 2014) and that humans typically perform better on tasks when the surface is illuminated by mixed directional and diffuse light (Schofield et al., 2011). Most vision research now uses spherical images of natural environments (light maps) to provide illumination for computer renderings. Although the complexity of the illuminant is not particularly important for realistic renderings of Lambertian materials (Ramamoorthi & Hanrahan, 2001), they are essential for the generation of compelling specular materials, which we will discuss in a coming section.

Patterns of shading may be represented as both a scalar field of varying image intensities (i.e., shading), or as a vector field of shading *flow*. The scalar field is generated by the normal component of the illumination direction and the vector field is generated by the tangent component for surfaces of varying orientation. Koenderink et al. (2007) refer to this vector field as "surface illuminance flow" and argue that in conjunction with the scalar field, provide information about the tangent and normal components of illumination direction. Illuminance flow may be estimated using either the squared gradient or Hessian of image intensity, which correspond to edge and line detectors present in visual cortex (Hanazawa & Komatsu, 2001). Human observers are sensitive to the illuminance flow cue (Koenderink et al., 2004), however estimates of illumination direction are often highly biased when viewing surfaces with anisotropic roughness (i.e., containing bumps that are stretched along one direction, Koenderink et al., 2007). Highly anisotropic shape features are approximately cylindrical in form, curving maximally in one direction and minimally in the orthogonal direction. These features generate shading gradients only in the direction of curvature, biasing estimates of the light vector toward the direction of the gradient. These findings raise important questions about interactions between surface geometry and the orientation of the illuminant, and whether changes to perception depend on inaccurate estimates of the illumination direction (i.e., caused by inaccuracy in an operation attempting to invert the generative function) or through a direct analysis of image gradients.

1.4.3. Role of Contours in Human Perception

Many computational approaches to shape-from-shading attempt to recover surface orientation from variation in luminance across arbitrary points in the image. Later approaches recognised that specific regions provide more reliable shape information, which may specify a scaffold from which shape may be estimated, namely smooth self-occlusions and regions that approximate their behaviour. Here we review the psychophysical evidence for the role these features play in human perception, examining how the visual system identifies a bounding contour as being caused by a smooth self-occlusion, and how shading along these contours influences perceived shape, material, and optical focus.

Several findings in the psychophysical literature suggest that the human visual system primarily uses a loose scaffold of prominent landmarks between which variations in surface orientation are estimated from shading flow. The segmentation of the surface into qualitatively distinct regions provided by image structure such as Kunsberg and Zucker's (2018) critical contours, predicts how and where ambiguities in shape-from-shading are perceived, such as the convex/concave ambiguity, and the inability of human observers to reliably extract metric depth (Todd & Reichel, 1989). Work by Mamassian and Kersten (1996) supports the view that human perception relies on only certain 'landmarks' in shading flow with much of the remaining structure seemingly irrelevant. Their observers provided local measurements of surface orientation and found highly similar judgements whether they viewed fully shaded images or a uniform grey silhouette. Observers in either condition appeared to simply interpolate surface orientation through the bounding contour where slant is unambiguously maximal, to somewhere in the figure where surface orientation would be frontoparallel before curving away again toward the bounding contour on the opposite side with minimal reference to the shading itself. For more complex surface geometries, perception may operate in a similar manner interpolating surface orientation between prominent landmarks in the shading flow without close attention to the point-by-point variation in image intensity. Rather than processing the metric orientation of a surface (i.e., a surface normal), there is evidence that human perception is organised by local, ordinal relations, or that we are sensitive to which point is deeper than another, and not the magnitude of the depth gradient (Todd & Reichel, 1989). This was demonstrated by poorer accuracy in relative depth judgements when there is a non-monotonic depth change between the compared points (e.g., comparing the depth of points on either side of a ridge).

To exploit smooth self-occlusions as a source of unambiguous shape information, the visual system must firstly determine whether contours in the image are indeed caused by a smooth self-occlusion. Other causes for contours in images are surface pigmentation (i.e., variation in albedo), occlusion by another surface (in which case, border ownership must be correctly assigned to the occluding surface), or a discontinuity in the surface due to a cut or sharp edge, or when viewing the bounding contour of a globally concave surface (e.g., a top-view of a cup). Smooth self-occlusions of a Lambertian surface may be identified by combining the two constraints that (1) surface orientation is unambiguously determined along a smooth self-occlusion where slant is constant (approximately 90°) and tilt is equal to the orientation of the outward pointing normal at each point, and (2) the intensity of reflected light declines monotonically as the surface normal faces away from the illumination direction. Given these two constraints, intensity along smooth self-occlusions will covary for most directional illuminations except when the illumination is aligned with the viewing axis (Marlow et al., 2019). The use of this photogeometric cue by the visual system is supported by the percepts elicited by stimuli that independently vary the frequencies of a sinusoidal intensity grating and bounding contour. Along the diagonal of Figure 18, contour orientation and intensity covary, resulting in a compelling percept of a 3D screw shape with Lambertian reflectance. The weakening or absence of this covariation in the off-diagonal cells generates a poor impression of both shape and Lambertian reflectance with the intensity variation appearing like changes in albedo.

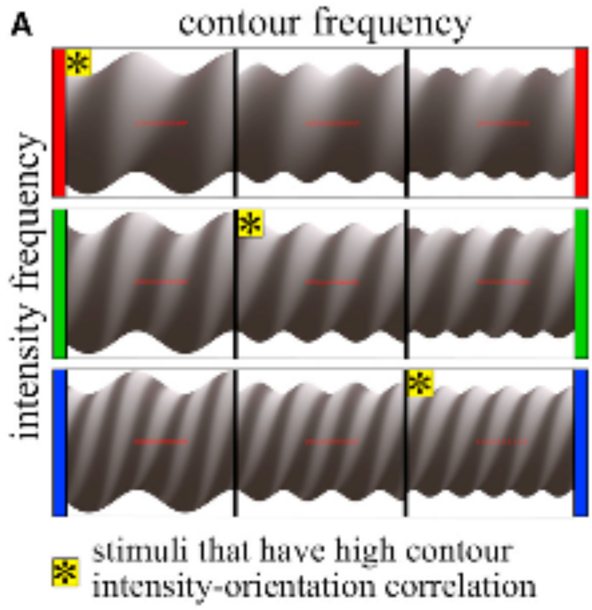


Figure 18. Effect of contour intensity-orientation correlation on shape and material perception. Frequency of sinusoidal contour varies along the columns and frequency of intensity variation varies along the rows. The asterisked diagonal presents stimuli with high contour intensity-orientation correlation and generate compelling percepts of a 3D screw shape with Lambertian reflectance that are largely absent in the remaining cells. Figure from Marlow et al. (2019).

The presence of smooth occlusions, or identification of features that approximate their behaviour, has been shown to provide constraints that are exploited by the visual system. Nefs et al. (2005) compared stability of perceived shape illuminated from different directions. Although non-affine changes in perceived shape were observed in all conditions, they were significantly reduced when the object was viewed from a direction that introduced smooth self-occlusions within their bounding contour. The complete absence of smooth self-occlusions, such as when viewing a bumpy plane from above, often produces an image that lacks definition of both shape and material, appearing defocussed or like a cloudy noise texture (Mooney et al., 2019). When these images are segmented by a planar cut parallel to the image plane, the remaining image structure is brought into sharp focus with compelling impressions of surface material and shape. Planar cuts differ from smooth self-occlusions in that they specify a locus of points of uniform *depth* but the surface slant may not be fixed, however they approximate the same orientation-intensity covariation as smooth self-occlusions for a wide range of rendering functions. Specular materials typically generate steeper gradients than Lambertian materials, depending on the degree of specular roughness and as such, identical intensity gradients can be perceived as either being generated by a Lambertian or specular surface when surface intensity along the contour varies faster than the cosine (Marlow et al., 2015, see Figure 19). Since smooth self-occlusions are generated where the surface curves out of view, shape along a smooth self-occlusion must either be saddle shaped (where the 2D contour is concave) or convex (where the 2D contour is convex). This constraint is not preserved for planar cuts where parts of the contour may also be formed by concavities.

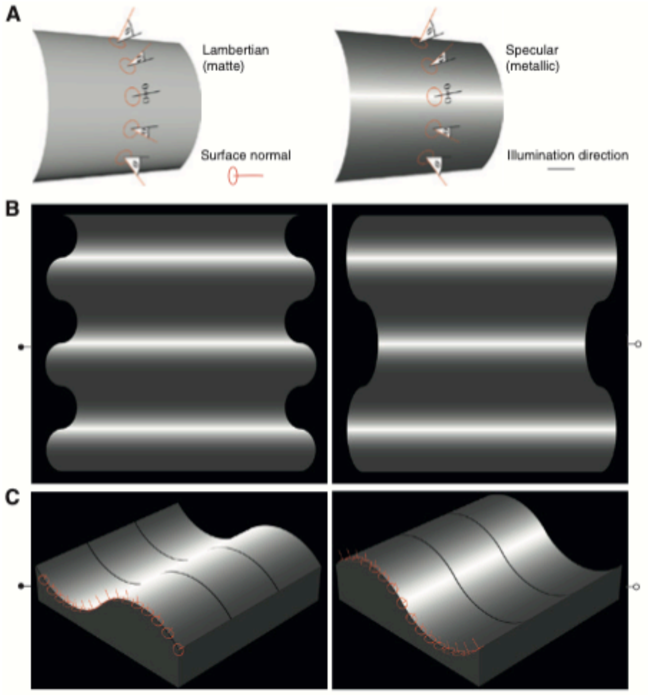


Figure 19. Varying the shape of the bounding contour affects perceived material and shape from identical intensity gradients. When intensity along the bounding contour varies with the cosine of contour orientation (left), the material appears Lambertian. When this covariation occurs at a rate that exceeds the cosine of contour orientation (right), the material appears specular. Figure from Marlow et al. (2015).

1.4.4. Shape Perception of Specular Materials

For natural illumination, specular reflections are systematically distorted by surface curvature, generating a texture-like pattern from the relative compression and stretching of reflections along the axes of maximal and minimal curvatures, respectively (Fleming et al., 2004, see Section 1.1.2.). Such cues offer a parsimonious mechanism by which the visual system extracts shape that does not require high-level knowledge of the illuminating environment and complex inverse computation. Furthermore, the mechanisms for extracting shape from specular reflections in this manner have well-established perceptual and neurophysiological models from work on the recovery of shape from texture (e.g., Gomez & Neumann, 2016). However, there are situations where it is not possible to extract shape from the deformations of specular reflections. For example, these cues do not stabilise the perception of specular bumpy planes, which exhibit no definitively discernible shape or material (Marlow et al., 2019; Marlow & Anderson, 2021), suggesting a necessity of bounding contours. The importance of bounding contours is further supported as Savarese, Fei-Fei, and Perona (2004) demonstrated that shape perception accuracy was only slightly better than chance, despite using highly contrived environments featuring regular patterns such as checkerboards and gridlines.

Specular reflections often occur in materials exhibiting a combination of reflectance functions called dielectric materials. Dielectric materials (e.g., plastic) reflect a portion of incident light at the surface interface (i.e., specularly) with the remaining portion penetrating the surface and undergoing subsurface scattering. They can be modelled by a weighted combination of a specular and diffuse component. Both types of reflectance often co-occur in natural materials, and the visual system may reasonably be sensitive to how they co-specify the underlying surface geometry. Since diffuse reflectance is dependent on the first derivative of shape and structured primarily by the low-order structure of the illuminant (Ramamoorthi & Hanrahan, 2001) and specular reflectance is dependent on the second derivative of shape and captures high-order structure of the illuminant, it is logical to expect that such materials would provide the most robust shape inferences in their cospecification of surface geometry. The combination of smooth shading with specular reflections has been found to promote the perception of shape and sharp focus, which is considerably weaker when viewing either component in isolation (Marlow & Anderson, 2021). However, evidence for this view is somewhat mixed. Both Norman et al. (2004) and Nefs et al. (2006) found no difference in shape judgements for matte and glossy objects. A possible explanation for these results is that their glossy stimuli only contained isolated specular highlights, omitting much of the structure present in stimuli rendered using natural light maps. Though the inclusion of specular highlights exaggerated perceived curvature (Todd & Mingolla, 1983) and improved the accuracy of judgements of the sign of curvature (convex or concave, Liu & Todd, 2004), these studies used simple cylinder or bump stimuli which may limit the generalisability of these findings. Furthermore, images rendered using natural light fields generate considerably more compelling impressions of gloss (Fleming et al., 2004). Early work primarily used simple area or point light sources, which only generate isolated specular highlights on rendered surfaces, lacking the rich source of information present in the pattern of deformation of a reflected natural environment (see Fleming et al., 2004). There is also evidence that beyond specular highlights, the visual system is sensitive to information contained in low-intensity specular reflections, or the "dark side of gloss" (Kim et al., 2012).

Further interactions between gloss and shape are observed when considering specular roughness. Introducing microscale roughness to a specular surface progressively blurs specular reflections due to small variations in local surface orientation, which can resemble diffuse reflectance in the limit (see Section 1.1.2.). Since smooth specular materials behave like a mirror, specular highlights form the highest intensity portions of an image, and their high contrast and sharpness provide strong cues for gloss (Marlow et al., 2012), and introducing specular roughness weakens these cues. Mooney and Anderson (2014) further demonstrated an interaction between specular roughness and ground truth surface curvature on shape perception. Blurred specular reflections may be misperceived as diffuse shading, leading to exaggeration of perceived surface relief. High curvature regions of a Lambertian surface generate steep shading gradients, increasing both their sharpness and contrast. In Mooney and Anderson's (2014) experiments, a high roughness specular component was apparently misperceived as diffuse shading, resulting in exaggerated judgements of surface relief. Furthermore, for stimuli with increased curvature, and therefore steeper diffuse shading, increased surface relief was also perceived with moderate specular roughness, indicating an interaction with surface geometry in the misperception of gloss as diffuse shading.

1.4.5. Bayesian Model of Visual Perception

Most work in computer vision relies on assumptions to constrain the inherently ill-posed problem of inverse optics (e.g., assuming known reflectance or illumination properties). Such assumptions are particularly problematic in the study of natural visual systems as it begs the question of how such prior knowledge could be acquired by the organism. Arguing that priors are innate and acquired over evolutionary history only defers the question as there is still the problem of how members of the species iteratively approximate the priors required for adaptive perception. The information necessary to shape a visual system must be present in some form within the image itself. Most of our visual perception interacts with images exclusively and opportunities to assess the accuracy of our percepts through direct exploration of the generative factors are highly uncommon (however there is evidence that the development of some perceptual abilities coincides with infants' motor development, e.g., Granrud et al., 1985; Yang et al., 2011). Bayesian approaches work with the assumption that perceptual abilities like identifying a surface's shape or material operates by identifying statistical regularities across images.

Bayesian theory has been applied with some success in modelling how the visual system may detect and exploit statistical regularities across scenes to effectively parse visual information to determine the most probable perceptual 'hypothesis' that explains the stimulus. Research promoting Bayesian models for visual perception rely on comparisons between the performance of Bayesian ideal observers and human observers. Bayesian models consist of three terms: the prior, the likelihood, and the posterior. The prior term, $P(H)$, contains what is known about the environment *prior* to viewing the stimulus, or evidence for candidate hypotheses, and can be understood as the relative frequencies of all possible scenes encoded by the organism through its experience or genetics. The prior term penalises scene properties that are infrequently encountered by the organism. The likelihood, $P(E|H)$, describes the conditional probability that an assumed hypothesis *H* would generate the image *E*. The likelihood term penalises degenerate hypotheses that could only produce the image under very specific or accidental viewing conditions, consistent with Freeman's (1994) 'generic viewpoint assumption'. For an estimate of the distal stimulus *H* to be probable, it must be robust to small changes in viewpoint and illumination direction (i.e., it should not be an 'accidental' image of the scene). In many situations, several perceptual hypotheses could be equally likely given probability distributions of the generative factors. In these cases, a generic viewpoint assumption penalises solutions that would generate substantially different images under small variations in the rendering function.

Bayesian models have yielded results consistent with many characteristics of the visual system. In terms of physiology, sensitivity ranges of cone receptors are closely predicted by Bayesian models and are well positioned to discriminate types of foliage for dichromatic animals (Lythgoe & Partridge, 1989) and identifying fruit in foliage for trichromatic primates (Regan et al., 2001). Freeman and Viola (1997) presented a series of simple synthetic images varying in subjective 'shadedness' or 'paintedness' by manipulating their sharpness and contrast and found high agreement in image rankings of human observers and a Bayesian ideal observer. Bayesian models also provide a parsimonious account of certain motion illusions. The motion of a translating rhombus is more likely to be misperceived as a diagonal direction with lower contrast and widths, which could be explained through an arbitrary application of different rules. However, a Bayesian observer with an a priori preference for slower velocities is able to qualitatively predict the occurrence of this motion illusion (Weiss et al., 2002). Bayesian models provide a mechanism by which prior knowledge can be updated by new information, accounting for the role of recent experience in the perception of stimuli presented serially. Judgements for a range of visual features demonstrate short-term attractive serial dependence (i.e., stimulus judgements are biased toward characteristics of recently-viewed stimuli) and long-term repulsive adaptation (see Fritsche et al., 2020 for a review). Fritsche et al. (2020) propose a mixed Bayesian and efficient observer model to account for these effects in which the longer-term repulsion arises from the recruitment of sensory resources prioritising efficient encoding of a typically stable environment, influencing the likelihood distribution, and shorter-term attraction represents an optimal strategy of decoding sensory information to promote perceptual stability, influencing the prior distribution.

Bayesian theories of perception adopt a computational approach in which the model is applied to optimise the performance of a given task or 'computation' (e.g., material categorisation, shape judgement, etc.) A problem with this approach is that Bayesian theory does not provide the hypothesis space itself making it unclear how the prior and likelihood distributions arise in the first place. Geisler and Diehl (2002) modelled how Bayesian decision theory can explain the natural selection of perceptual tasks. However, the definition of such natural tasks relies on the post hoc rationalisation that our ability to perform a task, and justification for how performance of this task improves fitness, proves that they are a direct consequence of natural selection. This approach presupposes some 'understanding' on the part of the organism of the generative process of image formation within the likelihood distribution, as well as definition of the problems the visual system is tasked with solving from the outset (Anderson, 2015). This prerequisite information cannot be directly ascertained from the input itself. Similar ideas are expressed by Gibson's (1979) theory of direct perception and Koenderink et al.'s (2012) "view from the inside". Gibson argued that images themselves contain sufficient information from which our environment can be perceived "directly" without need for priors or drawing inferences about the generative parameters. Similarly, the view from the inside contends that since visual systems evolved without knowledge of the generative physics, they should also be described in terms that do not invoke them.

Bayesian theories provide a compelling description for the perception of highly ambiguous stimuli, accounting for phenomena like the light-from-above prior and the effects of serially viewed stimuli. However, the computationally expensive and probabilistic models do not accord with our typically immediate and stable visual perception. Luccio (2019) proposed the separation of visual perception into primary and secondary processes. Primary processes govern the overwhelming majority of visual perception and provide the first and immediate segmentation of the perceptual field, typically producing a stable experience of the environment that does not necessitate probabilistic models. Secondary processes govern isolated instances where the input is highly ambiguous and are driven by cognitive processes, accounting for our ability to consciously ‘flip’ interpretations of bistable images, for example. It is more reasonable to suppose that Bayesian models account for these isolated cases once the general mechanisms of visual perception have been established by other means.

Chapter 2. The Role of Surface Geometry-Illumination Alignment on Shape Constancy

This thesis will examine conditions predictive of failures of shape constancy and describe the nature of these failures. Previous work demonstrated that perceived shape is often sheared in the direction of the illuminant (Christou & Koenderink, 1997), which is an affine transformation. Non-affine shape transformations have also been observed (e.g., Nefs et al., 2005), which tended to occur for stimuli lacking smooth self-occlusions. However, the specific nature of non-affine failures of shape constancy remains to be examined. We will investigate whether features of image structure may predict successes and failures of shape constancy directly, without necessitating perceptual priors or estimation of distal parameters.

Experiment 1

Experiment 1 used an exploratory approach to identify relationships between illumination direction and surface geometry that are predictive of failures of shape constancy. 'Potato' objects generated by randomly perturbing a sphere were rendered with Lambertian reflectance and a series of equally spaced illumination directions spanning the entire illumination sphere of possible directions. Low and high curvature versions of the shapes were included to test the stability of perceived shape for different surface curvatures. Highly curved surfaces generate steeper shading gradients and increase the likelihood of cast shadows and vignetting that may make the perception of prominent surface features more robust to variation in illumination direction. We also performed image analysis on our stimuli. Images were convolved using a filter selective for shading that was symmetrical (i.e., converging or diverging) and of high contrast. This filter is selective for features resembling Kunsberg and Zucker's (2018) critical contours, allowing us to test how well critical contours predict the judgements of human observers.

Methods

Observers

For the low curvature sample, twenty-four observers were recruited and given up to 60 minutes to complete the experiment using the low curvature stimuli described below. Observers provided informed written consent and were compensated with course credit. Research ethics was approved by the University of Sydney Human Research Ethics Committee. The final sample for analysis was reduced to ten, removing observers who completed less than 90% ($N = 12$) of the experiment and/or correctly responded to fewer than 8/10 of the catch trials ($N = 2$) described in the Procedure section. Twenty additional observers were recruited as above to view the high curvature stimuli. Fourteen observers who completed less than 90% of the experiment were removed from the analysis sample. The remaining six observers responded satisfactorily to the catch trials. Possible reasons for this substantial attrition will be addressed in the discussion.

Stimuli

The stimuli used in Experiment 1 were a set of three randomly perturbed spheres. The stimulus geometry was designed using a script written in MATLAB (version 2019a; Mathworks). MATLAB code by Ward (2015) constructed an icosahedron that had its faces recursively subdivided thirty times to generate a mesh containing 20480 triangular faces. The perturbations were generated by placing the mesh into a 3D noise volume and scaling the radial displacement of the vertices by the value of the noise at that vertex's position. A set of three unique meshes were created with these settings (see Figure 20), using a new randomly generated noise volume. High curvature versions of these three meshes were created by dividing the minimum radius value and multiplying the maximum radius value by 1.25 and rescaling the vertex radii across the new range.

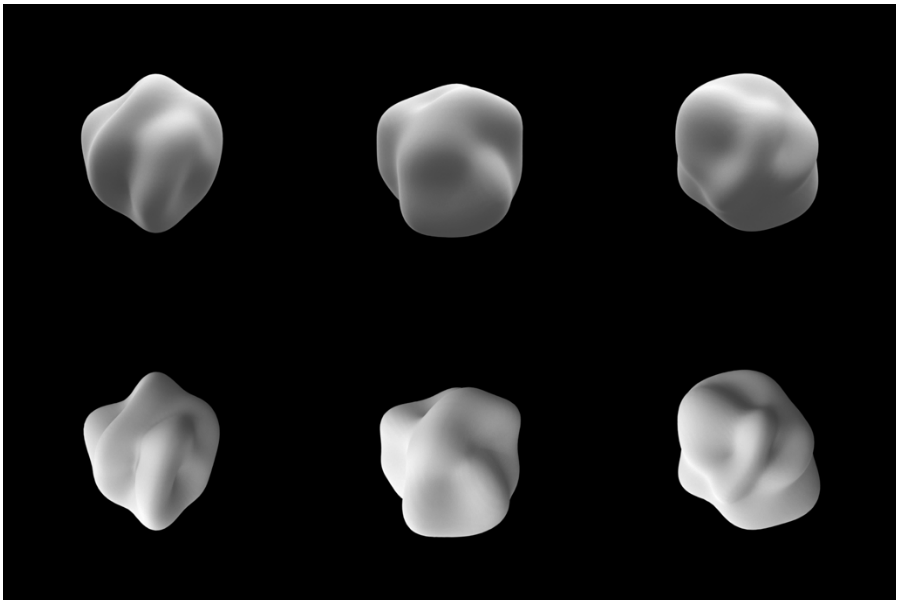


Figure 20. Examples of stimuli used for Experiment 1. From left to right, Shape 1, Shape 2, and Shape 3 are shown. The bottom row shows the high curvature versions of the corresponding shapes.

