Vision and expertise in remote sensing surveying

Emil Skog

Doctor of Philosophy

Aston University

July 2023

Aston University

Vision and expertise in remote sensing surveying

Emil Skog
Doctor of Philosophy
July 2023

Thesis abstract

Remote sensing surveyors are tasked with extracting information from aerial landscape images to build maps and geospatial models of landscapes. This thesis focuses on four broad hypotheses related to this task. First, that the unfamiliar aerial viewpoint is more difficult to process than the familiar ground viewpoint, but surveyors are better than novices at processing this viewpoint. Next, surveyors are experienced with using binocular disparity cues in stereoscopic aerial images, and thus make better use of this cue. Further, surveyors also adapt to the aerial imagery, and this can alter perceptual priors for shape-from-shading. Finally, surveyors develop expertise, and this can in part be explained by perceptual learning. An initial study established that expert surveyors have a superior ability to process the aerial viewpoint, and better recognise aerial-view features, compared to novices from the general population. Next, depth perception in stereoscopic aerial images was explored with two specific depth cues: binocular disparity, and luminance cues used to interpret shape-from-shading. This study required the innovation of a novel version of the classification image technique, that can estimate the simultaneous use of binocular disparity and luminance cues. Experts and novices classified stereoscopic aerial images, and group differences showed that: 1) Experts are better at prioritising and sampling binocular disparity cues, and 2) experts may have adapted to diminish the conventional lighting-from-above prior following experience with counter-conventional light source directions in aerial images. Finally, as a hallmark of expertise is better processing of binocular disparity cues, the classification images were employed to explore stereoscopic perceptual learning in novices. This study found evidence of learning that characterises how observers improve to better sample disparity cues in stereograms. This thesis provides novel evidence on the mechanisms involved in interpreting stereoscopic aerial images, and characterises expertise in remote sensing surveyors.

Keywords:
Vision, expertise, aerial images, remote sensing, depth perception, binocular disparity, shape from shading, classification images, viewpoint

To Brooke

Acknowledgements

Throughout this project, I have felt extremely lucky to be supervised by two excellent professors, Andrew Schofield and Tim Meese. I am very grateful for having had the opportunity to learn from your rigorous approach to research, and for your deep and consistent dedication to my doctoral education. Thank you so much for everything you have taught me during my dream PhD project.

Many thanks to Dr. Isabel Sargent and Andy Ormerod at Ordnance Survey for supporting my project and for the vital exchanges of ideas and resources that made this project possible. Thank you also to my fellow lab member Dr. Stella Qian, and former placement students Anisha Parmar and Rasna Chowdhury, for great days spent in the lab. Another thanks to my work-from-home colleagues, the cats Cleo and Smallcat.

Thank you to my family for supporting me. A special thanks also to Sandra, Max, Steffi, and Nicole from Hamburg for saving my cat who spent a week vacationing in a bush in Germany during my move from Sweden to England. Finally, I dedicate this thesis to my girlfriend, Brooke.

List of contents

Chapter 1: General introduction
- 1.1 Aims of thesis 8
- 1.2 Studying the world remotely 10
- 1.3 Introduction overview 12
- 1.4 Visual expertise 13
- 1.5 Depth perception 15
- 1.6 Stereoscopic judgements 17
- 1.7 Luminance cues to 3D shape 20
- 1.8 Cue combination of binocular disparity and luminance 23
- 1.9 Perceptual learning 24
- 1.10 Selecting a primary method 25
- 1.11 Classification images and the perceptual template 26
- 1.12 Templates in neurophysiology 29
- 1.13 Visual psychophysics with classification images 29
- 1.14 Classification image analysis 32
- 1.15 Bubbles 35
- 1.16 Aims of thesis summary 37

Chapter 2: Expertise for aerial images: Evidence from scene gist and object matching across ground and aerial viewpoints
- 2.1 Introduction 39
- 2.2 Experiment 1 43
  - 2.2.1 Method 43
  - 2.2.2 Results 46
  - 2.2.3 Discussion 51
- 2.3 Experiment 2 51
  - 2.3.1 Method 52
  - 2.3.2 Results 56
  - 2.3.3 Discussion 60
- 2.4 General discussion 62

Chapter 3: Pilot studies
- 3.1 Pilot 1: Evaluating experimental designs for classification image studies 65
- 3.2 Pilot 2: Developing and evaluating a novel method for generating 3D classification images 68
- 3.3 Pilot 3: Novel method for generating simultaneous 2D and 3D classification images 72
- 3.4 Conclusions on classification images 78
- 3.5 Post-hoc analysis of perceptual learning for disparity targets 79

