Modelling the human perception of shape-from-shading

by

PENG SUN

A thesis submitted to
The University of Birmingham
For the degree of
DOCTOR OF PHILOSOPHY

School of Psychology
The University of Birmingham
11/06/2010

Abstract

Shading conveys information on 3-D shape and the process of recovering this information is called shape-from-shading (SFS). This thesis divides the process of human SFS into two functional sub-units (luminance disambiguation and shape computation) and studies them individually. Based on results of a series of psychophysical experiments it is proposed that the interaction between first- and second-order channels plays an important role in disambiguating luminance. Based on this idea, two versions of a biologically plausible model are developed to explain the human performances observed here and elsewhere. An algorithm sharing the same idea is also developed as a solution to the problem of intrinsic image decomposition in the field of image processing.

With regard to the shape computation unit, a link between luminance variations and estimated surface norms is identified by testing participants on simple gratings with several different luminance profiles. This methodology is unconventional but can be justified in the light of past studies of human SFS. Finally a computational algorithm for SFS containing two distinct operating modes is proposed. This algorithm is broadly consistent with the known psychophysics on human SFS.

Dedication

To God and to my beloved parents

Acknowledgement

I would like to thank Dr. Andrew Schofield, my academic supervisor, a great motivator and also a compassionate personal tutor, for all the generous help, guidance, encouragement and all that I have learnt throughout my PhD research.

I would like to thank my parents. Without their support, this extraordinary journey would not be even imaginable. Above all that, I would give my whole hearted gratefulness to God, for being with me always and for having made me strong and courageous in front of all the hardships that I have encountered.

Table of Contents

1. Introduction

Both artificial and biological vision systems take as input 2-D arrays of light intensities transformed from the 3-D world according to the laws of optics. Interpreting these 2-D signals in terms of 3-D structures is an ill-posed, inverse problem but is nevertheless a crucial step in any visual processing algorithm. The ability of the human observer to see the world and understand it is so remarkable that no present machine vision algorithms are comparable in terms of versatility, robustness and accuracy.

Since the 1970's, great efforts have been made towards describing the human visual system as an information processing system which can reconstruct a 3-D representation of the world based on its corresponding 2-D projection onto the retina. Probably the most influential in this regard is Marr's theory of computational vision (Marr, 1982). In his theory, Marr proposes that visual information is represented at different levels. Between the level of 2-D image based representation (primal sketch) and the level of 3-D object-based representation there lies a transition layer called the 2.5-D sketch. This layer functions as a buffer to store information about the depth and orientation of local surface patches. The 2.5D sketch is the assembled output of many sub-modules each operating on separate sources such as shading, stereo, motion, texture and perceived contours. Underlying the 2.5D sketch is the idea that individual computational problems become solvable given constraints, and that they can be carried out more or less independent of each other (Marr, 1982, p103; Landy, Maloney, Johnston & Young, 1995; Bruce, Green & Georgeson, 1996, p137; Palmer, 1999, p200).

The aim of this thesis is to investigate one particular (putative) sub-module within the 2.5D sketch – shape-from-shading. This module is responsible for computing surface depth and orientation in space from the pictorial depth cue, monocular shading. In the language of Marr's theory, the thesis attempts to establish the following: what comprises the input, how information is represented, what constraints are needed in order to make the inverse problem solvable and finally what computations are carried out in each step to obtain the observed experimental output.

1.1 Background

1.1.1 3D vision

The optical signal that is received by the retina is inherently 2-D. During the projecting process, information in the 'distance' dimension is lost. But the fact that humans actually see a 3-D world rather than a 2-D image implies that one of the primary functions of our visual system is to reconstruct a 3-D space from the 2-D retinal image. But how this function is achieved had looked intractable until the emergence of Marr's computational theory of vision in the early 1980's (Marr 1982).

Unifying discoveries from neurophysiology, psychophysics and computer vision, Marr proposed that the visual system can be characterised in terms of an information processing system. At the early stage of visual processing, the system generates a representation of the input image, describing its important 2-D features (the primal sketch; see Figure 1.1). Information then progresses from the primal sketch to the 2.5D sketch which contains information about depth and surface orientation. The 2.5D sketch is a view centred representation of the 3-D world. At the next stage, this view centred representation is transformed to an object centred 3-D representation

which is invariant to the viewing direction. The introduction of the 2.5D sketch is a major contribution of Marr's theory, being a key step in getting from the 2D primal sketch to the 3D object centred view.

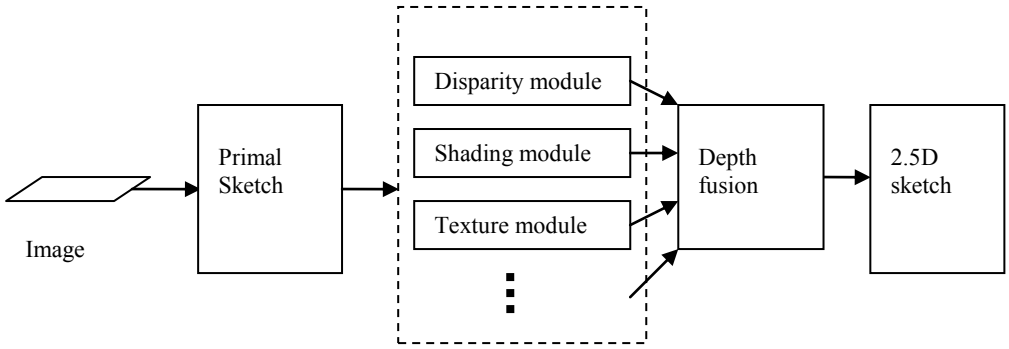


Figure 1.1 Modular description of the human visual system up to the level of 2.5D sketch, proposed by Marr (1982).

Marr assumed that surface orientation and distance in space are the essential building blocks to the final 3-D perception. However, some have questioned how vital surface orientation and distance in space are to 3-D perception. Pizlo (2008) pointed out that Marr's theory would fail to account for phenomena such as constancy of perceived 3-D shape. He proposed a computational theory of 3-D shape recovery which took a very different approach. Pizlo's model does not make use of surface orientations or depth information. Instead it is based on a few prior constraints such as symmetry and volume, i.e. most objects in the world are somewhat symmetrical and enclose a volume. Pizlo and his colleagues (Li, Pizlo & Steinman, 2009; Pizlo, Sawada, Li, Kropatsch & Steinman, 2010) then suggested that the human visual system could rely more on these priors to achieve a coherent visual representation of the 3-D world than on surface orientation and depth perception as suggested by Marr.

The conflict discussed above seems to be due to the different objectives of the two theories. While Pizlo's theory is mostly concerned with recovering 3-D shape of

concrete objects, Marr's proposed vision system has the more general purpose aim of solving a wide range of visual problems rather than just understanding and interpreting 3-D shapes. After all, Pizlo cannot deny the observation that humans can derive surface orientations and depth even when no solid shape is presented. For example, it has been shown that human observers can perceive slanted surfaces from random dot disparity stimuli which do not signal the shape of any meaningful 3-D objects (Julesz, 1960). Whether surface orientations and depth are the primary ingredients of 3-D vision may be uncertain but it is hard to believe that such information is not used at all during the process of 3-D reconstruction.

1.1.2 Depth perception

Although information in the depth (z) dimension is lost during the process of projection, the 2-D retinal image still contains "regularities" that reflect relative differences in distance between two points in a scene. These visual "regularities" are called depth cues. Known static depths cues include stereoscopic disparity, deformation of contours, texture gradients, and shading. Earlier studies have shown that human observers can make effective use of these cues to infer depth. For example, Julesz (1960) reported that disparity alone can generate a strong depth perception with very little involvement of other visual information. Human observers have also been found capable of perceiving depth from texture gradients (Gibson, 1950). It is worth emphasizing that the stimuli used in such studies contained the single cue of interest only. Despite this human observers were still able to obtain a depth percept from either disparity or texture gradient alone.

Motivated by such discoveries, Marr (1982) proposed that there exist a number of independent computational modules each operating on a particular depth cue. Each

module can be described as an information processing system which has tailored algorithms built in to work on the specific type of input signal. The outcome of each module is a point-wise depth map that contributes information to the generalized 2.5D sketch.

Marr's modular description of the early visual system is illustrated in Figure 1.1. By characterising the visual system as a modular system, one can divide it into many separate modules and study each independently. This so-called modular principle (Marr, 1982) is only a gross simplification of the complex system and does not prevent any possible interactions at a later stage where computed results from all modules are fused according to a certain scheme. Empirically, humans tend to be better at perceiving natural scenes containing multiple visual cues than experimental stimuli made of simple single cues, indicating the general plausibility of such a cue combination scheme.

Studies of cue combination have shown that human observers indeed choose to combine depth cues via a number of structured routines (Hills, Watt, Landy & Banks, 2004; Oruc, Maloney & Landy, 2003; Landy et al., 1995; Curran & Johnston, 1994; Bulthoff & Mallot, 1988). Moreover, it has been shown that modules are connected even before each computation is carried out such that the whole system can be described by a multivariate system with interactions existing between variables (Pankanti & Jain, 1995). For instance, Vuong, Domini and Caudek (2006) proposed that human observers use shading information to constrain the disparity module to arrive at a more precise estimation of depth. This cooperative relationship between modules is not surprising since each depth cue has its own "domain of expertise" in

depth computation. For example, stereoposis reveals relative depth directly but shading contributes more to the surface orientation and curvature. Thus modules can complement each other in the sense that one module could provide vital constraints for the computation carried out in another module. Therefore, building a complete integrated vision system requires not only the knowledge of any individual module but also the necessary / likely exchanges of information between modules. The latter demands an in-depth investigation of each module including its computational theory, the constraints required to complete each computation and any assumptions that are adopted should the necessary constraints be missing. This thesis tackles one of the depth modules: shape from monocular shading.

1.1.3 Shape from shading (SFS)

The definition of shading sometimes can be confusing. In most works shading is defined as the variations in the amount of reflected light as a direct result of variations in the orientation of the surface relative to a light source (for example see Palmer, 1999, p243). To see this process intuitively, imagine a curved surface lit by a single point light source (Figure 1.2). The parts of the surface facing towards the light source will appear brighter than the parts facing away from it. Thus surface orientation clearly plays a very important role in determining the surface brightness.

However, a broader definition of shading can also be found which refers to shading as variations in the amount of reflected light due to any source other than the reflectance properties of the surface material (Olmos & Kingdom 2004). Cast shadows and luminance variations caused by inter-reflections between surfaces are included in this definition of shading. To distinguish these two definitions, the former definition is often called “local shading” while the latter is termed “global shading” (Forsyth &

Ponce, 2003, p70). In this thesis, unless explicitly emphasised, the term shading referrers to "local shading", that is variations in light intensity that directly result from undulations of the surface in question.

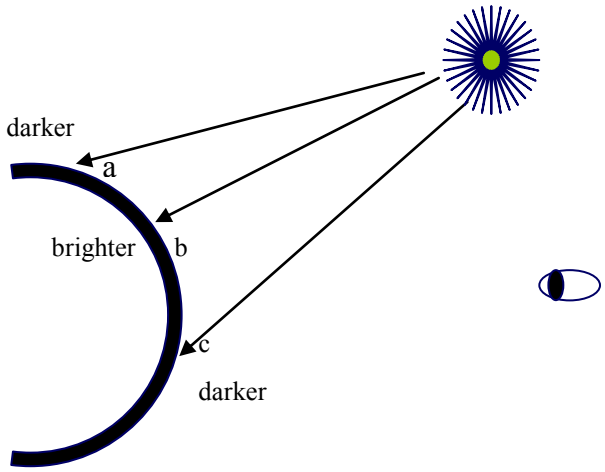


Figure 1.2 Surface brightness is dependent on surface orientation. Surface patches facing towards the light source (point b) will receive more irradiance thus look brighter than those facing away (point a and c).

Although the study of shape-from-shading has a long tradition (see for example, Rittenhouse, 1786; Brewster 1826), it was not until the 1970's that the computational analysis of shading was first proposed to quantitatively study the relation between shading and surface orientation and to apply it in computer vision. Horn published a series of papers (1975; 1977; Ikeuchi & Horn, 1981) leading to the formulation of the problem which he termed shape-from-shading (SFS). Strictly speaking, "surface orientation from shading" may better characterise the problem than "shape-from-shading" but the term SFS is used throughout this thesis in line with convention.

Like most other problems in computational vision, SFS is an ill-posed problem that requires the application of many constraints in order to make it mathematically well-posed. A typical way to solve SFS normally requires constraints on surface material, lighting direction, diffuseness of the lighting and so on. In early works on SFS constraints were often adopted to suit conditions that were not necessarily common in

daily life because early attempts to solve SFS were almost all dedicated to solving problems in specific areas such as aerospace, satellite surveillance and remote sensing. Some frequently imposed constraints include uniform surface material and uniform surface reflectance, Lambertian reflectance (Horn, 1975; Pentland, 1984; Pentland, 1988), distant light sources, and distant viewing positions (such that orthographic projection applies; Horn, 1975; Horn, 1977; Horn & Sjoberg, 1979; Ikeuchi & Horn, 1981). With the problem sufficiently constrained, SFS reduced to the solution of a relatively simple series of differential equations.

Since the mid of 80's, the subject of SFS has split into two different but related sub-fields. One stream continued to try to solve practical problems encountered in computer vision and thus focused on developing more powerful ways to solve the differential equations. For example, efforts have been made to ensure the robustness, existence and uniqueness of solutions by using methods such as numerical iteration, variational approaches, regularization and optimization (Durou, Falcone & Sagona, 2008). The other stream took a very different objective— to study SFS in the human visual system. That is, to study the processes by which human observers deduce surface orientation based on shading. The remainder of this chapter (indeed thesis) addresses SFS in the human visual system.

1.1.4 Perception of shape from shading

Our ability to perceive depth from luminance variations can be illustrated by the very simple stimulus in Figure 1.3a where a linear luminance ramp is bounded by a circular outline. This linear ramp will appear as a convex bump raised from the background – and is evidence of SFS operating in the visual system. Moreover, the process of SFS is quite fast such that it could happen at an early stage of visual

processing (Sun & Perona, 1996). This property of SFS was observed in a visual search tasks. If the luminance ramp in Fig 1.3a is presented up-side-down alongside several copies of the original, the up-side-down version tends to stand out as a concave dent with convex bumps forming the background (Fig 1.3b). The time needed to spot on the odd-one-out can be as fast as only a few hundred milliseconds and does not increase with the number of display items (Kleffner & Ramachandran, 1992); suggestive of pre-attentive, parallel search and the existence of a feature map for shape in early vision (Triesman & Gelade, 1980). On the other hand, when the 3-D impression of convexity v.s. concavity is not strong (e.g. opposite horizontally oriented luminance gradients), reaction times tends to be much longer and also increase drastically with display size (suggestive of attentive, serial search; Triesman & Gelade, 1980).

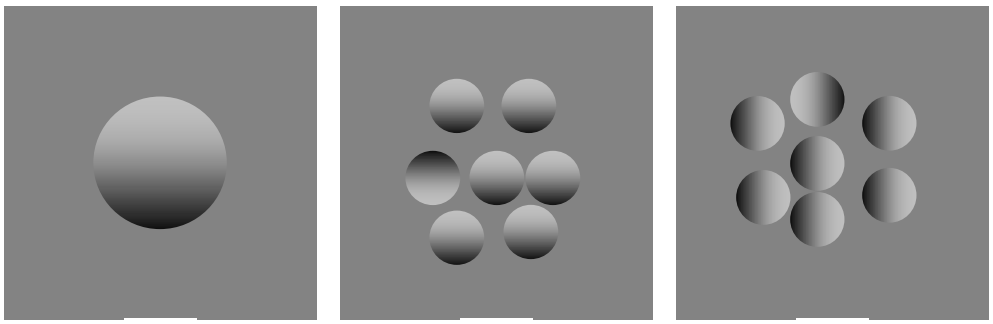


Figure 1.3 (a) An example of SFS in the human visual system. A luminance ramp bounded by a circular boundary will give rise to a perception of a bump raised from the grey background. (b) several linear ramps are placed together, one of which is vertically inverted. (c) horizontally oriented linear ramps. (After Kleffner and Ramachandran 1992).

1.1.4.1 Two early studies on the perception of SFS

Acknowledging the empirical evidence for human SFS, Todd and Mingolla (1983) were among the first to quantitatively examine the salience and the role of shading in the perception of 3-D depth. Benefiting from advances in computer graphics, they were able to use computer rendered realistic 3D surface to test human responses to

shading. More specifically, they tested how humans responded to the shading pattern of a cylindrical surface with a mixture of Lambertian (diffuse, matte) and mirror (glossy) reflectance lit by single distant light sources from a small number of distinct directions.

Participants were shown the physical object which the stimuli would depict and were then asked to rate how curved the surface in the stimuli appeared to be. They confirmed the human ability to understanding shading in terms of 3-D shape and showed that this ability was not subject to the same constraints (e.g. surface being Lambertian) as required by most machine vision algorithms at that time. In addition, surface curvatures tended to be underestimated for pure Lambertian reflectance but overestimated when a mirror reflectance was added. Observers also responded differently when the cylinder was lit by light from different directions. That is perceived surface shape is dependent on the light source and thus not veridical to the object being depicted.

In fear of having chosen an inappropriate measurement (curvedness) and a stimulus that was too simple, Mingolla and Todd (1986) used computer rendered ellipsoids as test stimuli and asked the participants to report the apparent slant and tilt at each of several measurement points on the simulated surface. Again, observers were able to interpret the shading pattern in a way that was coherent with the underlying 3D structure but settings also varied with the direction of the illumination. However, in this experiment, the glossiness of the surface did not have a noticeable impact on overall performance. Observers also responded differently as the eccentricity (deviation from spherecity) of the ellipsoid changed, with near spherical ellipsoids

being judged more accurately than ellipsoids with high eccentricity. For ellipsoids with high eccentricity, observer's judgements suggested that they could not even agree on overall perceived shape. Note, however, that the tasks used in these two early studies are both very abstract. In fact the researchers admitted that the participants found the tasks very difficult even after training had been provided.

These two early studies raise a number of interesting questions. First, human observers underestimated the curvature defined by Lambertian shading in the first study and showed large inter-observer variance when estimating the surface orientation in the second study. This questions the overall effectiveness and veridicallity of shading as a cue to the underlying 3D structure. Second, both studies reported that subjects responded differently under different illuminant directions, questioning the veridicallity of shape judgements, although it remains to be seen if there is a systematic relationship between illumination direction and perceived shape. Third, contradictory results appeared in terms of how subjects responded to surfaces with different reflectance properties (matte vs glossy). No safe conclusion can be drawn regarding whether or not changes in reflectance can alter shape perception in humans given the different tasks involved in these studies. Fourth, these studies raise questions as to the most appropriate stimuli for studying SFS. Simple stimuli may make the task too easy revealing little about the genuine characteristics of human SFS. But complex stimuli risk introducing potential confounding variables, making the result less valid. Finally, the shape estimates were different in the two studies. Todd and Mingolla (1983) measured curvedness (a second-order cue) whereas Mingolla and Todd (1986) measured surface orientation (a first-order shape cue). It is not clear which measurement is the better choice studying SFS.

Answering the above questions would give insight into the computation of SFS in the human brain. For example, the second and the third questions are closely related to how humans constrain the problem of SFS in terms of lighting and surface reflectance. The final question actually asks what comprises the immediate output of the computational module for SFS. The majority of studies on SFS following that of Mingolla and Todd tend to focus on a subset of these unresolved questions.

1.1.4.2 SFS is effective but shading cue is not

In a study of depth cue integration, Bülthoff and Mallot (1988) found that the depth percept generated from disparity vetoed shading when these two cues were put into conflict. That is, when shading suggested curvature while stereo edges suggested flatness, observers tended to base their perceptions on stereo cue only and ignore the effect of the shading cue. Thus shading appears to be carried less weight than disparity for deducing 3-D structure. Similar results have also been found by others that the effect of shading can be dominated by other cues such as edge contours and surface contours (Ramachandran, 1988; Knill, 1992). This down rating of the reliability of shading in the visual system may reflect some limitations of shading as a carrier of 3-D information in the physical world.

The computational analysis of shading reveals that it conveys only limited information on 3-D shape (Pentland, 1984). Assuming that all surfaces are approximately spherical, the sign of the principal curvatures cannot be determined by shading alone (Pentland, 1984). An immediate consequence of this limitation is that concavity, convexity, elliptic and hyperbolic shape can not be distinguished by shading alone (See Fig 1.4).

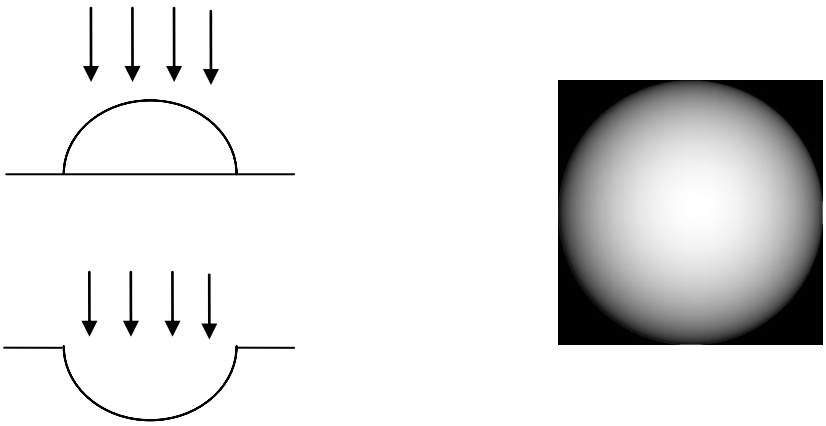


Figure 1.4 Example ambiguity in shaoe-from-shading. If a hemi-spherically convex surface with Lambertian reflectance is lit by directional light from above, it gives the same shading pattern as a hemi-spherically concave surface with Lambertian reflectance lit by the same type of light source.

In an experiment of local shape categorization (Erens, Kappers & Koenderink, 1993a), observers were very poor at differentiating elliptic shapes from hyperbolic shapes based on shading patterns when the occluding contours (outlines) of the shapes were obscured by random markings. Among the elliptic shapes, observers were unable to distinguish between convex and concave shapes. After adding information on illumination to the stimuli, observers were able to break the ambiguity between concavity and convexity but they were still unable to identify elliptic vs. hyperbolic shapes. These results clearly demonstrate one deficiency in SFS: humans can not differentiate between elliptic shapes and hyperbolic shapes when shading is the only available cue. However, this deficiency is due to the physical limitation of shading as a cue to depth. Put simply information about the sign of surface curvature is not contained in shading. Thus the deficiency cannot be blamed on any computational inability in the human visual system.

Mamassian, Kersten & Knill (1996) found a rather different result using a different experimental design to test observers' ability to categorize shapes based on shading.

In their experiment, observers viewed a computer generated croissant-shaped object rendered with Lambertian shading. The task was to label the points on the object surface according to whether they were in an elliptic or hyperbolic region. In this case, observers could separate the two regions very well and could even accurately locate where the parabolic curve (segregation line) was on the surface. However, a simple comparison between the two tasks reveals that the presence of object outlines as the cause of the discrepancy between the two results.

Given the observation that the shading can be down weighed in the presence of other cues, it is reasonable to question the capability of computing SFS in the visual system. In other words, does the brain allocate enough computing resources to the SFS module given that shading is relatively a poor cue to 3-D shape? Recall that in the experiment by Mingolla and Todd (1986), judgements of surface orientation showed large differences across observers. So it is possible that the visual system can not make effective use of shading at all to derive 3-D structures. However, human observers showed high sensitivity to changes of surface curvatures defined by shading in a curvature discrimination task (Johnston & Passmore, 1994a). Here, observers viewed a patch of test surface which formed a fraction of the standard spherical surface defined by Lambertian shading. The curvature of the test patch was varied systematically. The task was to indicate if the test patch was more curved or less curved than a comparison sphere. The discrimination threshold for curvatures increased as the curvature of the standard sphere, revealing a low Weber fraction of only 0.1. This performance is comparable to the Weber fraction of around 0.07 observed for curvature discrimination tasks in which curved surfaces defined by disparity were used (E B Johnston, 1991). The existence of a low Weber fraction

means that humans can detect a rather small change in curvature of a 3-D surface defined by shading. Thus, it can be concluded that humans use shading quite effectively to derive the underlying 3-D structure and this effectiveness is comparable to that of disparity in computing the curvature of 3-D surfaces. So perhaps the large variances in the results observed by Mingolla and Todd (1986) were not due to the inability of the visual system to analyse shading but rather, caused by some other factors such as the difficulty of the task and geometrical properties (slant and tilt) that were measured.

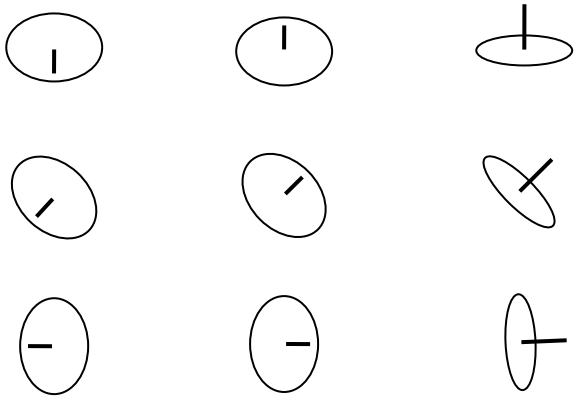


Figure 1.5 The probe image used in Koenderink's experiment (1992). The combination of a straight line and an oval depicts a circular disk with a needle erecting from the centre. The disk can be rotated in the three dimensional space.

Koenderink, vanDoorn and Kappers (1992) used a new method to evaluate human surface perception based on 2-D photographs of sculptures. The stimuli were composed of test photographs and probe images. The probe image consisted of an oval and a straight line starting from the centre of the oval and pointing towards the shorter axis of the oval (Fig 1.5). This combination depicts a circular disk in a three dimensional space which has a needle standing at the centre of the disk, pointing in the direction perpendicular to the surface of the disk. Observers adjusted the probe until they felt that the disk was sitting on the tangent plane of the perceived surface. The setting for each position on the photograph can be translated in terms of

perceived slant (rotation around the vertical axis of the image plane) and tilt (rotation around the axis of depth in space). Alternatively it can also be translated into the gradient vector of the perceived 3-D surface. Repeated settings made by each subject correlated well, indicating good reproducibility of the data for individual observers. More importantly, depth differences computed along closed triangles across the entire surface summed to zero, which is equivalent to zero curl for a continuous gradient field. This means that the settings conformed well to a perceived surface. However, there were large inter-observer differences. While the perceived shapes were quite similar across observers, the depth values were very different, leading to significant scaling effect between observers. Koenderink's method is a relatively easy task compared to the one used by Mingolla and Todd (1986) and even naïve observers could perform the task very quickly without training. Thus it has become a commonly adopted method in the study of shape-from-shading.

What is the significance of the large inter-observer variances reported in studies of SFS? Even when using an easier task and measuring more data points, researchers still failed to obtain consistent data across different observers (Koenderink et al., 1992). But interestingly, these inter-observer variations were not randomly distributed. Rather, they seemed to follow some systematic pattern such as the scaling effect. It has been argued that these variances may correspond to some ambiguities lying within the structure of 2-D shading and that resolving these ambiguities requires observers to apply their own "beholder's share" (Koenderink, vanDoorn & Kappers, 2001). In psychophysics, a good method should avoid this "beholder's share" as much as possible to exclude influences caused by individual differences. Unfortunately such

a method is not available for the study of SFS since no existing methods so far has managed to delivered consistent data across different observers.

Rather than trying to eliminate inter-observer differences, Koenderink et al. (2001) used these variances to provide useful information regarding the ambiguities associated with the perception of SFS in the hope that understanding such ambiguities would provide insight into the underlying visual mechanisms. Koenderink et al. (2001) tested four observers' perceived 3-D shapes based on photographs of four different smoothly curved object using three different tasks. One object had slight textures on its surface but shading was still the primary feature in the photograph. Three objects had relatively simple surfaces such as egg-shaped or vase-shaped ecliptics. The other was a more complex human mask shape. The three tasks were a probe disk task as used by Koenderink et al. (1992), pair-wise depth judgements between points on the surface, and the adjustment of a cross-sectional drawing to match the perceived surface. The pair-wise comparison task involved computing relative depth between two local points whereas the cross-section adjustment asked for global shape judgements. Linear regression revealed that participants largely agreed on the shapes of the three simple objects. But, once again, observers differed in the scale of the perceived depth: the scaling factor was up to 2.13. Shape estimates for the complex object correlated less well across observers. In addition, the depth scaling effect was also present, with a scaling factor up to 2.17. Correlations between different tasks were weak for any single participant. However, in a multiple regression of one participant's depths z_2 against depths z_1 of another participant as well as the image coordinates x and y, the resulting coefficients of determination were significantly improved between different participants, even for the photograph of the

complex surface. Significantly stronger coefficients were also found for different tasks after conducting a similar multiple regression. This result suggests that despite individual differences and inconsistencies between tasks, different tasks conducted by different participants nonetheless produced consistent shape estimates under affine transformations of perceived shape. That is, the shading defined 3-D surfaces are coded in the visual system as a functional of an affine transform $\hat{z}(x,y) = az(x,y) + bx + cy + d$ where $\hat{z}(x,y)$ is depth function estimated by an observer in a particular task, $z(x,y)$ is the depth function of the 3-D structure represented in the visual system, x, y are the coordinates of the image plane, constant a represents the scaling factor, while b, c and d control a shearing transforms of the 3-D surface (Koenderink et al., 2001). The constants defining depth scaling and shear represent the ambiguities that the observers must resolve by applying their "beholder's share". Thus each observer's response in any independent task was a sub-set of all the possible surface interpretations for affine transforms of the perceived 3-D structure. This theory can explain well the variances across observers reported by earlier studies as well as the variances of data obtained by different tasks.

1.1.4.3 Effect of illumination and surface material on SFS

Another interesting question about SFS concerns whether constancy can be achieved for the perceived 3-D structure under changes in lighting and surface reflectance. Lighting direction was varied in Johnston and Passmore's (1994a) curvature discrimination task, Sensitivity to changes in curvature did not vary as the illumination was rotated around the vertical axis of the image plane (tilt). But curvature thresholds increased as the illumination approached the viewing direction (reduced slant, frontal lighting). However this result can not safely support the

conclusion that the visual system analyses shading differently under different lighting conditions. Under frontal lighting, a Lambertian surface produces a luminance map that has considerably lower contrast than that produced by collimated lighting. Consequently the reduced sensitivity to curvature for frontal lighting condition could be due to poor detection of the luminance changes produced. On the other hand, an isotropic surface (sphere in this case) under collimated lighting with changing tilt direction will give rise to shading patterns that have about equal luminance contrast. Thus changes in surface curvature will produce similarly strong shading for all tilt directions. So it is not surprising that the curvature discrimination threshold were not affected when the tilt angle of the collimated lighting was varied.

In a related study, Curran and Johnston (1996) also had observers indicate which of two spheres was more curved. The surface reflectance could either be glossy or matte. Observers were more accurate when lighting was oblique than when it was frontal. For a frontal lighting, surface curvatures were consistently underestimated. For oblique lightings, observers were most accurate when the lighting was from above. Observers tended to underestimate curvatures as the light source was below the viewing axis. This was true for both types of surface reflectance but the trend was slightly weaker for glossy surfaces.

The effect of illuminant direction on SFS was also found in a complex scene understanding task (Koenderink, vonDoorn, Christou & Lappin, 1996a; 1996b). Observers adjusted pictorial reliefs of, respectively, photographs of sculptures (with shading), the silhouette of the original object, and a cartoon picture roughly equal to its contours (without shading). For the cartoon figure, observers produced a fully

articulated pictorial relief very similar to the actual photographs. But when viewing the photographs of the same sculpture under different lighting directions, the observers produced systematically deviated reliefs for individual stimuli. A similar phenomenon was also found by Todd, Koenderink, vonDoorn and Kappers (1996). The perceived picture relief from photographs of sculptures differed systematically between oblique and frontal light sources. Although large proportions of the variances (84%) could be accounted by affine transformations (cf, Koenderink et al., 2001), the residuals followed a systematic pattern. These residuals serve as an evidence that perceived shape is likely to vary with the lighting direction. Nefs, Koenderink and Kappers (2005; 2006) and Nefs (2008) also reported changes in perceived shapes from shaded objects under oblique lighting and frontal lighting. Applying an affine transform did not improve the coefficients of determination, suggesting substantial changes in perceived shape which could not be accounted by scaling or shear transforms. However, there were no obvious differences between matte and glossy surfaces

In an attempt to study the effects of lighting direction more quantitatively, Christou and Koenderink (1997) showed to observers stimuli of computer rendered ellipsoids with Lambertian reflectance. Perceived shapes differed for three different light source directions in that the perceived shapes all bulged towards the position of the light source. That is, the brightest point appeared closer to the observer than should be the case for a veridical interpretation. This effect was most pronounced for the lighting that was close to the viewing direction. Here, perceived depth was well predicted by an algorithm based on the linear regression between surface depth and the luminance gradient. For the other two lighting directions, the linear regression was also present

but not a dominant trend. The effect of illuminant direction was slightly weaker for surfaces with lighter albedo than that with darker albedo. This discovery confirms the speculation that little shape constancy could be achieved for the perception of SFS under different illuminations. In addition, it also suggests that the computational algorithms employed by the visual system could be different for oblique and frontal illumination. For the two oblique illuminations, the way that 3-D shape was derived seemed similar and the perceptual difference was due to the difference in luminance patterns. In a more expanded study, Khang, Koenderink and Kappers (2007) asked observers to judge the shape of computer rendered ellipsoids under various lighting conditions, surface materials and degree of specularity. Perceived shapes differed across the lighting conditions and surface materials but remained consistent when the degree of specularity was varied. Observers' judgements were most accurate for specular surfaces illuminated by collimated light farthest away from the viewing direction, although the judgment under all conditions was accurate overall.

To sum up, changes in illumination can influence perceived shape systematically. Therefore shape constancy should not be expected under changing illumination. But contradictory results have been reported for the effect of surface material. Matte and glossy surfaces are the most tested surface types. Perceptual differences were reported by some studies (Todd & Mingolla, 1983; Curran & Johnston, 1996; Khang et al., 2007) whereas others found no obvious effect when surfaces changed from matt to gloss (Nefs et al., 2006; Nefs, 2008). It should be noted that Nefs et al. used unusual stimuli with more complex edges and contours whereas the other studies all employed simple sphere and ellipsoid stimuli which had simpler outlines. Thus it is possible that

outlines and contours provided more information to help the observers to achieve a constant perception for different surface materials in the studies by Nefs et al.

1.1.4.4 Simple vs. complex stimuli: which are more suitable for SFS?

It has already been shown that different test stimuli can lead to inconsistent results. Humans are more likely to achieve accurate 3-D perception from more complex stimuli. For instance, observers were able to distinguish between elliptic and hyperbolic surfaces for shading patterns computed from a more complex object (Mamassian et al., 1996). Observers also managed to achieve shape constancy under changing surface materials for complex (Nefs et al., 2006; Nefs, 2008) but not simple stimuli (Todd & Mingolla, 1983; Curran & Johnston, 1996; Khang et al., 2007; Nefs et al., 2006; Nefs, 2008). One possible explanation for this difference is that edges and outlines in complex stimuli help to break inherent ambiguities associated with shading. However, the study of SFS in humans could be invalid if the effect of object outlines are not taken into full consideration. In a study of local surface perception (Mamassian & Kersten, 1996), observers consistently underestimated the surface slant and this bias increased as the real surface slant increased. But at the end of the report, they had to conclude that shading was probably not used by the observers during their experiment because observers' responses to the silhouette of the object followed a similar pattern. Some studies went even further using more complicated and more meaningful objects. In one example (Koenderink et al., 1996a), observers obtained a very similar shape judgement for a photographed shaded sculpture of human bodies and a cartoon figure of the same sculpture without shading, making the effect of SFS difficult to measure. Complex stimuli tend to be rich in other visual cues and contain information that can lead to higher level object recognition. Consequently, responses to complex stimuli may be confounded by judgements based on familiarity with the

objects. For example observers might implicitly reason “if it looks like a mug it must be cylindrical”.

Why did Koenderink et al. (1996a; 1996b) use complicated images and allow such apparently confounding factors to exist? After all, “cue-reduction” is a common strategy for studying perception based on single cues. One reason may be the potential ineffectiveness of shading compare to other cues. Prior to Koenderink’s study, it had been reported that the effect of shading could easily be overridden by other depth cues such as stereo (Bülthoff & Mallot, 1988), surface contours and outlines (Ramachandran, 1988; Knill, 1992). It was suspected therefore that alternative visual information had to be provided in order for SFS to function fully. Koenderink et al. (1996b) explained it with an analogue to clapping – it takes two hands to clap and shading alone may represent just one hand. Therefore it makes sense to use an object that is rich in visual information additional to shading to ensure the shading be made full use of.

Taken together the results discussed above suggest that shading needs other visual information to fully function as a cue to 3D shape. But the presence of too many additional visual cues could confound the measurement of the full effects of shading. Thus it is desirable to have a methodology in which information other than shading is just about enough to stop SFS from becoming a broken system. A realistic complex stimuli is perhaps less suitable for this purpose as information additional to shading in those stimuli can be more difficult to identify and control for.

1.1.4.5 Computational theories of human SFS

Although many computational algorithms can solve the SFS problem well under certain restricted conditions, these algorithms do not necessarily characterise human performance. No existing model of SFS takes adequate account of human perceptual responses in the whole process of SFS and very few have claimed any psychophysical plausibility. An exception is Pentland's biological model (1989) in which surface slant is linearly related to the underlying luminance. Pentland conducted a simple psychophysical experiment which proved that shape perception was consistent with this linear relationship for shading patterns composed of sine-wave functions. Having identified its psychophysical plausibility, Pentland proposed a method of implementation based on forward and inverse linear transforms which could be carried out by cells in visual primary cortex. But the validity of this linear relationship has not been extensively tested in human observers with other shading patterns. The geometry of lighting suggests that the linear relationship should only hold for oblique lighting directions where shading profiles are dominated by linear components. Quadratic components dominate when lighting is frontal with respect to surface undulations. Although slightly less accurate, shape perceptions for shading computed under frontal lighting can satisfactorily describe 3-D structures of the surfaces presented (Khang et al., 2007; Nefs et al., 2006; Christou & Koenderink, 1997; Todd et al., 1996; Koenderink et al., 1996b). Therefore, Pentland's theory is not a full account of human SFS.

In computer vision, solving SFS often involves finding the mathematical relationship between luminance and surface orientation. A classic way of describing such relationship is through a tool called the "reflectance map" which links luminance to

surface orientation on a gradient plane (Horn, 1977; Horn & Sjoberg, 1979). On the gradient plane, each spatial position corresponds to a two-element vector representing an orientation in the 3-D space. The value associated with each position is the luminance value in the shading image. A reflectance map can be uniquely determined for a surface of known material under a fixed distant point light source. Inspired by this idea, Seyama and Sato (1998) attempted to find the reflectance map assumed by humans so as to develop a psychologically plausible, computational theory of SFS. They tested observers with spherical and cylindrical surfaces with a light source at the viewing position. The obtained reflectance map was similar for all participants. Working in reverse, rendered images based on this reflectance map were perceived very accurately without the underestimation commonly found for surfaces rendered with Lambertian reflectance. Unfortunately, human reflectance maps were not obtained for surfaces under other lighting conditions. Therefore Seyama and Sato's method did not lead to a complete computational theory of human SFS.

Some hints as to how humans compute SFS can be drawn from past studies. Recall from section 1.1.4.3 that when the light source was close to the viewer, shape judgements could be explained by a linear regression model between the adjusted slant and decreasing luminance gradients, equivalent to the 'dark-is-deep' interpretation (Christou & Koenderink, 1997). That is, the brightest part of the image was seen as closest to the viewer. But for stimuli lit by oblique point light sources, a linear regression model could barely explain the data at all. This "dark is deep" strategy is similar to SFS algorithms developed to understand shading patterns under diffuse lighting (Langer & Zurker, 1992; Stewart & Langer, 1997). However, when testing human depth perception for shading patterns generated under diffuse lighting,

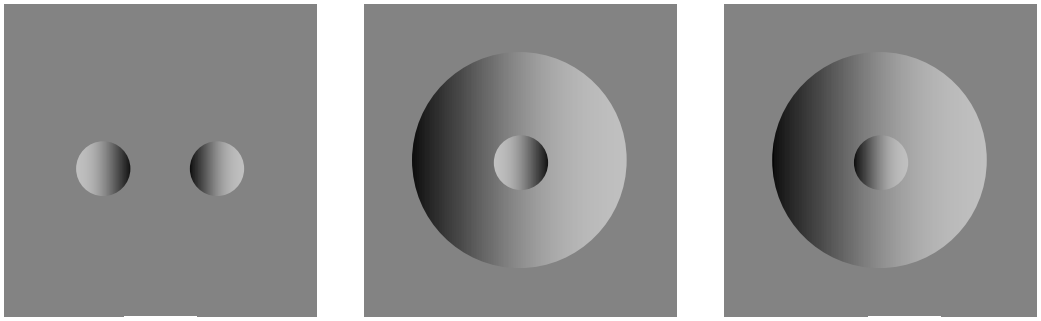
Langer and Bülthoff (2000) found that the observers utilized a strategy that was more powerful than the simple "dark is deep" Rule. Nevertheless, none of above studies has conclusively established a theory for shape from shading. One important reason is that most studies used complex object realistically rendered by computer programs. Admittedly, the choice of realistically computer rendered object does not undermine any of those qualitative conclusions discussed above. But it is hard to form firm, quantitative computational theories of SFS based on results which are potentially confounded due to the presence of edge contours.

1.1.5 Knowledge of Light source

The perception of SFS is often studied alongside the estimation of light source direction. This is because it is impossible to judge one without knowledge of the other unless one is assumed. Many SFS algorithms in computer vision require the illuminant direction to be known because shading is a function of the angle between the surface normal and the light source direction. However, whether or not knowledge of the lighting direction is a prerequisite in the visual system when solving SFS is an open question. Mingolla and Todd's (1986) found that error data of light source estimation did not correlate with that of surface perception, indicative of two independent processes. Further, Mamassian and Kersten (1996) found large errors for the tilt of the light source computed from observers' responses even when light source tilt could be very easily determined from the image. This result led them to conclude that the illumination direction was probably not used to aid in SFS tasks. But humans do seem to be able to infer the direction of a light source from cast shadows, specular reflections (Mingolla & Todd, 1986; Liu & Todd, 2004) and the second-order statistics of relief textures (ie finely rippled surfaces, Koenderink, vonDoorn & Pont,

2004; 2007; Pont & Koenderink, 2007). In addition, luminance gradients can also help human observers to indicate light source direction (Pentland, 1982).

It seems necessary to differentiate these two types of knowledge on light source. One, implicit lighting, is the light source suggested by perceived surface orientation following SFS (Mamassian & Kersten, 1996). In other words, the observer decides on the surface shape and interprets the lighting direction accordingly. The other, explicit lighting, is obtained directly from lighting cues in the image and can be assessed by tests of light source estimation. Humans also have a third type of knowledge regarding the light source, namely prior assumptions about where light is most likely to come from: lighting priors. Two known lighting priors are that lighting is directional (like the sun) and comes from above and slightly to the left of the observer (Ramachandran, 1988; Sun & Perona, 1998, Mamassian and Goutcher, 2001) and that lighting is diffuse and hemispherical, like the sky (Langer & Bülthoff, 2000; Tyler, 1998). The question then becomes how these three types of knowledge on light source are related and what role each type plays in SFS.



a b c

Figure 1.6 A demonstration of the global shading effect on breaking the convex & concave ambiguity. (a) Circular horizontal luminance ramps will appear a bump regardless of the direction of the gradient due to a bias of global convexity (Reichel & Todd, 1990; Langer & Bulthoff, 2001; Liu & Todd, 2004). (b) A smaller circular luminance ramp which has a gradient direction opposite to the larger circular appears a concave dent. (c) If the smaller circular is rotated 180 deg such that it has the same gradient orientation as the larger circular, it appears a convex bump (After Koenderink & vanDoorn, 2004)

A known effect of light priors is that the preference that light comes above rather than below helps to break the ambiguities between concave and convex surfaces. But it has been shown that the preferred light source direction for humans is a wide spread of directions centred at above left with large individual differences across observers (Adams, 2007). However the estimation of light priors is normally defined in terms of tilt angles. Light priors regarding the slant angle are seldom investigated. The effects of explicit light sources have been studied in the context of global shading effects (Erens, Kappers & Koenderink, 1993b; Koenderink & vanDoorn, 2004). Shading or shadows in areas surrounding an object could indicate the illuminant direction. If so the perception of the object in question would be affected by the surrounding shading patterns. It has been reported that the convex & concave ambiguity can be broken by global shading (Koenderink & vanDoorn, 2004), as demonstrated in Figure 1.6. However the presence of global shading did not improve the accuracy of SFS in another experiment (Erens et al., 1993b). It seems that explicit light source information is not used as a prerequisite in human SFS. Instead, it had a similar role to lighting priors; breaking ambiguities associated with shading.

So far there is no conclusive result available that could clarify the relationships between light priors, explicit light sources and implicit lighting. But a theory can be formulated to address this issue. Recall in section 1.1.4.3 it was shown that SFS relied on the luminance distributions. Since implicit lighting is computed from perceived shapes, it may be more related to luminance distributions as well. But explicit cues to lighting can be obtained from many sources such as cast shadow, specular highlight, edges, and 3-D structures induced by shading (e.g. Fig 1.6). Each of these cues has a different reliability and strength and cues can act against or in favour of each other. If

explicit light sources were not involved in shape computation in a point wise manner but helped to resolve shading ambiguities much as lighting priors do, then explicit and implicit light sources might appear mutually exclusive when the reliability assigned to the source of 3-D structure induced by luminance distribution is low. That is, explicit light sources and implicit light are drawn from independent sources of information. But when no other sources are available and the source of 3-D structure induced by shading is reliable, explicit light source should correlate with implicit light sources.

1.1.6 Disambiguating origins of luminance variations

Another constraint often imposed by SFS algorithms is that surface materials are Lambertian with constant reflectance. The benefit of applying such a constraint is that images contain shading only and so can be a direct input to the system. In reality, luminance variations can result from changes in reflectance as well as shading (surface orientation). The fact that the uniform reflectance constraint is seldom satisfied in natural scenes has significantly hampered the application of SFS algorithms in real world applications. On the other hand, human SFS seems to be robust to the natural environment. Does this mean that the human visual system has a stage responsible for disambiguating luminance in a scene? There is evidence to suggest that this is very possible.

Humans do not judge the lightness of a surface simply based on the perceived brightness rather lightness perception is often affected by contextual information and spatial arrangement (Gilchrist, 1988; Gilchrist, 1977). Induced lightness can not be explained by low level inhibition but seems to suggest an awareness of how illumination and transmitting atmosphere affect the perceived brightness of 3-D structures (Knill & Kersten, 1991; Adelson & Pentland, 1996; Anderson & Winawer,

2005; Adelson, 1993). In addition, colour perception is also influenced by 3-D layout (Bloj, Kersten & Hubert, 1999). Perceived colour and perceived lightness are closely related to reflectance (capturing albedo and pigment respectively). The fact that humans take illumination into account when judging the reflective properties of a surface indicates separate representations for illumination and reflectance in the visual system. A generic theory has been formulated for lightness perception, the perception of transparency, and the perception of shading and shadows. This theory states that at a certain stage of visual processing, the image is decomposed and represented in different layers according to sources of origin such as illumination, reflectance and optical medium (Kingdom, 2008; Gilchrist, 2006, p189; Anderson & Winawer, 2005) – a process similar to that described as extracting the intrinsic image in machine vision (Barrow & Tenenbaum, 1978).

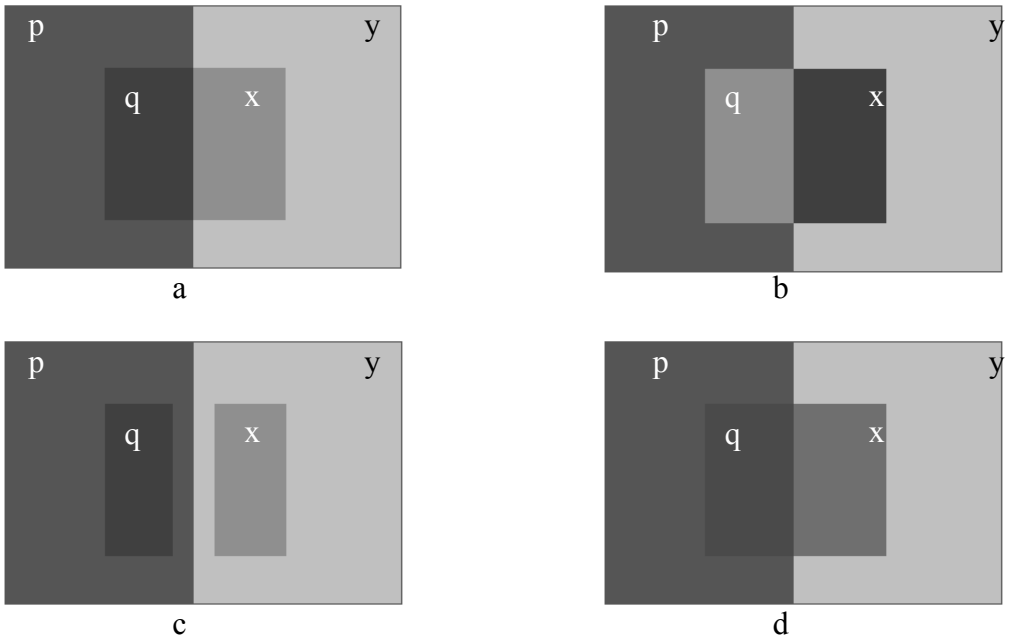


Figure 1.7 Effect of edge intersection. (a) luminance values along each edge obey the rule of "ratio invariance", i.e. $p/y = q/x$, giving a shadow impression to either the central square or the left half of the figure. (b) If the sign of edges or the contrast sign changes, in this case $p/y = (q/x)^{-1}$, the shadow impression disappears and both edges look more like reflectance changes. (c) If edge intersections are removed, the impression of changes in illumination is weakened. (d) The sign of edges is same as (a) but the luminance ratio is changed such that $p/y > q/x$. The central square now appears as a transparent surface over the background. (After Kingdom 2008)

Humans carry out layer decomposition with the help of a variety of cues (Kingdom, 2008). When viewing a grey target and a white paper half in a shadow, observers assigned the target with higher grey levels than when the surrounding area was obscured (Gilchrist, 1988). Gilchrist suggested that humans identify the darker half of the white paper as a less illuminated area when contextual information containing edge intersections were available. In the real world, the effect of illumination is multiplicative so that any luminance ratios remain constant even when the illumination changes (see Fig 1.7a). Thus edge intersections should obey the rule of "ratio-invariance", corresponding to the situation where illumination edges intersect with reflectance edges (Gilchrist, 1988; Kingdom, 2008). In contrast, if the sign of edges change across edge intersections, both edges are unlikely to be caused by illumination (see Fig 1.7b). Moreover, the perceptual decomposition does not occur if the spatial arrangement of edge intersection is destroyed (Fig 1.7c).

Figure 1.7a can be also perceived as a transparent square floating over the background. According to Metelli's transparency theory (1974), edge intersections with "ratio invariance" also signature a non-reflective transparency. The restriction of "ratio invariance" can be relaxed to achieve a perception of transparency as long as the signs of edges remain consistent across intersections (Fig 1.7d). This combination typically corresponds to a background surface seen through a transparency with a reflective component (Kingdom, 2008; Singh & Anderson, 2002). Gilchrist (2006, p192) argued that the process of edge classification is critical to the process of lightness perception. But the nature of the computation that follows edge classification to achieve lightness has not been made explicit, although Gilchrist, Delman and Jacobsen (1983) suggested a process of edge integration. There is also a

weakness in “ratio-invariance” as a cue for layer decomposition; while it signifies the existence of an illumination edge, it does not specify which edge of the intersection is due to reflectance and which is due to shading. As illustrated in Fig 1.7a, either the central square or the left half of the figure can be perceived as lying in the shadow.

Edge sharpness is considered by some researchers as another cue to layer decomposition (Land & McCann, 1971; Horn, 1974). Land and McCann’s Retinex theory (1971) assumes that illumination changes in a field are gradual and smooth such that they are invisible to a low-level edge detection scheme. Thus illumination and reflectance can be separated by thresholding luminance gradients. Horn (1974) extended the Retinex theory and developed an algorithm that could remove lightness from 2-D images. Horn’s algorithm is based on the Laplacian operator and its inverse which he believes behave similarly to some cells in visual cortex. One problem with classifying gradients is that one has to reintegrate them afterwards: the gradient process needs an inverse. Horn’s algorithm provides for reintegration and has served as a general framework for future algorithms for layer decomposition and intrinsic image separation. For example, Gilchrist et al. (1983) suggested that the rule of “ratio invariance” could be added to make an edge classification unit together with the thresholding scheme proposed by the Retinex theory. However, the notion of gradual and smooth nature of illumination changes is more empirical than ecologically plausible. Edge shadows can be very sharp (Fig 1.7a) and sharp luminance changes due to shading are also frequent in natural scenes, e.g. at the corners or vertices of 3-D objects (Sinha & Addelson, 1993). More importantly, humans have no problems in interpreting shadows and shading with sharp edges (Kingdom, 2008; Gilchrist, 2006; Gilchrist, 1979; Adelson & Pentland, 1996). From this perspective, the “illumination

change is smooth" rule is seen as more of a general guide than a reliable rule (Kingdom, 2008).

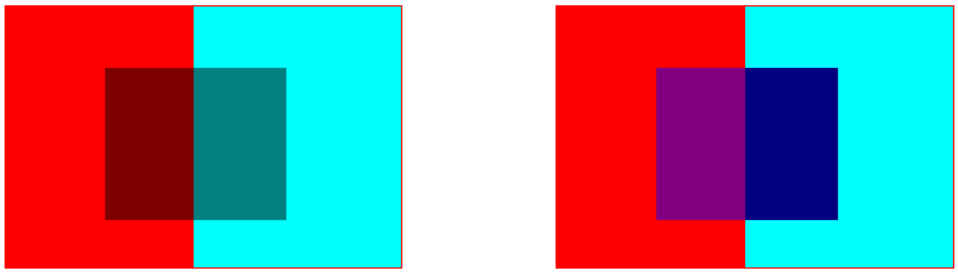


Figure 1.8 colour brings even more clarity to shadow/shading. (a) Luminance changes along the border between the background and the central square but hue remains consistent. This produces even stronger shadow impression than Fig1.7a. (b) Both hue and luminance change along the border between the central square and the background. The luminance in the central square is the same as in (a). (After Kingdom 2008)

The cues described above are both suggestive of illumination changes. Changes in colour, on the other hand, suggest material changes. Kingdom, Beauce and Hunter (2004) showed that adding colour to luminance edge intersections facilitated identifications of shadows. The effect of colour in shadow identification is illustrated in Figure 1.8a where luminance changes achromatically between the background and central regions. Hue changes only along the middle edge. This combination produces an even stronger impression of illumination changes than that seen in Figure 1.7a. The central square in Figure 1.8b has the same luminance level as that in Fig 1.8a, but appears as a patch with different reflectance from the background because hue changes across the luminance border.

Another related study (Kingdom, 2003) linked the human ability to disambiguate luminance variations in SFS. This study suggests that luminance variations classified as shading provide direct input into SFS. Stimuli consisted of a luminance-defined sinusoidal grating and a sinusoidal grating defined by isoluminant red-green shifts. The two components had the same orientation and were combined either in-phase or

out of phase. An orthogonal red-green grating was added to the main pair in a plaid configuration. The degree to which the luminance sinusoid appeared as shading was measured by ratings of the perceived depth of the apparent corrugation, a task that involves the process of SFS. Results showed that the perceived depth was enhanced when the phase alignment of the mixed colour and luminance component was destroyed or the contrast of an in-phase colour component was reduced. Thus luminance changes that are aligned with changes in hue are likely to be perceived as reflectance whereas non-aligned variations in hue and luminance trigger the impression of shading.

The two studies above prove the importance of colour in disambiguating luminance, but they reveal different aspects of the process. The results of the shadow experiment are consistent with Gilchrist's idea of edge classification (1983). Thus colour can be an effective addition to the edge classification unit within the layer decomposition framework. Olmos and Kingdom (2004) exploited this idea to develop an algorithm that separates shading from reflectance. This algorithm finds edges via a classic edge detection method and categories them into illumination and reflectance edges, by applying the rules discovered in the shadow experiment (Kingdom et al., 2004). The edge types can then be reintegrated separately to obtain the corresponding layers. However, Kingdom's (2003) shading experiment provides a strong argument that layer decomposition may not be based on classified edges. Edge information in the sinusoidal gratings could not be easily detected by known edge detectors, but the separation of shading and reflectance was still effective, suggesting that the decomposition could be based on correlations between channel outputs rather than just edges.

Like hue, texture amplitude (or luminance amplitude) can be used by humans to differentiate shading from reflectance (Schofield, Hesse, Rock & Georgeson, 2006). Here the authors were interested in the relationship between modulations of local mean luminance (LM) and local luminance amplitude (AM). AM was calculated as the standard deviation of a local patch of luminance values, making up a textured pattern. This is a measure of the absolute difference in pixel values rather than local contrast which is a relative measure. The physics of shading suggest that, low light intensities will reduce LM as well as AM such that the two components are positively correlated. Local contrast (CM) meanwhile is constant. Figure 1.9 illustrates this relationship.

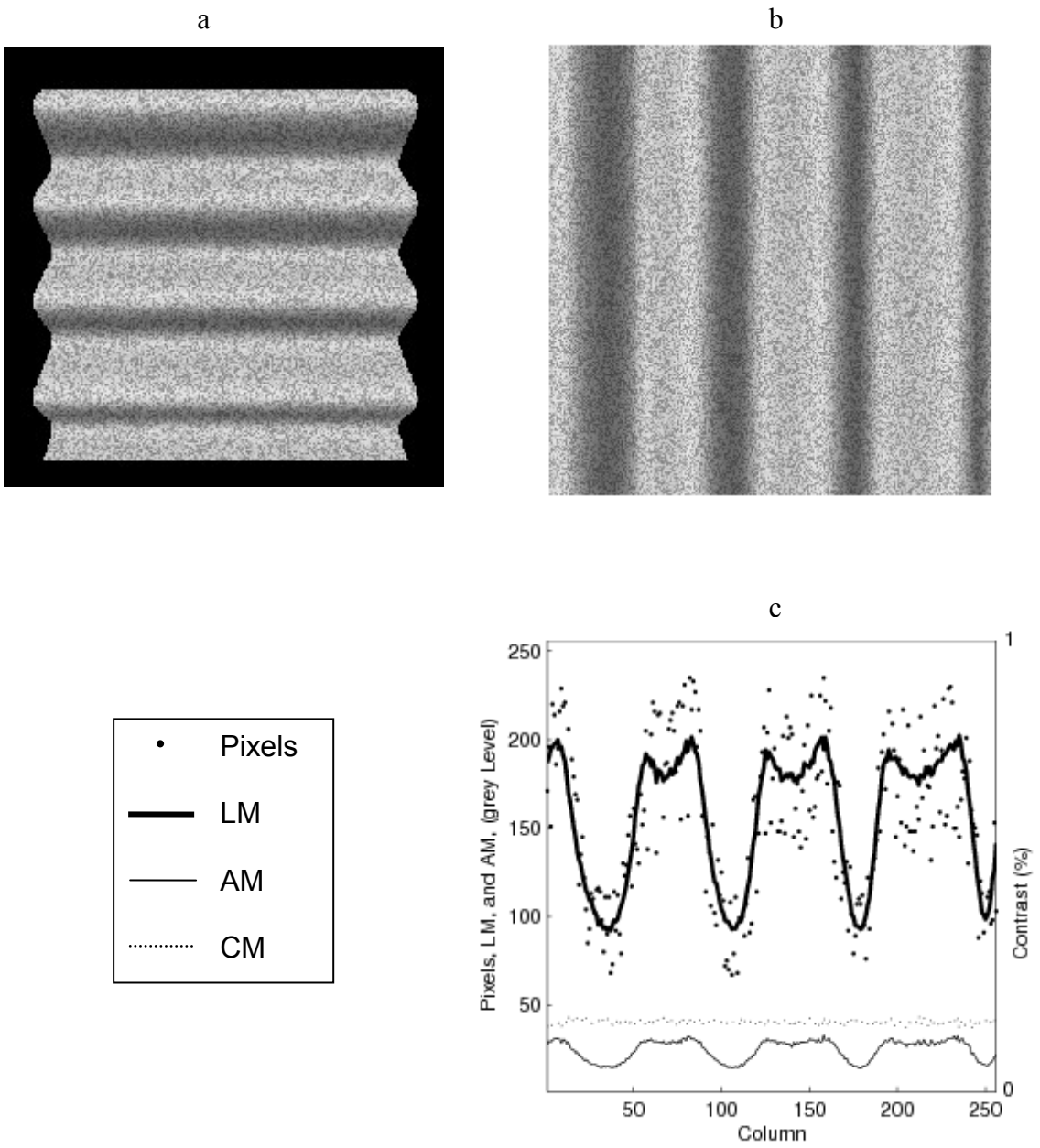


Figure 1.9 The relationship between LM, AM and CM under variations in light intensity. (a) A computer rendered image (256 by 256) depicting a corrugated surface with uniformly painted texture (the surface is smoothly corrugated) is lit by a single point source from above. (b) A portion of (a) cropped, rotated and magnified. (c) Cross sections along the central row through (b); thick dots represent the pixel values in the central row; the solid thick line represents the mean pixel values in each column (LM); the thin line represents the stand deviation of pixel values in each column (AM); the local contrast of pixel values in each column (CM) is defined by the ratio AM/LM and described by the thin dotted line. When the intensity of the light varies due to the surface corrugation, AM varies in pace with LM but CM remains almost constant. Images from Schofield et al., 2006, with permission from the authors.

The relationship between LM and AM was found to be an effective cue in differentiating between shading and reflectance. Figure 1.10 shows one of the stimuli

used in Schofield et al's experiments. It is composed of two visible sine wave gratings at two orthogonal orientations together with noise textures. In the right oblique, AM is varied in-phase with LM (LM+AM) such that the two signals are positively correlated, consistent with shading. In the left oblique, AM is varied in anti-phase with LM (LM-AM) in a way that is not consistent with variations in shading. When the two types of cue (LM+AM & LM-AM) are presented together (in a plaid), human observers tend to perceive LM+AM (right oblique in fig 1.10) as a shading pattern giving rise to the perception of a surface corrugated in one direction only. LM-AM (left oblique) is seen as flat stripes that are 'painted onto' the surface. These percepts were measured by assessing perceived depth amplitude and (like Kingdom's 2003 study) the result demonstrates that the disambiguated luminance variations are carried forward for the analysis of SFS in the visual system. Again this process does not seem based on edge operations. No algorithms have yet been developed to implement this kind of layer decomposition nor has a biologically plausible implementation been proposed with regard to the role of luminance amplitude in luminance disambiguation (Note Schofield, Rock, Sun and Georgeson, 2009 & Schofield, Rock, Sun, Jiang and Georgeson, 2010 in press, present such a model based on work, presented later in this thesis, carried out by the author).

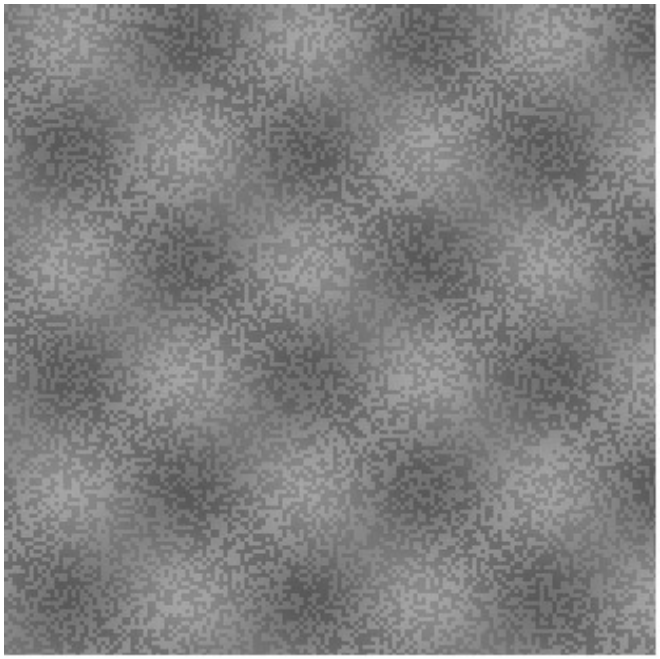


Figure 1.10 A plaid consisting of two orthogonal sine wave luminance gratings additively combined with two orthogonal amplitude modulations. From top let to bottom right, the luminance modulated sinusoid varies in-phase with the amplitude modulated sinusoid, equivalent (LM+AM). From bottom left to top right, the sinusoidal luminance grating varies in anti-phase with the amplitude modulated sinusoid (LM-AM). The right oblique was perceived as shading resulting from corrugated surface whereas the left oblique was perceived much flatter (image from Schofield et al., 2006, with permission from the autors).