The meshes were imported into Blender open-source graphics software for rendering (version 2.80; Blender Foundation). The meshes were first smoothed with Blender's Smooth modifier using the values Factor = 0.5, Repeat = 20. The surfaces were rendered using Blender's Cycles engine using a Lambertian BRDF with a reflectance value of 0.3.

Surfaces were rendered using the 'Eucalyptus Grove' light map from Debevec's (1998) light probe image gallery (Figure 21), a high dynamic range (HDR) image of a natural scene. Blender maps the HDR image onto an infinitely distant sphere by converting the vertical axis into longitude and horizontal axis into latitude. Twelve randomly chosen 3D coordinates were selected, evenly covering the entire illumination sphere. The light map was rotated such that the zenith was directed toward each of these coordinates, generating twelve illumination orientations. The even sampling of illumination directions allows us to comment on the role of priors in predicting successes and failures of shape constancy. Our treatment of illumination priors will be general in Experiment 1, but the light-from-above prior will be examined formally in Experiment 2.

Figure 21.'Eucalyptus Grove' light map used in the rendering of stimuli for Experiment 1. This figure depicts an angular map of a spherical image with pixel latitudes and longitudes derived from *x* and *y* pixel coordinates, respectively. The original HDR image used for rendering has a dynamic range of 5000:1.

The objects were placed against a black background and rendered using perspective projection. The three objects were each rendered using the twelve illumination orientations, producing thirty-six 8-bit greyscale images. Since this experiment was concerned specifically with 3D shape perception, the intensity values were normalised such that the maximum pixel value in each of the thirty-six images was 255. Without normalisation, the intensity values in each of the images varied considerably due to the surface being variably exposed to the primary light source in the light map depending on its orientation. Images were normalised by maximum pixel intensity as human judgement of albedo is reasonably well predicted by the highest image intensities (Toscani et al., 2017). Controlling the perceived albedo across our images may help observers better report their experience of 3D shape specifically.

Procedure

Observers were presented with three images of the rendered objects against a black background and instructed to select the object that had the most different shape relative to the remaining two. The experimental trials on which the data are based were constructed by selecting combinations of three images from the set of twelve images of a given shape and given curvature (i.e., the three presented stimuli varied only in illumination direction). Ten catch trials were included in which one of the three presented images was of a different shape type (see Figure 22 for an example of an experimental and catch trial). We would expect a reasonably attentive observer to consistently select the image of the shape type that differed from the remaining two. These trivial judgements were included to allow filtering of data from non-attentive observers. The trial order and positions of the three images were randomised. Each observer viewed two of the three shapes tested to reduce the number of trials each observer had to complete to a more manageable number. Trials were constructed from 12C3 illumination directions by two shape types, plus ten catch trials for a total of 450 trials.

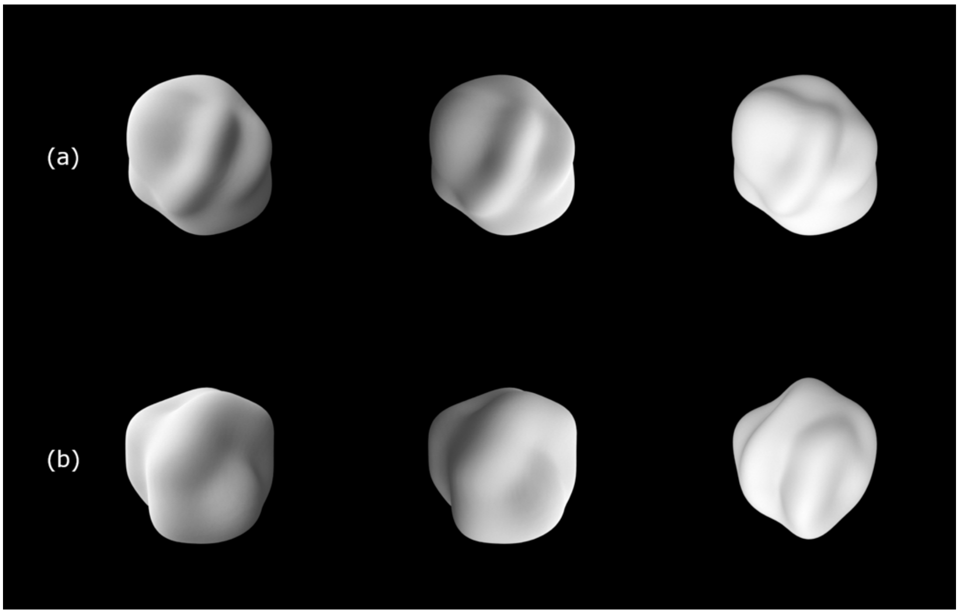


Figure 22. Examples of an experimental (a) and catch (b) trial. Experimental trials consisted of all combinations of three illumination directions but were of the same shape type and curvature. For catch trials, one of the three images were of a shape type that differed from the remaining two. Catch trials were included to screen data of inattentive observers, as reasonably attentive observers would consistently select the rightmost shape in the above example. Images of each trial set were randomly positioned.

Results and discussion

Data analysis

To account for missing data in the sample, proportions were calculated for how likely each observer was to select each stimulus as being the ‘odd-one-out’ in terms of perceived shape, based on how frequently they had viewed that stimulus during the experiment. Results are summarised in Figure 23 with the proportion of each stimulus being selected as being the odd-one-out presented on the y-axis and illumination direction presented on the x-axis. Visualisations of our illumination directions are provided as renderings of shaded spheres along the x-axis at the bottom of the Figure.

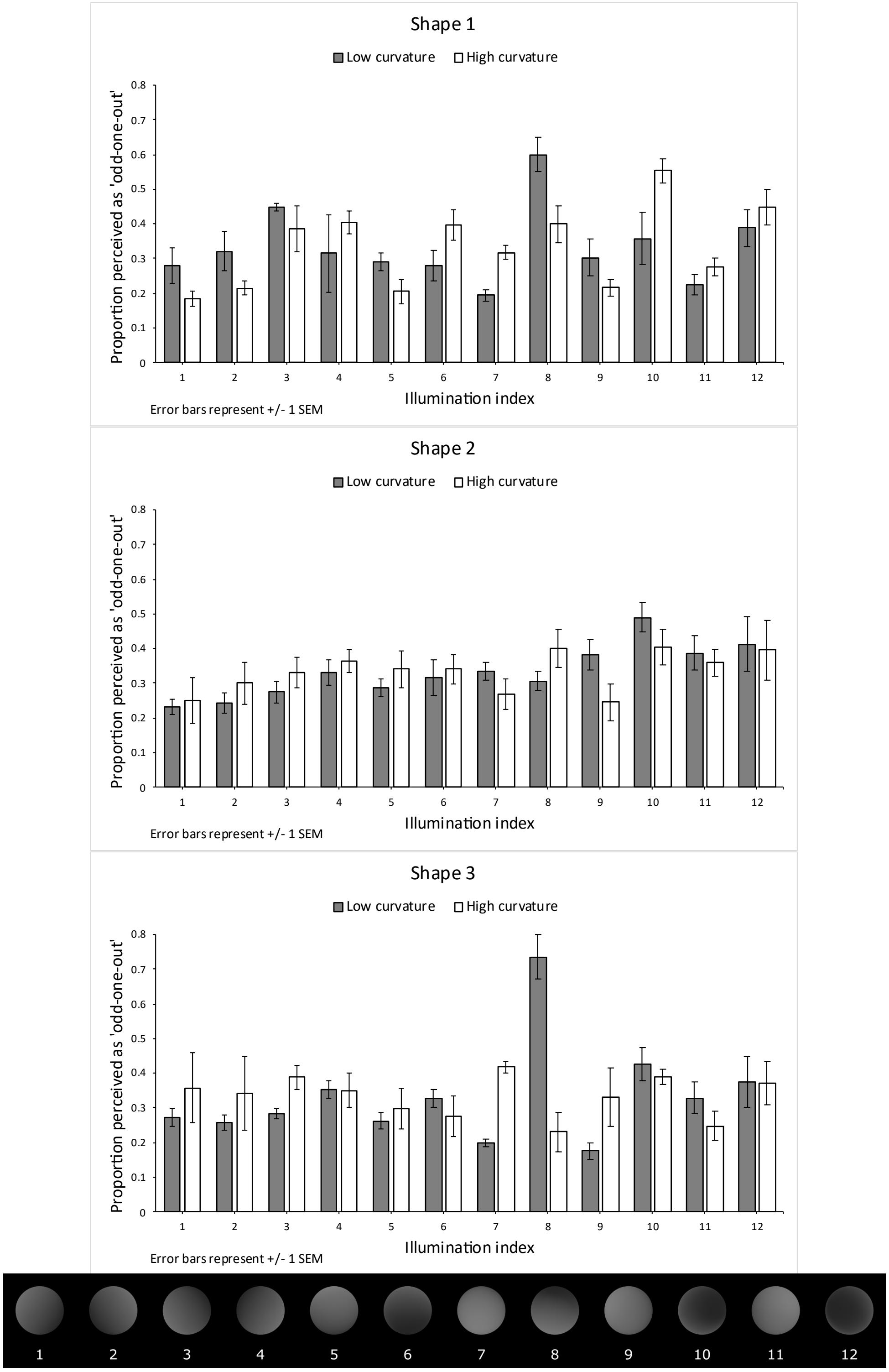


Figure 23. Observer data from the 'odd-one-out' task of Experiment 1. Graphs separate data for each type of shape with grey bars denoting low curvature and white bars denoting high curvature versions of the stimuli. Renderings of diffuse spheres along the bottom of the figure provide visualisations of the illumination orientations of the corresponding indices along the x-axis.

Data were analysed using ANOVA. Main effects of curvature were omitted since high and low curvature stimuli were not directly compared by observers during the experiment. The illumination direction by curvature interactions were analysed to test whether increasing curvature promotes stability across variation in illumination direction. Significant main effects of illumination direction were observed for all presented shapes, $p < .001$. The interaction between illumination direction and curvature was significant for both Shape 1 and Shape 3, $p < .05$ and $p < .001$ respectively, but was not significant for Shape 2, $p > .05$.

Image analysis

We firstly considered whether the pattern of observers' judgements could be explained by the zero-order differences between our stimuli. Correlations were calculated for each image pair from our set of stimuli, and then compared with observer judgements of shape stability (see Table 1). Image correlations ranged widely between strong positive/negative and near-zero values. However, the image correlations were not predictive of mean observer judgements. Since zero-order differences between images were not predictive of observer judgements, we proceed to analysis of image gradients.

Table 1.

Summary of image pair correlations and their correlations with mean observer responses.

Low curvature

Shape type	Image correlations		Correlation with perceived shape stability	
	Range	Min	r	p
1	-.800 - .890	.022	.165	.317
2	-.783 - .880	-.007	-.022	.484
3	-.805 - .878	-.006	.079	.414

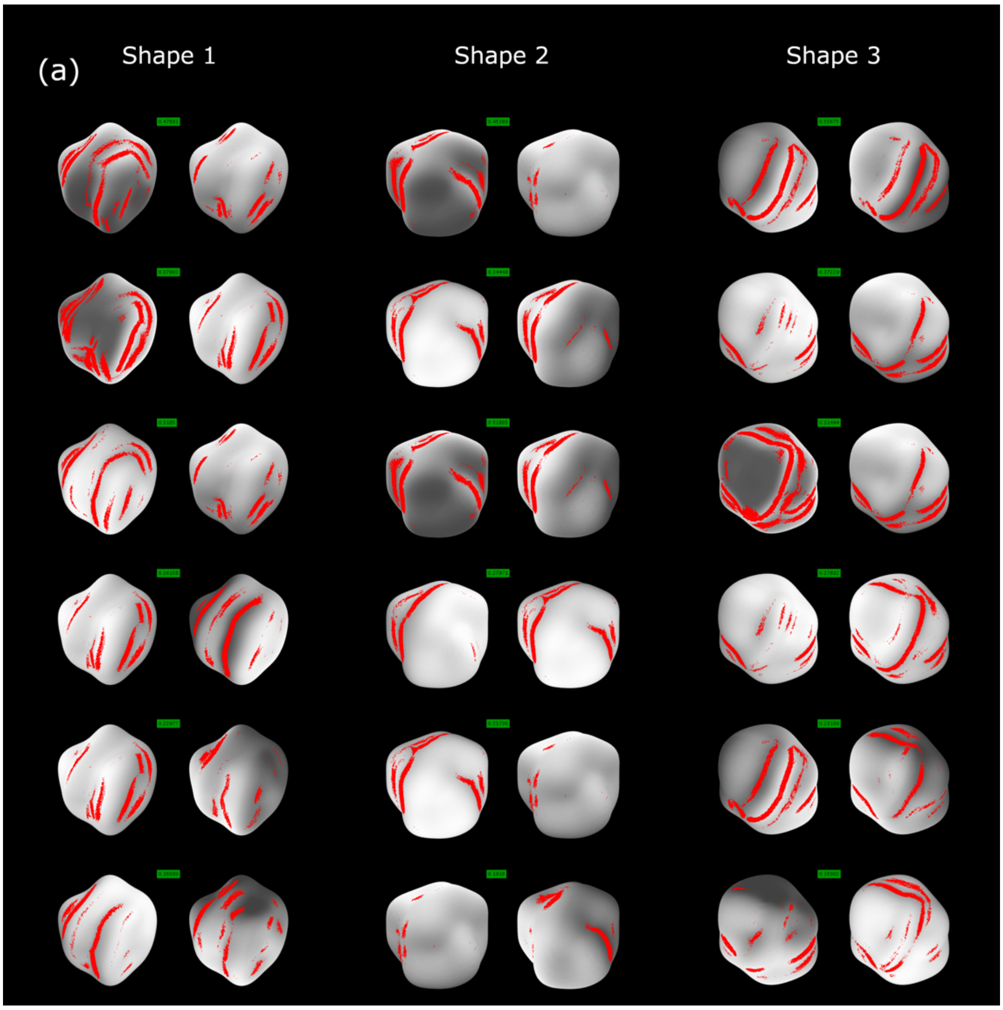
High curvature

Shape type	Image correlations		Correlation with perceived shape stability	
	Range	Min	r	p
1	-.795 - .980	-.004	.045	.478
2	-.758 - .995	-.002	-.042	.479
3	-.769 - .995	-.066	.083	.459

Note. The 'Image correlations' columns present the range of correlations between image pairs with the 'Min' sub-column presenting the minimum absolute correlation.

For our analysis of image gradients, stimuli were convolved using a filter selective for regions of high contrast symmetrical shading flow to capture image structure similar to critical contours described by Kunsberg and Zucker (2018). Our method of computation differs from that used by Kunsberg and Zucker (2018), which applies Morse-Smale theory. A Morse-Smale (MS) complex computes smooth paths of maximal descent of a vector field, from local maxima to saddle points, and from saddle points to local minima. These paths, or one-cells, segment the vector field into descending manifolds, or two-cells, demarcating regions where the vector flows converge on two distinct local minima. The vector field is therefore symmetrical about these paths. Critical contours are the subset of one-cells connecting a saddle point to a local minimum that are of "sufficiently high contrast". These two properties are captured by our method of computation (see Section 1.3.2. for a review of critical contours).

A thirty-by-thirty kernel was applied to each pixel in the image that calculated image intensity contrast and correlation between each half of cross-sections oriented between 0° and 165° in steps of 15°. Only kernels that were fully contained within the shape were computed. Negative correlations were assigned a value of 0 and variances were calculated by squaring the correlations. Contrast was calculated as the difference between the maximum and minimum pixel intensities along the cross-section. Each pixel was assigned the maximum variance value and this was multiplied by the contrast value calculated along the corresponding cross-section orientation. These values were then converted to 8-bit greyscale images. In Figure 24, we present a sample of image pairs with pixels strongly exhibiting characteristics of critical contours shown red, as defined above. Red denotes pixels for which the variance-contrast product exceeded a value of 14. Figure 24a presents the low curvature stimuli and Figure 24b presents the high curvature stimuli.


(a)
Shape 1
Shape 2
Shape 3

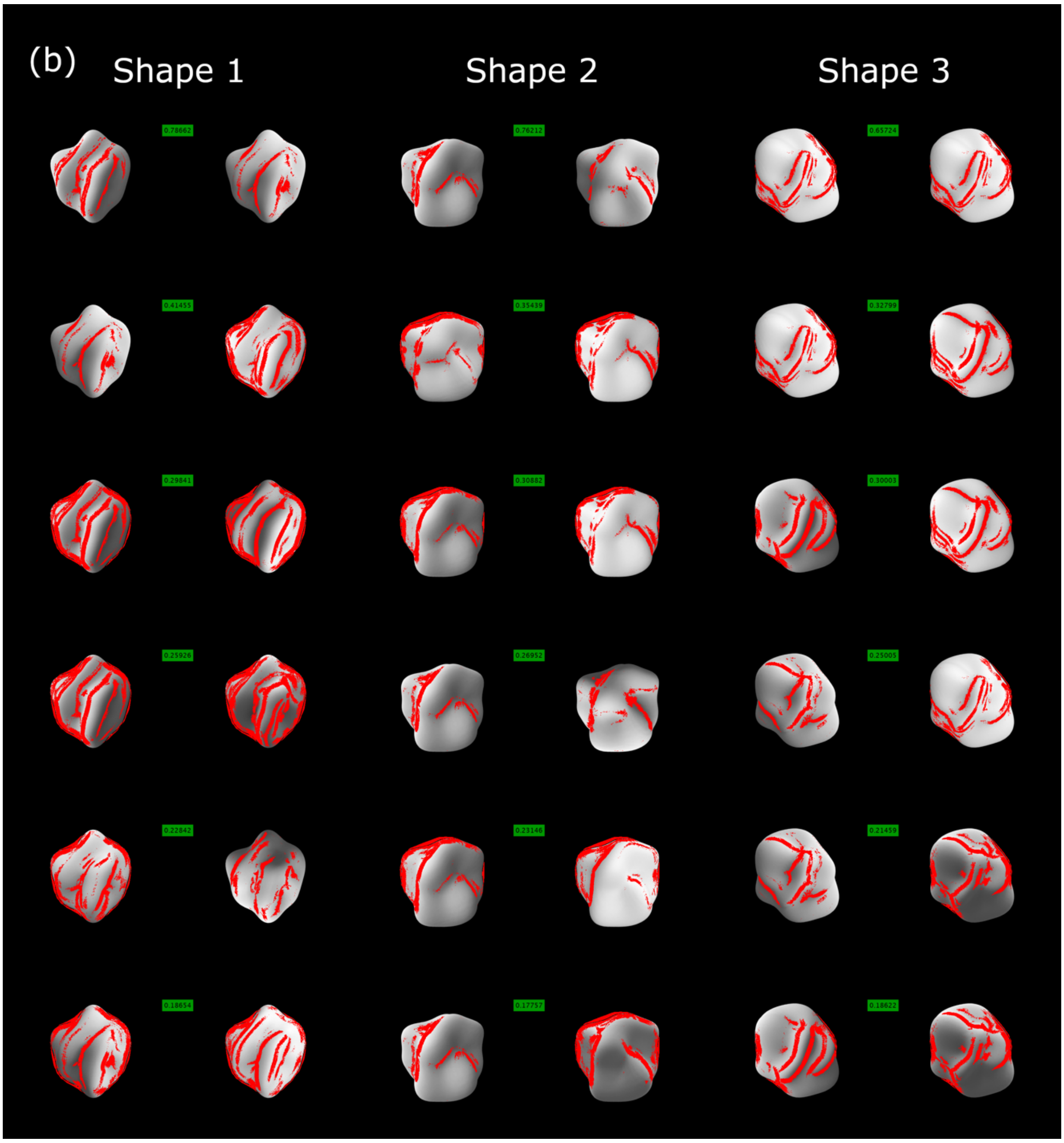


Figure 24. Sample of image pairs ordered in descending degree of stability of perceived shape from top to bottom for the low (a) and high (b) curvature stimuli. Regions resembling critical contours (Kunsberg & Zucker, 2018) are denoted in red and mean proportions that each pair was judged as having the most similar shape out of the three presented stimuli are shown in the green boxes. Of the fifty-five image pairs, each row presents the most to least stable pairs in steps of eleven pairs (i.e., first, twelfth, twenty-third, etc.) to provide a cross-section of our observations.

In many of our stimuli, the red pixels form contours demarcating the extent of ridges along the surfaces. Figure 25 shows heatmaps of the total frequencies of pixels marked as a critical contour summed over images rendered under each illumination direction for each shape. The frequency with which these pixels organise into critical contours validates an important argument from Kunsberg and Zucker (2018) that this structure is fairly robust to variation in illumination, and therefore provides a structure that the visual system may exploit to stabilise the perception of shape. In Figure 24, comparing the image structure generated by Shape 2 with the other two shapes provides an explanation for the lack of a curvature by illumination direction interaction for this shape. For both high and low curvature versions of Shape 2, key shape features are consistently represented by critical contour structure. For the other two shapes, the critical contours are considerably more variable in the low curvature versions and are more robust in the high curvature stimuli. This view is supported by the heatmaps in Figure 25, where the difference between the low and high curvature heatmaps for shape 2 is relatively small. For the remaining two shapes, critical contours more reliably demarcate the geometry for the high curvature version. Of note, the left boundary of the slightly oblique vertical ridge in shape 3 is more apparent in the critical contour structures of the high curvature images (compare top and bottom right images of Figure 25).

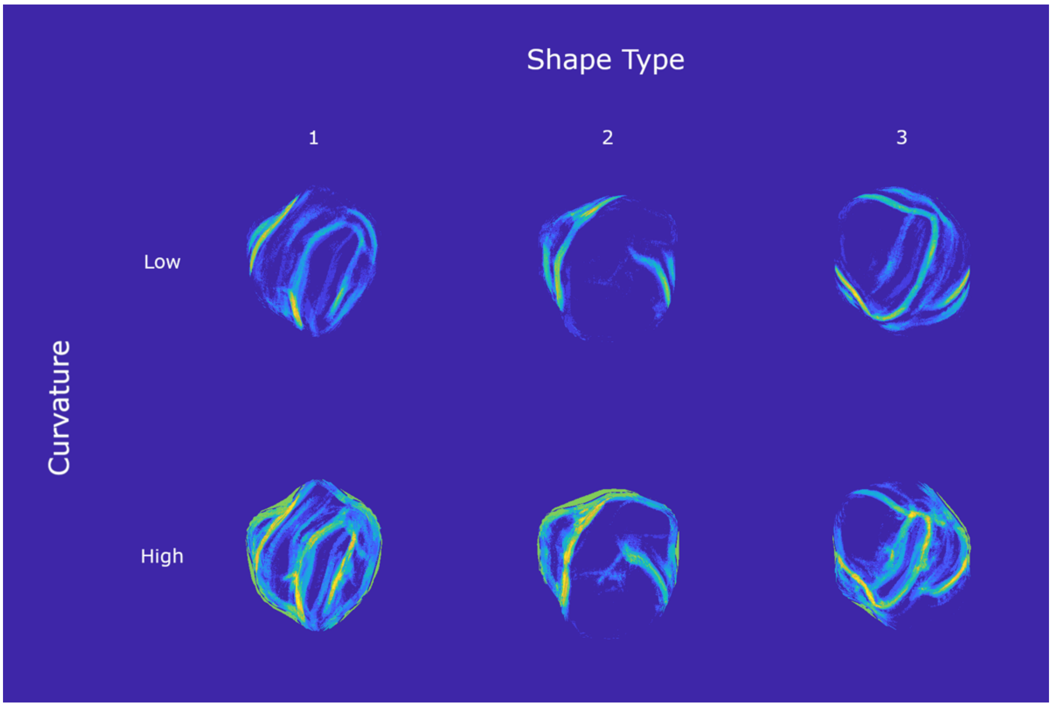


Figure 25. Heatmap showing frequencies with which pixels are marked as belonging to a critical contour (methodology described in text) summed over the twelve total illumination directions for each shape (columns) and curvature (rows). Low frequencies are shown in violet/blue and high frequencies in green/yellow.