Chapter 4: Classification images for aerial images capture visual expertise for binocular disparity and a prior for lighting from above
- 4.1 Introduction 84
- 4.2 Methods 87
- 4.3 Results and discussion 96
- 4.4 General discussion 114
- 4.5 Follow-up experiment on lighting direction priors in expert surveyors 120

Chapter 5: Characterising perceptual learning for stereopsis with stereoscopic classification images
- 5.1 Introduction 127
- 5.2 Pilot experiment 129
- 5.3 Main experiment 132
  - 5.3.1 Method 132
  - 5.3.2 Results 137
- 5.4 Discussion 143

Chapter 6: Discussion
6.1 Meeting the aims of the thesis 145
6.2 Overview of results 146
6.3 Future directions 151
6.4 Implications 154
6.5 Conclusions 156
References 158
Appendices 169

List of abbreviations

2AFC	Two-alternative forced-choice
2D	Two-dimensional
3D	Three-dimensional
CI	Classification image
CM	Confusion matrix
dB	Decibel
H	Hypothesis
Man	Man-made
Nat	Natural
OS	Ordnance Survey
PL	Perceptual learning
PNG	Portable Network Graphics
RDS	Random-dot stereogram
RFT	Random field theory
RT	Response time
SE	Standard error
SD	Standard deviation
SIBR	Single-interval binary-response
SNR	Signal-to-noise ratio
T	Transfer task
UK	United Kingdom

List of Tables

Table 2.1. Results of statistical testing of sectioned confusion matrices. 51
Table 2.2. Results of repeated measures ANOVAs to accuracy and RT. 60
Table 4.1. TNO thresholds. 94
Table 4.2. Gaussian parameters for fits to the average horizontal CIs. 103
Table 4.3. Gabor parameters for fits to the average vertical CIs. 104
Table 5.1. Individual fits of linear regression models. 139

List of Figures

Figure 1.1. Stereoscopic pair of two houses. 12
Figure 1.2. Photograph of a slanted surface. 17
Figure 1.3. Illustration diagram of the horopter. 19
Figure 1.4. Dense RDS-like image pair containing a binocular disparity target. 20
Figure 1.5. Images providing an interpretation of 3D shape from shading. 21
Figure 1.6. An analysis procedure for filter-based decomposition and statistical testing of CIs. 35
Figure 2.1. Example stimulus from all scene categories from both aerial and ground viewpoints. 44
Figure 2.2. Average correct responses for the two groups and viewpoints. 47
Figure 2.3. Average confusion matrices for the two groups and viewpoints. 48
Figure 2.4. Example stimulus from the main experiment and two control experiments. 54
Figure 2.5. Accuracy results from the different experiments. 57
Figure 2.6. Response time results from the different experiments. 58
Figure 3.1. Luminance CIs from the Pilot 1. 68
Figure 3.2. Algorithm for generating disparity noise with an example image. 70
Figure 3.3. Disparity CIs from Pilot 2. 72
Figure 3.4. Cross-sections of disparity CIs. 72
Figure 3.5. CIs for disparity, luminance, and the compound condition from Pilot 3. 75
Figure 3.6. Cross-sections of disparity CIs. 76
Figure 3.7. Cross-sections of luminance CIs. 76
Figure 3.8. Disparity thresholds across sessions from Pilot 2. 81
Figure 3.9. Disparity thresholds across sessions from Pilot 3. 82
Figure 4.1. Aerial images of hedges and ditches used to create stimulus images. 88
Figure 4.2. Example Z-coordinate texture used to map random disparities. 89
Figure 4.3. Procedure for making stimulus images with both luminance and disparity noise. 91
Figure 4.4. Hexagonal lattice ‘honeycomb’ stimulus. 92
Figure 4.5. CIs for disparity and luminance. 98
Figure 4.6. Cross-sections of disparity CIs. 101
Figure 4.7. Cross-sections of luminance CIs. 102
Figure 4.8. Curve fits to averaged cross-sections of CIs. 103
Figure 4.9. Interactions of CI amplitudes across cue type and groups. 105
Figure 4.10. Individual biases. 107
Figure 4.11. Individual sensitivities to three aspects of the stimulus images. 109
Figure 4.12. Correlation between disparity CI amplitudes and sensitivities to disparity profiles. 111
Figure 4.13. Correlation between sensitivities to lighting-from-above and offsets in CI data. 114
Figure 4.14. Example honeycomb, hedge, and ditch with light coming from above and below. 117
Figure 4.15. Example stimulus images from follow-up experiment. 122
Figure 4.16. Results from follow-up experiment. 124
Figure 5.1. Individual correct responses across sessions. 131
Figure 5.2. Procedure for generating stimulus images with disparity noise. 135
Figure 5.3. Individual thresholds across sessions. 138
Figure 5.4. Individual partial disparity CIs. 141
Figure 5.5. Cross-sections of partial disparity CIs. 142
Appendix A. Disparity CIs from different image manipulation conditions. 169
Appendix B. Luminance CIs from different image manipulation conditions. 170
Appendix C. Individual fits to vertical cross-sections of the disparity CIs. 171
Appendix D. Individual fits to vertical cross-sections of the luminance CIs. 172
Appendix E. Cross-sections of partial luminance CIs. 173