To summarize, human SFS is very robust in the natural environment in that it seldom confuses shading with reflectance variations. Accumulating evidence points towards the idea that image intensities are disentangled and represented in different layers according to their origin. The illumination layer can serve as a disambiguated input to SFS, as suggested by some studies (Kingdom, 2003; Schofield et al., 2006). A general framework has been proposed to tackle the algorithmatic level of the visual process of layer decomposition (Horn, 1974; Gilchrist et al., 1983; Gilchrist, 1988; 2006). Central to this framework is edge detection and edge classification followed by reintegration. Various cues are contained in the edge classification unit to guide the process. A number of algorithms have been implemented under this framework based on cues such as edge sharpness and colour alignment. But the framework will fail to

explain the observation that humans can separate shading from reflectance for luminance variations where edges are obscure and hard to detect.

1.2 Towards a model of SFS in human

The challenge of constructing a model of SFS in the human visual systems can be tackled in different stages and at different levels. More specifically, the whole process of SFS can be divided into a series of functional sub-units. For each sub-unit, a computational theory should be identified and a method by which neural mechanisms could implement it determined. But first, we should specify the role of each functional sub-unit.

The most obvious sub-unit is the unit that computes surface orientation from shading (Shape recovery unit). It seems reasonable that this sub-unit should be preceded by luminance disambiguation which only passes shading variations into the shape recovery unit. Then, like most other visual models, there should be a pre-processing stage which mimics the very lowest level of visual processing: feature extraction. This framework is shown in Figure 1.11. In the following subsections, each sub-unit will be analysed and their transfer functions identified based on the required input-output relations.

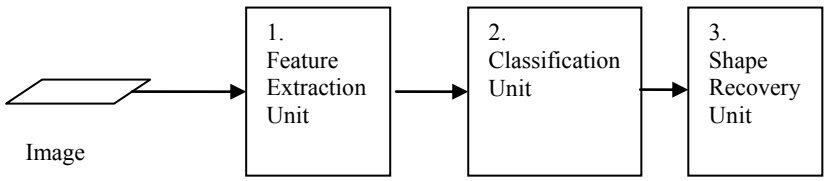


Figure 1.11 Proposed framework of SFS in human vision. The retinal image is first coded and represented as features. Coded representations are then classified into shading and reflectance with shading signals being passed onto the next stage of processing. The last unit operates on the shading signal and derives surface orientations from it.

1.2.1 The feature extraction unit

1) Specifying the input and output

It is widely accepted that one of the tasks involved in the early stage of feature representation in the visual system is to make explicit the important information contained in the retinal image (Bruce et al., 1996, p76). A representation of these features such as local changes in luminance is normally called the primal sketch (Marr, 1982) and obtaining such representation has become a common practice in both computer vision and human vision studies. At this stage, the input signal is the original retinal image and the output signal should contain a full representation of the input under some coding scheme. Ideally these representations can fully characterise all the luminance variations present in the image. Furthermore, for the purpose of the next unit, the output should also provide information that is required to disambiguate the origin of luminance variations. Although the achromatic features serving to disambiguate luminance variations are not well specified, some hypotheses can be proposed. Recall that in Figure 1.10 the two luminance gratings were perceived differently but what made them distinct was the phase of AM. Thus it is very likely that the process of luminance disambiguation involves detecting AM: a second-order cue (see Schofield et al., 2006 & among others Schofield & Gerogeson, 1999). The hypothesis proposed here is that, as a second-order entity, AM is detected by second-order mechanisms in visual systems and is exploited to help with the luminance disambiguation process in the next stage. This hypothesis isn't restricted to that particular type of stimuli only. It can be generalized to other achromatic cues as well. For example, the heuristic classification based on edge intersections discussed in section 1.1.6 (Fig 1.7) can be also thought as a second-order processing: local edge contrast is computed and then compared at a more global scale.

As a relatively independent processing in human vision (Zhou & Baker, 1996; Schofield & Gerogeson, 1999), second-order vision shows a number of characteristics distinctive from processing first-order luminance defined stimuli such as its modulation frequency dependency and carrier frequency dependency (Sutter, Sperling & Chubb, 1995; Dakin & Mareschal, 2000; Schofield & Georgeson, 2003). If this hypothesis is true, the effectiveness of the layer decomposition should show similar frequency dependencies as does the second-order vision. Chapter 2 and 3 of this thesis is dedicated to testing predictions based on this hypothesis and the result will be helpful in the formation of a complete the model for the feature extraction unit.

2) Specifying the computational algorithm

How the visual system codes the retinal image is a well studied subject and both the computational theory and neural implementation have been extensively explored. The process is typically modelled as a series of filtering processes which decompose the retinal image into different frequency channels and orientation bands. In this way, the entire luminance variations are fully coded by the energies (also called coefficients) in those frequency channels and orientation bands. Second-order signals, also known as non-Fourier cues (Chubb & Sperling, 1988), were first used by Cavanagh and Mather (1989) to describe modulations of a carrier signal that are themselves defined by non-luminance variations such as contrast and orientation. Several models for detecting second-order cues have been proposed. A typical computational mechanism for detecting second-order signal contains two filtering processes; one responsive to the carrier and the other responsive to the modulation. These two filtering process are normally separated by a non-linear rectification stage. For this reason, models of second-order vision with the similar structure are called a Filter-rectifier-filter (FRF).

3) Possible implementation by known neural mechanisms

The behaviour of some cells in area V1 can indeed be modelled as linear summation across their receptive field and the responses of such cells to visual stimuli can be predicted by a filtering process (Heeger, 1993; Campbell et al., 1968; Hubel & Wiesel, 1962). These cells are tuned to different orientations and frequencies and their responses are likely to correspond to important features such as edges and bars in real images (Marr & Hildreth, 1980). Cells responsive to second-order stimuli also have been found in early visual areas (Zhou & Baker, 1996). These cells tend to be tuned to lower frequencies and could conduct the same computation as the FRF channels in models of second-order vision.

1.2.2 The classification unit

1) Specifying the input and output

One of the findings in Schofield et al's (2006) study is that as a cue for shading, the relation between LM and AM is most effective when LM+AM and LM-AM are seen together, intertwined within a single stimulus. That is, LM-AM is more likely to get rejected as shading when presented with LM+AM. Although slightly less depthy, LM-AM can be perceived as shading as well as LM+AM when they are presented individually. In a later experiment (Schofield et al., 2009 & 2010 in press), observers' perceived depths were recorded for LM+AM single oblique, LM-AM single oblique and plaids formed of the two combinations. The results are shown in Figure 1.12. The *x* axis represents the modulation depth of AM. Negative values indicate LM-AM. The perceived depth for the combination of LM and AM in a plaid appear to be a sigmoidal function with LM-AM being seen as flat. However, single oblique stimuli appear more depthy in general and decline only slightly when AM is out of phase

with LM. Thus it seems that at this stage the visual system operates on all the LM signals and picks up the signal that it believes most likely associated with shading. Since an LM signal is equivalent to the response of a stimulus to a filter, it can be regarded as a coefficient representing the energies at a particular frequency channel and orientation band. Thus this unit probably takes all those coefficients obtained from the previous unit as input and applies certain rules to enhance the energy in some channels while suppressing others.

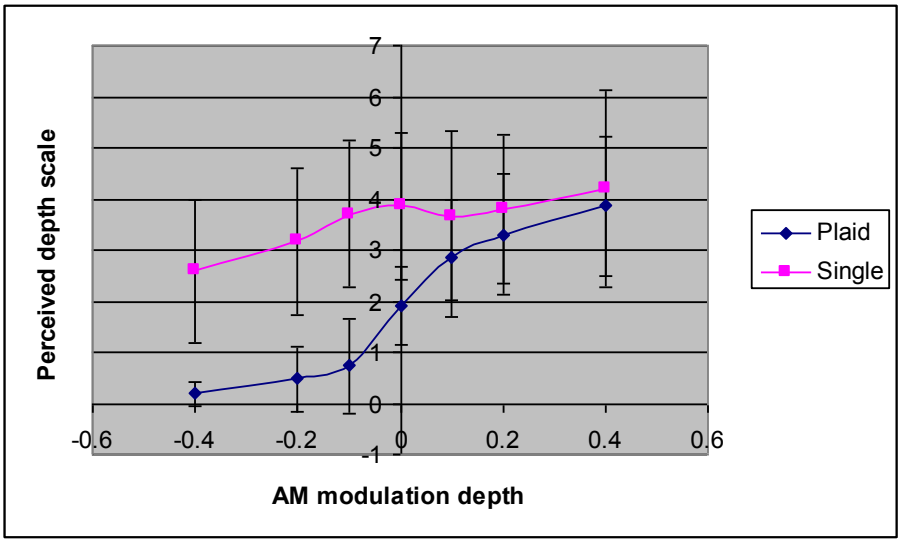


Figure 1.12 The perceived depth as of function of AM modulation depth for plaid (diamond) and single (square). Negative AM indicates LM-AM. The plot is reproduced from data taken from Schofield, Rock, Sun and Georgeson, 2009 (VSS poster) . Note although the current author devised a model for these data presented later in this thesis he was not involved in data collection.

2) Specifying the computational algorithm

The result of Figure 1.12 indicates that there might be a selection scheme based on the relationship of LM and AM. Schofield and Georgeson (1999) found no sub-threshold summation between LM and AM (they use the term CM) which is a strong implication of two separate channels for the processing of luminance modulations and contrast modulations. But in a later study, Georgeson and Schofield (2002) reported transfer of aftereffects between the two signals, indicating a later stage of processing at which the two signals were integrated. So it is psychophysically plausible to

introduce some sort of integration between LM and AM. The sigmoid shape of perceived depth for the plaid stimuli suggests that this integration may be followed by an inhibitory network working across orientation bands. Such cross-orientation inhibition has been discovered in other behavioural studies. For example, human observers demonstrate similar cross-orientation masking for purely first-order stimuli (Foley, 1994; Meese & Hess, 2004; Meese & Holmes, 2007).

3) Possible implementation by known neural mechanisms

Some cells in cat areas 17 and 18 are responsive to both first-order and second-order stimuli (Zhou & Baker, 1996; Mareschal & Baker, 1998; Zhan & Baker, 2008). These cells respond to combinations of LM and AM as if computing a linear sum between the two cues (Hutchinson, Baker and Ledgeway, 2007) although their sensitivity to AM is much lower than that for LM. Furthermore, simple cells in V1 respond non-linearly to single (Albrecht & Hamilton, 1982) as well as superimposed pairs of gratings (Bonds, 1989), which may be the neural basis for the aforementioned cross-orientation inhibition observed behaviourally (Foley, 1994; Meese & Hess, 2004).

1.2.3 The shape recovery unit

1) Specifying the input and the output

As the name suggests, this unit takes the shading information from the previous stage and computes the surface orientation for each point in the image which then leads to the computation of depth. The output of such a unit is a viewer-centred 3D representation equivalent to the 2.5 sketch proposed by Marr (1982).

2) Specifying the computational algorithm

The shape recovery unit is the hardest of the three proposed units to characterise. To create a psychophysically plausible model, experimental data on human SFS must be available. Unfortunately, much of the data collected to date is not suitable for the purposes of this thesis. One of the objectives of this thesis is to obtain data that is not confounded by object outlines and reflects more directly the computation that the visual system conducts to recover surface orientation from luminance variations (see section 1.1.4.5). The choice of test stimuli is vital: a computer generated, realistic object will provide unwanted visual information such as self-shadow, outlines, and object identity and hence will confound the results. In addition, any realistic shading pattern will be produced by some pre-defined mathematical rendering model, which is not necessarily the one that is assumed by the visual system. In theory, one could test many shading patterns produced by various mathematical models and find the one that is most consistent with observers' responses. But doing so would be impractical. In this thesis, a different methodology is proposed. Instead of viewing realistic objects, observers judged the orientation of the apparent surface based on luminance variations alone. These luminance variations are not subject to any pre-defined shading model, and did not represent objects, or present contour or occlusion cues. Thus the results presented later reflect an un-confounded mapping between shading and perceived surface orientation.

1.3 Thesis structure

Chapters 2 and 3 of this thesis test human performance with respect to luminance disambiguation. If second-order vision is indeed involved at this stage, we should expect to see an influence that is consistent with known properties of second-order

vision. Chapter 4 uses data from Schofield, et al., (2009 & in press) [data not collected by the author] to construct a model of the classification stage. This model is believed to be biologically plausible because 1) it fits well the psychophysical data, 2) it can predict the data obtained in chapter 3 reasonably well, and 3) it is consistent with known neurophysiology in early visual area of monkey and cat. Chapter 5 introduces an algorithm which decomposes a real image into its shading and reflectance components. This algorithm is built upon the same principles as the model of chapter 4 but uses the edge classification framework. Experimental results on some real images show that the algorithm can separate shading and reflectance when a texture is present and the degree of shading is not so great as to reduce texture contrast below usable levels. Chapter 6 examines human shape judgments based on luminance variations only. A computational theory of shading analysis in the visual system is then proposed and some predictions made. Chapter 7 concludes the thesis highlighting possible improvements to the model and computational algorithm.

2. The role of carrier frequency in shape-from-shading

This chapter links the perception of shape-from-shading to second-order vision by showing that the carrier frequency of a texture affects the impression of shape-from-shading in human observers. Second-order signals such as AM are detected by mechanisms that are sensitive to the composition of the carrier signal. Hence changing the carrier frequency may affect the detection of AM signal and, where the AM signal is rendered weak, reduce the perceptual difference between LM+AM and LM-AM.

2.1 Introduction

2.1.1 Second-order vision

In the context of human vision, second-order signals refer to stimuli that are defined by local properties (e.g. contrast and texture) of first-order luminance defined carrier signals. Many studies have suggested that such variations are detectable by the visual system in both humans (Chubb & Sperling, 1988; Cavanagh & Mather, 1989; Wilson et al., 1992; Sutter et al, 1995; Schofield & Georgeson, 1999; Dakin and Mareschal, 2000; Ellemberg, Allen & Hess, 2006) and other animals (Zhou & Baker, 1993; Zhou & Baker, 1996; Mareschal & Baker, 1999; Mareschal & Baker, 1998a; 1998b; Zhan & Baker, 2008). There is also evidence to suggest that the mechanisms for detecting second-order stimuli have similar behaviour to first-order mechanisms. For example, Albright (1992) reported that certain neurons responded similarly to stimulus irrespective of the physical cues defining it, of which he termed the phenomenon *form-cue invariance*. Testing with moving second-order stimuli, Mareschal and Baker (1998b; 1999) recorded similar optimal orientation tuning and similar spatial and temporal bandwidth to envelope (second-order) and corresponding luminance (first-

order) signals. In psychophysical studies, Schofield and Georgeson (1999) found similarities in the shape of the modulation sensitivity functions (MSFs) for second-order contrast modulations and first-order luminance modulations of the same type of carrier noise. Both MFS's were low pass. Jamar and Koenderink (1985) measured detection thresholds for sinusoidal amplitude modulations carried by noise patterns that had been band pass filtered according to the contrast sensitivity function. Modulation threshold increased with the spatial frequency of modulation, suggesting a reduction in sensitivity for high frequency modulations. More recently, in a discrimination task at detection threshold that was used to determine the number of channels making up early spatial frequency processing, Ellemberg et al. (2006) reported the same number of second-order channels and first-order channels at frequencies up to 2.0 c/d but fewer second-order channels at higher frequencies. Reconciliation of these findings suggest that mechanisms for processing second-order modulations probably have very similar behaviour, but are tuned to lower spatial frequencies compare to their first-order counterparts.

The detection of second-order signals does not only depend on the properties of the envelope; detection also depends on the first-order signal that carries the second-order modulation (Mareschal & Baker, 1999; Sutter et al, 1995; Dakin & Mareschal, 2000; Schofield & Georgeson, 2003; Song & Baker, 2006; Zhan & Baker, 2008). There is some evidence showing that second-order mechanisms in human vision are tuned to carrier frequency such that each channel is responsive to its own optimal carrier frequency (Sutter et al, 1995). However this idea has been challenged by physiological studies in cat areas 17 and 18 where no fixed optimal carrier frequencies have been found (Mareschal & Baker, 1999). Moreover, later

psychophysical studies (Dakin & Mareschal, 2000) have also failed to find optimal tuning for carrier frequency. If second-order vision is mediated by a filter-rectifier-filter structure (as suggested by Wilson et al., 1992), then Dakin and Mareschal's results suggest that the second-stage filter is connected to a broad range of first stage filters whose frequencies lie at least 3 octaves above the preferred frequency of the second-stage filter. Above this ratio (3~4 octaves as suggested in their work), the second stage filter receives input from first order stage filters across a broad range of orientations. Below this ratio, the second stage filter seems only wired to the first stage filter with orientations orthogonal to that of the second stage filter.

2.1.2 Effect of textures on shape-from-shading

Sakai (2006) has shown that adding random textures to luminance gradients can facilitate depth perception. In this experiment, the texture was band-pass noise with spatial frequencies distinct from that of shading patterns. Sakai hypothesised that facilitation might not have occurred had the texture been more low frequency such that the texture and the shading had similar Fourier spectra. The frequency dependency of LM & AM mixes as cues to shape-from-shading (Schofield et al., 2006) has not been tested. However, given that AM is closely related to the contrast modulated signals used to study second-order vision (Schofield & Georgeson, 1999), and that second-order mechanisms have a preference towards high frequency carriers (Dakin & Mareschal, 2000), it can be predicted that the reliability of such cues depends on carrier frequency. Here I extend the previous work of Schofield et al. (2006) to include more carrier frequencies. Doing so is also valuable because (a) the results may verify the Sakai's hypothesis that low frequency textures might not facilitate depth perception and (b) it would help to marry the literature on second-

order vision to recent shape-from-shading results giving a possible explanation as to why the human visual system is sensitive to second-order cues.

2.2 General methods

The method was similar to that of Schofield et al. (2006) except that binary noise textures were replaced with noises made of Gabor patterns. The dominant frequencies of the textures were varied to test the consistency of the role of AM in shape-from-shading in relation to carrier frequencies.

2.2.1 Stimuli

All images were composed from the following basic components:

First-order, luminance modulations (LM signal)

$$LM = nN(x,y) + l_a \cos(2\pi f(\cos\theta_a x - \sin\theta_a y) - \phi_a) + l_b \cos(2\pi f(\cos\theta_b x - \sin\theta_b y) - \phi_b), \quad (1)$$

Second-order, amplitude modulations (AM signal)

$$AM = nN(x,y) \times \left(m_c \cos(2\pi f(\cos\theta_c x - \sin\theta_c y) - \phi_c) + m_d \cos(2\pi f(\cos\theta_d x - \sin\theta_d y) - \phi_d)\right), \quad (2)$$

where f is the spatial frequency of the modulation, 0.5 c/d for all experiments in this chapter, l_a and l_b are the contrasts of LM component, m_c and m_d are the modulation depths of AM component. Having two LM and two AM terms means that each components can be presented as single oblique or cross-oriented plaid stimuli. AM modulation depths and LM contrast were made equal, as is the case when a corrugated uniform albedo texture surface is illuminated (Schofield et al., 2006), and fixed at 0.2. θ_a and θ_b are the orientations of LM obliques, ϕ_a and ϕ_b are their spatial phase, θ_c and θ_d are orientations of AM obliques, ϕ_c and ϕ_d are their spatial phase. $\theta_a = \theta_c = -45^\circ, \theta_b = \theta_d = 45^\circ$. Note that the AM component multiplies the noise texture (contrast modulates it) whereas the luminance component is added to it.

The *N(x,y)* term in the above equations represents Gabor noise texture carriers constructed in the following way:

1) Create two Gabor patterns using the formula below:

$$g(x,y)=\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\times\cos\left(2\pi f_G\left(x\cos\theta+y\sin\theta\right)+\phi\right)$$
$$\sigma\times f_G=\frac{1}{\pi}\times\sqrt{\frac{\ln 2}{2}}\times\frac{2^b+1}{2^b-1} \quad (3)$$

Note that the two Gabor patterns are orthogonal to each other (θs differ by 90°). ϕ was fixed to 0°. Two sets of Gabor orientations were used: $\pm 45^\circ, 0^\circ \& 90^\circ$. Along with two bandwidth values b (1.5 & 0.5 octaves), these parameters were introduced to control the masking power of the noise, see section 2.3.

2) Compute the Fourier transform of the image containing the two Gabor patterns

3) Randomize the phase spectrum of the Fourier image.

4) Compute the inverse Fourier transform.

The resulting stimuli represent uniform textured surfaces composed of randomly displaced Gabor patterns whose frequencies matched the dominant frequencies of the carrier. In practice, two frequencies were tested: high frequency textures based on 4.0 c/deg Gabors and low frequency textures based on 1.5 c/deg Gabors. These frequencies were chosen because significant variations in performance seemed to occur within that frequency range during the pilot study. See Figure 2.1 for demonstrations of texture carriers with these two dominant frequencies. Note that these Gabor textures were intended to represent reflectance or albedo textures not bumpy surfaces although it is possible to interpret them as the latter, see section2.6. for further discussion.

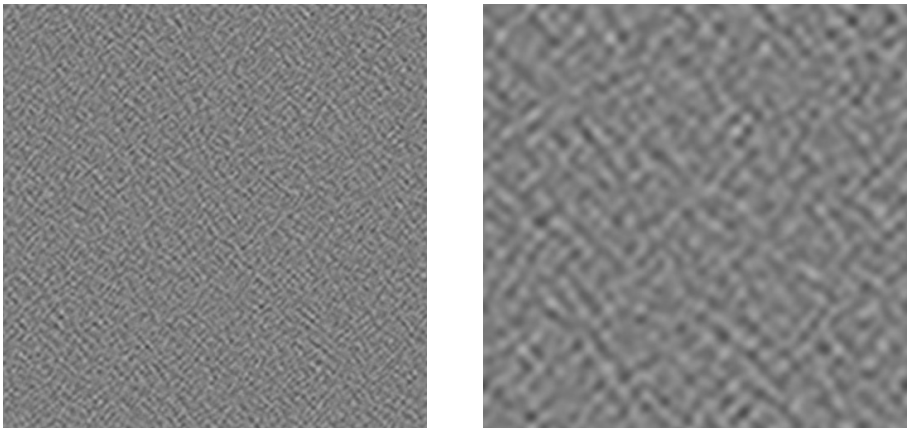
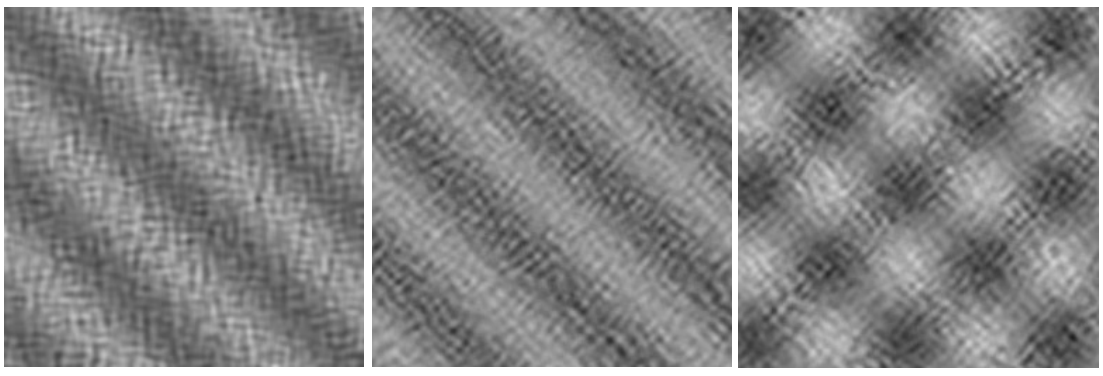


Figure 2.1 Carrier textures generated by randomizing phases of the Fourier coefficients of two Gabor patterns. Carrier dominant frequencies are determined by their corresponding Gabor spatial frequencies: 4.0 c/d (Left), 1.5 c/d (Right).

The components listed above were combined according to the formula below:

$$L(x,y) = L_0(1 + LM + AM) \quad (4)$$

where L_0 is the mean luminance of the monitor. The effect is to add noise, contrast modulated noise and luminance modulations together. LM and AM can be applied in phase to create a LM+AM component (that is, LM and AM are positively correlated) or they can be applied out-of-phase to create a LM-AM component (that is, LM and AM are negatively correlated). Both components can be presented alone or they can form a plaid. Figure 2.2 gives examples stimuli for LM+AM, LM-AM component presented alone and a plaid configuration stimulus.



(a) (b) (c)

Figure 2.2 example stimuli for LM+AM (a), LM-AM (b) and the mix of the two combinations forming a plaid (c). Images are showing only a few cycles of the original stimuli for demonstration purpose.

2.2.2 Equipment and calibration

Stimuli were generated using VSG2/5 graphics card (Cambridge Research System, CRS Ltd, UK) and presented on a 21" Sony Flexscan GDM –F520 CRT monitor. Responses were made via a CRS-CB3 response box connected to the VSG. Images measured 13.312 by 13.312 degrees of arc (512 by 512 pixels) displayed inside a central window. Outside of the central window the display was set to mean luminance to the limits of the monitor. Viewing distance was 1 m, in a darkened room where the experimental monitor was the only significant light source.

The calibration was based on the four parameter CRT model proposed by Brainard, Pelli and Robson (2002)

$$\frac{L-k}{L_{\max}-k}=\left(\frac{j-j_0}{j_{\max}-j_0}\right)^{\gamma} \qquad (6)$$

where L is the luminance output of the monitor, j is the output or entries of the look-up table (LUT), $L_{\max}, k, j_0, \gamma$ are parameters to be fitted. A set of luminance values was first measured from the monitor screen using a linear LUT and a CRS Colour Cal Luminance meter, for a range of j s including 0 and $j_{\max}$. These values were used to estimate the four parameters and a new LUT generated. The process of calibration and parameter estimation was carried out with an in-house software.

2.3 Control for masking

Masking is the (normally inhibitory) affect of one stimulus on the detection of another where the stimuli are coincident in space and simultaneous in time (Legge & Foley 1980). According to Harmon and Julesz (1973), noise frequencies that are adjacent to or overlapped with the picture spectrum, suppresses the detection of the target feature.

When put into the context of the current study, the texture carrier used in AM may mask the detection of luminance signal thus inhibiting shape-from-shading via an uninteresting route. The problem is illustrated in Figure 2.3, which shows the Fourier spectra of a 0.5 c/deg sine wave and examples of our two texture elements. Therefore, masking power was controlled for by varying the orientation and spatial frequency bandwidth of the textures: textures with their dominant orientations tilted away from that of the luminance modulation should mask it less as channels are known to be orientation sensitive (Campbell & Kulikowski 1966). I varied carrier orientation as follows: 'in-line' textures were made from Gabors with orientations +-45° to match the modulation, whereas the Gabors in the 'out-of-line' textures were oriented at 0 and 90°. Similarly, reducing the spatial frequency bandwidth of the textures should reduce the spectral overlap between signal and texture thus mitigating the effects of masking. More specifically, textures with bandwidth of 0.5 octaves should have less masking power than that with bandwidth of 1.5 octaves.

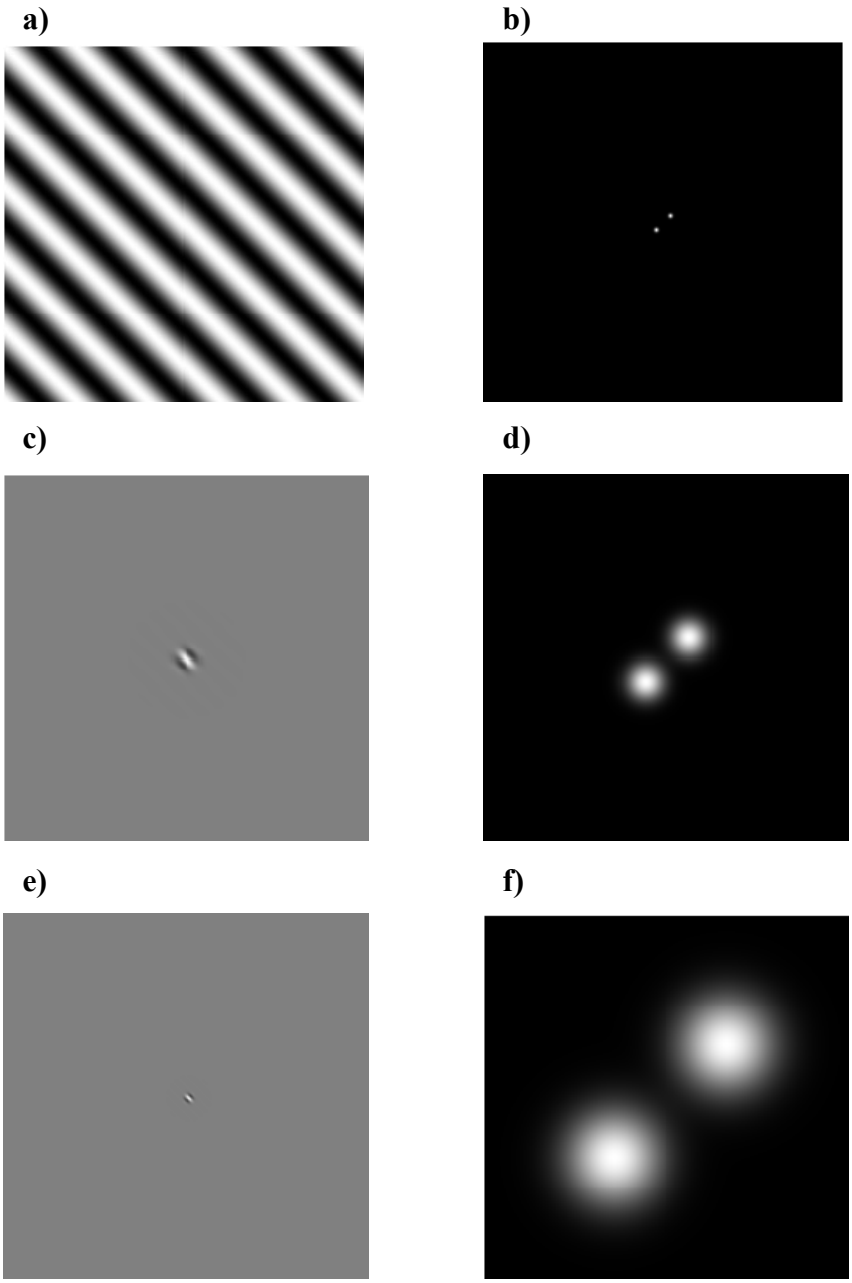


Figure 2.3 Demonstration of masking problems: a) example of sinusoidal luminance signal with spatial frequency of 0.5 c/d. b) spectrum of a), note that the two dots were slightly enlarged only for demonstration purposes. C) Gabor pattern with spatial frequency of 1.5 c/d and bandwidth of 1.5 octaves. d) spectrum of c), note that d) has a high risks of overlapping b). e) Gabor pattern with spatial frequency of 4.0 c/d and bandwidth of 1.5 octaves. f) spectrum of e), which has comparably small risks of overlapping b). Thus masking alone could affect human performances for the two testing frequencies.

2.4 Experiment 1: single oblique

2.4.1 Procedure

The procedure was also similar to that used by Schofield et al. (2006). Observers viewed single oblique images and indicated which of two marked positions appeared closer to them (e.g. Figure 2.4; Marks were coloured in red or blue in practice but are shown as black and white on the figure). The effective distance (the phase difference within a cycle) between marked positions was 1/18th of a period (shown by black and white crosses in Figure 2.4, which were not shown in experiments) along one or other orientation (called the test diagonal). In practice, the distance between markers was increased by a (random) integer number of periods along both orientations in order to encourage global processing.

Only one diagonal was tested in each trial. That is, the *effective* distance between markers took non-zero values in one direction while being fixed at zero along the orthogonal direction. One combination of LM and AM was presented alone (single oblique) on one diagonal while no modulation was present in the orthogonal direction. Only the modulated direction was tested. The absolute phase of each oblique was chosen at random. Then the markers were placed according to the following:

1) First a reference location was given by the absolute phase of the oblique.
2) The phase of each diagonal was added by an offset (phase of the test position) along the diagonal to get the nominal test location. Offsets were a set of 8 possible distances at 1/8th of a cycle intervals relative to the reference point. Due to the periodic nature of the modulation, only 8 test locations were required to span a full cycle of modulation. The 0 and 1 whole cycle offsets

were represented by the same nominal test position. Offsets were chosen separately for each diagonal.

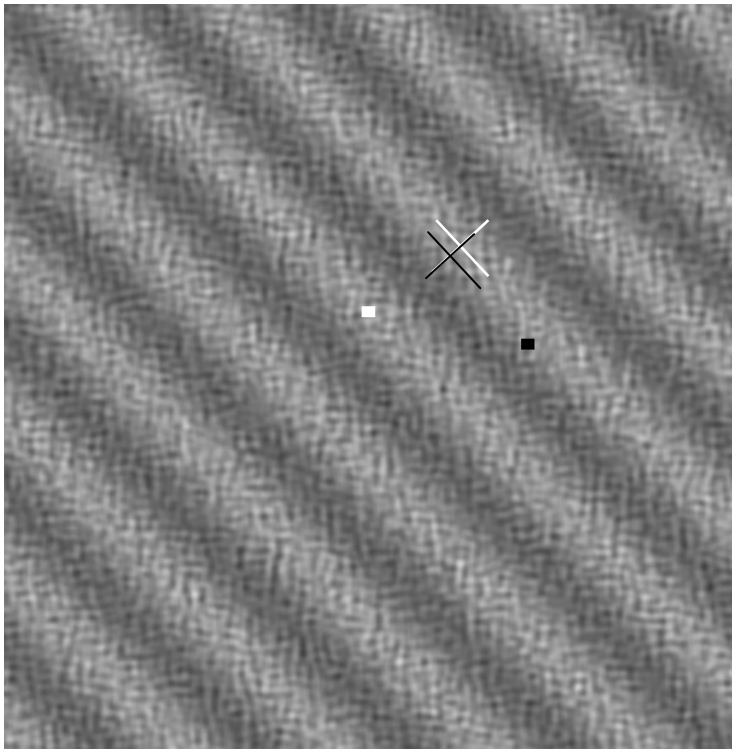
3) Nominal marker positions were chosen to be 1/36th of a cycle on each side of the nominal test location along the test diagonal. Along the non-test diagonal there was no displacement between the two marker positions.
4) A further displacement of a random integer multiple of a cycle was added to both marker positions along both diagonals, to enforce a depth comparison at a more global scale.
5) Finally, marker locations were rounded to the nearest pixel.

In addition, all positions and offsets were measured diagonally working from top-left to bottom-right or top-right to bottom-left depending on the diagonal under test. Masking was controlled for by applying the techniques described in section 2.3. Overall, there were

$$8\,(\text{positions}) \times 2\,(\text{modulation orientations}) \times 2\,(\text{phase combinations}) \times 2\,(\text{orientations of Gabor patterns}) + 8\,(\text{positions}) \times 2\,(\text{modulation orientations}) \times 2\,(\text{phase combinations}) \times 2\,(\text{bandwidths of Gabor patterns}) = 128$$

trials per session and participants completed 8 sessions each.

a)



b)

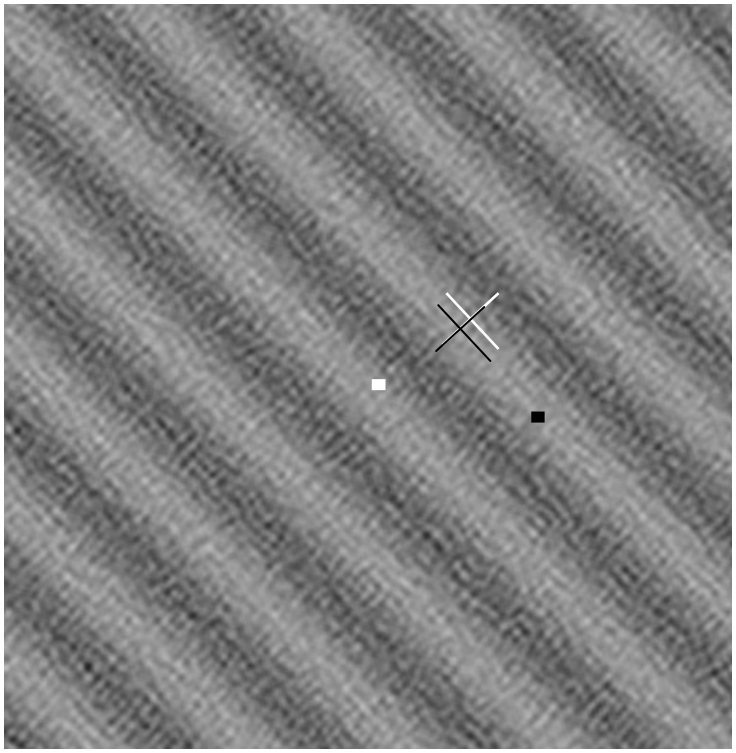


Figure 2.4 Example single oblique stimuli: a) LM+AM alone. The test diagonal is from top right to bottom left (modulation diagonal). The white and black cross are shown to aid understanding the underlying offset between two marker positions but were not shown on the experiment stimuli. b) LM-AM alone. The test diagonal is again the modulation diagonal from top right to bottom left. The apparent effective distance between the two markers made here are for demonstration only. They are not representing the true distance values made in the experiment.

In total 3 observers took part in this experiment, all being naïve to the purpose of the experiment. All had normal or corrected to normal vision. Observers were asked to press an appropriately coloured key on a button box in response to which of the marked locations they thought appear closer to them in depth. Each condition was tested equally often in random order. Each individual undertook a short training session containing 50 random trials prior to testing. There was no restriction on viewing time although observers were encouraged to give their best guess 'without thinking too much'. No feedback was given.

2.4.3 Analysis

Recalling that all positions and offsets were measured working from top to bottom, the marker located lower down the screen (before the application of the integer wavelength displacement) was regarded as the positively shifted marker. A positively shifted marker seen as closer in depth indicates a positive value in gradient and was scored +1. Likewise, -1 was scored when a negatively shifted marker was seen closer. Average scores served as a metric for the perceived surface gradient for each test location. Observers may have been biased towards pressing one key more often than the other. Such biases would produce a non-zero DC gradient and were removed by taking the Fourier transform of each gradient profile and setting its DC component to zero. After applying the inverse Fourier transform, the resulting gradients were integrated to recover the perceived surface shape. The amplitude of the fundamental component for each recovered depth profile was recorded as a measure of the strength of the shape-from-shading percept. Phase shifts of the fundamental (relative to a cosine) were also recorded for further analysis. A minus 90° phase shift means that the fitted cosine function perfectly coincides with the underlying sinusoidal luminance.

2.4.4 Results

Results for single oblique are shown in Figures 2.5, 2.6 and Tables 2.1 and 2.2. Figure 2.5 gives some example traces for one participant. Thick solid lines indicate the underlying sinusoidal luminance modulation. Dots represent the perceived depth at each test location. Traces are grouped in two rows with the top row being for LM+AM and bottom row being for LM-AM respectively. In the top 8 panels, traces are divided into two columns of which the left associates with inline Gabor texture (more masking power) and the right associates with out-of-line Gabor texture (less masking power). In the bottom 8 panels, traces are divided into two columns of which the left associates with Gabor texture of narrower bandwidth (less masking power) and the right associates with Gabor texture of broader bandwidth (more masking power). Figure 2.6 shows mean depth amplitudes for both carrier frequencies and orientations. The left most bar of each frequency group represents the perceived depth for any particular combination when Gabor texture orientations are in-line with orientations of luminance signal and Gabor bandwidth is relatively large, thus producing more masking effects. The middle bars correspond to perceived depth when Gabor texture orientations are out-of-line with the luminance signal. The right most bars correspond to Gabor textures with relatively small bandwidth (0.5 compared to 1.5). In both cases, carrier textures should produce less masking power. Tables 2.1 and 2.2 give details of depth amplitudes and phases for each individual observer in response to single oblique component under all conditions. Titles in the first column indicate the test cue and its orientation, as well as the underlying carrier frequency and masking condition. Although there were individual differences between absolute values of observers' perceived amplitudes, the drop in amplitudes seemed to be consistent.

In the case of higher carrier frequency, the results are consistent with that of Schofield et al., (2006). Briefly, observers interpreted corrugated surface from the sinusoidal luminance signal. The phase information in table 2.1 shows that perceived surface peaks tend to be below luminance peaks, indicating the operation of the lighting from above assumption. However the perception of shape-from-shading deteriorated when the carrier frequency was 1.5 c/d. Depth amplitude went down significantly. It's also noted that reducing masking power had very little effect, even in the low frequency condition. That is, regardless of the changes in masking power, the impression of shape-from-shading was considerably reduced on lower frequency carriers compared to high-frequency ones. Moreover, the inter-observer variability in phase was high for the lower carrier frequency. For example, Observer WXG's phase estimates at lower frequencies have a standard deviation of 36.4 deg while those at higher frequencies have a standard deviation of 15.8 deg, this further confirms the degradation of depth perception; people are less sure where the peaks lie. There is no significant difference between the LM+AM and LM-AM data, although the perceived depth amplitude for LM-AM was slightly lower and the phase for LM-AM contained larger inter-observer variability (consistent with Schofield et al., 2006).

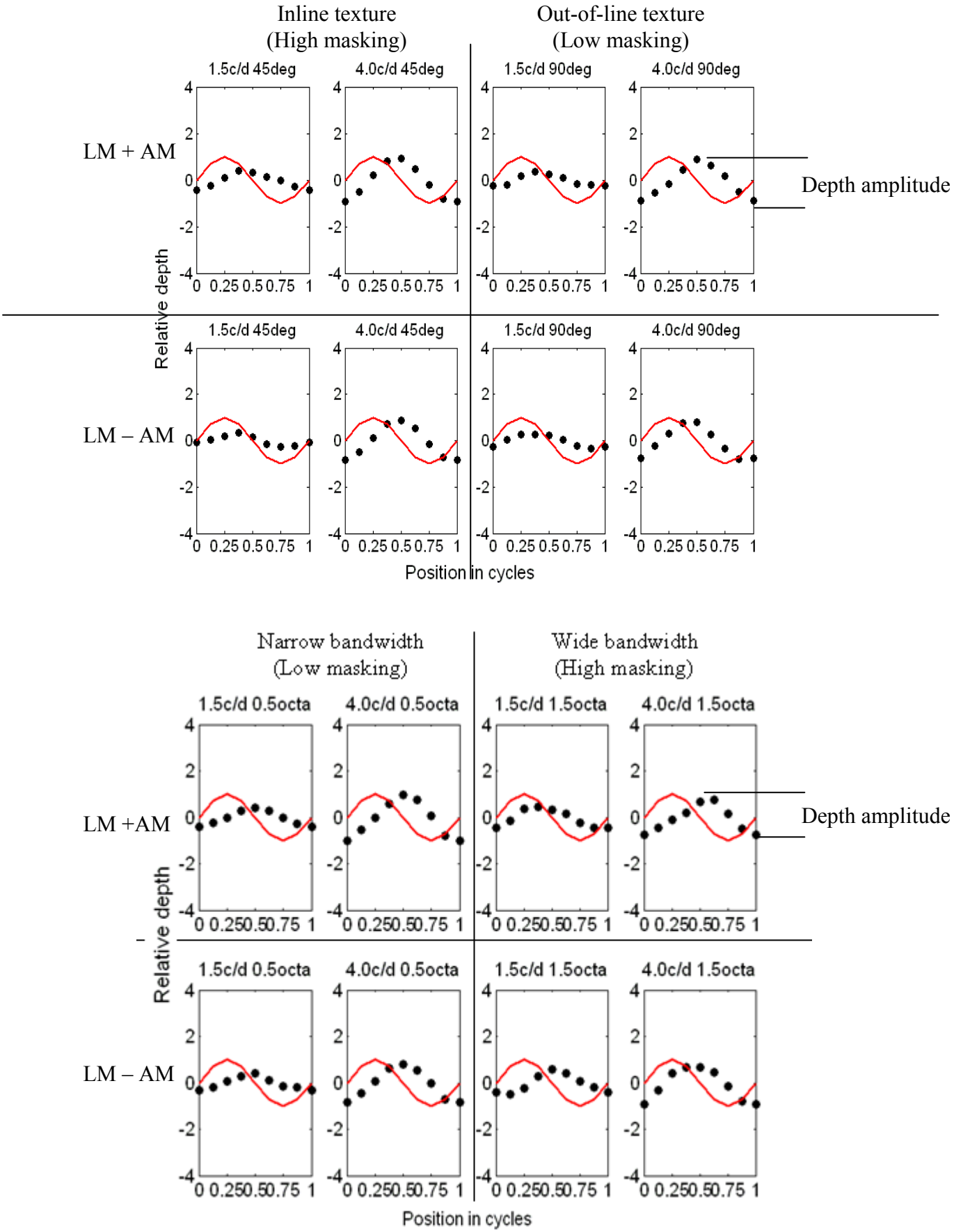
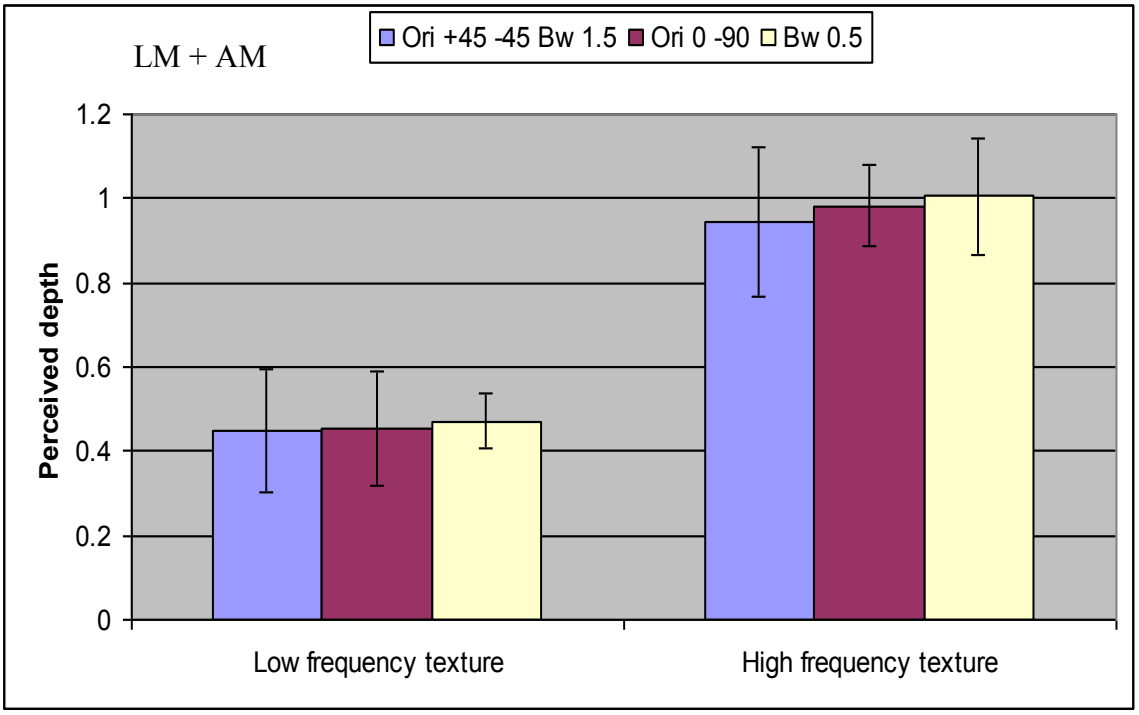


Figure 2.5 Recovered depth traces under two masking conditions: change in orientation (top half) and change in bandwidth (bottom half). Thick solid lines indicate underlying luminance

modulations. Traces are divided into 4 slots and each contains two traces for two carrier frequencies to compare one against the other.



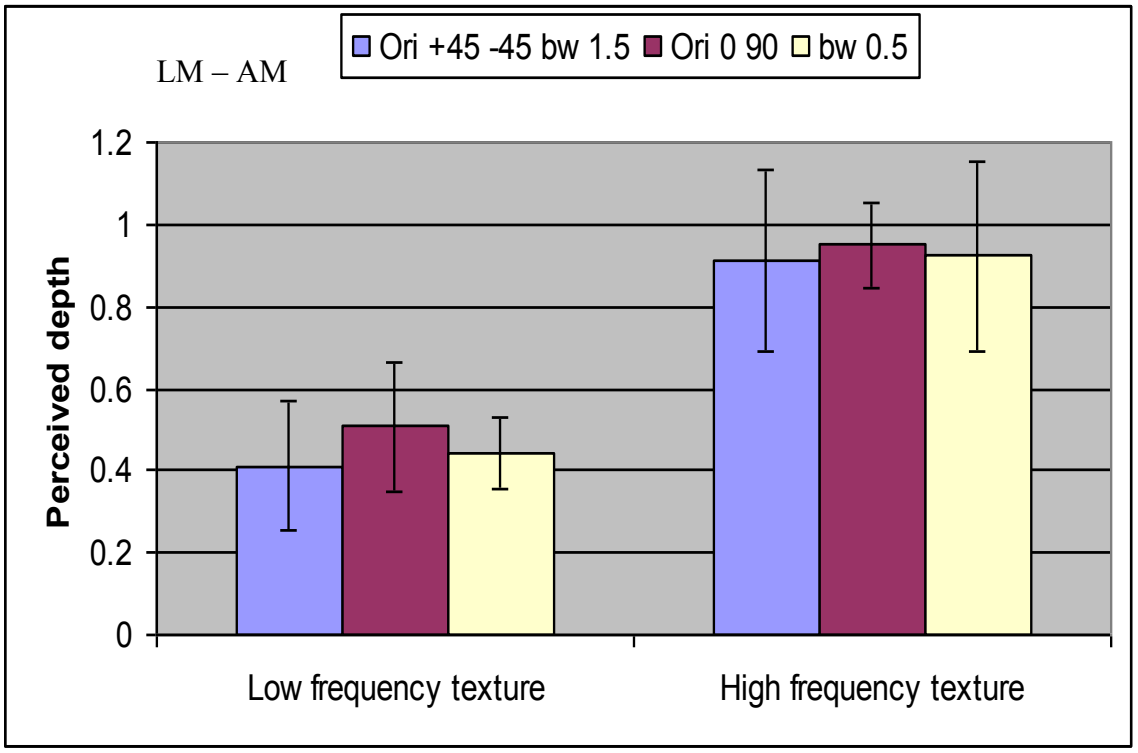


Figure 2.6 Averaged depth amplitudes for high and low carrier frequencies textures, under three masking conditions. The left most bar in each cluster corresponds to the texture with most masking power.

Testing Conditions		JCY	WYK	WXG
LM+AM Left Low frequency High masking	Amplitude	0.429	0.833	0.221
	phase	-76.9	-152.3	-157.5
LM+AM Left High frequency High masking	Amplitude	0.844	0.806	1.238
	Phase	-168.2	-193.4	-154.3
LM+AM Left Low frequency Out-of-line texture	Amplitude	0.241	0.442	0.649
	phase	-122	-129.3	-82.5
LM+AM Left High frequency Out-of-line texture	Amplitude	0.583	1.239	0.989
	Phase	-167.2	-154.9	-130.6
LM+AM Left Low frequency Narrow bandwidth	Amplitude	0.465	0.259	0.415
	Phase	-141.6	-141.9	-169.5
LM+AM Left High frequency Narrow bandwidth	Amplitude	1.103	0.854	1.150
	Phase	-172.1	-126.3	-139.9
LM+AM Right Low frequency High masking	Amplitude	0.361	0.418	0.326
	Phase	-169.4	-93	-172.6
LM+AM Right High frequency High masking	Amplitude	0.886	1.069	1.159
	Phase	-173.1	-160.9	-171.4
LM+AM Right Low frequency Out-of-line texture	Amplitude	0.434	0.292	0.615
	Phase	-142.1	-196.8	-175.0
LM+AM Right High frequency Out-of-line texture	Amplitude	1.096	0.867	0.998
	Phase	-149.3	-182.1	-168.2
LM+AM Right Low frequency Narrow bandwidth	Amplitude	0.325	0.621	0.675
	Phase	-203.4	-104.9	-126.9
LM+AM Right High frequency Narrow bandwidth	Amplitude	0.537	1.137	1.170
	Phase	-181.0	-156	-152.7

Table 2.1 Properties of perceived surfaces inferred from LM + AM single oblique experiment. The testing conditions are listed in the head for each row. Values are given for the amplitude and phase of the fundamental component for individual depth profiles.

Testing Conditions		JCY	WYK	WXG
LM – AM Left Low frequency High masking	Amplitude	0.569	0.279	0.452
	phase	-135	-338	-194
LM – AM Left High frequency High masking	Amplitude	0.973	1.152	1.245
	Phase	-142	-170	-150
LM – AM Left Low frequency Out-of-line texture	Amplitude	0.281	0.285	0.827
	phase	-184.5	-191	-132
LM – AM Left High frequency Out-of-line texture	Amplitude	0.877	0.83	1.061
	Phase	-196	-156.3	-151
LM – AM Left Low frequency Narrow bandwidth	Amplitude	0.407	0.39	0.817
	Phase	-153.4	-7.86	-151.8
LM – AM Left High frequency Narrow bandwidth	Amplitude	0.754	0.518	1.103
	Phase	-165.01	-153	-142.9
LM – AM Right Low frequency High masking	Amplitude	0.431	0.169	0.568
	Phase	-161.3	-161.6	-160.9
LM – AM Right High frequency High masking	Amplitude	0.773	1.027	1.034
	Phase	-200	-159.8	-145.2
LM – AM Right Low frequency Out-of-line texture	Amplitude	0.409	0.72	0.51
	Phase	-133.2	-136.8	-165.3
LM – AM Right High frequency Out-of-line texture	Amplitude	0.782	1.174	0.974
	Phase	-180.7	-185.9	-163.3
LM – AM Right Low frequency Narrow bandwidth	Amplitude	0.447	0.342	0.258
	Phase	-201	-290.5	-278.3
LM – AM Right High frequency Narrow bandwidth	Amplitude	1.238	0.775	1.126
	Phase	-189.5	-168.8	-171.8

Table 2.2 Properties of perceived surfaces inferred from LM – AM single oblique experiment. Details are as for table 2.1

2.4.5 Discussion

Lower frequency textures suppressed shape-from-shading. The dominant orientation and spatial-frequency bandwidth of the textures were varied so as to reduce their ability to mask the shading pattern. But neither manipulation had any effect. Thus simple masking did not seem to account for the decline in depth percept. There maybe two other factors contributing to this suppression:

(a) As described in chapter 1, AM helps the human visual system to isolate shading signal and hence improve shape-from-shading impression. AM represents second-order information and so may require second-order mechanism for detection. Such mechanism has been described by many using an F-R-F model (Wilson, Ferrera & Yo, 1992; Kingdom, Prins & Hayes, 2003). If second-order mechanisms are more sensitive to high frequency carriers, then AM was most likely detected less well and therefore the shading signal was less well isolated. (b) Alternatively, low frequency texture elements themselves could look like shading/shadows, which have interfered with the probe tasks: adding noise to individual judgements reduced the amplitude of the interpolated depth profile. In this case, large-scale undulations in the surface might still be observed but judgement of relative depth between two fine locations might be interrupted by small-scale undulations produced by the texture. However at this stage, it is not possible to ensure the action of either of the two, hence it is difficult to assess the role of second-order processing. Experiment 2 attempts to investigate this.

2.5 Experiment 2: plaid configuration

One reason that results from previous section are inconclusive is that the role of AM in shape-from-shading is less obvious for single oblique stimuli than it is for plaids (Schofield et al., 2006). When LM+AM and LM-AM are presented together in a plaid the latter cue looks flat despite the strong luminance signal. The procedure described above was then applied to the plaid configuration with the prediction that the influence of AM on plaid stimuli could be affected by changes in carrier frequency.

2.5.1 Procedure

For the plaid experiment stimuli consisted of a LM+AM signal presented on one oblique and a LM-AM signal presented on the other oblique, either cue could be

placed under test. As with the single oblique experiment, only one diagonal was tested on each trial. There was no phase offset between markers on the non-test diagonal. For example in Figure 2.7, the effective displacement of markers (offset between white and black crosses) is in the bottom left to top right direction. Hence, it is the LM+AM grating whose depth is being tested. No control against masking was included for the plaid experiment since it had already been shown that masking was unlikely to be one of the major causes of suppressed depth perception in these experiments. In fact, result for the plaid configuration is a further weight to the argument that masking is not an issue. All other experiment settings were the same as the previous experiment. Orientations of Gabor patterns were ±45° and had the bandwidth of 1.5 octaves. Overall there were

$$8\,(\text{positions}) \times 2\,(\text{modulation orientations}) \times 2\,(\text{phase combinations under test}) = 32$$

trials in each session and each observer completed 8 trials all together.

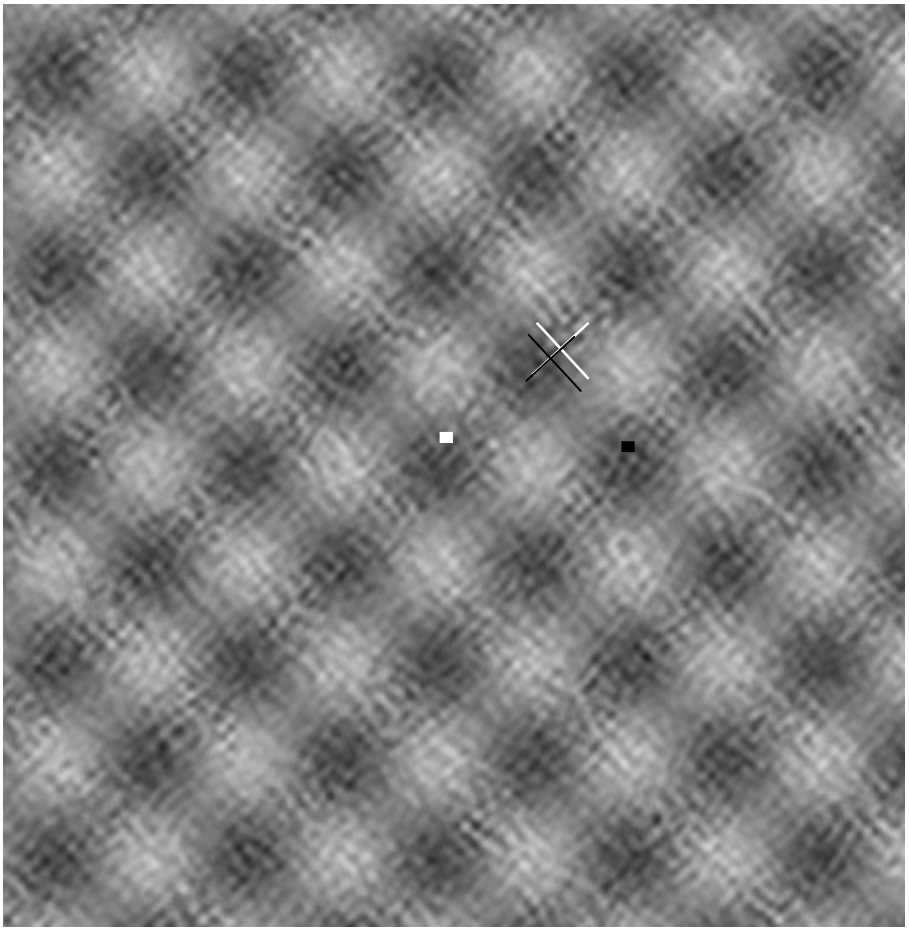


Figure 2.7 Example stimuli for plaid configuration: LM+AM on the right oblique and LM – AM on the left oblique. The test diagonal is from top right to bottom left thus LM+AM is being tested. The white and black cross are shown to aid understanding the underlying offset between two marker positions but were not shown on the experiment stimuli.

Results are shown in a similar format to those of section 2.4. Recovered depth profiles for one observer are shown in Figure 2.8. Thick solid lines indicate underlying luminance modulations. Each dot represents a recovered depth relative to 0 at each test location. Depth profiles for the two combinations are grouped into two columns. The amplitude of the fundamental component was recorded as a measure of depth amplitude. Figure 2.9 shows mean depth amplitude calculated across all observers. Within each frequency group, the left and right bars correspond to conditions where out-of-phase and in-phase combinations were under test respectively.

2.5.2 Result

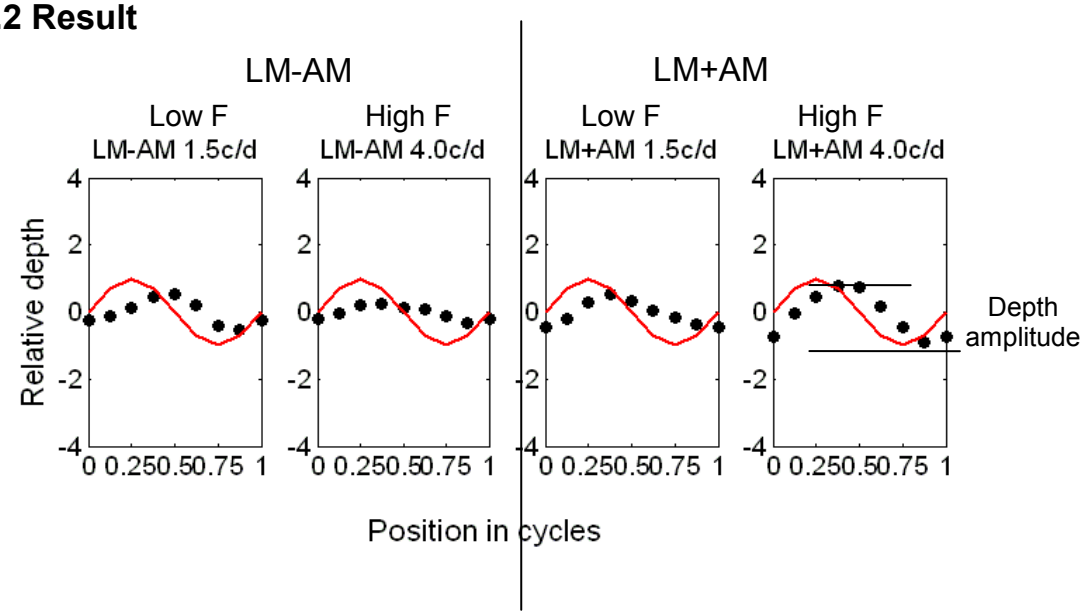


Figure2.8 Example of perceived depth profile when LM+AM and LM – AM are presented on a plaid. Thick solid lines indicate underlying luminance modulations. Slots are divided into two columns, with the left and right columns show perceived depth profiles when LM – AM and LM+AM were under test respectively.

When tested against each other LM+AM had a much higher perceived depth amplitude than LM-AM, for high frequency carriers. Similar to what was found for single oblique stimuli, depth profiles for LM+AM on a plaid peaked below the luminance peak and was very stable across observers. It is worth pointing out that not only did LM-AM have much lower perceived amplitude than LM+AM on higher frequencies, but the position of the perceived peaks also varied considerably between observers. This is further evidence that LM-AM was perceived to be less corrugated than LM+AM. However the perceived depth amplitudes are higher than those obtained by Schofield et al. (2006).

In contrast to the above result, when the modulations were carried by low frequency textures perceived amplitude for LM+AM dropped, although depth profiles still peaked below the luminance peak, much as for the single oblique case. Its counterpart LM-AM signal produced a similar result. That is, the perception of shape-from-

shading from LM+AM was reduced to be more like that for LM-AM when the texture frequency was reduced.

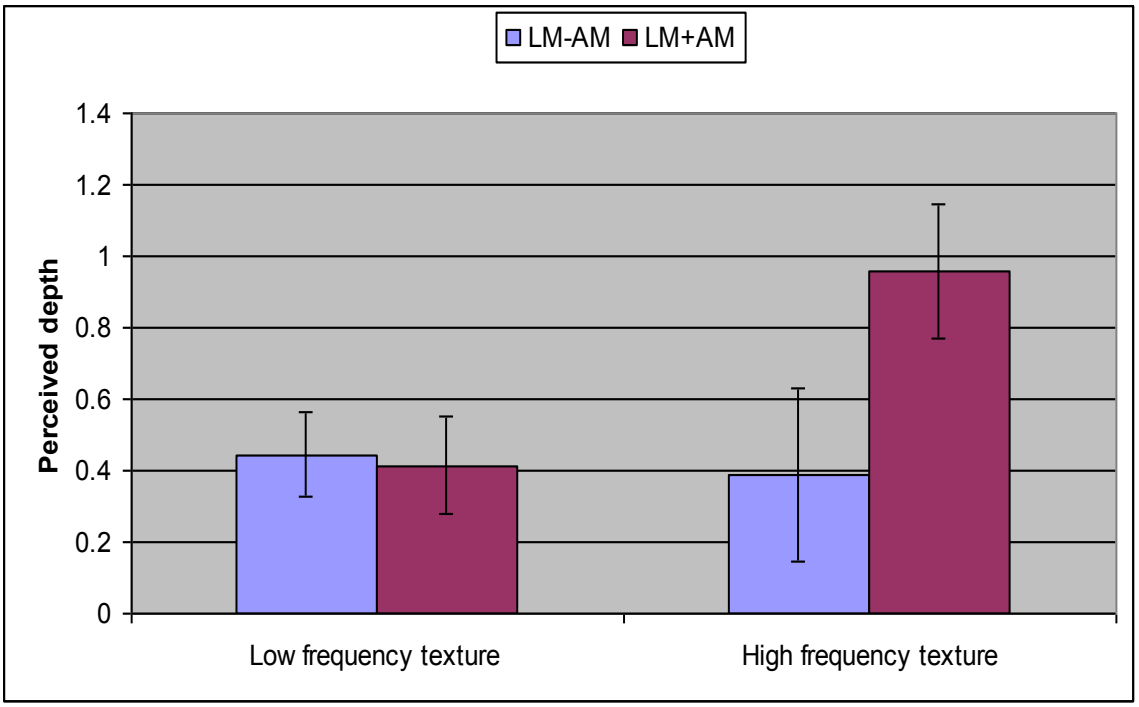


Figure 2.9 Averaged depth amplitudes for low and high frequency carriers.