These results suggest that anisotropic shape features (i.e., features that mostly curve in one direction) may predict significant deviations in shape perception under specific illumination conditions. Since the relative intensity of diffuse shading is determined only by the slant angle between the surface normal and illumination direction, there will be minimal variation in intensity along the axis of minimal curvature where the magnitude of this curvature approaches zero (e.g., the flat axis of a cylinder). Under generic conditions, an image of a Lambertian cylindrical ridge generates a shading gradient only in the perpendicular direction. However, when the illumination direction runs parallel to the flat axis of a ridge, the image structure is largely abolished, resulting in the perception of a flat surface patch. This is a substantial and localised (i.e., non-affine) deviation of perception from ground truth.

Our sample attrition is almost entirely attributable to observers completing insufficient trials. The difficulty of this task for naïve observers was underestimated as experiment timing was determined by pilot studies using two experienced psychophysical observers. This was an exploratory study that did not test any specific hypotheses. As a result, our task instructions of selecting the image with the most different shape relative to the remaining lacked specificity. With the insights obtained from this study, stimulus and task design of subsequent experiments will be more targeted and suitable for naïve observers.

Subsequent experiments are designed to investigate our hypothesis of the role of anisotropic shape features in failures of the stability of perceived shape. Our first experiment tested three randomly generated objects that were not specifically designed for testing this hypothesis. In Experiment 1, we did not control the orientations of illumination and surface features and therefore cannot conclude whether the alignment of anisotropic surface features with the illumination explains the effects we observed. Our image analysis also requires further examination due to the arbitrary thresholds that were selected to generate critical contours. Pixels where the product of local contrast and shading symmetry exceeded some value were marked in red, and this value was selected such that it captured the general pattern of perception of the experimenters.

Experiment 2a

The second pair of experiments tests the hypothesis that the alignment of illumination with the axis of minimum curvature of anisotropic shape features is predictive of failures of shape constancy. Random 'potato' shape bases were configured to contain mainly isotropic bumps or anisotropic ridges, and illumination directions were sampled to observe shape perception as the illuminant approached alignment with the anisotropic ridges. Previous work has shown that planes containing anisotropic bumps are poorly resolved when the illuminant is aligned with the axis of elongation, and that qualitative shape is largely stable with off-axis illumination with differences in relief height due to increased contrast when surfaces are illuminated in the perpendicular direction (Koenderink, Doorn, et al., 2007). Anisotropic bumps generate shading gradients mainly in the direction perpendicular to the axis of elongation, biasing observer judgements of illumination orientation toward this direction. This often results in highly inaccurate estimates of the illuminant, potentially accounting for the variability in perceived shape. Here, we examine the perception of anisotropic shape features embedded in globally convex surfaces, which provide a robust cue for the illumination direction, namely intensity variation along the bounding contour. Similar failures of shape constancy in this context are unlikely to be explained by inaccurate estimation of the illuminant.

As in Experiment 1, we also tested whether any effects would translate to stimuli in which the surface features were of higher curvature. Since higher curvature shapes generate higher contrast shading gradients and increase the incidence of shadowing and vignetting, we may predict a main effect of curvature promoting stability of perceived shape. We hypothesise a couple of outcomes regarding the interaction between shape curvature and isotropy. The stabilising effect of higher curvature may be greater for anisotropic shapes given that anisotropic shapes would otherwise present greater failures of shape constancy than the isotropic shapes, which we predict would be largely robust to variation in illumination direction. However, given that the failure of shape constancy for anisotropic features arises specifically due to the near-zero curvature of the elongated axis (which is near-zero in both the low and high curvature stimuli), we may find that increasing curvature has a minimal effect on promoting stability of anisotropic shape.

Methods

Stimuli

Stimuli were generated by perturbing a sphere by rescaling vertex radius according to the value of a 3D noise volume at that point to generate a 'potato' shaped base, similar to the technique used to generate the stimuli used in Experiment 1. For the 'isotropic' stimuli, low amplitude bumps were introduced to the base shape. Then, several random vertices were selected as locations for higher amplitude Gaussian bumps, which were the prominent features of the shape. The amplitude and sigma varied between bumps. For the 'anisotropic' stimuli, the 3D noise was modified such that its mean frequency was lower along the x-axis, causing a horizontal stretching of the bumps generated. Two unique shapes were generated for each of the stimulus categories. Low and high curvature versions of each shape were produced by rescaling the minimum and maximum radii by a factor of 1.1, similar to the technique in Experiment 1 (see Figure 26).

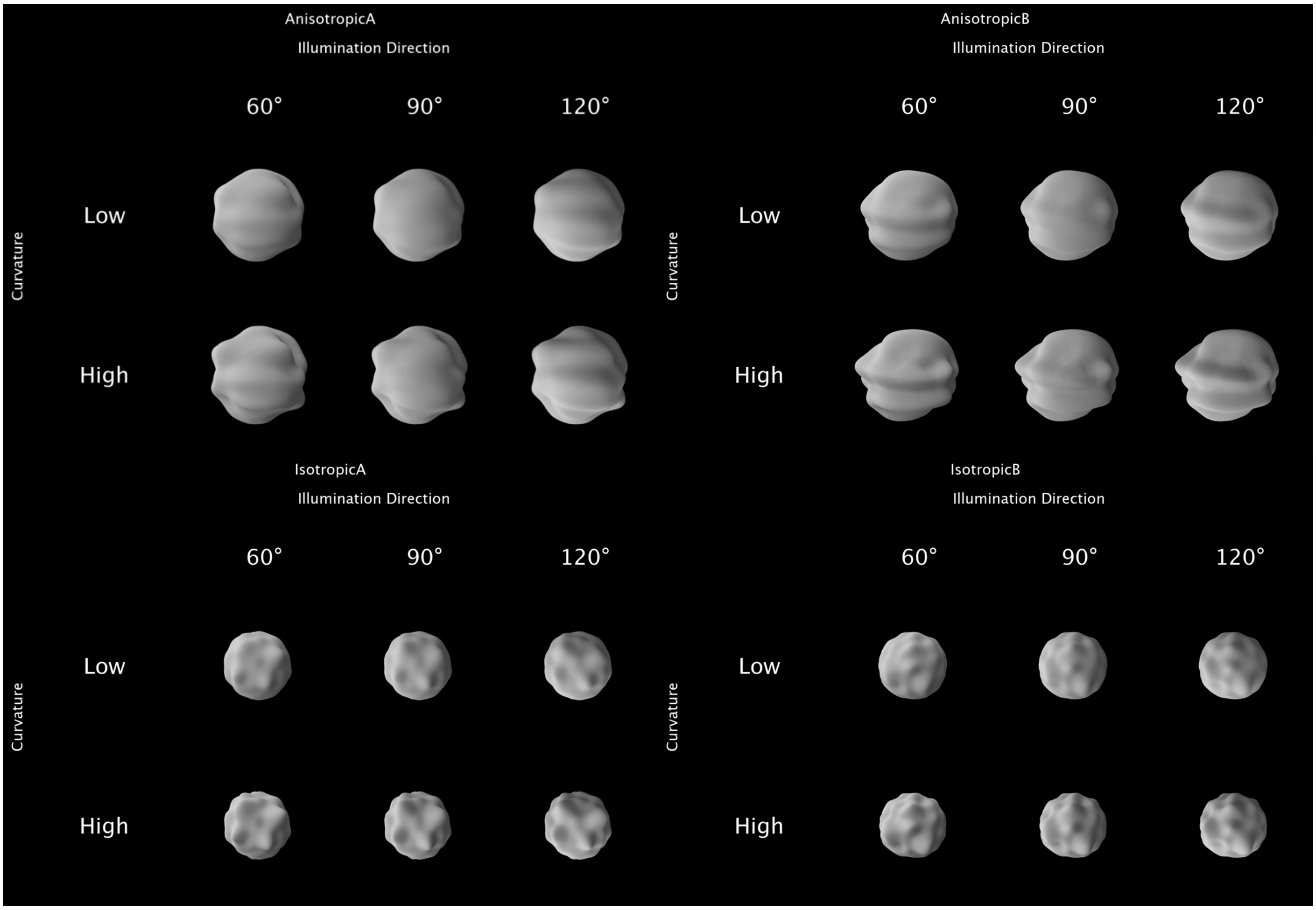


Figure 26. Examples of stimuli used in Experiment 2.

The surfaces were rendered in Blender using the same camera and material settings as in Experiment 1. The same 'Eucalyptus Grove' light map was used with illumination direction varied by rotating the light map anticlockwise in 15° steps from 45° to 135°. Intensity values were normalised such that each image had a maximum intensity of 255.

Observers

Twenty-three observers were recruited and given 30 minutes to complete the experiment online. Observers provided informed consent and were compensated with course credit. Research ethics was approved by the University of Sydney Human Research Ethics

Committee.

Procedure

Observers were presented two pairs of images (top and bottom) and instructed to select the pair that had the most similar 3D shape. The tested illumination pairs all differed by a 30° clockwise rotation (i.e., 45° and 75°, 60° and 90°, etc.), generating five illumination pairs. The four images presented on each trial were of the same shape and all combinations of illumination pairs were compared between the top and bottom pair of stimuli. Frequencies that each illumination pair was selected as having the most similar shape for each of the four shapes were analysed.

Results and discussion

Since the levels of the Curvature and Isotropy independent variables were never compared in an individual trial, the data for the low/high (an)isotropic stimuli were analysed separately with within-subjects ANOVAs over the Illumination Direction variable. For each analysis, we tested our hypothesis that pairs containing 90° illumination direction with a planned contrast, comparing the mean of the 60° and 90°, and 90° and 120°, with the mean of the remaining illumination pairs. These subsets of data are presented in Figure 27 with the frequency that a stimulus pair was selected as having higher shape similarity varying along the y-axis, and its illumination direction pair varying along the x-axis.

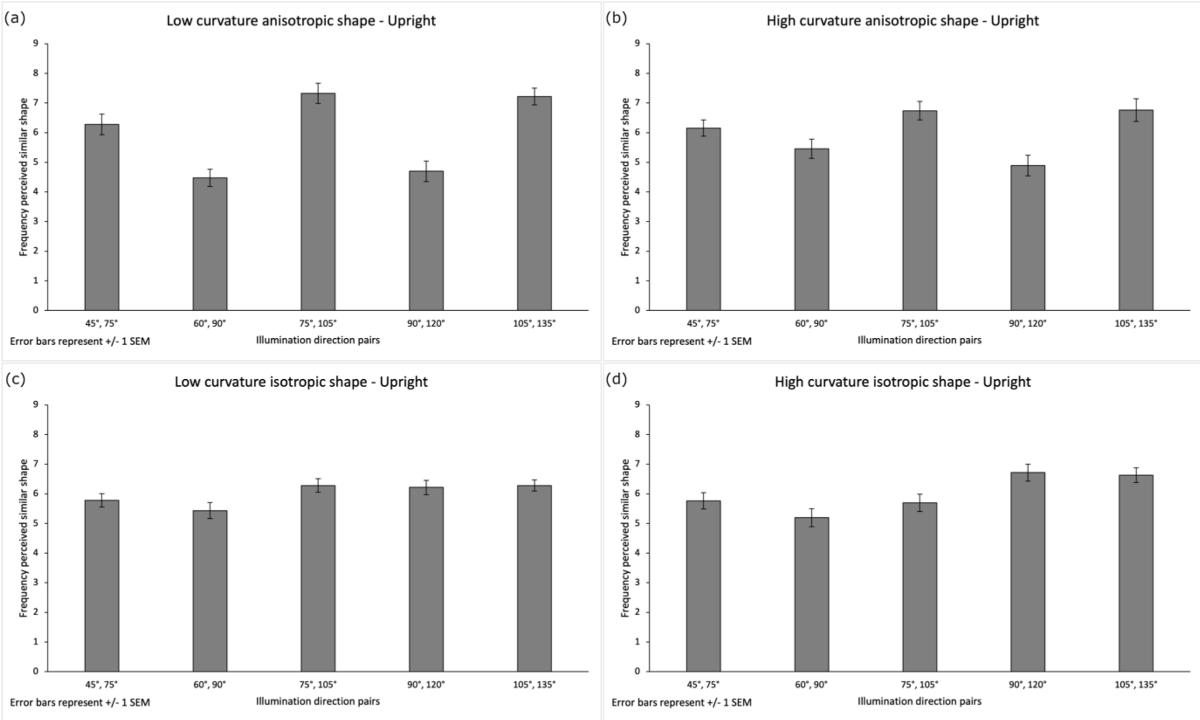


Figure 27. Data from Experiment 2a's paired comparison task. Observers selected which of two pairs of stimuli were more similar in shape. Each pair was composed of a shape illuminated at two different directions. Graphs separate data for each shape category, with mean frequencies of similar perceived shape shown on the y-axis, and the illumination pairs shown on the x-axis.

For the low curvature anisotropic stimuli, there was a significant effect of illumination direction on perceived shape stability, $F(4,88) = 14.13$, $p < .001$. Pairs containing 90° illumination were selected as having similar shape significantly less frequently than the remaining illumination pairs, $F(1,22) = 28.60$, $p < .001$. We also observed a significant effect of illumination direction for the high curvature anisotropic stimuli, $F(4,88) = 4.88$, $p < .001$. Likewise, pairs containing 90° were selected as having similar shape significantly less frequently than the remaining illumination pairs, $F(1,22) = 13.78$, $p < .001$. This suggests that increasing curvature had a negligible effect on promoting stability of anisotropic shape features, however our experimental design did not allow for testing of the interaction effect of curvature. There was no significant effect of illumination direction on stability of perceived shape for the low curvature isotropic shapes, $F(4,88) = 2.14$, $p = .083$. Curiously, we did observe a significant effect of illumination for the high curvature isotropic shapes, $F(4,88) = 4.31$, $p < .005$, however unlike the anisotropic stimuli, the comparison of pairs containing 90°

with remaining pairs was not significant, $F(1,22) = 0.09$, $p < .05$. The non-significant contrast testing for whether pairs containing 90° illumination direction is unsurprising as this effect would only be anticipated for anisotropic shape features stretched along the horizontal direction. The main effect of illumination direction specifically for the high curvature versions of the isotropic stimuli could be explained by the more pronounced shearing of the bumps toward the illumination direction, which is more apparent for bumps of higher relief.

In Experiment 2a, we demonstrated significant transformations of perceived shape specifically for anisotropic shapes where the illumination direction is aligned with this axis of minimal curvature. These effects occurred despite the presence of a robust illumination direction cue provided by the self-occluding rim of a globally convex surface. Unlike in Experiment 1, these stimuli were specifically designed to align with one of the tested illumination directions at 90° to the ridges in these stimuli. This effect is attenuated when the anisotropic shape features are of a higher curvature. Although our illumination direction variable specifies the orientation of the dominant light source, complex light maps illuminate the surface from all directions, introducing sources of illumination off the axis of minimal curvature. This generates shading flow perpendicular to the axis of minimal curvature, which is of higher contrast and more apparent when shape features are of higher curvature. Higher curvatures also increase the incidence of shadowing and vignetting, which may also promote the visibility of the ridges' boundaries. We suspect that the changes in perceived shape observed may be non-affine as they appear to be localised specifically to anisotropic shape features. However, since our data do not describe the magnitudes of changes to perceived depth, we cannot determine whether observed shape changes are better described as an affine uniform scaling or shearing in depth, or a non-affine localised deletion of specific shape features.

As predicted, shape was robust to variation in illumination direction for all isotropic shapes. The Gaussian bumps used in this stimuli curve equally in all directions and therefore generate largely identical shading structure across variation in illumination direction, varying only in the orientation of the shading gradient.

Our result of shape constancy failure specifically for anisotropic stimuli illuminated from the left could also be explained by the atypicality of the illumination. Most scenes we encounter are illuminated from above, be that under harsh sun or indoor ceiling lights, or more diffuse hemispherical lighting on a cloudy day. Evidence suggests that humans typically exhibit a light-from-above prior, shifted slightly to the left (e.g., Ramachandran, 1988; Sun & Perona, 1998). In Experiment 2b, we present a control experiment for the effect of the light-from-above prior.

Experiment 2b

Methods

Stimuli

The stimuli from Experiment 2a were rotated 90° anti-clockwise for this experiment and were otherwise identical. This was done to control for the role of the light-from-above prior, which could also predict changes to shape perception for illumination directions further from the vertical axis.

Observers

Twenty-six observers were recruited and given 30 minutes to complete the experiment online. Observers provided informed consent and were compensated with course credit.

Procedure

This experiment followed an identical procedure to Experiment 2a.

Results and discussion

As with Experiment 2a, the data for the low/high (an)isotropic stimuli were analysed separately with within-subjects ANOVAs over the Illumination Direction variable. The data for the rotated stimuli are presented in Figure 28. Note that the labelling scheme for illumination direction refers to the same stimuli as the unrotated set from Experiment 2a (i.e., 90° now describes light from above).

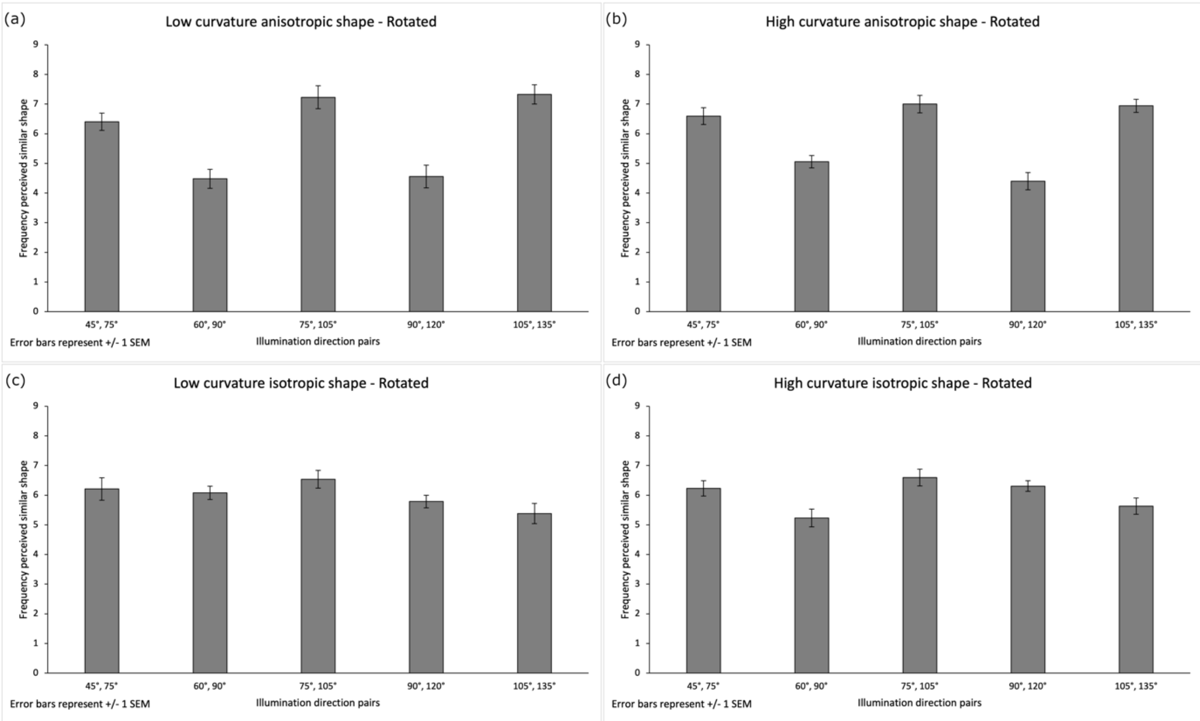


Figure 28. Data from Experiment 2b's paired comparison task. An identical procedure was used to Experiment 2a, however the stimuli were rotated 90° counter-clockwise such that our 90° stimuli were now illuminated from above rather than from the left.

The data for the rotated stimuli presented the same pattern of effects as the original stimuli, suggesting that these effects are not due to the illumination directions themselves (e.g., relying on priors such as illumination coming from above) and depend specifically on the illumination direction relative to surface geometry. There was a significant effect of illumination direction on the stability of perceived shape for both the low curvature, $F(4,100) = 13.39$, $p < .001$, and high curvature, $F(4,100) = 16.45$, $p < .001$, anisotropic stimuli. Similarly, pairs containing 90° illumination were selected as having similar shape significantly less frequently than the remaining illumination pairs for both the low curvature,

$F(1,25) = 35.20$, $p < .0001$, and high curvature, $F(1,25) = 62.55$, $p < .001$, anisotropic stimuli. No significant effects were observed for the low curvature isotropic stimuli, $F(4,100) = 1.72$, $p = .152$. However, we replicated the results for the high curvature isotropic stimuli from Experiment 2a. There was a significant effect of illumination direction, $F(4,100) = 7.99$, $p < .01$. There was no significant difference between pairs containing 90° and remaining pairs, $F(1,25) = 1.97$, $p = .173$.

We replicated the results from Experiment 2a on stimuli where the 90° illumination was now directed from above, and therefore typical illumination conditions consistent with the light-from-above prior. This supports our argument that instability of perceived shape is explained by the alignment of illumination with surface geometry for these stimuli. We also replicated the surprising significant main effect of illumination direction on perception of the high curvature isotropic stimuli. Given this replication and the fact that this appears to be driven mainly by dissimilarity between the 60° and 90° pair in both experiments, this cannot be dismissed as an idiosyncratic result. According to the experimenters, shape appears largely stable across all isotropic stimuli. This result could be an artefact of the experimental design, where observers were forced to judge which pair of shapes appeared most similar, and may be relying on some other quality of the images in judging the high curvature isotropic stimuli. It is unclear what this may be, however we can exclude the role that violation of the light-from-above prior may play as observers consistently underreported the similarity of the same illumination pair even when stimuli were rotated by 90° in Experiment 2b.

One qualification we provide for these experiments is that we did not control for the curvatures of the anisotropic and isotropic features of our stimuli. This was due to the design method of the stimuli, where the isotropic stimuli contained embedded Gaussian bumps while the anisotropic stimuli were generated by stretching a noise volume along the horizontal axis. Furthermore, the mean curvature of surface features is confounded with the isotropy manipulation. Stretching a bump into a ridge has the consequences of both introducing anisotropy and reducing the mean curvature of the feature, due to flattening of the shape along the direction of stretching. Although our hypothesis for the role of alignment of surface geometry and illumination on failures of shape constancy is compelling given that these failures arose specifically for the 90° stimuli, we have not accounted for the reduction in mean curvature of our anisotropic features.

Experiment 3

As with the two preceding experiments, Experiment 3 tested for changes to the stability of perceived shape as a function of feature isotropy and illumination direction. This experiment parametrically varied the embedded shape features and controlled for the effects of feature size and mean curvature. Introducing curvature anisotropy by elongating a bump presents a confound of reducing the mean curvature of the bump through its flattening along one of its axes. Here, we compared the perception of a small and highly curved isotropic bump, a big and lowly curved isotropic bump, and an anisotropic elongated bump. In Experiment 3, we used an ordinal depth judgement task to more directly probe the shape changes our observers perceived.

Methods

Stimuli

Stimuli were generated by embedding an anisotropic, and two sizes of isotropic bump in the centre of a randomly perturbed sphere constructed using methods similar to Experiments 1 and 2. Each stimulus category used the same base shape and only the embedded feature varied. The embedded features were constructed by intersecting pairs of sigmoid curves aligned vertically and horizontally. The sigmoid curves were generated using the normcdf function in MATLAB with a sigma parameter of 0.2 and mu parameter varying between the stimulus type. The mu parameter controls the displacement between the pairs of curves, with a larger mu value generating a wider bump in the vertical or horizontal direction.

The vertically aligned curves had a negative weight, constraining the resultant bump toward the centre of the base shape vertically. The horizontally aligned curves had a positive weight and dictated the amplitude of the bump. The displacement between the pairs of curves were varied to generate the small isotropic, big isotropic, and anisotropic stimuli. This was controlled by the mu parameter summarised in Table 2.

Table 2.

Sigmoid function parameters for stimuli.

Stimulus	Sigmoid Function Parameters		
	Sigma	Horizontal Mu	Vertical Mu
Anisotropic	0.2	0.5	0.2
Small Isotropic	0.2	0.2	0.2
Big Isotropic	0.2	0.35	0.35

The anisotropic stimulus was generated by applying a higher mu value for the horizontal pair of curves, resulting in an elongated ridge in the horizontal direction. Two isotropic stimuli were generated to test for the effect of shape anisotropy while controlling for the size of the feature. The small isotropic ridge had a vertical and horizontal mu equal to the vertical width of the anisotropic ridge and the big isotropic ridge used the average of the horizontal and vertical mu of the anisotropic ridge.