Chapter 1

General introduction

1.1 Aims of thesis

This thesis explores the visual mechanisms associated with expertise in the remote sensing surveying of stereoscopic aerial landscape images. The body of literature on expertise for remote sensing surveying is rather small, but studies have shown expertise in, for example, visual recognition memory (Šikl et al., 2019) and visual search (Lansdale, Underwood & Davies, 2010). An elaboration on this body of literature is provided below. The studies in this thesis focus on previously unexplored aspects expertise.

The studies in this thesis were developed following discussions with remote sensing surveyors at the Ordnance Survey (OS), who co-funded the work. The OS creates maps and geospatial models of the United Kingdom (UK). The discussions with the surveyors were informative on the nature of the aerial imagery and tasks, and how the surveyors work with photogrammetry. The surveyors also provided multiple examples of challenging landscape features, elaborated in the section below. A primary challenge in remote sensing surveying is the use of aerial images, as the aerial viewpoint provides an unusual view of landscapes. The first study in this thesis thus sought to explore if, and to characterise how, experts perform better than novices when viewing aerial images. Two experiments were designed to explore scene processing and object recognition from both the ground and aerial viewpoints. Natural images of scenes and objects provided a rich stimulus set, and Chapter 2 shows performance measures for e.g., processing scene gist in multiple scene categories, and identity matching objects, across both ground and aerial viewpoints. This study also explored mental rotation, and whether aerial images of objects have preferred orientations.

In addition to the general processing of aerial images described above, aerial images are provided as stereograms to OS surveyors to aid the perception of 3D shape in the landscape. The surveyors are thus very experienced with interpreting stereoscopic cues in aerial images. A primary aim of this thesis was therefore to characterise how experienced surveyors use various visual cues to support depth perception in aerial images. Chapter 4 describes how experts and novices classify landscape features based on different stereoscopic cues. The discrimination of hedges and ditches provided a suitable task for this investigation. Discriminating hedges and ditches is challenging because they are similar when viewed from above and require stereoscopic judgements based on binocular disparity and/or luminance cues. Hedges have crossed disparity and are lighter, while ditches have uncrossed disparity and are darker (Chapter 4).

In addition to these factors, stereoscopic judgements of hedges and ditches can also be influenced by sunlight direction in an interaction with human priors for lighting direction. The discussions with OS surveyors also revealed the notable observation that the OS typically orients their aerial images to face north-up. This convention is congruent with the traditional orientation of maps where the bottom of the page represents south, and the top of the page represents north. The UK is in the northern hemisphere, so the sun is always to the south of the camera. This means that the direction of the sun is typically from below the line of sight when viewed by surveyors. This produces a lighting-from-below structure that conflicts with the well-known prior for lighting-from-above (Ramachandran, 1988; Sun & Perona, 1998). Given a conflict between lit-from-below images and the lighting-from-above prior, aerial images of hedges (convex) and ditches (concave) could lead to switched interpretations of their 3D profiles. For example, in terms of shape from shading, hedges lit from below will have a ditch-like shading profile if interpreted as lit from above, and vice versa. As the typical lighting-from-above prior conflicts with lit-from-below images, it is possible that the experts have developed an exception to this prior for the aerial images, to avoid misinterpretations of 3D shape from shading. The experts might thus show evidence of interpreting aerial images with diminished or switched lighting direction priors. Inverting hedge and ditch images vertically should flip their interpretation and, exploiting such inversions, Chapter 4 included measures of lighting direction priors to explore any effects of the experts' unusual experience with lit-from-below images.