LM + AM Left LM – AM Right Low frequency	Amplitude	0.335	0.477	0.315
	phase	-165.4	-64.6	-144.7
LM – AM Left LM + AM Right Low frequency	Amplitude	0.44	0.517	0.496
	Phase	-166	-139.9	-167.6
LM + AM Left LM – AM Right High frequency	Amplitude	0.849	0.702	0.917
	phase	-169.76	-156.5	-128.3
LM – AM Left LM + AM Right High frequency	Amplitude	0.89	0.254	0.367
	Phase	-183.4	-102.1	-96.9
LM + AM Right LM – AM Left Low frequency	Amplitude	0.085	0.567	0.576
	Phase	-186.2	-105.4	-153.4
LM – AM Right LM + AM Left Low frequency	Amplitude	0.683	0.308	0.525
	Phase	-122.2	-69.7	-132.6
LM + AM Right LM – AM Left High frequency	Amplitude	0.88	0.969	0.86
	Phase	-171.6	-156.5	-161
LM – AM Right LM + AM Left High frequency	Amplitude	0.569	0.359	0.359
	Phase	-180.9	-188.5	-188.5

Table 2.3 Properties of perceived surfaces inferred from plaid experiment. The testing conditions are listed in the head for each row. The first line of each head indicates the component under test. e.g. in the first condition row, in-phase combination was under test. The phase values represent the phase shifts of fit cosine functions. Thus a minus 90° phase shift means that the fit cosine function perfectly coincides with the underlying sinusoidal luminance.

2.5.3 Discussion

In the plaid configuration, a strong shape-from-shading percept was found for LM+AM signals when the carrier frequency was high (Fig2.9). In contrast, LM-AM was seen as much less corrugated in this condition, though not as flat as what was found by Schofield et al. (2006). For example, in a similar plaid configuration, LM-AM produced an even weaker depth percept as suggested by an even lower fit amplitude (average 0.1) obtained by the same method of analysis (Schofield et al., 2006). So it can be argued that observers still gained considerable depth perception in the LM-AM direction during the experiment presented here. However LM+AM and LM-AM produced similar perceived depth profiles for low frequency texture carriers, suggesting that the distinction between these two signals was weakened in this case.

Meanwhile perceived depth for LM-AM was not further reduced as the carrier frequency decreased. Considering that the perceived depth for LM-AM can be even flatter as found elsewhere (Schofield et al., 2006), any masking effect would have simply reduced perceived depth amplitude and position stability for LM+AM while further weakening any depth precept gained from LM-AM as well. The same is true for any influence due to the fact that low frequencies textures can look like shading in their own right; shape-from-shading should be disrupted for both LM-AM and LM+AM not just LM+AM. Neither straight forward masking nor interference from apparent undulations in the texture can account for the reduction in perceived depth for LM+AM in the absence of a reduction for LM-AM, such that the two cues become indistinguishable. It can be argued that the information that makes them distinct is conveyed less well by low frequency carriers. As a second-order cue, AM requires a high frequency carrier for good detection (Dakin & Mareschal 2000). For low frequency carriers, AM may not have been detected well enough to help the HVS to distinguish the two signals. Thus both cue types were perceived as weakly corrugated.

2.6 General discussion

Together, the results from the single oblique and plaid experiments suggest that changing carrier frequency may affect shape-from-shading in human observers. In general, textures whose frequencies are below a certain level would give less support to shape-from-shading. Masking did not seem to account for this suppression. The degree of suppression was not reduced when the masking power of low frequency textures was reduced. Hence this suppression was probably carried out via one of two alternative routes: a), support to shape-from-shading that would normally arise from underlying texture is weakened; b), as the texture frequencies go down, the underlying

texture becomes more like shading/shadow, thus interfering with the global depth percept. Experiment 1 suggests that at least one of the above possibilities is true. Results from the plaid experiment suggest that a) dominates: the distinction between the percept for LM+AM and LM-AM vanished for low frequency carriers (this can be concluded from the similarity between their depth profiles). Although the decrease in depth amplitude for LM + AM on low frequency carriers was most likely due to b), b) alone is not sufficient to explain the absence of a reduction in perceived depth for LM – AM on low frequency carriers. Thus, a) must have been a factor also. The findings confirm the hypothesis by Saikai (2006) that low frequency textures do not facilitate depth perception. Instead, they have a negative impact on the perception of shape-from-shading. As a second-order entity, AM is conveyed less well by low frequency carriers, which is consistent with the idea that second-order vision is most sensitive to high frequency carriers (Sutter et al.,1995; Dakin and Mareschal, 2000; Mareschal and Baker, 1999; Zhan and Baker, 2008).

3. The frequency dependency of AM cue in shape-from-shading

The experiments reported in chapter 2 tested perceived depth amplitude for LM+AM and LM-AM mixes based on just two carrier frequencies. The experiments described in this chapter tested the same mixes against a larger range of carrier frequencies. The relationship between LM and AM signals seems to determine perceived depth in the stimulus. Presumably AM must be detected if it is to have any influence on shape-from-shading. If the AM component is detected by a second-order mechanism we should expect the influence of the AM signal to follow the known characteristics of second-order vision. Specifically in cases where the carrier signal is not able to act as an effective carrier for AM signals we should expect LM+AM and LM-AM cues to produce similar depth percept – because the AM cue will be ineffective in such cases. The results presented in this chapter show that this is the case.

3.1 Introduction

Results of chapter 2 showed qualitatively how shape-from-shading may be affected by carrier frequencies. The choice of the two frequencies was somewhat arbitrary, and it is not clear what carrier frequency should be considered as the division between 'high' and 'low'. In the experiments of this chapter, more carrier frequencies were tested in order to characterise more fully the influence of carrier frequency on shape-perception.

The strength of an AM signal determines the perceptual difference between LM+AM and LM-AM in a shape-from-shading task when the two are presented simultaneously

(Schofield et al., in press). AM is a second-order entity, closely related to contrast modulation. Hence the detection of AM should be dependent on the carrier frequency (Sutter et al, 1995; Dakin and Mareschal, 2000). If the visibility of the AM signal varies with carrier frequency then its effect on shape-from-shading should also vary. Testing shape-from-shading in LM/AM mixes using a wider range of carrier frequencies is not only interesting in terms of the shape-from-shading task itself; the result will also further expose the characteristics of the human second-order mechanism.

There is some debate as to whether second-order signals are processed at all when conveyed by low frequency carriers. The disagreement arises from the argument that a second-order signal such as abutting line gratings could potentially activate conventional linear receptive fields thus would not require a non-linear detection mechanism (Skottun, 1994). For abutting line gratings stimuli, low frequency carriers have more visible luminance contrast and produce stronger luminance edges at the terminations of lines, which could serve to detect the modulation gratings (Song & Baker, 2006). Verification of this hypothesis came from the physiological study by Song and Baker (2006) which reported that a large population of cells in cat area 18 responded bi-modally to abutting line gratings with one peak at low frequency carriers and the other at high frequency carriers. Responses to stimuli based on low frequency carriers varied with carrier phase, indicating that these cells were in fact responding to the luminance edges rather than second-order modulations. Although the detection of second-order signal was not discounted completely in this study, the involvement of a non-linear mechanism was not obvious in this context. On the other hand, Dakin and Mareschal (2000) believed that the detection of Gabor modulations conveyed by low

frequency noise combined the detection of both first-order artefacts and genuine second-order cues. These first-order artefacts were also called side-band effects. Although side-band effects were controlled in their experiment and little effect was found for high frequency carriers, the role of first-order luminance artefact could not be entirely discounted. An example of how first-order luminance feature can lead to detection of modulation is illustrated in Figure 3.1.

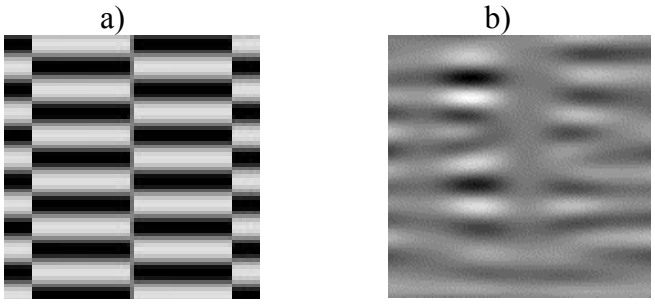


Figure 3.1 Illustrations how first order luminance defined features may lead to modulation detection. a) abutting line gratings with low frequency carrier, image taken from a sample stimuli in Song and Baker's study (2006). Edges at the terminations of lines could serve to detect the vertical modulation. b) Low pass horizontal noise contrast modulated by a Gabor pattern. The image is taken from Dakin and Mareschal (2000). Luminance defined edges are visible in b).

The experiments discussed in following sections address the question of whether the effect of AM on depth perception varies in accordance with any reported carrier frequency dependency in second-order vision. In addition, unlike in detection tasks where existence of luminance defined edges could well lead to an observer's decision, the perceptual difference of LM+AM and LM-AM in a probe-task is unlikely to be triggered by local luminance defined edges, because local edges would not boost or suppress a global impression of shape-from-shading. Hence, observers' performance in this task is an alternative verification of the existence of the processing of AM.

3.2 General method

Methods were same to those of described in chapter 2 except that more texture frequencies were tested. The same two point depth comparison method was used.

Four naïve observers took part in this experiment. One of them had done the plaid configuration experiment described in chapter two. The remainder had no previous experience of this type of experiment.

3.2.1 Stimuli

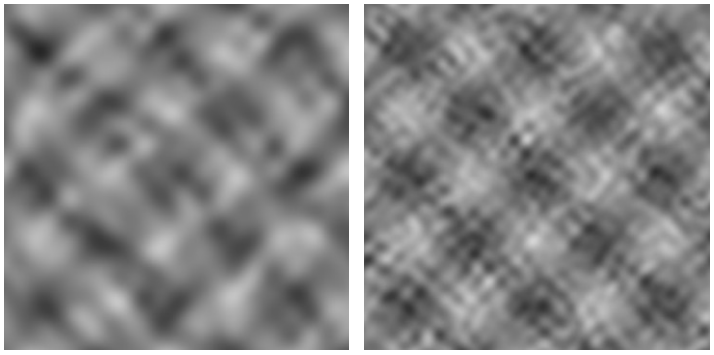
Images were made following the procedure outlined in chapter 2. Overall five carrier frequencies were tested: 1.0 c/d, 2.0 c/d, 4.0 c/d, 8.0 c/d and 16.0 c/d. Textures were made of ±45° Gabor elements as in chapter 2. It was not possible to test at higher frequencies due to the Nyquist sampling limit of the display system. Both modulation frequencies were fixed at 0.5 c/d.

3.3.2 Equipment and calibration

Monitors were calibrated using the same method as in chapter two. The viewing was changed to 2m to cater for a larger range of carrier frequencies.

3.4 Experiment 1 Plaid configuration

In this experiment, LM+AM and LM-AM mixtures were presented in a plaid. The procedure is same to that of plaid experiment in chapter two. Examples of stimuli are shown in Figure 3.2



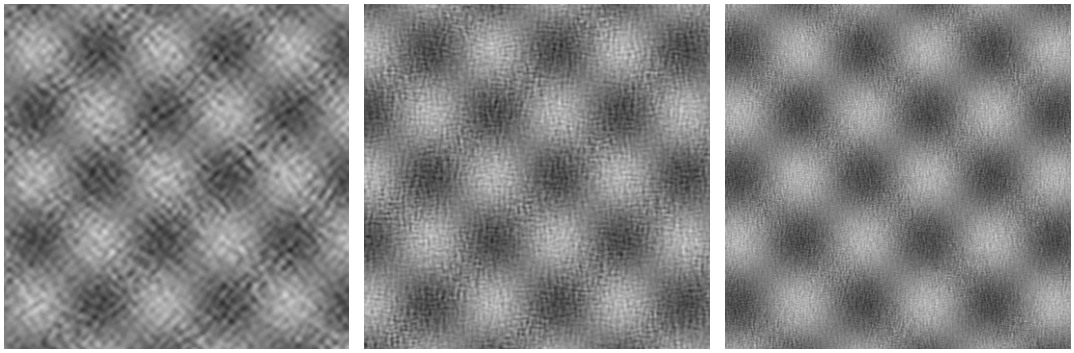


Figure 3.2 Stimuli used in plaid configuration experiment. Top row from left to the right are textures with dominant carrier frequencies of 1.0 c/d and 2.0 c/d respectively. Bottom row from left to right are textures with dominant carrier frequencies of 4.0 c/d, 8.0 c/d and 16.0 c/d respectively. These examples are draw to give the correct spatial frequencies at a 50cm viewing distance.

3.4.1 Results

As before perceived surface gradient was measured at each test location. After removing biases, the gradients were then reintegrated to produce a perceived surface shape. The amplitude of the fundamental component was recorded as a measure of depth amplitude. Figure 3.3 shows mean depth amplitude calculated across all observers calculated as the amplitude of the fundamental component of the reconstructed depth profile. The perceptual difference between LM+AM and LM-AM in a plaid configuration was measured by the difference in their perceived depth amplitudes, which was done separately for each participant. Since any masking effect should produce same reductions in perceived depth in the two phase relationships, taking the difference between the two should remove this uniform effect while retaining the influence of the AM cue. The mean difference across four participants is depicted in Figure 3.4. The distribution of perceived surface phase (position) across participants can also provide information about the reliability of perceived depth. A broad phase distribution together with low mean depth amplitude is a signature of a flat perceived surface. Phase can thus be combined with fundamental amplitude to produce a more reliable single measure of the perceived surface shape. A simple way

to achieve this is to add all fundamental sine wave functions together and divide the resulting sine wave function by the number of participants. Surface profiles that vary in phase will tend to cancel one another reducing the amplitude of the combined trace. If a surface is perceived flat, the phase of its fit sinusoidal function doesn't reveal anything meaningful but is evenly distributed among the entire phase range. On the other hand if a surface appears corrugated, although observers may differ in the position of the perceived surface peak (measured by the phase of its fit sinusoidal function), inter-observer variances tend to be relatively small compare to when the surface is flat. This combined measure is shown in Figure 3.5. The difference of LM+AM and LM-AM in the combined measurement is also provided in Figure 3.6. Table 3.1 gives details of depth amplitudes and phases for each individual observer in response to plaid configurations under all conditions.

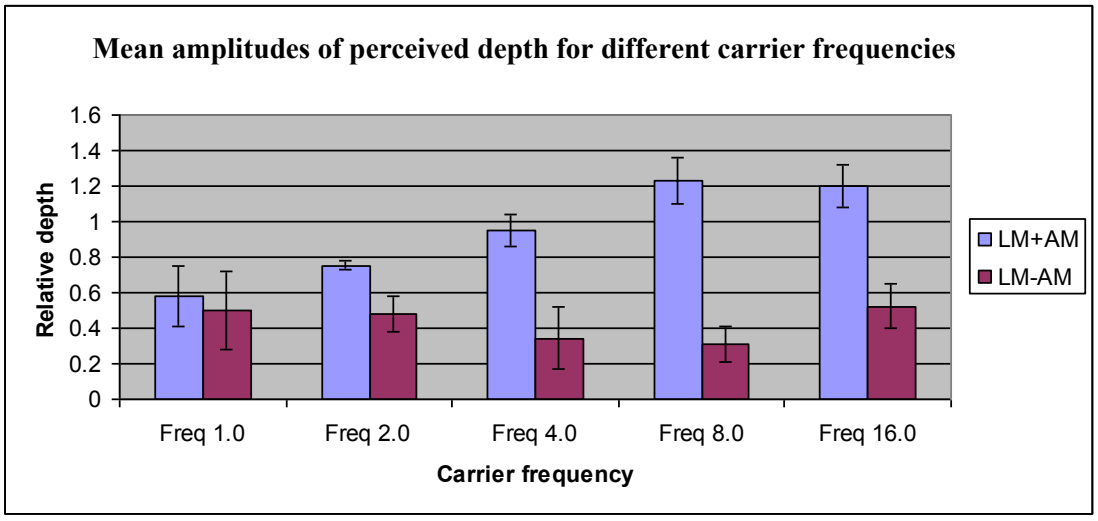


Figure 3.3 Mean depth amplitude calculated across four participants for five frequency conditions. Phase information is not considered. Error bars represent 95% confidence level.

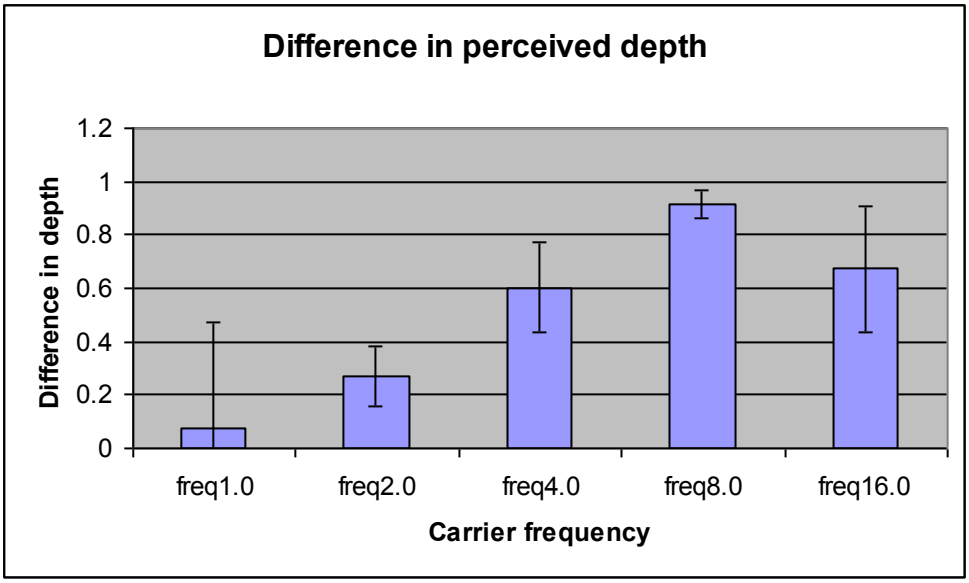


Figure 3.4 Perceptual difference between LM+AM and LM-AM presented in plaid configuration. Error bars represent 95% confidence level; n=4.

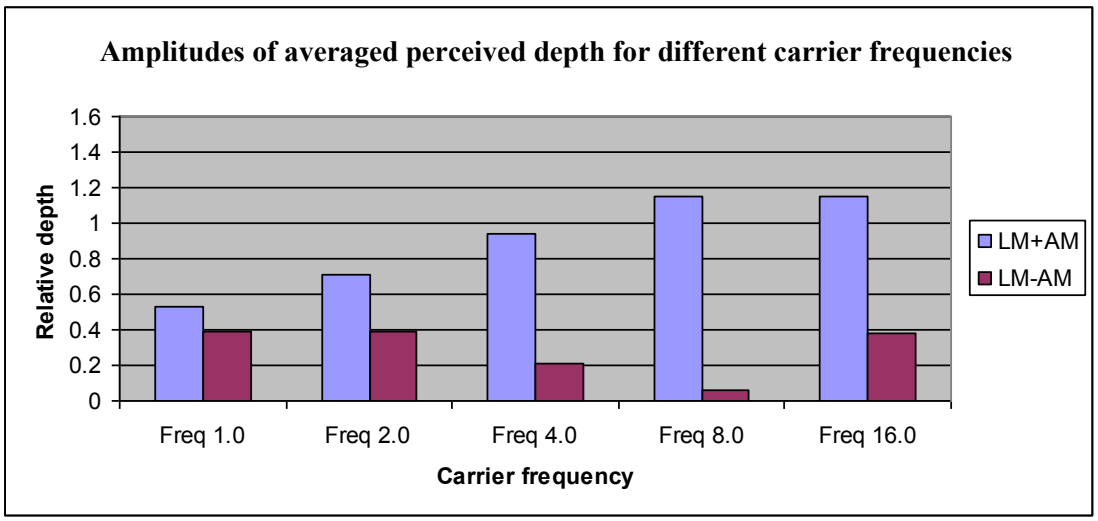


Figure 3.5 The amplitude of the sine function resulted from averaging the depth profiles.

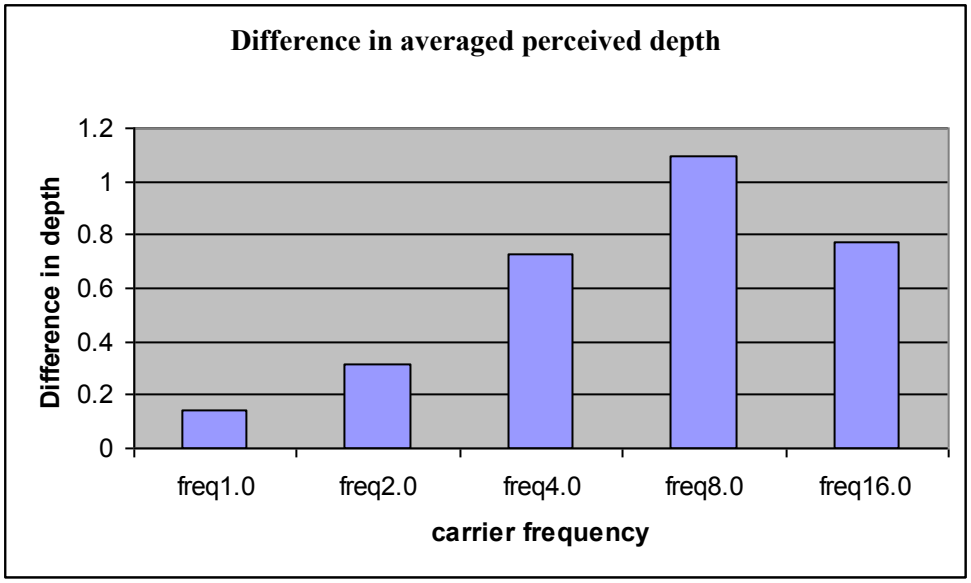


Figure 3.6 Perceptual difference between LM+AM and LM-AM across different carrier frequencies.

Carrier frequency 1.0 c/d						
		LM+AM		LM-AM		Difference Fundamental Amplitude
ID		Fundamental Amplitude	Fundamental phase	Fundamental Amplitude	Fundamental Phase	
SL	45	0.148	-151.3	0.787	-105.7	-0.639
	-45	0.558	-132	0.845	-116.2	-0.287
WH	45	0.529	-167	0.485	-183.7	0.044
	-45	0.852	-174.9	0.47	-190	0.382
WXX	45	0.624	-106.1	0.366	-140	0.258
	-45	0.429	-121.9	0.507	-117.7	-0.078
YJY	45	0.778	-150.2	0.226	-19.5	0.552
	-45	0.710	-159.8	0.315	-125.3	0.395
Carrier frequency 2.0 c/d						
		LM+AM		LM-AM		Difference Fundamental Amplitude
ID		Fundamental Amplitude	Fundamental phase	Fundamental Amplitude	Fundamental Phase	
SL	45	0.757	-108.1	0.54	-95.7	0.217
	-45	0.699	-117.2	0.346	-126.1	0.353
WH	45	1.017	-161.4	0.874	-152.1	0.143
	-45	0.449	-177.2	0.392	-101.7	0.057
WXX	45	0.735	-151.1	0.283	-53	0.452
	-45	0.839	-140.2	0.618	-107.1	0.221
YJY	45	0.928	-157.5	0.311	-166.000	0.617
	-45	0.600	-137.1	0.494	-182.6	0.106

Carrier frequency 4.0 c/d						
		LM+AM		LM-AM		Difference Fundamental Amplitude
ID		Fundamental Amplitude	Fundamental phase	Fundamental Amplitude	Fundamental Phase	
SL	45	1.178	-127.7	0.51	-104.6	0.668
	-45	0.707	-110.6	0.602	-84.9	0.105
WH	45	1.004	-151.2	0.472	-168.1	0.532
	-45	1.026	-166	0.387	-131.7	0.639
WXX	45	1.261	-172.9	0.315	-22.5	0.946
	-45	0.773	-147.8	0.121	-90	0.652
YJY	45	1.011	-154.200	0.160	-266.700	0.851
	-45	0.620	-158.600	0.187	-166.400	0.433

Carrier frequency 8.0 c/d						
		LM+AM		LM-AM		Difference Fundamental Amplitude
ID		Fundamental Amplitude	Fundamental phase	Fundamental Amplitude	Fundamental Phase	
SL	45	1.207	-120.5	0.162	-262.1	1.045
	-45	0.957	-113.2	0.305	-98.3	0.652
WH	45	1.277	-176.6	0.312	-184.1	0.965
	-45	1.116	-168.3	0.335	-255.4	0.781
WXX	45	1.387	-163.3	0.546	-160.5	0.841
	-45	1.421	-159.1	0.356	-48.6	1.065
YJY	45	1.219	-156.100	0.174	-352.700	1.0
	-45	1.221	-145.700	0.296	-25.800	0.9

Carrier frequency 16.0 c/d						
		LM+AM		LM-AM		Difference Fundamental Amplitude
ID		Fundamental Amplitude	Fundamental phase	Fundamental Amplitude	Fundamental Phase	
SL	45	1.228	-117.1	0.751	-91.7	0.477
	-45	0.829	-139.6	0.657	-83.4	0.172
WH	45	1.448	-149.2	0.697	-135	0.751
	-45	1.185	-163.7	0.226	-225	0.959
WXX	45	1.319	-154.4	0.435	-155.3	0.884
	-45	1.133	-124.6	0.603	-78.5	0.53
YJY	45	1.409	-158.000	0.491	-196.300	0.9
	-45	1.023	-150.300	0.337	-116.600	0.7

Table 3.1 details of depth amplitudes and phases for each individual observer in response to plaid configurations under all conditions. The difference in amplitude is the subtraction of LM-AM

from LM+AM. The phase values represent the phase shifts of the fitted cosine functions. Thus a minus 90° phase shift means that the fit cosine function perfectly coincides with the underlying sinusoidal luminance.

Results are consistent with those of the plaid experiment described in chapter two. LM+AM had a much higher perceived depth amplitude than LM-AM on higher frequency carriers (4.0 and 8.0). In addition, depth profiles for LM+AM were offset from the luminance peaks by about 1/8~1/4th wavelength below the luminance peak and was very stable across observers whereas the position of the perceived peaks for LM-AM varied considerably between observers. Levene's test for equality of variance gives:

$STD_{LM-AM} = 113.4, STD_{LM+AM} = 22.6, p = 0.005$ for 8.0c/d carrier

$STD_{LM-AM} = 73.0, STD_{LM+AM} = 20.4, p = 0.033$ for 4.0 c/d carrier

On the other hand, LM+AM and LM-AM were less distinguishable on lower frequency carriers: their fundamental amplitudes were more similar and the phase of LM-AM became more stable, as if the LM-AM condition became more like the LM+AM condition for low frequency carriers. This can be concluded by the decreasing standard deviations of phase values for LM-AM with the decrease in carrier frequency, as plotted in Figure 3.7. The difference in the standard deviations of LM-AM phases is significant between 8.0 c/d and 1.0 c/d carriers ($p = 0.044$) and between 8.0 c/d and 2.0 c/d carriers ($p = 0.025$), revealed by the Levene's test for equality of variance.

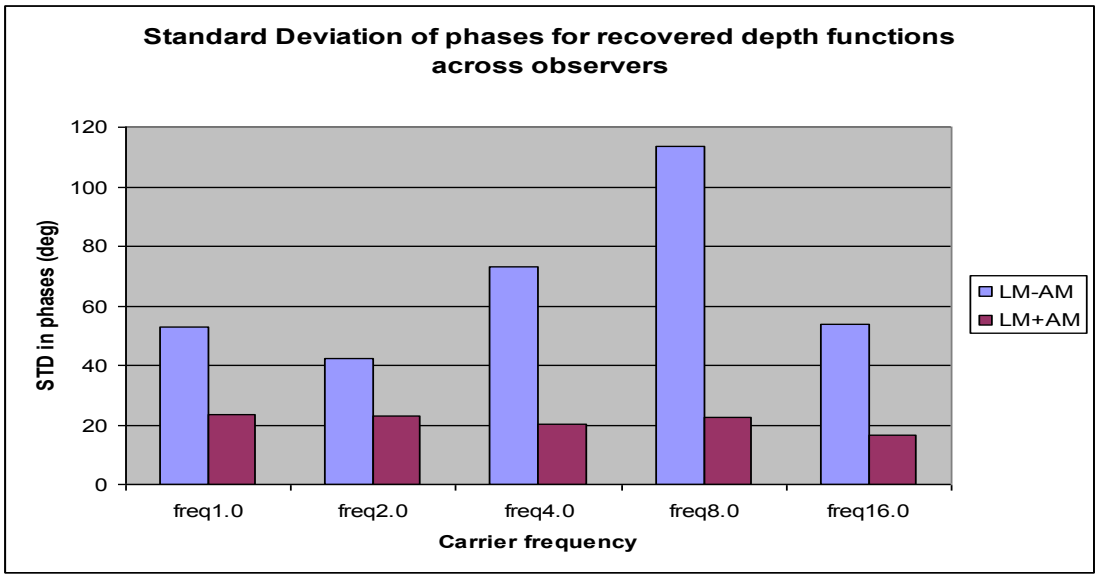


Figure 3.7 Standard deviations of phase values for recovered depth functions for all carrier frequencies. Phases of LM-AM were more sparsely distributed among observers for 8 c/d carrier.

Figure 3.5 combines phase and amplitude information and provides a better illustration of perceived depth across all frequencies. Note that perceived depth in Figure 3.5 for LM-AM on 8.0 c/d carriers was even more reduced compare to that in Figure 3.3 whereas on 1.0 c/d and 2.0 c/d carriers, it was almost unaffected. The perceptual difference between LM+AM and LM-AM was most significant when carrier frequency was at 8.0 c/d and was least significant when carrier frequency was at 1.0 c/d (Figure 3.5). This distinction steadily declined with the decreasing carrier frequency. Figure 3.5 also shows that on 16.0 c/d, LM-AM seemed to appear more corrugated than that on 8.0 c/d.

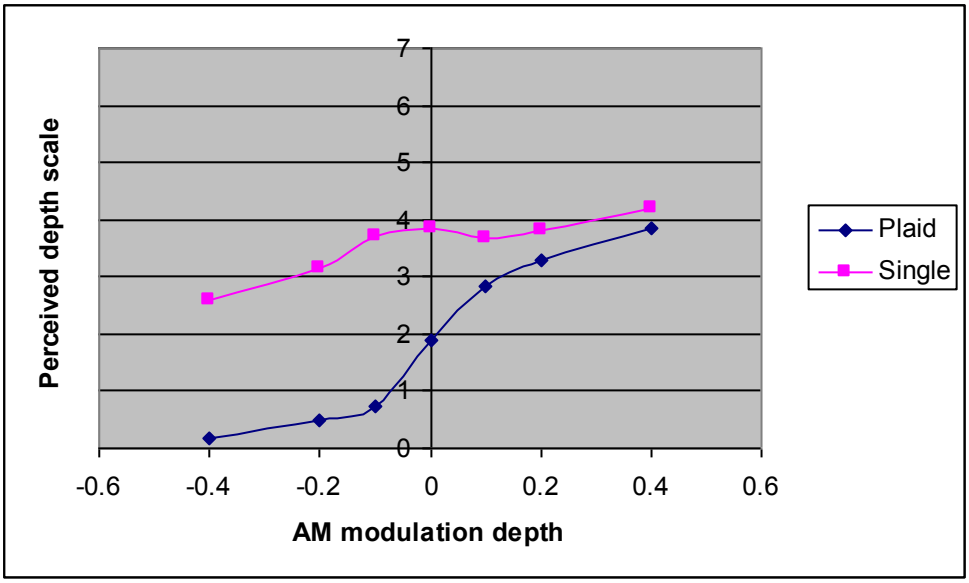


Figure 3.8 Perceived depth plotted as functions of AM modulation depth. Negative AM values indicate anti phase combination i.e. LM-AM. Positive values indicate in phase combination i.e. LM+AM. (diamond) Perceived depth for plaid configuration. (square) Perceived depth for single oblique. Data taken from Schofield et al in press, not collected by the author.

3.4.2 Discussion

Even on very low frequency carriers (2.0 c/d), observers still perceived LM+AM to be more corrugated than LM-AM, suggesting that AM detection was functioning even at such low carrier frequencies. Whether AM was processed at all on carriers with 1.0 c/d frequency is not clear due to the large errors.

Schofield et al (in press) have shown that AM modulates shape-from-shading in textured surfaces. This modulation depends on the strength of AM signal. As the strength of AM approaches zero, the perceived depth of LM+AM reduces whereas that for LM-AM was enhanced so that they became less distinguishable and eventually meet at a medium depth level when AM is zero. Figure 3.8 describes this dependency. The x-axis represents the modulation depth of AM. Negative values indicate LM-AM. The perceived depth for the combination of LM and AM in a plaid appear to be a sigmoidal function with LM-AM being seen as flat. However, single

oblique stimuli appear more corrugated in general and decline only slightly when AM is out of phase with LM. This pattern has been produced in this experiment by varying the carrier frequency instead of AM signal strength. LM+AM and LM-AM best distinguished when AM is carried by 8.0 c/d Gabor textures, 4 octaves above the modulation frequency. The gap between the two closed with decreasing carrier frequency. This suggests that carrier frequency affects the strength of the de-modulated AM signal.

As shown in Figure 3.3, perceived depth for LM-AM was gradually enhanced as carrier frequency decreased, which excludes the possibility that masking or any other first-order artefacts simply inhibited the detection of LM. In Figure 3.4 and 3.6, the influence by AM seems to suggest a band pass characteristic with an optimal carrier to modulation ratio of 16. The result shown here is consistent with results from Sutter et al (1995) and partially similar to results reported by Dakin and Mareschal (2000) although the latter did not report a deterioration in performance at high ratios of carrier to modulation frequency (above 32:1). However, despite testing the same maximum carrier**:**modulation frequency ratio, the highest carrier frequencies tested in the two studies were different. Dakin and Mareschal only tested carrier frequencies up to 8.0 c/d whereas Sutter et al tested carrier frequencies up to 16.0 c/d and only reported a deterioration at such high frequencies. Indeed, in the present study, the highest carrier frequency tested is same as that of Sutter et al (1995) and a similar decline in AM visibility was found. Sutter et al attributed the band pass property to a specific-mapping between carrier processing mechanism and modulation processing mechanism. On the other hand, because Dakin and Mareschal did not find deterioration at high ratios, they suggested a general mapping between these two

mechanisms and a broader tuning of the carrier frequency selectivity. To reconcile these studies, I argue that the detection of second-order signal will drop after it reaches its maximum but the deterioration is unlikely to be dependant on the ratio of carrier to modulation frequency. Instead, it depends on the absolute value of carrier frequencies (16.0 c/d as suggested by Sutter et al 1995 and the shape-from-shading task reported here). There are two possible explanations: One is that the carrier processing mechanism is band-pass in frequency. This idea has some support from physiological studies which reported that in cat area 18, cells responsive to second-order stimuli were selective to a band of high carrier frequencies (Zhou & Baker, 1996; Song & Baker, 2006). The other explanation is that the second-order stimuli are not detected well at high carrier frequencies due to the reduced visibility of the carrier itself at very high frequencies. Note that human contrast sensitivity drops considerably between 8.0 c/d to 16.0 c/d (Campbell & Robson, 1968).

The reduced influence of AM at lower carrier frequencies is consistent with results from both studies and could be accounted by the idea that when the preferred frequencies of modulation processing and carrier processing mechanisms differ by less than 3 octaves, connections are made between first- and second-stage filters with orthogonal preferred orientations only (Schofield, 2000; Dakin & Mareschal, 2000) thus reducing the effective power of the carrier and hence its ability to support the detection of AM.

Note that Schofield et al. (in press) report near symmetrical changes for both LM+AM and LM-AM when gradually reducing the strength of AM to zero. Using a haptic match method, they tested LM+AM and LM-AM in both a plaid configuration

and individually. Their results are shown in Figure 3.8 which depicts the perceived depths as a function of the strength of AM. Negative AM values indicate LM-AM while positive values indicate LM+AM. In Figure 3.3 however, the depth perceptions for LM+AM and LM-AM did not change symmetrically: the reduction in perceived depth for LM+AM was greater than the enhancement of perceived depth for LM-AM cues. This might be caused by the interference from the first-order carrier on the LM signal. The texture elements were enlarged as a result of reducing the carrier frequency, these elements may look like depth ripples at low frequencies. If this is the case, reducing the carrier frequency may have reduced the distinction between LM+AM and LM-AM due to inadequate AM detection and also reduced the overall reliability of the depth percept due to interference from the carrier. If this hypothesis is true then single oblique stimuli on low frequency carriers will also result in suppressed depth perception. This time however the suppression will be dominated by interference from the carrier. The next experiment attempts to verify this hypothesis.

3.5 Experiment 2 Effect on single oblique

The asymmetry of changes in perceived depth for LM+AM and LM-AM suggests that the carrier directly interferes with the perception of shape-from-shading process; affecting both LM-AM and LM+AM. If this was the case, we would expect to see a suppression in perceived depth for both LM+AM and LM-AM when presented as single oblique stimuli. This was tested in experiment 2. All experimental details were the same as the previous experiment except for the stimuli tested.

3.5.1 Results and discussions

Figure 3.9 shows mean depth amplitude averaged across four participants. Results for plaids are also included so as to make comparison easier between these two

configurations. As predicted, single obliques were perceived less reliably corrugated when carried by lower frequencies, regardless of the phase relationships between the components. Perceived depth for LM+AM and LM-AM dropped at the same rate. The effect of AM phase was not measurable. Therefore, any visible variations in perceived depth could well be due to the same source of interference which would enforce same impact on both combinations. Similar to what was found by Schofield et al. (in press), the perceived depth for single oblique was generally higher than that for plaid configurations.

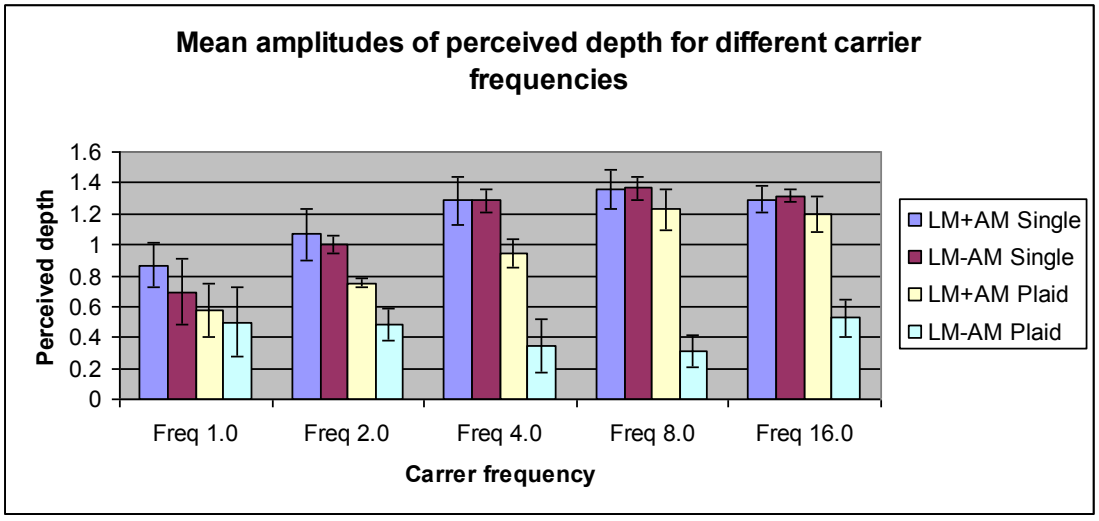


Figure 3.9 mean depth amplitude averaged across four participants.

3.6 General discussion

3.6.1 Carrier frequency modulates depth perception

Prior to estimating shape-from-shading, humans are likely to conduct a process to disambiguate luminance variations and select only those that are most likely due to shading in natural scenes (see Introduction). Along with other cues (e.g. colour), AM is believed to be involved in this selection process such that luminance signals that are

correlated with AM are preferentially weighted for later shape-from-shading analysis (Schofield et al., 2006). In this chapter, it has been shown that varying carrier frequencies has an impact on this disambiguation process via two identifiable routes.

1) Classification based on frequencies of luminance variations

Participants seemed to base their surface perception on luminance modulations while ignoring luminance variations caused by high frequency textures. However when the carrier was low frequency, carrier elements started to interfere the judgment of the surface gradient: they appear as random undulations in their own right. This was true for both plaid and single oblique configurations. Results for single oblique stimuli suggest that this interference starts when carrier frequency is less than 4 times the modulation frequency and continues to grow as carrier frequency decreases. Based on this observation, it is proposed that humans are able to exclude high frequency luminance variations from any subsequent shape analysis but retain low frequency luminance variations. The classification may be achieved by conventional linear spatial channel with a low pass band. In the single oblique experiment, low frequency carriers were not excluded but were carried through to future shape-from-shading analysis, due to carriers leaking through the channel that processes the low frequency luminance modulation signal. The idea that humans assume a low frequency characteristic for changes in illumination intensity is in agreement with a number of classic machine vision algorithms separating illumination from reflectance. Algorithms such as Retinex (Land & McCann, 1971) and its refined versions (Horn, 1974; Blake, 1985) were based on the same assumption and are still in wide uses in many real world applications.

2) Classification based on accompanied luminance amplitude modulations

Secondly, although AM helps to disambiguate luminance variations as either shading or reflectance changes, the effectiveness of the AM based classification is determined by the carrier frequency. This selection process is most effective when carrier to modulation frequency ratio falls into the range of 8:1~32:1, with a peak at 16:1.

3.6.2 Implications for second-order vision

Since it is the relationship between LM and AM that makes LM+AM and LM-AM distinct, examination of the perceptual difference of the two combinations reveals some characteristics of AM processing mechanisms in visual systems.

1) Does second-order vision exist at all for low frequency carriers?

First-order luminance artefact due to side-band signals may act as a cue for presence of second-order signal in a detection task, e.g. the luminance edges present in Figure 3.1 (Henning, Hertz & Broadbent, 1975). Although this effect was controlled in many psychophysical studies, side-band effects could not be entirely excluded for low frequency carriers (Dakin & Mareschal, 2000). In the current study, it has been shown that the perceptual difference, although much less than its maximum value, still exists for carriers with frequencies as low as 2.0 c/d, 4 time the modulation frequency. Due to the nature of the probe task, the perceptual difference was unlikely due to luminance defined edges thus indicating that second-order vision operates under this condition. For 1.0 c/d carriers, evidence is not strong enough to support a processing of AM.

2) The processing of AM is tuned to high frequency carriers

The influence of AM peaked for 8.0 c/d frequency carriers which is 16 times that of the modulation frequency. It was steadily reduced when carrier frequencies went below this value. Qualitatively, this finding is consistent with data obtained by Dakin and Mareschal (2000) and Sutter et al. (1995) which reported that contrast modulation processing was tuned to high frequency carriers and there was a smooth transition from low detection threshold for high carrier frequencies to high detection threshold for low carrier frequencies. Both studies suggested that the decline started when carrier frequencies dropped to around 8 times the modulation frequency, similar to what was reported in the present study. This ratio seems to be scale invariant since it holds true for both modulation frequencies tested in Dakin and Mareschal's experiment (0.35 c/d and 0.7 c/d) and the modulation frequency tested in current study (0.5 c/d). Whether it is true for even lower modulation frequencies remains untested. Thus a more modest conclusion is that the processing of AM is tuned to carrier frequencies that are at least 2 octaves above the modulation frequency. This ratio seems to be scale invariant for at least a range of modulation frequencies based on data from both present and previous studies.

3) The carrier frequency tuning is also band-limited

For 16.0 c/d carriers (carrier/modulation: 32/1), the influence of AM seemed to be reduced relative to 8.0 c/d carriers, which is consistent with the finding of Sutter et al. (1995). However this ratio does not seem to be scale-invariant as no deterioration was found at such ratio in a later psychophysical study (Dakin & Mareschal, 2000). The discrepancy could be due to the fact that the high end frequency tested (8.0 c/d) in the later study was not high enough to be significantly attenuated by the first order contrast sensitivity function.

Together with *1)* and *2)*, it is thus proposed that the processing of AM shows carrier frequency dependence. Generally, the mechanism that processes AM is tuned to a band of higher frequencies. The lower bound of such pass band is at least 2 octaves above the modulation frequency so that the two frequency values remain in a fixed ratio. There should exist an upper bound of this pass band, although it was not quantitatively identified in this study. However, the upper bound should be above 16.0 c/d and does not depend on modulation frequency. It may be due to the contrast sensitivity function or bandwidth restrictions in early visual processing. Some supporting evidence can be found from studies of envelope responsive cells in cat area 17/18 which demonstrated that responses driven by envelope signals were selective to carrier frequencies ranging from 4 or 5 times of the modulation frequencies to the upper resolution limit of the X-retinal ganglion cells at the same retinal eccentricity (Zhou & Baker, 1996; Mareschal & Baker 1999; Song & Baker 2006; Song & Baker 2007).

4. A model that can account for human's ability to disambiguate luminance changes for shape-from-shading analysis

This chapter describes a model to explain observers' performance in two different shape-from-shading tasks: a haptic matching experiment and the previously described two-point probe task. The model constitutes the feature extraction- and luminance classification units introduced, as a part of the general framework for the shape-from-shading, in the introduction. The influence of AM on the perceived depth of LM signals was modelled by a summation between LM and AM channels. Inhibition across orientation channels models the exaggerated suppression of perceived depth for LM-AM when placed against LM+AM, as compared to when presented alone. The model predicts performance in a haptic depth matching experiment. With some further adjustments, it can also predict the results form the probe-point experiments of chapter 3.

4.1 General structure

The proposed general framework for shape-from-shading is illustrated in Figure 4.1, with the section modelled here enclosed in dashed lines. In unit 1, the retinal image is decomposed and represented as features at different frequencies and orientations. Conventionally this stage of visual processing is modelled by a bank of linear filters spanning a range of spatial frequencies and orientations, mimicking the known property of cells in area V1 of primate visual cortex (Marcelja, 1980). Another way to think about this first stage is that early processing in the visual system conducts a windowed Fourier transform and codes the retinal image with coefficients representing energies at different frequencies and orientations. In the current model,

second-order features are also extracted from the retinal images as they are important at the classification stage. The early extraction of such second order features is consistent with neurophysiology (Mareschal & Baker, 1998a; 1998b; Mareschal & Baker, 1999; Zhou & Baker, 1996). In unit 2, the features extracted by unit 1 are classified according to rules that have been discussed previously. The output of unit 2 (the output of the model discussed here) represents the strengths of shading components from which the surface shape can be computed. The following subsections present the two units under discussion in detail.

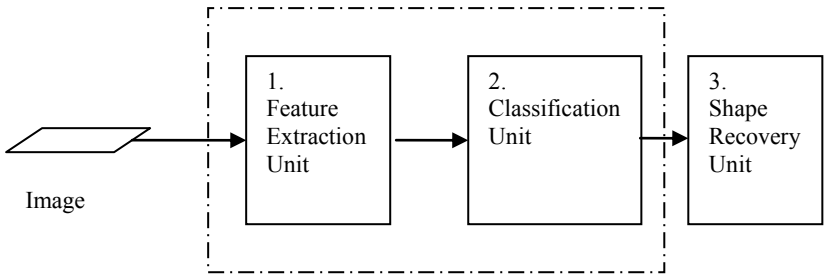


Figure 4.1 Shape-from-shading framework (redrawn from Fig 1.11). The model to be described in this enclosed within the dashed lines.

4.2 Feature extraction unit

4.2.1 First-order feature extraction

First-order features are extracted by convolving the retinal image with a series of linear filters with a subsequent compressive nonlinearity limiting the amplitude of the response. This process has been accepted as a way to model the processing of first-order stimuli in the early stage of visual perception (Carandini, Heeger & Movshon, 1999) although it does not capture cross-channel inhibition.

4.2.2 Second-order feature extraction

Second-order features are extracted using a separate second-order mechanism. Hypothetical models have been developed in recent decades that can process second-

order information. Wilson et al. (1992) proposed a filter-rectify-filter (FRF) model for second-order vision which comprised two filtering stages separated by a nonlinearity (see figure 4.2). The nonlinearity provides a full-wave or half-wave rectification to the responses from the first-stage filters such that simulated neural responses are positive. This nonlinear rectification is thought to model the responses of ON and OFF receptive fields (Malik & Perona, 1990) and functions as a demodulator for the carried signals (Schofield, 2000). The FRF model has been proposed to mediate the detection of illusory contours (Song & Baker, 2006) and the detection of signals that are modulations of orientation, contrast and spatial frequencies (Kingdom & Keeble, 1996; Kingdom et al, 2003; Arsenault, Wilkinson & Kingdom, 1999). Although differing from each other in the specific choice of parameters, all of the above implementations seem to agree on the relative sizes of the two filters. The first-stage filter was normally tuned to relatively high frequencies such that the high frequency carrier components will be processed and low frequency modulations can be passed on to the second-stage filter.

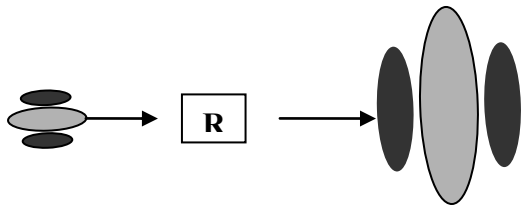


Figure 4.2 A FRF model that can process second-order information. The first stage filter is tuned to relatively high frequencies and processes carrier components while blocking low-frequency first-order signals. The rectifier (R) demodulates the second order signal. The second stage filter is tuned to relatively low frequencies to reject high frequency carrier component and pass the modulation components.

4.2.3 An elaborated FRF model

The original FRF model has been elaborated as psychophysical and physiological evidence regarding the nature of second-order vision has accumulated (Sutter et al,

1995; Graham & Sutter, 1998; Dakin & Mareschal, 2000; Graham & Sutter, 2000; Ledgeway, Zhan, Johnson, Song & Baker, 2005).

The intermediate nonlinearity (rectifier) represents a gross nonlinear process in the second-order channel. Whilst the rectifier is piecewise linear in many FRF implementations (Malik & Perona, 1990; Schofield, 2000; Johnson & Baker, 2004), psychophysical studies on visual texture segregation suggest that it is probably an expansive power function with an exponent between 3 and 4 (Graham & Sutter, 1998). In physiology, Ledgeway et al. (2005) recorded spike rates of cells in area 18 of cat that were responsive to moving contrast modulations (second-order motion). When plotted as functions of either modulation contrast or carrier contrast, responses of these cells were expansive. Neurones typically require considerably more second-order modulation than first-order luminance contrast to elicit the same response. Further whereas first-order responses are compressive at high contrast no such saturation is found for second-order signals. Based on a comparison of the two contrast response functions (CRFs), Ledgeway et al. (2005) proposed three versions of the FRF mechanism. In the first version (shown here in Fig 4.3b), the intermediate nonlinearity is piecewise linear but the contrast response of the second stage filter has a much higher threshold than that of linear channels with similar tuning (e.g. preferred spatial frequency and orientation). Alternatively, the observed CRF for second-order motion could be a result of second-order channel being less sensitive than first-order channel by a scaling factor. The second stage filter in the second-order channel has the same contrast response as its counterpart in a first-order channel (Fig 4.3a) and the intermediate rectifier is also linear. However the responses from the first-stage filters are multiplicatively reduced such that only the expansive part of the CRF curve is

observed (see Fig 4.3c). The third version favours an expansive rectifier obeying a steep power law (Fig 4.3d). Ledgeway et al. (2005) were inclined to the third version as it was consistent with human psychophysics (Graham & Sutter, 1998).

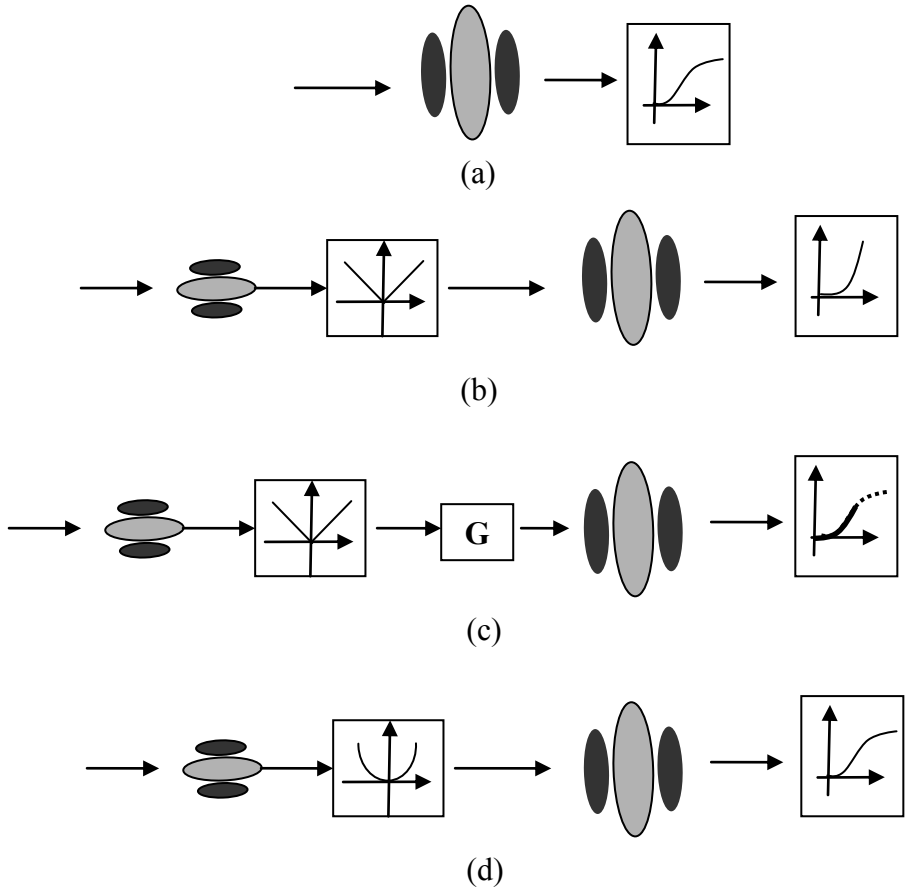


Figure 4.3 Diagram of a first-order channel (a) and possible second-order channel structures (b-d). (a) The response in first-order channel is intensively nonlinear at lower contrast, followed by an immediate acceleration and saturation at high contrasts. (b) The intermediate rectifier is a piecewise linear function. But the transfer function of the second-stage filter has a higher threshold and a deeper accelerating curve than that of (a). (c) The second filter has a transfer function similar to (a) but its input signal is reduced so that the second-stage filter only operates over the lower half of its transfer function. (d) The rectifier obeys a deep power law giving significant suppression for low contrast carriers or signals having weak modulation depths but the net transfer function of the mechanism is expansive. (After Ledgeway et al. 2005)

Note that the intermediate nonlinearity is only a conceptual unit existing in the cortical process of second-order vision. It is not necessarily a unique neural mechanism nor does it have to functionally lie between the two filtering processes. It could in principle arise from any known nonlinearities existing in the relatively early visual processes (e.g. inhibition among channels), although it is unlikely to be due to

very early nonlinearities (e.g. light adaption) in LGN and retina (Graham & Sutter, 2000).

There is now evidence suggesting that the first-stage filter should be a bank of orientation selective filters rather than one single isotropic filter (Dakin & Mareschal, 2000). Further, the dependence of contrast modulation on carrier frequency has implications for the connections between the first- and second-stage filters (Sutter et al, 1995; Dakin & Mareschal, 2000). Two elaborated versions of the FRF model were implemented by Schofield (2000). In one of the models, first-stage filters were only connected to second-stage filters with preferred frequencies at least two octaves below their own. Connections were made between orthogonal filters only when the difference in the two preferred frequencies was three octaves or less. Above this threshold, second-stage filters received input from multiple orientation selective first-order filters such that the assembly of the first-stage filters had broad orientation selectivity. This design is in agreement with the finding that when the carrier-to-envelope frequency ratio drops below 3 octaves, the underlying second-order information becomes harder to detect (Sutter et al, 1995; Dakin & Mareschal, 2000; see also Chapter 3), and the mechanism as a whole becomes tuned to the carrier orientation – preferring those orientations orthogonal to the modulation (Dakin & Mareschal, 2000).

I have adapted this model with a slight modification to the rules governing inter-connections between first- and second-stage filters. In the new model, when the preferred frequencies of first- and second-stage filters differ by more than 3 octaves, the connections are as described by Schofield (2000). However, when the frequency

difference is exactly 3 octaves orientation tuning makes a smooth transition towards narrower tuning by the application of weights to each first-stage filter. The weights are calculated from a Gaussian function with a mean value at the orientation orthogonal to the modulation. This Gaussian function has standard deviation of 45°, reflecting Dakin and Mareschal's (2000) data. For frequency ratios below 3 octaves, the second-stage filters only receive input from orthogonal first-stage filters as per Schofield's model (2000). When the frequency difference is exactly 1 octave, second-stage filters still receive input from orthogonal first-stage filters but a lower weight is applied to reflect the fact that the sensitivity of second-order vision reduces monotonically with carrier frequency (Dakin & Mareschal, 2000; Sutter, 1995). A graphical illustration of the adapted mode is shown in Figure 4.4 while Figure 4.5 provides a summary of the feature extraction unit as a whole.

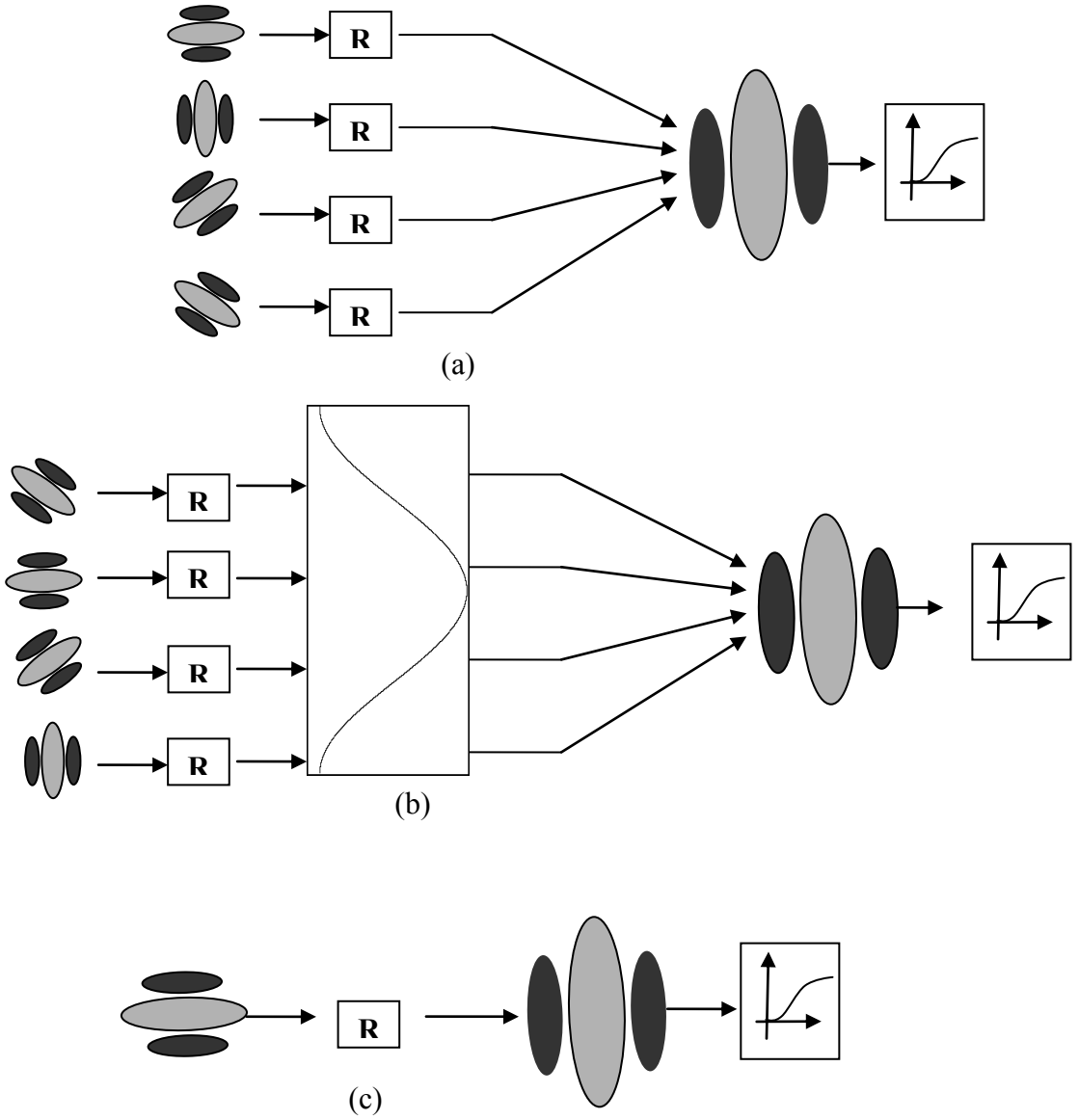


Figure 4.4 Model for second-order feature extraction used in this chapter. (a) when the two preferred frequencies differ by more than 3 octaves, second-stage filters receives input from a broad band of orientation selective first-stage filters. (b) when two frequencies differ by exactly 3 octaves, input from first stage filters are weighted according to a Gaussian function with a mean value at the orientation orthogonal to that of the second-stage filters. (c) Below 3 octaves, second-stage filters are wired to orthogonal first-stage filters only. The sigmoid functions at the end of each channel represent possible nonlinear transfer functions and do not represent a particular shape. R provides full-wave rectification but the choice of its shape will be further discussed later.

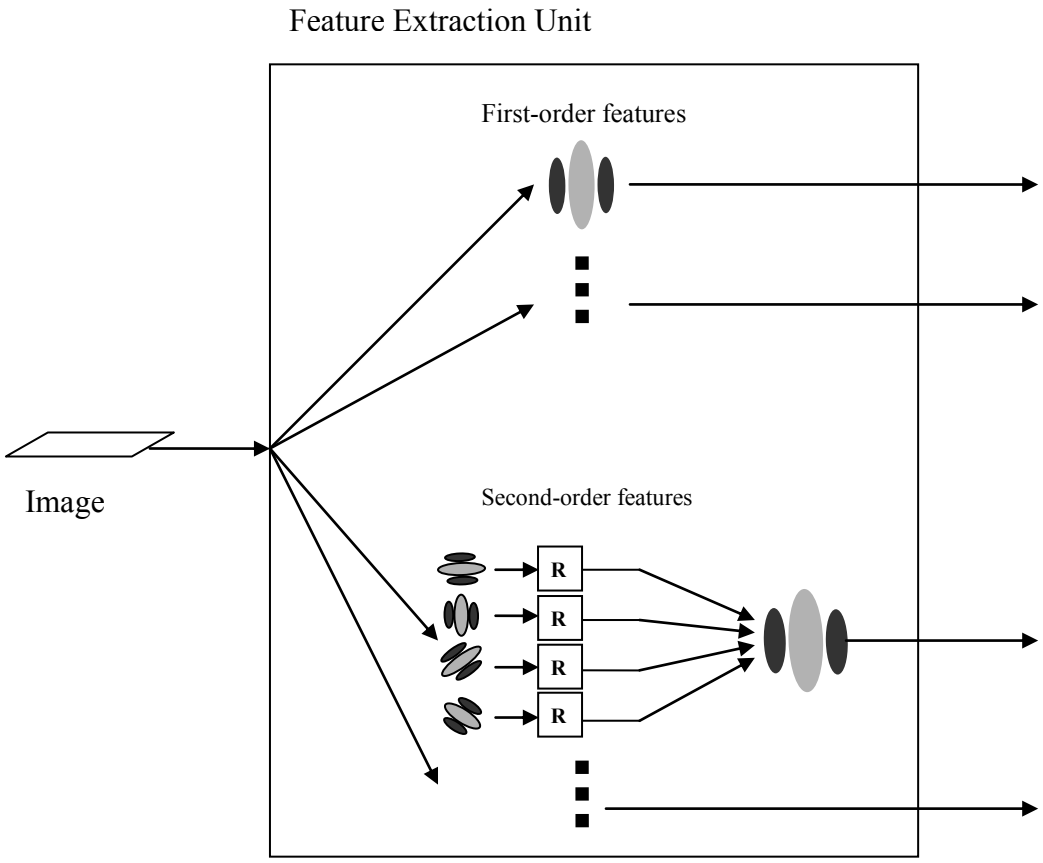


Figure 4.5 The content of the Feature Extraction Unit introduced in Section 4.1.

4.3 Classification Unit

The feature extraction unit does not differentiate between shading cues and reflectance changes; rather it treats all sources of luminance variation the same. The purpose of the classification unit is to mimic the ability of humans to disambiguate these features prior to a subsequent shape-from-shading analysis. Psychophysical studies have shown that humans use many cues to help with this disambiguation task but the current model only concerns the rules that were introduced in chapters 2 and 3.