Images were rendered in Blender using a diffuse BRDF with an albedo of 0.3 and illuminated by the ‘Eucalyptus Grove’ light map, smoothing the meshes as in previous experiments. Three illumination directions were used ranging between illumination from above (0°), upper left (45°), and left (90°).

Observers

Forty observers were recruited and given up to 60 minutes to complete the experiment online. Observers provided informed consent and were compensated with course credit. Five observers who provided incomplete data were removed from the analysis, leaving a final sample of thirty-five.

Procedure

On each trial, observers were presented all pairs of probe points superimposed on the stimulus randomly sampled from a vertical cross section of seventeen points through the bump in the middle of each stimulus (see Figure 29). They were instructed to report perceived ordinal depth by clicking the point that lies over the closest part of the 3D shape (i.e., the point at which the perceived surface relief is greater). Once observers had reported on all point pairs, they were presented with a second screen that showed the tested stimulus with all seventeen probe points superimposed in red alongside a vertical line plot depicting a depth profile generated from their responses on the first task. On this screen, observers could scale the height and smoothness of the plot until they were satisfied that the depicted relief resembled the perceived 3D shape in the stimulus. This was repeated for each of the nine stimuli, presented in a random order.

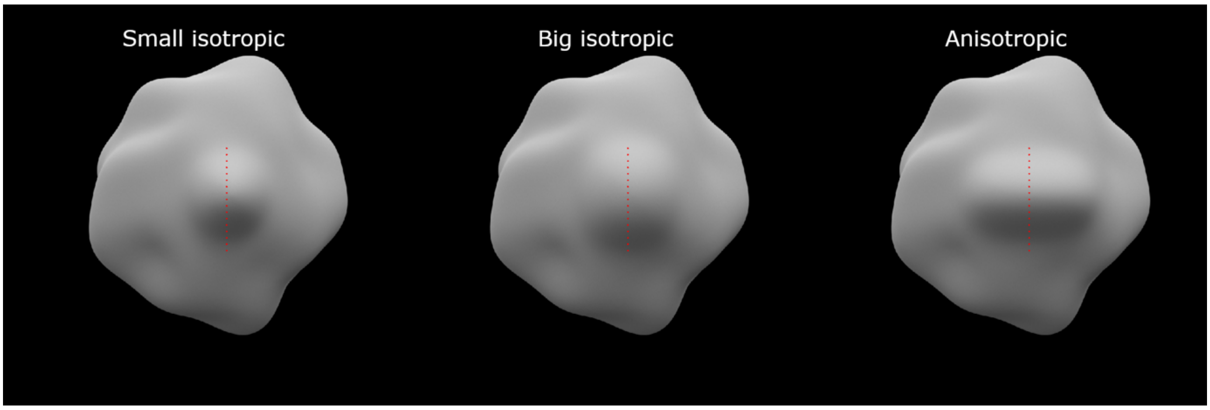


Figure 29. Examples of stimuli for Experiment 3 illuminated from above. Probe points are indicated by the red dots. Observers were presented with pairs of these dots and instructed to report which dot appeared over the part of the surface with higher relief.

Results and discussion

Prior to analysis, each observers' mean perceived relief height was normalised to equal the overall mean perceived relief height across the entire experiment. This eliminates much of the inter-observer variance accounted for by an affine transformation. A within-subjects ANOVA was conducted on the overall data testing the interaction effects between probe location and the shape and illumination direction factors. In this design, shape transformations are captured by the interaction between the probe location variable and the remaining two factors. We are specifically interested in how the isotropy of shape features affects the stability of perceived shape across variation in illumination direction and therefore only report the three-way shape by illumination direction by probe location interactions. This was followed up with two planned contrasts that tested interactions between probe location comparing the larger isotropic shape to the smaller isotropic shape, and comparing the average of the isotropic shapes to the anisotropic shape. The normalised relief data are presented in Figure 30. Note that since the surface features are embedded in a globally convex surface, the reduction in perceived relief at either end of the plot is partly attributable to the closer proximity to the bounding contour. However, the region surrounding the surface feature is identical for each stimulus, meaning that differences between data plots are largely due to perception of the shape feature specifically.

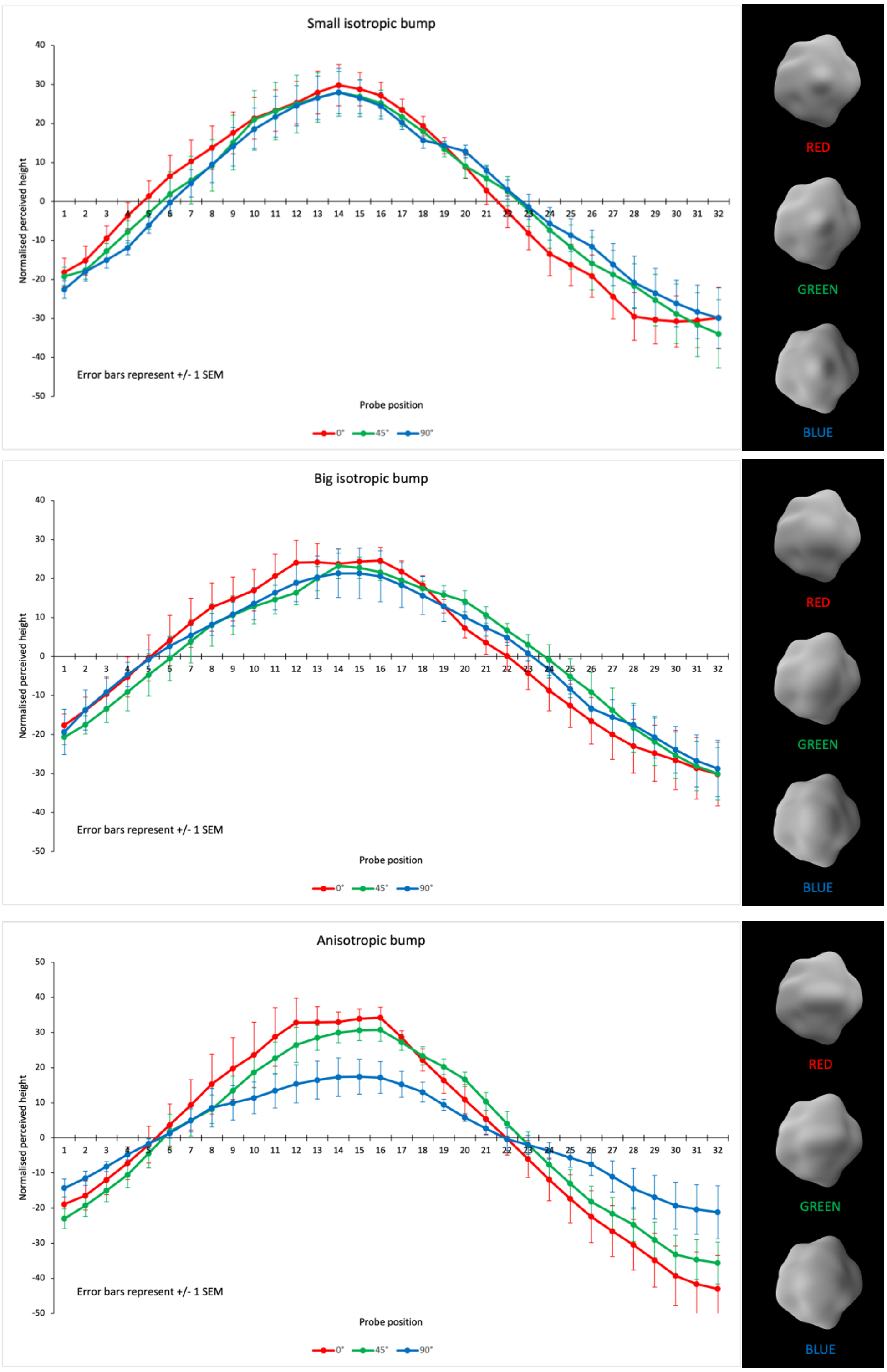


Figure 30. Normalised relief data for Experiment 3. From top to bottom, graphs separate data for each shape category (small/big isotropic and anisotropic), and each plot separates data for each illumination direction (0°, 45°, and 90°).

For the overall data, the two-way interactions between probe location and shape, $F(62,2108) = 1.74$, $p < .001$, and between probe location and illumination direction, $F(62,2108) = 2.90$, $p < .001$, were both significant. We also observed a significant three-way interaction between probe location, shape, and illumination direction, $F(124,4216) = 1.88$, $p < .001$.

We next analysed planned contrasts on the interaction effects with probe location, comparing the small isotropic and big isotropic shapes. The three-way interaction between probe location, shape, and illumination direction was not significant, $F(2,6696) = 1.15$, $p = .318$. The two-way interaction between probe location and shape was not significant, $F(1,6696) = 0.476$, $p = .490$, suggesting that the perceived shape of both isotropic stimuli behaved consistently across the tested illumination directions. However, the two-way interaction between probe location and illumination was significant, $F(2,6696) = 6.85$, $p < .005$. Though this result is not pertinent to our comparison of differences in stability of perceived shape across isotropic and anisotropic shapes, it could be explained by the consistent finding that perceived shape is sheared toward the direction of the light source (e.g., Christou & Koenderink, 1997).

Finally, the interaction effects with probe location and illumination comparing the small isotropic and big isotropic, and the average of the isotropic stimuli to the anisotropic stimulus ,were analysed. The interaction contrast comparing the two isotropic shapes was not significant, $F(2,1044) = 1.12$, $p > .05$. However, the interaction contrast comparing the average of the isotropic shapes to the anisotropic shapes was significant, $F(2,1044) = 10.41$, $p < .0001$. Referring to Figure 30, we may interpret this result as explaining the prominent perceived flattening of the anisotropic ridge when the main illumination direction was aligned to the ridge's axis of minimal curvature (i.e., at 90°), while the perceived shape of the two isotropic figures remained largely stable across the tested illumination range.

The results of Experiment 3 suggest that our data to this point are not likely to be caused by variation in feature size or mean curvature. Although our stimuli are impossible to ideally match on these two dimensions, the tested comparisons suggest that the effects observed are not likely to be accounted for by variations in mean curvature or feature size, since the interaction contrast comparing the two isotropic shapes was not significant. The data from the ordinal depth judgement task allow us to comment more specifically on the nature of shape changes our observers perceived. The first point to clarify is that since these are ordinal data, absolute differences in magnitude should be interpreted with caution since increments along our perceived height metric may not be uniform in size. Plots for the anisotropic bump shape suggest an affine scaling of depth, with the bump being present in all three illumination conditions although of significantly diminished magnitude under 90° illumination. However, this interpretation comes with two qualifications. Firstly, if the scaling of perceived depth is confined to the central bump, it would describe a non-affine transformation of the global shape. We have reason to believe that perceived shape changes are indeed confined to the central bump as the globally similar big and small isotropic shapes that had identical surrounding shape did not demonstrate effects of illumination direction. Secondly, since our ordinal measure prevents us from interpreting observed depth magnitudes directly, it is difficult to distinguish an affine reduction in curvature from a non-affine near-abolition of surface features. Furthermore, part of the increasing depth toward the centre of our stimuli is attributable to the reduction of surface depth relative to the image plane as the self-occluding rim is approached.

Experiment 4

The preceding experiments demonstrated situations in which alignment between illumination direction and surface geometry lead to the abolition of shading structure and subsequent flattening of surface features. Our findings may be more generally explained by variation in the image contrast generated by slant variation occurring parallel and perpendicular to the illumination direction. Diffuse shading is a function of the slant angle between the surface normal and illumination direction, and image intensity is proportional to the projection of the surface normal in the direction of illumination. For this reason, variation in surface slant (or lack thereof) parallel to the illumination direction is maximally represented in the image, whereas variation in surface slant in the orthogonal direction receives zero (in the case of a single collimated light source) or minimal representation in the image. In the preceding demonstrations, alignment of the illuminant with a surface patch's axis of minimum curvature accentuated the near-zero curvature along this axis, resulting in the percept of a near-zero curvature (i.e., flat) surface patch.

Experiment 4 will test this general description by observing effects of illumination direction on the perception of elongated surface features with qualitative differences in slant variation along the two axes. To this end, we designed a surface that contained an elongated ridge with a small depression in its centre such that it resembled two humps when viewed from profile rather than a smooth, cylindrical ridge. The surface feature was still 'ridge-like' in that the centre depression was of a higher relief than the surrounding surface in which it was embedded.

In Experiment 4, we also test for effects of material on the stability of perceived shape. In addition to a pure Lambertian reflectance function, we test smooth and rough dielectric glossy materials constructed by a weighted summation of a purely Lambertian and purely specular reflectance function. The inclusion of specular reflections has been shown to promote the perception of shape and optical focus of shading gradients (Marlow & Anderson, 2021) demonstrating that the cospecification of surface geometry by distinct reflectance functions reduces perceptual ambiguity. Diffuse and specular reflection convey different information about surface geometry, potentially explaining the value of their combination. Specular reflections are increasingly compressed along directions of high surface curvature, providing information about the second derivative of shape (see Fleming et al., 2004 and Section 1.1.2.). The intensity of diffuse reflection depends on the slant angle between the surface and illumination, providing information about the first derivative of shape, although shading gradients, being the derivative of shading, can also provide information of the second derivative of shape. An important distinction between diffuse and specular reflection is that diffuse reflection depends specifically on the slant angle between the surface and illuminant, meaning that surface geometry varying in the perpendicular direction to the illuminant may be obscured, as we have demonstrated in previous experiments. The structure generated by specular reflections is generally robust to variation in illumination direction, with the exception of regions of the light field that contain minimal structure (e.g., a clear sky). Some form of 'texture' must be present in the illuminating environment for a surface's geometry to be represented through the pattern of deformation of reflected features. Blurry specular reflections in some situations are misperceived as diffuse shading (Mooney & Anderson, 2014), which may also promote the stability of perceived shape through the apparent increased contrast of otherwise diminished shading gradients. This causes an exaggeration of perceived curvature, which may change previously observed near-abolition of surface features to affine scaling of depth (see Figure 31).

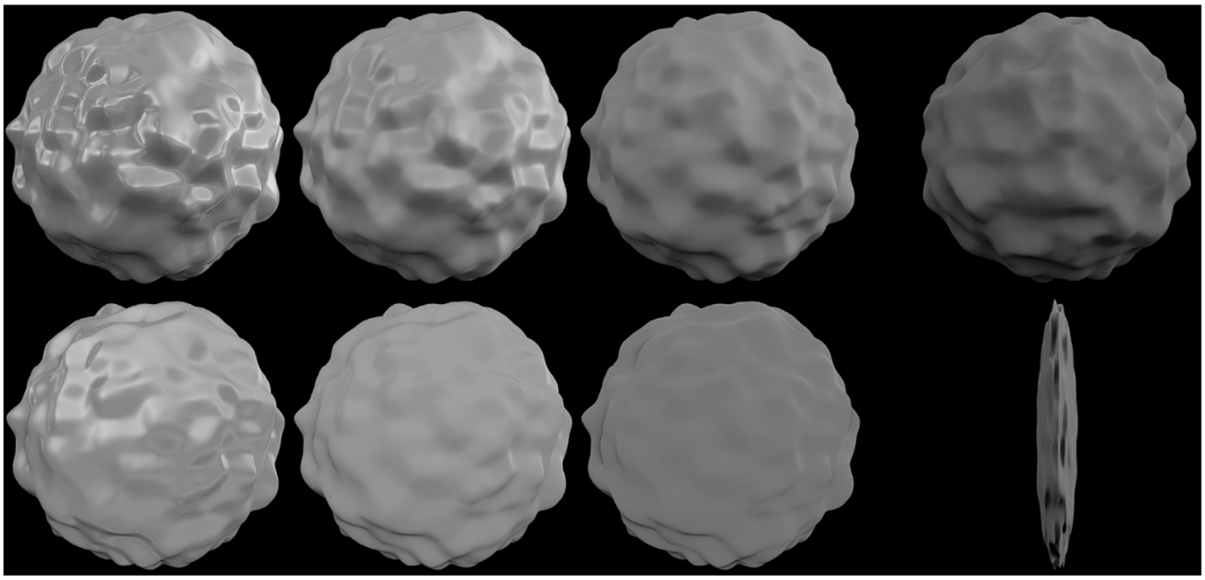


Figure 31. Images of bumpy spheres rendered with different levels of specular roughness and ground-truth depth. From left to right, the first three columns of images show surfaces rendered using low and moderate specular roughness, and diffuse shading only. The rightmost side view images show the pronounced difference in ground-truth curvature between the top and bottom rows. The presence of specular reflections causes an exaggeration of perceived curvature, which is particularly apparent for the 'squashed' bottom row of images. Figure from Mooney and Anderson (2014).

Methods

Stimuli

Stimuli were generated by modulating the relief of the elongated ridge of the anisotropic stimulus used in Experiment 4 such that it had a small depression in its centre. Stimuli were rendered in Blender using a diffuse BRDF with the same settings as previous experiments. Rough and smooth gloss stimuli were also rendered using the Cycles engine specular BRDF with 0.3 and 0.1 roughness, respectively. The rough and smooth gloss images were composites of the pure diffuse and specular rendered images that were additively combined using a weight of 0.2 for the specular component and 0.8 for the diffuse component. The same Eucalyptus Grove light map was used with illumination directions ranging between -30° (top right) and 120° (bottom left) in 15° increments. See Figure 32 for examples of our stimuli and Figure 33 for a side profile of the central modulated ridge, showing the small depression in its centre.

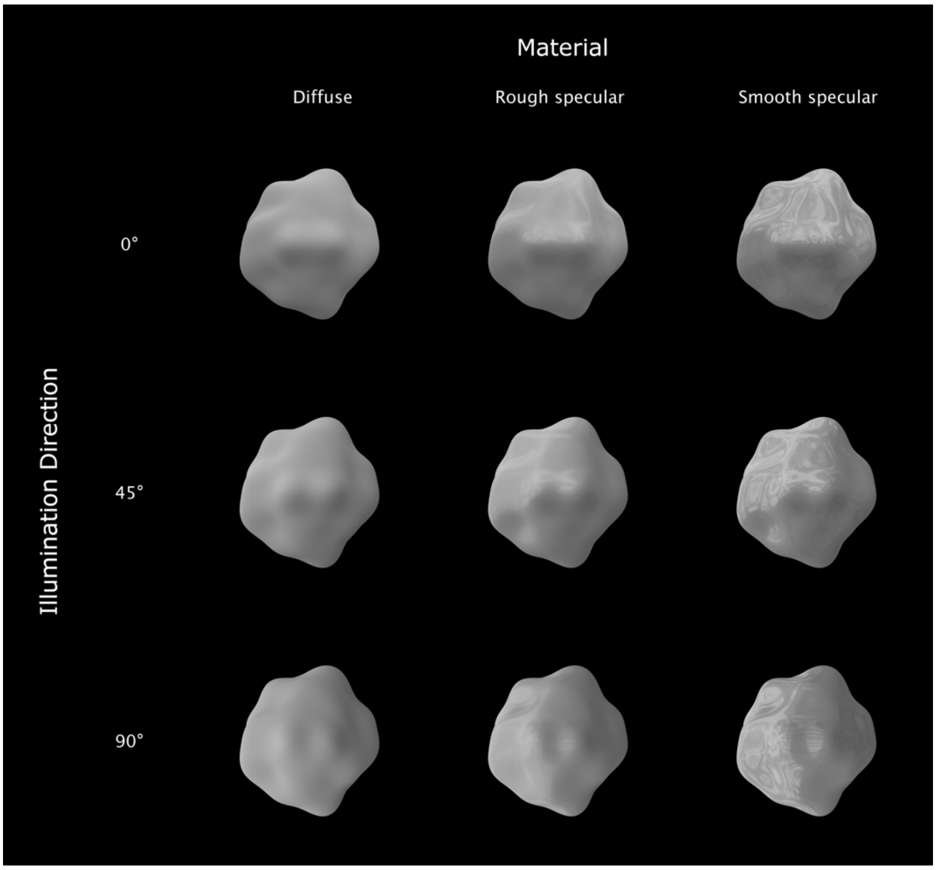


Figure 32. Examples of stimuli used in Experiment 4.

Figure 33. Side profile of central modulated ridge of Experiment 4's stimuli (see Figure 31). The feature is approximately ridge-like, with a small depression in its centre to introduce variation in surface orientation along the axis of elongation.

During the training phase, one of the two presented stimuli was of a shape with an unmodified smooth continuous ridge, as used in Experiment 3. These were presented to familiarise observers with the shape features they should be judging and to filter out data from participants that either misunderstood the task or did not perform the task with due care.

Observers

Twenty-one observers were recruited and given 30 minutes to complete the experiment online. Observers provided informed consent and were compensated with course credit. Four observers that scored less than 8/11 in the training trials, or spent under 15 minutes doing the task, were removed from the analysis, leaving a final sample of seventeen. Research ethics was approved by the University of Sydney Human Research Ethics Committee.

Procedure

Observers were initially presented eleven training trials that contained one shape with a modified bump and one shape with a smooth continuous bump. During the training phase, both stimuli had the same illumination direction. These trials were designed to provide a simple judgement to explain the shape changes they should be sensitive to during the experiment, and to provide a means of filtering observers that either misunderstood or were not adequately engaged in the task.

During the experimental phase, observers were presented all illumination pairs for each material condition and were instructed to select the image that had the stronger impression of a segmented ridge, or ridge consisting of two bumps, in its centre. Stimuli were organised in a factorial design consisting of 11C2 illumination pairs by three materials by three repetitions of each trial, generating a total of 495 trials.

Results and discussion

Since the different materials were never directly compared in the experiment, data for each of the three material conditions were analysed separately using within-subjects ANOVAs. The frequency that each stimulus was rated as having a stronger impression of two segmented bumps in its centre is presented in Figure 34 with illumination direction varying across the x-axis and the separate plots describing the data for each of the three materials.

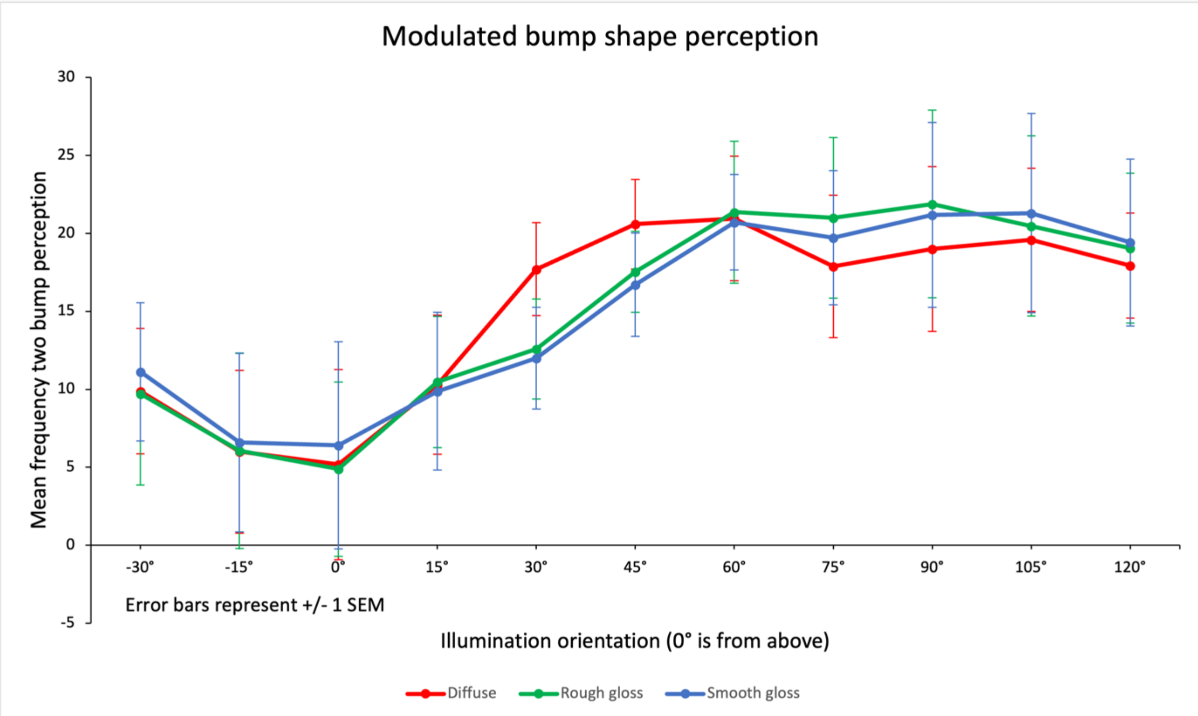


Figure 34. Data from Experiment 4. Mean frequency that observers selected stimulus as generating a more compelling impression of two bumps in its central ridge is presented on the y-axis with illumination direction varying along the x-axis.

Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated for the matte, X^2 (54) = 154.28, $p < .001$, rough glossy, X^2 (54) = 179.28, $p < .001$, and smooth glossy, X^2 (54) = 171.05, $p < .001$ stimuli. The following statistics are reported with adjusted degrees of freedom using Greenhouse-Geisser correction. There was a significant effect of illumination direction on the perceived shape of the diffuse stimuli, $F(1.76,28.14) = 28.06$, $\eta_p^2 = .637$, $p < .001$. Significant effects were also observed for the rough glossy stimuli, $F(1.67,26.66) = 25.17$, $\eta_p^2 = .611$, $p < .001$, and the smooth glossy stimuli, $F(1.81,28.92) = 21.54$, $\eta_p^2 = .574$, $p < .05$. Considering the data in Figure 34, the effects of illumination direction may be explained by a greater tendency to misperceive the modulated bump stimuli as containing a smooth continuous ridge when the illumination is coming from above (i.e., perpendicular to the axis of depth variation across the bump), and that this relationship is largely consistent across the material categories (i.e., there is no evidence that the addition of a specular component promoted shape stability in these stimuli).

Experiment 4 demonstrates a novel non-affine transformation of shape caused by the alignment of illumination with an axis of curvature of a surface feature. If consistent with the experimenters' view, our observers also perceived the near-complete abolition of the two bumps when the surface was illuminated from above. Unlike preceding demonstrations, this failure of shape constancy is not caused by the complete destruction of image structure, but rather from a changing ratio in the contrast of shading gradients generated by surface orientation variation relative to the dominant illumination direction. Here, the failure of shape constancy occurs when the illuminant is perpendicular to the ridge's axis of elongation, causing an accentuation of the overall cylindrical form of the ridge that masks the appearance of the subtle isotropic bumps along the ridge. The angular difference between the illuminant and surface normals generating the two bumps is smallest when illumination is directed from above, causing a significant reduction in the contrast representing these bumps, particularly in comparison to the high contrast gradient generated by the overall cylindrical form in this configuration. Similarly to our previous demonstrations, this failure of shape constancy requires a particular configuration of the illuminant and surface geometry, occurring only when the illumination is approximately coming from above.

This pattern of shape misperception was present for all material categories. Since there were no cross-material comparisons in this experiment, main and interaction effects of material were not analysed, however we note the increasing variance of our data and decreasing effect sizes through the progression to rough and smooth glossy material. We avoided cross-material comparisons in this experiment to promote observers' attention toward shape rather than other differences between stimuli.

Image analysis

We examined how the ratio of shading gradient contrasts describing a smooth ridge or a 'mountain range' containing two peaks by sampling image intensities across a vertical and horizontal cross-section of the region of interest investigated by this experiment for our diffuse stimuli, for each tested illumination direction (see Figure 35). For the vertical cross-section, which captures the overall cylindrical form of the shape feature, we calculated the contrast as the difference between the maximum and minimum pixel value. There was a strong negative correlation, $r = -.79$, between these shading gradient contrasts and observer judgements (i.e., a high contrast along the vertical cross-section was less likely to be perceived as two bumps).

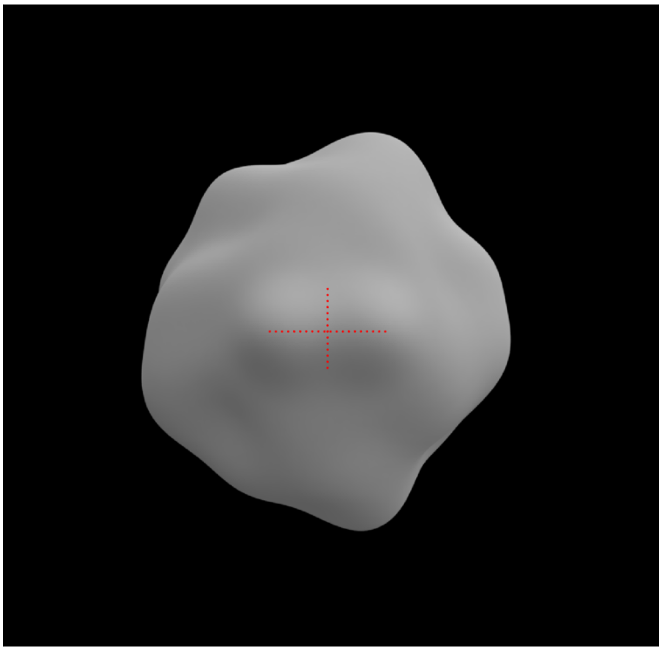


Figure 35. Cross-sections from which shading contrast were calculated. Higher shading contrast across the vertical cross-section was negatively correlated with observer perception of a ridge containing two bumps (i.e., more like a smooth ridge), whereas contrast across the horizontal cross-section was positively correlated with a percept of two bumps.

Since our horizontal cross-section contains two local maxima of image intensity, contrasts were calculated in a different manner to better capture the overall variation across its entire length. Plots of intensity variation across the horizontal cross-section were ‘flattest’ when the stimulus was illuminated from overhead (i.e., 0°, see Figure 36). All plots possessed the same sinusoidal form and varied in amplitude as illumination orientation approached 90°. We calculated contrast here with reference to the flattest plot as the sum of absolute deviation from the mean of the intensity profile of the stimulus illuminated at 0°, which was a pixel value of 138.75. These contrasts were more strongly correlated with observer judgements, r = .92, than the correlation observed for the vertical plot. Due to the differences in the image intensity profiles across each cross-section, as well as the different length they span, we were unable to directly analyse whether the ratio of their contrasts is more predictive of observer judgements. Despite this, the contrast of gradients along either cross-section were highly predictive of our data, suggesting that they may both play an important role in the perception of these stimuli.

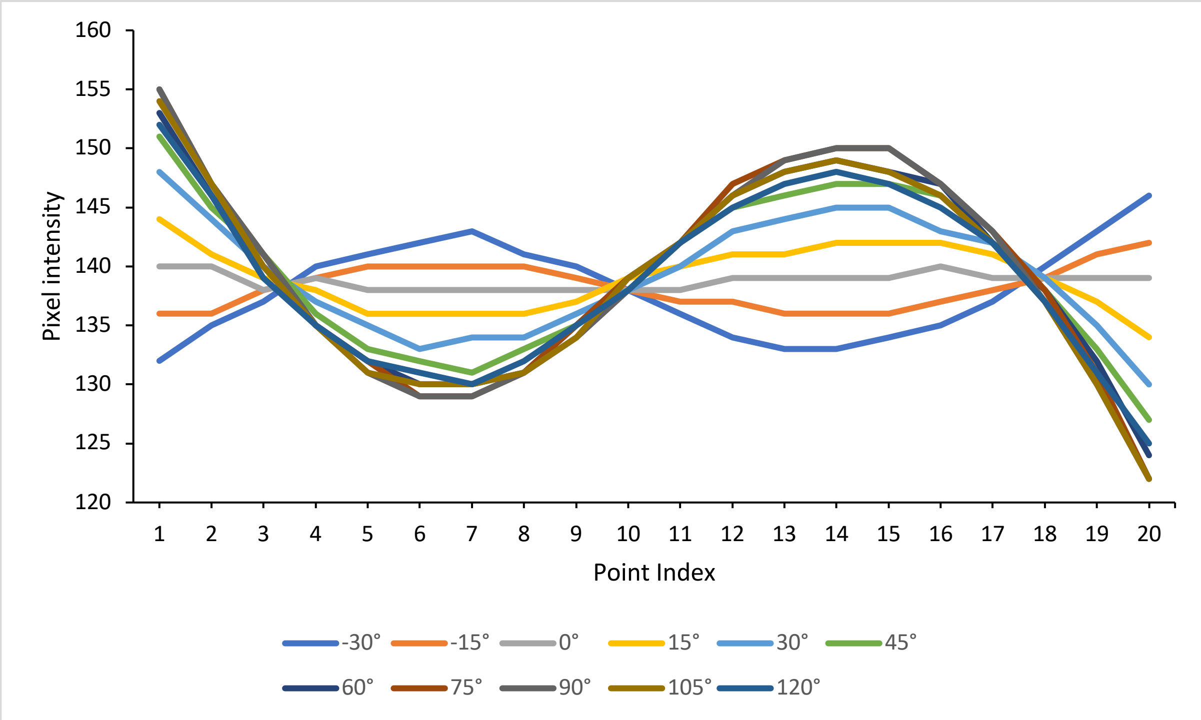


Figure 36. Plot of pixel intensities along horizontal cross-section (see Figure 35. Cross-sections from which shading contrast were calculated. Higher shading contrast across the vertical cross-section was negatively correlated with observer perception of a ridge containing two bumps (i.e., more like a smooth ridge), whereas contrast across the horizontal cross-section was positively correlated with a percept of two bumps.). Plots are separated by the illumination orientation of the sampled stimulus.

We repeated the image analysis conducted in Experiment 1 on Experiment 4's stimuli. Experiment 4's stimuli differ from those in Experiment 1 in that the shape changes observed depend on relatively small variations in contrast ratio for gradients specifying the overall ridge-like appearance of the surface from the finer bumps along the ridge, and we also test glossy materials that contain high frequency structure absent in pure matte surfaces. We varied the values of two parameters in this analysis, which we refer to as 'neighbourhood' and 'cut-off' for brevity. The neighbourhood value specifies the size of the square kernel as the distance of the computed pixel from the kernel's horizontal and vertical edges (i.e., a neighbourhood value of fifteen specifies a thirty-by-thirty kernel). The cut-off value specifies the minimum value for a pixel to be marked in red, following conversion of the image analysis output to an 8-bit greyscale image. In Figure 37, we present the output of this analysis for a subset of our stimuli. Each image matrix varies the size of the kernel neighbourhood by row, and the pixel cut-off value by column. The left columns of image matrices show stimuli illuminated from above that generated the strongest percept of a smooth ridge, and the right columns show stimuli illuminated from the left that generated the strongest percept of two bumps. Rows of matrices separate the different surface materials, from top to bottom, pure matte, rough glossy, and smooth glossy. The first row of matte images prompted our closer analysis of how the values of the free parameters affect the output of our critical contour computation. The original cut-off and neighbourhood values of fifteen used in Experiment 1 yielded minimal structure in the output likely due to the lower contrast of shading gradients in Experiment 2's stimuli. The middle cells of the matte matrices (cut-off = 10, neighbourhood = 15) marked by green boxes produced the best reproduction of perceived shape changes, with a smooth ridge on the left and segmented ridge on the right both clearly defined. Results were considerably poorer for the two sets of glossy stimuli where high intensity specular reflections of the top half of the illuminant introduced structure that would not be interpreted as shape by human observers. This clutter is present in all images but using high cut-off and neighbourhood values begins to capture the key shape changes occurring in the central ridge (bottom right cells).

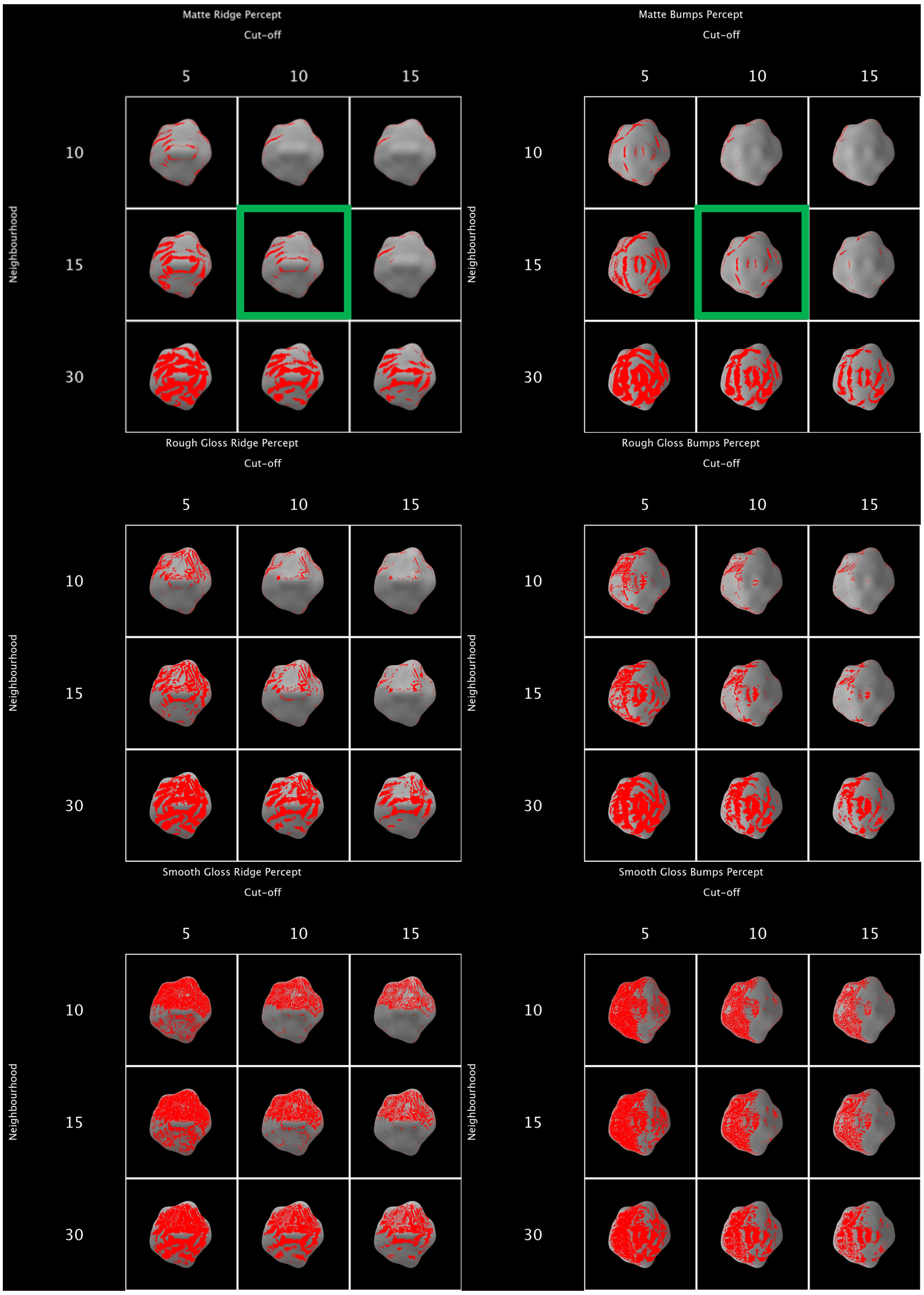


Figure 37. Summary of our critical contour image analysis for Experiment 4's stimuli. The left column of image matrices presents stimuli illuminated from above, in which the central feature more closely resembled a smooth ridge, and the right column presents stimuli illuminated from the left, in which the central feature more closely resembled a pair of segmented bumps. Surface material varies by rows of image matrices. Within each matrix, the cut-off value for marking a pixel red varies along the columns and the size of the kernel varies along the rows. The two images marked by green boxes best demonstrate how critical contours describe shape perception. On the left, the central bump is bounded by an elliptical contour, capturing the weaker perception of two bumps (or stronger perception of a smooth ridge) reported by our observers. On the right, vertical contours appear to segment the ridge in three parts (a left bump, a depression, and a right bump), consistent with observers' compelling perception of two bumps for these stimuli.

Kunsberg and Zucker (2021) address the extraction of critical contours from glossy and purely specular materials. For purely specular materials, they refer to demonstrations by

Fleming et al. (2004, see Figure 38) and show that tracing integral curves through the structure tensor direction achieves similar results to analyses performed on images of matte surfaces. However, their demonstrations using dielectric glossy materials sampled from Norman et al. (2020) bear little resemblance to our stimuli, possessing only specular highlights with minimal discriminable reflections of the illuminating environment. This could be due to different methods of tone mapping the specular reflections. The spherical images of natural environments used in many perception experiments often have high dynamic ranges (e.g., the 'Grove' light map used in this experiment has a dynamic range of 5000:1), far exceeding what may be reproduced with current display technology. This causes images of specular reflections to be 'blown out' with much of the high frequency, low intensity structure of the illuminating environment becoming too compressed to be perceived, resulting in isolated glossy highlights. Tone mapping involves a non-linear scaling of image intensity to better capture the true appearance of the photographed environment, managing the substantial clipping of intensities when attempting to display an unadjusted HDR image on a monitor. These images lack much of the high frequency structure that would have interfered with computation of critical contours, as in our stimuli. Although Kunsberg and Zucker's demonstration using specular materials shows that the issue posed by high frequency specular reflections is tractable, in the case of dieletric glossy materials, this would require a non-trivial segmentation of the image structure generated by the specular and diffuse components. This problem has received some attention in the computational literature (e.g., Roman et al., 2020; Zokay & Saylani, 2022), however it is unclear how this work would translate to human vision, particularly given the frequency with which human observers conflate distal causes of image formation (e.g., Mooney & Anderson, 2014; Todd & Mingolla, 1983).

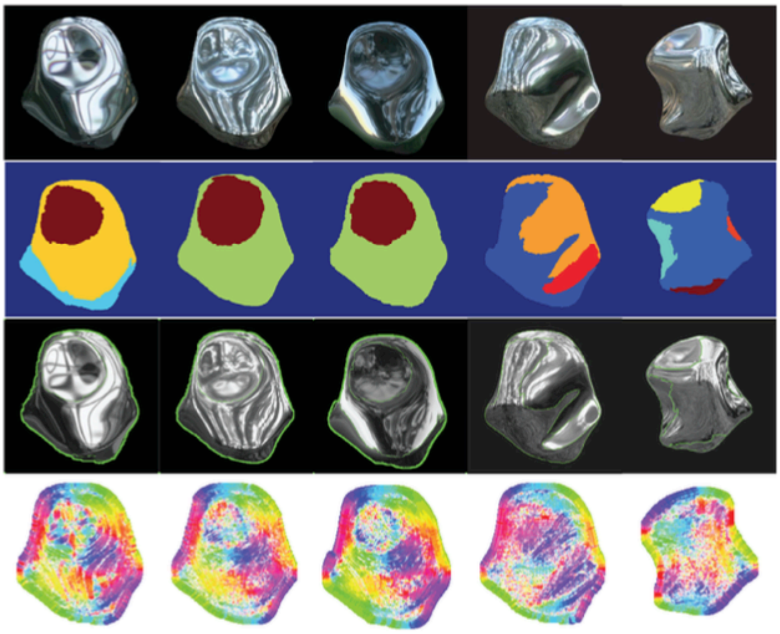


Figure 38. Extraction of critical contours from specular images (from Fleming et al., 2004). Row 2 shows a segmentation of images according to integral curves through the structure tensor direction (row 3). This segmentation closely corresponds to the apparent bumps and dents on the surfaces. Row 4 shows the colourised orientation map. Figure from Kunsberg and Zucker (2021).

Experiment 5

In Experiment 4, we demonstrated compelling transformations of perceived shape due to the alignment of illumination direction with surface geometry, beyond the simple deletion of surface features. In Experiment 5, we return to testing the perception of ridges with near-zero curvature along the minimal axis, this time testing how the deletion of certain shading information due to illumination-geometry alignment affects the perception of adjoining surface features. Stimuli consist of a cross shape constructed by the intersection of two smooth ridges at 90° to each other. In these stimuli, aligning the illuminant with one of the ridges will result in the deletion, or minimal shading, of that ridge, while the ridge running in the orthogonal direction will be maximally shaded. Of particular interest in this experiment is what will occur at the point of intersection between these ridges. We predict perception to resemble ground-truth (i.e., a smooth, continuous cross) under oblique illuminations. When the illuminant is aligned with one of the ridges, we predict a perceived flattening of the aligned ridge and a resultant segmentation of the unaligned ridge into two halves due to an apparent depression where the ridges meet in the centre.

Methods

Stimuli

Stimuli were generated using the same base shape as Experiments 3 and 4. Intersecting vertical and horizontal sigmoid ridges were embedded in the centre of the shape, generating a smooth, continuous cross. Catch stimuli were generated by making either the horizontal or vertical ridge have a higher relief than the other ridge, smoothly descending downward where it intersected with the other ridge. In these stimuli, the higher 'ridge' was segmented in half in the centre, while the lower ridge had a level height across its length. Diffuse and glossy versions of the stimuli were rendered using the same procedure as Experiment 5 (see Figure 39 for a sample of stimuli used in this experiment).

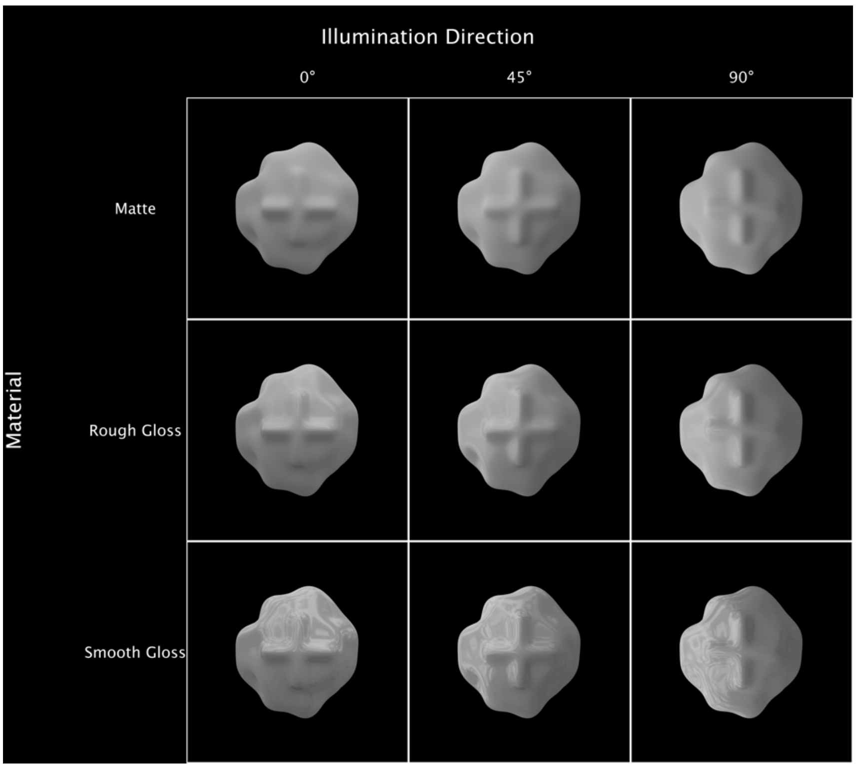


Figure 39. Sample of the cross ridge stimuli used in Experiment 5. Illumination direction varies by column and surface material varies by row. At 0° and 90°, when the illuminant is aligned respectively with the vertical and horizontal ridges, results in a substantial reduction in perceived relief of that ridge. This also results in the apparent segmentation of the unaligned ridge into two parts where the ridges intersect at the centre.

Observers

Sixty-three observers were recruited and given 30 minutes to complete the experiment online ($N = 15$) or in-person ($N = 48$). Observers provided informed consent and were compensated with course credit. Thirty-eight observers that correctly responded to fewer than 9/10 of the training trials and/or fewer than 18/20 on the catch trials were removed from the analysis, leaving a final sample of twenty-five observers. Research ethics was approved by the University of Sydney Human Research Ethics Committee.