Multiple stereoscopic cues to 3D shape can thus support the discrimination of hedges and ditches, i.e., binocular disparity, diffuse luminance ('dark-is-deep'), and shape-from-shading / lighting direction priors. A primary aim was to capture how surveyors use these 3D cues without imposing tight constraints on the stimuli and task. A psychophysical technique called classification images (CI) affords the possibility of measuring the visual cues that are sampled from stimulus images during behavioural tasks. This technique was suitable as it relies on a random component that modulates visual cues in the stimulus, and the analysis focuses on how random visual cue patterns influence the task. The CI technique is elaborated on in detail below. A novel version of this technique was required to simultaneously estimate CIs from binocular disparity and luminance. Thus, an innovation supporting this thesis was this novel CI technique that is later implemented to study mechanisms involved in visual expertise for remote sensing surveying. The development of this CI technique is described in a series of pilot experiments in Chapter 3, and the technique is later applied in the study of Chapter 4.

The results in Chapter 4 suggest that experience with stereoscopic aerial images contribute to significant learning for processing disparity cues. Visual expertise can develop through long-term engagement with domain-specific stimuli through perceptual learning (PL). The study in Chapter 5

was therefore designed to characterise stereoscopic PL for improving novices' ability to sample disparity cues in stereograms. This study used a PL intervention on novice participants to characterise learning. Continuing with the CI technique, this study sought to provide a link between training to improve processing of stereograms, and how such training might change internal templates.

1.2 Studying the world remotely

Extracting information from photographs is a key component in modern map making and geospatial modelling. A primary challenge with making a detailed map or model of the world is that of scalability. A neighbourhood can be efficiently mapped in detail by a single on-site observer. But mapping a whole city or country would no longer be viable using observers who travel to see locations in-person. This solution does not scale up due to logistical restrictions and is now only used when aerial photographs cannot resolve the necessary detail. Tasked with mapping an entire country, the main strategy employed by the Ordnance Survey (OS) is photogrammetry, where countrywide aerial photographs are used for landscape feature classification. The photogrammetric tasks require observers, called remote sensing surveyors, who can accurately classify the contents of the photographs. Remote sensing is the acquisition of information without physical contact, through remote observation.

At the OS, remote sensing surveyors work to create and update map data of the UK. The remote sensing tasks are often difficult, as aerial images are unfamiliar to human observers (Lloyd, Hodgson & Stokes, 2002; Loschky et al., 2015). Almost every landscape object appears different when viewed from above compared to a normal ground view. Scenes and objects seen from the ground viewpoint afford image configurations that we are very experienced with, and our vision typically works effortlessly and accurately. For example, we are very experienced with looking at common objects such as buildings from a side-view, where the building's façade is emphasized. But in the aerial viewpoint the emphasized feature might be the roof, changing the general appearance of the object. Such an 'atypical viewpoint' can impair the ability to detect and recognise objects (e.g., Center et al., 2022). We are also experienced with the image-statistical regularities in familiar scenes seen from the ground viewpoint. For example, we expect the typical view of a beach to contain a large stretch of sand or rock in front of our feet, followed by water, and the sky horizon. This view contains well-defined horizontal edges, where sand meets water and where water meets the horizon, and constitutes the typical image statistics of a beach scene (Oliva & Torralba, 2001). But a beach from the aerial viewpoint appears different, as every surface is frontoparallel to the observer and the sand-water edge occurs in an arbitrary orientation.

In typical ground viewpoints, the 'gist' of scenes is processed rapidly despite the complexity of natural images, where for example, a basic distinction of 'natural' vs 'man-made' scenes can be discriminated in as little as 8 milliseconds, or one frame of a 120hz monitor (Furtak, Mudrik & Bola, 2022). But the gist of the scene is processed with more difficulty in aerial viewpoints (Chapter 2; Loschky et al., 2015). Objects are also recognised more easily from typical viewpoints, but recognition from atypical viewpoints is likely to be slower or more erroneous (Center et al., 2022; Lawson, 1999; Tarr et al., 1998).