4.3.1 Separation of shading and texture channels

A number of studies have suggested separate shading- and texture-processing channels (Georgeson & Schofield, 2002; Saki, 2006). The existence of such channels can also be inferred from the results reported in previous chapters. At present too little is known about how the visual system makes such categorizations but some intuitive yet hypothetical rules can be established to fulfil the intended purpose. Intuitively, shading information tends to be low frequency (or smooth edged), therefore low frequency first-order features should be weighed more strongly as shading than high frequency ones. Ideally a threshold function should be applied to determine the cut off point between shading and non-shading components but at present there is no data available to constrain such a function. The machine vision algorithms mentioned in Section 1.1.6 (Retinex and similar) do not have well-defined values for such a threshold but rather determine an appropriate threshold value from example stimuli. Here the separation of the two signal types was done in a rather ad-hoc manner: for the sake of simplicity, the model contained only one frequency channel tuned to the luminance modulation. All other frequency channels were assumed to be associated with textures.

4.3.2 Summation between shading and texture channels

Classifications based on feature spatial frequency alone are not sufficient to explain a favourable weighing for LM when associated with in-phase AM. Whether a shading component is boosted or suppressed depends on the phase relationship between the shading component (i.e. LM) and the accompanying AM. Such interactions can be modelled by a weighed summation between the LM and AM channels with the same orientation selectivity and preferred spatial frequency. This model echoes previous

reports that shading and texture are processed in initially separate information channels but are then integrated at a later, but still relatively early stage (Georgeson & Schofield, 2002). Note that the summation serves to enhance or suppress shading components but does not mean that the information from the two channels is merged, they may provide separate inputs to other processes. Note that Baker (1999) proposed a model structure for cells responsive to both first- and second-order motion in which first- and second-order responses are summed. But the summation in that model was to provide a concept of a combined response rather than arithmetic operation.

4.3.3 Need for a contrast gain control scheme

A simple summation between LM and AM channels falls short of a complete account of the data in Figure 3.8 regarding the influence of AM in a shape-from-shading task. First, a summation would produce a function expansive at both ends whereas the perceived depth of LM+AM and LM-AM mixtures saturates for LM+AM. Second, the perceived depth of a single component is always higher than that for the same component when presented as part of a plaid; a simple summation would result in a consistent depth percept regardless of the context in which a cue is presented. Third, LM-AM stimuli are perceived as having much less depth when presented in a plaid with LM+AM than when presented alone. This also could not arise from simple summation.

The behaviour mentioned above is reminiscent of similar nonlinear aspects of simple cell responses in area V1. For example, the amplitude of responses of simple cells saturate (Albrecht & Hamilton, 1982) similar to the saturation of perceived depth for plaids. Additionally, the fact that single components were perceived more depthy than components within a plaid is similar to the cross-orientation inhibition phenomenon

found in simple cells whose response to a superimposed pair of gratings is about half that for one grating alone (Bonds, 1989). To explain these nonlinearities Heeger and his colleagues (1993; 1994; 1996) have proposed a normalization model of simple cell responses that successfully predicts simple cell nonlinearities. I propose that a similar gain control scheme could account for the data plot in Figure 3.6.

4.3.4 Heeger's normalization model of simple cells

This subsection introduces Heeger's normalization model of simple cells. The electrical behaviour of a cell's membrane can be typically modelled by a compartment circuit with conductors and capacitors (Carandini & Heeger, 1994; Carandini et al, 1999), which is illustrated in Figure 4.6. The membrane potential changes over time and obeys Equation 4.1:

$$\begin{aligned} -C\frac{dV}{dt} &= g_i(V-V_i)+g_e(V-V_e)+g_{shunt}(V-V_{shunt})+g_{leak}(V-V_{leak}) = gV-I_d \\ I_d &= g_eV_e+g_iV_i+g_{shunt}V_{shunt}+g_{leak}V_{leak} \\ g &= g_e+g_i+g_{shunt}+g_{leak} \end{aligned} \tag{4.1}$$

where C represents the membrane capacitance, V_e, V_i and V_{shunt} are excitatory, inhibitory and shunt equilibrium potentials, g_e, g_i and g_{shunt} are the corresponding variable conductance resistors, and V_{leak}, g_{leak} together determine the leak current. The shunt variable resistor represents shunting inhibition which has been proposed to model how a cell's conductance changes with stimulation (Carandini & Heeger, 1994; Carandini et al, 1999). At the steady state, i.e. when $\frac{dV}{dt}=0$, the differential equation in 4.1 becomes:

$$V = I_d / g \tag{4.2}$$

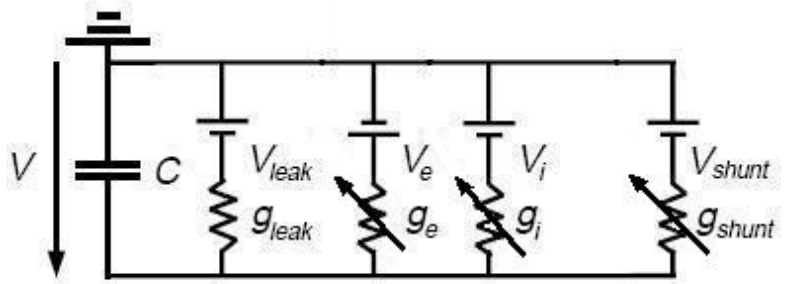


Figure 4.6 Circuit model of a cortical cell. The capacitance of the membrane is represented by the capacitor C. V_e, V_i and V_{shunt} are excitatory, inhibitory equilibrium and shunt equilibrium potentials. g_e, g_i and g_{shunt} are corresponding variable resistors. V_{leak} and g_{leak} determine the leak current. (After Carandini & Heeger, 1994)

g_e, g_i are varied in a push-pull manner such that the linear inputs trade off against one another as in equation 4.3:

$$g_i + g_e + g_{leak} = g_0 \tag{4.3}$$

where g_0 is a constant, representing the cell's conductance when there is no visual input. Then the cell's total conductance only depends on the shunt conductance g_{shunt}, which varies with the normalization resulting from the activation of all the cortical neurons in the assembly. The activity of a cell, i.e. its firing rate, is approximately related to the membrane potential by equation 4.4:

$$r \approx [\max(0, V)]^2 \tag{4.4}$$

In Heeger's normalization model, the authors also assume that shunt equilibrium potential equals a cell's resting potential and assert this as the reference potential:

$$V_{shunt} = V_{rest} = 0 \tag{4.5}$$

which suggests that I_d in equation 4.1 only depends on the visual input. Now it is clear that in the steady state, a cell's membrane potential depends on two sets of inputs: I_d, the linear input from the visual stimuli, and g, the cell's total conductance which in turn depends on the activation of all cells in the assembly. This term represents

divisive inhibition. Because the activity of a cell is related to its membrane potential by equation 4.4, these variables are then dependant on each other in a recursive manner, as described in equation 4.6:

$$R_i = K \frac{C_i}{\sigma^2 + \sum_j C_j} \qquad (4.6)$$
$$C = [\max(0, L)]^2$$

where C is the squared response of the conventional linear model of a simple cell. The denominator is the sum of the squared responses of all cells in the normalization pool plus a non-zero constant σ^2 which is related to g_0 in equation 4.3. The existence of σ^2 stops division-by-zero when there is no visual stimulus present. K is an overall scaling factor. Figure 4.7 depicts the circuit diagram of equation 4.6:

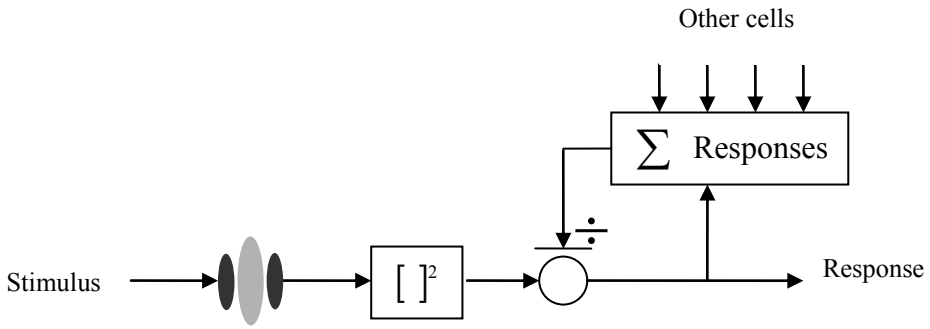


Figure 4.7 A circuit diagram for a normalization model of a simple cell at its steady state. The linear response is half squared and is normalized by responses from many other cells. (After Carandini & Heeger, 1994)

4.3.5 The contrast gain control scheme after weighed summation

A normalization loop similar to Heeger's normalization circuit was implemented after summing each first-order LM channel with its corresponding, weighed second-order AM channel. The *amplitude* of the response from a shading channel was squared and divided by responses from other shading channels tuned to different orientations and spatial frequencies. The final output represents the strength of the shading component

at each frequency and orientation. A diagram for the complete model is shown in Figure 4.8 illustrating the content of the classification unit.

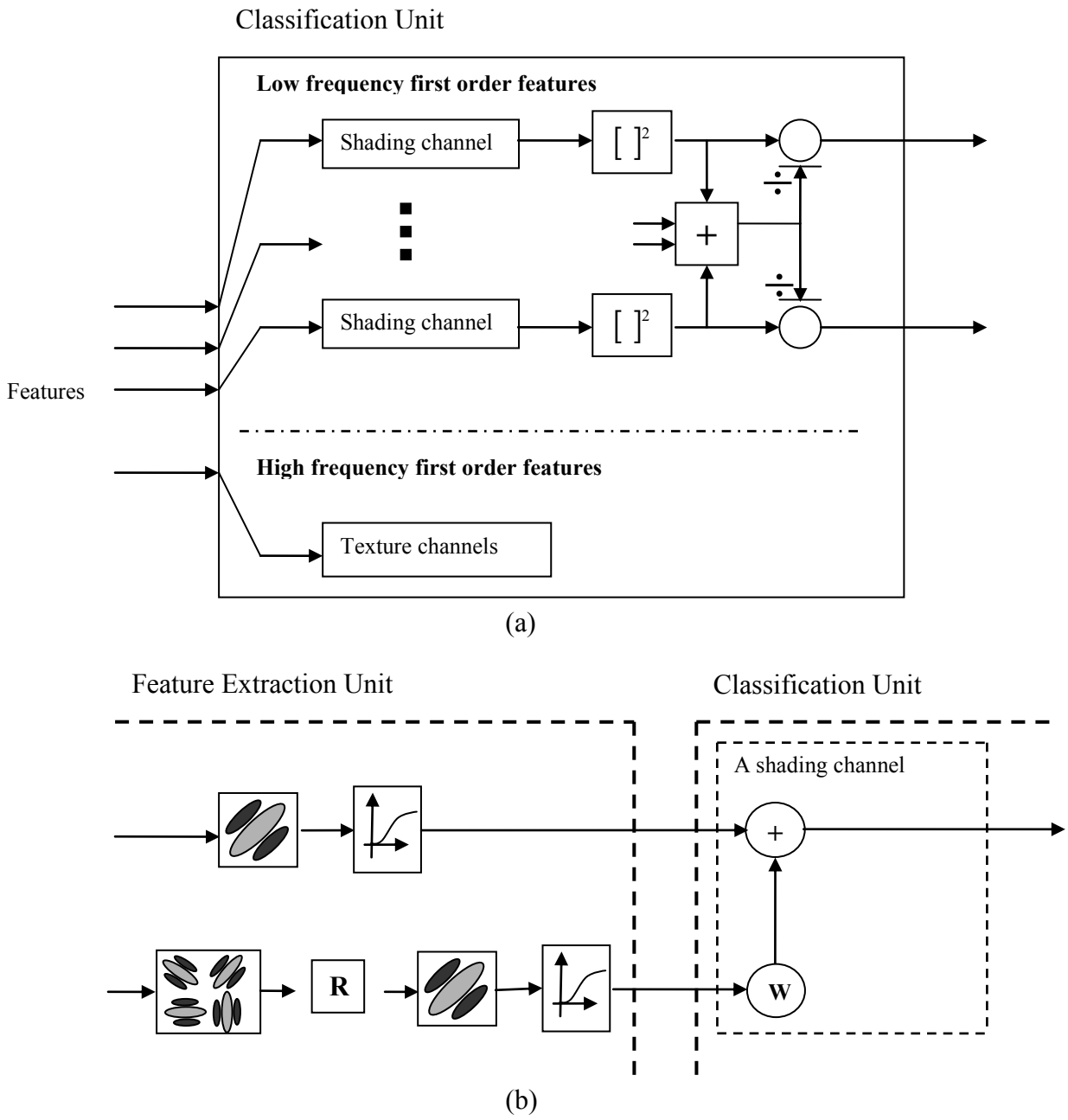


Figure 4.8 (a) The Classification Unit receives first-order features from the preceding unit and undertakes a crude shading and texture separation based on spatial frequency. Responses from cells with larger receptive fields are categorized as shading features, forming shading channels. The amplitude of the response from a shading channel is then normalized by response amplitudes of all the shading channels. (b) The content of a shading channel. In each shading channel, the constituent first-order feature is added to a weighed second-order feature tuned at the same orientation and frequency.

4.3.6 Possible neural basis of the proposed model structure

In cat areas 17 and 18, some cells are responsive to both first-order and second-order stimuli (Zhou & Baker, 1996; Mareschal & Baker, 1998a; Mareschal & Baker, 1998b; Mareschal & Baker, 1999; Zhan & Baker, 2008). When responding to first-order luminance gratings these cells tend to have a unique pass-band. These cells can also respond to a second-order modulation carried by first-order gratings which normally fall out of the first-order pass-band and would not excite the cells alone. When responding to second-order signals, these cells can have two separated pass-bands, one tuned to carriers and one to the modulation. Although often different, the pass-band for the modulation is close to the first-order pass-band. Moreover, when responding to the combination of LM and AM, the responses of these cells peaked when the two components were combined in-phase (LM+AM) and was much weaker for phase shifts of 180° (LM-AM), as if computing a linear sum of the two cues (Hutchinson, Baker and Ledgeway, 2007). Thus, these cells could serve as the neural mechanisms as described in Figure 4.8b and underlie the proposed computations.

4.4 Using experimental data to fit the model

The model illustrated in Fig 4.8 was implemented in two forms with different rectifying nonlinearities in the FRF network and subsequent nonlinear transfer functions.

Model implementation: version one

In this implementation, the intermediate rectifier obeyed a power law with an exponent of 3. Second-order channels had the same contrast response functions as first-order channels tuned to the same spatial frequencies and orientations. Second-

order channels constructed in this way are similar to the structure illustrated in Fig 4.3d.

The normalized strength of each shading component in Figure 4.8 (a) can be expressed in equation 4.7:

$$C_{Ri} = K\frac{C_{ri}}{\sigma^2 + \sum_j C_{rj}} \qquad (4.7)$$
$$C_r = [Var(r(x,y))]^2$$

where $r(x,y)$ is the response from one shading channel, $Var(\ )$ calculates its standard deviation as a measure of its amplitude, C_r takes the squared amplitude, C_R is the resulting amplitude after normalization which is in similar format to a normalized simple cell response described in equation 4.6. In practice, $Var(r(x,y))$ was approximated by taking the linear combination of the standard deviations of LM and AM signals. Since the relation of LM and AM in question was either anti-phase or in-phase, the sign of AM was either positive or negative accordingly:

$$\begin{aligned} &Var(r(x,y)) = Var(LM(x,y) + g \times AM(x,y)) \\ &\approx Var(LM(x,y)) \pm g \times Var(AM(x,y)) \end{aligned} \qquad (4.8)$$

where $LM(x,y)$ and $AM(x,y)$ are responses from LM and AM channels respectively. Before the weighed summation, the responses from both LM and AM channels are subject to saturation with the following squashing function:

$$SAT\{\} = \frac{e^{x/v}}{1+e^{x/v}} - 0.5 \qquad (4.9)$$

where v determines the saturation rate. The shape of the function is drawn in Figure 4.9:

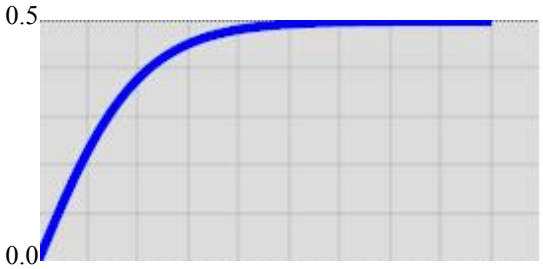


Figure 4.9 The sigmoid function that was used to saturate amplitudes of LM and AM channels. Parameter v controls the saturating rate or steepness of the function.

Overall, there are 3 free parameters to be determined: g is the multiplier of the second-order channel, v adjusts the steepness of the saturating function and σ^2 prevents division by zero in the contrast gain control stage. K is an overall scaling factor making the system output fall into the region of human data. Noting that the maximum output prior to K is 1, K was fixed at 4 to match the human data presented in Schofield et al. (in press).

Data from a haptic matching experiment was provided by Schofield (private communication; Schofield at al., in press) to fit those parameters. The depth amplitudes in the data were used to measure the strengths of corresponding shading components. In order to obtain $LM(x, y)$ and $AM(x, y)$, images that were used in the haptic experiment were regenerated in a similar way to the stimuli described in previous two chapters, except that modulations were carried by binary noise instead of Gabor patterns. In each image the contrast of the luminance modulation at each orientation was fixed at 0.2, and the modulation depth of the amplitude modulation at each orientation was varied from 0.0 to 0.4. *LM* and *AM* signals had 6.5 cycles per image. The scale of binary noise was two-pixel wide thus the fundamental frequency of a square wave made by two adjacent white and black noise samples was approximately 128 cycles per image. Examples of these images are shown in Figure

4.10 and Figure 4.11. $LM(x,y)$ in equation 4.8 was produced by filtering the image with a Gabor filter tuned to 6.5 cycle/image frequency on ±45º orientations. $AM(x,y)$ was obtained by implementing a FRF model with a second-stage filter tuned to the modulation frequency and four first-stage filters all tuned to 128cycle/image spatial frequency but each tuned to 0º, 45º, 90º and 135º. Gabor filters' bandwidths were all fixed to 1.5 octaves, consistent with V1 cells (De Valois, Albrecht & Thorell, 1982).

The search for optimal parameters was done by implementing the function *fminsearch()* iteratively in MATLAB subject to a cost function defined by the squared difference between the model responses and the data. The parameter set which resulted in the least cost values is as follows:

g =70 the multiplier of second-order channel

v =1.5 the steepness of the saturating function

σ =0.13 prevents division by zero

Note that g is not the overall gain of the second-order channel. The large value of g means that the signal strength in the second-order channel after the nonlinear rectification is so small that the signal has to be amplified to meet the requirements. Given equal strengths of LM and AM both at modulation depth of 0.1, the response of AM channel is about $1/5^{th}$ of that of LM channel. This ratio is broadly consistent with psychophysical (Schofield & Georgeson, 1999) and physiological data (Ledgeway et al., 2005), if a little low.

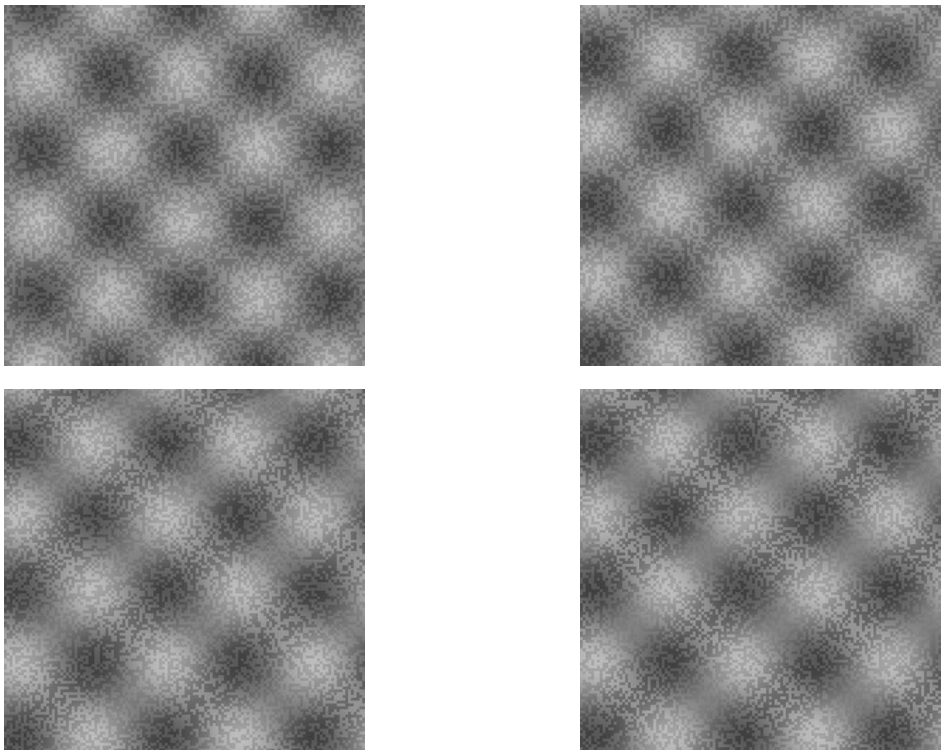


Figure 4.10 images that contain the orthogonal mix of LM+AM and LM-AM. The strength of LM was fixed to 0.2. The strength of AM was varied from 0 to 0.1 (left to right on the top row) and 0.3 to 0.4 (left to right on the bottom row). Only half of the total cycles are shown here for demonstration purposes.

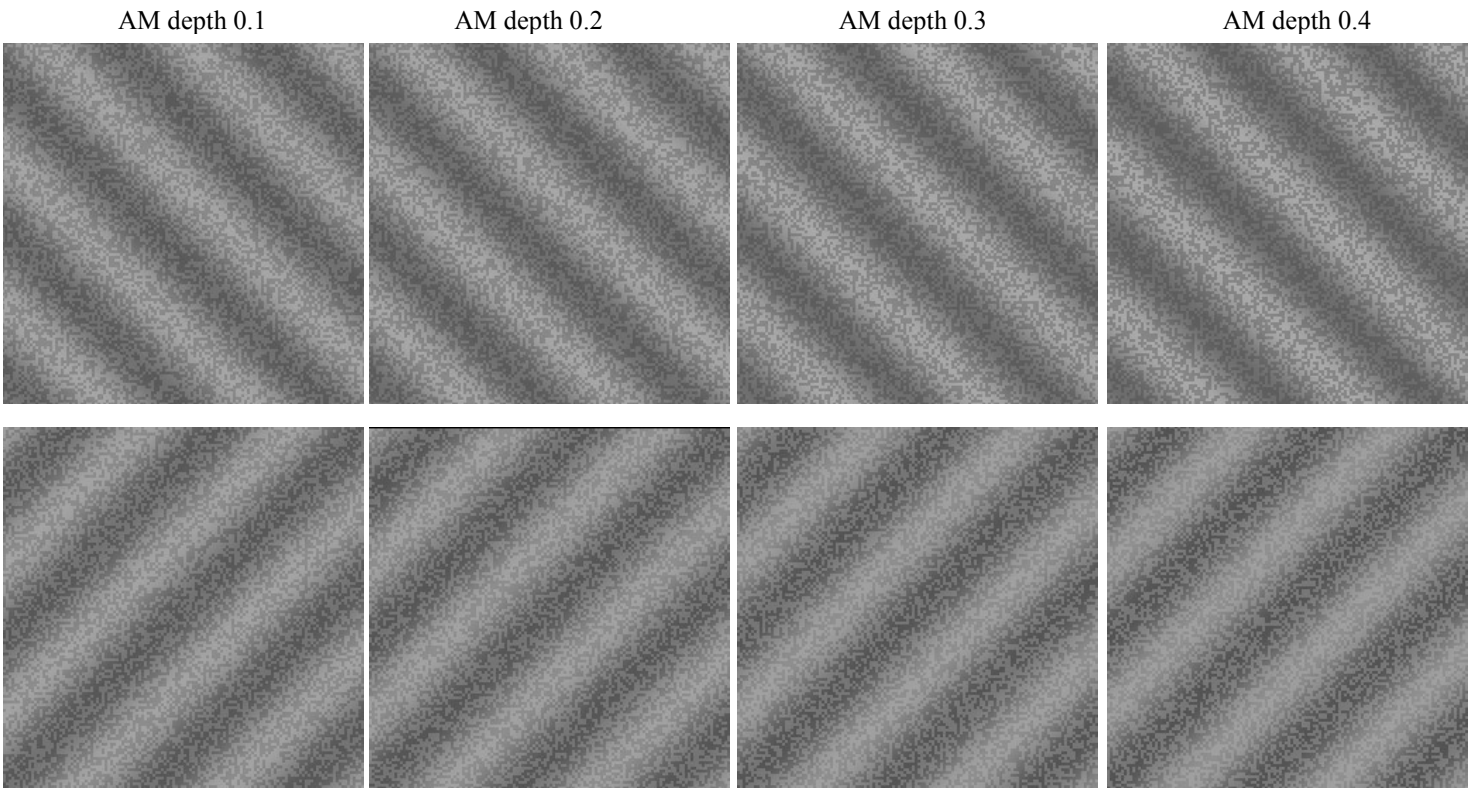


Figure 4.11 Images that contain LM+AM only (top row) and LM-AM only (bottom row). From left to right, the strength of AM was varied from 0.1 to 0.4. Only half of the total cycles are shown here for demonstration purpose.

Model implementation: version two

The nonlinear transfer function of simple cells can be largely accounted by a divisive normalization among the cells (Carandini et al., 1999). Similarly, psychophysical evidence suggests that the nonlinearities associated with second-order vision are caused by similar normalizations among channels (Graham & Sutter, 2000). In the second implementation, I removed the nonlinearity from the intermediate rectifier (making it piecewise linear) as well as the nonlinear transfer functions at the end of both types of channels. Another contrast gain control network was added between the first-order channel and the first stage filters in the second-order channel. This early normalization is expected to make both first- and second-order channel outputs nonlinear but the resulting transfer functions are not necessarily the same for the two channels. Figure 4.12 shows the modified model structure.

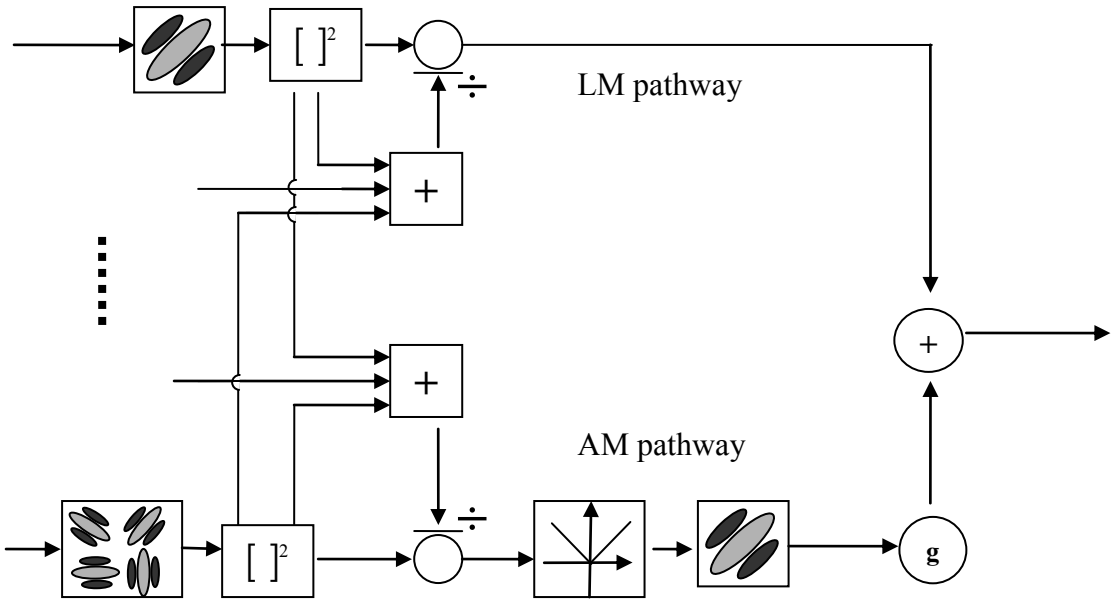


Figure 4.12 Model implementation version two (in place of Fig 4.8b). Explicit nonlinearities are removed and additional early normalization network are added. Early normalizations take place among simple cells before information is passed on to the second filtering stage.

The fitting was done analytically. $Var(\ )$ in Equation (4.7) was replaced with a more general operator representing the magnitude of the response in each shading component:

$$C_{Ri} = K\frac{C_{ri}}{\sigma^2 + \sum_j C_{rj}} \qquad (4.10)$$
$$C_r = [Mag(r)]^2$$

Since LM signals and AM signals were combined either in-phase or out-of-phase, the magnitude of the linear combination of LM and AM responses can be written as:

$$Mag(r) = Mag(rLM) \pm g \times Mag(rAM) \qquad (4.12)$$

where rLM and rAM are responses of the LM pathway and AM pathway respectively.

The magnitude of rLM is the direct product of the early normalization and can be expressed in a format similar to Eq 4.7 and 4.10:

$$Mag(rLM) = \frac{C_{Li}}{\sigma_E^2 + \sum_j w_j C_{Lj} + \sum_f w_f C_f} \qquad (4.13)$$
$$C_L = [Mag(L(x,y))]^2$$

where σ_E^2 is the equivalent of σ in Eq 4.7 and Eq 4.10 for the early normalization, $L(x,y)$ is the linear response of a simple cell. w is the weight for the contribution of other simple cells to the normalization pool. w should vary with the spatial frequency (w_f) and orientation tuning (w_j) of the contributing cells (Foley, 1994). In this implementation, the weights for cells tuned to the same spatial frequency as the excitatory cell were fixed to 1 (w_j =1), regardless of their orientation tuning. This is to reflect that the orientation tuning in the inhibitory term in the denominator of Eq 4.13 is very broad (Foley, 1994) and that substantial suppression can still be found in cross-orientation masking paradigm where mask and target differ significantly in orientation (Meese & Holmes, 2007). Studies concerning the weights for cross-frequency interactions are rare so w_f was made a free parameter and depended on the differences in spatial frequency between the channels. Empirically the inhibitory power of a simple cell over a given excitatory cell is determined by the similarities of the two cells: the excitatory cell receives most inhibition from cells similar to itself (but see Meese & Hess, 2004). C_f was the mean response of simple cells to the noise carrier thus its value was chosen to be the noise contrast.

The next step is to analytically derive $Mag(L(x,y))$. LM signals were generated using the formula below:

$$LM = I_0(1 + nN + mM) \qquad (4.14)$$

where I_0 is the mean luminance in the look-up table, N is the pattern of binary noise, n is the noise contrast, M is the modulation signal, in this case sinusoidal grating, m is the modulation depth. Suppose that noise does not fall into the passband of the filter tuned to the modulation. That is, the term nN in Eq 4.14 will not contribute to $Mag(L(x,y))$. Let us introduce another symbol denoting the signals which will contribute to $Mag(L(x,y))$:

$$LM' = I_0(1 + mM) \tag{4.15}$$

Suppose that linear filters are perfectly DC balanced. Then the magnitude of the response of a linear filter tuned to LM is a linear function of the signal strength of LM' (one without binary noise). Here I take the difference between the maximum and minimum values of LM' as a measure of the signal strength. That is:

$$Mag(L(x,y)) = k(LM'_{Max} - LM'_{Min}) = 2kI_0 m \tag{4.16}$$

here k is a constant. For the sake of simplicity, I take the assumption that $2kI_0 = 1$. So Eq 4.16 can be rewritten as:

$$\begin{aligned} &Mag(L(x,y)) = m \\ &2kI_0 = 1 \end{aligned} \tag{4.17}$$

Note, however, that Eq 4.17 *does not* mean that the magnitude of the linear response of a simple cell to its preferred optical pattern is the contrast of that particular pattern. The linear response is also dependent on the mean value I_0. We can substitute Eq 4.17 back to Eq 4.13 to get $Mag(rLM)$.

To derive the magnitude of rAM, we start from the linear responses of simple cells at the first filtering stage. AM signals were generated using the formula below:

$$AM = I_0(1 + nN \times (1 + mM)) \tag{4.18}$$

Note that the addition of low frequency luminance modulation to the AM signal will not go through AM pathway, assuming linear filters are perfectly DC balanced. Thus in the AM pathway, the signal contributing to $Mag(rAM)$ is exactly AM. It is clear from Eq 4.18 that the minimum and maximum amplitudes in the AM signal are:

$$\begin{aligned} amMin &= n(1-m)\times I_0 \\ amMax &= n(1+m)\times I_0 \end{aligned} \tag{4.19}$$

Hence the corresponding minimum and maximum signal strengths are twice Eq 4.19:

$$\begin{aligned} sMin &= 2I_0 n(1-m) \\ sMax &= 2I_0 n(1+m) \end{aligned} \tag{4.20}$$

According to Eq 4.16 and Eq 4.17, the minimum and maximum magnitudes of the linear responses of the first stage filters are:

$$\begin{aligned} lMin &= k \times sMin = 2kI_0 n(1-m) = n(1-m) \\ lMax &= k \times sMax = 2kI_0 n(1+m) = n(1+m) \end{aligned} \tag{4.21}$$

These linear responses will go through the same normalization network as does C_L in Eq 4.13. Let $rMin$ and $rMax$ denote the normalized minimum and maximum responses of simple cells in the AM pathway. Then we have:

$$\begin{aligned} rMin &= \frac{C_{\min}}{\sigma_E^2 + \sum_f w_f C_f + C_{\min}} \\ rMax &= \frac{C_{\max}}{\sigma_E^2 + \sum_f w_f C_f + C_{\max}} \\ C_{\min} &= [lMin]^2 \\ C_{\max} &= [lMax]^2 \\ C_f &= [Mag(L(x,y))]^2 \end{aligned} \tag{4.22}$$

σ_E is same as in Eq 4.13. w_f also had the same value in the practice because only two frequencies were involved and w_f was equal for the two interactions. The magnitude of the response after the second stage filter can be determined by the formula below:

$$Mag(rAM) = k\frac{rMax - rMin}{2} \quad (4.23)$$

The derivation of Eq 4.23 is explained graphically in Figure 4.13. k was absorbed into gain term g in Eq 4.12. In total there were 4 free parameters: σ, σ_E, g, w. The search for optimal parameters was done in the same way as for version one. The parameter set which resulted in the least cost values is as follows:

$\sigma_E = 0.029, \sigma = 0.24, g = 3, w = 0.23$

Again, g is not an overall gain of the AM pathway but a parameter adjusting the relative strength of the two pathways. Given LM and AM of equal modulation depth, the response of AM pathway is about 1/10th of LM pathway under these parameter setting, matching the relative sensitivity to the cues for noise contrast 0.1 as found psychophysically (Schofield & Georgeson, 1999).

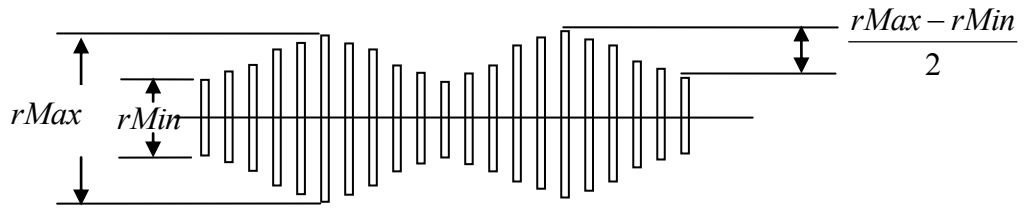


Figure 4.13 Illustration of the derivation of Mag(rAM). Responses of simple cells in the AM pathway are rectified about the mean (the line in the middle). Thus the signal that will be picked up by the second stage filter is the variations in amplitudes. The contrast of this variation is half of the difference between the maximum and minimum responses.

4.5 Model predictions

The model was implemented in two ways as discussed in the last section. The results were compared with experimental data described in previous chapters.

4.5.1 The perceived depth as a function of AM depth

Figure 4.14 shows experimental data and the model prediction. Details of the haptic experiment can be found in Schofield et al. (2009 and in press) but is briefly described

as follows. Note that the current author did not collect this data although the model described here is reported in the paper. Observers felt a surface undulating in one direction only using a haptic force feedback arm. The frequency of the surface matched that of the gratings with the peaks of the surface matched to each observer's preferred location. Observers were asked to adjust the amplitudes of the haptic surface to match perceived surface depth. Amplitudes of the haptic surface were recorded as a measurement of perceived depth amplitude.

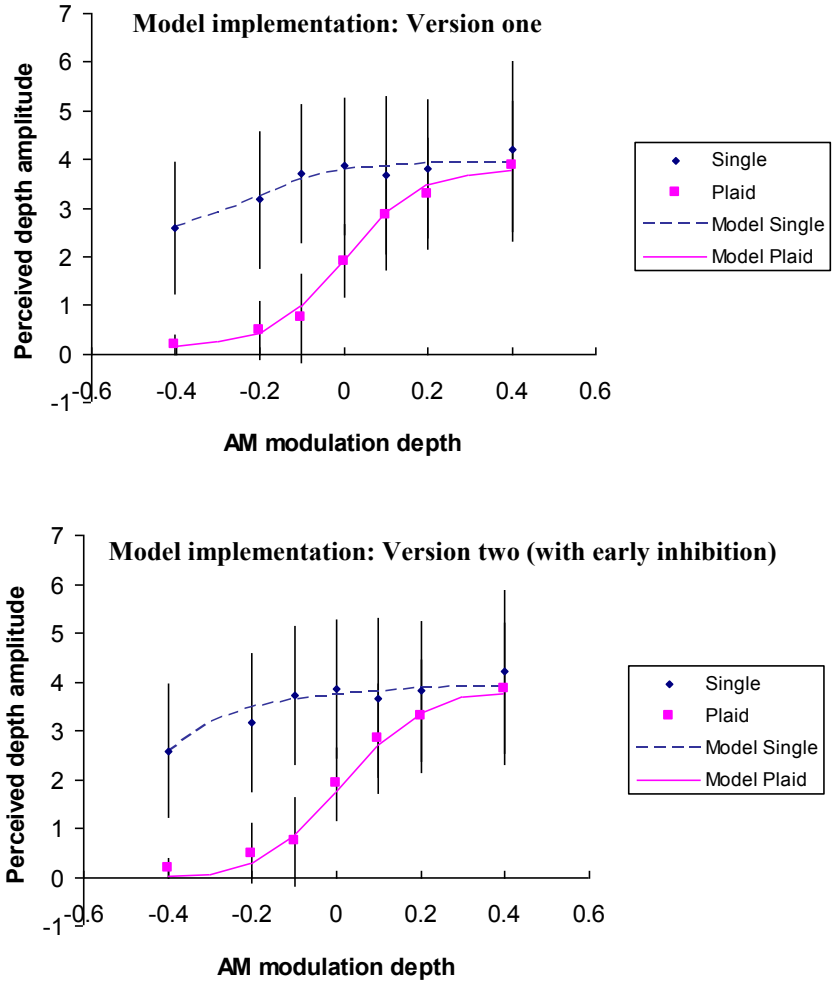


Figure 4.14 Model predictions for perceived depth as a function of AM strength. Experimental data from Schofield et al (2009; in press) are provided to facilitate comparisons. Human perceived depth amplitudes for single oblique and plaids stimuli were given by diamond and squares symbols respectively. Model predictions are shown by the lines with the dashed line representing single oblique and the solid line plaids.

Figure 4.15 shows the resulting images based on the output of the model version one in response to some of the test images in Figure 4.10 and 4.11. Output images for model version two would look very similar if presented here. As discussed earlier on, the output of the classification unit represents strengths of shading component at various frequencies and orientations. Thus the shading image can be generated by rescaling each shading component to its normalized magnitude. Results show that when AM was weak (2a), the LM+AM stripe was preferentially weighed but the LM-AM stripe still produced as identifiable shape-from-shading component (2b). However when AM was strong (3a), LM+AM completely dominated the output and the LM-AM stripe was almost completely flattened (eradicated from the output image) (3b). For a single oblique, the LM-AM stripe (4a) still gave rise to a shading map on its orientation (4b), consistent with the observation that a LM-AM alone is perceived more corrugated than when it is in a plaid.

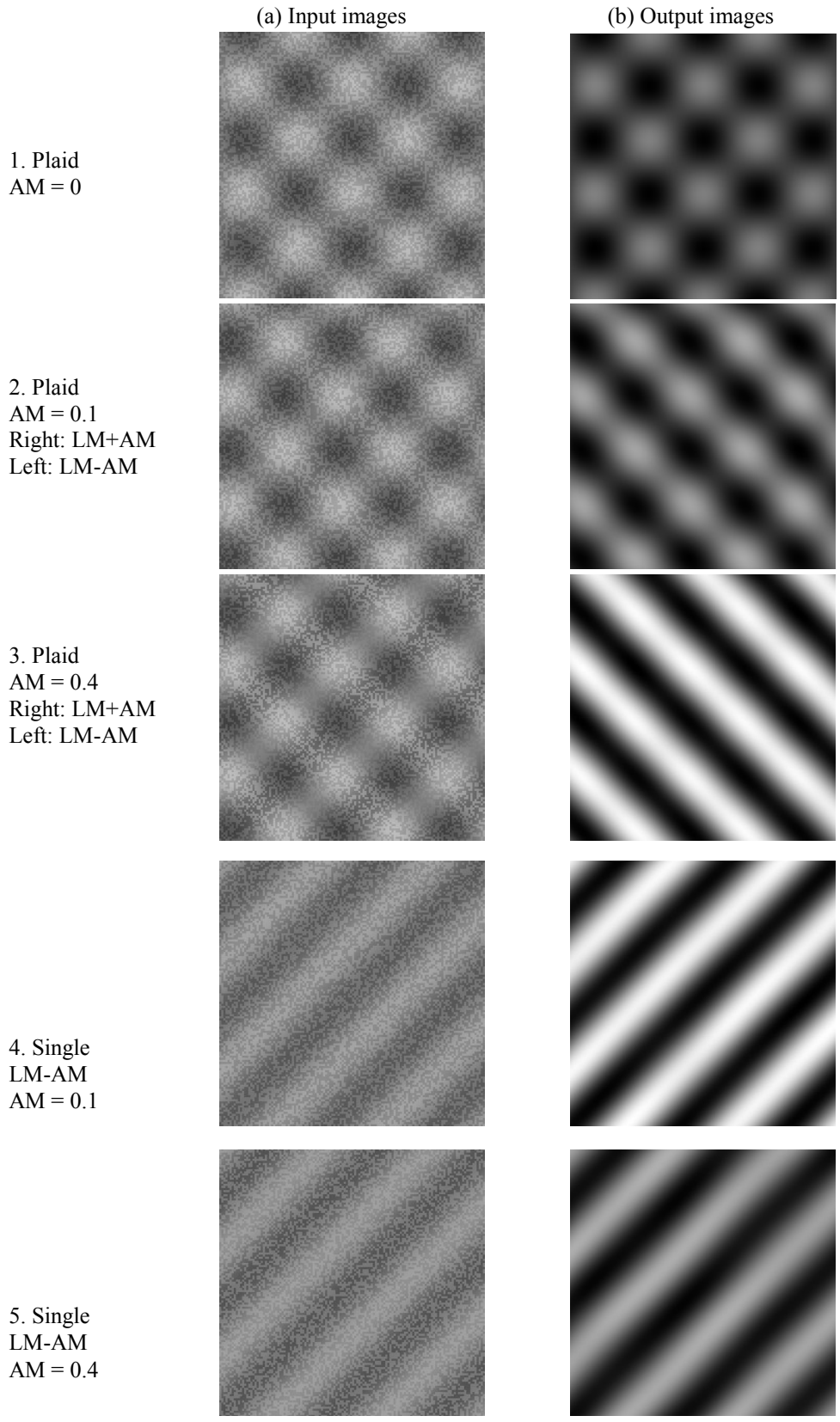


Figure 4.15 Input images (only half of total cycles are shown here) and images generated from the output of the model.

4.5.2 Perceived depth as a function of carrier frequency

Version one of the model was applied to the stimuli presented in chapter 3. Version 2 was not exposed to these stimuli as a lot of the assumptions made in implementing this version do not hold for these stimuli. For example, the assumption leading to Eq 4.15 no longer holds because low frequency carriers will also go through the LM pathway. Moreover, psychophysical evidence suggest that the cross-frequency weighting term w would need to vary for carriers with different spatial frequencies (Meese & Hess, 2004) leading to a lot of additional free parameters that would weaken the predictive power of the model.

The stimuli described in chapter 3 were processed by version one of the with similar parameter settings except that the first-stage filters were tuned to the dominant frequencies of the constituent Gabor patterns. The inter-connection between the two filter stages were established according to the rules described in section 4.2.3. As the carrier frequency approached the modulation frequency, interference due to the carrier leaking through the 'high' frequency LM channel was no longer negligible. The model output is drawn in Figure 4.16.

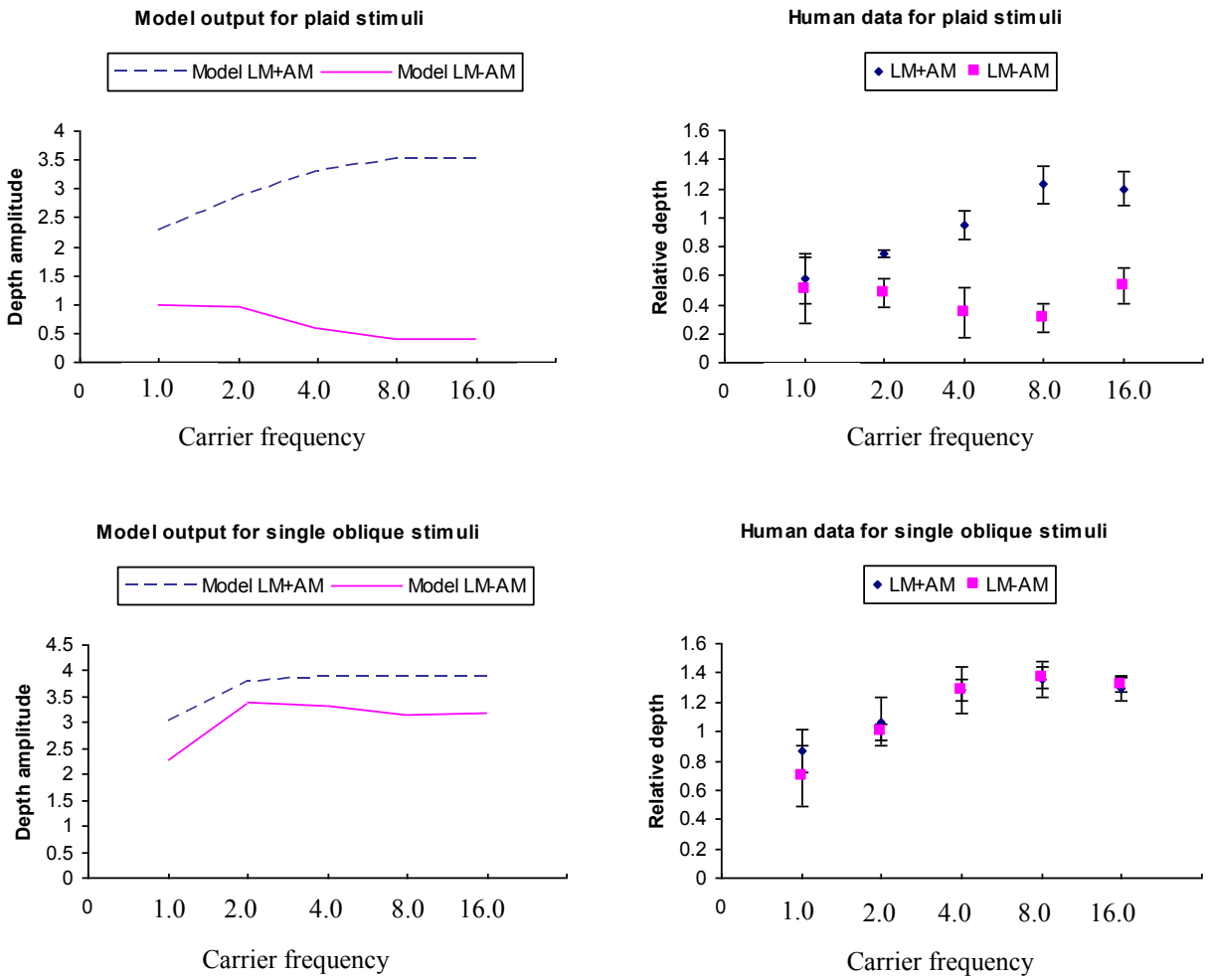


Figure 4.16 Comparison of model predictions and human performance.

The model captures the overall trends in the probe point data (see Fig 4.16). For example, the model output for single components is reduced at lower frequency carriers regardless of how LM and AM are combined. For the plaid configuration, the model output of LM+AM and LM-AM is well separated when carrier frequencies are high but start to merge as the carrier frequency is reduced. Direct comparison between human performance and model predictions are difficult for the probe point experiment for reasons outlined below. Thus in the next section I derive a suitable conversion algorithm that allows such a comparison.

4.5.3 Assessment of the model prediction

In assessing the model's predictions it is important to note that the two-point probe task (See chapter 3) and the haptic match task used by Schofield et al. (in press) are very different. The model output C_{Ri} cannot be used to directly predict the perceived depth in the two-point probe task. A notable distinction between the two experiments is that in a two-point probe task, observer's decision regarding to which point on the test orientation appeared closer could be affected by the grating on the other orientation in the plaid condition. In the probe task two dots were placed with a small offset along the test orientation but no net offset along the orthogonal orientation. However observers could base their response on the non-test grating instead of the test pattern. Further the probe tasks measures relative depth not absolute depth and may also be affected by uncertainty such that estimated depth amplitudes are a measure of how reliable the depth percept is rather than perceived depth per say. Therefore the data from the two experiments cannot be compared on a piecewise basis and the model output (designed to match the haptic data) should not be compared directly to data obtained in a two-point probe task.

It is possible however, to make a quantitative link between the model output and the human data from the two-point probe task, which requires estimating the distribution of the latter. The following subsections will introduce further corrections needed in order to link the model output with human performance in the two-point probe task.

It is a common practice to model human responses by a joint likelihood function. In the problem of interest, observers' responses were dependent on three source of information: the gratings along both test (T) and foil (F) orientations and any texture

(N) that leaks through the LM filter. When the modulations were carried by high frequency textures, the two gratings were the major contributor to observer's response. Let us consider the case when the testing grating was the only information available. Assume first that an observer could tell the offset with the probability *p* at the three testing positions within half cycle of a sinusoidal grating (position B, C and D in Figure 4.17). Then the response to a given trial is given by a random variable *x* which obeys the Bernoulli distribution:

$$\Pr(x=1)=p, \Pr(x=-1)=1-p, E(x)=2p-1, Variance(x)=4p-4p^2, 0.5<p<1$$

The accumulated score for all three positions is a random variable *X* which is a sum of all attempts made in the $8\times 3=24$ repetitions. According to the central limit theorem, the distribution of X can be approximated by a Gaussian distribution with $\mu=48p-24, \sigma^2=96p-96p^2$. For position A and E, the accumulated score for each position Y follows a similar Gaussian distribution but with $p=0.5$. That is $Y\sim N(0,8)$. Thus the surface height that an observer reported would follow a Gaussian distribution. During the original data analysis, the final surface height was divided by half of the total number of samples within a cycle and approximated by the amplitude of a fitted sine function. So the final reported height (see Fig 4.17) is given by, $h_T=\frac{height}{2\times 4}=\frac{X+Y}{8}: h_T\sim N\left(6p-3, 0.5-1.5(p-0.5)^2\right)$, but note that this only holds if the test grating were the only information available.

When viewing single gratings carried by high frequency textures, the reported depth values can be well described by h_T. Thus *p* can be estimated by solving the equation below:

$$Mean(h_T)=6p-3=\overline{H} \qquad (4.24)$$

where $\overline{H}$ is the average perceived depth in the human data for single gratings carried by 8.0c/d and 16.0c/d textures. Solving Eq 4.24 yields $p = 0.723$, which literally says that, on average, the chance of human observers making correct decisions on the depth comparison between two adjacent positions is 72.3% even when no other sources interfered with their decision.

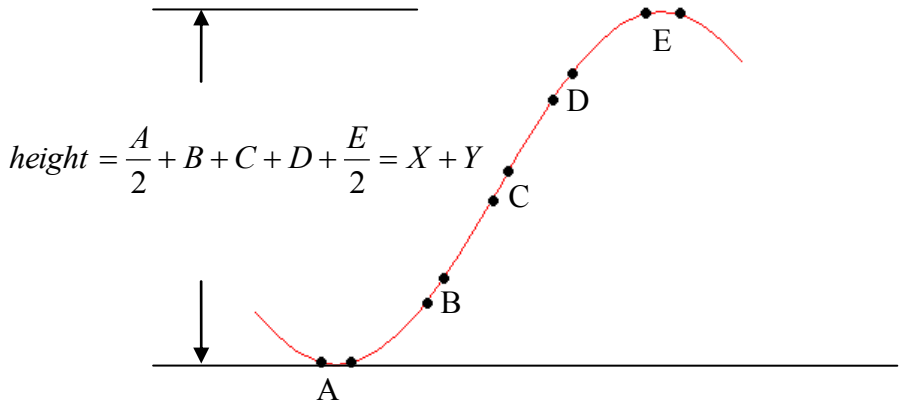


Figure 4.17 Surface height was computed by discrete integration along the testing direction. The sinusoidal trace represents a sinusoidal surface that an observer perceives from a sinusoidal grating.

If the foil grating is the only source of information, then observer's attempts at each trial will again obey the same Bernoulli distribution as x due to the zero offset between the two positions in the direction of the foil grating. Thus the score for each position after eight repetitions also roughly followed the same Gaussian distribution as X. As a result, the reported surface height in this case (h_F) would follow a similar Gaussian distribution to that of h_T but with $p = 0.5 : h_F \sim N(0,0.5)$.

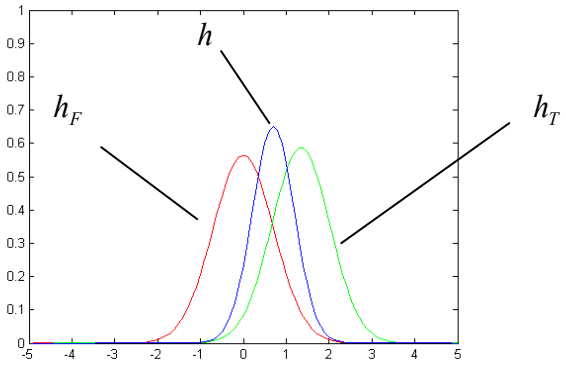


Figure 4.18 Human behavioural response is a joint distribution of all source of information including testing grating and false grating (distribution of the response based on textures is not shown). The distribution of h is scaled for demonstration purpose.

The behavioral response h is a joint distribution of h_T and h_F (plus h_N when textures start to interfere), as shown in Figure 4.18. The probability density function (pdf) for h has a shape close to Gaussian lying between h_T, h_F. Its mean is a linear combination of that of h_T and h_F. The weights are inversely related to the variance of each distribution. Figure 4.18 describes the situation where the signal strengths of the two sources are equal. When they are not equal, the mean of the grating with greater signal strength should be weighted more. Taken together, the mean of h can be obtained using the formula below:

$$\mu_h = w_T \mu_T + w_F \mu_F$$
$$w_T = \frac{\sigma_F^2 C_T}{\sigma_T^2 C_F + \sigma_F^2 C_T}, w_F = \frac{\sigma_T^2 C_F}{\sigma_T^2 C_F + \sigma_F^2 C_T} \quad (4.25)$$

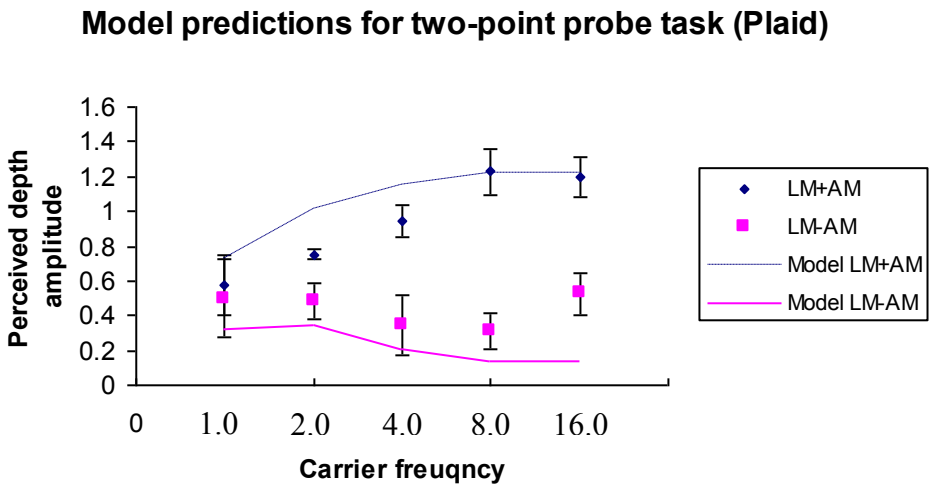
where the C s are the channel response for either test gratings or foil gratings, the μ s and σ s are means and standard deviations of the estimated distributions.

When the carrier signal leaks though the LM filter, this signal would act as a third source of information affecting observer's response. Denoted h_N, the reported surface height if the observer only responded to the carrier texture. It is easy to see

that h_N obeys the same distribution as h_F except that the variance of h_N should be scaled properly based on its relative strength. Incorporating the influence of h_N into Eq 4.25 gives:

$$\begin{aligned}
\mu_h &= w_T\mu_T + w_F\mu_F + w_N\mu_N \\
w_T &= \frac{\sigma_F^2\sigma_N^2 C_T}{\sigma_T^2\sigma_F^2 C_N + \sigma_T^2\sigma_N^2 C_F + \sigma_F^2\sigma_N^2 C_T} \\
w_F &= \frac{\sigma_T^2\sigma_N^2 C_F}{\sigma_T^2\sigma_F^2 C_N + \sigma_T^2\sigma_N^2 C_F + \sigma_F^2\sigma_N^2 C_T} \\
w_N &= \frac{\sigma_T^2\sigma_F^2 C_N}{\sigma_T^2\sigma_F^2 C_N + \sigma_T^2\sigma_N^2 C_F + \sigma_F^2\sigma_N^2 C_T}
\end{aligned} \tag{4.26}$$

Using Eq 4.26 and the estimated distributions for h_T, h_F and h_N, a quantitative link between the model output and the human data can be established. Figure 4.19 shows the model curves as computed from Eq 4.26.



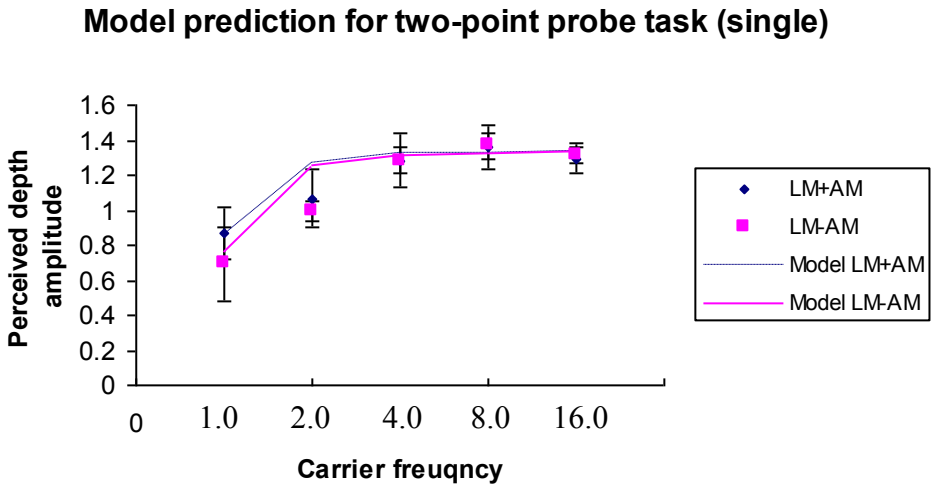


Figure 4.19 The comparison between the model prediction and the experimental data after the model data have been transformed into the same 'space' as the human data. Squares and diamonds represent human data. Predictions made by applying Eq 4.26 together with the model outputs (Fig 4.16) are represented by dashed and solid lines.

Model predictions and human data can now be compared directly. In the plaid configuration, the tendency for the depth amplitudes of LM+AM and LM-AM to merge at low carrier frequency is retained, although the perceived depth amplitude of LM-AM signals is somewhat underestimated by the model. The difference between the model cues is slightly overestimated at low carrier frequencies and this may indicate that the reduction in the signal strength of second-order vision in the model is not as strong as that in humans. For single gratings, predicted depth amplitude is high for both LM+AM and LM-AM on high frequency carriers and starts to decrease as the carrier frequency approaches the frequency of the modulation. However the decrease takes place one octave sooner in the human data. Recall that cells responsive to both first-order and second-order signals have separated pass-bands (Zhou & Baker, 1996; Song & Baker, 2006). Although the first-order pass-band is close to the second-order pass-band and both are relatively low-frequency, the first-order pass-band is often selective to slightly higher frequencies than the second-order pass-band (Zhou & Baker, 1996; Song & Baker, 2006). To reflect this in the model, the

preferred frequency of the filter in the LM pathway should be slightly higher than the second filter in the AM pathway. This adjustment would shift the point at which the carrier starts leak through the LM pathway upwards in frequency and hence predicted depth amplitude would start to decrease at relatively higher carrier frequencies.

4.6 Discussions

4.6.1 Comparisons of the two versions

The two versions of the model differ in how the nonlinearities in the second-order channel are achieved. Version one applies a deep power law to the intermediate rectifier and a nonlinear function to the channel response. Version two replaces those explicit nonlinear functions with an early normalization network. With less free parameters, version one provides a slightly better fit and can also be easily extended to predict human performance in the multi-carrier frequency experiment (Chapter 3). Version two however, provides an insight into the origins of the nonlinearities in the AM pathway and may be more biologically plausible. In comparison, version one only gives a functional description of the nonlinearity and achieves response saturation in a rather unrealistic way. However, both versions have support from human psychophysics in terms of the characteristics of the nonlinearities associated with second-order vision (Graham & Sutter, 1998; Graham & Sutter, 2000).

4.6.2 The nonlinearities in the second-order vision

The contrast transfer function measured for contrast responsive cells in cat area 18 is low for second-order contrast modulations with weaker signal strengths and expansively accelerates without saturation for stronger signals (Ledgeway et al., 2005). Ledgeway et al. (2005) favoured an explanation in which weaker signals were

suppressed by an intermediate rectifier obeying a deep power law. However the large value of the multiplier g in version one suggests that this arrangement may over-suppress the information such that it has to be amplified again by a great deal. This could risk a poor signal to noise ratio at the implementation level. In fact, from Eq 4.14 and 4.18, it is clear that the ratio of the signal strengths of AM and LM is $n:1$ (where n is the noise contrast), given that their modulation depths are equal. Thus the high threshold found for second-order vision in early studies is more likely due to the inherently weak signal strengths in the stimuli, rather than some internal attenuation process within second-order vision. In version two, the deep power law rectifier was replaced with an early normalization network, which (perhaps counter intuitively) also provides acceleration for stronger signals. To verify the validity of this early normalization network, the second-order contrast response function of the model was constructed by plotting the AM channel response as a function of AM modulation depth based the parameters obtained earlier. As shown in Figure 4.19, the AM transfer function contains an early suppression followed by acceleration but with no saturation at high modulation depths. These properties are consistent with the nonlinear characters of the contrast response function of second-order cells (Ledgeway et al., 2005).

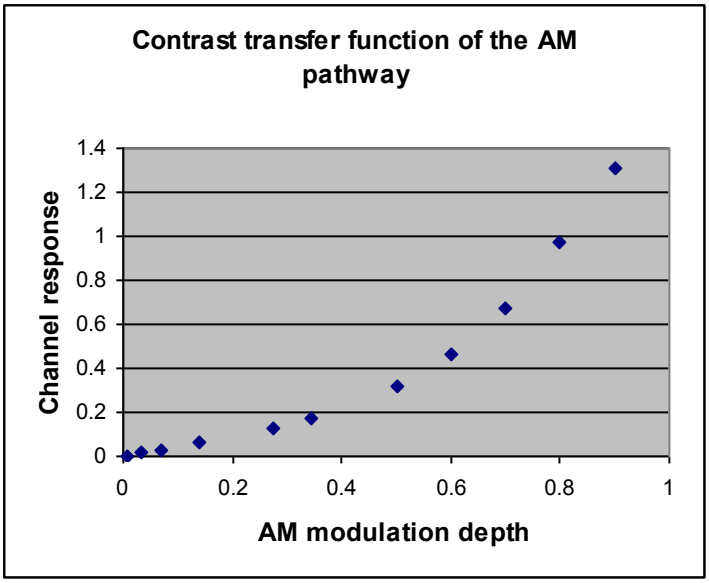


Figure 4.20 The response of AM pathway as a function of AM modulation depth.