Procedure

Observers were presented all illumination pairs for each material condition and were instructed to select the image that had the stronger impression of a smooth, continuous cross in its centre. To demonstrate the task and clarify questions, observers initially completed a training phase of ten trials in which each pair consisted of an experimental stimulus and catch stimulus to ensure that one of the two stimuli always had a clear segmented ridge. Both stimuli in the training and catch trials had the same illumination direction. The training phase was done under the supervision of the experimenter for experiments running in-person where any necessary clarifications about the task could be provided. Examples of clarifying instructions were providing hand-drawn side views of a smooth, continuous ridge, and a segmented ridge dipping in the centre. To explain how to make judgements in cases where neither stimulus had a smooth, continuous cross, observers were instructed to select the stimulus that more closely resembled this ideal case, or if you were to imagine these images as being clay objects, which would require less work to mould it to the ideal case.

Once observers were satisfied that they understood the task, they proceeded to the main experiment. Twenty catch trials were included amongst the experimental trials. The experimental trials presented all illumination pairs for each material condition.

Results and discussion

Since the different materials were never directly compared in the experiment, data for each of the three material conditions were analysed separately using within-subjects ANOVAs. The frequency that each stimulus was rated as having a stronger impression of a smooth continuous cross in its centre is presented in Figure 40 with illumination direction varying across the x-axis and the separate plots describing the data for each of the three materials.

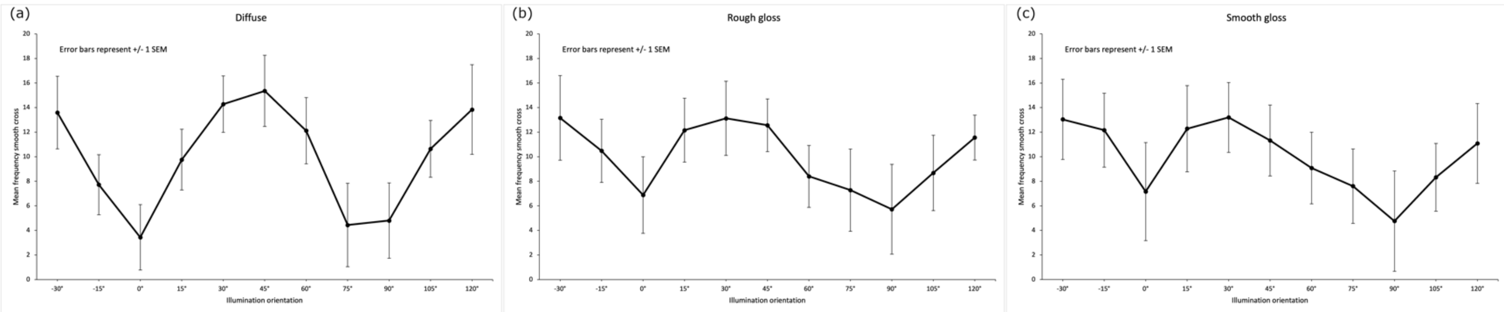


Figure 40. Data from Experiment 5's paired comparison task. The mean frequency that the stimulus was selected as having a stronger impression of a smooth, continuous cross is presented on the y-axis and the illumination orientation varies along the x-axis. Each graph shows data for each material condition. Error bars denote +/- 1 SEM.

Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated for the diffuse, X^2 (54) = 94.56, $p < .001$, rough glossy, X^2 (54) = 109.73, $p < .001$, and smooth glossy, X^2 (54) = 95.51, $p < .001$ stimuli. The following statistics are reported with adjusted degrees of freedom using Greenhouse-Geisser correction. There was a significant effect of illumination direction on shape perception for the diffuse stimuli, $F(5.15,123.61) = 52.55$, $\eta_p^2 = .686$, $p < .001$, the rough glossy stimuli, $F(4.73,113.52) = 19.88$, $\eta_p^2 = .453$, $p < .001$, and the smooth glossy stimuli, $F(5.05,121.09) = 16.51$, $\eta_p^2 = .408$, $p < .001$.

As with Experiment 4, significant effects of illumination direction were observed for all tested materials. Judging the plotted data, the perception of a smooth continuous cross is consistently disrupted when the illumination is orientated to the axes of minimal curvature of either ridge (i.e., illumination from above at 0° and from the left at 90°). Unlike Experiment

4, the disruption to perceived shape is particularly strong for the diffuse stimuli relative to either glossy material suggesting that in this instance, the addition of specular reflections stabilised perceived shape across changes in illumination to some degree. This is supported by the decreasing effect size, which is particularly pronounced between the diffuse and two glossy material categories.

The apparent depression in the centre of the unaligned ridge for the 0° and 90° illuminations implied by the reduced perception of a smooth continuous cross is an interesting result since there is no variation in image intensity along the crest of this ridge and should therefore indicate no variation of slant (i.e., a smooth, continuous ridge). The appearance of the shading running orthogonal to the ridgelines appears more informative here, which is of significantly lower contrast for the aligned ridge, and does discontinue in the centre of the unaligned ridge, where it intersects with the other ridge.

Image analysis

We first tested whether the relative contrast produced by each ridge could explain the pattern of our observers' judgements. We sampled image intensities across each ridge at the locations shown in Figure 41 for our matte stimuli under each tested illumination direction. We calculated the contrast of each cross-section as the difference between the maximum and minimum pixel value and their ratio by dividing the smaller value by the larger value, limiting these ratios to a value of 1 or less. These ratios strongly correlated with observer judgements, $r = .96$, with higher ratios predictive of a percept of a segmented ridge with a depression at the point of ridge intersection. This supports our view that relative depth judgements of intersecting surface geometry can signal a change of slant (in our case, a depression) in the absence of shading variation *in the direction* of the perceived change of slant.

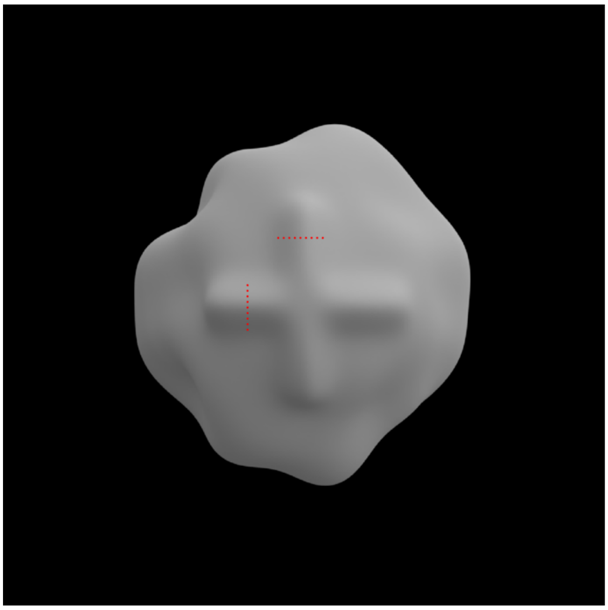


Figure 41. Cross-sections from which shading contrast across each ridge were calculated. Observer percepts of a segmented ridge were stronger where shading contrast along each ridge differed significantly.

Image analysis was conducted using the same method as in Experiment 4. The output is presented in Figure 42. As with previous applications of this image analysis, useful output was generated for images of matte materials. However, the parameter values that produced the best results for these stimuli differed from those in Experiment 4 (cut-off = 10, neighbourhood = 10), despite considerable similarities between the two sets of stimuli. Both sets were normalised such that each image's highest pixel value was matched and the base shape used was identical for both experiments, differing only in the shape feature embedded in their centres. This demonstrates that development of a method to select 'appropriate' values for these parameters from the outset will not be able to rely on global features of the images, instead requiring consideration of salient local features in the image. These salient features would presumably need to be identified by human observers, further diminishing the prospect of this calibration being performed from the outset.

Our analysis was much less successful in processing both glossy materials with the only success we identify being in the middle right matrix at parameter values cut-off = 15, neighbourhood = 15. Although our results for Experiment 4's glossy stimuli were also not particularly compelling, our results here are even less salvageable. A possible reason for this is that the directions of surface normal variation in our cross stimuli are less variable than the Experiment 4's modulated bump, varying primarily in the horizontal and vertical directions. This may make the contours generated by the vertical and horizontal ridges more 'sticky' as the specular reflections may only stretch perpendicular to the single direction of curvature. However, this 'failure' reflects an important difference in the pattern of results observed across Experiments 4 and 5. Effects of material were only observed in the latter, where our image analysis is of lower quality. The visual system does not appear to compute local surface orientation from the point-by-point variation in image intensity (Mamassian & Kersten, 1996), instead generating a coarser, qualitative description of shape. Some of our image analyses generate structures corresponding to Kunsberg and Zucker's (2021) definition of a 'bump'. A typical 2-cell begins with a local maximum that follows integral paths to two saddle points, which then converge on a local minimum (see Figure 43). These saddles will gradually merge as the surface feature becomes more 'bump-like', with an exemplary bump defined as a 2-cell in which the two saddles have merged into one, pushing the local minimum of the slant function (or maximum of the intensity function) to the centre or edge of the closed loop. We can see a good approximation of this behaviour in the cell located in the first row, second column of the top left image matrix in Figure 42. The horizontal ridge is clearly bounded by the red critical contours, which begin to curve inward as the horizontal ridge approaches the vertical from either direction, almost generating two enclosed loops. This is consistent with observer judgments of a ridge segmented into two distinct halves.

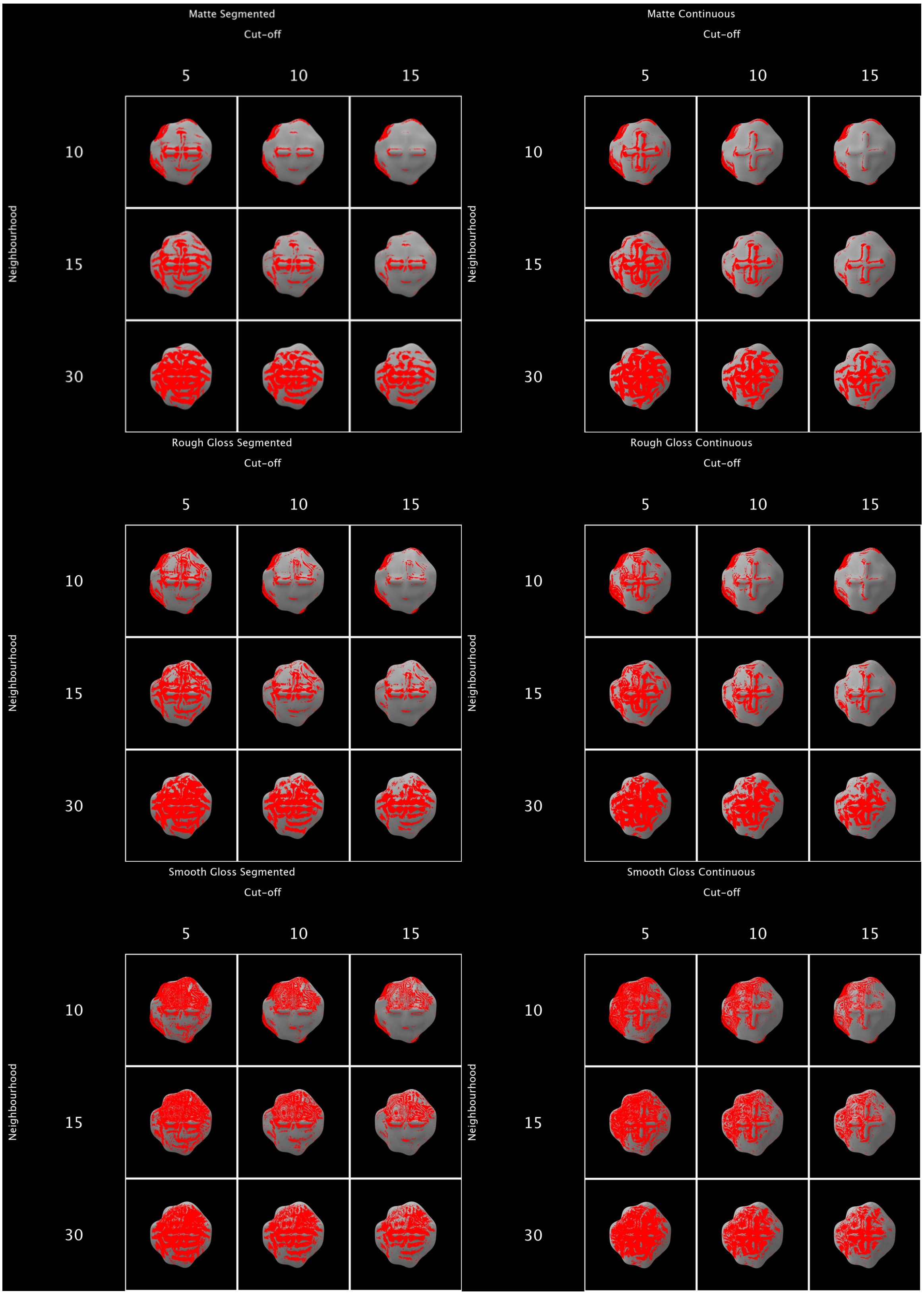


Figure 42. Summary of our critical contour image analysis for Experiment 4's stimuli. The left column of image matrices presents stimuli illuminated from above, in which the central feature more closely resembled a smooth ridge, and the right column presents stimuli illuminated from the left, in which the central feature more closely resembled a pair of segmented bumps. Surface material varies by rows of image matrices. Within each matrix, the cut-off value for marking a pixel red varies along the columns and the size of the kernel varies along the rows.

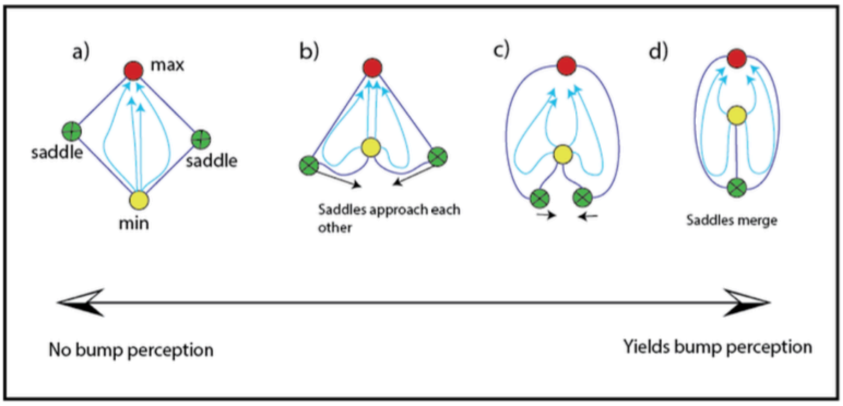


Figure 43. Kunsberg and Zucker's (2021) definition of a 'bump' as a special 2-cell of an MS complex. As a surface feature more closely resembles a bump, the two saddle points of the 2-cell are gradually drawn together, with the exemplary bump defined as a 2-cell in which the saddles have merged completely.

Experiment 6

This experiment aimed to investigate characteristics of illumination other than direction on the stability of perceived shape. Work by Zhang et al. (2015, 2018, 2019) observed effects of illumination type, defined by the relative strengths of their spherical harmonic orders, on the perception of material properties. Comparing 'focus' (i.e., strongly directed illumination with low strength zeroth order) and 'brilliance' illuminations (i.e., multidirectional illumination where the ratio of the sum of the higher orders is high relative to all orders), Zhang et al. (2019) found that objects rendered using a 'brilliance' light map were perceived as significantly more glossy. Since 'brilliance' light maps are composed of multiple strong light sources, they increase the coverage of the surface by specular highlights, increasing perceived glossiness (Marlow et al., 2012).

We propose two hypotheses for how a 'brilliance' light map may influence shape perception. Our previous experiments demonstrated significant failures of shape constancy dependant on the alignment of the illuminant and surface geometry. Since 'brilliance' light maps contain multiple strong light sources, the surface is illuminated from multiple directions, even when the primary light source is aligned with surface geometry. This may introduce additional shading gradients otherwise absent in a more directional light map, as well as shadowing and vignetting, thereby promoting shape constancy across variation in the orientation of the illuminant. However it is possible that a 'brilliance' light map will not significantly promote shape stability of matte surfaces, since the appearance of matte surfaces is largely determined by the lower order spherical harmonics (Ramamoorthi & Hanrahan, 2001), thereby discounting the high frequency structure present in the light map that promoted the perception of gloss in Zhang et al.'s (2019) study. Where smooth, specular materials reflect light at an angle equal to the angle of incidence, the appearance of matte materials depends on all incident light, meaning a diffuse convolution of a light map better predicts the intensity of light reflected by a matte surface. In Figure 44, the 'Ennis' (top row) and 'Grace-new' (bottom row) light maps are presented, along with their diffuse convolutions in the right column. Notice that the diffuse convolution eliminates the high frequency structure of both light maps, meaning that the net illumination reflected by a matte surface may be more directional than would appear in the unaltered images.

Another consideration is how the higher diffuseness of the 'Grace-new' light map may influence judgements of shape. Of the light maps analysed by Zhang et al. (2019), the 'Grace-new' light map had the highest brilliance metric, but was also moderately diffuse (see also Xia et al., 2017). Shading intensity varies with the angle subtended by the surface normal and illumination vector, providing information about shape. This cue is weakened as the directedness of the illuminant is weakened (or diffuseness strengthened) as points all over the surface receive illumination from a widening range of directions. As the illuminant becomes increasingly diffuse, image intensities vary as a function of surface aperture, or the amount of light reaching each point, making concavities darker than convexities (see Section 1.1.2.). Since shading variation as a function of illumination orientation decreases as a function of illuminant diffuseness, we may observe greater shape constancy under the 'Grace-new' light map. Photographic lighting techniques manipulate the directedness of the illuminant to soften undesirable stark shadows or appearance of skin texture. Larger light sources produce softer shadows as points on the subject are illuminated from a wider range of angles. Furthermore, positioning a light source at a greater distance produces harsher shadows as light rays reaching the subject are compressed toward a collimated beam (see Hunter et al., 2021 for a review).

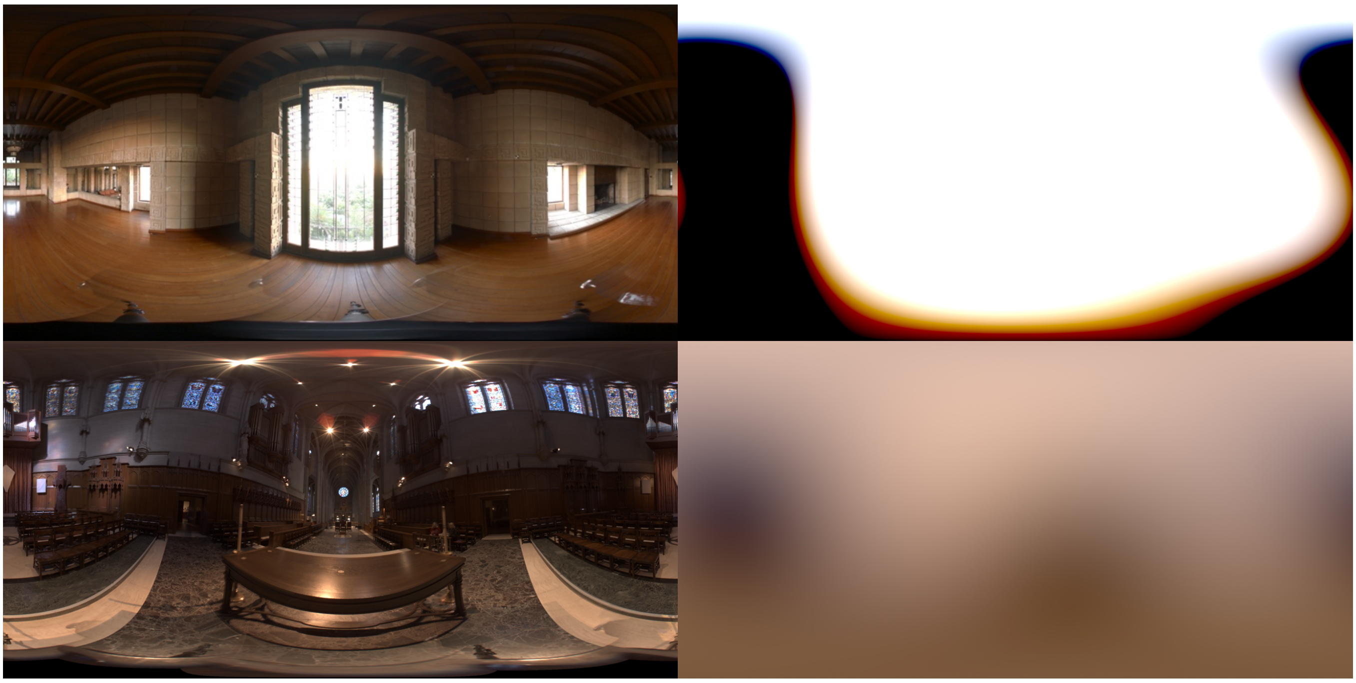


Figure 44. Light maps used in Experiment 6 and work by Zhang et al. (2019). The top row shows the 'Ennis' light map, defined as the 'focus' light map, where most light is directed in a tight range of directions. The bottom row shows the 'Grace-new' light map, defined as the 'brilliance' light map, containing multiple, distributed prominent light sources. The right column shows a diffuse convolution of the respective light map, and visualises the net incident light from each direction that influences the amount of light reflected by a matte surface. Notice that the diffuse convolution eliminates the high frequency structure of the light maps, and the 'Grace-new' light map resembles a more directional illuminant.

Methods

Stimuli

Stimuli were generated by rendering a randomly perturbed sphere from Experiment 1 illuminated by the 'Ennis' and 'Grace-new' light maps used by Zhang et al. (2015, 2018, 2019, 2020) as the 'focus' and 'brilliance' light maps, respectively. Twelve equally-spaced illumination directions were used spanning the entire range of spherical coordinates and surfaces were rendered using pure Lambertian and dielectric smooth glossy material, using the same settings as previous experiments. Examples of our stimuli are presented in **Error! Reference source not found.**.

Observers

Fourteen observers were recruited and completed the experiment within 30 minutes. Observers provided informed consent and were compensated with course credit. Three observers who failed to correctly respond to at least 8/10 catch trials were removed from analysis for a final sample of eleven. Research ethics was approved by the University of Sydney Human Research Ethics Committee.

Procedure

On each trial, observers were presented with two pairs of images and instructed to select the pair that had the least similar shape. 12C2 illumination pairs for each material, with two repetitions, yielding a total of 264 experimental trials. Images on each trial had the same material and same pair of illumination directions, differing only in illumination type ('Ennis' or 'Grace-new'). Ten catch trials were included where one of the two presented pairs contained a different shape.

Results and discussion

Since material categories were never directly compared in this experiment, separate analyses were conducted for the matte and glossy stimuli. We performed z-tests to determine whether the mean probability observers selected each type of illumination significantly differed from chance. In Figure 45, the frequency that each stimulus pair was selected as having the least similar shape varies along the y-axis with bars denoting stimuli illuminated using the 'Ennis' and 'Grace-new' light maps. The graph on the left shows data for the matte stimuli and the graph on the right shows data for the glossy stimuli.

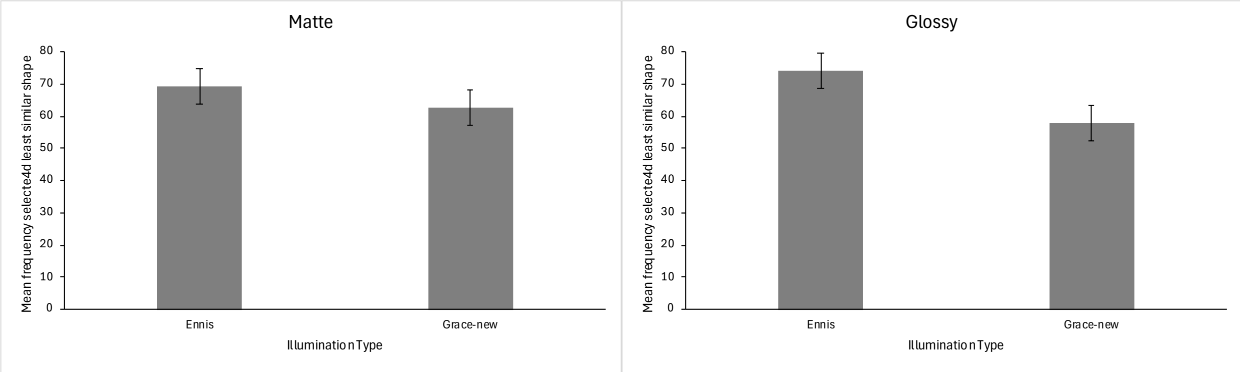


Figure 45. Data for the paired comparison task of Experiment 6. On each trial, two pairs of stimuli were presented and observers selected the pair that had the least similar shape. The mean frequency that each stimulus category was slected as having the least similar shape varies along the y-axis with bars representing the two illumination categories. Graphs separate data for the matte and glossy stimuli. Error bars denote +/- 1 SEM.