4.6.3 The role of the model in shape-from-shading

The model described in this chapter constitutes the feature extraction and luminance classification units of the general framework proposed for shape-from-shading in human vision. The feature extraction unit decomposes the input image into different frequency and orientation bands. In the mean time, second-order features were also extracted for future use. Coefficients at each band were subject to suppression or facilitation depending on the phase relationship between the underlying luminance signal and the corresponding second-order information. The output of the model represents the *strengths* of shading components at different orientations and different frequencies. This architecture attempts to explain the neural mechanisms underlying the known phenomenon of layer segmentation (Kingdom, 2008) with respect to the AM cues. The implementation could be carried out in early visual areas being broadly consistent with known psychophysics and physiology and does not require top down control. For example, some cells in cat area 18 are responsive to both first-order stimuli and second-order-stimuli (Mareschal and Baker, 1996) and seem to sum these signals linearly (Hutchinson et al, 2007). In studies of human psychophysics, first-

and second-order channels seem to integrate at some relatively early stage while retaining their own identities (Georgeson & Schofield, 2002). In the model, first-order channels and their corresponding second-order channels were summed, reflecting these links between the two channels. There is also ecological validity for the existence of a hard-wired connection between first- and second-order channels. Responses of biologically inspired first- and second-order channels were found to correlate in natural scenes (Johnson & Baker, 2004) but the sign of the correlation may vary between images (Schofield, 2000). This observation indicates that co-varying first- and second-order signals convey valuable information in natural scenes (Schofield et al., in press). The summation between the two channels provides a solution to code this covariance thus extracting the information conveyed by the relationship between the cues.

The model presented in this chapter could lead to a useful image processing algorithm working within the spatial frequency domain. The output of each shading channel represents the strength of shading component at the corresponding frequencies and orientations. The shading image can be recovered by multiplying each base component by their strengths, similar to the inverse operation of the linear decomposition at the initial stage. Natural images often contain numerous frequency and orientation components; thus a useful algorithm would require extra frequency and orientation channels to function well. However the relationship between LM and AM can be also be exploited in the spatial domain to serve as an image processing solution to real images. Such an image processing algorithm is described in next chapter.

5. Recovering shading and reflectance information from real images using texture

This chapter presents a machine vision algorithm for separating shading components from reflectance components in greyscale images. The rule used to distinguish between these two components is similar to that used by humans to assist in shape-from-shading tasks: luminance changes that are coincident with contrast changes are likely to be due to reflectance changes whereas those that are not associated with a change in contrast are likely to be due to shading (Schofield et al., 2006). This in turn arises from the multiplicative nature of shading (see section 5.2). Examples where the algorithm has been applied to experimental and real images are provided in the end of the chapter.

5.1 Introduction

The idea of separating the retinal image into layers in human visual processing (see section 1.1.6 in Introduction) has an equivalence in computer vision —intrinsic image decomposition. The term intrinsic image, first introduced by Barrow and Tenenbaum (1978), is used to describe information resulting from independent characteristics of the scene such as illumination, object / surface shape / orientation, and surface reflectance (see also Tappen et al 2005). The major difficulty of decomposing intrinsic images resides in the ill-posed nature of the problem–solving two unknowns (illumination and reflectance) with one known variable (pixel intensity). But natural scenes often contain visual regularities that could help to constrain the problem. Attempts had been made to achieve a similar purpose before the concept was formally developed. Land and McCann (1971) proposed the Retinex theory for removing lighting effect in photos of Mondrian patterns. The central idea was that the changes

between Mondrian patches form sharp edges whereas illumination causes gradual variations in luminance. The Retinex theory discounted these gradual variations while reintegrating sharp luminance changes to obtain only the reflectance component. The earliest Retinex theory was a 1-D implementation. Later on Horn (1974) extended it to be applicable to 2-D images. The process of finding luminance changes was modelled by filtering the input image with a 2-D Laplacian filter. The identification of reflectance changes from changes by illumination was based on the same idea as the original Retinex theory. The reintegration was conducted by applying an inverse 2-D Laplacian operator. Horn's extended Retinex algorithm has become a popular framework for intrinsic image decomposition. That is, a process of reconstruction from classified luminance edges or luminance derivatives. Many later studies on this topic tend to focus on developing new rules for classifying luminance edges. For example, a few studies have attempted to retrieve intrinsic images based on correlations with hue alone (Olmos & Kingdom 2004; Funt, Drew & Brockington, 1992). In these methods, separation was based on the observation that co-incident (positively correlated) changes in hue and luminance tend to indicate a reflectance change whereas a luminance change without a co-incident change in hue tends to indicate shading. The algorithms first extract luminance gradients and hue edges from the original image. The luminance gradients are then classified as being due to shading or reflectance changes based on the existence of co-located hue edges. Having been classified as either due to shading or reflectance, luminance changes can then be reintegrated to recover the corresponding intrinsic components.

In an alternative method, Finlayson, Hordley, Lu & Drew (2006) derived an illumination-invariant representation of a colour image based on a colour-calibrated

camera. By projecting the RGB values of a pixel in an image into a 2-D chromatic space, a direction in the space can be observed on which pixels remain constant under changing illumination, for any given surface for a calibrated camera. The illumination-invariant representation provides an additional constraint to help with the disambiguation of luminance derivatives. In fact Finlayson's illumination-invariant feature is a generalization of the rules employed by classic colour-based lightness recovery algorithms discussed above but delivers better performance for outdoor scenes taken by a specific camera at the cost of the additional calibration process. The improved performance results from the fact that Finlayson's illumination-invariant chromatic feature is immune to illumination colour and that natural and artificial lights often contain colour tints that confuse other algorithms. The hue based classification methods have a degree of biological plausibility. For example, Kingdom (2003) and Kingdom et al. (2004) has shown that changes in hue can help the human vision system to determine whether luminance changes are due to changes in reflectance or shadings.

The general success of colour based separation methods is attributed to the constraint provided by the additional chromatic measurement associated with each pixel in the image. When colour is not available, constraints on pixel level are hard to determine. However illumination also causes regularities on a more global level in terms of spatial relationships between regions and such regularities have been proved helpful either in combination with colour or alone for intrinsic image decomposition. Sinha and Adelson (1993) proposed a strategy for separating reflectance from illumination in painted polyhedra. Their strategy first computed the 3-D layout of the surface by employing a number of heuristic rules applicable to 3-D objects that are made of

planar surfaces. Then each edge in an image can be assigned with an identity of either reflectance of illumination according to the 3-D layout of the object. Hence it is a typical example of using global information to constrain the local luminance changes. Tappen et al. (2005) developed an algorithm with gradient classification rules based on both hue and the spatial relationships between pixels in the corresponding greyscale image. The rules linking spatial layout to illumination were learned through a training algorithm, though the rules that were learnt were not explicit in the application.

In another approach, Li, Tan & Lin (2008) observed that some global features of textures could be used as a cue to reflectance identification in addition to colour. Given a colour image, the global feature was obtained by assessing the similarities within like-textured regions. Thus each pixel in the image was assigned with a label in terms of which texture group it belonged to and a weight indicating the probability that this association could occur. Each texture group was assumed to have a unique reflectance value. The labelling and the weights could then help to further constrain the process of luminance classification. However, the number of different reflectances (i.e. number of different patches) must be determined in advance. This algorithm is another example of using global information to constrain the ill-posed problem.

Finally, some researchers have proposed solutions to the separation problem using multiple, registered images (Weiss, 2001; Agrawl, Raskar & Chellappa, 2006). This family of methods take the advantage of having multiple measurements of image intensities under various illuminations which relieves the ill-posed problem. However, these methods require multiple images of the same scene under different illuminations

and hence are less favoured, for pragmatic reasons, than approaches requiring only a single image.

Some of the methods described above echo the human ability to attribute luminance variations to the lightness of a surface and variations in surface norms based on the global 3-D layout (Knill, 1991; Sun & Perona, 1996). In this chapter, I will present an algorithm that also employs a new, non-local texture feature to classify luminance changes. The algorithm does not require colour information and thus provides a solution for intrinsic image decomposition when colour is not available. But, similar to the method proposed by Li et al. (2008), it can be used together with local features such as colour to provide further constraints on the inverse problem of image separation. In human vision studies, image texture provides a cue for interpreting luminance modulations as either due to shading or reflectance variations in a way that it analogous to the role of hue (Schofield et al., 2006).

5.2 Generative model

Assuming Lambertian surfaces, whenever a surface is shaded, the luminance at each point $I(x,y)$ is the product of the shading $S(x,y)$ and the reflectance $R(x,y)$:

$$I(x,y) = S(x,y) \times R(x,y) \tag{5.1}$$

The goal is to recover $S(x,y)$ and $R(x,y)$ from a gray image $I(x,y)$ where texture is the dominant reflectance feature. When a texture that is purely visual (i.e. painted onto the surface) and statistically uniform is shaded, the resulting change in luminance is accompanied by a correlated change in the local luminance properties of the texture such as the standard deviation of local luminance values (Schofield et al 2006). This cue is, basically, the same as the AM variations discussed in chapters 2-4.

Dark and light pixels in the texture are multiplied by the illumination such that as illumination varies both the maximum and minimum luminances change resulting in a change in both mean luminance and in the range of luminances present. Hence dividing the image by its local mean luminance profile will give rise to an image matrix with nearly uniform luminance properties (both mean luminance and standard deviation) so long as the texture is uniform. The division process will have effectively removed any variations due to shading. However, if there is more than one texture present in the scene then dividing by mean luminance will also remove luminance changes due to reflectance changes (albedo), *but* it will not remove changes in luminance standard deviation due to differences between textures. Let us define this residual luminance standard deviation as *intrinsic texture amplitude* (intrinsic here indicates that we are dealing with an intrinsic property of the scene). Intrinsic texture amplitude (ITA) is a measure of contrast; defined as the standard deviation of the pattern divided by the mean. The aim is to separate reflectance changes from luminance changes and we can now formulate a rule for this distinction, based on ITA, which is similar to that based on hue:

Co-incident (positively correlated) changes in ITA (contrast) and luminance changes in the original image tend to indicate a reflectance change whereas a luminance changes without a co-incident change in ITA (contrast) tend to indicate shading.

By applying the generative model in reverse, one can determine the origin of any luminance change in an image, and this information can be used to recover the intrinsic properties (images) for the scene. I have chosen to isolate luminance changes due to shading first and then apply an inverse method to recover the shading component. A graphical description of this process is illustrated in Figure 2.1.

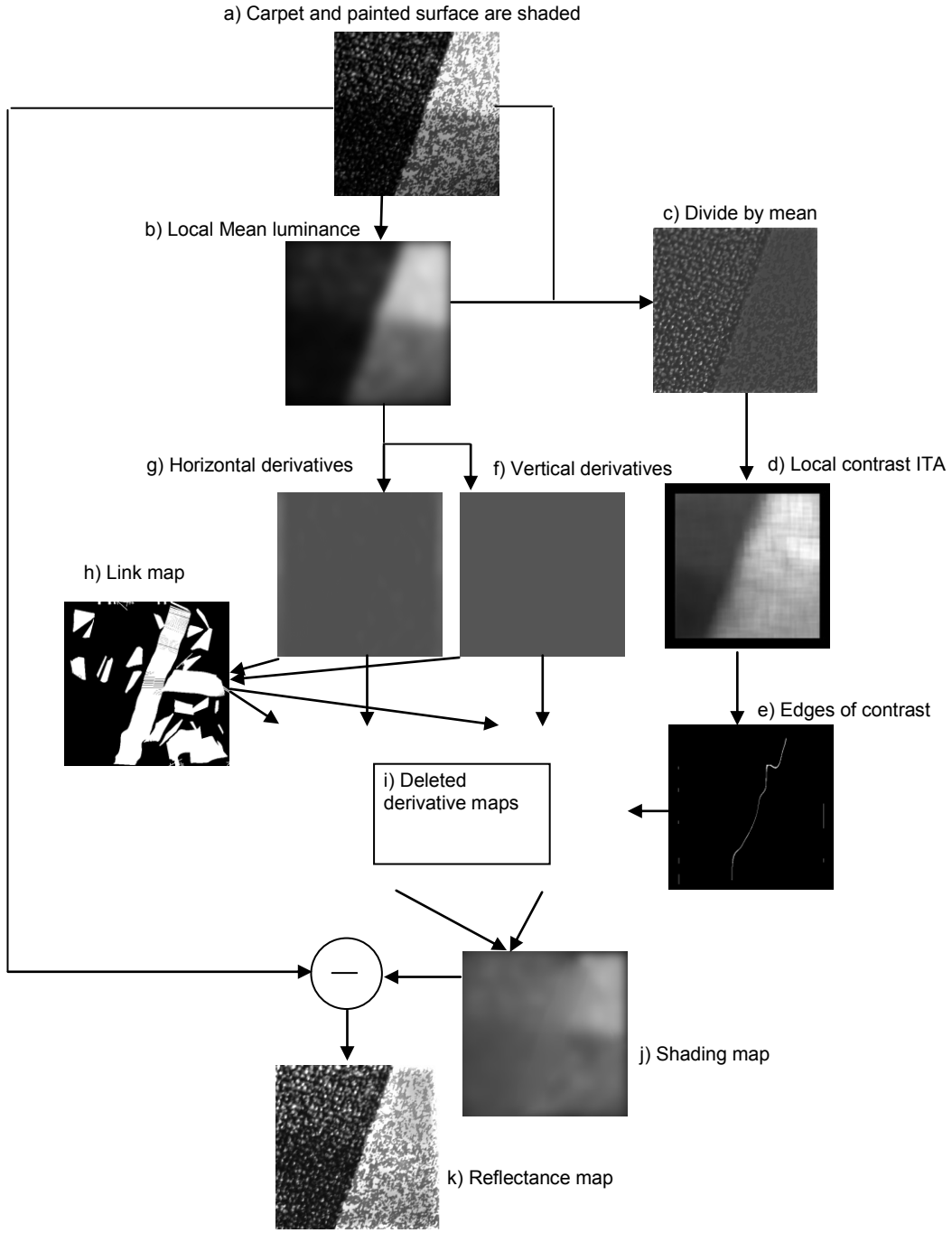


Figure 2.1 Graphical illustration of decomposing intrinsic images based on Intrinsic Texture Amplitude (ITA). The original image (a) is decomposed into its mean luminance (b) and (c) resulting from dividing (a) by (b). Local contrast (d) is calculated from (c) and its edge is extracted to form (e). Partial derivatives (g) and (f) are then calculated from (b). These partial derivatives are linked to produce a link map (h) according to the region that the edge spans. (g) and (f) are classified according to information in (e) and (h) to give (i), from which the shading image (j) can be computed. Subtraction (j) from (a) will give (k).

5.3 Image pre processing: low pass filtering

It is believed that shading signals in natural images normally present a low spatial frequency profile (Land & McCann 1971; Horn 1974). While preserving most shading signals in the image, this step removes the fine luminance changes due to texture elements for future computational efficiency. As will be discussed later, every luminance change will go through a classifying process. Reducing the number of luminance changes to be processed, by removing those which are unlikely to be due to shading, increases the efficiency of the classification process. The original image is filtered with the use of a normalised Gaussian kernel such that the filtering process does not introduce any luminance scaling. The resulting image is one that mainly contains large scale luminance variations due to either shading or changes in texture. This process is described by Equation 2.2:

$$I_{blur}(x,y) = I(x,y) * G(x,y) \tag{2.2}$$

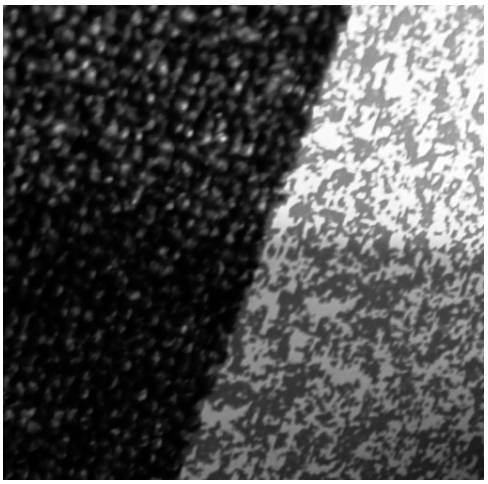
5.4 ITA and its variations

Recall that the term ITA refers to the standard deviation of local luminance values after dividing the original image by its local mean luminance map. If we denote $I_{div}(x,y)$ as the image matrix after the division, an operator mask can be generated to move continuously across $I_{div}(x,y)$ to calculate standard deviation (i.e. ITA) based on overlapped regions at each point. Equation 2.3 describes this operation:

$$\begin{aligned} I_{div}(x,y) &= \frac{I(x,y)}{I_{blur}(x,y)} \\ ITA(x,y) &= I_{div}(x,y) \circ f_{std} \end{aligned} \tag{2.3}$$

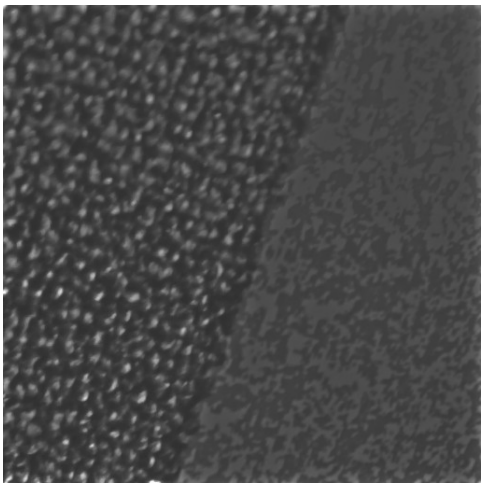
where $I_{blur}(x,y)$ is the low pass filtered image after conducting equation 2.2 and f_{std} is the operator mask calculating standard deviation of luminance values within the size

of the mask. The effect of dividing an image by its local mean luminance is illustrated in Figure 2.2.



(a)

(b)



(c)

Figure 2.2 Effect of dividing an image by its low pass profile. (a) original image where two texture patches are shaded $I(x,y)$. (b) $I_{blur}(x,y)$ low pass profile of (a). Note that the two large variations in luminance are not distinguishable. (c) $I_{div}(x,y)$ (a) has been divided by (b).

As explained in section 2.2, ITA is invariant to large scale luminance changes but preserves the property of the underlying texture and therefore can be used as a feature to find boundaries of different textures. Note that ITA is not the reflectance component of the original image. The division removes albedo changes as well as illumination effects. These boundary locations indicate where reflectance changes

take place in low pass filtered images. For example in Figure 2.2, large scale variation in ITA arises from the boundary of the carpet and the painted texture.

The next step is to locate these variations in ITA which is equivalent to detecting edges in the ITA map. This is accomplished by finding zero-crossings of the second derivatives of the ITA map. Subject to thresholding and appropriate choice of filter size, texture segmentation can be achieved which is invariant to illumination. The accuracy of this boundary localization process is not crucial for reasons that will become clear later. Let's term the resulting edge-map $TxtEdge(x, y)$. Recall from section 2.2 that any luminance change without a co-incident change in ITA tends to indicate shading. Thus edge-map $TxtEdge(x, y)$ is useful for disambiguating luminance changes in $I_{blur}(x, y)$.

It is worth pointing out that the actual edge detection algorithm to be employed is not critical. In fact, any method that is able to segment $ITA(x, y)$ will suffice. The key point is to find locations where one texture-defined patch abuts another while disregarding any illumination effects. In the case of this implementation, it was ITA that served as a defining feature for locating the genuine texture boundaries.

5.5 Classification of luminance changes

Ideally, each luminance change will be labelled as due to shading if there is no corresponding edge in $TxtEdge(x, y)$. However, the accuracy of the edge locations obtained at previous stage cannot be guaranteed. Furthermore, due to the characteristics of low spatial frequency variations, luminance changes induced by a

texture boundary in $I_{blur}(x, y)$ tend to span in the direction orthogonal to the actual boundary. All these factors combined suggest that it make more senses to discount not only the luminance changes which have accompanied edges at the exact locations in $TxtEdge(x, y)$ but also the luminance changes close to or associated with them. This problem can be solved by constructing a group of links each of which consists of a set of associated luminance changes. If any element in a link is labelled as NOT due to shading, the entire link is rejected as candidates for shading. The proposed linking rule is that luminance changes, which form a smooth ramp in the gradient direction, are grouped to form one link. Figure 2.3 illustrates this idea.

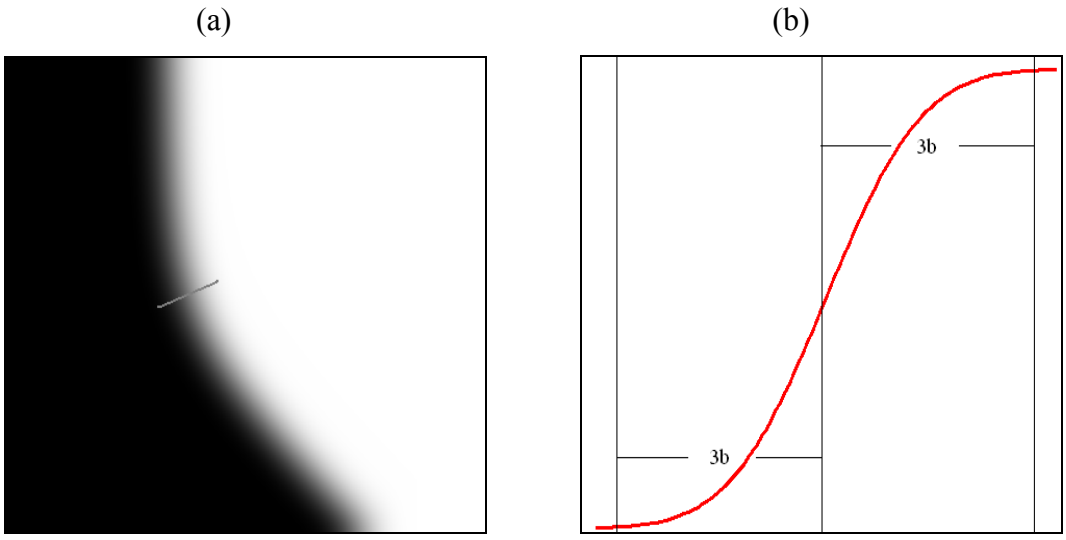


Figure 2.3 Illustration of one link consisting of associated luminance changes: (a) a Gaussian blurred curvature edge. The gray bar runs across the edge in the direction of its local norm (gradient). (b) Luminance values on the gray bar in (a). The gray bar is bounded and its length should not exceed the width of the edge. In this demonstration, all the luminance changes falling inline with the gray bar should be linked together.

Step1 The widths of Gaussian edges

An edge width estimation method (Georgeson, Freeman & Hess, 2007; Lindeberg 1998; Lindeberg 1993) was used in order to construct the links described above. Traditionally, features such as edges in an image can be extracted using Gaussian

derivative filters with appropriate scales (Georgeson et al 2007; ter Haar Romeny 2003; Lindeberg 1993). In the case of one dimensional signal, such filters can be expressed as

$$L_n(x,\sigma) = I(x) * \frac{\partial^n G(x,\sigma)}{\partial x^n}$$
$$G(x,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\sigma^2}\right) \tag{2.4}$$

where n represents the order of the derivative operators. Without any prior knowledge about the scale of a feature, the choice of the filter scale can be arbitrary. Lindeberg (1998) has devised a framework for edge detection with automatic scale selection. Within Lindeberg's framework, responses of image features are multiplied by a scale-dependent normalization factor σ^γ such that the normalized responses will peak at true scales of the features. Equation (2.4) then becomes

$$N_n(x,\sigma) = \sigma^\gamma I(x) * \frac{\partial^n G(x,\sigma)}{\partial x^n} \tag{2.5}$$

where γ can be set equal to $\frac{n}{2}$ when applied to Gaussian blurred edges, as is a reasonable assumption to make for $I_{blur}(x,y)$. Georgeson et al. (2007) used this method to explain how human vision system might code the blur of a given Gaussian edge in one dimension. More precisely, they have implemented the third derivative response $N_3(x,\sigma)$ with a more biologically plausible model to locate the position of a Gaussian edge as well as estimate its blur (i.e. width). Since convolution and differentiation are linear operators, they can be applied in any order. Thus $N_3(x,\sigma)$ can be expressed as:

$$
\begin{aligned}
N_3(x,\sigma) &= \sigma^{\gamma}\partial I(x,b) * \frac{\partial^2 G(x,\sigma)}{\partial x^2} \\
&= \sigma^{\gamma} G(x,b) * \frac{\partial^2 G(x,\sigma)}{\partial x^2} \\
&= \sigma^{\gamma} \frac{\partial^2 G(x,s)}{\partial x^2} \\
s &= \sqrt{\sigma^2 + b^2}
\end{aligned}
\tag{2.6}
$$

where b is the standard deviation of the Gaussian function which had generated the edge (assuming the edge is Gaussian). Note that the two Gaussian expressions become one from step 2 to step 3 because Gaussian variances add under convolution. Applying some basic calculus to (2.6) and setting $x = 0$ (edge location) gives

$$
N_3(0,\sigma;b) = \frac{-\sigma^{\gamma}}{\left(\sigma^2 + b^2\right)^{1.5}\sqrt{2\pi}}. \tag{2.7}
$$

From (2.7), it is apparent that $N_3(0,\sigma,b)$ peaks when $\sigma = b$, as its derivative with respect to σ reaches zero at that point. Thus the value of σ at which $N_3(0,\sigma,b)$ achieves a local extrema can be used to estimate the width of a Gaussian edge.

Step2 Construct linked coordinates

Lindeberg (1998) has utilized both normalized first derivatives and normalized third derivatives measures and has claimed that both achieve the goal of automatic scale selection for diffuse edges. Georgeson et al. (2007) argued that $N_3(0,\sigma,b)$ has better resolution than $N_1(0,\sigma,b)$. Moreover, they have made a modification to $N_3(0,\sigma,b)$ in which the differentiation is split into two stages and only positive parts of the response are transmitted at each stage (half rectification). They thus solved the problem that two extra peaks are generated by a third derivative operator in response to each edge. In the current algorithm, zero-crossings of second derivative filters are

first used to derive edge locations. At each edge location, a normalized third derivative response is examined across all scales. Where the normalized third derivative response achieves a local extrema, the value of the scale is recorded and serves as an estimate of the width. For example in Figure 2.2, the actual width of a diffuse edge is $2 \times 3\sigma_{Max}$.

In order to compute a higher order directional derivative of a 2D image $I(x, y)$, it is more convenient to introduce two local orthogonal directions u and v with v parallel to the local gradient at each point and u orthogonal to it (Lindeberg 1998; Lindeberg 1993). Thus derivatives in these two directions can be expressed in terms of partial derivatives in the original Cartesian coordinates system

$$\begin{aligned} \partial_v &= \partial_x \cos\alpha + \partial_y \sin\alpha \\ \partial_u &= \partial_x \sin\alpha - \partial_y \cos\alpha \end{aligned} \tag{2.8}$$

where α is the angle between the gradient and x axis and can be determined using the following formula

$$\begin{aligned} \cos\alpha &= \frac{I_x}{\sqrt{I_x^2 + I_y^2}} \\ \sin\alpha &= \frac{I_y}{\sqrt{I_x^2 + I_y^2}} \end{aligned} \tag{2.9}$$

Here I use simplified notations I_x and I_y for $\frac{\partial I}{\partial x}$ and $\frac{\partial I}{\partial y}$. Similarly I_{xx} I_{xxx} will denote higher order derivatives in the following discussion. The problem of finding zero-crossings of second order derivative in gradient directions can hence be expressed as

$$\begin{cases} I_{vv} = 0 \\ I_{vvv} < 0 \end{cases} \tag{2.10}$$

Note that the n th order directional derivative of a 2D function I along the v axis is

$$\partial_v^n I = \left(\partial_x \cos\alpha + \partial_y \sin\alpha\right)^n I \tag{2.11}$$

Substituting I_{vv} I_{vvv} in (2.10) with (2.11) and (2.9), gives:

$$\begin{cases} \dfrac{I_x^2 I_{xx} + 2I_x I_y I_{xy} + I_y^2 I_{yy}}{I_x^2 + I_y^2} = 0 \\ \dfrac{I_x^3 I_{xxx} + 3I_x^2 I_y I_{xxy} + 3I_x I_y^2 I_{xyy} + I_y^3 I_{yyy}}{\left(I_x^2 + I_y^2\right)^{1.5}} < 0 \end{cases} \tag{2.12}$$

Equation (2.12) shows that I_{vv} I_{vvv} are combinations of partial derivatives in the Cartesian coordinate system which can be computed by convolving with corresponding Gaussian derivative filters at appropriate scales. For instance, I_{xy} is the convolution of an image I and a Gaussian partial derivative operator $\dfrac{\partial G(x, y, \sigma)}{\partial x \partial y}$.

Once edge positions are located, the normalized third derivatives at those points are assessed across all the candidate scales and the true scale σ_{Max} can be easily estimated. With the edge point, gradient direction α and estimated width all available, a link such as the gray bar in Figure 2.3 can be constructed. If the same process is carried out for every edge point, a group of such links will be established for the entire low pass filtered image $I_{blur}(x, y)$. If we compare $TxtEdge(x, y)$ with this group of linked positions, we will have estimation in regard to where luminance varies due to changes in texture and where luminance varies due to shading.

Step 3 Labelling luminance changes

In this section, we will look at how the classified luminance changes should be processed. Luminance changes can be modelled using linear operators such as gradient operator or Laplacian. Equation (2.1) can also be written in the log domain as:

$$\hat{I}(x,y) = \hat{S}(x,y) + \hat{R}(x,y) \tag{2.13}$$

where $\hat{S}(x,y) = \log S(x,y)$ and $\hat{R}(x,y) = \log R(x,y)$ such that the two intrinsic components are linearly separable. Consequently, applying a linear operator to $\hat{I}(x,y)$ is equivalent to applying the same operator to the intrinsic components individually and then adding the results together:

$$L\hat{I}(x,y) = L\hat{S}(x,y) + L\hat{R}(x,y) \tag{2.14}$$

where L is a linear operator, representing changes in luminance. Those non-zero values in $L\hat{I}(x,y)$ whose locations have been classified as places where luminance varies due to changes in texture, are set to zero. All the rest of the non-zero values in $L\hat{I}(x,y)$ are retained. In doing so, we hope to eliminate the reflectance component (the second term in (2.14)) and only retain the shading component (the first term in (2.14)).

However problems may arise from treating all points lying on a link equally. Imagine at a location that is very close to the two boundaries of a link, the local gradient may lie on a direction very different to that of the link. This often occurs when two types of luminance variations intersect. Hence it is useful to compare the local gradient direction with the direction of the link before carrying out any labelling process. Only when the difference of the two directions falls below a threshold, are the corresponding derivatives set to zero.

The remaining non-zero values in $L\hat{I}(x,y)$ should mostly represent $L\hat{S}(x,y)$ and should be ready for the reconstruction process.

5.6 Reconstruction: Inverse filtering

Given an estimated $L\hat{S}(x,y)$, the recovery of $\hat{S}(x,y)$ involves inverting a system:

$$L\hat{S}(x,y) = C\left(L\hat{I}(x,y)\right) \tag{2.15}$$

where $C(\ )$ represents the classification process. The problem of finding the inverse of an imaging system is very often an ill-posed one. Weiss (2001) and others (Olmos & Kingdom 2004, Tappen et al 2005) used the gradient operator ∇ in place of L. Solving (2.15) then involves calculating the integral:

$$\hat{S}(x,y) = \int_{R^2} C(\nabla \hat{I}(x,y)) \tag{2.16}$$

where ∇ is a vector field. For discrete functions, differentiation can be approximated with the difference between the two adjacent samples. Written in the format of filtering and broken down to two scalar equations, (2.15) reads as follows

$$\begin{cases} \hat{S}(x,y) * f_x = \hat{D}_x(x,y) \\ \hat{S}(x,y) * f_y = \hat{D}_y(x.y) \end{cases} \tag{2.17}$$

where $D_x(x,y)$ and $D_y(x,y)$ are classified horizontal and vertical derivatives respectively. The two filters f_x and f_y are simply $[-1,1]$ and $[1,-1]^T$. Equation (2.17) can also be understood as an over-strained linear system where each scalar equation places a linear constraint on the image. The solution to (2.17) can be found by working out its pseudo-inverse:

$$\begin{cases} \hat{S} = g * \left(f_x^T * \hat{D}_x + f_y^T * \hat{D}_y\right) \\ g * \left(f_x^T * f_x + f_y^T * f_y\right) = \delta \end{cases} \tag{2.18}$$

where f_x^T f_y^T are the transpose of f_x and f_y, δ is a *Kronecker Delta Function*-like metric with value 1 in the centre and 0 elsewhere. In practice, the calculation is often carried out in the frequency domain. Gradient based methods carry out the

reintegration along the fixed path (i.e. horizontal and vertical directions) hence are anisotropic. They take no consideration in the actual integrability of the gradient field as the operation is always valid given an initial condition. But in situations where the underlying gradient field is not conservative (cannot be integrated), the integrating path is vital and the horizontal and vertical direction may not be the path that gives rise to the best reintegration.

Another widely used method is based on the Laplacian operator ∇^2:

$$\nabla^2 f(x,y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{2.19}$$

Replacing L with ∇^2 transfers the task into solving Poisson's equation:

$$\nabla^2 \hat{S} = E \tag{2.20}$$

where E is the classified output of Laplacian operator (i.e. classified edges). The advantage of working with Laplacian operator is that finding the solution to Poisson's equation has been well studied. Second, being both a scalar function and isotropic, the inverse of a Laplacian is relatively straightforward to compute. However, from the information theory point of view, if a system results in loss of information, then the solution to the system is not unique because the lost information corresponds to many possible solutions. The more information the system dismisses, the more solutions it will have. Being a second-order derivative operator, a Laplacian throws away first order linear changes as well as the constant DC term of an image. For example, Horn's algorithm (1974) would fail to recover the reflectance of an image if it is not constant at the border. To counter this, Blake (1985) based the classification process on the gradient field (image obtained by applying gradient operator) but ran the inverse process by solving Poisson's equation similar to (2.20). He proved that when

the gradient field was a conservative field and Neumann's condition was met, solving (2.16) is equivalent to solving:

$$\nabla^2 \hat{S} = \hat{E} \qquad (2.21)$$

where $\hat{E}$ is the divergence of the classified gradient field and is a scalar function. The algorithm introduced in this chapter was based on Blake's method but with a fast Poisson solver to solve the Poisson equation defined in (2.21). After having recovered the shading component of an image, its reflectance component can be obtained by subtracting the shading component from the original image in the log domain (equivalent to the division of linear images).

5.7 Examples and discussion

Some examples are shown at the end of the chapter. The results are not numerically accurate but still qualitatively acceptable. Limitations in each step of the algorithm have introduced different types of error as discussed below.

- Reflectance edges must not coincide with shading edges, i.e. luminance changes must either be due to changes in reflectance or shading not both. This may be the major limitation of the algorithm, which is shared with its counterpart in algorithms based on hue. The consequence is obvious in Figure 2.4 (1b) where the near-horizontal edge is disrupted in the middle where the two boundaries intersect. In fact, luminance changes in this area are combinations of the two factors. If transformed into the log domain, a luminance change is an addition of two vectors, each representing a change in reflectance or shading. However the current technique sets all luminance derivatives to zero, resulting in errors in the classification of luminance

changes. Another related limitation is that major luminance edges should not lie too close to each other – This is due to the limitation of the width estimation method. If two edges are too close to each other (e.g. ridges), pixels between the edges could be linked to points on both of them. In consequence, corresponding points on both edges are assigned with the same identity and will either both be deleted or retained. Perhaps two adjacent edges will have the same identity (i.e. both are shading or both are reflectance). But if not, one will be misclassified. A possible solution to this problem is that instead of setting luminance derivatives to zero, only remove the contribution of one component. The local curl values of the gradient field may be a guide to find the best way of decomposing the gradient vector to ensure the resulting curl is minimized.

- The classification rule in this algorithm is based on the constant local contrast under changing illumination. However two different surface patches could have similar contrast and in this case they will be treated as the same surface by the algorithm. This problem could be overcome by examining more features that are invariant to illumination. For example, texture elements or textons are another feature which could be helpful to differentiate textured surfaces. More generically, any texture segmentation algorithm that can produce a successful segmentation to an image after dividing by its mean will provide a good solution. A more accurately determined texture edge will in turn improve the classification accuracy. Note however that texture segmentation in natural images is itself difficult.

- The resultant shading image is actually a low-pass filtered version of the true shading component. This tends to cause distortions in reflectance images in places along the shading/shadow edges. The problem deteriorates when shading/shadow edges in the original image is very sharp, as in the case of Fig 2.4 (2c and 4c).

In general the performance is satisfactory especially considering that this algorithm requires no colour information. The algorithm provides a solution at a global level but distortions are present in local regions. This was expected since no local constraints were applied to guide the classification process. When combined with local constraints such as colour, a better classification should be expected.

5.8 Conclusion

Luminance changes in a scene are often due to many sources such as shading and reflectance. Similar to the use of hue to assist the separate these two components, texture information can also be useful in this kind of task. An algorithm for separating shading/shadows from reflectance changes has been presented. This separation algorithm is based on characteristics of the textures in the scene. The idea of using texture to accomplish the task is inspired by the fact that humans also use textures to help with shape-from-shading tasks. The performance of the algorithm is satisfactory when testing images contain large areas of shading and reflectance. Decompositions at local regions may suffer distortions due to the lack of local constraints such as colour. However the texture based algorithm could serve as a global constraint complementary to other local features to form a better solution to intrinsic image decomposition.

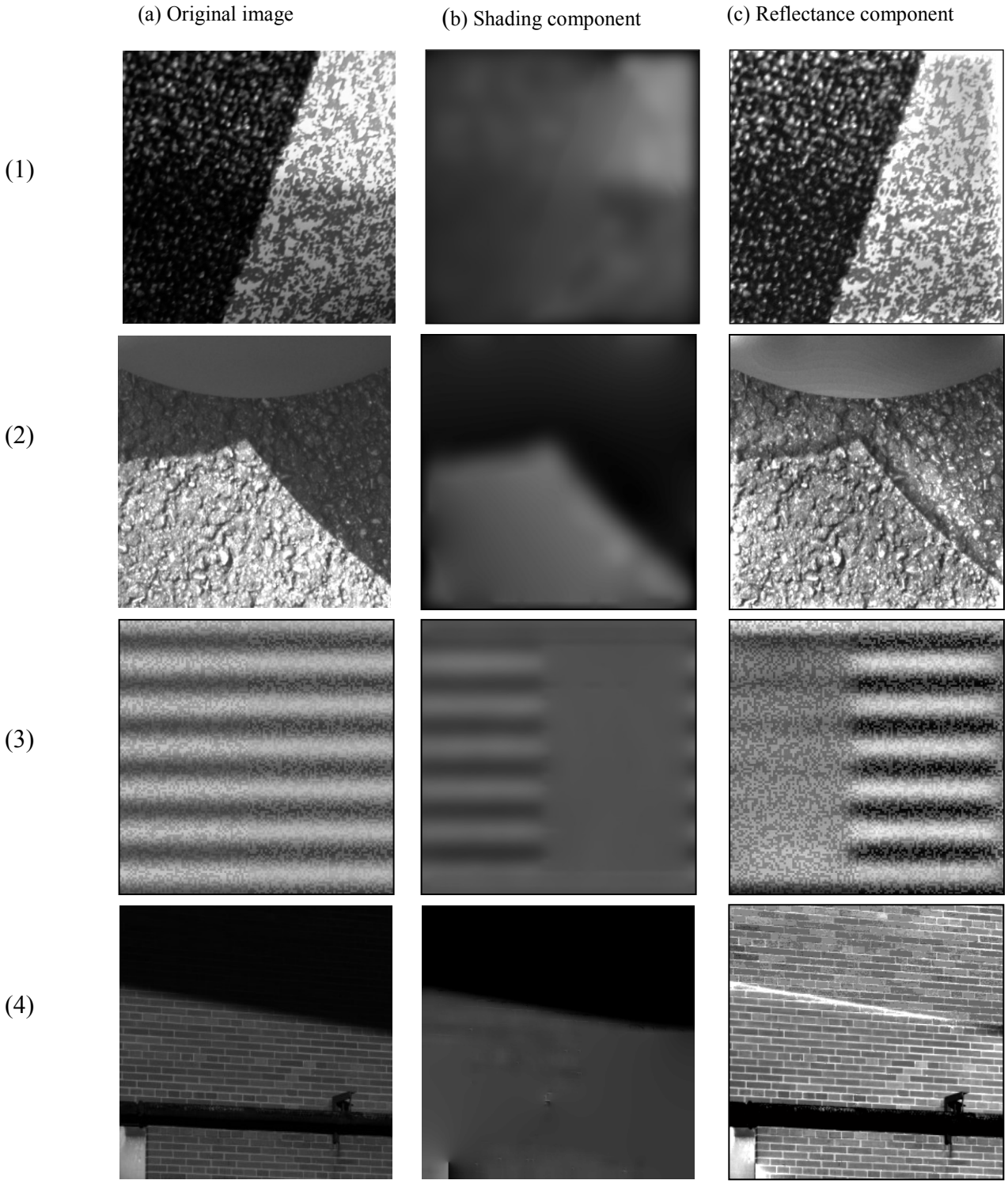


Figure 2.4 Examples of intrinsic image decomposition produced by the algorithm. Image set (1): a carpet surface meets a pained flat surface, both casted by a shadow. Image set (2) part of a ball hides in the shadow. Image set 3: Synthesized sinusoidal gratings containing binary noise. Local contrast is constant on the left half but undergoes the same undulation as the luminance on the right half. Image set (4) a brick wall. Shadow is casted over the top of the image.

6. Perception of shape-from-shading

Chapter 4 and 5 introduced a method for extracting shading cues from an image. This Chapter addresses the issue of how human vision deduces surface shape from such shading cues. The chapter addresses the built-in rules that humans use to derive 3-D shape based on shading alone. In particular, the long-held assumption that perceived slant is proportional to luminance is tested experimentally. To test the validity of this assumption, the perceived shape of various types of luminance grating, including sine wave, square wave, periodic saw-tooth and cropped saw-tooth, were tested using a gauge figure task. The results show that the slant = luminance relationship only holds for gratings which are bounded by edges of equal strength and polarity. In the second experiment, the square wave and sine-wave gratings were cropped such that luminance variations were not bounded by edges with same polarity. Observers perceived cropped gratings differently from those surrounded by like-polarity edges. The interaction between shading and edges is further discussed at the end of the chapter, followed by a new theory for human SFS.

6.1 Background

This section reviews the historical background of shape-from-shading in human vision, starting with the formulation of the problem in the physical world, followed by an account of our understanding of shape-from-shading in humans. The section ends by outlining the motivation for the experiments in this chapter.

6.1.1 Formulation of shading

The subject of shape-from-shading has been extensively studied and is still an active area of research in both computer vision and human perception. Theories in both areas are based on the fact that shading (variations in reflected light intensity) on a surface

in 3-D space depends on the angle of the surface normal relative to the light source. Consider Figure 6.1, here a surface is inclined (with angle i) with respect to the incident light and (with angle e) with respect to the observer, this latter angle being termed the angle of reflectance. The incidence reflectance vectors form an angle σ. If we denote incident light intensity by I_1 per unit area perpendicular to the incident ray and the reflected light intensity by I_2 per unit solid angle per unit area perpendicular to the reflected light (Horn, 1975). Then the *reflectivity function* $\phi(i,e,\sigma) = I_2 / I_1$ determines the relation between the incident light and the light received by a viewer (Horn, 1975). When the properties of light source and the surface reflectance are known, $\phi(i,e,\sigma)$ becomes a mapping between the image intensity and the three angles. That is, image intensity provides information about the 3 dimensional form of a surface and this information can be characterised by the reflectivity function $\phi(i,e,\sigma)$. In computer vision, the study of shape-from-shading is concerned with establishing a mathematical mapping between these variables allowing one variable (the surface norm) to be solved given the other (image intensity) (Horn, 1975; Horn, 1977; Ikeuchi & Horn, 1981; Pentland, 1984; Pentland, 1988; Horn & Brooks, 1989; Horn, 1989). One of the most well established shading models is the uniform Lambertian surface lit by a distant light source. A perfect Lambertian surface reflects light equally in all directions. More intuitively, it means the image intensity at a point on the surface is constant regardless of the viewing direction, (i.e. independent of the reflectance angle e). Under these restrictions the image intensity depends only on the angle between the surface norm and the incident ray (Horn, 1977; Marr 1982; Pentland, 1988). Further, under these conditions, image intensity is proportional to the cosine of the angle i in Fig 6.1: that is, $L = K \cdot I \cos i$ where L is the image intensity, I is the light

source intensity, K is a constant. This model will be discussed in more depth later in this chapter.

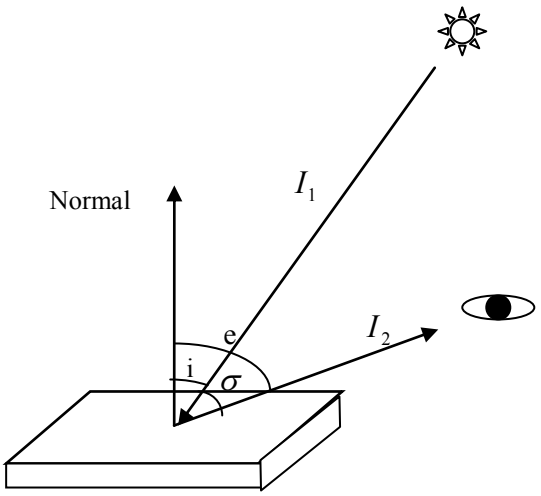


Figure 6.1The formulation of shading. The reflected light is related to the incidence angle i, reflectance angle e, resulting in degradation of luminance according to the normal of the surface (after Horn, 1975).

6.1.2 Ambiguities of shading

As the 2-D projection of a 3-D structure, shading is inherently ambiguous. A well-known ambiguity of shading is that principal curvatures of surfaces (assuming surface is locally spherical) can not be revealed by shading information alone (Pentland, 1984). The traditional view regarding the shading ambiguity is that shading is a product of surface orientation, light source and surface reflectance. Any particular shading image can be due to infinite possible combinations of the three variables (See Fig 6.2). Fortunately the ambiguities of shading have been extensively investigated for cases where the surface is Lambertian. Belhumeur et al. (1999) proved mathematically that the ambiguities obey an affine transform or, as the authors called it, a "Generalized bas-relief transformation (GBR)". Under this affine transform,

$$\hat{f}(x,y) = \lambda f(x,y) + \mu x + \nu y \tag{6.1}$$

where x, y are co-ordinates of the image plane, $f(x,y)$ is the depth function of the real 3-D structure, λ is a scaling factor, μ and v control shearing. In matrix form, a point $\mathbf{p} = [x \quad y \quad f(x,y)]$ on the surface becomes $\hat{\mathbf{p}} = G\mathbf{p}$ where $G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & v & \lambda \end{bmatrix}$.

Given a light source $\mathbf{l}$, the luminance intensity at a point $\mathbf{p}$ can be defined as $L(x,y) = a(x,y)\mathbf{n} \cdot \mathbf{l}$ where $\mathbf{n} = [-f_x \quad -f_y \quad 1]^T$ is the unit surface normal and $a(x,y)$ is the surface albedo. Belhumeur et al. (1999) proved that there exists a light source $\hat{\mathbf{l}}$ illuminating a GBR transformation of the original surface with albedo $\hat{a}(x,y)$ such that the luminance intensity at a point $\hat{\mathbf{p}}$ on the transformed surface:

$$\hat{L}(x,y) = \hat{a}(x,y)\hat{\mathbf{n}} \cdot \hat{\mathbf{l}} = L(x,y) \tag{6.2}$$

where $\hat{\mathbf{n}} = \mathbf{n}G^{-1}, \hat{\mathbf{l}} = G\mathbf{l}$ and $\hat{a}(x,y) = \frac{a}{\lambda}\left(\frac{(\lambda f_x + \mu)^2 + (\lambda f_y + v)^2 + 1}{f_x^2 + f_y^2 + 1}\right)^{1/2}$. Equation (6.2) means that given a shading image, the solution of the 3-D shape can be determined only up to a space of affine transformations.

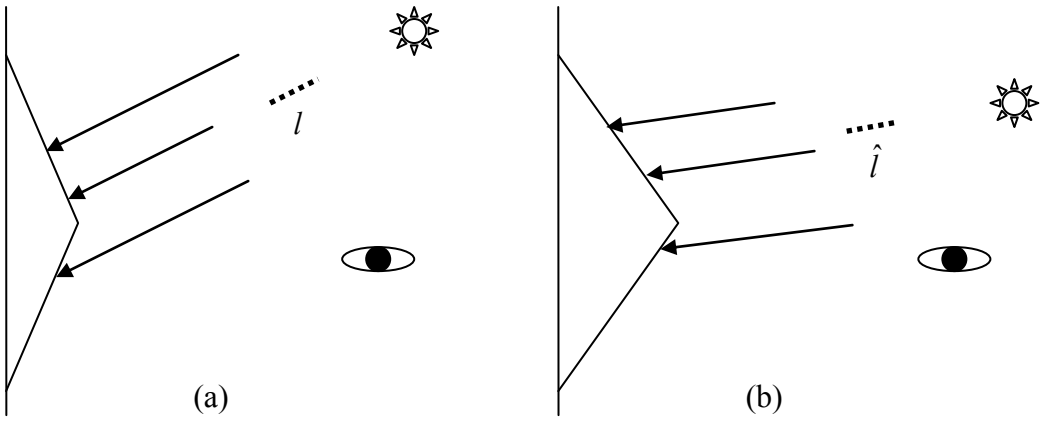


Figure 6.2 Demonstration of shading ambiguity—Generalized bas-relief transformation (GBR). (a) A triangular surface with Lambertian reflectance is lit by a directional illumination l. The surface is frontally viewed. It is easy to see that the resulting shading pattern consists of two gray levels. (b) The surface shape is scaled. The illumination vector is manipulated to become $\hat{l}$ so that the ratio of the two grey levels in the shading appearance remain constant. By adjusting the surface albedo, the shading image of (b) can be made identical to that of (a).

6.1.3 Human perception of shape-from-shading (SFS)

Humans are capable of interpreting qualitative 3D-shapes from shading but the mechanism for this is poorly understood. The complexity of this visual function can be shown by the fact that results from previous studies on this topic often lead to contradictory conclusions. Major discoveries on human SFS are summarised in the following paragraphs.

Ineffectiveness vs. effectiveness

Shading has been considered a relatively weak cue to depth compared to other cues such as disparity and texture gradient and the effect of shading appears minor when those other cues are present (Bülthoff & Mallot, 1988). There is also evidence suggesting that the three dimensional structure inferred from shading by humans is inaccurate: Humans tend to underestimate surface slant in shading patterns compared to the slant that would be required in a Lambertian model (Todd & Mingolla, 1983; Mingolla & Todd, 1986; Mamassian & Kersten, 1996; Hann, Erens & Noest, 1995; Norman & Todd, 1996). But when it comes to shading patterns containing highlights, the perceived slant tends to be overestimated (Todd & Mingolla, 1983).

Further, human SFS is also ineffective in response to the shading ambiguities mentioned in 6.1.2. As expected, human observers cannot differentiate between an elliptic shape and a hyperbolic shape based on shading alone (Erens et al., 1993a). In addition, shape judgements from different observers differ significantly but in a systematic way; shape judgements for simple objects often differ by a scaling factor whereas those for complex objects can normally be accounted for by an affine transformation (see 1.1.4.2). This affine transformation echoes the bas-relief

ambiguity defined by equation (6.1). Koenderink et al. (2001) argues that human SFS is operational such that humans resolve the ambiguities by applying their "beholder's share" when they respond to 2-D shading. In a parallel of the "Generalized bas-relief transformation (GBR)", human SFS can also be characterised by a space of affine transformations:

$$\hat{z}(x,y) = az(x,y) + bx + cy + d \qquad (6.3)$$

where $\hat{z}(x,y)$ is depth function estimated by an observer in a particular task, $z(x,y)$ is the depth function of the 3-D structure as represented in the visual system prior to the affine transformation, x, y are the coordinates of the image plane, constant a represents the scaling factor, b, c and d control a shear transformation of the 3-D surface. Note that $z(x,y)$ does not have to be the ground truth depth profile of the surface. Rather, it may reflect a common coding strategy shared by all participants before they apply their "beholder's share" (see also Battu, Kappers & Koenderink, 2007).

The ineffectiveness of shading in 3D tasks can easily lead to the conclusion that shading is not a very useful cue and that it could be irrelevant in the tasks of understanding complex scenes rich in other visual information. But results from later studies have confirmed that this conclusion is not warranted. Using a curvature discrimination task, Johnston and Passmore (1994a) reported a low discrimination threshold (Weber fraction of 0.1), indicating that the observers made effective use of shading during the task. Furthermore, in contrast to their inability to differentiate between elliptic and hyperbolic shapes, humans have no problem segmenting the surface of a croissant-shaped object according to whether the region is hyperbolic or elliptic (Mamassian et al., 1996). The most pronounced evidence against the view that shading is irrelevant comes from a series of complex scene understanding tasks

published by Koenderink et al. (1996a; 1996b). In these tasks, human observers viewed photographs of human statues lit from various directions. The photographs of a particular stature contained very different shading patterns but all depicted the same relief. The observers demonstrated qualitatively accurate interpretations of the relief for all lighting directions. But the quantitative measurements of surface for each testing position on the sculpture appeared to undergo a systematic variation with regard to the shading pattern. This phenomenon is a strong sign that shading is not irrelevant, even in the situation where other visual information is rich, and that shading has a systematic effect on the perception of 3D surfaces.

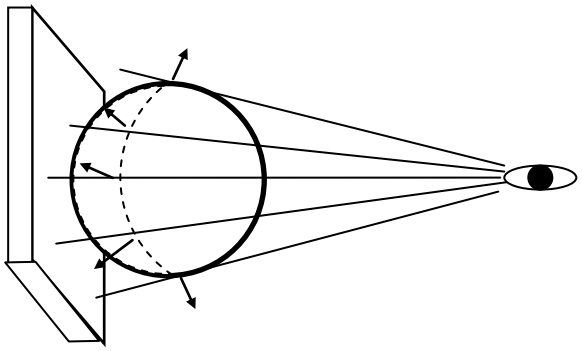
Shape constancy does not hold for SFS

If the perceived shape of an object remains unaffected under different conditions, the perceptual output is said to possess shape constancy (Khang et al., 2007). While desirable for many visual functions, shape constancy does not seem to exist in SFS. The overestimation of the perceived slant of a shiny surface and its underestimation for Lambertian surfaces (Todd & Mingolla, 1983) demonstrate a lack of shape constancy: perceived shape changes with material properties. In addition, changes in lighting direction can also lead to changes in perceived shape (Christou & Koenderink, 1997). More recently, Khang et al. (2007) tested observers with objects under various lighting conditions and surface treatments. Consistent with previous findings, perceived shape varied with both lighting and material, suggesting that little shape constancy was achieved except when the degree of specularity was varied. However, from an experimental point of view, the lack of constancy in SFS means that it is possible to characterise SFS in the visual system. The SFS algorithm employed by the visual system is most likely simpler than that which would be

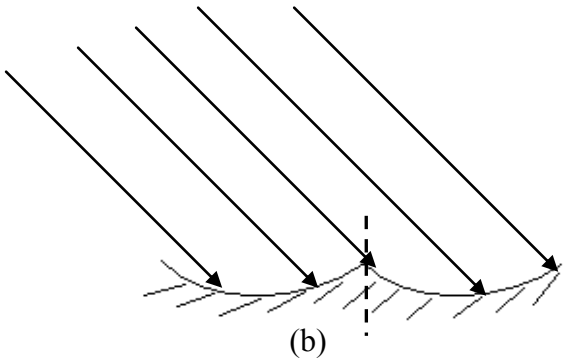
required to achieve shape constancy. The logic of this argument will be explained in section 6.1.4.

Edges and shading

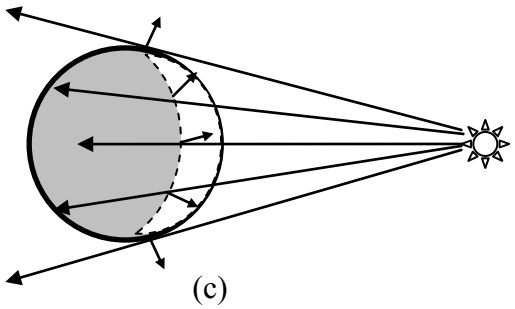
Edges can arise from a variety of causes. Marr (1982) summarised the origins of edges as the following: 1) reflectance changes, 2) discontinuities in depth such as occluding boundaries 3) discontinuities in surface orientation and 4) illumination effects such as shadows and highlights. Edges caused by reflectance changes are irrelevant in SFS and have been dealt with in previous chapters: they are not considered further here. Occluding boundaries (see Figure 6.3a) are a direct result of discontinuities in depth but can be a cue to surface orientation. Edges falling into this category are thought to be the points where the surface normal is perpendicular to the viewing direction (Marr, 1977; Barrow & Tenenbaum 1981; Malik, 1987; DeCarlo et al., 2004; Lawlor et al., 2009). In fact there exist computer vision algorithms which produce edge maps of a 3-D object by searching for the points that meet the criteria for occlusion edges (DeCarlo et al., 2004). Edges due to changes in surface orientations are more relevant in the context of shape-from-shading. Edges of this type are formed by the same principle as shading and can be understood as special instances of shading for which the variations in luminance are more abrupt as they arise from discontinuities in surface orientation (Figure 6.3b). Edges caused by illumination effects are view point independent. An example is the boundary between an illuminated area and area of self-shadow (Figure 6.3c). A strict definition of this kind of edge requires that edges occur where the surface normal is orthogonal to the direction of the incident light (Marr, 1982; Barrow & Tenenbaum 1981; Ikeuchi & Horn, 1981).



(a)



(b)



(c)

Figure 6.3 Three types of edges in the context of shape-from-shading. (a) Edges that are due to discontinuities in depth. The occluding edge (dashed line) marks the two visible surfaces which are different in depth. (b) Edges that are caused by discontinuities in surface orientation. The surface is lit by directional light source. The crease in the middle (labelled with a dashed line) will produce a discontinuity in luminance as a result of a discontinuity in surface orientation. (c) Edges that segment the surface into illuminated area and self-shadows (after Palmer, 1999, p245). The dashed line represents the edge points at which the surface normal is orthogonal to the incidence.

The aforementioned edge types (2, 3 & 4 in Marr's 1982 classification) constitute object boundaries and edge contours which together can be termed outlines. Object outlines are important cues to surface-shape (Ramachandran, 1988; Todd, 2004) and can be exploited to compute the 3-D shape of an object (Guzman, 1969; Clowes, 1971; Barrow & Tenenbaum 1981; Waltz, 1975; Marr, 1982; Malik, 1987). The shape cue provided by outlines is so strong that it can override other cues such as shading (Ramachandran, 1988; Bülthoff & Mallot, 1988). In such cases, object outlines alone can produce a 3D shape percept without any shading (see Figure 6.4). Indeed, humans can articulate a pictorial relief similar to that based on photographs from outlines alone (Koenderink et al., 1996a). Thus when outlines dominate shape perception, shading appears almost immaterial and its effect is either hard to measure or completely confounded by the outlines (Mamassian & Kersten, 1996). For this reason it is tempting to remove the confounding effect of outlines by cue reduction, but Koenderink et al. (1996b) argue that the methodology of cue reduction is inappropriate. In particular, Koenderink et al. (1996b) argue that SFS requires some other visual information in order to fully function, although the authors did not specify the form of the additional information required.

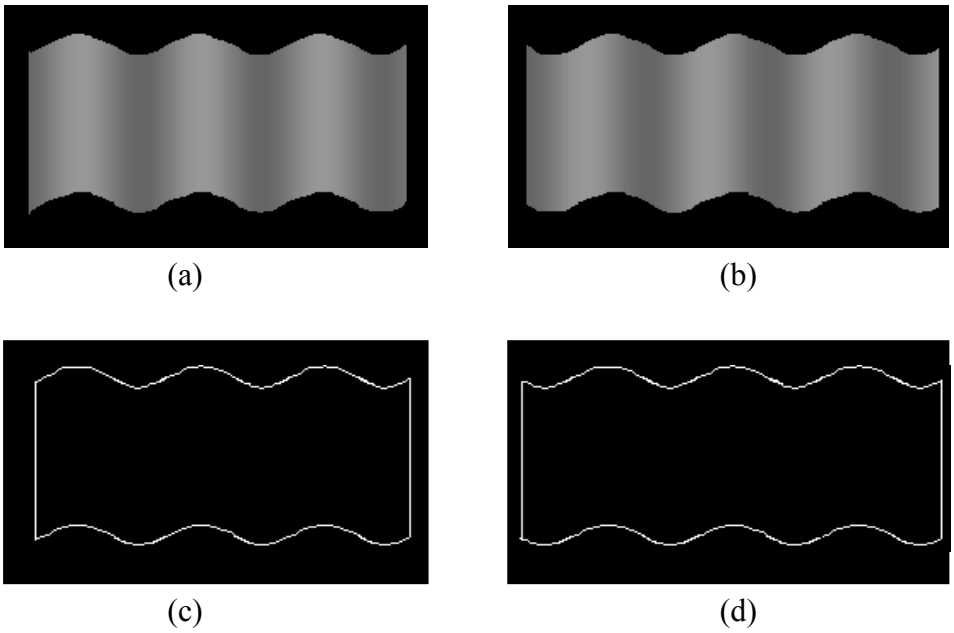


Figure 6.4 Outlines determine how humans interpret luminance variations. (a) and (b) have the same pattern of luminance variations, but these variations are interpreted differently: the brightest points in (a) appear the highest on the surface while in (b) the points with median brightness appear to be the peak of the surface. Their 3-D interpretations seem to follow what are suggested by their outlines (c) and (d), suggesting that outlines could act as a confounding factor in the perception of SFS (After Ramachandran, 1988)

The three types of edges most closely related to shading are good candidates for the complementary visual information required to make SFS function. The interaction between edges and shading has also been utilised in computer vision. For example, classical computational approaches for shape-from-shading often involve solving partial differential equations. For these methods, edges and occluding boundaries can serve as initial curves or boundary conditions because the orientations of surface norm at these locations are known to be perpendicular to the viewing direction (Ikeuchi & Horn, 1981). The complementary relationship between edges and shading has not been thoroughly examined in terms of human perception. Figure 6.5 illustrates the importance of outlines in the perception of SFS even when outlines alone do not support unambiguous 3D perception.

In summary, shading needs other visual information to fully function as a cue to 3D shape and edges are likely to be a candidate for this information, since the formulation of certain types of edges is very closely related to that of shading. But the presence of excessive outlines could easily dominate shape perception, making the effect of shading difficult to measure. What is required is a methodology which introduces edges in a controlled way allowing the effects of pure shading and edges to be differentiated without allowing the edges to fully dominate the percept.

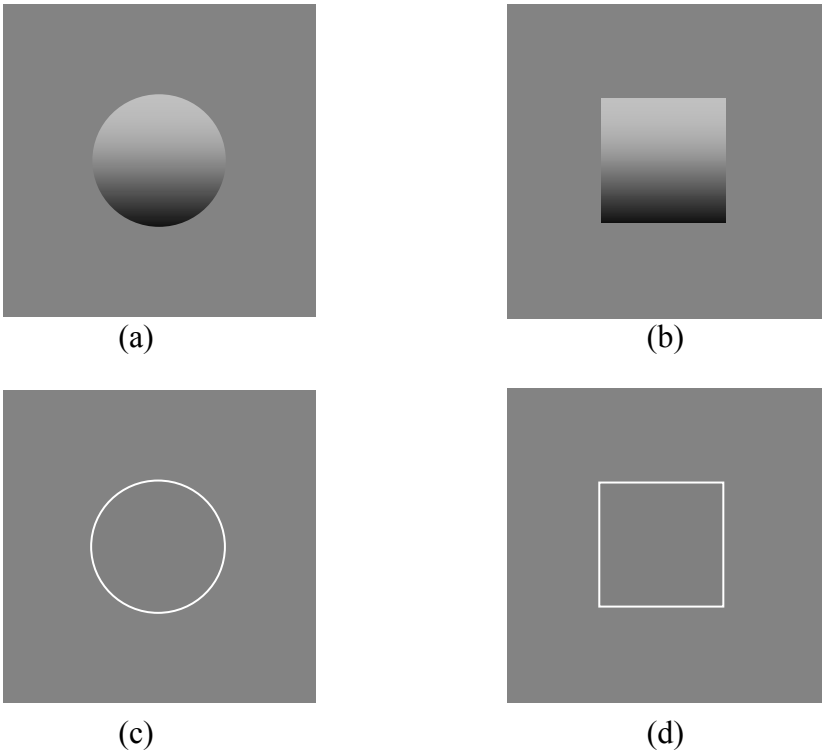


Figure 6.5 Outlines (edges contours and object boundaries) modulate the perception of SFS. (a) A linear luminance ramp bounded by a circle appears to be a bump. (b) The same linear luminance ramp bounded by a square appears to be a cylinder. Neither outline (c, d) produces the impression of 3-D shape in the absence of shading.