There was no significant effect of illumination type observed for the matte stimuli, $z(10) = 0.60$, $p = .547$. There was no significant effect of illumination type observed for the glossy stimuli, $z(10) = 1.47$, $p = .141$.

We conclude that illuminating a surface using a multi-directional ‘brilliance’ light map does not promote stability of shape perception. Although our analysis of the glossy stimuli approached significance, the mean difference observed was in the opposite direction to the reasonable hypothesis that the ‘Grace-new’ light map would promote stability of shape perception based on our argument that a multi-directional illuminant may generate more consistent shading flow across variation in illumination orientation. It is likely that in the absence of compelling variation in shape stability across our stimulus pairs, observers may have attended to other sources of variation in the images to make their judgements, specifically the specular reflections. Given the higher frequency structure present in the ‘Grace-new’ light map, the distribution of specular reflections would be highly variable, accounting for our glossy stimuli results.

The similarity between the diffuse convolutions of the two light maps may explain our null result for the matte stimuli (see Figure 44). The intensity of light reflected by a matte surface is dependent on incident light from all directions and the total irradiance from each point on the sphere of all orientations is better visualised by the diffuse convolution of a light field. There is substantial overlap in the angular distribution of the primary light source between the two convolved light maps with the 'Grace-new' light map being more diffuse and perhaps slightly bi-directional. Despite these differences, the 'focus' and 'brilliance' light maps used in the present and Zhang et al.'s experiments appear to generate similar patterns of shading with matte materials.

Chapter 3. General Discussion

Despite decades of research, it remains unclear precisely how the visual system perceives shape from shading. Inverse optics is an intractable problem requiring solutions to three variables (material, shape, and illumination) from a single input (the image). Research in visual perception attempts to identify cues to these distal causes that are directly accessible from the images themselves to circumvent the issue of inverting the generative processes. This too has proved to be a challenging problem as images are highly variable even though our perception is robust to much of this variability. This suggests that perception is specified by invariant structures that are preserved across such changes to the rendering function. Under certain circumstances, failures of shape constancy also occur. Surfaces can appear sheared toward the illuminant (Christou & Koenderink, 1997) and the addition of blurred specular reflections can cause exaggeration of perceived curvature due to the conflation of gradients produced by specular and diffuse reflectance (Mooney & Anderson, 2014). If certain image invariants are used by the brain to extract shape, it suggests that they should be able to predict these changes.

This thesis further examines conditions predictive of failures of shape constancy. Part of the variation in perceived shape between observers (Todd et al., 2001) and variation due to the orientation of the illuminant (Christou & Koenderink, 1997) has been attributed to affine transformations consisting of scaling in depth and shearing toward the illuminant. However, variation in illumination direction can also cause non-affine changes in perceived shape (Nefs et al., 2005) and the causes and nature of these changes are not well understood. This thesis documents several changes of perceived shape due to the relative orientation of surface geometry and the illuminant. We present six experiments that identify some of the generative conditions that lead to the misperception of shape and provide insights into the image features (or absence thereof) responsible for these failures. Experiments 1 to 3 demonstrated that alignment between the illumination direction and axis of minimal curvature of anisotropic shape features results in the deletion of shading structure (critical contours, Kunsberg & Zucker, 2018) leading to the deletion of those shape features. This effect was not explained by variation in size of mean curvature of shape features (Experiment 3). Experiment 4 showed that variation in the relative contrast of shading gradients results in non-affine failures of shape constancy due to alignment of illumination with directions of different slant variation. Experiment 5 showed that the deletion of shape features leads to a misperception of adjacent surface geometry. These results were robust to qualitative variation in the illumination (i.e., directionality of the illuminant, Experiment 6) and for specular reflectance functions, although we observed reduced effect sizes with the inclusion of specular reflections (Experiments 4 to 6). I summarise our experimental results below.

3.1. Summary of experimental results

Our first experiment suggested that anisotropic shape features that are elongated in one direction, like a cylindrical ridge, may cause substantial failures of shape constancy when illuminated along its axis of minimal curvature. Since the intensity of light reflected by a Lambertian surface is proportional to the angle subtended by the illumination direction and surface normal, illuminating a surface feature along a direction of near-zero curvature results in the local deletion of shading gradients. This observation was supported by analysing the critical contour structure (Kunsberg & Zucker, 2018), demonstrating their overall robustness to variations in the rendering function, and importantly their disruption when failures of shape constancy occur. Experiment 2 demonstrated that the observed effects were not caused by violation of the light-from-above prior. Experiment 3 showed that this was still the case while partially controlling for the size and mean curvature of surface features.

These experiments demonstrated a reduction of the perceived curvature of anisotropic shape features when illuminated along their axis of minimal curvature. The angular variation between surface normals and the illuminant is maximised in the direction of the dominant light source, which accentuates apparent surface curvature in this direction. The deletion of surface features when the illuminant is aligned with the axis of elongation of a cylindrical ridge occurs because there is minimal variation of surface orientation in this direction. Since the surface normal variation along the axis of maximum curvature is approximately parallel to the illuminant in this situation, the contrast of the shading gradient they generate is minimised. Considered locally, this would constitute an affine scaling in depth. However, affine transformations require a uniform scaling and/or shearing in depth across the entire surface, so if the reduction in perceived curvature we observed is confined to the anisotropic shape features, it would qualify globally as a non-affine transformation. A simple transformation of scale is achieved by multiplying the depth of all points by some value such that the additive increase in relief of surface peaks (i.e., regions of high depth value) is larger than that of troughs (i.e., regions of low depth value). A shear is achieved if the multiplicative scaling occurring at all points is oriented in a specific direction (i.e., stretching the surface in that direction). We are unable to verify the nature of these perceived shape changes because did not sample shape judgements over a broader region of the surface. However, informal observation of our stimuli suggests that these shape changes were restricted to a sub-region of the surface.

We also presented examples of non-affine perceived shape changes related to the alignment of surface geometry with the illuminant beyond an affine reduction of perceived relief of anisotropic shape features. In Experiment 4, we presented stimuli containing a ridge-like feature with a slight depression in its centre, resembling a subtle pair of mountains along a mountain range. Since diffuse reflection is a function of the slant angle between the surface normal and illumination direction, variation in surface slant along the direction of illumination is maximally represented in the image through higher contrast shading gradients.

Illuminating the surface from above increases the relative contrast of shading gradients specifying the coarse ridge-like form of the surface feature, masking the pair of bumps that constitute it, whereas the shading gradients specifying the pair of bumps are maximally represented when illumination is oriented in the direction of this slant variation (i.e., horizontally). In Experiment 5, we presented stimuli containing a pair of intersecting ridges (a cross). As with our initial experiments, alignment of the illuminant with a ridge's axis of minimal curvature caused the near-deletion of that ridge. However, this experiment also demonstrated a resultant illusory segmentation of the non-aligned ridge. This was an interesting result as there was no variation in image intensity along the ridge specifying its apparent depression where it intersects with the other ridge, and seems to be caused by the interruption to the steep shading gradients around the ridge's boundary. In Experiment 3, deletion of the anisotropic ridge occurred in the degenerate configuration of near-perfect alignment of illumination with the axis of minimal curvature. Since the humps were approximately isotropic, they were less sensitive to variation in illumination direction and were perceived across most of the tested range of illumination directions. However, when the surface was illuminated from approximately overhead, the percept of a smooth, cylindrical ridge dominated (see Figure 34) due to the relative contrast of shading representative of the overall ridge-like form masking the subtler gradients generated by the humps.

The addition of specular reflections to diffuse shading was found to reduce the influence of illumination orientation on perceived shape although significant failures of shape constancy were also observed for these material categories. Previous research found that diffuse shading and specular reflections provide complementary information, compellingly demonstrated using images of bumpy planes containing no bounding contour. A clear image of a surface is perceived when specular and diffuse reflections are combined where either source viewed in isolation is poorly resolved (Mooney et al., 2019). Furthermore, 'blurred' specular reflections (generated by surface roughness) become conflated with diffuse shading in certain situations (Mooney & Anderson, 2014), increasing the apparent contrast of shading gradients and potentially preserving perception of surface features that may otherwise be deleted. Both Experiments 4 and 5 demonstrated failures of shape constancy for all material categories, systematically occurring where there was some alignment between the illumination direction and surface geometry. However, the effect size of illumination direction on perceived shape was reduced, particularly with the addition of any specular reflection (i.e., the reduction in effect size was largest between the purely diffuse and rough glossy stimuli).

The role of gloss in promoting shape stability is particularly apparent in the data of Experiment 5, with the difference between the aligned horizontal/vertical and remaining illumination directions being more pronounced for the purely diffuse stimuli than the glossy conditions. Since our experiments contained no cross-material comparisons, we were unable to statistically verify material interactions, and the lack of such comparisons also rendered our design less sensitive to effects of material. All comparisons were made within each material category, potentially causing observers to weigh the diffuse component higher in making their judgements. We still believe that such a design was necessary to encourage naïve and inexperienced psychophysical observers to attend to the experiment's dimension of interest, namely shape.

There is evidence suggesting that the effects we observed are robust to variation in the qualitative structure of the illuminant and is largely determined by the orientation of the dominant light source, captured by the first order spherical harmonic of the light field. Zhang et al. (2019) demonstrated an exaggeration of perceived gloss for surfaces illuminated using a multidirectional 'brilliance' light map compared to directional 'focus' and diffuse light maps. We therefore thought it was possible that 'brilliance' light maps would promote shape constancy in our demonstrations given their fine dependence on alignment of illumination with surface geometry. Since 'brilliance' light maps consist of multiple distributed strong light sources, shading gradients may be produced by these additional light sources that would otherwise be absent under a more directional illuminant. We did not find this to be the case with the perceived shape of both purely diffuse and dielectric glossy surfaces largely determined by the lower order structure of the illuminant.

The definition for a failure of shape constancy requires some qualification given the considerable inter-observer variability reported by many studies on shape perception. This inter-observer variability primarily emerges in observers' depth estimations (Todd et al., 2001) and comes as a consequence of the inherent ambiguity in the recovery of 3D depth from 2D images. There is insufficient information in a 2D image to recover metric depth with fidelity prompting theories that focus on the derivation of qualitative shape in both the computational and perceptual literature. Todd et al. (2001) showed that inter-observer variability is almost entirely explained by affine transformations between shape judgements characterised by a shearing or scaling in depth. Given that the human visual system does not appear to extract metric depth, discounting variability across observers that are explained by an affine transformation may be prudent. However, it is important to remember that objects related by an affine transformation are experienced as different shapes for individual observers, such as how material properties predict exaggeration of perceived relief (Mooney & Anderson, 2014), or how spheres and ellipsoids are perceived as distinct shapes.

3.2. Relation to previous work and future considerations

Much of the computational and psychophysical work on shape from shading from the last two decades has shifted from focussing on the recovery of local surface orientation to the identification of salient image features that are robust to variation in the rendering function to account for stability of perceived shape. A number of these approaches generalise the unambiguous information provided along the silhouette of objects by the smooth occluding contour to structures occurring within the object. Smooth occluding contours are curves of maximal surface slant, constraining one of the two surface orientation parameters to 90°, with the remaining tilt parameter being simple to recover from the normal direction to the contour at each point. Suggestive contours (DeCarlo et al., 2003), apparent ridges (Judd et al., 2007), and critical contours (Kunsberg & Zucker, 2018) all attempt to recover curves of high surface slant from shading flow because these regions approximate characteristics of smooth self-occlusions. The rich information provided by contour-like structures in shading is compellingly demonstrated by the efficiency with which line drawings convey shape. Koenderink et al. (2012) demonstrated that observers were capable of interpreting densely sampled surface orientations from only sparse line drawings with striking consistency. This suggests that human perception of shape from shading may be structured by the segmentation of shading flow into qualitative regions rather than integrating over local estimations of surface orientation. Experiments presented in this thesis provide compelling evidence that contour-like structures in shading are predictive of both when failures of shape constancy occur as well as the nature of these failures.

Prior research has begun to describe some of the causes for variation in perceived shape, as well as the nature of these changes. Belhumeur et al. (1997) formally described the inherent ambiguity due to the fact that changes in an image may be the result of a change of surface geometry or illumination conditions. The authors demonstrated that the 3D interpretations of an image could be constrained by a family of shapes related by an affine transformation, such that an affinely transformed geometry produces an indentical image with a corresponding change to the illumination direction. However, this mathematical equivalence may not explain all of human vision. Todd et al. (1996) collected dense samples of local surface orientation of sculptures of human torsos using a gauge figure adjustment task. They compared perceived shape for torsos illuminated frontally and from an oblique direction and found that 85% of the variance in observers' judgements was accounted for by an affine transformation. Clusters in the scatterplot were identified that were related to specific regions of the torso (e.g., the arm, buttocks, shoulderblade), and running the analysis in a piecewise manner increased the explained variance to 95%. This suggests that each part of the surface was affected differently by variation in illumination, and therefore is not explained by an affine transformation of global shape. Our experiments support this view by demonstrating that variation in illumination direction can cause the local deletion of shading gradients (Experiments 1 to 3, and Experiment 5) and changes in the contrast ratio of gradients specifying slant variation in different directions (Experiment 4). Furthermore, the deletion of gradients can affect the perception of adjacent surface geometry (Experiment 5). Our cross ridge stimuli show that deletion of shading specifying one of the ridges results in illusory segmentation of the intersecting ridge.

Most of our experiments describe failures of shape constancy related to the deletion of shading gradients specifying local shape. However, Experiment 4 demonstrated that the relative contrast of gradients specifying slant variation along different directions. Slant variation parallel to the illumination direction is maximally represented in shading gradients and slant variation in the perpendicular direction generates no gradient, explaining the deletion of cylindrical ridges in our experiments. In Experiment 4, we showed that the perception of two humps, specified mostly by slant variation in the horizontal direction, or a smooth ridge, specified by slant variation in the vertical direction, was explained by the degree of alignment of the illuminant with these two directions. The effect we demonstrated occurred over a relatively large region of the surface, and there are possible effects of illuminant-geometry alignment on more local shape judgements. For example, judgements of saddle regions, which are positively and negatively curved along their major axes may be more likely to be judged as convexities or concavities depending on the alignment of the illuminant. This in turn may have consequences on shape judgements of surrounding regions.

The failures of shape constancy reported in this thesis were robust to variation in reflectance function but were attenuated for the glossy stimuli. It is possible that the stabilising effect of specular reflections is underrepresented by our experimental designs which only contained within-material comparisons. This was done to help our naïve observers base their judgements on perceived shape rather than other sources of variation between our stimuli. Although we observed the same patterns of shape constancy failures for diffuse and glossy stimuli, the effect sizes were progressively lower for rough glossy and smooth glossy stimuli. Since effect sizes were consistently smallest for the smooth glossy stimuli, it is unlikely that our results are explained by the exaggeration of perceived curvature due to the conflation of blurred specular highlights with diffuse shading (Mooney & Anderson, 2014), but rather by how the two reflectance functions cospecify surface geometry. However, our experiments did not titrate magnitudes of surface curvature and specular roughness, and it is possible that the effects observed by Mooney and Anderson (2014) could occur for the more complex surface geometries used in our experiments.

The experiments reported in this thesis suggest an important role of contour-like structures within shading flow in structuring human perception of shape. In this section, we provide demonstrations following up from the cross ridge stimuli of Experiment 5. Our result of the near-deletion in the perception of the ridge aligned with the illumination direction and subsequent segmentation of the other ridge in two parts appears to be explained by a reduction in contrast of the convergent shading flow that defines a contour around the ridges' boundaries. This reduction in contrast informs the visual system of a reduction in curvature and surface relief. A question for future investigation is whether depth cues provided by the ridge boundaries may be preserved across variation in illumination direction through modification of surface geometry close to the ridge boundary. In Figure 46, we demonstrate an approach of introducing small bumps along the ridge boundary. Since these bumps are fairly isotropic, they generate strong shading gradients that are robust to variation in illumination direction. In these images, the depth cue generated by the bumpy edge weakly propagates to the ridge centre. Perception is poorly resolved and the bumpy edge can appear segmented from the ridge to which they are smoothly connected, rather they appear as small bumps separated by a trough.

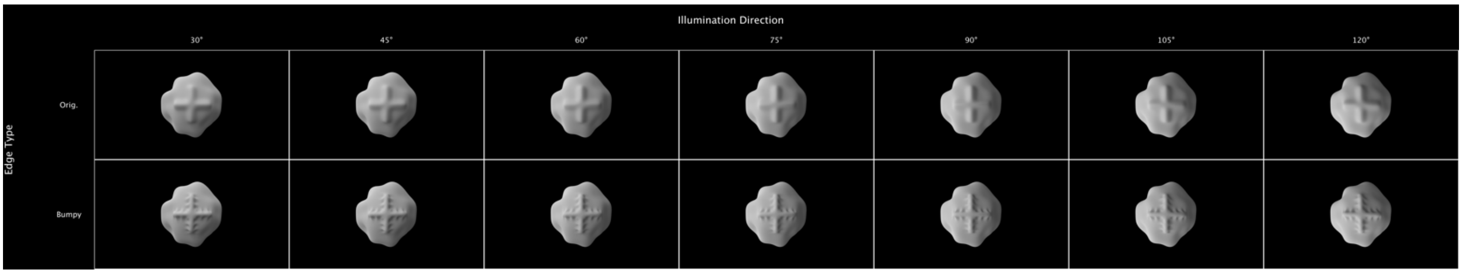


Figure 46. Original cross ridge stimuli from Experiment 5 (top row) compared with modified bumpy edge stimuli (bottom row).

Another condition we examined is whether the presence of mesoscale texture would promote stability of perceived shape. Few naturally occurring materials are perfectly smooth; they typically contain some surface texture of small bumps that are considerably smaller in scale to surface features (‘mesotexture’). Both albedo and mesoscale texture provide shape cues due to the pattern of compression they produce, similar to specular reflections as reviewed in Section 1.1.2. Due to foreshortening, the spatial frequency of surface texture increases proportionally to surface slant such that texture is compressed in the direction of slant. Mesotexture is often isotropically curved, and therefore will generate image structure that is robust to variation in illumination direction. While shading depends on the perpendicular component of the light vector, mesotexture also provides information about the tangential component (Koenderink, Doorn, et al., 2007). Although observers are capable of exploiting this cue to accurately estimate illumination direction (Koenderink et al., 2004), it remains unclear whether this information can stabilise shape perception. Koenderink et al. (2007) showed that estimation of illumination direction is often grossly inaccurate when viewing anisotropically-textured surfaces due to how these surfaces polarise the illuminance flow cue (see Section 1.4.2.). Despite the presence of an illumination direction cue in our experiments in the form of the self-occluding rim of a globally convex surface, we still observed significant deviation in the perception of anisotropic shape features. This suggests that in our experiments, information about the illuminant did not significantly influence perception of shape. A comparison with mesotextured surfaces where the illuminant cue is present over the entire surface is useful to address the question of the degree to which observer estimates of distal factors influence perception overall.

Since surface texture is robust to variation in illumination direction, it may provide information that stabilises shape perception in our demonstrations. In Figure 47, we present the cross ridge stimuli illuminated from above (left column) and the left (right column), resulting in the deletion of the vertical and horizontal ridge, respectively. The top row shows our original smooth stimuli and the following rows show two examples of these shapes with mesoscale texture of different scales. Contrary to the above, mesoscale texture does not appear to promote stability of perceived shape, and may even strengthen the failure of shape constancy due to the texture's obscuring of the low contrast gradients specifying the boundary of the deleted ridge. However, this does not settle the matter for the role of mesotexture in stabilising shape perception as the scales of mesotexture and macro variations in surface orientation may interact where some configurations may obscure information provided by shading gradients, while other configurations may provide useful information for the visual system.

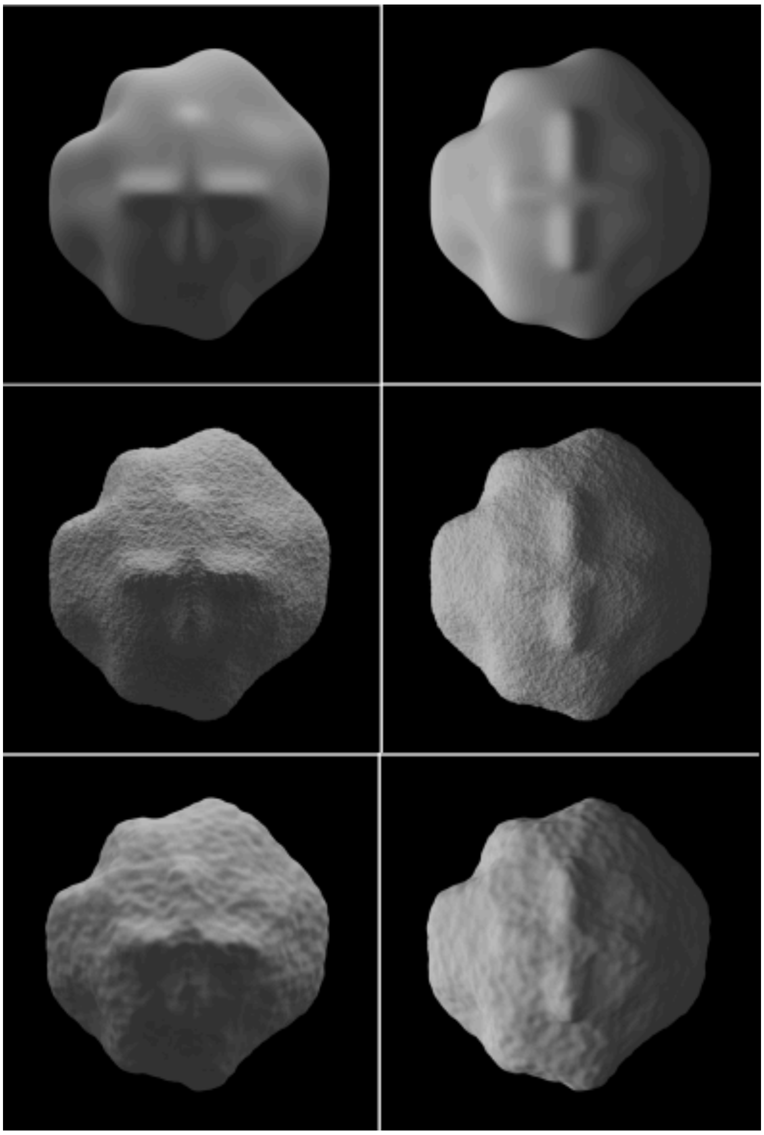


Figure 47. Original smooth cross ridge stimuli (top row) and versions rendered with mesoscale bumps of different scales (second and third rows). The images in the left and right columns are illuminated from above and from the left, respectively, resulting in the deletion of the aligned ridge. The presence of mesoscale bumps seems to further promote failure of shape constancy. Images courtesy of Phillip Marlow.

Several studies have demonstrated that smooth self-occlusions, and contours resembling their behaviour, can play an important role in stabilising perception. These demonstrations provide important tests for critical contours as they are local image features (e.g., generally occurring around an object's silhouette in the case of a smooth self-occlusion) that have compelling effects on the global perception of an image. In many of these cases, most of the image structure is identical and critical contours may fail to capture perceptual changes. Frontally viewed bumpy planes are poorly resolved by the visual system with shape and material being difficult to discern, and the image appearing defocussed (Mooney et al., 2019). Images appear more defined as the surface is rotated in slant likely due to the emergence of occluding contours. In Figure 48, images of bumpy planes rendered using a Lambertian BRDF (left column) along with their critical contour structure in isolation (middle column) and superimposed on the shaded image (right column). The top row shows the bumpy plane viewed frontally and the bottom row shows a slanted view. The illumination direction relative to the plane is constant. The frontally viewed plane appears defocussed and this is reduced as the plane is viewed with increased slant. The critical contour structure seems to capture this change in perception with the contours appearing shorter in length and sparser in the top row. Since diffuse shading is viewpoint invariant, all variation between the images is due to a combination of foreshortening and parts of the surface coming in and out of occlusion. However, critical contours fail to account for Mooney et al.'s (2019) demonstration using level cuts. This manipulation involves the deletion of image structure with the remaining structure being identical to the original image and is a phenomenon that cannot be explained using critical contours alone. This suggests an important role of sharp contours approximating the intensity-orientation covariation along smooth self-occlusions, provided by level cuts.