Estimating the direction of the light source

An analysis of the generation of shading reveals that the information that shading conveys directly is an angle relative to the direction of the incident light. Thus from a computational point of view, the 3-D structure of the surface cannot be determined without the knowledge of the illumination. In computer vision, assumptions (often

unrealistic ones) have to be made about the properties of the illumination for the shape-from-shading algorithm to deliver a unique solution of the 3-D shape. Alternatively, illumination can be estimated using a method which is based on the patterns of luminance gradients in a scene (Pentland, 1982). Either way, the illumination has to be specified before a solution is obtained. As to the function of SFS in the visual system however, the role of light source estimation is rather complex: knowledge of the light source helps to break the convex / concave ambiguity (Ramachandran, 1988) but perceived shape can also affect the light source estimation (Koenderink et al., 2004; 2007). Humans can acquire the information about the light source through analysing shadows and highlights (Mingolla & Todd, 1986; Liu & Todd, 2004), luminance gradients (Pentland, 1982) and second-order statistics of relief texture (Koenderink et al., 2004; Koenderink et al., 2007; Pont & Koenderink, 2007). But the 3-D structure in such stimuli is probably estimated along side the light source direction. Thus, light source estimation and 3-D shape perception are likely to be two products of a full functional SFS method; arguments as to which is conducted first are likely to prove unproductive

6.1.4 Algorithms for human SFS suggested by psychophysics

Linear reflectance model (LRM)

A commonly held (but often only implicitly articulated) view in the study of SFS is that the perceived slope is proportional to the luminance values of the shading pattern (slant $\propto$ luminance). The linear relationship between the luminance variation and the perceived slant can explain some observed characteristics of human SFS. For example, perceived slant is overestimated when the slant of a surface generated by a Lambertian shading model is small, but underestimated when the surface is more slanted and this bias rises as the real surface slant increases (Mamassian & Kersten,

1996). Suppose a Lambertian surface is lit by a distant source and has slant θ relative to the vertical axis in the image plane. Suppose the viewing vector overlaps with z axis and let incidence angle be i and the angle between incidence ray and the viewing direction be σ (see Fig 6.6). Then we have $i + \theta = \sigma$. σ should be less than 90° to avoid cast shadows. Since perceived slant is linear to luminance, we have $\cos i = \tan\hat{\theta} \Rightarrow \sin(90° - \sigma + \theta) = \tan\hat{\theta}$, where $\hat{\theta}$ is the slant angle estimated by observers. If σ equals 90°, then $\tan\hat{\theta} = \sin\theta \Rightarrow \sin\hat{\theta} = \sin\theta\cos\hat{\theta} \Rightarrow \sin\hat{\theta} < \sin\theta$, so the slant angle should always be underestimated. As σ varies and let $\alpha = 90° - \sigma > 0$, then $\tan\hat{\theta} = \sin(\theta + \alpha) > \tan\theta$ when θ is very small, but $\tan\hat{\theta} = \sin(\theta + \alpha) < \tan\theta$ as θ increase and the difference becomes even larger as θ approaches 90°, i.e. perceived slant is overestimated when the Lambertian surface is only slightly slanted but becomes underestimated when the slant gets larger. The underestimation will increase as the slant of the surface.

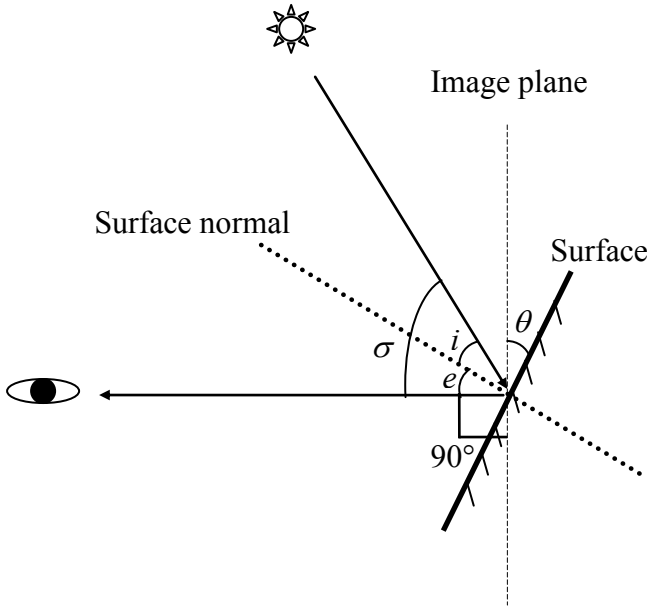


Figure 6.6 The relation of image intensity and the orientation of a Lambertian surface lit by a single point source. The process is illustrated in 1D. e, i and σ represent the same angles as in Figure 6.1. θ is the angle that the surface is inclined with in respect to the image plane. Without the loss of generalization, the viewing direction is set to perpendicular to the image plane. Under this setting, e equals θ.

Surprisingly however, very few studies have experimentally investigated the empiric of slant = luminance. Some marginally related work can be traced to Pentland's biologically inspired model for recovering surface height from shading (1988) which is outlined below. Assuming a Lambertian surface lit by a distant light source and viewing direction fixed to be perpendicular to the image plane, the normalized image intensity will be:

$$I(x) = \cos i = \bar{n} \cdot \bar{I} = \frac{p \sin\sigma + \cos\sigma}{\sqrt{p^2 + 1}} \tag{6.4}$$

where σ is the angle between the incident ray and the viewing direction, $\bar{I} = (\cos\sigma, \sin\sigma)$ is the vector of the incident ray, p is the slope of the surface along the image plane, i.e. $p = \tan\theta$ and $\bar{n} = (1, p)$ is the vector of the surface norm (Fig 6.6). Note that the image plane has been simplified to be a 1-D signal in this expression. Taking the Taylor series expansion of equation 6.4 up to its quadratic term will give:

$$I(x) \approx \cos\sigma + p \sin\sigma - \frac{\cos\sigma}{2} p^2 \tag{6.5}$$

Pentland (1988) then argued that when $|p| << 1$ (leading to a negligible quadratic term $\frac{\cos\sigma}{2} p^2$) and ignoring the DC term $\cos\sigma$, the relationship between image intensity and the surface slope is linear.

When this linear relationship holds, the shading image of a sinusoidal surface is a sinusoidal profile of the same frequency as the surface corrugation but with a 90° phase shift. But if the quadratic term in Eq 6.5 dominates, a sinusoidal surface will give rise to a sinusoidal luminance variation with twice the surface frequency (Pentland, 1988). In a follow-up experiment, Pentland showed that human observers

inferred a near sinusoidal surface from a sinusoidal shading and that the fundamental frequencies of the inferred surface profile and the shading are similar. It was as if human observers ignored the quadratic term and assumed a linear relationship between perceived slant and luminance.

The derivation of this linear relationship can serve as a theoretical support for the LRM. Conversely, the LRM should operate most optimally when the conditions that lead to Eq 6.5 are satisfied and when Eq 6.5 can be best approximated by the linear relationship. That is to say, the LRM corresponds to the situation in which the illumination is directional and oblique ($\frac{\cos\sigma}{2}$ is small) and $|p| << 1$. However, omitting the DC term in Eq 6.5 is a weakness in Pentland's model (1988). In fact the DC term is important as will be explained in section 6.1.4.

The "dark is deep" rule

In some circumstances perceived surface height correlates with luminance – a computation often described as "dark is deep". For example, Christou and Koenderink (1997) reported that observers' slant judgement correlated with decreasing luminance gradients when viewing a rendered sphere with Lambertian shading—equivalent to "dark is deep". Langer and Bülthoff (2000) measured the accuracy of depth comparison between two positions on a surface rendered under collimated lighting and diffuse lighting. They called "correlated" the condition where a brighter point on the surface happened to be higher and "anti-correlated" the condition where a darker point is higher. Results showed that for surfaces rendered by collimated light, the accuracy was very high regardless of the correlation condition and human performances for "correlated" and "anti-correlated" conditions were similar. When

judging surfaces rendered by diffuse light however, observers favoured the "correlated" condition with much higher accuracy than the "anti-correlated" condition. The accuracy for the "correlated" condition under diffuse lighting was comparable to that under collimated lighting. This means that when surfaces were rendered under a diffuse light source, observer's depth setting correlated (to the first approximation) with the "dark-is-deep" rule.

Application of the "dark is deep" rule has been found in a range of shape-from-shading tasks (Nefs, Koenderink and Kappers, 2005; Christou & Koenderink, 1997; Langer & Bülthoff, 2000), although more so under some conditions than the others From this point of view, the "dark is deep" rule seems to comprise part of the entire human SFS algorithm and may dominate in some circumstances. Unlike LRM however, "dark is deep" lacks a solid theoretical support, although to a certain degree it is descriptive of a shading model under diffuse lighting. According to a model proposed by Langer and Zucker (1992), image intensities generated under diffuse lighting depends on how much a surface position is exposed to the "sky". Thus under the diffuse lighting conditions, a periodical sinusoidal surface will generate a luminance trace that is a periodic grating with the same fundamental frequency and phase as the surface (Wright & Ledgeway, 2004). In this case, "dark is deep" gives a qualitatively good description of the model (see Fig6.7a). However in many other cases, "dark is deep" only provides a partial generalization. For example in the case of a single cycle of sine-wave (Fig6.7b), although the top half of the surface obeys "dark is deep", the bottom half of the surface gives a near uniform luminance which is also the minimum luminance value of the whole shading image. Figure 6.7c and d give

two more examples where the diffuse model can not be generalized by the “dark is deep” rule.

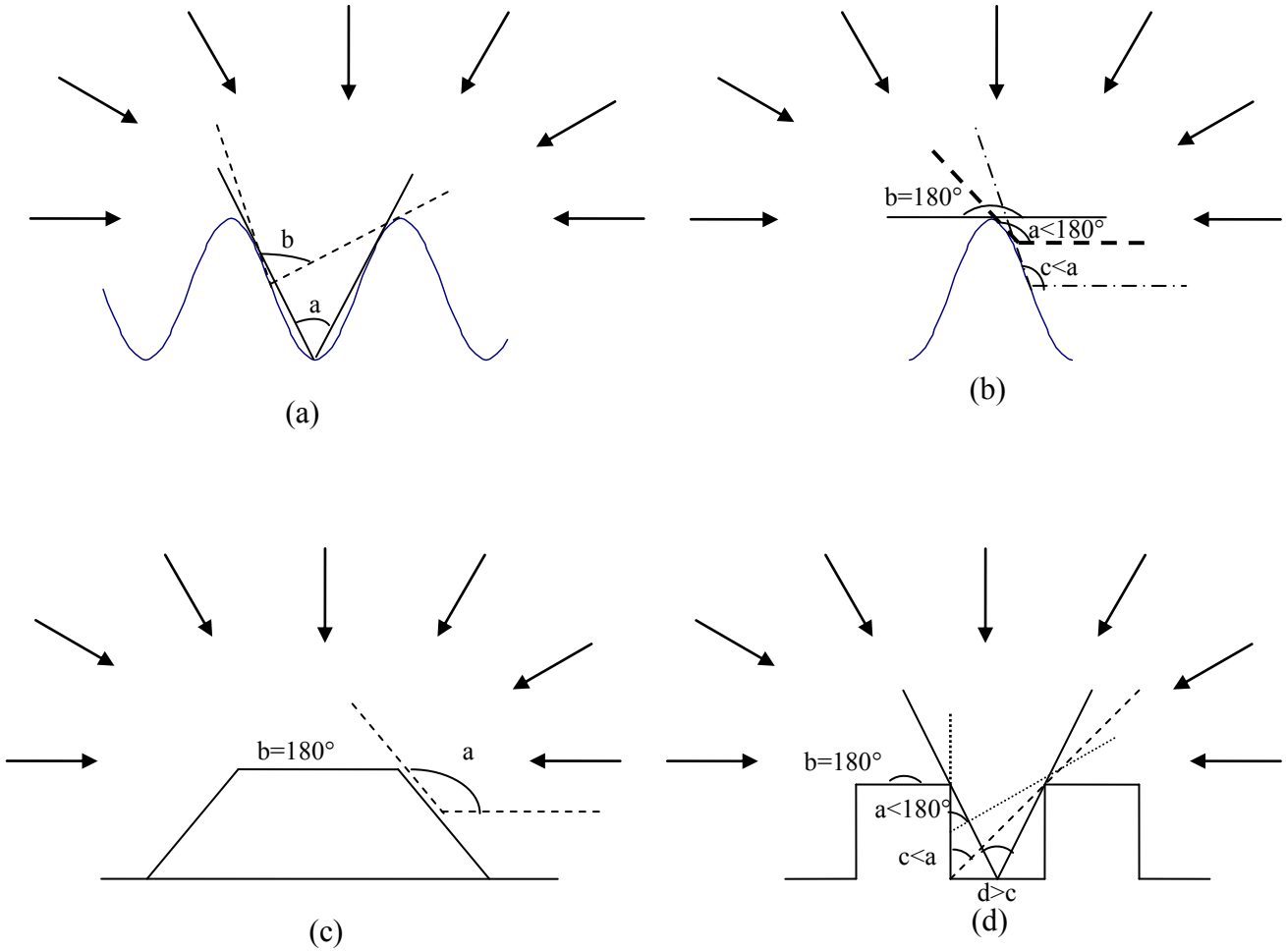


Figure 6.7: (a) periodical sinusoidal surface is illuminated by diffuse light. The valley sees a portion of the sky which subtends angle a. From the valley to the hill, the subtended angle increases and reaches the maximum at the peak. (b): a single sinusoidal ripple is illuminated by diffuse light. The top of the hill sees all of the sky hence is the brightest. Moving towards the valley, surface positions only see a portion of the sky and subtended angle a. This angle decreases and reaches its minimum at half of the ripple height. (c): trapezoidal surface is illuminated by diffuse light. The top plane is exposed to the entire hemisphere while the side surface only sees part of the light source. (d): a surface of square wave under diffuse light source. The top plane is exposed to the entire sky. The exposure decreases as the height of the position until the height reaches the bottom. As the measuring position moves across the valley, the exposure increases again and achieves a local maximal at the centre of the valley.

6.1.5 Motivation and aim of the study

As mentioned in section 6.1.3, the study of shape-from-shading in computer vision is primarily interested in establishing a mathematical mapping, “reflectance map”

(Horn, 1977) between the surface orientation and the shading pattern presented in the image. However, very few studies have had the clear aim of investigating whether humans assume a particular mapping or what mapping might be employed. The cause of this discrepancy probably comes from two sources. First, the same surface orientation will give rise to different patterns of shading under different lighting conditions and material properties, each corresponding to one particular mapping between shading and shape. If people are able to achieve shape constancy, the number of mappings available to humans is bound to be infinite: trying to measure any one mapping seems futile. Fortunately, Khang et al. (2007) discovered that when lighting or material properties change, leading to changes in the resultant shading patterns, observers' shape judgement also changes: humans do not have shape constancy. They concluded that 3-D shape judgment largely depends on the luminance pattern and less so on any other factors. This result suggests that humans may only utilise a limited number of mappings. Hence it is worth trying to characterise the mappings involved. Recall that human SFS is subject to an affine transformation defined by Equation 6.3. Thus the key question in studying human SFS is to find the common internal 3-D representation of Equation 6.3; that formed prior to the affine transformation ($z(x, y)$ in Eq 6.3).

Another reason why the study of SFS is less interested in characterising the built-in reflectance map is that even if a robust reflectance map does exist for humans, the linear mapping suggested by Pentland (1988) is taken for granted in spite of having not been sufficiently tested. A key problem with Pentland's mapping resides in the direct removal of the DC term $\cos\sigma$ in equation 6.5 which seems rather ad-hoc. A more justifiable way to decouple the DC term and the linear term is to differentiate

the two sides of the equation (supposing the quadratic term $\frac{\cos\sigma}{2}p^2$ is small enough to be ignored):

$$I'(x) \approx p' \sin\sigma \approx C \cdot \frac{d^2 z(x)}{dx^2} \qquad (6.6)$$

where $\sin\sigma$ is substituted by a constant C, $z(x)$ is the height of the surface. Therefore instead of associating the absolute value of image intensity to the slope of the underlying surface, equation 6.6 suggests that the first derivatives of image intensity and second-derivatives of the surface height are proportionally related. Compared to equation 6.5, equation 6.6 is more biologically plausible because the human visual system is more sensitive to changes in image intensity than to absolute image intensities (Pentland, 1982). If the perception of shape-from-shading is indeed based on equation 6.6, then the commonly held linear mapping is itself only one of an infinite number of possible mappings, each a solution to equation 6.6. For equation 6.6 to have a single solution, two boundary values are required. Furthermore, in order for the solution to be a linear mapping between $I(x)$ and $z'(x)$, the two boundary values have to be equal – the so called fix-fix condition in solving ordinary differential equations. Inspired by the use of edges as boundary conditions in computer vision, I speculate that edges may provide necessary boundary information for human vision to resolve the problem posed by equation 6.6 in such a way as to produce a linear mapping in many cases.

Distinct from many other studies, the research presented in this chapter attempts to investigate a fundamental question in the subject of human shape from shading. That is, what are the characteristics of the mapping used by humans to link surface orientation and luminance? Do we assume a linear relationship between the surface

orientation and luminance? Conventional approaches which involve testing with computer generated realistic 3-D objects are not practical for this purpose as they would require testing the many possible rendering models and find the one that is most consistent with human data. Besides, realistic 3-D objects contain outlines which could determine perceived shape, undermining the effects of shading. Therefore a different methodology has been taken: Instead of using computer rendered realistic 3D objects, I have tested human observers with stimuli made up of several very simple luminance profiles without contextual outlines. By doing so, the underlying mapping can be revealed in a way that is not subject to any particular rendering model, while excluding the influence of other cues to surface shape. Using luminance profiles is also psychologically plausible because not only does human SFS mostly depend on luminance patterns (Khang et al., 2007) but also is it quite stable (See 1.1.4.3).

6.2 Experiment 1

In the first experiment, observers viewed four patterns of luminance variations in three different orientations. The aim is to verify the commonly held view in respect to the linear relationship between perceived surface slants and the underlying image intensity.

6.2.1 Equipment and calibration

Monitors were calibrated using the same method as in chapter two. The viewing distance was 1 meter. Images measured 13.312 by 13.312 degrees of arc (512 by 512 pixels) displayed inside a central window. Outside of the central window the display was set to mean luminance to the limits of the display.

6.2.2 Stimuli

The stimuli were luminance gratings with or without superimposed isotropic textures (see Figure 6.8). The grating stimuli were luminance sine waves, square waves, periodic saw-tooth or cropped saw-tooth functions with only 2 cycles of modulation visible. The textures were made in the same way as those used in previous experiments. All gratings had the same minimum and maximum luminance values. Without loss of generality, the median values for all functions were referenced as value zero. Within each cycle of the saw-tooth function the luminance profile formed a straight line running from minimum to maximum (Figure 6.8 c). The frequency of the sine wave was fixed to 0.2 c/d. All luminance profiles had the same wavelength so the square wave and saw-tooth function both had a fundamental frequency of 0.2 c/d. Thus for all types of profile except cropped saw-tooth, a display image contained 3~4 cycles. In this configuration, each stimulus contained either step edges in luminance or edges defined by zero crossings of the second derivative of luminance. Further, each periodical grating had at least two edges that were equal in magnitude and contrast polarity. The cropped saw-tooth stimuli, however, had only one edge between the two modulation cycles (although it also contained two edges that were shared between the figure and the background). When textures were superimposed, the combination of shading and texture was multiplicative such that the AM signal conveyed in the textures was positively correlated with the luminance profile conveyed in the shading. This is consistent with the shading of a Lambertian texture. The central frequency of the textures was 8 c/d. Stimuli were presented at three orientations (horizontal and ±45° relative to the right half of the horizontal axis). Figure 6.8 gives one example for each type of grating at 45° as well as their corresponding luminance cross-section measured on the diagonal indicated.

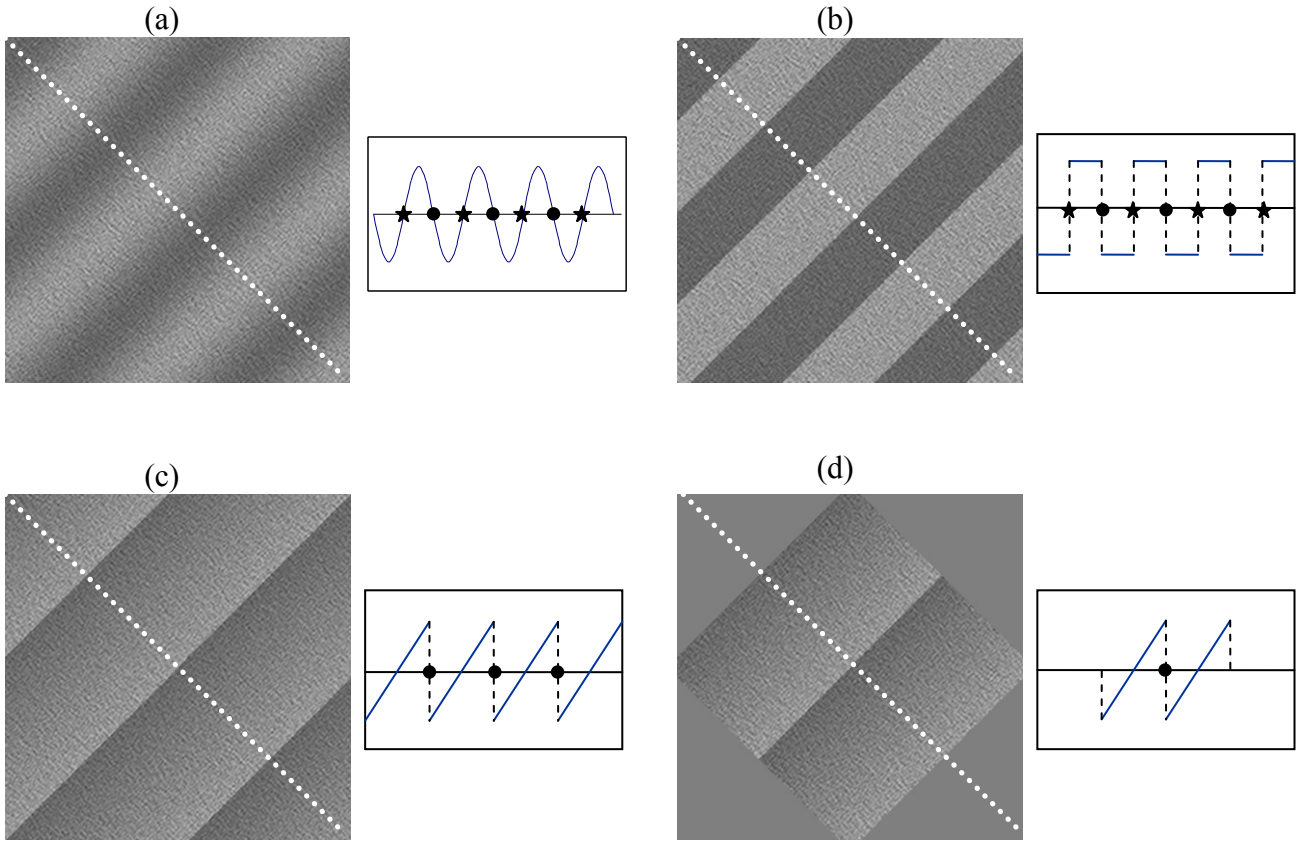


Figure 6.8 Four types of textured luminance profiles that observers viewed. The diagonal cross-sections (white dotted lines) of their LM component are plotted on the right of each stimulus. All gratings are at the orientation of 45deg. Sine wave (a) has a group of four (stars) and a group of three (circles) identical edges defined by zero-crossings of second-order derivatives. The two groups are different in the sign of the corresponding first-derivatives. Square wave (b) has a group of four (stars) and a group of three (circles) identical edges of the same contrast and the same polarity. One group differs from the other by the polarities of the edges (e.g. from dark to light vs. from light to dark). Periodical saw-tooth function (c) has three identical edges (circles) but the cropped saw-tooth (d) contains no identical edge pairs. The above four luminance patterns were also shown without textures and were tested separately.

6.2.3 Procedure

The cross-section of perceived shape was measured by using a gauge-figure comprising a disk (diameter 0.533 deg) and a perpendicular needle drawn at the centre of the disk (Koenderink et al., 1992). The aspect ratio of the disk and the direction and length of the needle were varied so as to represent the gauge-figure drawn at different slants according to linear perspective.

Stimuli consisted of gratings (Figure 6.9) onto which the gauge-figure was pasted. The slope of the gauge-figure was randomly initialized and was adjustable only in the

direction of the luminance variation. Observers viewed the images and were asked to adjust the apparent slope of the gauge-figure so that it matched that of the underlying surface (Figure 6.9). The step size for these adjustments was 10°. One cycle of sine wave and square wave modulations were measured (from a circle to the next circle in the cross-section profile in Figure 6.8) but for the saw-tooth and the cropped saw-tooth grating, two consecutive cycles of modulations were measured in order to make a valid comparison between the two saw-tooth functions. Testing points close to the edges in saw-tooth stimuli were moved by $1/24^{th}$ of the wavelength to avoid testing directly at the edge points. The measuring points were sampled at multiples of $1/8^{th}$ of a cycle of the grating (0.625 deg) but randomly displaced along the orthogonal direction. Thus the diameter of the disc (0.533 deg) was less than the sampling distance (0.625 deg). For the stimuli of sine-wave and square-wave, the measuring points also had a shift of integer cycles. The integer was randomly drawn from the set $\{-1,0,1\}$.

Each participant made 4 settings for each test position and the mean value of the 4 gradients were taken as the perceived slant at that location. The mean gradients were also integrated to get an estimate of the perceived depth profiles. In total there were:

$2\,(\text{textured or non-textured}) \times 3\,(\text{orientations}) \times 8\,(\text{positions}) \times 4\,(\text{repetitions}) = 192$

trials in the testing of sine wave and square wave luminance profile or

$2\,(\text{textured or non-textured}) \times 3\,(\text{orientations}) \times 18\,(\text{positions}) \times 4\,(\text{repetitions}) = 432$

trials in the testing of saw-tooth and cropped saw-tooth profiles. For each trial the image was generated online and trials were presented in a random order.

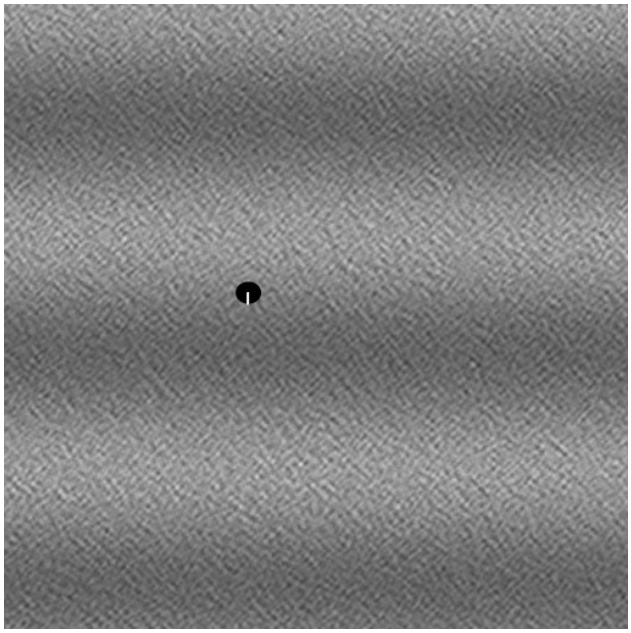


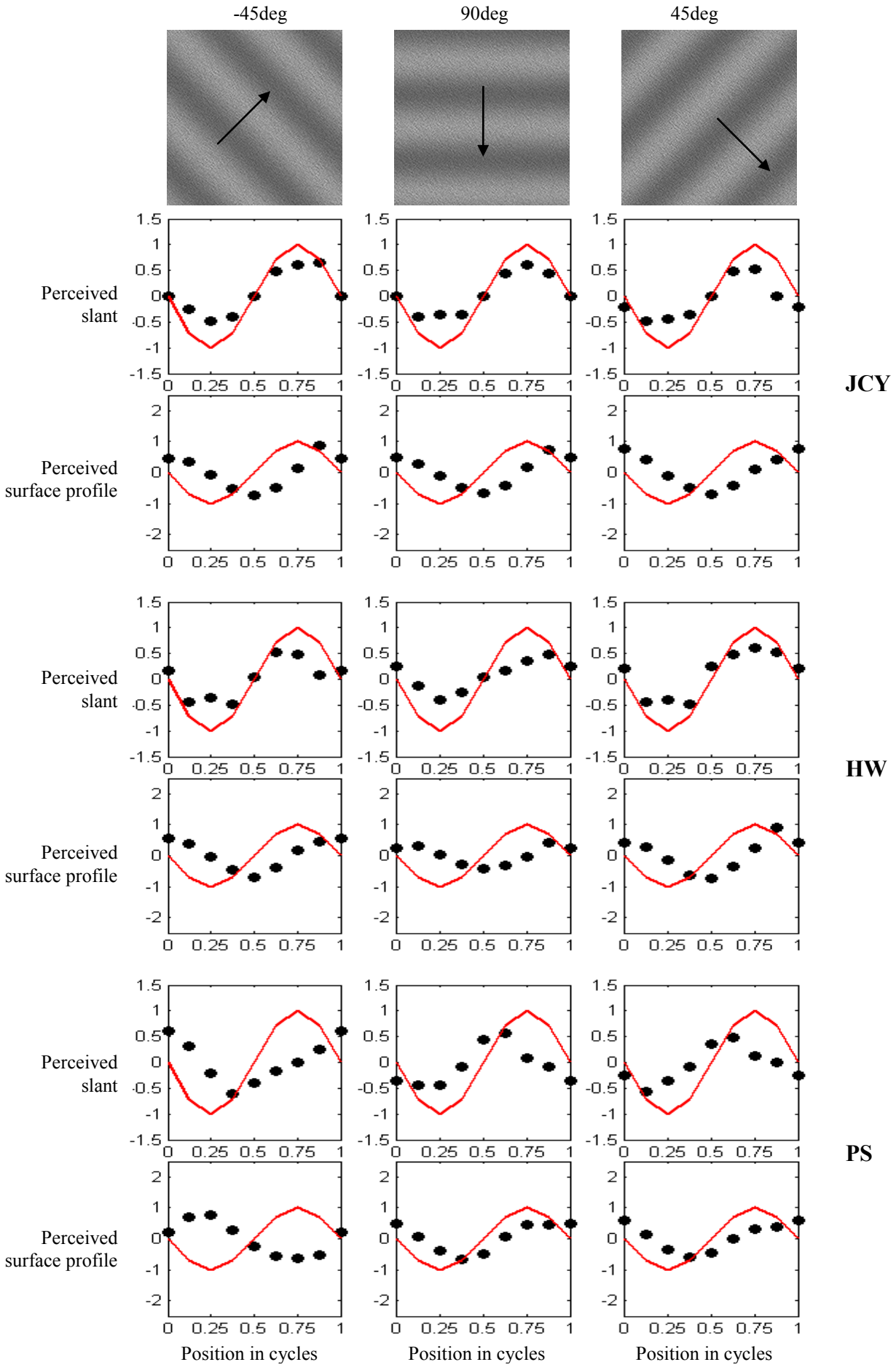
Figure 6.9 Images contained a stimulus and a probe. The probe was made adjustable along the direction in which the luminance is undergoing a variation (sinusoidal variation in this case).

6.2.4 Results

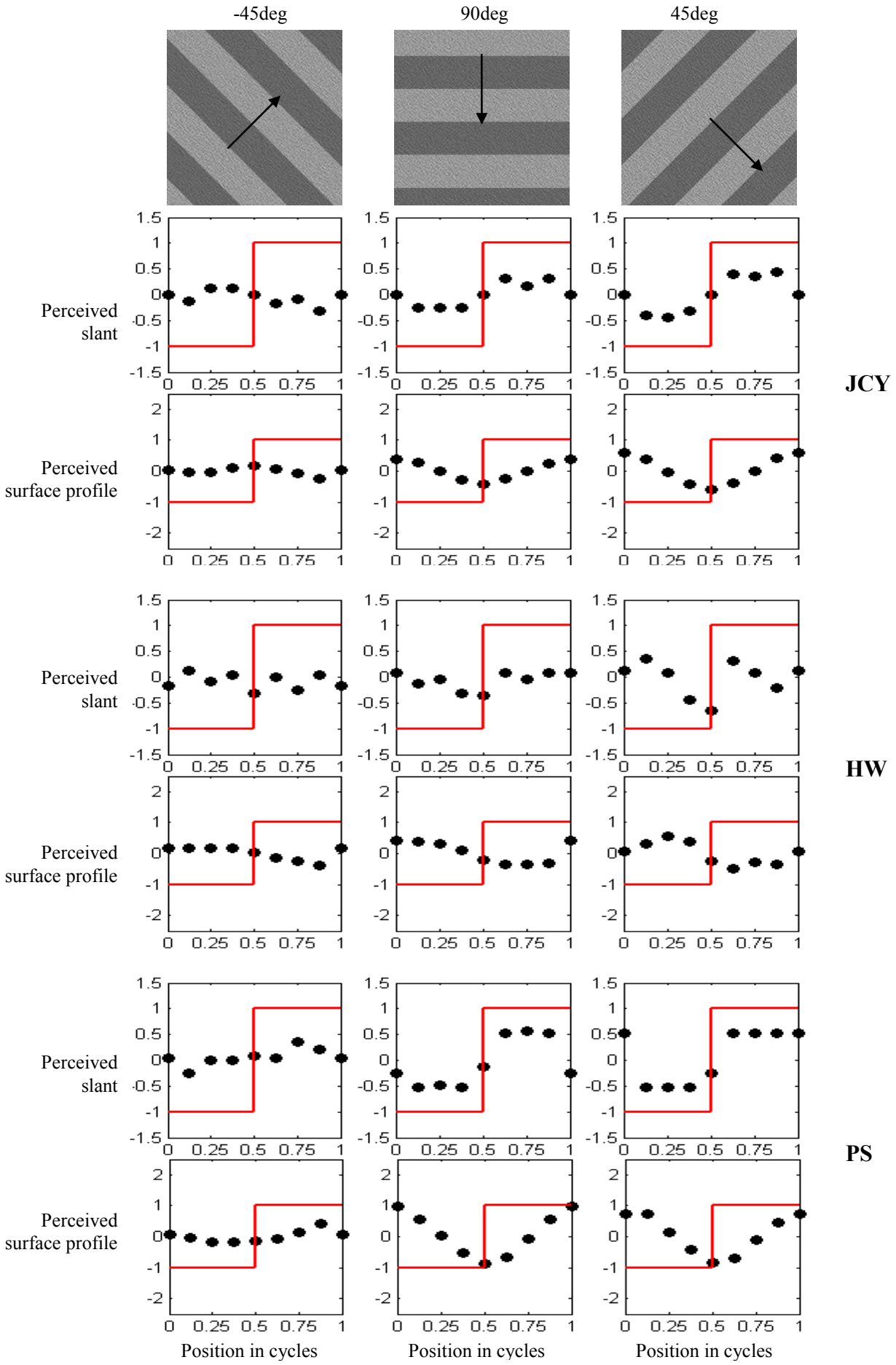
Three people took part in the experiment including two naïve participants (HW and JCY) and the author (PS). The naïve participants were paid for their efforts. The data for textured stimuli is given in Figure 6.10. The data for the non-textured stimuli are very similar to that for the textured case and are therefore not shown. The linear relationship between perceived slants and luminance as well as that between recovered surface heights and luminance were measured by calculating Pearson's correlation coefficients; see Tables 6.1 and 6.2.

All three participants agreed qualitatively on the surface shape for periodical saw-tooth gratings except for an ambiguity between concave and convex interpretations. The perceived slants appeared proportional to the luminance profiles of the stimuli (mean correlation 0.96). For the two naïve subjects, the sign of the relationship switched from positive to negative when the orientation of the saw-tooth gratings changed from 90° and +45° to -45° so that 90° and 45° periodical saw-tooth were

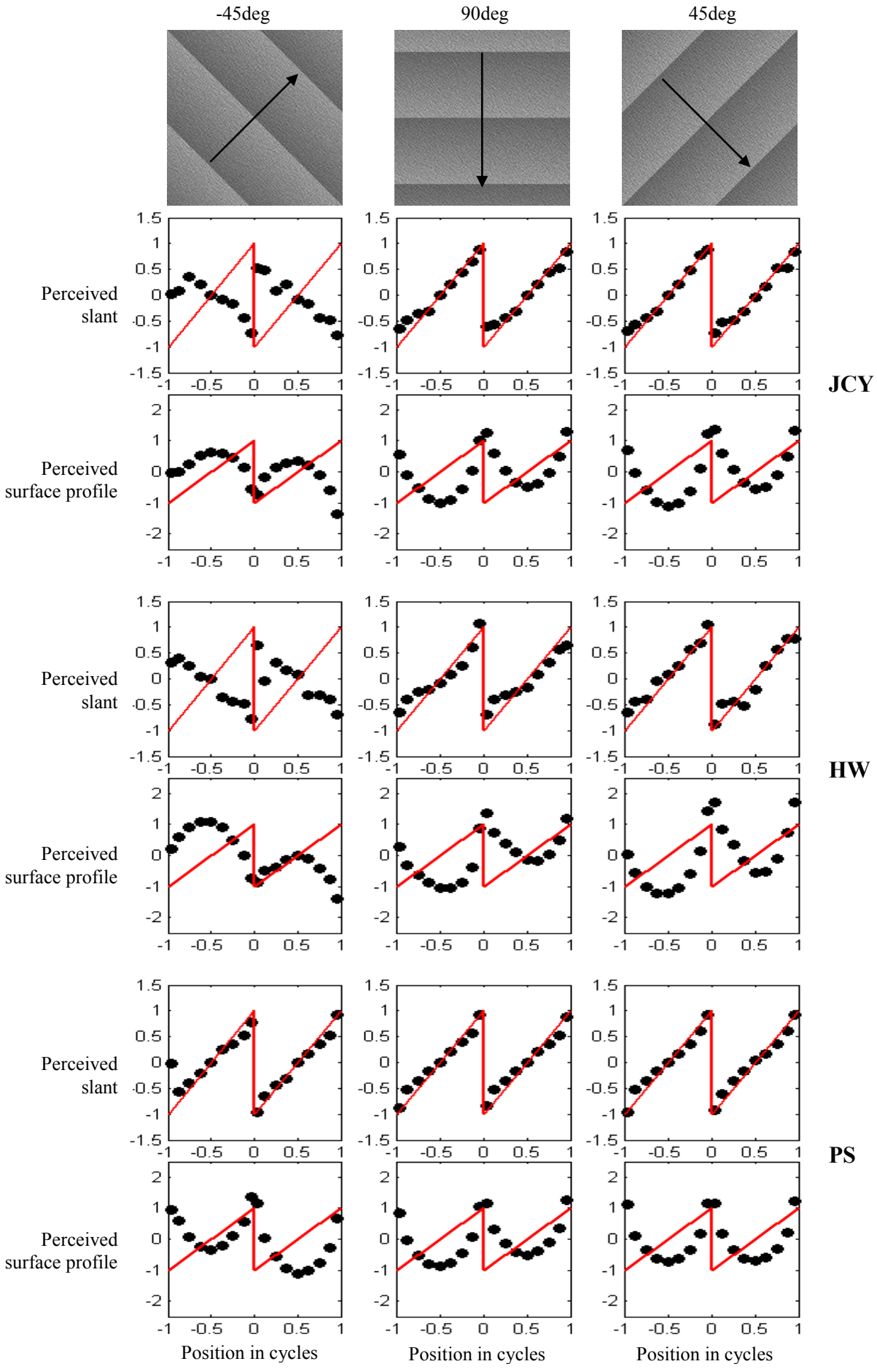
perceived as broad deep valleys with sharp ridges while -45° periodical saw-tooth was perceived as broad mounds with sharp valleys (Figure 6.10c). The other participant (PS) did not demonstrate this sign switch. For cropped saw-tooth gratings, when participants assumed concavity (the 90° and 45° gratings for JCY and HW but gratings in all three orientations for PS), gratings were no longer perceived as broad deep valleys with multiple ridges. The recovered surface looked more like a single crease formed by two curved surfaces. Departing from the middle ridge, gradients were initially proportional to luminance but quickly deviated from linearity towards stimulus borders. However when observers assumed convexity (the -45° grating for JCY and HW) gratings were still perceived as mounds with multiple values and the perceived gradients were still negatively proportional to the luminance.


-45deg
90deg
45deg
Perceived slant
Perceived surface profile
JCY
HW
PS
Position in cycles
Position in cycles
Position in cycles

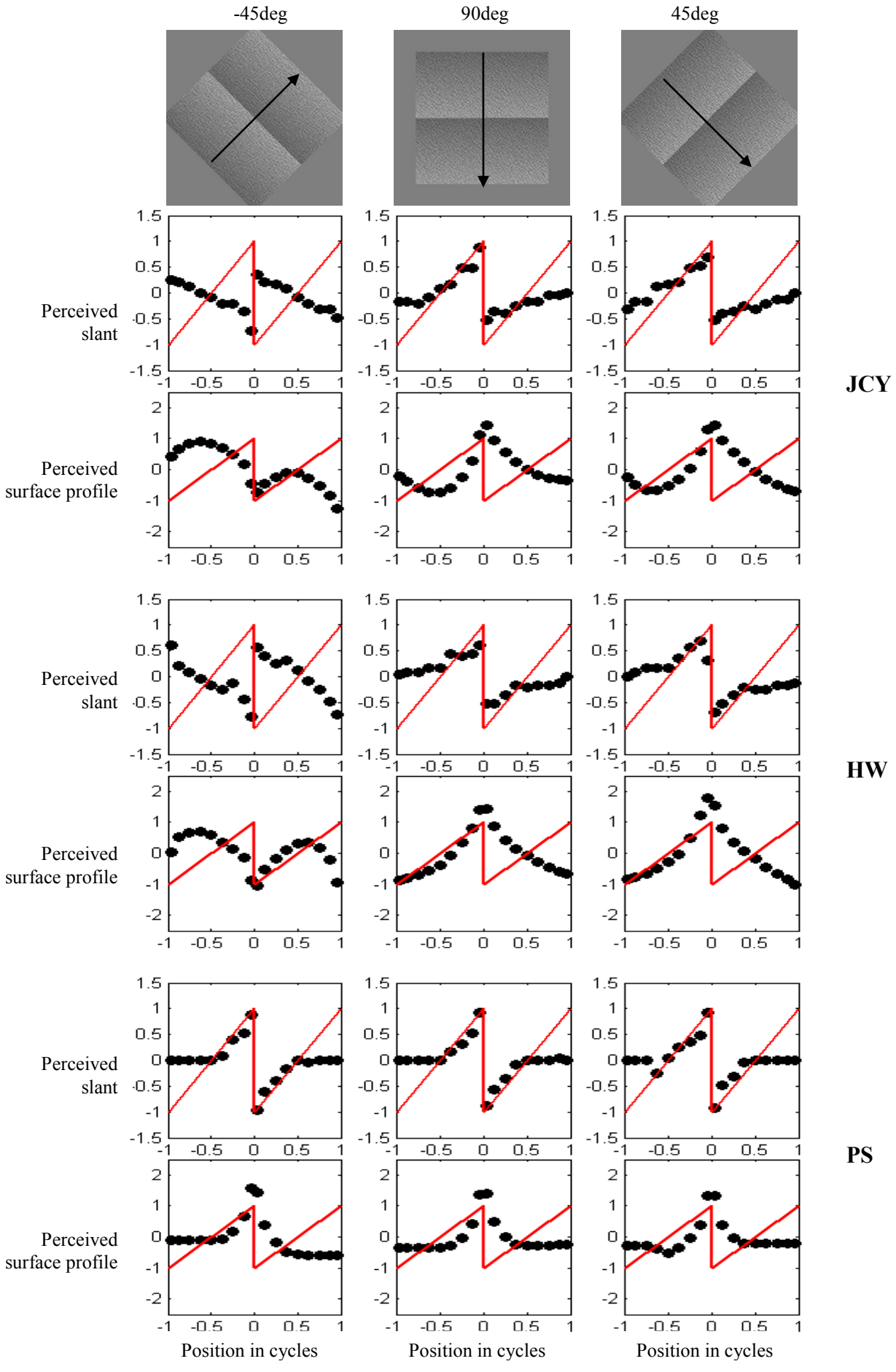
(a) sine wave



(b) square wave


-45deg
90deg
45deg
Perceived slant
Perceived surface profile
JCY
HW
PS
Position in cycles
Position in cycles
Position in cycles

(c) sawtooth



(d) cropped saw-tooth

Figure 6.10 Three participant's perceived slant and perceived surface profile for sine wave gratings (a), square wave gratings (b), periodical saw-tooth (c) and cropped saw-tooth (d). Results for stimuli with the same orientation are grouped in the same column. Solid lines represent the luminance profile. The observer's response is represented by dots. The horizontal axis is the spatial location in the unit of grating cycles. The black arrow indicates the direction of the luminance variation.

For sine-wave gratings, the two naïve subjects also assumed an approximately linear relationship between the perceived slants and the luminance (mean correlation 0.94). The recovered depth profiles for these two participants look like phase-shifted sine waves, which is consistent with what was reported in the previous two-point probe experiments (Chapter 2 and 3). For -45° sine-wave gratings, the other participant (PS) assumed a mapping that could not be accounted by a linear relationship between the perceived slants and the luminance values (correlation = 0.2). The recovered surface height for PS, however, appears to be in proportion to the luminance (correlation = -0.95). For 90° and 45° sine-wave grating, the correlations for PS's perceived slant and luminance values are 0.58 and 0.67 and those for PS's perceived height and luminance are 0.62 and 0.5.

PS and JCY perceived 90° and 45° square-wave stimuli as a triangle surface, a result of the linear relationship between the perceived slants and the luminance of the stimuli (mean correlation = 0.81). HW did not assume a linear relationship between the perceived slants and the luminance of 90° and 45° square-wave stimuli as the other two participants do (mean correlation = 0.25). But the relationship between the recovered surface height and the luminance is roughly linear (correlation = -0.73). For the -45° square wave, the recovered surface heights for all three participants do not display any observable patterns and the perceived slants appear to be distributed around zero.

To summarize, from table 6.1, data for saw-tooth gratings are very consistent across all participants. All three people set their perceived gradients proportional to the luminance profile, as indicated by the correlation coefficients extremely close to either 1 or -1. The sign of the relationship varied with the orientation of the grating. When the same saw-tooth gratings were cropped such that no equal edge pairs were present in the figure, the recovered surface profiles were qualitatively different for all participants and the linear relationship between the perceived slants and the luminance did not always hold.

Sine wave gratings were perceived quite depthy too, with the perceived gradients roughly proportional to the luminance profile under most conditions. The exception is PS's data for -45° sine-wave where the perceived heights rather than gradients were correlated to the luminance. The square wave looked the least depthy compare to the other two gratings but when they did look depthy to the observers, either their perceived gradients or perceived heights were still highly correlated with the luminance. Responses for the -45° square wave are as if they contained very little signal at all.

Luminance / Participants	Sine wave			Periodical saw-tooth			Cropped saw-tooth			Square wave		
	-45°	0°	45°	-45°	0°	45°	-45°	0°	45°	-45°	0°	45°
JCY (Naïve)	0.98	0.99	0.90	-0.9	0.99	0.99	-0.97	0.74	0.70	-0.61	0.85	0.86
HW (Naïve)	0.9	0.9	0.96	-0.94	0.97	0.97	-0.95	0.53	0.46	-0.25	0.44	0.06
PS (Author)	0.2	0.58	0.67	0.93	0.99	0.99	0.7	0.71	0.7	0.66	0.84	0.75

Table 6.1 Correlation coefficients between each observer's perceived gradients and the luminance profiles for all three types of stimuli. Most coefficients are quite high, suggesting the perceived gradients are highly correlated with the luminance profiles.

Luminance / Participants	Sine wave			Periodical saw-tooth			Cropped saw-tooth			Square wave		
	-45°	0°	45°	-45°	0°	45°	-45°	0°	45°	-45°	0°	45°
JCY (Naïve)	0.2	0.23	0.09	-0.23	0.05	0.04	-0.33	-0.17	-0.15	-0.28	0.002	0.03
HW (Naïve)	0.1	0.004	0.34	-0.29	-0.03	0.17	-0.19	0.04	0.04	-0.73	-0.66	-0.81
PS (Author)	0.95	0.62	0.5	-0.05	0.07	0.03	-0.08	0.02	0.02	0.58	0.04	0.16

Table 6.2 Correlation coefficients between each observer's perceived surface heights and the luminance for all three types of stimuli. Most coefficients are quite low but high correlations are found for when the corresponding cells in table 6.1 are low.

6.3.5 Discussion

The direction of luminance variations provides a cue for the direction of the illumination (Pentland, 1982). For the stimuli used in this experiment, the suggested illumination should be inline with the direction of the luminance variations (that is, perpendicular to lines of constant luminance in the non-textured stimuli). However, the sign of the direction of the illumination should be determined by the lighting assumptions of individual observers in order to resolve problems such as the convex / concave ambiguity (Ramachandran, 1988; Sun & Perona, 1998; Mamassian & Goutcher, 2001). Thus the direction of the assumed light source can be obtained from the convexity or concavity of the perceived surface. When interpreting horizontal gratings, results showed that all observers clearly used a light from above prior (see Figure 6.10). For all 45° gratings, the suggested direction of the illumination (inline with the direction of the luminance variation) should be either above-left or below-right. The results suggest when perceiving 45° gratings, observers preferred light from above-left. For -45° gratings, the suggested illumination directions should be either from above-right or below-left to be inline with the luminance variation. But results were not very consistent for this orientation. For -45° sine wave gratings, observer JCY and HW seemed to prefer light from below-left. In contrast JCY preferred the light from above-right for -45° square wave and saw-tooth gratings. HW also

preferred the light from above-right for -45° saw-tooth gratings. PS's judgments for -45° saw-tooth gratings were consistent with light from below-left but his light assumption for -45° sine-wave could not be explained by oblique lighting. Generally speaking, neither light from above-right nor below-left is the most favourite lighting prior so observers did not demonstrate a strong preference for one against the other. In fact, the flattened recovered surface for -45º square wave grating may well be due to the action of two contradictory lighting assumptions working against one another. Since a new image was generated during each trial, there might be a flip between the two contradicting lighting assumptions, resulting in the perceived gradients cancelling one another out making mean gradients much lower than might have been the case on individual trials.

Edges played an important role for perceiving saw-tooth gratings. When luminance gradients were bounded by equal polarity edges, human performance can be predicted by a linear solution to equation 6.6. For cropped saw-tooth stimuli, edges between figure and background may be still active during the task but the strengths of the edges were not equal. Under this condition, the linear relationship broke for concavely perceived surfaces but still held for when surfaces were perceived as convex. For sine waves and square waves, for which both of equal edge pairs and edge pairs with opposite signs coexist, some observers mapped luminance to perceived slant with a linear function, but not all of the time. However in the cases where this linear relationship did not hold and perceived depth was not flat, the recovered surface height tended to be proportional to the luminance, consistent with the "dark is deep" rule.

6.4 Experiment 2

It was shown above that for the three types of luminance variations that were bounded by equal polarity edges, perceived gradients were approximately proportional to luminance. But when the boundary condition was violated as was the case of cropped saw-tooth stimuli, the linear relationship no longer held, at least when they were perceived as concave. In some cases, however, perceived gradients did not correlate with luminance, but luminance did correlate with perceived height. These two relationships represent two distinct computations associated with human SFS: slant $\propto$ luminance and dark is deep. It is expected that edges are important in deciding which computation to carry out. However in experiment 1, equal edges and edges with opposite polarity coexist in periodical sine-wave and square-wave gratings, which could have been the cause of the fact that both types of computations were observed for those stimuli. The effect of edge polarities were further investigated in this experiment.

6.4.1 Stimuli

The stimuli were made from the same sine-wave and square wave gratings as in experiment 1 but were always superimposed with isotropic textures. Some gratings were cropped and the retained section contained 1.2 cycles such that the only remaining visible edges had opposite polarities (see Figure 6.11). For cropped gratings, phases were fixed such that peak luminance always appeared in the centre of the screen and the background was set to the medium luminance. Stimulus orientation was fixed at 45°.

6.4.2 Procedure

Perceived shape was measured by the same gauge figure task as in experiment 1 except that the disk had a smaller diameter of 0.48 deg. Steps of adjustment were made either 1° or 10° so that observers could choose either the coarse or fine adjustment. The measuring points were sampled at multiples of $1/10^{th}$ of a cycle of the grating (0.5 deg) but randomly displaced along the orthogonal direction. Thus the diameter of the disc (0.48 deg) was less than the sampling distance (0.5 deg). The measuring positions were arranged in a way that edges in square wave gratings were excluded. Measuring positions started at $1/20^{th}$ of a cycle from the top left edge of the cropped stimuli and at a similar position relative to the centre of the un-cropped stimuli (see Figure 6.11). Each type of stimulus remained on the screen and the gauge figure appeared in random order at the measuring positions. Observers made four disk settings per position. Unlike experiment 1, each stimulus remained on the screen until the participant completed all the trials for that stimulus. The four types of stimuli were displayed in random order. Mean gradients were integrated to provide an estimate of the perceived depth profiles. In total each participant completed:

$$10\,(\text{positions}) \times 4\,(\text{repetitions}) \times 4\,(\text{stimuli}) = 160$$

trials during the experiment. Observers were likely to reset their "beholder's share" when viewing new images in each trial during experiment 1. The effect of such resetting was minimised in experiment 2 to avoid the possible cancellations found in the results of experiment 1.

6.4.3 Results and discussions

Three naïve participants were tested in experiment 2. None of them had participated in experiment 1. Figure 6.12 describes the data in a similar format to Figure 6.10.

Table 6.3 gives the Pearson correlation coefficients between perceived slant and luminance, and between perceived height and luminance respectively.

Perceived surface profiles for un-cropped sine wave gratings were similar to those of experiment 1: Two observers (TT and ZXQ) produced a gradient profile linearly related to luminance (correlations = 0.98 and 0.87). The correlations between their perceived heights and luminance were both low (-0.44 and -0.28). The two coefficients for the other observer (KL) were both at a medium level (0.67 and 0.58). For cropped sine waves, no participants assumed a linear relationship between perceived gradients and luminance (correlations = -0.23, -0.2 and -0.27). However the correlation between perceived height and luminance all increased (correlations = 0.96, 0.76 and 0.94). When viewing un-cropped square waves, all participants agreed on a linear relationship between gradients and luminance (correlations = 0.98, 0.97 and 0.98). The correlations between heights and luminance were consistently low (correlations = -0.32, -0.64 and -0.5). But for the cropped square-wave this pattern was destroyed (correlations between gradient and luminance = -0.26, -0.23 and -0.37). Instead perceived height and luminance were correlated (0.76, 0.69 and 0.73).

For sine wave gratings, a linear gradient model will produce a sinusoidal surface with a 90° phase shift to the luminance whereas a "dark is deep" model will produce a near sinusoidal profile that is in-phase with luminance. The perceived shape of un-cropped sine wave could be explained by the linear gradient model for two observers however they both switched to a "dark is deep" model when the sine wave was cropped. From the graph (bottom left in Fig 6.12a), shape judgment for the other participant also appeared as a sinusoidal surface but with a smaller phase shift than predicted by a

linear gradient model. As if it were a combination of the two model predictions. But this participant also switched to the "dark is deep" model when judging cropped sine waves.

For un-cropped square waves, all participants' performance could be explained by the linear gradient model. Removing edge pairs with equal polarities made all observers changed their strategy. Although correlations between luminance and perceived surface heights increased significantly, they were not as high as for cropped sine wave gratings. Graphically, it is also very clear that shape perceptions for cropped square waves did not exactly follow the luminance trace. However observers qualitatively agreed on their perceived shapes which appeared approximately as trapezoidal surfaces. This also suggests that the adopted new strategy should be consistent across all observers.

Whether the new strategy was the same as that for cropped sine waves is open to discussion. By comparison, perceived shapes for these two types of stimuli were qualitatively similar except that one was smoothly curved and the other was made of planar surfaces. Considering the similarities of the two luminance traces, it is possible that observers switched to the same strategy when edges with equal polarities were removed. If this was true, the "dark is deep" rule would not serve as a perfect model to characterise the unknown strategy, though it might provide an approximate model.

Luminance / Participants	Sine		Cropped Sine		Square		Cropped square	
	Gradient	Height	Gradient	Height	Gradient	Height	Gradient	Height
TT	0.98	-0.44	-0.23	0.96	0.98	-0.32	-0.26	0.76
ZXQ	0.87	-0.28	-0.2	0.76	0.97	-0.64	-0.23	0.69
KL	0.66	0.58	-0.27	0.94	0.98	-0.5	-0.37	0.73

Table 6.3 Pearson coefficients between each observer's perceived gradients and the luminance, as well as between perceived surface heights and luminance for all stimuli.

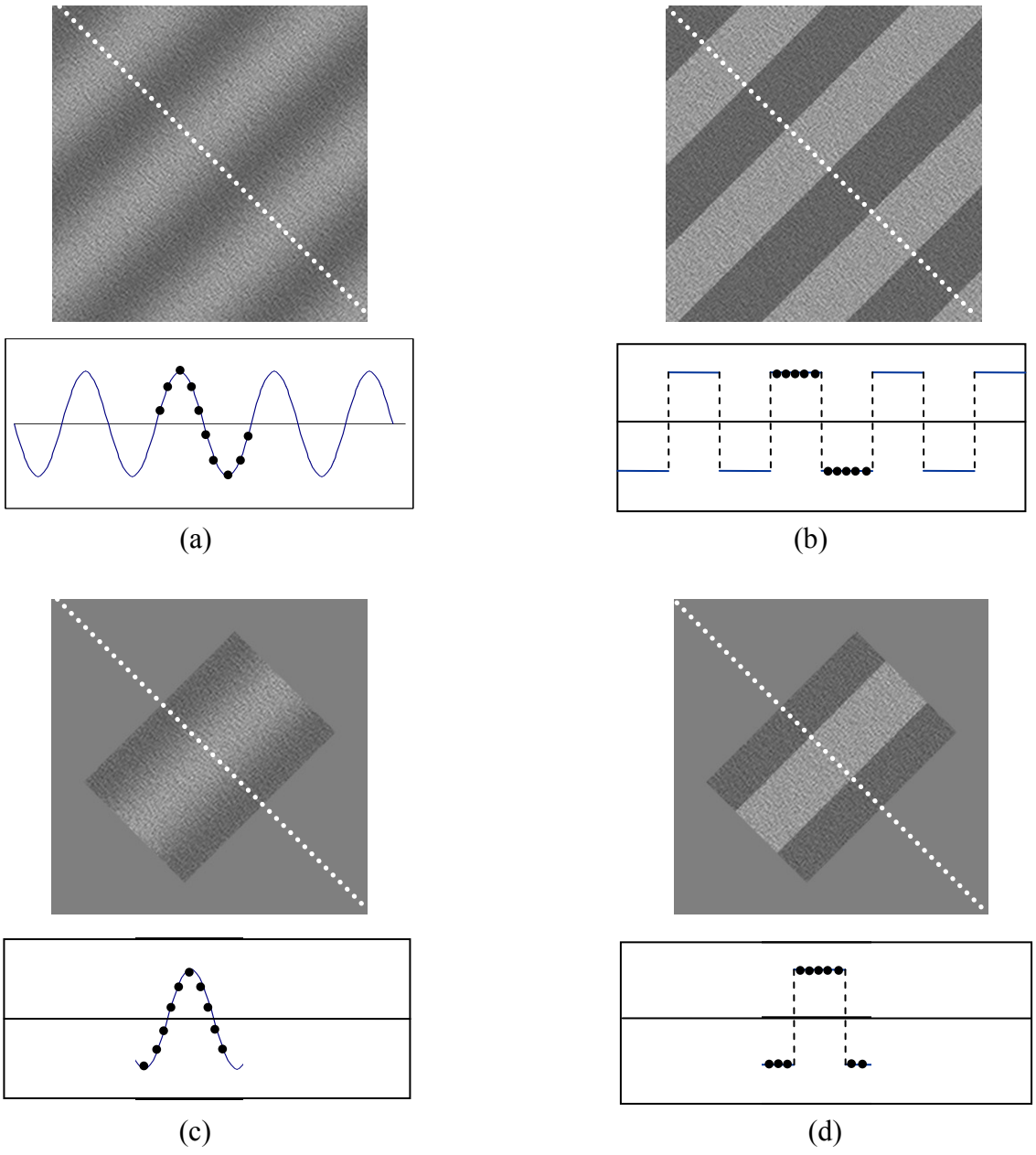
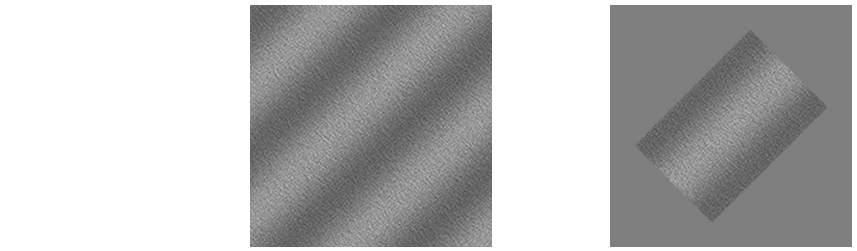
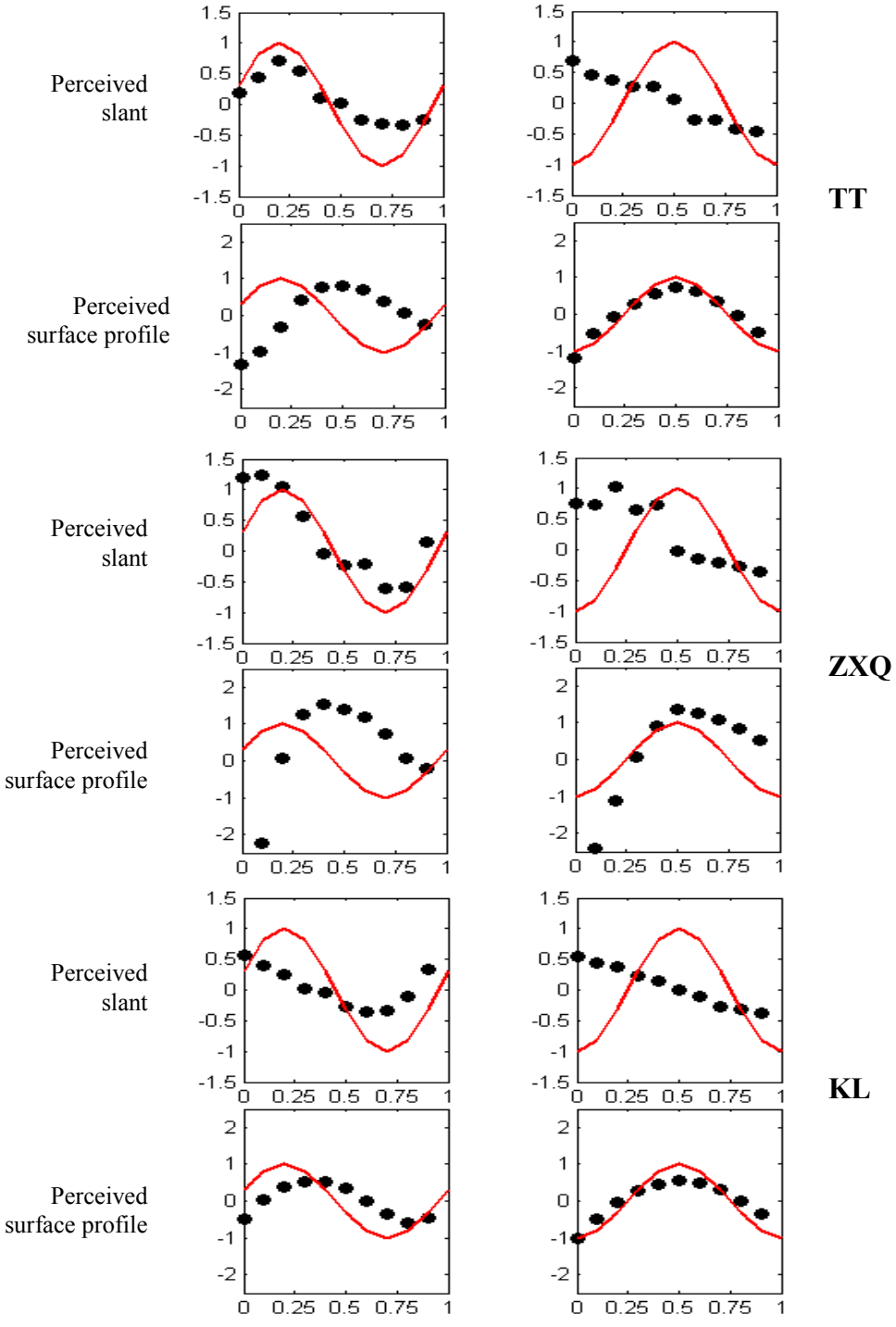
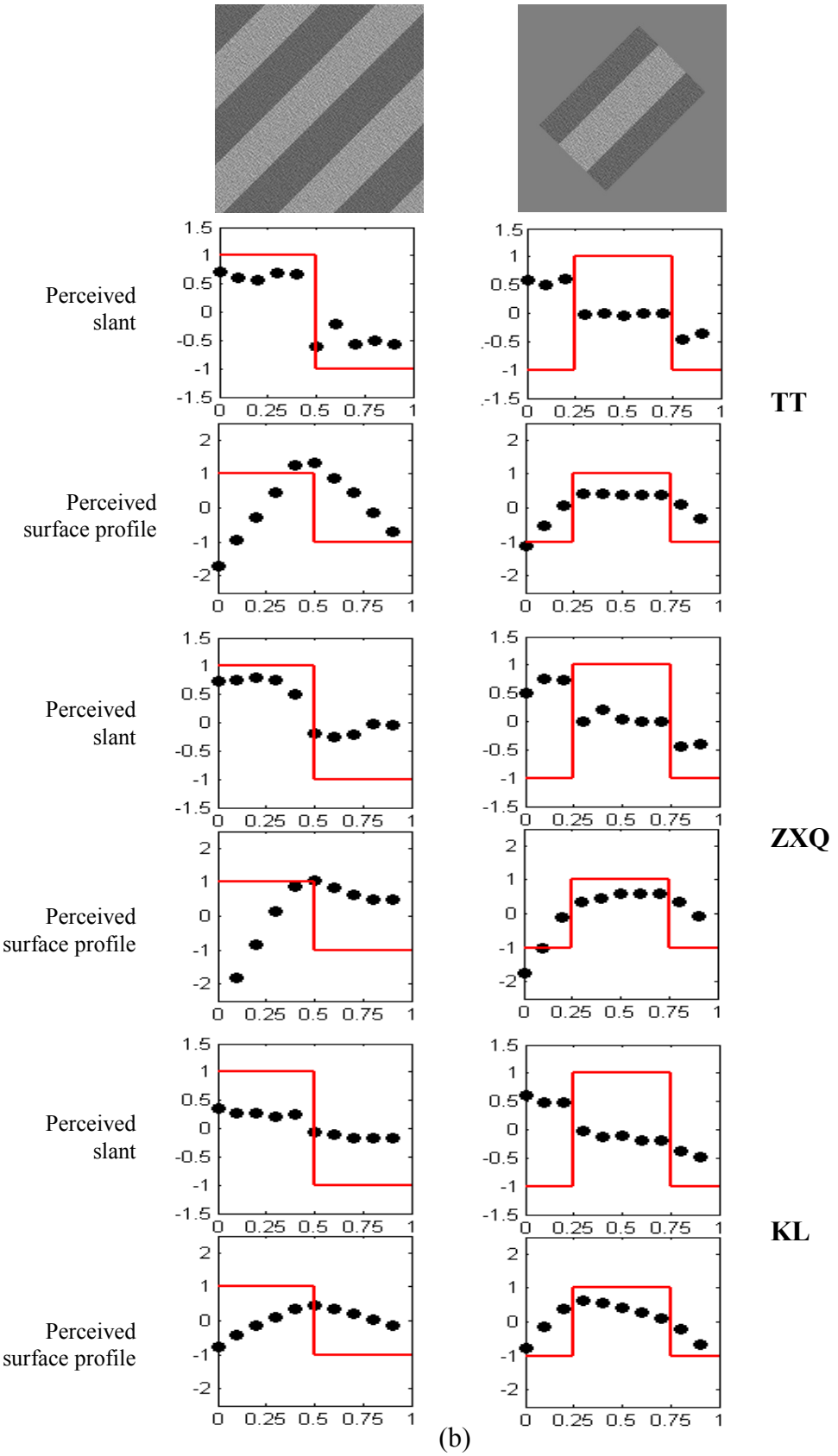


Figure 6.11 Stimuli in experiment 2. sinewave (a) and square wave (b) are the same as in experiment 1 except their phase were fixed during the experiment. (c) and (d) are cropped version of (a) and (b) respectively. The visible portions in (c) and (d) are 1.2 cycles of the periodical gratings. (a) and (c), (b) and (d) are shifted by 90°. The dots mark the ten measuring positions within a cycle of the test gratings.




Perceived slant
Perceived surface profile
TT
Perceived slant
Perceived surface profile
ZXQ
Perceived slant
Perceived surface profile
KL

(a)



(b)

Figure 6.12 Three participants' perceived slants and perceived surface profiles for sine wave gratings (a), square wave gratings (b). Legends are same as in Fig 6.8

6.5 General discussion

The major finding in experiment 1 was that observers assumed a linear relationship between luminance and perceived slant for periodic saw-tooth gratings when luminance was bounded by equal edges. The relationship no longer held when these boundary condition were violated, at least when the surface was perceived as concave. The linear relationship for the other two gratings was most pronounced when the suggested light source directions were consistent with the light from above left prior. Results for other stimuli may have been compromised by the concave / convex ambiguity. But other patterns of behaviour were also found; some suggesting a "dark is deep" model. In these stimuli, edge pairs with equal and opposite polarities coexisted and this may have confounded the results. Experiment 2 was conducted to examine the role of edge polarities in determining the computation of SFS, while reducing the cancellations caused by unfavourable assumed lighting directions. Results in experiment 2 suggested that when edges with equal sign were removed, perceived slant was no longer linear but indicated a computation that can be approximated by a "dark is deep" model. Taken together, two types of computations could be identified and are discussed in the following subsections.

6.5.1 The linear reflectance mode (LRM)

Human SFS is operated a linear reflectance model (LRM) at least when the polarities of two boundary edges are the same. This model is based on solving equation

$$C\frac{d^2 z(x)}{dx^2} = I'(x). \tag{6.7}$$

The solution is given by:

$$z(x) = a\int I(x)dx + bx + c \tag{6.8}$$

where $z(x)$ is the depth function of a surface, $I(x)$ is shading image, a, b, c are coefficients to be determined by observers. Equation (6.8) is a 1-D version of the ambiguity function of human SFS defined in equation (6.3). Humans must resolve the ambiguities during SFS. To determine b, the difference in height between two boundary positions is required. With equal boundary conditions, the perceived gradients will be linearly related to luminance, i.e. b equals zero. On a surface, equal boundary condition means that two surface positions are at the same height relative to the image plane. Other things being equal, edges with similar contrast under LRM are likely to be treated by humans as being at roughly equal height. a, c are left to individuals to resolve using their "beholder's share". But when information on the relative heights of two boundary positions is not available, all three parameters are left completely to the individuals "beholder's share". This "Beholder's share" appeared quite different across observers when cropped saw-tooth grating appeared concave, as can be concluded by the different relative distances between the two boundary positions on the recovered surface. When surfaces appeared convex, observers still resolved the ambiguity by assigning roughly same surface height to boundary positions, resulting in linear relationships between perceived slant and luminance. The central idea of LRM is that surface shape is coded in the format of equation (6.7). The behavioural response to SFS tasks under this mode is concerned with a specific realization of equation (6.8), that is, assigning values to the three coefficients a, b, c based on visual cues in the image as well as observers' "beholder's share". This idea is consistent with the claim that the visual system codes surface curvature in the process of SFS (Johnston & Passmore, 1994b), because $z''(x)$ is a good approximation of surface curvature.

The idea that slant is proportional to luminance when bounded by equal edges is also consistent with the bumpy perception of 1D luminance gradient bounded by a circular contour. Take the luminance gradient as the gradient of the surface and integrate column by column along the vertical direction between the boundaries. The integration will give rise to a series of quadratic curves with domes at different height, approximating a bumpy sphere. In comparison, when a 1D gradient is bounded by a square, the same process will give rise to a series of quadratic curves with domes at the same height, leading to a cylindrical perception (See Figure 6.13).

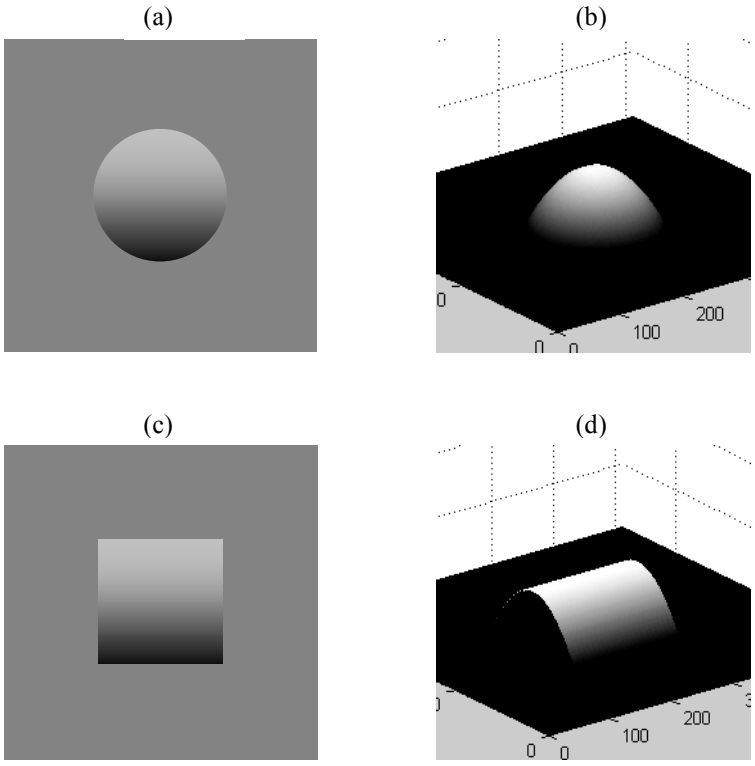


Figure 6.13 1D luminance gradient can be perceived as bump when it is bounded by a circular contour (a) but also be perceived as cylindrical when it is bounded by a square (c). (c) and (d) were obtained by solving the ordinary differential equation 6.7 with equal boundary conditions at their surrounding contours. Results were produced using a simple algorithm based on the descriptions in the text.

6.5.2 "Diffuse or frontal" lighting mode

When luminance variations are not bounded by equal polarity edges, observers are likely to adopt a different strategy. Cropped sinusoids were perceived as sinusoidal

surfaces without a phase shift, in contrast to the phase shifted sinusoidal surfaces observed elsewhere (Schofield et al., 2006; Pentland, 1988; chapter 2, 3). Cropped square waves were perceived as trapezoid shapes in contrast to triangle shaped surfaces found for un-cropped square waves. Perceived shape for cropped sine-wave can be explained by a "dark is deep" model. Perceived shape for cropped square-wave didn't fully obey "dark is deep" but was broadly consistent with a Lambertian surface illuminated by diffuse lighting (see Fig6.7c). Under diffuse lighting a trapezoidal surface will produce three patches of uniform luminance because points on each planar surface see the same portion of the lighting hemisphere. But the planar surface on top will appear lighter than the other because it is exposed to the entire light source. The two surfaces on the side are less bright as back planes stop light coming in from behind. However this doesn't mean that the new computational strategy is designed exclusively for the condition of diffuse lighting. For example a trapezoidal surface will produce similar shading patterns under collimated frontal lighting as well. Also, the "dark is deep" rule reported by Christou and Koenderink (1997) is most pronounced when the direction of the light source was close to the viewing direction (frontal lighting). But what is certain is that this new strategy corresponds to a lighting condition which is either diffuse or, if collimated, frontal.

The computation under the new strategy is not very clear either. While the "dark is deep" rule predicted sinusoidal stimuli well, it does less well for square waves. In Langer and Bülthoff's experiment (2000), the accuracy of the depth comparison task performed with diffusely lit surface was still above chance level for the "anti-correlated" condition (see 6.1.4). If human SFS under diffuse lighting completely

obeyed “dark-is-deep” rule, the accuracy would have been close to zero. Thus “dark-is-deep” rule doesn’t fully describe human SFS under diffusing lighting.

6.5.3 Human SFS operates in distinct modes

Given a shading image, the visual system should first decide through which computational strategy the shading will be interpreted. The two known operational modes correspond to two lighting conditions: collimated oblique illumination and an illumination that is either diffuse or collimated but frontal. Edge polarities are likely to play a role in making the decision. Luminance variations bounded by edges with same polarity are likely to trigger the implementation of LRM but otherwise a “dark-is-deep” rule or a variant of it might prevail. An example is the equilateral triangle wave (Fig 6.14) which has similar luminance profiles as a sine-wave grating but does not have any zero crossing in the second-derivative and therefore does not have any edges. As shown in Figure 6.14, the perception of this luminance variation seems to follow the “dark is deep” rule instead of the “slant proportional to luminance” rule. When edge pairs with both equal an opposite polarities are present in an image, humans may decide in accordance with probabilities of each mode in natural scenes. Data suggests that LRM tends to be preferably weighed which is consistent with what was found for a natural scene interpretation task (Pentland, 1988). But some participants seemed to combine the two modes. The product of LRM operating on a sinusoid is a sinusoid with 90° phase shift. But under the other mode, no phase shift is obtained. A linear combination of the two operations will give rise to a sinusoid with a phase shift in the range of 0°-90°, consistent with PS and KL’s responses for 45° sine wave gratings.

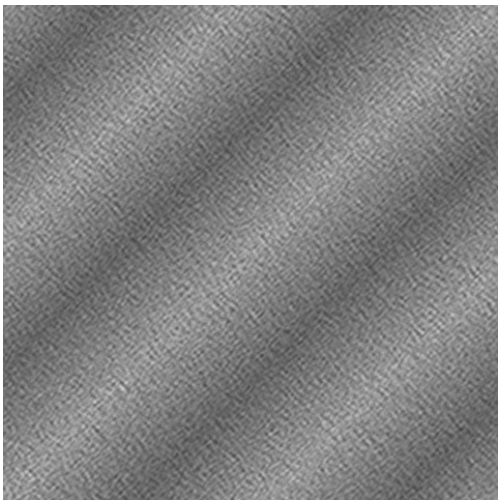


Figure 6.14 An equilateral triangle wave appears an equilateral triangle surface.

6.5.4 Psychological plausibility of distinct modes in human SFS

Shading is ambiguous. For each possible lighting direction, there exists a corresponding surface in a family of affine transformation to generate the same shading pattern (Belhumeur et al., 1999). To obtain a unique solution of the surface, humans must have a unique and stable prior knowledge on light source tilt and slant (together they form light source direction). Unfortunately, human lighting priors are thought to span a wide range of tilt angles and priors for slant remain unknown. When estimating light source direction, humans demonstrate very poor accuracies and individual differences are huge. Thus it is unlikely that human SFS can achieve a unique surface representation with a specific light source direction. Rather it is more plausible that human SFS interprets shading in terms of a set of 3-D surfaces. To achieve this, the interpretation has to be conducted without precise knowledge of light source. In other words, human SFS is mostly dependent on shading patterns and is insensitive to small changes in light source directions. This is exactly what has been reported regarding to the lack of shape constancy under changing lighting directions in the literature (Khang et al., 2007; Christou & Koenderink, 1997; Todd et al., 1996).

On the other hand, humans should be more sensitive to changes in lighting patterns or large changes in lighting directions under which the formulation of shading may also change. Thus it is reasonable that human SFS switches its operational mode in response to apparent changes in the illumination pattern. Indeed, different behaviours have been reported for different lighting conditions during a curvature discrimination task (Johnston & Passmore, 1994a; Curran & Johnston, 1996), surface attitude judgement tasks on rendered images (Christou & Koenderink et al., 1997; Langer & Bulthoff, 2000; Nefs, 2008) and surface attitude judgement tasks for photographs of real objects (Todd et al., 1996). For simple images like those used here, the decision on which mode to operate is based on the polarities of edge pairs bounding the luminance variations. But it may not be as straight forward for natural images. However it is still possible that switching between the operational modes is cued by distributions of edges. There is evidence suggesting that the activities of edge detectors in a complex images made up of Gaussian textures can be decisive in light field estimation tasks (Koenderink et al., 2007).

7. Conclusion

The human visual system is thought to comprise a series of modules each specializing in a particular task, one of which is SFS. The aim of the thesis was to investigate the operation of the two sub-modules within the SFS module. The two stages studies are luminance disambiguation and the estimation of surface height from shading components. The major findings about each computational stage and their validity are summarised in the following sections.

7.1 Second-order vision in luminance disambiguation

This stage is closely related to the theory that luminance variations are separated into layers by visual system according to their origins in the scene (e.g. changes due to illumination and surface reflectance might be separated at this stage; Kingdom, 2008). The theory of layer segmentation coincides well with SFS as ideally human SFS is based on intrinsic shading instead of raw luminance variations. Among many others, texture amplitude is an effective cue used by humans to differentiate changes in reflectance from illumination (Schofield et al., 2006). The first three chapters of this thesis were dedicated to further examining the characteristics of this cue as well as proposing a neural mechanism to explain the computations involved.

It is well known that humans are sensitive to stimuli consisting of second-order signals (Chubb & Sperling, 1988; Cavanagh & Mather, 1989; Wilson et al, 1992). Moreover, it is now clear that the visual system dedicates a separate multi-channel mechanism to processing second-order signals (see for example, Schofield & Georgeson, 1999). Studies of the distribution of first-order and second-order signals in natural scenes point towards the idea that second-order signals (more precisely the

relationship between first- and second-order signals) may convey important information about the scene. However, the role of second-order vision in our daily experiences is not well understood. Schofield et al. (2006) proposed that the relationship between first-order luminance signal (LM) and a second-order entity AM determine whether the luminance variations have the appearance of shading or reflectance. Further, the effectiveness of differentiating illumination from reflectance changes was found to vary with the underlying strength of the AM signal (Schofield et al., 2010), suggesting second-order vision could play a role in luminance disambiguation. To further verify this hypothesis, chapters 2 and 3 tested the effect of carrier frequency on the impression of SFS. The results showed that the impression of corrugations versus flatness varied with the carrier frequency in a similar way to second-order vision, providing further evidence of the active role of second-order vision in the process of luminance disambiguation. Through another route, reducing the frequency of the texture components gradually made them appear more like shading. This finding is consistent with another heuristic that shading in natural scenes are normally made of low frequency components (Kingdom, 2008). Taking these results together, it is proposed that layer decomposition based on texture amplitude is conducted by retrieving second-order signals through a second-order channel.

Based on Schofield et al.'s data (2010), Chapter 4 established a computational strategy to differentiate between changes in illumination and reflectance. First-order luminance variation and second-order amplitude modulations were extracted separately and were then combined at a later stage. A contrast gain control circuit then applied cross-inhibition among multiple channels. The output of the model is a set of

scalar values representing the strength of shading at each frequency and orientation (the model was only implemented with two such channels). The inverse operation then multiplied the component strengths with their corresponding basis functions to recover the full shading image. The whole process is analogue to applying a linear operation (e.g. Fourier transform) decomposing the image into bands of different frequencies and orientations. The coefficients of these bands are either suppressed or boosted according to the accompanying second-order information, followed by a final cross-inhibition stage before transforming the retained components back into the spatial domain. A parallel process can be used to extract reflectance components to form a reflectance image. Chapter 4 also suggested a neural mechanism which could conduct the proposed computation. The neural mechanism consisted of multiple shading channels each containing two separate sub-channels to retrieve first-order and second-order information respectively. The two sub-channels within each shading channel were tuned to the same frequency and orientation. Responses of both sub-channels were summed and the squared energy of the summation was taken as the strength of the shading channel. The proposed neural mechanism is consistent with known physiology in early visual area. Cells have been found in cat area 17 and 18 that are responsive to both first-order gratings and second-order contrast modulated envelopes (Mareschal and Baker, 1998a; 1998b). Further, when presented with a combination of first-order and second-order signals, the response of such cells varied with the phase relationship of the two components with response peaks at zero phase differences and troughs at 180° of phase shift (Hutchinson et al., 2007). The characteristic of such cells are similar to the proposed neural mechanism.

Not only did the model provide a good fit to Schofield et al.'s data (2010), it could also capture the trend of the data obtained in Chapters 2 and 3, after some adjustment. The adjustment was made to the output of the model without any changes to its inner structure and parameter settings. The analysis of the difference between the two types of studies proved that such adjustments were justifiable.

7.2 Application in Intrinsic image separation

The biologically inspired model proposed in Chapter 4 was competent for images consisting of one or a very small number of frequency bands, as demonstrated by the output images in Chapter 4. But it is not a mature solution for real images which are broadband in frequency and orientation. The reason for this is that the experimental data are only available for stimuli made of single frequency component and two orientation components. The study of cross-frequency inhibition is also rather incomplete in the literature (see Meese, 2004). Therefore the parameters for inhibitory terms acting across frequency channels can not be determined.

To provide a solution for image processing, Chapter 5 adopted a framework similar to the classic Retinex algorithm (Land & McCann 1971; Horn 1974) and replaced the original gradient classification rule with the one derived from psychological experiments on second-order cues (Schofield et al., 2006). The algorithm assumed that local contrast should be constant within a uniform flat surface under changing illumination. Thus any changes in local contrast should be due to reflectance. The algorithm compared luminance edges in the original image with edges in local contrast and deleted those luminance edges whereby edges in local contrast were also found. Due to the fact that edges in local contrast and edges in luminance never coincided exactly, a width estimation algorithm was used which provided tolerance

for such mismatch. Results showed that the algorithm performed reasonably well on images containing large patch of shadows, with distortions along sharp shadow boundaries. Possible improvements include using a more accurate texture segmentation algorithm to find the edges in local contrast or adding other local features and using texture edges as a global constraint. Note that the output from either type of the model comprised components due to generalised changes in illumination. The models do not distinguish shading from cast shadows.

7.3 Computing 3-D shape from shading

The module for computing of 3D shape from shading assumes that its input contains only shading information. Human SFS has been an active research topic for more than two decades. Yet it is still far from determining the computational algorithm for this aspect of human vision. The impression that humans can achieve a coherent representation of the 3-D world under changing illumination and surface material suggests that the computation of human SFS is not unique (no single and simple computation can suit all situations) and may be too complicated to determine. However studies on the shape constancy of SFS have alleviated this concern as humans are incapable of deriving a constant shape perception under changing illumination and surface reflectance (Khang et al., 2007; Christou & Koenderink, 1997). Instead, SFS was largely dependent on the underlying shading patterns (Khang et al., 2007). This suggests that human SFS may in fact rely on a rather simple, if fallible, computation. Thus chapter 6 used a different methodology aimed at establishing a computational theory for human SFS—testing human shape perception with non-naturalistic luminance variations. These stimuli did not provide any information about the identity of the object thus prevented interferences from high

level object recognition. Image outlines were not indicative of any 3-D information either refraining other depth cues to override shading.

Chapter 6 also introduced a computational scheme to explain the SFS data. The scheme proposed two distinct computational modes for human SFS. In the linear reflectance model (LRM), the recovered surface height is one of a family of solutions to an ordinary differential equation. When human observers assumed equal height at the two boundaries, the solution is consistent with the traditionally held view that "perceived slant is proportional to luminance". This mode is consistent with collimated lighting from an oblique angle. In the other mode, recovered surface height is indicative of a surface under a lighting that is not "collimated and oblique". To some extent, the computation under this mode could be accounted for by the "dark is deep" rule. Switching between these two modes was related to the sign of the two edges at the boundaries of the stimulus. LRM was switched on when two boundary edges had the same sign of contrast. The dark-is-deep mode operated when two boundary edges had oppositely signed contrasts. When both types of edge boundaries existed, human SFS preferred LRM but could demonstrate a combination of the two operations.

The proposed theory could explain a number of known characteristics of human SFS. The computations in the two modes do not require precise knowledge of the lighting direction. This is consistent with the discovery that the process of SFS was independent of light source estimation (Mingolla & Todd, 1986; Mamassian et al., 1996) and that perceived curvature remained constant under small changes in lighting directions as long as the lighting was not frontal (Curran & Johnston, 1994). But

humans do need a rough idea of the lighting in terms of whether it is directionally oblique or not. One cue for obtaining this rough knowledge could be the sign of edge boundaries. This is consistent with the report that distributions of edges were decisive in the light source estimations by humans (Koenderink et al., 2007). Human observers were found to overestimate surface slant when the actual slant was small but immediately started to underestimate it when the actual slant increased. This can be explained by the LRM model. An examination of those stimuli for which such performance was reported reveals that those stimuli were indeed under oblique lighting conditions and were bounded by edges with same polarities.

Due to the bas-relief ambiguity (Belhumeur et al., 1999), any given shading pattern corresponds to a family of depth functions and a set of illuminations. Only when the precise direction of the illumination is available, can the solution of the 3-D shape be uniquely determined. Thus the proposed SFS theory implies that the exact 3-D shape is not represented uniquely in the visual system. Given shading alone human SFS must derive a family of functions to describe the 3D shape. Under LRM, this family of functions are solutions to an ordinary differential equation which codes the second derivative of the surface with differences in luminance, consistent with the claim that it is the surface curvature that is coded in the 3-D vision (Johnston & Passmore, 1994b). Humans then need to place further constraints to choose from among the family of solutions. This can explain the large individual differences often found during SFS experiments despite the tendency for observers to agree on the qualitative shape perceived.

Questions still remain regarding to the exact computation of the other mode and how the two modes should be combined. However the proposed theory is pioneering in that it is the first attempt to establish a computational theory for human SFS and it disassociates the shape computation with precise light source estimation and surface material, which is seemingly how human behaved in the reported studies of SFS.

Reference List

Adams,W.J. (2007). A common light prior for visual search, shape and reflectance judgments. *Journal of Vision*, **7**, 1-7.

Adelson,E.H. (1993). Perceptual organization and the judgment of brightness. *Science*, **262**, 2042-2044.

Adelson,E.H. & Pentland,A. (1996). The Perception of Shading and Reflectance. In *Perception as Bayesian Inference*, eds. Knill,D.C. & Richards,W., pp. 409-423. New York: Cambridge University Press.

Agrawl,A., Raskar,R., and Chellappa,R. Edge suppression by gradient field transformation using cross-projection tensors. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. *Proceedings CVPR'2006* **2**, 2301-2308. 2006.

Albrecht,D.G. & Hamilton,D.B. (1982). Striate cortex of monkey and cat: contrast response function. *Journal of Neurophysiology* , **48**, 217-237.

Albright,T.D. (1992). Form-cue invariant motion processing in primate visual cortex. *Science*, **255**, 1141-1143.

Anderson,B.L. & Winawer,J. (2005). Image segmentation and lightness perception. *Nature*, **434**, 79-83.

Arsenault,A.S., Wilkinson,F. & Kingdom,F.A. (1999). Modulation frequency and orientation tuning of second-order texture mechanisms. *Journal of the Optical Society of America*, **16**, 427-435.

Baker,C.L.Jr. (1999). Central neural mechanisms for detecting second-order motion. *Current Opinion in Neurobiology*, **9**, 461-466.

Barrow,H.G. & Tenenbaum,J.M. (1978). Recovering Intrinsic Scene Characteristics from Images. In *Computer Vision Systems*, eds. Hanson,A. & Riseman,E., pp. 3-26. New York: Accademic Press.

Barrow,H.G. & Tenenbaum,J.M. (1981). Interpreting line drawings as three-dimensional surfaces. *Artificial Intelligence*, **17**, 75-116.

Battu,B., Kappers,A.M.L. & Koenderink,J.J. (2007). Ambiguity in pictorial depth. *Perception*, **36**, 1290-1304.

Belhumeur,P.N., Kriegman,D.J. & Yuille,A.L. (1999). The bas-relief ambiguity. *International Journal of Computer Vision*, **35**, 33-44.

Blake,A. (1985). Boundary conditions for lightness computation in Mondrian World. *Computer Vision, Graphics, and Image Processing*, **32**, 314-327.

Bloj,M.G., Kersten,D. & Hurlbert,A.C. (1999). Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature*, **402**, 877-879.

Bonds,A.B. (1989). Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neuroscience*, **2**, 41-55.

Brainard,D.H., Pelli,D.G. & Bobson,T. (2002). Display characterization. In *Encyclopedia of Imaging Science and Technology*, ed. Hornak,J., pp. 172-188.

Brewster,D. (1826). On the optical illusion of the conversion of cameos into intaglios, and intaglios into cameos, with and account of other analogous phenomena. *Edinburgh Journal of Science*, **4**, 99-108.

Bruce,V., Green,P.R. & Georgeson,M. (1996). *Visual Perception: Physiology, Psychology and Ecology (3rd Ed)*. Psychology Press Ltd.

Bulthoff,H.H. & Mallot,H.A. (1988). Integration of depth cues: stereo and shading. *Journal of the Optical Society of America*, **5**, 1749-1758.

Campbell,F.W. & Kulikowski,J.J. (1966). Orientational selectivity of the human visual system. *Journal of Physiology London*, **187**, 437-445.

Campbell,F.W., Cleland,B.G., Cooper,G.F. & Enroth-Cugell,C. (1968). The angular selectivity of visual cortical cells to moving gratings. *Journal of Physiology London*, **198**, 237-250.

Campbell,F.W., Cooper,G.F. & Enroth-Cugell,C. (1969). The spatial selectivity of visual cells of the cat. *Journal of Physiology London*, **203**, 223-235.

Carandini,M. & Heeger,D.J. (1994). Summation and division by neurons in primate visual cortex. *Science*, **264**, 1333-1336.

Carandini,M., Heeger,D.J. & Movshon,J.A. (1999). Linearity and gain control in V1 simple cells. In *Cerebral Cortex: Volumes 13: Models of Cortical Circuits*, eds. Ulinski,P.P., Jones,E.G. & Peters,A., pp. 401-436. New York: Plenum Publishers.

Cavanagh,P. & Mather,G. (1989). Motion: the long and short of it. *Spatial Vision*, **4**, 103-129.

Christou,C. & Koenderink,J.J. (1997). Light source dependency in shape from shading. *Vision Research*, **37**, 1441-1449.

Chubb,C. & Sperling,G. (1988). Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America*, **5**, 1986-2007.

Clows,M.B. (1971). On seeing things. *Artificial Intelligence*, **2**, 79-116.

Curran,W. & Johnston,A. (1994). Integration of shading and texture cues: testing the linear model. *Vision Research*, **34**, 1863-1874.

Curran,W. & Johnston,A. (1996). The effect of illuminant position on perceived curvature. *Vision Research*, **36**, 1399-1410.

Dakin,S.C. & Mareschal,I. (2000). Sensitivity to contrast modulation depends on carrier spatial frequency and orientation. *Vision Research*, **40**, 311-329.

De Valois,R.L., Albrecht,D.G. & Thorell,L.G. (1982). Spatial frequency selectivity of cells in Macaque visual cortex. *Vision Research*, **22**, 545-559.

DeCarlo,D., Finkelstein,A., and Rusinkiewicz,S. Interactive rendering of suggestive contours with temporal coherence. Third International Symposium on Non-Photorealistic Animation Rendering. *Proceedings NPAR'2004* , 135-145. 2004.

Durou,J.D., Falcone,M. & Sagona,M. (2008). Numerical methods for shape-from-shading: A new survey with benchmarks. *Computer Vision and Image Understanding*, **109**, 22-43.

Ellemberg,D., Allen,H.A. & Hess,R.F. (2006). Second-order spatial frequency and orientation channels in human vision. *Vision Research*, **46**, 2798-2803.

Erens,R.G.F., Kappers,A.M.L. & Koenderink,J.J. (1993a). Perception of local shape from shading. *Perception and Psychophysics*, **54**, 145-156.

Erens,R.G.F., Kappers,A.M.L. & Koenderink,J.J. (1993b). Estimating local shape from shading in the presence of global shading. *Perception and Psychophysics*, **54**, 334-342.

Finlayson,G.D., Hordley,S.D., Lu,C. & Drew,M.S. (2006). On the removal of shadows from images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **28**, 59-68.

Foley,J.M. (1994). Human luminance pattern-vision mechanisms: masking experiments require a new model. *Journal of the Optical Society of America*, **11**, 1710-1719.

Forsyth,D.A. & Ponce,J. (2002). *Computer Vision: A Modern Approach*. Prentice Hall.

Funt,B.V., Drew,M.S., and Brockington,M. Recovering shading from color images. Second European Conference on Computer Vision. *Proceedings ECCV'92* , 124-132. 1992.

Georgeson,M. & Schofield,A.J. (2002). Shading and Texture: separate information channels with a common adaptation mechanism? *Spatial Vision*, **16**, 59-76.

Georgeson,M., May,K.A., Freeman,T.C. & Hesse,G. (2007). From filters to features: Scale-space analysis of edge and blur coding in human vision. *Journal of Vision*, **7**, 1-21.

Gibson,J.J. (1950). The perception of visual surfaces. *The American Journal of Psychology*, **100**, 646-664.

Gilchrist,A.L. (1977). Perceived lightness depends on perceived spatial arrangement. *Science*, **195**, 185-187.

Gilchrist,A.L. (1979). The perception of surface blacks and whites. *Scientific American*, **240**, 112-123.

Gilchrist,A.L., Delman,S. & Jacobsen,A. (1983). The classification and integration of edges as critical to the perception of reflectance and illumination. *Perception and Psychophysics*, **33**, 425-436.

Gilchrist,A.L. (1988). Lightness contrast and failures of constancy: A common explanation. *Perception and Psychophysics*, **43**, 415-424.

Gilchrist,A.L. (2006). *Seeing Black and White*. Oxford University Press.

Graham,N. & Sutter,A. (1998). Spatial summation in simple (Fourier) and complex (Non-Fourier) texture channels. *Vision Research*, **38**, 231-257.

Graham,N. & Sutter,A. (2000). Normalization: contrast-gain control in simple (Fourier) and complex (non-Fourier) pathways of pattern vision. *Vision Research*, **40**, 3737-2761.

Guzman,A. (1969). Decomposition of a Visual Scene into Three-dimensional Bodies. In *Automatic Interpretation and Classification of Images*, ed. Grasselli,A., pp. 243-276. New York, London: Academic Press.

Hann,E.D., Erens,R.G.F. & Noest,N.J. (1995). Shape from shaded random surfaces. *Vision Research*, **35**, 2985-3001.

Harmon,L.D. & Julesz,B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. *Science*, **180**, 1194-1197.

Heeger,D.J. (1993). Modeling simple cell direction selectivity with normalized, half squared, linear operator. *Journal of Neurophysiology*, **70**, 1885-1898.

Heeger,D.J., Simoncelli,E.P., and Movshon,J.A. Computational models of cortical visual processing. Vision: From Photon to Perception. *Proceedings National Academy of Sciences 93*, 623-627. 1996.

Henning,G.B., Hertz,B.G. & Broadbent,D.E. (1975). Some experiments bearing on the hypothesis that the visual system analyses spatial patterns in independent bands of spatial frequency. *Vision Research*, **15**, 887-897.

Hills,J.M., Watt,S.J., Landy,M.S. & Banks,M.S. (2004). Slant from texture and disparity cues: optimal cue combination. *Journal of Vision*, **4**, 967-992.

Horn,B.K.P. (1974). Determining lightness from an image. *Computer Graphics and Image Processing*, **3**, 277-299.

Horn,B.K.P. (1975). Obtaining Shape from Shading Information. In *The Psychology of Computer Vision*, ed. Winston,P.H., pp. 115-155. New York: McGrawHill.

Horn,B.K.P. (1977). Understanding image intensities. *Artificial Intelligence*, **8**, 201-231.

Horn,B.K.P. & Sjoberg,R.W. (1979). Calculating the reflectance map. *Applied Optics*, **18**, 1770-1779.

Horn,B.K.P. & Brooks,M.J. (1989). *Shape from Shading*. MIT Press.

Horn,B.K.P. (1989). Height and gradient from shading. *International Journal of Computer Vision*, **5**, 37-75.

Hubel,D. & Wiesel,T. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology London*, **160**, 106-154.

Hutchinson,C.V., Baker,C.L.Jr., and Ledgeway,T. Response to combined first-order and second-order motion in visual cortex neurons. Perception (Supplement) 36, 305-306. 2007.

Ikeuchi,K. & Horn,B.K.P. (1981). Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, **17**, 141-184.

Jamar,J.H.T. & Koenderink,J.J. (1985). Contrast detection and detection of contrast modulation for noise gratings. *Vision Research*, **25**, 521.

Johnson,A.P. & Baker,C.L.Jr. (2004). First- and second-order information in natural images: a filter-based approach to image statistics. *Journal of the Optical Society of America*, **21** , 913-925.

Johnston,A. & Passmore,P.J. (1994a). Shape from shading: Surface curvature and orientation. *Perception*, **23**, 169-189.

Johnston,A. & Passmore,P.J. (1994b). Independent encoding of surface orientation and surface curvature. *Vision Research*, **34**, 3005-3012.

Johnston,E.B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, **31**, 1351-1360.

Julesz,B. (1960). Binocular depth perception of computer-generated patterns. *Bell System Technical Journal*, **39**, 1125-1162.

Khang,B.G., Koenderink,J.J. & Kappers,A.M.L. (2007). Shape from shading from images rendered with various surface types and light fields. *Perception*, **36**, 1191-1213.

Kingdom,F.A. & Keeble,D.R.T. (1996). A linear systems approach to the detection of both abrupt and smooth spatial variations in orientation-defined textures. *Vision Research*, **36**, 409-420.

Kingdom,F.A. (2003). Colour brings relief to human vision. *Nature Neuroscience*, **6**, 641-644.

Kingdom,F.A., Prins,N. & Hayes,A. (2003). Mechanism independence for texture-modulation detection is consistent with a filter-rectify-filter mechanism. *Visual Neuroscience*, **20**, 65-76.

Kingdom,F.A., Beauce,C. & Hunter,L. (2004). Colour vision brings clarity to shadows. *Perception*, **33**, 907-914.

Kingdom,F.A. (2008). Perceiving light versus material. *Vision Research*, **48**, 2090-2105.

Kleffner,D. & Ramachandran,V.S. (1992). On the perception of shape from shading. *Perception and Psychophysics*, **52**, 18-36.

Knill,D.C. & Kersten,D. (1991). Apparent surface curvature affects lightness perception. *Nature*, **351**, 228-230.

Knill,D.C. (1992). Perception of surface contours and surface shape: from computation to psychophysics. *Journal of the Optical Society of America*, **9**, 1449-1464.

Koenderink,J.J., van Doorn,A.J. & Kappers,A.M.L. (1992). Surface perception in pictures. *Perception and Psychophysics*, **52**, 487-496.

Koenderink,J.J., van Doorn,A.J., Christou,C. & Lappin,J.S. (1996a). Shape constancy in pictorial relief. *Perception*, **25**, 155-164.

Koenderink,J.J., van Doorn,A.J., Christou,C. & Lappin,J.S. (1996b). Perturbation study of shading in pictures. *Perception*, **25**, 1009-1026.

Koenderink,J.J., van Doorn,A.J. & Kappers,A.M.L. (2001). Ambiguity and the 'mental eye' in pictorial relief. *Perception*, **30**, 431-448.

Koenderink,J.J., van Doorn,A.J. & Pont,S.C. (2004). Light direction from shad(ow)ed random Gaussian surfaces. *Perception*, **33**, 1405-1420.

Koenderink,J.J. & van Doorn,A.J. (2004). Shape and Shading. In *The Visual Neurosciences*, eds. Chalupa,L.M. & Werner,J.S., pp. 1090-1105.The MIT Press.

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

Land,E.H. & McCann,J.J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America*, **61**, 1-11.

Landy,M.S., Maloney,L.T., Johnston,E.B. & Mark,Y. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, **35**, 389-412.

Langer,M.S. and Zurcker,S.W. Qualitative shape from active shading. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. *Proceedings CVPR'92* , 713-715. 1992.

Langer,M.S. & Bulthoff,H.H. (2000). Depth discrimination from shading under diffuse lighting. *Perception*, **29**, 649-660.

Langer,M.S. & Bulthoff,H.H. (2001). A prior for global convexity in local shape-from-shading. *Perception*, **30**, 403-410.

Lawlor,M., Holtmann-Rice,D., Huggins,P., Ben-Shahar,O. & Zurcker,S.W. (2009). Boundaries, shading, and border ownership: A cusp at their interaction. *Journal of Physiology Paris*, **103** , 18-36.

Ledgeway,T., Zhan,C., Johnson,A.P., Song,Y. & Baker,C.L.Jr. (2005). The direction-selective contrast response of area 18 neurons is different for first- and second-order motion. *Visual Neuroscience*, **22**, 87-99.

Legge,G.E. & Foley,J.M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, **70**, 1458-1471.

Li,S., Tan,P., and Lin,S. Intrinsic image decomposition with non-local texture cues. 2008 IEEE Conference on Computer Vision and Pattern Recognition. *Proceedings CVPR'2008* , 1-7. 2008.

Li,Y., Pizlo,Z. & Steinman,R.M. (2009). A computational model that recovers 3D shape of an object from a single 2D retina representation. *Vision Research*, **49**, 979-991.

Lindeberg,T. (1993). Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction. *International Journal of Computer Vision*, **3**, 349-376.

Lindeberg,T. (1998). Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, **20**, 117-154.

Liu,B. & Todd,J.T. (2004). Perceptual biases in the interpretation of 3D shape from shading. *Vision Research*, **44**, 2135-2145.

Malik,J. (1987). Interpreting line drawings of curved objects. *International Journal of Computer Vision*, **1**, 73-107.

Malik,J. & Perona,P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America*, **7**, 923-932.

Mamassian,P., Kersten,D. & Knill,D.C. (1996). Categorical local shape perception. *Perception*, **25**, 95-107.

Mamassian,P. & Kersten,D. (1996). Illumination, shading and the perception of local orientation. *Vision Research*, **36**, 2351-2367.

Mamassian,P. & Goutcher,R. (2001). Prior knowledge on the illumination position. *Cognition*, **81**, B1-B9.

Marcelja,S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America*, **70**, 1297-1300.

Mareschal,I. & Baker,C.L.Jr. (1998). Temporal and spatial response to second-order stimuli in cat area 18. *Journal of Neurophysiology*, **80**, 2811-2823.

Mareschal,I. & Baker,C.L.Jr. (1998). A cortical locus for the processing of contrast-defined contours. *Nature Neuroscience*, **1**, 150-154.

Mareschal,I. & Baker,C.L.Jr. (1999). Cortical processing of second-order motion. *Visual Neuroscience*, **16**, 527-540.

Marr,D. & Nishihara,H.K. (1980). Theory of edge detection. *Proceedings of the Royal Society of London B*, **207**, 187-217.

Marr,D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: W. H. Freeman.

Meese,T.S. & Hess,R.F. (2004). Low spatial frequencies are suppressively masked across spatial scale, orientation, field position, and eye of origin. *Journal of Vision*, **4**, 843-859.

Meese,T.S. & Holmes,D.J. (2007). Spatial and temporal dependencies of cross-orientation suppression in human vision. *Proceedings of the Royal Society London B*, **274**, 127-136.

Metelli,F. (1974). The perception of transparency. *Scientific American*, **230**, 90-98.

Mingolla,E. & Todd,J.T. (1986). Perception of solid shape from shading. *Biological Cybernetics*, **53**, 137-151.

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**, 275-287.

Nefs,H.T., Koenderink,J.J. & Kappers,A.M.L. (2006). Shape-from-shading for matte and glossy objects. *Acta Psychologica*, **121**, 297-316.

Nefs,H.T. (2008). Three-dimensional object shape from shading and contour disparities. *Journal of Vision*, **8**, 1-16.

Norman,J.F. & Todd,J.T. (1996). The discriminability of local surface structure. *Perception*, **25**, 381-398.

Olmos,A. & Kingdom,F.A. (2004). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, **33**, 1463-1473.

Oruc,I., Maloney,L.T. & Landy,M.S. (2003). Weighed linear cue combination with possibly correlated error. *Vision Research*, **43**, 2451-2468.

Palmer,S.E. (1999). *Vision Science: Photons to Phenomenology*. The MIT Press.

Pankanti,S. & Jain,A.K. (1995). Integrating vision modules: stereo, shading, grouping and ling labeling. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **17**, 831-842.

Pentland,A. (1982). Finding the illuminant direction. *Journal of the Optical Society of America*, **72**, 448-455.

Pentland,A. (1984). Local shading analysis. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **6**, 170-187.

Pentland,A. (1989). Shape information from shading: a theory about human perception. *Spatial Vision*, **4** , 165-182.

Pizlo,Z. (2008). *3D Shape: Its Unique Place in Visual Perception*. The MIT Press.

Pizlo,Z., Sawada,T., Li,Y., Kropatsch,W.G. & Steinman,R.M. (2010). New approach to the perception of 3D shape veridicality, complexity, symmetry and volume. *Vision Research*, **50**, 1-11.

Pont,S.C. & Koenderink,J.J. (2007). Matching illumination of solid objects. *Perception and Psychophysics*, **69**, 659-668.

Ramachandran,V.S. (1988). Perception of shape from shading. *Nature*, **331**, 163-166.

Reichel,F.D. & Todd,J.T. (1990). Perceived depth inversion of smoothly curved surfaces due to image orientation. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 653-664.

Rittenhouse,D. (1786). Explanation of an optical deception. *Transactions of the American Philosophical Society*, **2**, 37-42.

Sakai,K., Narushima,K. & Aoki,N. (2006). Facilitation of shape-from-shading perception by random textures. *Journal of the Optical Society of America*, **23**, 1805-1813.

Schofield,A.J. & Georgeson,M. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: separate mechanisms with similar behaviour. *Vision Research*, **39**, 2697-2716.

Schofield,A.J. (2000). What does second-order vision see in an image? *Perception*, **29**, 1071-1086.

Schofield,A.J. & Georgeson,M. (2003). Sensitivity to contrast modulation: the spatial frequency dependence of second-order vision. *Vision Research*, **43**, 243-259.

Schofield,A.J., Hesse,G., Rock,P. & Georgeson,M. (2006). Local luminance amplitude modulates the interpretation of shape-from-shading in textured surfaces. *Vision Research*, **46**, 3462-3482.

Schofield,A.J., Rock,P., Sun,P., and Georgeson,M. The role of second-order vision in discriminating shading versus material changes. Journal of Vision 9. 2009 (VSS Poster).

Schofield,A.J., Rock,P., Sun,P., Jiang, XY., and Georgeson,M (in press). What is second order vision for? Discriminating illumination versus material changes, Journal of Vision.

Seyama,J. & Sato,T. (1998). Shape from shading: estimation of reflectance map. *Vision Research*, **38**, 3805-3815.

Singh,M. & Anderson,B.L. (2002). Toward a perceptual theory of transparency. *Psychological Review*, **109**, 492-519.

Sinha,P. and Adelson,E.H. Recovering reflectance and illumination in a world of painted polyhedra. Fourth International Conference on Computer Vision. *Proceedings ICCV'93*, 156-163. 1993.

Skottun,B.C. (1994). Illusory contours and linear filters. *Experimental Brain Research*, **100**, 360-364.

Song,Y. & Baker,C.L.Jr. (2006). Neural mechanisms mediating responses to abutting gratings: Luminance edges vs. illusory contours. *Visual Neuroscience*, **23**, 181-199.

Stewart,A.J. & Langer,M.S. (1997). Toward accurate recovery of shape from shading under diffuse lighting. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **19**, 1020-1025.

Sun,J. & Perona,P. (1996). Early computation of shape and reflectance in the visual system. *Nature*, **379**, 165-168.

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

Sutter,A., Sperling,G. & Chubb,C. (1995). Measuring the spatial frequency selectivity of second-order texture mechanisms. *Vision Research*, **35**, 915-924.

Tappen,M.F., Freeman,W.T. & Adelson,E.H. (2005). Recovering intrinsic images from a single image. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **27**, 1472.

ter Haar Romeny,B.M. (2003). *Front-end vision and multi-scale image analysis*. Springer.

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**, 583-595.

Todd,J.T., Koenderink,J.J., van Doorn,A.J. & Kappers,A.M.L. (1996). Effect of changing viewing conditions on the perceived structure of smoothly curved surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, **22**, 695-706.

Todd,J.T. (2004). The visual perception of 3D shape. *TRENDS in Cognitive Science*, **8**, 116-121.

Treisman,A.M. & Gelade,G. (1980). A feature integration theory of attention. *Cognitive Psychology*, **12**, 97-136.

Tyler,C.W. (1998). Diffuse illumination as a default assumption for shape-from-shading in the absence of shadows. *Journal of Imaging Science and Technology*, **42**, 319-325.

Vuong,Q.C., Domini,F. & Caudek,C. (2006). Disparity and shading cues cooperate for surface interpolation. *Perception*, **35**, 145-155.

Waltz,D. (1975). Understanding Line Drawings of Scenes with Shadows. In *The Psychology of Computer Vision*, ed. Winston,P.H., pp. 19-91. New York: McGrawHill.

Weiss,Y. Deriving intrinsic images from image sequences. 2001 IEEE International Conference on Computer Vision. *Proceedings ICCV'2001* **2**, 68-75. 2001.

Wilson,H.R., Ferrera,V.P. & Yo,C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, **9**, 79-97.

Wright,M. & Ledgeway,T. (2004). Interaction between luminance gratings and disparity gratings. *Spatial Vision*, **17**, 51-74.

Zhan,C. & Baker,C.L.Jr. (2008). Critical spatial frequencies for illusory contour processing in early visual cortex. *Cerebral Cortex*, **18**, 1029-1041.

Zhou,Y.X. & Baker,C.L.Jr. (1993). A processing stream in mammalian visual cortex neurons for non-Fourier responses. *Science*, **261**, 98-101.

Zhou,Y.X. & Baker,C.L.Jr. (1996). Spatial properties of envelope-responsive cells in area 17 and 18 neurons of the cat. *Journal of Neurophysiology*, **75**, 1038-1050.

Appendix 1: Published Journal Article

What is second-order vision for? Discriminating illumination versus material changes

Andrew J Schofield[1], Paul B Rock[1], Peng Sun[1], Xiaoyue Jiang[1], Mark A Georgeson[2]

[1]School of Psychology,
University of Birmingham, Birmingham, B15 2TT, UK.

[2]School of Life & Health Sciences,
Aston University, Birmingham, B4 7ET, UK

Contact information
Schofield: Email a.j.schofield@bham.ac.uk Web www.vision.bham.ac.uk
Rock: Email pr@star.sr.bham.ac.uk Web www.vision.bham.ac.uk
Sun: Email PXS315@bham.ac.uk Web www.vision.bham.ac.uk
Jiang: Email x.y.jiang@bham.ac.uk Web www.vision.bham.ac.uk
Georgeson: Email m.a.georgeson@aston.ac.uk Web
http://www1.aston.ac.uk/lhs/staff/az-index/georgema/

Abstract

The human visual system is sensitive to second-order modulations of the local contrast (CM) or amplitude (AM) of a carrier signal. Second-order cues are detected independently of first-order luminance signals; however it is not clear why vision should benefit from second-order sensitivity. Analysis of the first- and second-order content of natural images suggests that these cues tend to occur together but their phase relationship varies. We have shown that in-phase combinations of LM and AM are perceived as a shaded corrugated surface whereas the anti-phase combination can be seen as corrugated when presented alone or as a flat, material change when presented in a plaid containing the in-phase cue. We now extend these findings using new stimulus types and a novel haptic matching task. We also introduce a computational model based on initially separate first- and second-order channels that are combined within orientation and subsequently across orientation to produce a shading signal. Contrast gain control allows the LM+AM cue to suppress responses to the LM-AM when presented in a plaid. Thus the model sees LM-AM as flat in these

circumstances. We conclude that second-order vision plays a key role in disambiguating the origin of luminance changes within an image.

Introduction

The human visual system is sensitive to variations of second-order cues such as modulations of the local contrast (CM) of textured stimuli. This is true for both moving (see Baker, 1999 for an early review) and static (Badcock, Clifford & Khuu, 2005; Dakin & Mareschal, 2000; Georgeson & Schofield, 2002; Graham & Sutter, 2000; Henning, Hertz and Broadbent, 1975; Larsson, Landy & Heeger (2006);Nachmias, 1989; Nachmias & Rogowitz, 1983; Schofield & Georgeson, 1999, 2003; Sutter, Sperling, & Chubb, 1995) stimuli, although here we concentrate on static cues. There is strong psychophysical evidence to suggest that static CM is detected separately from first-order luminance modulations (LM). For example, there is no sub-threshold facilitation between the cues (Schofield & Georgeson, 1999), they can be distinguished at detection threshold (Georgeson & Schofield, 2002), lateral interactions are different for the two cues (Ellemberg, Allen, & Hess, 2004), their channel structure is different (Ellemberg, Allen & Hess, 2006), noise masking is doubly-dissociated (Allard & Faubert, 2007), they make separate contributions to global form detection (Badcock, et al., 2005) and different contributions to contour linking processes (Hess, Ledgeway & Dakin, 2000). Finally although most retinotopic visual areas respond to both LM and CM there is preferential fMRI adaptation for CM in the higher areas (specifically VO1, LO1 and V3a; Larsson, et al., 2006).

It is also clear, however, that CM and LM are integrated or partially integrated in some cases. For example, contrast modulations of a high-contrast grating carriers mask LM signals (Henning et al., 1975; Nachmias & Rogowitz, 1983) but modulations of low contrast noise carriers do not (Schofield & Georgeson, 1999). LM masks the detection of CM in noise carriers but not vice versa (Ellemberg, Allen, & Hess, 2006; Schofield & Georgeson, 1999), and similar asymmetric interference has been found for global form detection (Badcock, et al., 2005). The orientation of first order stimuli affects the perceived orientation of second-order stimuli (Morgan, Mason & Baldassi, 2000). The signal types combine at low contrasts to improve perceptual accuracy (Smith & Scott-Samuel, 2001). Further, tilt and contrast reduction after-effects transfer between LM and CM (Georgeson & Schofield, 2002),

as does the tilt illusion (Smith, Clifford & Wenderoth, 2001). Finally, we have previously shown that LM and CM interact in the perception of shape-from-shading (Schofield, Hesse, Rock & Georgeson, 2006).

The physiological evidence for independent first- and second-order mechanisms is less clear-cut and comes mainly from studies using moving stimuli. Mareschal & Baker (1998) found cells in cat area 18 that are responsive to second-order stimuli, but these also responded to first-order stimuli: suggesting early integration. However, typically, preferred frequencies for the two cues were slightly different. They concluded that such cells were likely to take their input from independent first- and second-order sub-mechanisms (see also Zhou & Baker, 1996, and Song & Baker, 2006). Further, in physiology, it is common to search for cells using first-order stimuli. Any cell that is then found to be sensitive to second-order cues will, by definition, also be sensitive to first-order stimuli. Finally, Second-order signals may be extracted in another visual area; V3a has been implicated in second-order processing for both static (Larsson et al, 2006) and moving stimuli (Ashida, Lingnau, Wall & Smith, 2007). Perhaps second-order signals are extracted in V3a and fed back to V1/V2.

Despite the above evidence for separate but interacting first- and second-order mechanisms, psychophysically human vision is an order of magnitude less sensitive to CM than LM (Schofield & Georgeson, 1999) and similar, if less extreme, results have been found for motion in cat area 17/18 (Mareschal and Baker, 1998; Zhou and Baker, 1996; Ledgeway, Zhan, Johnson, Song & Baker, 2005; Hutchinson, Baker and Ledgeway, 2007) and monkey MT (Albright, 1992). This suggests that CM is something of a secondary cue, and it is not yet clear why the independent detection of static second-order cues is beneficial to human vision. We now address this question.

Human vision presumably obtains some advantage from processing first- and second-order cues independently and indeed from detecting second-order cues at all. Johnson and Baker (2004) measured the relationship between patterns of LM and CM in natural scenes and found the two cues to be highly correlated on an unsigned magnitude metric. This implies that CM variations tend to occur alongside LM. However, Schofield (2000) performed a similar analysis using a signed metric and

found that whereas the two cues may be strongly correlated within a single image the sign of the correlation varies between images, such that they are uncorrelated over an ensemble of images. Taken together these results suggest that CM is an informative cue in natural images but that information may be conveyed by its relationship with LM rather than its mere presence.

In this paper (as previously, Schofield, et al., 2006) we prefer to use the term amplitude modulation (AM) over CM because although they are mathematically equivalent when presented alone, when combined with LM they can be interpreted as distinct image properties with AM being the better description for our purposes. Schofield et al. (2006) showed that LM and AM are yoked whenever an albedo-textured surface is shaded or in shadow (see Figure 1 for a natural example of such shading and Schofield et al., 2006, for a full account of the yoking between these cues). Albedo textures represent locally smooth surfaces whose local reflectance changes creating a visual texture. So LM+AM represents a strong cue to shading / shadows when certain textured surfaces are present.

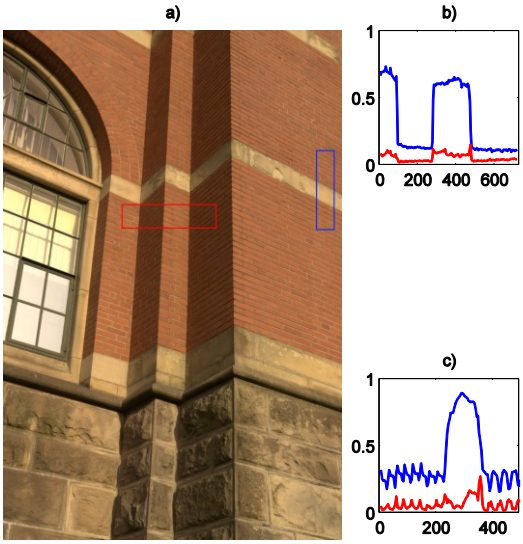


Figure 1. a) a natural image showing part of a building on the University of Birmingham campus. The building 'steps' out twice working left to right and the orientation of the faces produces shading but not cast shadows. The brick sections are, approximately, a reflectance texture of the type described in the text. The image also shows gross reflectance changes, most notably the strips of sandstone among the red brick sections. The red and blue boxes show approximate sampling regions for the traces of panels b) and c) respectively. The red section of a) was extracted and rotated so that the shading edges were vertical. The blue section of a) was extracted and rotated

so that the sandstone edges were vertical. Sample sections were also converted to greyscale. b) Mean (blue line) and standard deviation (red line) of the gray level values in each column of the rotated red section. Mean pixel values are a measure of luminance whereas their standard deviation measures luminance amplitude or range. Transitions of high to low luminance (LM) are clearly mimicked by changes in luminance amplitude (AM) and the two cues are positively correlated. c) mean and standard deviation for the columns in the rotated blue section of a) here the transition to high luminance in the sandstone section is not mirrored by a change in standard deviation.

People see sinusoidal shading patterns as sinusoidally undulating surfaces (Kingdom , 2003; Pentland, 1988; Schofield et al., 2006; Schofield, Rock & Georgeson, submitted) even though such surfaces only give rise to sinusoidal shading in restricted circumstances. We presume that the luminance component of the LM+AM signal is coded as a shading pattern and then interpreted as a corrugated surface via shape-from-shading (Christou & Koenderink, 1997; Erens, Kappers & Koenderink, 1993; Horn & Brooks, 1989; Kleffner & Ramachandran, 1992; Langer & Bülthoff, 2000; Ramachandran, 1988; Todd & Mingolla, 1983; Tyler, 1998) whereby luminance level is equated with surface gradient such that the parts of the surface that are most luminous are seen as being oriented towards the illuminant. When the direction of the illuminant is unknown humans assume a lighting-from-above prior (Adams, Graf & Ernst; 2004; Brewster, 1826; Mamassian and Groucher (2001); Ramachandran, 1988; Rittenhouse, 1786; Sun & Perona, 1998). Our earlier results (Schofield et al, 2006) with LM+AM sinusoids are consistent with this interpretation, except that we now propose an illumination prior that is a mixture of diffuse and point source lighting (Schofield, Rock. Georgeson & Yates, 2007; Schofield, Rock, & Georgeson, submitted).

The filter-rectify-filter model used by Schofield (2000) to extract second-order cues from natural images was sensitive to AM, and it seems likely that natural images containing positively correlated first- and second-order cues are dominated by shadows and shading. But what of those images that contain negatively correlated cues?

Transparent overlays also give rise to second-order cues in natural stimuli (Fleet and Langley, 1994). The specific case of a semi-opaque, light (or milky) transparency is

pertinent here. Those parts of a textured surface that are obscured by such a transparency suffer an increase in mean luminance (e.g. if the base colour of the overlay is white its luminance will be higher than the mean luminance of the texture) but a decrease in local amplitude (the difference between the light and dark parts of the texture will fall due to the blurring caused by the semi-transparent medium). This configuration exhibits negatively correlated LM and AM (LM-AM: Note however that if the transparency is dark LM and AM will again be positively correlated). The notion that LM-AM is a possible cue for transparency is supported by the qualitative description of such stimuli given by Georgeson & Schofield (2002; they used the term LM-CM). If LM-AM is seen as a cue to transparency then the overall perception is likely to be of flat surfaces although the semi-transparent regions may be seen as being in front of the main surface. LM-AM might also be interpreted as a material change, as there is no restriction on the relationship between LM and AM when two surfaces comprising materials with different textures are abutted (see Figure 1).

The idea that LM-AM may be interpreted as either a material change or as an overlaid transparency was given empirical support by our previous finding that this cue is seen as flat when presented in a plaid with LM+AM (Schofield et al., 2006). LM+AM is, by contrast, seen as a shading cue and is therefore perceived as corrugated in depth via shape-from-shading. However, when presented alone LM-AM is also seen as corrugated albeit less strongly (less reliably) than LM+AM. Why might LM-AM be seen as flat in some cases and corrugated in others? There are cases where undulating surfaces *can* produce negatively correlated LM and AM. An example of such a surface would be a physically textured (rough) surface under certain illumination conditions (see Figure 2 of Schofield et al., 2006). Thus we previously concluded that whereas LM+AM is a strong cue to shading, LM-AM is rather ambiguous when seen alone. However, when intimately associated with LM+AM as in the case of a plaid stimulus where the two cues are necessarily presented with the same texture carrier the interpretation of LM+AM as being due to shading seems to force the interpretation of LM-AM as being due to some sort of material change (Schofield et al., 2006).

The notion that the relationship between LM and AM provides a key for separating shading and shadows from material changes has important implications for human

vision and applications in machine vision. In principle, a given image can arise from an infinite number of scene and lighting combinations. Human vision may make considerable use of stored knowledge about the world in a top-down fashion to correctly interpret visual scenes. However, natural images may also contain cues that can be used to disambiguate the incoming luminance variations via bottom-up processes. Specifically, luminance variations are ambiguous; they may result from changes in illumination (shadows and shading) or changes in surface reflectance. If human vision were only sensitive to luminance its ability to distinguish these possibilities on the basis of low-level cues would be greatly restricted. Barrow and Tannenbaum (1978) showed how some progress can be made towards the separation of illumination and reflectance in a 'luminance only' system, but they also highlighted the potential benefits of being sensitive to other cues and the importance of understanding how cues relate to one another in real world stimuli. Others have shown that hue can be used to separate illumination from reflectance changes (see for example Kingdom 2003; Olmos and Kingdom, 2004; Tappen, Freeman and Adelson, 2005). Here we consider the use of AM as a cue to separate the luminance changes due to variations in surface reflectance from those due to variations in illumination or shading, and we provide a simple bottom up model - based on both the filter-rectify-filter model of second order vision (Wilson, Ferrera, & Yo, 1992) and the processing scheme for envelope neurons proposed by Zhou & Baker (1996) - that can account for our psychophysical results.

In our earlier study (Schofield et al., 2006) we asked observers to make relative depth judgements about pairs of probe points from which we derived normalised gradients before reconstructing perceived surface profiles: we did not measure perceived depth directly. Thus we were unable to express perceived depth in absolute terms, unable to measure differences in depth between stimuli with very different signal strengths and unable distinguish between low-relief and unreliable depth percepts. Further, participants in our earlier experiments reported that the depth probe task felt artificial because the probe markers did not appear to be attached to the surface. We avoided these problems here by asking observers to match the properties of a haptic surface to the perceived corrugations in a co-located visual stimulus. This task felt natural to participants and gave direct, and absolute estimates of perceived depth amplitude.

We report three experiments. In the first experiment we fixed the position of the haptic cue based on the results of a pilot study and asked observers to set the amplitude of the haptic undulations to match the perceived surface undulations. Our previous study (Schofield et al., 2006) only measured depth profiles at two levels on LM (for fixed AM) and found little difference between these conditions. We now measure perceived depth amplitude (PDA) as a function of signal strength, varying LM and AM together (Experiment 1) yielding a better understanding of how LM and AM interact at different signal strengths. In Experiments 2 and 3, we fixed the contrast of the LM cue and measured PDA as a function of AM signal strength in both plaid (Experiment 2) and single component (Experiment 3) stimuli, exploring the role of AM in more detail. We also present a biologically plausible model providing a good fit to the data suggesting that human performance in this task can be explained by a bottom up system that first detects and then integrates first- and second-order information.

General methods

We introduce a new method for assessing shape-from-shading. Observers viewed sinusoidal visual stimuli while stroking a sinusoidally corrugated haptic stimulus and were asked to set the depth amplitude of the haptic stimulus to match the visually perceived surface. Visual stimuli comprised various combinations of LM and AM as described below. After a short training session this method felt very natural to the observers. However, the method relies on the assumption that observers would perceive sinusoidal luminance patterns as sinusoidal corrugations with the same spatial frequency. This assumption is supported by our previous depth mapping experiments (Schofield et al., 2006), the findings of Pentland (1988), and results from a gauge figure experiment reported elsewhere (Schofield et al., submitted). There is also a danger that the haptic stimulus might alter the visual experience, perhaps acting as a training stimulus (Adams, Graf and Ernst, 2004). We think that this is unlikely partly because results from the haptic match task are similar to those obtained with other methods (Schofield et al., 2006 and Schofield et al., submitted). Further, while we do not doubt that haptic stimuli can be used to alter visual perception we see no reason why such cross-modal influence should be mandatory. Here we made it clear

that observers should treat the visual stimulus as the fixed reference and set the haptic stimulus to match it. Other than being a sinusoid of the same frequency as the visual cue, there was no systematic manipulation of the haptic stimulus to entrain the visual percept.

Visual stimuli.

We follow Pentland (1988), and Kingdom (2003) in using sinusoidal shading patterns with no occluding boundaries. Stimuli were not rendered surfaces. Studies of shape perception more typically use images of rendered (or real) objects, irregular shapes, or sections thereof. We used grating stimuli and random noise textures for the following reasons; 1) Shading is known to be a relatively weak or secondary cue to shape and can be dominated by other cues including object outlines. Thus the outlines of rendered objects or blobs can influence both the perceived surface shape (see Knill, 1992) and the strength of the depth percept. 2) We need to simulate textured surfaces in our stimuli, but if these had been rendered then geometric distortions in the texture would have been an additional cue to shape. Our noise textures were isotropic, providing no cue to shape. 3) With gratings it is very easy to control the phase relationship between LM and AM and the amount of AM. 4) The use of gratings made it easy for us to cue which component was to be matched to the haptic probe.