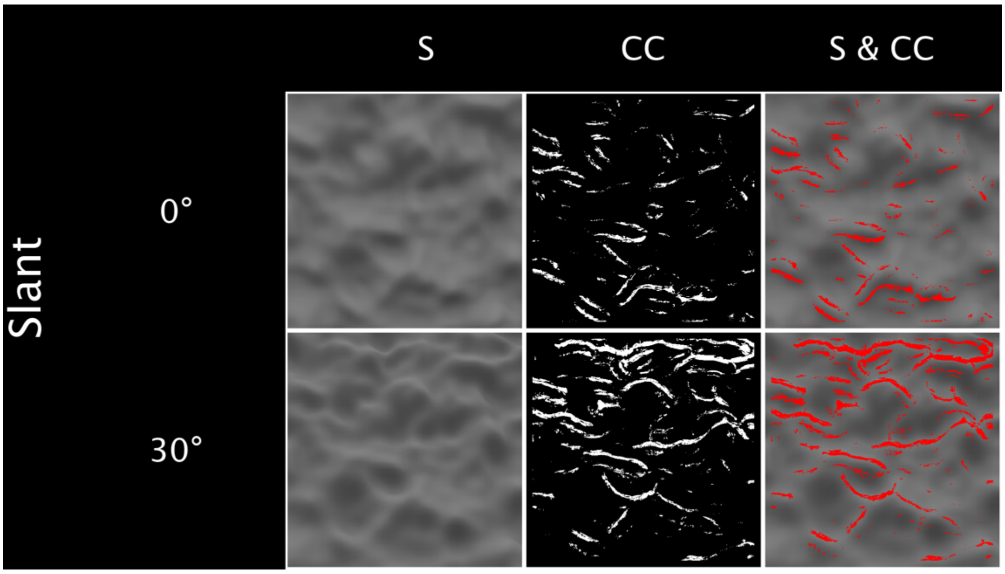


Figure 48. Frontally (top row) and obliquely (bottom row) viewed bumpy planes. Due to the lack of smooth self-occlusions, frontally viewed bumpy planes (top left) may appear defocussed and poorly resolved. Viewing the plane obliquely (bottom left) increases the presence of regions approaching smooth self-occlusions, promoting perception of a sharp image. The middle and right columns show critical contours in isolation and superimposed on the shaded images, respectively. The increased presence of critical contours in the obliquely-viewed plane appears to explain the clearer percept in this condition.

Another issue to reconcile between studies on self-occluding contours and critical contours is that they argue an opposite relationship between shape perception and illumination direction. The fundamental argument for the validity of critical contours as an explanation of human perception of shape from shading is the close correspondence between extremal contours of slant (a property of shape) with critical contours (a property of the image). Over a wide range of rendering functions, the MS complex generated from the image of the shading function (i.e., the critical contours) closely approximate the MS complex generated from the image of the slant function (i.e., the extremal contours). Under frontal illumination, the shading function is almost identical to the slant function, with some deviation introduced by shadowing, vignetting, or variation in illuminance energy in natural environments. This almost perfect correspondence between the shading and slant functions may predict optimal performance under these conditions. However, another criterion of critical contours is high contrast of shading gradients, and frontal illumination typically produces images of the lowest contrast. This raises the question of how critical contour criteria (contrast and symmetry of shading flow) should be weighted. Several psychophysical demonstrations suggest that perception is typically poorer for frontally-illuminated scenes. The highly-informative intensity-orientation covariation along smooth self-occlusions is a reliable cue that is present under all illumination orientations, with the exception of frontal and rear illumination (Marlow et al., 2019). In most situations, image intensities and surface geometry will have the same qualitative shape (Todd et al., 2023). Along a given cross section, plots of image intensity and surface geometry contain the same number of convexities and concavities, although their positions may be shifted relative to one another due to image intensities on the surface being skewed toward the illumination direction (e.g., a sphere illuminated from the left will appear lighter on the left). For example, the shading produced by a series of bumps may be light at the convex regions and dark in the concave regions, and the image intensities will follow this pattern for each reversal of surface concavity. However, frontal illumination can cause a doubling in frequency of image intensities as intensity maxima will now appear in both the peaks and troughs of the surface (Pentland, 1988).

Critical contours and related image analyses contain several free parameters that are typically set post hoc to generate structures that resemble human perception. Although this thesis and other research (Kunsberg & Zucker, 2018, 2021) demonstrate that these structures effectively capture human perception of shaded surfaces, the present lack of a principled approach to determine values for the algorithm's parameters poses a significant issue. Tuning the algorithm to conform to a pre-existing solution (i.e., our perception of the surface) fails to explain how the visual system arrives at this solution in the first place. For critical contours to provide a valid model of the visual system's segmentation of shading flow, it is necessary to investigate how features of the image such as contrast and spatial frequency may be used to set 'appropriate' values for these parameters from the outset. An adaptive visual system requires a means by which it can determine which variations of image intensity are attributable to changing surface orientation, or other causes such as variation in local illuminance.

3.3. Conclusions

The experiments in this thesis have identified some of the generative conditions predictive of failures of shape constancy and described the nature of theses failures. We demonstrated that these failures are predicted by salient image features that approximate smooth self-occlusions contained within the surface silhouette. The contrast of shading gradients produced by Lambertian materials is influenced by the relative orientation of surface geometry and illumination direction. Surface slant variation parallel to the illumination direction are maximally represented in the image and the contrast of shading gradients is minimised for slant variation in the perpendicular direction. For surface features in which slant variation along different direction varies, the relative alignment of the illuminant can influence the qualitative nature of shape perception. This is particularly the case for anisotropic shape features such as cylindrical ridges that vary dramatically in the nature of their slant variation in different directions (i.e., flat in one direction and highly curved in the perpendicular direction). Aligning the illumination with a ridge's axis of elongation results in the near-deletion of that image feature, which can influence perception of adjacent surface geometry. The failures of shape constancy presented in this thesis were largely robust to variation in the qualitative characteristics of the light field and for glossy materials, although the size of the effect was reduced when specular reflections are present. We conclude that variation in illumination direction affects shape perception of Lambertian and glossy materials in a complex piecewise manner. This opens directions for further research as variation in the perception of local shape features can in turn influence judgements of adjacent geometry. This thesis showed that deletion of a local shape feature results in the illusory segmentation of adjacent geometry, however beyond deletion, geometry-illumination alignment may influence judgements of a surface patch's qualitative shape index, be it convex, concave, or saddle shaped. How this affects the visual system's extraction of coherent surface geometry remains to be examined.

References

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the "light-from-above" prior. *Nature Neuroscience*, *7*(10), 1057–1058. https://doi.org/10.1038/nn1312

Anderson, B. L. (2015). Can Computational Goals Inform Theories of Vision? *Topics in Cognitive Science*, *7*(2), 274–286. https://doi.org/10.1111/tops.12136

Belhumeur, P. N., Kriegman, D. J., & Yuille, A. L. (1997). The bas-relief ambiguity. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1060–1066. https://doi.org/10.1109/CVPR.1997.609461

Breton, P., & Zucker, S. W. (1996). Shadows and shading flow fields. *Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 782–789.

Brooks, M. J., & Horn, B. K. P. (1985). Shape and source from shading. *International Joint Conference on Artificial Intelligence*, 932–936.

Christou, C. G., & Koenderink, J. J. (1997). Light Source Dependence in Shape from Shading. *Vision Research*, *37*(11), 1441–1449. https://doi.org/10.1016/S0042-6989(96)00282-9

David, C., & Zucker, S. W. (1990). Potentials, valleys, and dynamic global coverings. *International Journal of Computer Vision*, *5*(3), 219–238.

Debevec, P. (1998). Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-based Graphics with Global Illumination and High Dynamic Range Photography. *Conference Proceedings*, 10.

Debevec, P. E., & Malik, J. (1997). Recovering high dynamic range radiance maps from photographs. *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, 369–378.

DeCarlo, D., Finkelstein, A., Rusinkiewicz, S., & Santella, A. (2003). *Suggestive Contours for Conveying Shape*. 8.

Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, *3*(5), 3. https://doi.org/10.1167/3.5.3

Fleming, R. W., Torralba, A., & Adelson, E. H. (2004). Specular reflections and the perception of shape. *Journal of Vision*, *4*(9), 10. https://doi.org/10.1167/4.9.10

Freeman, W. T. (1994). The generic viewpoint assumption in a framework for visual perception. *Nature*, *368*(6471), 542–545. https://doi.org/10.1038/368542a0

Freeman, W. T., & Viola, P. A. (1997). Bayesian Model of Surface Perception. *Proceedings of the 10th International Conference on Neural Information Processing Systems*, 787–793.

Fritsche, M., Spaak, E., & De Lange, F. P. (2020). A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. *eLife*, *9*, e55389. https://doi.org/10.7554/eLife.55389

Geisler, W. S., & Diehl, R. L. (2002). Bayesian natural selection and the evolution of perceptual systems. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *357*(1420), 419–448. https://doi.org/10.1098/rstb.2001.1055

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Gomez, O., & Neumann, H. (2016). Biologically Inspired Model for Inference of 3D Shape from Texture. *PLOS ONE*, *11*(9), e0160868. https://doi.org/10.1371/journal.pone.0160868

Granrud, C. E., Yonas, A., & Opland, E. A. (1985). Infants’ sensitivity to the depth cue of shading. *Perception & Psychophysics*, *37*(5), 415–419. https://doi.org/10.3758/BF03202872

Hanazawa, A., & Komatsu, H. (2001). Influence of the Direction of Elemental Luminance Gradients on the Responses of V4 Cells to Textured Surfaces. *The Journal of Neuroscience*, *21*(12), 4490–4497. https://doi.org/10.1523/JNEUROSCI.21-12-04490.2001

Horn, B. K. P. (1970). *Shape From Shading: A Method for Obtaining the Shape of a Smooth Opaque Object From One View*.

Hunter, F., Biver, S., Fuqua, P., & Hunter-Reid, R. (2021). *Light, science and magic: An introduction to photographic lighting* (Sixth edition). Routledge.

Ikeuchi, K., & Horn, B. K. P. (1981). Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, *17*(1–3), 141–184. https://doi.org/10.1016/0004-3702(81)90023-0

Judd, T., Durand, F., & Adelson, E. (2007). Apparent ridges for line drawing. *ACM Transactions on Graphics*, *26*(3), 19. https://doi.org/10.1145/1276377.1276401

Kartashova, T., Pas, S. F. T., Ridder, H. D., & Pont, S. C. (2019). Light Shapes: Perception-Based Visualizations of the Global Light Transport. *ACM Transactions on Applied Perception*, *16*(1), 1–17. https://doi.org/10.1145/3232851

Kim, J., Marlow, P. J., & Anderson, B. L. (2012). The dark side of gloss. *Nature Neuroscience*, *15*(11), 1590–1595. https://doi.org/10.1038/nn.3221

Koenderink, J. J., Doorn, A. J. V., & Pont, S. C. (2007). Perception of illuminance flow in the case of anisotropic rough surfaces. *Perception & Psychophysics*, *69*(6), 895–903.

Koenderink, J. J., Pont, S. C., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (2007). The Visual Light Field. *Perception*, *36*(11), 1595–1610. https://doi.org/10.1068/p5672

Koenderink, J. J., & van Doorn, A. J. (1980). Photometric Invariants Related to Solid Shape. *Optica Acta: International Journal of Optics*, *27*(7), 981–996. https://doi.org/10.1080/713820338

Koenderink, J. J., & van Doorn, A. J. (1992). Surface shape and curvature scales. *Image and Vision Computing*, *10*(8), 557–564. https://doi.org/10.1016/0262-8856(92)90076-F

Koenderink, J. J., Van Doorn, A. J., & Pont, S. C. (2004). Light Direction from Shad(ow)ed Random Gaussian Surfaces. *Perception*, *33*(12), 1405–1420. https://doi.org/10.1068/p5287

Koenderink, J. J., van Doorn, A. J., & Wagemans, J. (2014). Local Shape of Pictorial Relief. *I-Perception*, *5*(3), 188–204. https://doi.org/10.1068/i0659

Koenderink, J. J., van Doorn, A. J., & Wagemans, J. (2015). Local Solid Shape. *I-Perception*, *6*(5), 204166951560406. https://doi.org/10.1177/2041669515604063

Koenderink, J. J., Van Doorn, A., & Pont, S. (2012). Shading, a View from the Inside. *Seeing and Perceiving*, *25*(3–4), 303–338. https://doi.org/10.1163/187847511X590923

Koenderink, J., Van Doorn, A., & Wagemans, J. (2012). Picasso in the mind's eye of the beholder: Three-dimensional filling-in of ambiguous line drawings. *Cognition*, *125*(3), 394–412. https://doi.org/10.1016/j.cognition.2012.07.019

Kunsberg, B., Holtmann-Rice, D., Alexander, E., Cholewiak, S., Fleming, R., & Zucker, S. W. (2018). Colour, contours, shading and shape: Flow interactions reveal anchor neighbourhoods. *Interface Focus*, *8*(4), 20180019. https://doi.org/10.1098/rsfs.2018.0019

Kunsberg, B., & Zucker, S. W. (2018). Critical Contours: An Invariant Linking Image Flow with Salient Surface Organization. *SIAM Journal on Imaging Sciences*, *11*(3), 1849–1877. https://doi.org/10.1137/17M1145525

Kunsberg, B., & Zucker, S. W. (2021). From boundaries to bumps: When closed (extremal) contours are critical. *Journal of Vision*, *21*(13), 7. https://doi.org/10.1167/jov.21.13.7

Lambert, J. H. (1760). *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. E. Klett C.P. Detleffsen.

Langer, M. S., & Bülthoff, H. H. (2001). A Prior for Global Convexity in Local Shape-from-Shading. *Perception*, *30*(4), 403–410. https://doi.org/10.1068/p3178

Langer, M. S., & Zucker, S. W. (1994). Shape-from-shading on a cloudy day. *Journal of the Optical Society of America A*, *11*(2), 467. https://doi.org/10.1364/JOSAA.11.000467

Liu, B., & Todd, J. T. (2004). Perceptual biases in the interpretation of 3D shape from shading. *Vision Research*, *44*(18), 2135–2145. https://doi.org/10.1016/j.visres.2004.03.024

Luccio, R. (2019). Limits of the Application of Bayesian Modeling to Perception. *Perception*, *48*(10), 901–917. https://doi.org/10.1177/0301006619868125

Lythgoe, J. N., & Partridge, J. C. (1989). Visual pigments and the acquisition of visual information. *Journal of Experimental Biology*, *146*(1), 1–20. https://doi.org/10.1242/jeb.146.1.1

Mamassian, P., & Kersten, D. (1996). Illumination, Shading and the Perception of Local Orientation. *Vision Research*, *36*(15), 2351–2367. https://doi.org/10.1016/0042-6989(95)00286-3

Marlow, P. J., & Anderson, B. L. (2021). The cospecification of the shape and material properties of light permeable materials. *Proceedings of the National Academy of Sciences*, *118*(14), e2024798118. https://doi.org/10.1073/pnas.2024798118

Marlow, P. J., Kim, J., & Anderson, B. L. (2012). The Perception and Misperception of Specular Surface Reflectance. *Current Biology*, *22*(20), 1909–1913. https://doi.org/10.1016/j.cub.2012.08.009

Marlow, P. J., Mooney, S. W. J., & Anderson, B. L. (2019). Photogeometric Cues to Perceived Surface Shading. *Current Biology*, *29*(2), 306-311.e3. https://doi.org/10.1016/j.cub.2018.11.041

Marlow, P. J., Todorović, D., & Anderson, B. L. (2015). Coupled computations of three-dimensional shape and material. *Current Biology*, *25*(6), R221–R222. https://doi.org/10.1016/j.cub.2015.01.062

Mooney, S. W. J., & Anderson, B. L. (2014). Specular Image Structure Modulates the Perception of Three-Dimensional Shape. *Current Biology*, *24*(22), 2737–2742. https://doi.org/10.1016/j.cub.2014.09.074

Mooney, S. W., Marlow, P. J., & Anderson, B. L. (2019). The perception and misperception of optical defocus, shading, and shape. *eLife*, *8*, e48214. https://doi.org/10.7554/eLife.48214

Morgenstern, Y., Geisler, W. S., & Murray, R. F. (2014). Human vision is attuned to the diffuseness of natural light. *Journal of Vision*, *14*(9), 15. https://doi.org/10.1167/14.9.15

Mury, A. A., Pont, S. C., & Koenderink, J. J. (2007a). Light field constancy within natural scenes. *Applied Optics*, *46*(29), 7308. https://doi.org/10.1364/AO.46.007308

Mury, A. A., Pont, S. C., & Koenderink, J. J. (2007b). Spatial properties of light fields in natural scenes. *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization - APGV '07*, 140. https://doi.org/10.1145/1272582.1272617

Mury, A. A., Pont, S. C., & Koenderink, J. J. (2009). Structure of light fields in natural scenes. *Applied Optics*, *48*(28), 5386. https://doi.org/10.1364/AO.48.005386

Nefs, H. T., Koenderink, J. J., & Kappers, A. M. L. (2005). The Influence of Illumination Direction on the Pictorial Reliefs of Lambertian Surfaces. *Perception*, *34*(3), 275–287. https://doi.org/10.1068/p5179

Nefs, H. T., Koenderink, J. J., & Kappers, A. M. L. (2006). Shape-from-shading for matte and glossy objects. *Acta Psychologica*, *121*(3), 297–316. https://doi.org/10.1016/j.actpsy.2005.08.001

Norman, J. F., Todd, J. T., & Orban, G. A. (2004). Perception of Three-Dimensional Shape From Specular Highlights, Deformations of Shading, and Other Types of Visual Information. *Psychological Science*, *15*(8), 565–570. https://doi.org/10.1111/j.0956-7976.2004.00720.x

Norman, J. F., Todd, J. T., & Phillips, F. (2020). Effects of illumination on the categorization of shiny materials. *Journal of Vision*, *20*(5), 2. https://doi.org/10.1167/jov.20.5.2

Pentland, A. P. (1984). Local Shading Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*(2), 170–187. https://doi.org/10.1109/TPAMI.1984.4767501

Pentland, A. P. (1988). Shape information from shading: A theory about human perception. *Second International Conference on Computer Vision*.

Ramachandran, V. S. (1988). *Perception of shape from shading*. 4.

Ramamoorthi, R., & Hanrahan, P. (2001). On the relationship between radiance and irradiance: Determining the illumination from images of a convex Lambertian object. *Journal of the Optical Society of America A*, *18*(10), 2448. https://doi.org/10.1364/JOSAA.18.002448

Regan, B. C., Julliot, C., Simmen, B., Viénot, F., Charles–Dominique, P., & Mollon, J. D. (2001). Fruits, foliage and the evolution of primate colour vision. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1407), 229–283. https://doi.org/10.1098/rstb.2000.0773

Roman, S., Viacheslav, V., Nikolay, G., Nadykto, A., Pižurica, A., & Zelensky, A. (2020). Automated visual inspection algorithm for the reflection detection and removing in image sequences. In W. Osten & D. P. Nikolaev (Eds.), *Twelfth International Conference on Machine Vision (ICMV 2019)* (p. 80). SPIE. https://doi.org/10.1117/12.2559362

Savarese, S., Fei-Fei, L., & Perona, P. (2004). What do reflections tell us about the shape of a mirror? *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization - APGV '04*, 115. https://doi.org/10.1145/1012551.1012571

Schofield, A. J., Rock, P. B., & Georgeson, M. A. (2011). Sun and sky: Does human vision assume a mixture of point and diffuse illumination when interpreting shape-from-shading? *Vision Research*, *51*(21–22), 2317–2330. https://doi.org/10.1016/j.visres.2011.09.004

Sun, J., & Perona, P. (1998). Where is the sun? *Nature Neuroscience*, *1*(3), 183–184.

Todd, J. T., Koenderink, J. J., van Doom, A. J., & Kappers, A. M. L. (1996). Effects of Changing Viewing Conditions on the Perceived Structure of Smoothly Curved Surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(3), 695–706.

Todd, J. T., & Mingolla, E. (1983). Perception of Surface Curvature and Direction of Illumination From Patterns of Shading. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(4), 583–595.

Todd, J. T., Oomes, A. H. J., Koenderink, J. J., & Kappers, A. M. L. (2001). On the Affine Structure of Perceptual Space. *Psychological Science*, *12*(3), 191–196. https://doi.org/10.1111/1467-9280.00335

Todd, J. T., & Reichel, F. D. (1989). Ordinal Structure in the Visual Perception and Cognition of Smoothly Curved Surfaces. *Psychological Review*, *96*(4), 643–657.

Todd, J. T., Yu, Y., & Phillips, F. (2023). Qualitative perception of 3D shape from patterns of luminance curvature. *Journal of Vision*, *23*(5), 10. https://doi.org/10.1167/jov.23.5.10

Toscani, M., Valsecchi, M., & Gegenfurtner, K. R. (2017). Lightness perception for matte and glossy complex shapes. *Vision Research*, *131*, 82–95. https://doi.org/10.1016/j.visres.2016.12.004

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604. https://doi.org/10.1038/nn0602-858

Xia, L., Pont, S. C., & Heynderick, I. (2017). Separate and Simultaneous Adjustment of Light Qualities in a Real Scene. *I-Perception*, *8*(1), 204166951668608. https://doi.org/10.1177/2041669516686089

Xia, L., Pont, S. C., & Heynderickx, I. (2014). The Visual Light Field in Real Scenes. *I-Perception*, *5*(7), 613–629. https://doi.org/10.1068/i0654

Yang, J., Otsuka, Y., Kanazawa, S., Yamaguchi, M. K., & Motoyoshi, I. (2011). Perception of Surface Glossiness by Infants Aged 5 to 8 Months. *Perception*, *40*(12), 1491–1502. https://doi.org/10.1068/p6893

Zhang, F., De Ridder, H., Barla, P., & Pont, S. (2019). A systematic approach to testing and predicting light-material interactions. *Journal of Vision*, *19*(4), 11. https://doi.org/10.1167/19.4.11

Zhang, F., De Ridder, H., Barla, P., & Pont, S. (2020). Effects of light map orientation and shape on the visual perception of canonical materials. *Journal of Vision*, *20*(4), 13. https://doi.org/10.1167/jov.20.4.13

Zhang, F., De Ridder, H., & Pont, S. (2015). *The influence of lighting on visual perception of material qualities* (B. E. Rogowitz, T. N. Pappas, & H. De Ridder, Eds.; p. 93940Q). https://doi.org/10.1117/12.2085021

Zhang, F., De Ridder, H., & Pont, S. C. (2018). Asymmetric perceptual confounds between canonical lightings and materials. *Journal of Vision*, *18*(11), 11. https://doi.org/10.1167/18.11.11

Zhang, R., Tsai, P., Cryer, J. E., & Shah, M. (1999). Shape-from-shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(8), 690–706. https://doi.org/10.1109/34.784284

Zokay, M., & Saylani, H. (2022). Removing Specular Reflection in Multispectral Dermatological Images Using Blind Source Separation. In G. Yang, A. Aviles-Rivero, M. Roberts, & C.-B. Schönlieb (Eds.), *Medical Image Understanding and Analysis* (Vol. 13413, pp. 734–750). Springer International Publishing. https://doi.org/10.1007/978-3-031-12053-4_54

Zucker, S. W. (1985). Early orientation selection: Tangent fields and the dimensionality of their support. *Computer Vision, Graphics, and Image Processing*, *32*, 74–103.

Zucker, S. W., Dobbins, A., & Iverson, L. (1989). Two stages of curve detection suggest two styles of visual computation. *Neural Computation*, *1*, 68–81.