Visual stimuli were formed from isotropic, binary visual noise with a Michelson (and r.m.s.) contrast of 0.1, onto which we imposed sinusoidal modulations of luminance and amplitude. Noise elements comprised 2x2 screen pixels and subtended 0.06 degrees of arc at the 57cm viewing distance. We imposed five types of sinusoidal modulation onto these noise textures: (a) LM-only (Figure 2a) comprising luminance modulations added to the noise pattern with no variation in AM, (b) AM-only (Figure 2b) comprising amplitude modulated noise, (c) LM+AM alone (Figure 2c), (d) LM-AM alone (Figure 2d), and (e) plaid stimuli comprising LM+AM on one oblique and LM-AM on the other (Figure 2e). Except when AM modulation depth was zero we did not test plaids composed of the same cues (ie both LM+AM) on both diagonals. In the case of plaids either the LM+AM or LM-AM component could be designated as the test cue making a total of 6 test conditions in all (but not all conditions were tested in every experiment). Test cues were presented in one of two orientations; left oblique or right oblique (±45°). The wavelength of the modulations was 25mm (spatial

frequency = 0.4 c/deg). The contrast of the LM signals and the modulation depth of the AM signals varied between experiments and conditions. Stimuli were presented in a modified ReachIN^{TM} haptic workstation (Reachin AB, Sweden) depicted in Figure 3. Visual stimuli were presented on a 17" Sony Trinitron CPD G200 CRT monitor (Sony Inc, Japan) mounted at an angle of 45° above a horizontal half-silvered mirror. Observers looked into the mirror at a downward angle and thus perceived the visual stimulus to be beneath the mirror and approximately perpendicular to their line of sight. A hood prevented the observer from viewing the monitor directly. Observers were asked not to tilt their heads to one side but, except for the need to sit close to the workstation and the limitations imposed by the hood, viewing position was not physically constrained. Stimuli were viewed in the dark such that observers could not see their own hand beneath the mirror. Viewing was binocular and so the visual stimulus provided stereoscopic cues to flatness. However, a robust percept of shape-from-shading can be derived from such stimuli (Schofield at al., 2006).

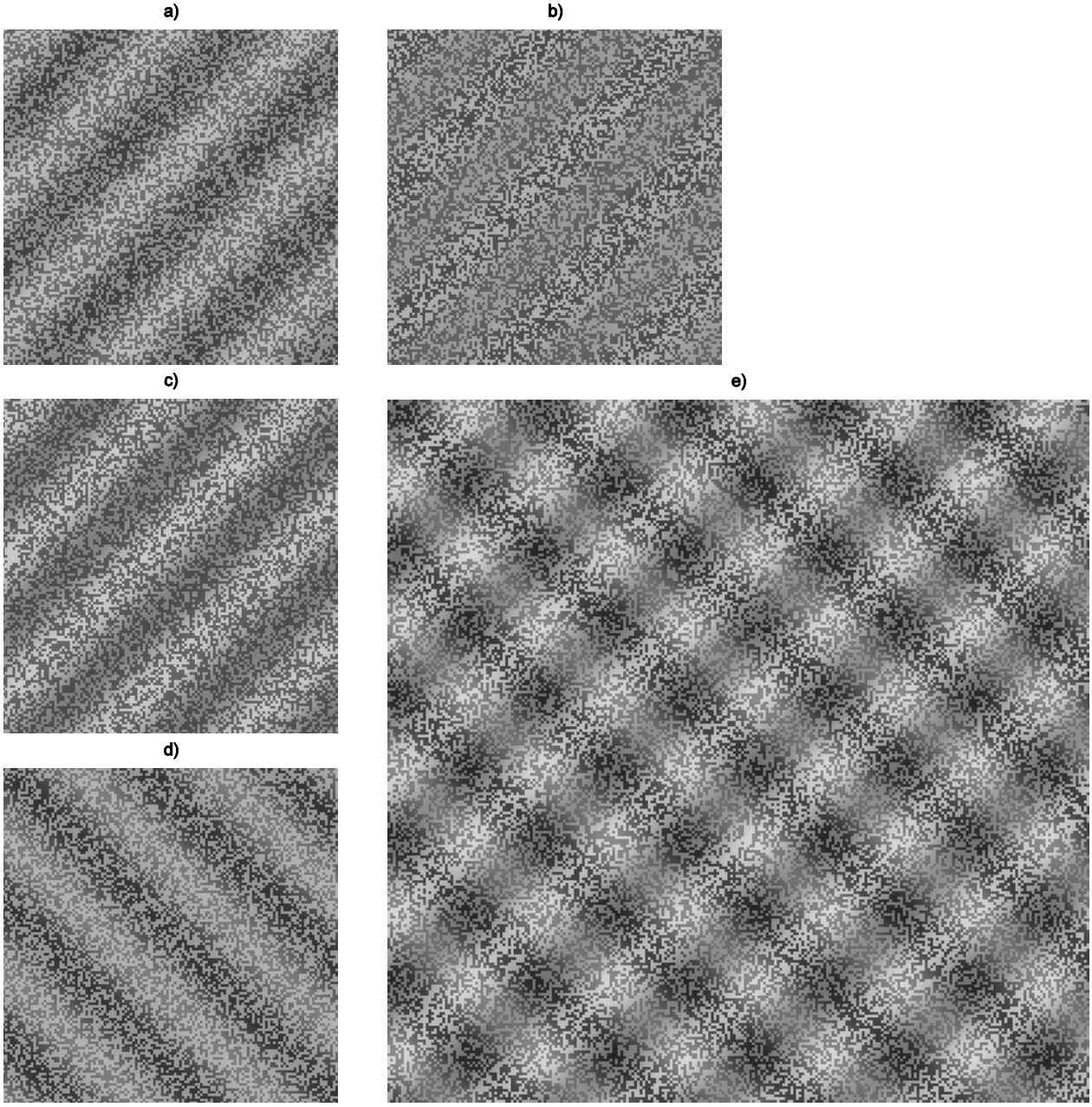


Figure 2. Extracts from example stimuli: a) LM-only, formed by arithmetically adding a luminance grating to spatial, binary noise; b) AM-only, formed by modulating the amplitude (standard deviation) of the noise; c) LM+AM only, formed by combining the cues of a) and b) in-phase, equivalent to multiplicative shading; d) LM-AM only, formed by combining the cues of a) and b) in anti-phase; e) LM+AM and LM-AM in a plaid configuration; here LM+AM is on the right oblique. Note noise contrast has been increased to from 0.1 to 0.3 to aid presentation.

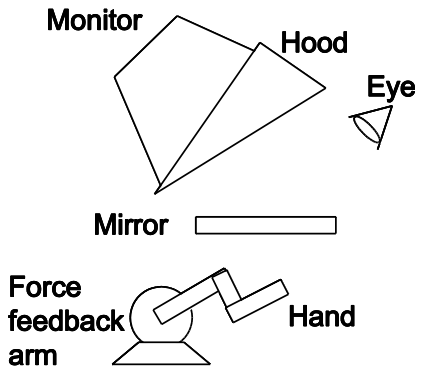


Figure 3. Sketch of the ReachIN™ workstation with additional hood; support structure not shown.

Stimuli were calibrated against the monitor's gamma characteristic using look up tables in a BITS++ attenuation device (CRS Ltd, UK) which also served to enhance the available grey level resolution to the equivalent of 14 bits. Values in the look up tables were determined by fitting a four-parameter monitor model to luminance readings recorded with a CRS ColourCal photometer. Problems in presenting AM stimuli associated with the adjacent pixel non-linearity (Klein, Hu, & Carney, 1996) were avoided by using a high bandwidth monitor, and noise samples with relatively low contrast, but relatively large element size. However, the noise elements were unlikely to be large enough to produce a noticeable clumping artefact (Smith and Ledgeway, 1997; see Schofield & Georgeson, 1999, for a full discussion of these issues).

Haptic stimuli.

Haptic stimuli were presented via a Phantom-Desktop™ (SensAble Technologies Inc, MA, USA) force feedback device located beneath the mirror and consisted of a virtual surface collocated with the visual stimulus. Haptic surfaces had sinusoidal undulations in the direction of the visual test cue. The spatial frequency of the undulations matched that of the visual stimuli. Observers held the Phantom's stylus like a pen with their dominant hand and stroked the surface. The Phantom provided physical resistance whenever the observer tried to move the stylus tip through the virtual surface. Three markers were added to the visual stimulus: one at the centre and two at opposite corners of the stimulus, so that the alignment of the three markers indicated the direction in which observers should stroke the haptic surface in order to feel the undulations. We verified that distances specified in the haptic stimuli were

faithfully reproduced by the Phantom. Visual and haptic stimuli were generated on the same PC.

Visual cursor.

We ensured that the location, orientation and spatial frequency of the haptic stimuli matched the visual stimuli well. However, we also conducted a pilot experiment to verify that observers could reliably match the position of the haptic undulations to visual features. In this experiment the visual stimuli consisted of a horizontal luminance grating and observers were asked to adjust the position of the peaks in the haptic stimuli to match the position of the luminance peaks. In the absence of any visual feedback as to the location of the stylus tip observers were unable to match the positions on the visual and haptic stimuli with any reliability (standard deviation of match positions = 0.288 wavelengths). However, reliable position matches were possible on the introduction of a visual cursor that tracked the tip of the stylus (standard deviation of match positions = 0.041 wavelengths). A cursor was therefore included in all the experiments. We conclude that co-registration of the haptic and visual stimuli is not sufficient to allow reliable position matching in the absence of visual feedback. Further, although we have not tested this directly, we suspect that precise co-registration is not necessary if feedback is provided. We note, for example, that computer users can reliably place a pointer at a specified screen location despite a gross mismatch between the physical positions of the pointer and 'mouse'.

Position of haptic stimulus.

Prior to the main experiments we asked observers to adjust the position of a haptic stimulus to match that of the perceived corrugations in the visual stimuli. These settings were then used to determine the precise relative position of the visual and haptic stimuli in the main experiments such that haptic peaks were always aligned with perceived surface peaks. Typically perceived surface peaks (and hence haptic peaks) are offset from the luminance peaks (see Schofield et al., 2006). Details of how these measurements were performed can be found in experiment 1 of Schofield et al. (submitted). We measured offsets (the difference between the position of the luminance peaks and the haptic peaks) for LM+AM, LM-AM, LM-only & AM-only in the single oblique condition and LM+AM when presented as part of a plaid stimulus. AM-only offsets were measured relative to peaks in the amplitude signal.

We then applied the appropriate offsets between our visual and haptic stimuli on a per condition and observer basis. However, we could not measure offsets for LM-AM stimuli in the plaid configuration as observers saw this cue as flat and therefore could not identify any surface peaks against which to make a match. Instead we used the LM+AM offsets when testing LM-AM in a plaid.

Main adjustment task.

The text experiments reported below observers adjusted the amplitude of the haptic surface up or down by pressing one of two keys on a numeric keypad. A third key toggled the step size for adjustments between 2 and 0.5 mm (half-height amplitude). Observers heard a long tone for each 2mm adjustment and a short tone for each 0.5mm adjustment. Observers could not drive the amplitude of the haptic surface below zero and received an auditory warning of any attempt to do so. Estimates of PDA were calculated as the median of at least 5 measurements.

Observers.

Five observers took part in the experiments. With the exception of author PR, observers were naïve to the purposes of the experiment and were paid for their time. Author PS was a naïve observer at the time of the study. Author AJS contributed some additional data to Experiment 2. All observers had normal or corrected-to-normal vision and no physical disability or injury. Observers held the stylus in their dominant hand: JG is left handed; the remaining observers are right handed.

Experiment 1: Perceived depth amplitude versus overall signal strength

In this experiment we considered the effect of overall signal strength on the PDA of visual stimuli. We also varied the relative phase of the LM and AM cues at the test orientation, and we compared two components (plaids) with single component stimuli (gratings). The LM contrast and AM modulation depth were equal in any given stimulus, consistent with multiplicative shading for in-phase pairings.

Method

Signal strength, governing both LM component contrast and AM component modulation depth, was varied in multiples (0.1, 0.4, 0.8, 1.6, 3.2 & 4.0) of each observer's AM detection threshold as measured in separate sessions using a staircase method (Levitt, 1971) and a two interval forced choice design. In this pilot experiment, stimuli consisted of AM gratings presented alone. Note that our AM gratings are identical to the CM gratings often used to study second-order vision. The mean AM threshold across observers was 0.086, and this is consistent with the literature on second-order vision (Schofield & Georgeson, 1999). Stimuli consisted of plaids comprising LM+AM on one diagonal and LM-AM on the other (Figure 2e), LM+AM presented alone (Figure 2c), or LM-AM presented alone (Figure 2d). Because they contain two orientation components, plaids had greater overall contrast and modulation depth than single component stimuli. Many of the stimuli in this experiment contained sub-threshold levels of AM, but their LM components were likely to be supra-threshold because thresholds for LM in visual noise are about an order of magnitude lower than AM (CM) thresholds (Schofield & Georgeson, 1999).

Results and discussion

Figure 4 shows the results of Experiment 1 averaged over the five observers. Mean PDA was low for weak stimuli regardless of their composition and remained low for LM-AM at all signal levels when this cue was part of a plaid (squares in Figure 4b). However, when LM-AM was presented alone (squares in Figure 4a) PDA increased with signal strength. PDA also increased with signal strength for LM+AM whether presented alone (circles in Figure 4a) or in a plaid (circles in Figure 4b). Although the variances were high, we note that PDA rises to a level significantly above zero for all cues except LM-AM presented in a plaid (error bars on Figure 4 represent 95% confidence intervals). PDAs for strong LM+AM gratings tend to be greater than those for LM+AM presented as part of a plaid despite the fact that overall luminance contrast was higher for the latter stimulus. This trend can also be seen in weaker stimuli where components of a plaid produced lower PDAs than single grating stimuli. For single obliques, strong LM+AM gratings produced somewhat greater PDAs than LM-AM gratings, but only when AM was above threshold. Perceived depth for LM+AM was also greater than for LM-AM in plaid stimuli and this seemed to hold down to signal levels where AM was below threshold (between 0.4 and 1 x AM-threshold). Lines in Figure 4 show predictions of the model described later.

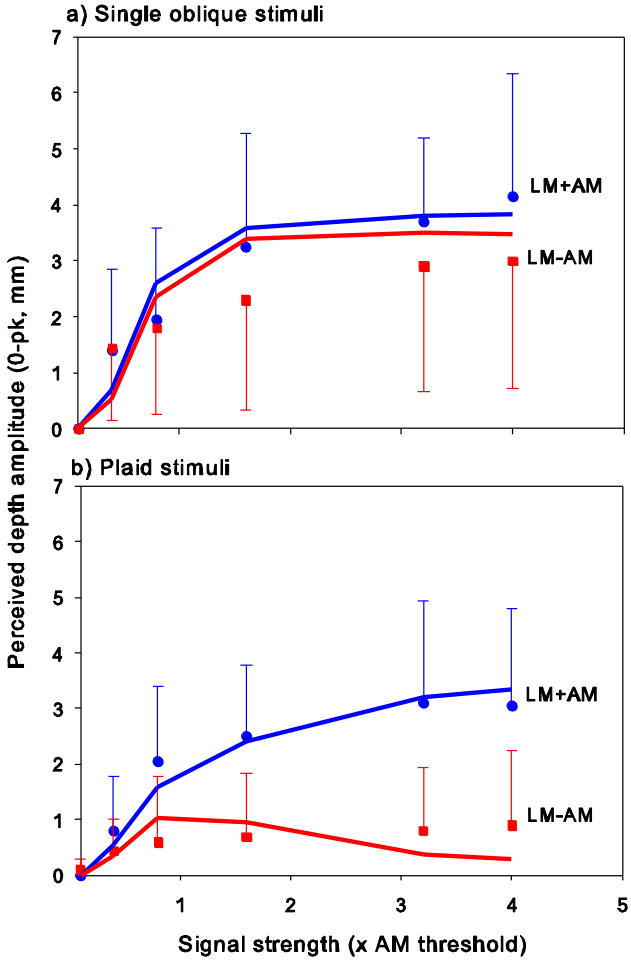


Figure 4. Experiment 1. Perceived depth amplitude as a function of overall signal strength: (a) single oblique stimuli, (b) plaid stimuli. Blue circles show the perceived amplitude of LM+AM mixes; Red squares LM-AM mixes. X-axis shows signal strength as a multiple of AM threshold. Error bars represent 95% confidence intervals and are drawn single-sided to aid interpretation. Lines show predictions of the 'shading-channel' model; blue and red for LM+AM and LM-AM respectively (see description of model for details).

Noting that the plots of Figure 4 are approximately linear against log signal strength, we estimated (with linear regression) the slope of the relationship between log signal strength and PDA separately for each participant and each stimulus type. Figure 5 plots the mean slope for each stimulus type and their associated 95% confidence intervals. Slopes for LM-AM were not significantly different from zero regardless of the configuration used (one-sample, one-way t-test: LM-AM only, t=2.55, df=4, p>0.05; LM-AM in plaid, t=1.16, df=4, p>0.05). LM+AM stimuli produced significant slopes (LM+AM only, t=4.26, df=4, p<0.05; LM+AM in plaid, t=3.74, df=4, p<0.05). A repeated measures ANOVA (with Greenhouse-Geisser correction)

showed that there were significant differences between the mean slopes across the four conditions (F=8.57, df=1.6,6.38, p<0.05). Bonferroni corrected post-hoc paired comparisons showed that slopes for LM-AM in a plaid were significantly lower than those for the LM+AM conditions (LM-AM in a plaid vs LM+AM in plaid, t=6.2, df=4, p<0.05; LM-AM in plaid vs LM+AM only, t=5.32, df=4, p<0.05). The difference in slopes between LM-AM in a plaid and this cue presented alone was significant prior to Bonferroni correction but not after (t=3.1). None of the other pairings were significantly different suggesting that LM-AM presented alone produces behaviour similar to that of LM+AM.

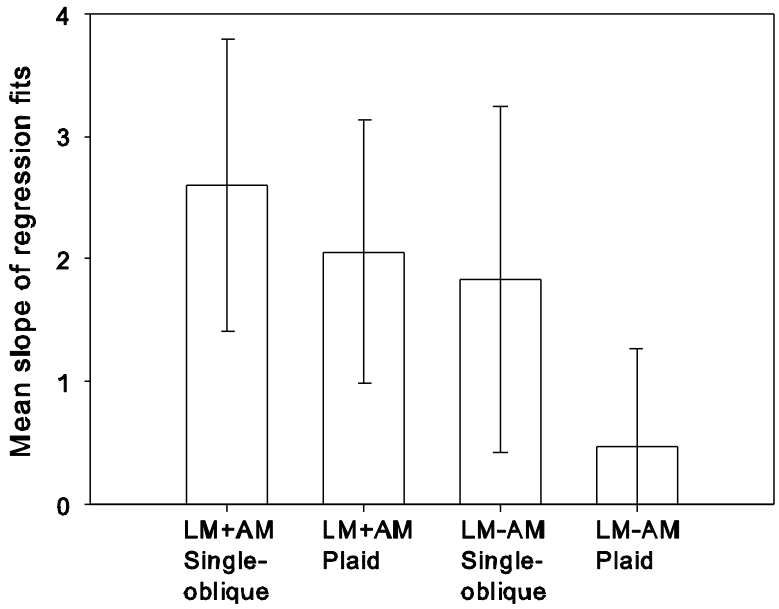


Figure 5. Mean slopes for regression fits to individual data from Experiment 1 for each of the four test conditions. Error bars represent 95% confidence intervals.

Taken together, these results show that LM-AM is seen as a shape-from-shading cue when presented on its own. PDAs for this cue are about the same as those for LM+AM in a plaid but below those for LM+AM presented alone. When LM-AM is presented as part of a plaid, however, it is seen as quite flat. Inspecting individual data revealed that most observers saw this condition as almost completely flat even at high signal strength and that the slope observed in Figure 4 is largely due to one observer who saw this stimulus as conveying some depth. By contrast LM-AM alone was seen as quite corrugated by all but one observer and the two LM+AM conditions were seen as corrugated by all observers. PDAs naturally converge toward zero as signal

strength is reduced. PDAs for single components converge at about the point where the AM signal falls below threshold. PDAs for the two members of a plaid converge at a point below the measured AM detection threshold; this could be due to probability summation which may serve to increase the visibility of AM in plaid stimuli above that of single orientation components. It is clear that LM is the dominant cue for depth perception in shaded textures but that its relationship with AM and the overall configuration of the stimulus is also important. We now investigate the specific role of AM in more detail.

Experiments 2 and 3. Effect of AM modulation depth on perceived depth amplitude.

In these experiments we varied AM strength while keeping LM contrast constant. We thus assessed the ability of AM to influence perceived depth.

Method.

Visual stimuli were diagonally oriented gratings and plaids with a fixed LM contrast of 0.2 and several AM modulation depths (0, 0.1, 0.2, 0.4). Again we varied the phase relationship between LM and AM. In Experiment 2 we tested plaid stimuli only. Experiment 3 tested single component stimuli including AM-only gratings (see Figure 2b). When we devised Experiment 2 we considered the LM+AM and LM-AM components to be distinctly different stimulus types. We therefore did not test the case where the AM signal was zero (i.e. an LM-only vs LM-only plaid). We later realised that these cues form a continuum running from strong negative AM to strong positive AM, with LM-only (AM modulation depth = 0) representing the midpoint on this continuum. We thus added the AM=0 case to the test battery for Experiment 3 and tested an additional observer in Experiment 2 including the AM=0 case.

Results.

Figure 6 shows PDA as a function of AM modulation depth. Blue squares show the results for plaid stimuli (Experiment 2); Red circles and green triangles the single component results (Experiment 3). There was no effect of test orientation (left or right oblique) so we averaged across this condition.

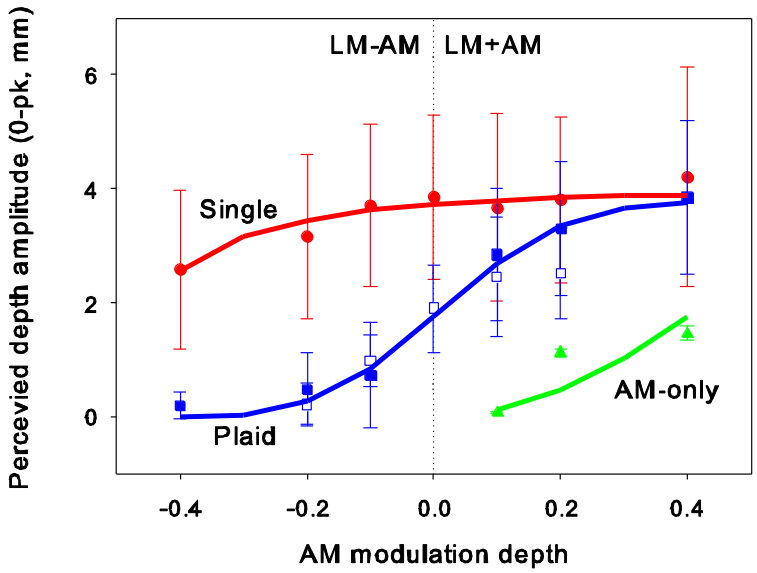


Figure 6. Experiments 2 & 3. Perceived depth amplitude as a function of AM modulation depth and sign. X-axis shows AM modulation depth; negative values mean that the AM cue was in anti-phase with the LM cue (LM-AM). Green triangles, AM-alone. Red circles, single oblique LM and AM signals. Blue squares, LM and AM presented as a plaid. Note that when AM was in anti-phase with LM on the test oblique (negative values) the non-test oblique had an in-phase mix with an equally strong AM cue (and vice versa). Open squares, results for observer AJS for plaid stimuli including the case where AM modulation depth was zero – ie. LM-only on both obliques. Lines represent model fits for the 'shading-channel' model. Except for the open squares, data points are the means of 5 observers and error bars represent 95% confidence intervals. For AJS (open blue squares) error bar represents the standard deviation of individual depth estimates.

Experiment 2: Plaids. For plaid stimuli PDA increased with signed modulation depth such that stimuli were seen as increasingly flat for negative modulation depths (LM-AM) and increasingly corrugated for positive modulation depths (LM+AM). There was a pronounced increase in PDA around AM=0. A repeated measures ANOVA (with Greenhouse-Geisser correction) showed that the overall change in PDA was significant (F=42.468, df=1.493,7.464, p<0.01) and Bonferroni corrected post-hoc paired-samples t-tests showed that antiphase stimuli (LM-AM) produced significantly lower PDAs than in-phase stimuli (LM+AM). Results from the one observer (AJS) tested with AM=0 (open square symbols in Figure 6) suggest that PDAs for LM-only plaids fall nicely on the continuum from LM-AM to LM+AM.

Experiment 3: Single components. There was much less variation in PDA with AM modulation depth in the single component stimuli. Here we found only a gradual increase in PDA with AM modulation depth and hardly any increase at all among LM+AM stimuli. The overall trend was not significant (Greenhouse-Geisser corrected ANOVA, F=4.013, df=1.583,7.916, p=0.069). There were no significant differences between any of the levels tested for the single component stimuli (based on Bonferroni corrected paired t-tests). Paired sample t-tests between AM-only stimuli (triangles) and single component mixed stimuli (filled circles) with equivalent levels of AM suggest that the AM-only simuli were seen as significantly flatter than LM/AM mixes regardless of the phase relationship in the mix (based on paired samples t-tests corrected using Horn's multistage Bonferroni method). Similarly PDAs for LM+AM in a plaid were significantly greater than their AM-only counterparts. In contrast, PDAs for LM-AM stimuli in a plaid were not significantly greater than those for AM-only. Finally we note that LM-AM stimuli in a plaid are seen as significantly less corrugated than the equivalent single component stimuli but that the differences between LM+AM in plaid and single component configurations are not significant.

Discussion.

Taken together the results of Experiments 2 and 3 show that LM-AM was seen as flat when shown in a plaid with LM+AM but was seen as corrugated otherwise. PDAs for LM-AM and LM+AM stimuli tend to be similar at low AM modulation depths. This result is to be expected because these cues become identical as AM modulation depth approaches zero. However, while PDAs for the LM+AM and LM-AM gratings (at a single orientation) were almost identical for AM modulation depth in the range -0.1 to +0.1, those for the plaid stimuli varied significantly over this range.

We note that LM+AM stimuli also appear a little less corrugated in a plaid than they do as single components and although these differences are not significant some discussion is merited. We note particularly that plaid stimuli with little or no AM signal have a doubly corrugated or 'egg box' appearance. The PDA of such stimuli in a given direction is likely to vary with position along the orthogonal axis and this may reduce the average PDA. Single component stimuli appear as single corrugations whose PDA does not vary with position in the direction orthogonal to the

modulations. It is possible that the 'egg-box' effect accounts for the observed difference between plaid and single component stimuli in the LM+AM case. However, there is an alternative explanation based on mutual suppression between obliques and we discuss this next.

Model

We constructed a model to explain our data. The purpose of the model is to demonstrate that the observed effects can be predicted by bottom up mechanisms involving biologically plausible second-order processes. The model (shown in Figure 7) is intended to represent one spatial frequency tuned 'shading channel' within a multi-channel scheme. It is based on the processing scheme for envelope sensitive neurons proposed by Zhou & Baker (1996) and the filter-rectify-filter (FRF) model of second-order vision (Wilson, Ferrara and Yo, 1992), and has similarities with the three stage model proposed by Henning et al. (1975). The first-stage comprises a bank of linear filters tuned to multiple spatial frequencies and orientations. These filters share a gain control mechanism. The second-stage consists of a bank of rectifiers followed by linear filtering (the RF of the FRF scheme) taking their input from high-frequency first-stage filters. This stage extracts the AM cue and is not directly subject to gain control. At the third-stage we take a weighted sum of the outputs of like-oriented linear and FRF channels; producing behaviour like that of Zhou & Baker's (1996) envelope neurons. This final stage is subject to gain-control. We envisage that separate signals for first- and second-order cues are available at the points marked LM and AM respectively and that these signals support the detection of these cues.

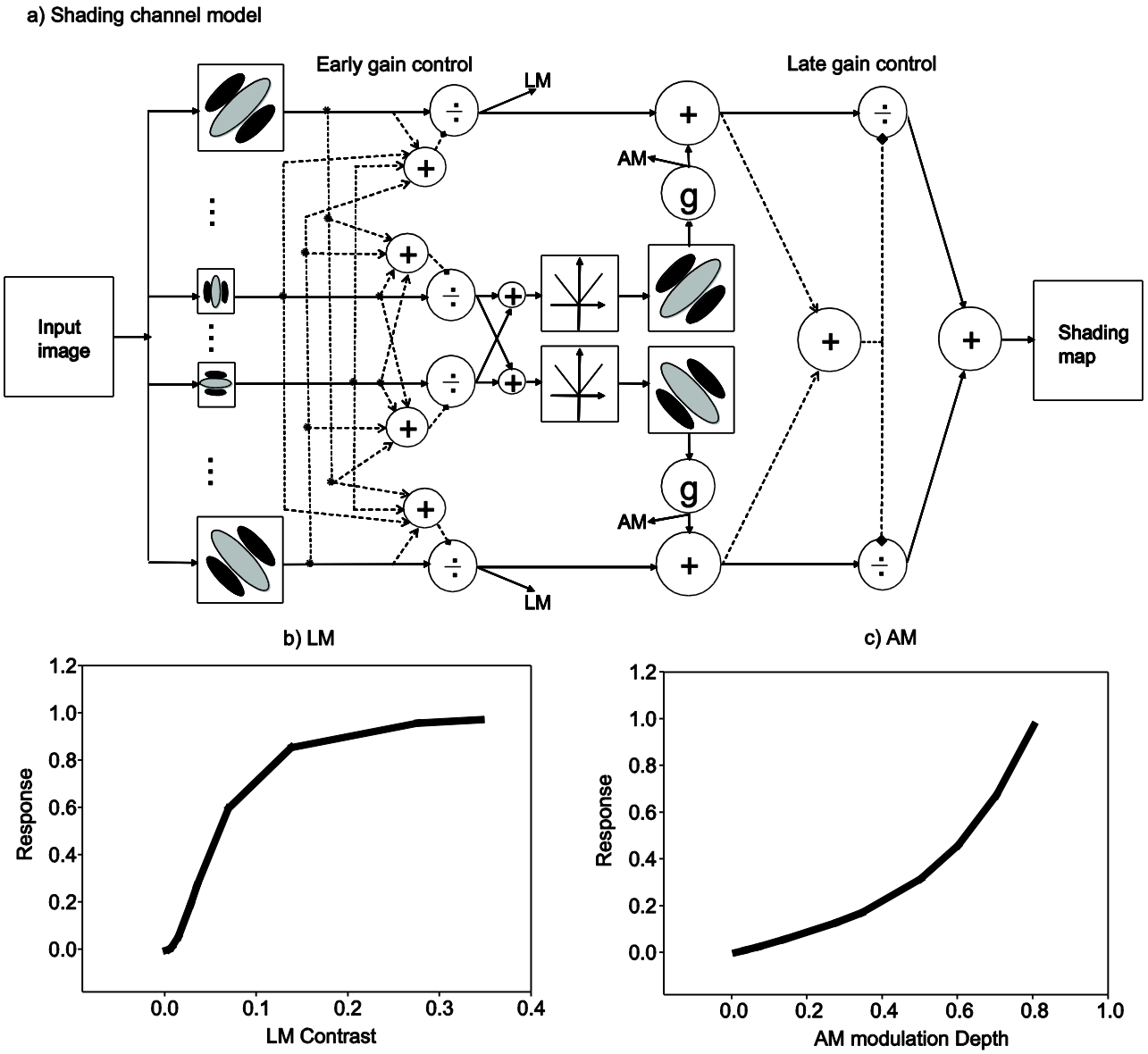


Figure 7. a) Schematic diagram of the 'shading-channel' model; see text for description. b) input-output response for first-order (LM) sub-channel. c) input-output response for second-order (AM) sub-channel.

We now address the biological plausibility of the proposed scheme, considering the following components: Linear first-stage filtering with gain control, rectification, independent outputs, weighted summation between sub-mechanisms, final gain control.

Linear first-stage filtering with gain control: Linear spatial frequency channels were first proposed by Campbell and Robson (1968) and are now accepted as the basis for early visual processing. More recent evidence suggests that while such mechanisms are approximately linear they have a non-linear transfer function which is expansive

for low-input values and compressive for larger inputs (Legge & Foley, 1980). This compression is now thought to be due a contrast gain control mechanism that pools input from many channels and across space (Foley, 1994) and has been proposed as an explanation for the compressive behaviour of simple cells in primary visual cortex (Albrecht & Geisler, 1991; Heeger, 1992). However, the pooling process is far from uniform: masking (and indeed facilitation) depends on the relative, frequency, orientation and spatial locations of the test and mask stimuli giving rise to complex patterns of behaviour (Foley, 1994; Meese, Challinor, Summers & Baker, 2009; Meese, 2004). Specifically, a given channel receives most masking from channels tuned to similar frequencies and orientations although the orientation tuning of masking is very broad (Foley, 1994). Thus we apply cross-channel gain control to our first-stage filters. Each filter has its own gain control pool with equal weight being given to all orientations in the pool but less weight given to frequencies distant from the preferred frequency of the filter in question. Because of the simple nature of our stimuli, we only modelled first-stage filters tuned to the image equivalent of 0.4 and 16c/deg and ±45°. First-stage responses are given by Equation 1.

$$R_i = \frac{C_i^p}{s_1^q + (\sum C_a^q + w\sum C_b^q)}, \qquad (1)$$

Where C_i is the pre-gain control response of the ith filter, C_a the response of all filters with the same preferred frequency as the ith filter, C_b the response of filters with preferred frequency different to that of the ith filter, w is the weight applied to off-frequency filters in the gain pool, p and q represent exponents on the forward and gain control terms respectively and s_1 is the semi saturation constant. In line with other similar models we set p and q to 2.0 (e.g. Meese et al, 2009); s_1 and w were free parameters. Application of this gain control mechanism results in a first-stage transfer function that initially accelerates and then saturates (Figure 7b) broadly consistent with both psychophysical 'dipper' experiments (Legge & Foley, 1980) and physiology (Albrecht & Geisler, 1991; Ledgeway et al., 2005) .

Rectification: Non-linear, FRF channels similar to our rectification stage (where the first filters are found in the first-stage of our model) have been proposed to explain

the detection of contrast modulations (our AM; Wilson, Ferrara and Yo, 1992) and various texture segmentation phenomena (Landy & Bergen, 1991; Graham & Sutter, 2000). Although the FRF mechanism is now widely accepted as the means by which second-order cues are detected, debates continue about the wiring between first- and second-stage filters and the shape of the rectifying non-linearity. Within the context of our limited model and following Sutter, Sperling & Chubb (1995) and Dakin & Mareschal (2000) we connect our second-stage filters to only the high-frequency first stage filters according to Equation 2.

$$S_i = f_i\left(\left|\sum f_{hf}(I)\right|^{\gamma}\right), \quad (2)$$

Where f_i is a second-stage filter with the same spatial frequency and orientation as the ith first-stage filter (but only low frequency second-stage filters are implemented), f_{hf} are the high frequency first-stage filters, $|\cdot|$ represents rectification and γ governs the shape of the rectifier. We sum first-stage filter responses across orientation and after application of the gain control (Equation 1). Graham & Sutter (2000) suggest that γ should be about 3.5 however this is based on psychophysical results that depend on the operation of the whole mechanism. Ledgeway et al. (2005) note that cells responsive to second-order cues demonstrate an accelerating transfer function and do not saturate. We used a linear rectifier (γ=1) but tested the transfer function of our model in respect of AM signals and found it to accelerate as the cube of input strength with no saturation (see Figure 7c). This lack of saturation can explain why CM stimuli do not mask themselves (Schofield & Georgeson, 1999). We believe that the early gain control mechanism and linear rectifier serve to produce the a cubic transfer function in the FRF network. It should be noted that cell responses to second-order stimuli are likely to saturate at some point if both the carrier and modulation signals are high enough. Due to the simplicity of our stimuli we only implemented second-stage filters at 0.4 c/deg and ±45°.

Independent outputs: It should be noted at this point that second-order detection could in principle be achieved by a single stage of non-linear filtering but that this would prevent the independent processing of first- and second-order cues. In the introduction

we describe a considerable body of evidence to suggest that the cues are detected independently. We will not rehearse that argument here but it is our basis for proposing a separate second-order mechanism. However, the finding that cells responsive to first- and second-order cues have different preferred frequencies for the two cues strongly suggests the existence of separate sub-mechanisms (Mareschal & Baker, 1998). Given that we will shortly propose the integration of first- and second-order cues the evidence for independent detection also leads us to propose that the outputs of the mechanisms are separately available. If the first-order signals were extracted prior to the summation stage this would explain why CM does not mask LM as second-order signals have no direct access to the first-stage gain control mechanism. This 'separate signals' hypothesis is somewhat at odds with the physiological evidence. Although cells responsive to only first-order and both first- and second-order cues have been found there is little or no physiological evidence for the existence of cells responsive to second-order signals only, but (as discussed in the introduction) this may be due to sampling biases.

Weighted summation between sub-mechanisms: For motion at least there is compelling physiological evidence for cells which linearly sum first- and second-order information (Mareschal and Baker, 1998; Zhou and Baker, 1996; Ledgeway, et al., 2005; Hutchinson, et al., 2007). Hutchinson, et al. (2007) explicitly tested for interactions between the two cues and found that cell responses were dependent on the phase relationship between the two cues, strongest for in-phase stimuli and considerably weaker for anti-phase stimuli. They used stimuli that produced equally strong responses when presented alone. Our AM cues were weaker (compared to threshold) than our LM cues so we should expect a weaker interaction. We note that our second-order mechanism is inherently insensitive. That is, by the time our relatively weak carrier has been filtered and the envelope extracted the response to the AM cue is very low - about $1/30^{th}$ of the equivalent LM response. In order to provide some differentiation between LM+AM and LM-AM and to give the model more flexibility we introduced a gain term (or weight) on the output of the second-stage filters. However, it is the overall sensitivity to AM relative to that for LM which matters. The output of each 'shading channel' after the sum is given simply by:

$$D_i = R_i + gS_i \quad (3)$$

where g is the gain term for the second-order mechanisms. Only low frequency first-stage filters and their corresponding second-stage filters are included at this stage.

Final gain-control: The final gain-control process is the most speculative part of the model but its existence and position are fundamental to the successful operation of the model. It is this mechanisms which turns the relatively poor differentiation between LM+AM and LM-AM for single gratings into the relatively strong differences found for plaids. Its position , after summation, is key to this. If it acted before LM and AM were summed then there would be no difference in signals to drive the 'winner take all' behaviour that the model needs to describe the plaid data. External justification for late gain control is provided by late interactions between the cues as noted in the introduction; most notably the transfer of the contrast-reduction after-effect and the tilt after-effect (Georgeson & Schofield, 2002). Several authors have linked simultaneous masking with sequential adaptation (Foley & Chen, 1997; Meese & Holmes, 2002). So evidence for a cross-over of adaptation could be taken as evidence of gain control. But, based on the evidence for independent detection this would have to take place after an initial detection stage. The final response of the model is given by Equation 4,

$$U_i = K \cdot \frac{D_i^{\,p}}{s_2^q + \sum D_j^q}, \quad (4)$$

where D_i is the output from the ith 'shading channel', D_j is jth channel's input to the gain control pool, s_2 is the semi-saturation constant and exponents p and q were again set to 2.0. K is a final scaling factor used to equate the range of model outputs to the human data but with no influence on the shape of the model output curves.

Implementation

For the purpose of fitting the data, the model was implemented analytically. That is we calculated ideal filter responses based on the stimulus parameters: we did not actually filter images. We subsequently implemented a 'filter-based' version of the

model that was capable of processing natural images (see later text). A final consideration is how to relate model output to measured PDAs. If we assume that the final output of the model described above is fed into a shape-from-shading module then the model output up to that point can be thought of as a conditioned shading signal. That is, LM is assumed to be a shading signal but its efficacy is modulated both by the presence of AM with the same orientation and the context provided from other orientations. For the purposes of model fitting we assume a linear relationship between the input and output of the hypothesised shape-from-shading module (Pentland, 1988) such that the contrast of the input signal at any orientation gives the perceived depth of surface undulations in that direction up to a scale factor; K in Equation 4.

Operation of the model

When an LM/AM mix is presented on only one oblique the action of the normalisation stage is largely irrelevant as there are only two channels, one of which has no output. In this case AM will have a slight modulatory effect on the shading signal determined by the overall sensitivity of the AM channel. LM-AM will hence be seen as less corrugated than LM+AM but the difference will be small. When an LM/AM plaid is presented to the model the stronger LM+AM signal will dominate the weaker LM-AM signal at the final gain control stage, driving its output down but the mutual inhibition will also limit the LM+AM signal to a value below that which would be obtained for LM+AM alone.

Model fits

The model described above has four free parameters: w, the weight applied to off-frequency maskers in the gain control of Equation 1, the semi-saturation constants s_1 and s_2, and the second-stage gain term g. Noting that, due to arbitrary scaling, the maximum theoretical output of the model prior to the multiplier K is 1 we simply set K=4 to match the maximum mean PDA. The remaining parameters were fit to the data for Experiments 2 & 3 using the *fminsearch* function in *Matlab* (The Mathworks Inc, MA). Fitted parameter values are shown in Table 1 and the fits are shown as lines in Figure 6. The model fits the data well. A key characteristic of the model is that it allows LM-AM to be seen as relatively strongly modulated in depth when presented alone but flat when presented in a plaid. The model highlights the continuous nature

of the relationship between LM and AM. Even in the plaid case adding weak AM does not produce an abrupt change in perceived depth amplitude.

Parameter	Value
w	**0.23**
s_1	**0.029**
s_2	**0.25**
g	**3.0**

Table 1: Model parameters

We also used the model to predict the results of Experiment 1. Here PDA was measured as a function of AM threshold. The model has no concept of threshold so we added an extra parameter T which represents the base AM modulation depth from which model 'threshold' multiples were calculated. This parameter was used to fit the model to the data of Experiment 1 but with no further adjustment of the other parameters. Model predictions are shown as lines in Figure 4. The model provides a good fit to the data.

The gain term g is of interest only because it relates to the overall sensitivity of the second-order mechanism. Of more interest is the relative sensitivity of the two mechanisms. We recorded output strengths for LM-only and AM-only gratings at contrast / modulation depth = 0.2. These were 0.93, and 0.09 respectively, making second-order sensitivity $1/10^{th}$ that of first-order, and correctly predicting the ratio found by Schofield & Georgeson (1999) on noise carriers with contrast = 0.1 (as used here).

Processing natural images

It is useful to fit an analytical model to data, as done here. In particular restricting the complexity of the model reduces the number of free parameters and this is useful for fitting purposes. However, it does not follow that the model will produce meaningful results when applied to real world images such as that in Figure 1. Even if implemented with filters the model described above would be useless in such an application because it has only two oriented channels at one spatial frequency. At best

it would produce plaid-like outputs for every image. We therefore implemented a more complete model with multiple orientation and frequency channels (both first- and second-order) carried through to the final output. We used 3 frequency bands and 16 orientations; 48 channels in all. Apart from having multiple channels the structure of the model was very similar to that of Figure 7, a key difference being that we dispensed with the early gain-control stage and replaced it with a simple sigmoidal transfer function. We did this because we felt unable to model the subtle spatial interactions required of a full blown gain control mechanism (Meese 2004). This model captures the spirit of the 'shading-channels' described above. As might be expected we find the model to be most effective in cases where LM+AM and LM-AM co-exist in the same scene. Figure 8a shows an example input image and the resulting model output (Figure 8b). Figure 8c show the result of processing the stimulus example shown if Figure 2e. In both cases the model successfully separates shading (or perceived shading) from reflectance changes.

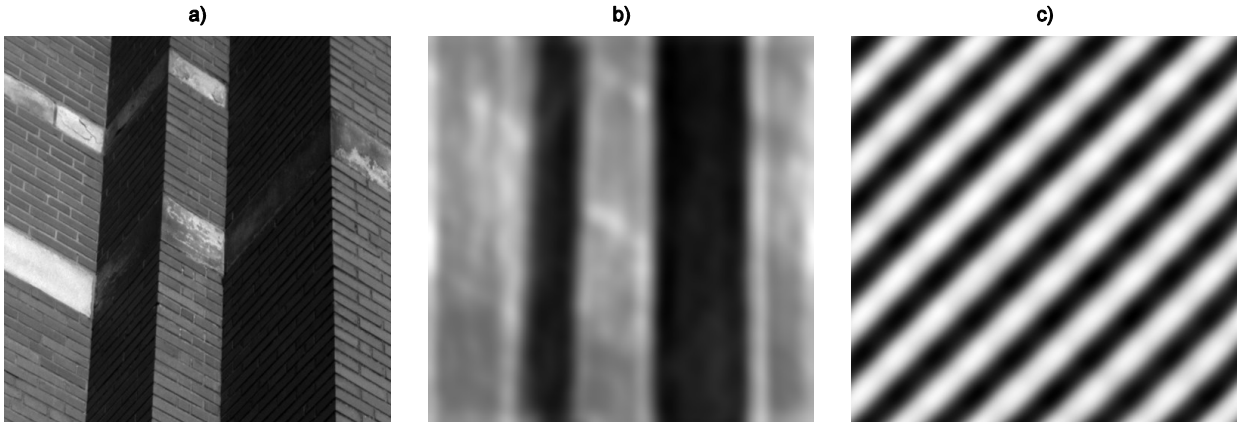


Figure 8. b) Results of applying the multi-channel shading model to an image of a section of wall (a) similar to that shown in Figure 1. c) results of applying the model to the plaid stimulus of Figure 2e.

General discussion

The results presented here extend those of Schofield et al. (2006) by introducing a more natural depth matching task, new test conditions, and a computational model. Observers had to set the amplitude of haptic stimuli to match the properties of a visually perceived surface. Perceived depth amplitude increased with overall modulation strength (Experiment 1) for all stimuli containing LM except LM-AM in a plaid. LM-AM in a plaid was perceived as nearly flat across a range of signal

strengths but, consistent with our previous findings, LM-AM was seen as modulated in depth when presented alone. Note that, as we found previously, LM-AM alone was seen corrugated, but less so than LM+AM alone. This difference is smaller when measured with the haptic task. Keeping LM contrast constant while varying AM modulation depth (Experiments 2 & 3) allowed us to study the influence of AM on LM cues. Increased AM modulation depth did not greatly affect the PDA of LM when the two were presented in-phase and alone (LM+AM, circles to right of Figure 6). Anti-phase AM did reduce the PDA of the associated LM signal (LM-AM) but only slightly (circles to left of Figure 6). However, AM had a more marked influence on PDAs in the plaid configuration. Here increasing AM in-phase with LM produced a marked but saturating increase in PDA while anti-phase AM reduced PDA (squares in Figure 6). We stress that in these plaids LM+AM and LM-AM were seen together such that as AM was stronger in the LM-AM component it also became stronger in the associated LM+AM component and vice versa. The pattern of results observed would not necessarily hold if say the LM-AM member of a plaid were fixed while the AM part of the LM+AM cue was allowed to vary, although the model would allow us to make predictions for this case. Amplitude modulations presented alone produced only a weak depth percept but perceived depth amplitude did increase a little with AM modulation depth (triangles in Figure 6).

It is tempting to suggest that higher-level cognitive processes must be at work in the interpretation of stimuli when, as here, the stimulus context is relevant to the interpretation of a particular cue: here LM-AM was seen as flat only when present in a plaid with LM+AM. However, we have successfully modelled the data with an architecture that requires no top down control and which could well be implemented in early visual areas such as V1 or V2 with the possible aid of V3a to process AM. The model combines LM and AM responses in an additive fashion within a given orientation / frequency band and then combines those responses across different orientations with gain control governing the balance between them. The resultant shading signal tends to be stronger when AM is presented in-phase with LM, but is very weak when the anti-phase combination occurs in a plaid alongside an LM+AM component. A multi-channel version of the model was tested on natural images and worked well in conditions were LM+AM and LM-AM cues co-existed.

The model presents some challenges to our previous work on cue independence. We have previously argued quite strongly that LM and CM (in our current terminology AM) are detected independently (Schofield & Georgeson, 1999; Georgeson & Schofield, 2002), but our current model suggests relatively early summation and a lack of independence. We suggest that LM and AM are indeed detected independently and are thus (for example) discriminable at threshold but that they are summed for the purpose of disambiguating the role of the luminance cue at some stage beyond simple detection. Such a configuration would allow the two cues to interact in various ways both with each other and with other cues such as disparity and texture. Our proposal here is that the two cues are summed to aid the computation of shape-from-shading, and perhaps in other situations too, but we don't suppose that this summation is either ubiquitous or mandatory.

The model makes some clear predictions about interaction of LM and AM in shape-from-shading. If such processing is based on the early channel-like mechanisms with gain control then we should expect interactions along the lines of those described above for a variety of interleaved stimuli. For example, we might expect it to be possible for LM-AM to be seen as corrugated if presented alone in one part of a stimulus but flat in some other part of the same stimulus if it overlapped with LM+AM in that region. We might expect some degree of spatial overlap to be necessary between LM+AM and LM-AM for the latter cue to be seen as flat but that the overlap need not be complete. We predict that plaids should behave as described above when their components are not orthogonal, but only if there is sufficient separation between the orientations that they fall into different orientation channels. We similarly expect LM+AM and LM-AM to dissociate if handled by different spatial frequency channels. Finally, adding an additional LM+AM component at another orientation should further suppress PDA for an LM-AM cue. We have yet to test these interesting predictions.

We presume that if AM is used to disambiguate LM in the way described above then this interaction should be driven by ecologically valid constraints. That is, LM-AM should be a reliable cue to a material change but only in the context of LM+AM cues. We have previously noted that visual texture can arise from a variety of sources and that the yoking of LM and AM (LM+AM) is only guaranteed for shaded albedo

textures (Schofield et al., 2006). LM-AM can arise when a rough, corrugated surface is shaded, although such an outcome is not guaranteed. However, it is highly unlikely that a doubly corrugated, locally rough surface could give rise to LM-AM on one oblique and LM+AM on the other. We therefore conclude that the co-presentation of LM+AM and LM-AM confirms the former cue as shading of an albedo texture and the latter cue as due to reflectance changes within that texture.

Conclusion

In conclusion, second-order modulations (specifically modulations of local luminance amplitude / contrast) can affect the perception of shape-from-shading from luminance-modulated textures. In some cases this influence is profound with the phase relationship between LM and AM determining the perceptual role of the luminance cue, flipping it from being used as a shading cue to a cue for material change. Given that luminance changes are ambiguous about their environmental causes, second-order vision may play an important role in the interpretation of luminance variations. Perhaps the need to compare these two cues is one reason why human vision is configured to detect AM (CM) cues separately from LM in the first place. In general, when AM varies in anti-phase with LM (LM-AM) surfaces are seen as flatter than when the two cues co-vary in phase (LM+AM). The flattening observed in LM-AM stimuli is most pronounced when it is presented in a plaid configuration with an LM+AM cue. However, this context effect does not require a top-down interpretation because it was possible to model key features of our data using bottom-up channel-like mechanisms.

Acknowledgement

This work was supported by EPSRC grants GR/S07254/01 & EP/F026269/1 to AJS and GR/S07261/01 to MAG. We thank the two anonymous reviewers for their helpful comments.

References

Adams, W.J., Graf, E.W. & Ernst, M.O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience, 7,* 1057-1058.

Albrecht, D.G., & Geisler, W.S. (1991). Motion selectivity and the contrast-response function of simple cells in visual cortex. *Visual Neuroscience, 7*, 531-546.

Albright, T.D. (1992). Form-cue invariant motion processing in primate visual cortex. Science, 255, 1141-1143.

Allard, R., & Faubert, J. (2007). Double dissociation between first- and second-order processing. *Vision Research, 47,* 1129-1141.

Ashida, H., Lingnau, A., Wall, M.B., & Smith, A.T. (2007). *Journal of Neurophysiology, 97*, 1319-1325

Barrow, H.G., & Tenebaum, J.M. (1978). Recovering intrinsic scene characteristics from images. In: A. Hanson and E. Riseman, Eds, *Computer Vision Systems*, Academic Press, New York.

Badcock, D.R., Clifford, C.W.G., & Khun, S.K. (2005). Interactions between luminance and contrast signals in global form detection. *Vision Research, 45*, 881-889.

Baker, C. L. Jr. (1999). Central neural mechanisms for detecting second-order motion. *Current Opinion in Neurobiology, 9,* 461-466.

Brewster, D. (1826). On the optical illusion of the conversion of cameos into intaglios, and intaglios into cameos, with and account of other analogous phenomena, *Edinburgh Journal of Science, 4,* 99-108.

Cavanagh, P. & Mather, G. (1989). Motion: the long and short of it. *Spatial Vision, 4,* 103-129.

Campbell, F.W., & Robson, J.G. (1968) Application of fourier analysis to the visibility of gratings. *Journal of Physiology, London. 181*, 576-445.

Christou, C. G. & Koenderink, J. J. (1997). Light source dependence in shape from shading. *Vision Research, 37,* 1441-1449.

Dakin, S. C. & Mareschal, I. (2000). Sensitivity to contrast modulation depends on carrier spatial frequency and orientation. *Vision Research, 40,* 311-329.

Ellemberg, D., Allen, H.A., & Hess, R.F., (2004). Investigating local network interactions underlying first- and second-order processing. *Vision Research, 44*, 1787-1797.

Ellemberg, D., Allen, H.A., & Hess, R.F., (2006). Second-order spatial frequency and orientation channels in human vision. *Vision Research, 46*, 2798-2803.

Erens, R. G. F., Kappers, A. M. L. & Koenderink, J. J. (1993). Perception of local shape from shading. *Perception & Psychophysics, 54,* 145-156.

Fleet, D. J. & Langley, K. (1994). Computational analysis of non-Fourier motion, Vision Research 34,

3057–3079.

Foley, J. (1994). Human luminance patter-vision mechanisms: masking experiments require a new model. *Journal of the Optical Society of America A,*11, 1710-1719.

Foley, J.M., & Chen, C.C. (1997) Analysis of the effect of pattern adaptation on pattern pedestal effects: A two-process model. Vision Research, 37, 2781-2788.

Georgeson, M. A. & Schofield, A. J. (2002). Shading and texture: separate information channels with a common adaptation mechanism? *Spatial Vision, 16,* 59-76.

Graham, N & Sutter, A. (2000). Normalization: contrast-gain control in simple (Fourier) and complex (non-Fourier) pathways of pattern vision, Vision Research, 40, 2737-2761.

Heeger, D., (1992) Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*, 181-197.

Heeger, D. (1993) Modelling simple-cell direction selectivity with normalised, half-squared, linear operators. *Journal of Neurophysiology*, 70, 1885-1898.

Henning, G.B., Hertz, B.G., & Broadbent, D.E. (1975) .Some experiments bearing on the hypothesis that the visual system analyses spatial patterns in independent band of spatial frequency. *Vision Research, 15*, 887-897.

Hess, R.F., Ledgeway, T., & Dakin, S. (2000). Impoverished second-order input to global linking in human vision. *Vision Research, 40*, 3309-3318.

Horn, B.K.P, & Brooks, M.J., (1989) *Shape from shading*. Cambridge, MA: MIT Press.

Hutchinson, C.V., Baker, C.L., Jr., & Ledgeway, T. (2007) Response to combined first-order and second-order motion in visual cortex neurons. *Perception, 36*, 305,306.

Johnson, A. P. & Baker, C. L. (2004). First- and second-order information in natural images: a filter-based approach to image statistics. *Journal Of the Optical Society Of America A-Optics Image Science and Vision, 21,* 913-925.

Kingdom, F.A.A. (2003) Colour brings relief to human vision*, Nature Neuroscience*, 6, 641-644.

Kleffner, D. A. & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception & Psychophysics, 52,* 18-36.

Klein, S. A., Hu, Q. J., & Carney, T. (1996). The Adjacent Pixel Nonlinearity: Problems and Solutions. *Vision Research, 36,* 3167-3181.

Knill, D. C. (1992) The perception of surface contours and surface shape: from computation to psychophysics. *Journal of the Optical Society of America A*, 9 (9), 1449-1464.

Landy, M. S. & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research, 31,* 679-691.

Langer, M. S. & Bulthoff, H. H. (2000). Depth discrimination from shading under diffuse lighting. *Perception, 29,* 649-660.

Larsson, J., Landy, M.S., & Heeger, D.J. (2006) Orientation-selective adaptation to first- and second-order patterns in human visual cortex, *Journal of Neurophysiology, 95*, 862-881.

Ledgway, T., Zhan, C.A., Johnson, A.P., Song, Y., & Baker, C.L., Jr. (2005). The direction-selective contrast response of area 18 neurons in different for first- and second-order motion. *Visual Neuroscience, 22*, 87-99.

Legge, G., & Foley, J. (1980). Contrast masking in human vision. *Journal of the Optical Society of America, 70*, 1458-1471.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49,* 467-&.

Mamassian, P. & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition, 81,* B1-B9.

Mareschal, I. & Baker C.L., Jr. (1998). Temporal and spatial response to second-order stimuli in cat area 18. *Journal of Neurophysiology*, 80, 2811-2823.

Meese, T.S. (2004). Area summation and masking. *Journal of Vision, 4*, 930-943.

Meese, T.S., Challinor, K.L., Summers, R.J., & Baker, D.H. (2009). Suppression pathways saturate with contrast for parallel surrounds but not for superimposed cross-oriented masks. *Vision Research, 49,* 2927-2935.

Meese, T.S., & Holmes, D.J. (2002). Adaptation and gain pool summation: alternative models and masking data. *Vision Research, 42*, 1113-1125.

Morgan, M.J., Mason, A.J.S., & Baldassi, S. (2000). Are there separate first-order and second-order mechanisms for orientation discrimination? Vision Research, 40, 1751-1763.

Nachmias, J. (1989). Contrast modulated maskers: test of a late nonlinearity hypothesis. *Vision Research, 29*, 137-142.

Nachmias, J., & Rogowitz, B.E., (1983). Masking by spatially-modulated gratings. *Vision Research*, 23, 1621-1629.

Olmos, A. & Kingdom, F.A.A. (2004) A biologically inspired algorithm for the recovery of shading and reflectance images, *Perception*, 33, 1463-1473.

Pentland, A. (1988) Shape Information From Shading: A Theory About Human Perception,
Second International Conference on Computer Vision, 404-413.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, 331,* 163-165.

Rittenhouse, D. (1786). Explanation of an optical deception. *Transactions of the American Philosophical Society, 2,* 37-42.

Schofield, A. J. (2000). What does second-order vision see in an image? *Perception, 29,* 1071-1086.

Schofield, A. J. & Georgeson, M. A. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: separate mechanisms with similar behaviour. *Vision Research, 39,* 2697-2716.

Schofield, A. J. & Georgeson, M. A. (2003). Sensitivity to contrast modulation: the spatial frequency dependence of second-order vision. *Vision Research, 43,* 243-259.

Schofield, A. J., Hesse, G., Rock, P. B., & Georgeson, M. A. (2006). Local luminance amplitude modulates the interpretation of shape-from-shading in textured surfaces. *Vision Research, 46,* 3462-3482.

Schofield, A. J., Rock, P. B., & Georgeson, M. A. (submitted). Is there a sun? Probing the default illuminant for shape from shading. Submitted to Journal of Vision.

Schofield, A. J., Rock, P. B., Georgeson, M. A., & Yates, T. A. (2007). Humans assume a mixture of diffuse and point-source lighting when viewing sinusoidal shading patterns. *Perception, 36, Supplement,* 108-109.

Smith, A. T. & Ledgeway, T. (1997). Separate detection of moving luminance and contrast modulations: Fact or artifact? *Vision Research, 37,* 45-62.

Smith, A. T., Scott-Samuel, N. E. (2001). First-order and second-order signals combine to improve perceptual accuracy. *Journal of the Optical Society of America A. 18*, 2267-2272.

Smith, S., Clifford, C.W.G., & Wenderoth, P. (2001). Interaction between first- and second-order orientation channels revealed by the tilt illusion: psychophysics and computational modelling. *Vision Research, 41*, 1057-1071.

Song, Y. & Baker, C.L., Jr. (2006) Neural mechanisms mediating responses to abutting gratings: Luminance edges vs. illusory contours. *Visual Neuroscience*, 23, 181-199.

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

Sutter, A., Sperling, G. & Chubb, C. (1995). Measuring the spatial frequency selectivity of second-order texture mechanisms. *Vision Research, 35,* 915-924.

Tappen, M.F., Freeman, W.T., & Adelson, E.H. (2005). Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*, 1459-1472.

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,* 583-595.

Tyler, C. W. (1998). Diffuse illumination as a default assumption for shape-from-shading in graded images. *Journal of Image Science and Technology, 42,* 319-325.

Wilson, H.R., Ferrera, V.P. & Yo, C. (1992) A psychophysically motivated model for two-dimensional motion perception, *Visual Neuroscience*, 9, 79-97.

Zhou, Y.-X. & Baker, C.L., Jr. (1996) Spatial properties of envelope-responsive cells in area 17 and 18 neurons of the cat, *Journal of Neurophysiology*, 75, 1038-1050.

Appendix 2: Published Conference Abstracts

High frequency textures provide better support for shape-from-shading than low frequency textures

Peng Sun and Andrew Schofield

Abstract. Observers perceive a sinusoidally shaded texture as a corrugated surface even when the texture elements themselves undergo no geometric distortions (Schofield, Heese, Rock & Georgeson, 2006, *Vision Research, 46*, 3462–3482). Using a similar two-point probe task but Gabor noise textures, we varied the dominant spatial frequency of the texture (from 1.5 to 12 c/deg) and found that high frequency textures support a more robust percept of shape-from-shading than do low frequency textures. Given that our sinusoidal shading patterns were themselves low frequency (0.5 c/deg) we were concerned that this difference may be due to masking. That is, the low frequency textures might simply have reduced the visibility of the shading patterns. To control for this we varied the dominant orientation of the textures so as to reduce their ability to mask the shading pattern; this had no affect. Reducing the spatial-frequency bandwidth of the textures, which should reduced masking, also had no affect. Multiplicative shading of an albedo textured surface produces a change in local mean luminance coupled with a change local luminance amplitude (AM). Schofield et al. (2006) showed that this AM cue modulates the perception of shape-from-shading. Given that AM is a second-order cue requiring comparisons across pairs of pixels, our results are consitent with the idea that second-order processes receive most of their input from high-frequency channels (Dakin & Mareschal, 2000, *Vision Research, 40*, 311–329). We speculate that when the carrier texture is high frequency, AM is detected well and thus supports shape-from-shading. When the carrier is low frequency AM is detected less well and consequently shape-from-shading is inhibited.

Shape-from-shading for grating stimuli: Slant is proportional to luminance, with some exceptions

Andrew Schofield and Peng Sun

Abstract. Humans are able to interpret luminance variations as changes in shading which are in turn interpreted as due to undulations of an illuminated surface. In general, we seem to adopt the implicit assumptions that surfaces are Lambertian and illuminated by a point source such that luminance in proportional to the angle between the surface normal and the direction of the illuminant. Thus, perceived surface slant depends on luminance. Most studies of shape-from-shading use stimuli based on simulations of solid objects viewed under a specified light source. We took an alternative approach; measuring the perceived shape of a range of grating stimuli (horizontal sine-wave, square-wave, and saw-tooth gratings). Observers set the slant of a probe disk to match the slant of the perceived surface at various points on each grating. In most cases perceived slant was proportional to luminance with mean luminance equal to zero slant (surface locally fonto-parallel). Sinusoidal luminance modulations produced sinusoidal perceived surfaces even though sinusoidal corrugations seldom produce sinusoidal shading patterns in real scenes. Square-wave

luminance profiles produced triangular perceived surface profiles. Saw-tooth luminance profiles with several repetitions produced perceived surfaces that were dished or bowed (depending of the direction of the luminance ramps) with surface sections meeting at localised ridges/troughs. We found one notable exception to the general result that slant is proportional to luminance. Stimuli consisting of just two linear ramps in a saw-tooth configuration were mapped as a largely flat surface with a single central crease. The regions at the top and bottom of such stimuli were perceived to have zero slant even though luminance varied linearly in these regions and was not close to mean luminance. This result suggests that luminance edges and boundaries affect the perception of shape-from-shading even for relatively simple grating stimuli.

Using texture amplitude to recover shading and reflectance image

Peng Sun and Andrew Schofield

Abstract. In the computer vision society, shape-from-shading is a process to recover surface orientation from luminance changes in a scene. In the real world however, luminance changes due to real shading are often confounded with changes in surfaces reflectance such as hue and texture. Such ambiguity in luminance changes has been a difficulty that is confronted by many shape-from-shading algorithms which would always assume uniform surface albedo. Here we present an algorithm for separating the shading and reflectance components in grayscale images. Our algorithm exploits the same rule as appear to be used by humans to assist in shape-from-shading tasks: luminance changes that are coincident with contrast changes are likely to be due to reflectance changes whereas those that are not associated with a change in contrast are likely to be due to shading (Schofield et al, 2006). This in turn arises from the multiplicative nature of shading. The mean luminance of an image is computed first and then classified by changes in contrast, which can be obtained by applying a texture segmentation algorithm. Compare to its counterpart which is based on hue alone, this method faces the difficulty resulted from the unreliability and inaccuracy of any existing texture segmentation algorithm. We have solved this problem by introducing an edge width estimation mechanism which provides tolerance to the inaccuracy of the texture segmentation algorithm employed. The final shading component is obtained by reconstructing the classified mean luminance map, while the reflectance component is obtained by subtracting the shading component from the original image.