To discover the types of remote sensing surveying tasks that can be especially challenging at the OS, discussions were held with less experienced surveyors (<1 year of experience) and more experienced surveyors (>1 year of experience) on separate occasions. Additional discussions were also held with a very senior surveyor throughout the project. These discussions helped to characterise how the surveyors work and what is especially challenging with photogrammetry of aerial images. The OS places high priority on accurate classification of landscape features. Remote sensing surveyors are trained and experienced with classifying features which can be very difficult to even detect for untrained observers, such as fences, drainage ditches, culverts, and bus stop stands. Surveyors also discriminate between confusable features. According to a common task specification, permanent objects should be mapped but temporary objects should be disregarded. Discriminating 'permanent vs temporary' can be difficult in examples such as telling apart a greenhouse and a polytunnel, a pile of dirt waiting to be removed and a permanent dirt mound, a lake and a temporarily flooded field, or a large tent and a building. Another challenging task specification can be to disregard man-made features. Challenging examples of discrimination between 'natural vs man-made' include embankments, rivers and drainage ditches, or hedgerows and natural shrubbery. Other challenging discriminations include hedges and ditches, bridges and tunnels, hedgerows and Cornish hedgerows (which are built into walls), pavements and bike lanes, sheds and garages, pastures and crop fields, and material properties such as asphalt and gravel in roads. Surveyors also update existing map data, where differences between an older map and a newer aerial landscape image must be detected (Lansdale, Underwood & Davies, 2010). Changes can be subtle in comparing an older map and a newer aerial image, and humans are prone to change blindness, where even relatively large changes can be missed (Rensink, 2002; Simons & Levin, 1997).

To make these photogrammetric tasks easier, images are presented stereoscopically, providing 3D viewing through binocular disparity cues. An elaboration on depth perception is provided below. Both newly recruited and experienced surveyors at the OS self-report a strong reliance on stereopsis in images. Stereograms are created by pairing appropriately spaced aerial photographs. See Figure 1.1 for an example. The stereogram landscape images are presented to the

surveyors via a pair of polarized monitors viewed through a glass frame that provides similar light across the monitor pair (3D PluraView Monitor). Polarized glasses are worn for dichoptic viewing, separating most of the received light from the two monitors to each eye. The landscape photographs are captured by aircraft, providing higher resolution imagery than typical satellite imagery. Individual photographs typically cover 2.5 x 1.5 kilometres (450 megapixels), and surveyors zoom in on these high-resolution images to reveal features more closely.

Figure 1.1: Stereoscopic pair of two houses. Binocular fusion is achievable by crossing the eyes so that the left-hand and right-hand image is seen by the right eye and left eye, respectively. *© Crown copyright and database rights 2023 OS*, used with permission.

Surveyors with years of experience tend to be better at the various photogrammetric tasks compared to less experienced colleagues (Lansdale, Underwood & Davies, 2010; Šikl et al., 2019). To help manage the challenging visual tasks, the OS explicitly trains newly recruited surveyors to learn how to classify features in aerial landscape images, and difficult tasks are often tackled in collaboration with colleagues. Newly recruited surveyors are believed to improve through work experience, developing expertise within this domain.

1.3 Introduction overview

This introduction is organised into topics that reflect four broad hypotheses in this thesis. These hypotheses regard the mechanisms associated with expertise in remote sensing surveying of stereoscopic aerial landscape images. The first hypothesis states that the unfamiliar aerial viewpoint is more difficult to process, but expert surveyors are better at processing the aerial viewpoint. Next, surveyors at OS are experienced with using binocular disparity cues in stereoscopic aerial images, and they are better able to process this cue. The surveyors also adapt to the aerial imagery, and this can alter perceptual priors for interpreting shape from shading. Finally, the surveyors develop expertise from experience, and this can in part be explained by PL.

This thesis studies expertise in remote sensing surveyors, and this general introduction continues with a discussion of visual expertise in general, and in surveyors. The four broad hypotheses described above reflect different topics in vision research that will be discussed in this chapter. The first topic, relating to processing difficulties with aerial viewpoints, was briefly covered in the previous section, and is the subject of the study in Chapter 2. The second and third topics relate to depth perception with binocular disparity and luminance cues. These topics will be extensively studied in this thesis, and this chapter will contain an elaboration on how these cues contribute to depth perception. Cues supporting depth perception will be covered after the section on visual expertise. Following this, the fourth topic regards mechanisms of learning that help surveyors improve with experience. To conclude this chapter, a discussion focuses on the method that is used to study how expert surveyors use binocular disparity and luminance cues. This method was the CI technique, and a novel version of CIs was required that could simultaneously estimate the use of binocular disparity and luminance cues. The CI technique is a central method to this thesis, and later parts of this chapter provide an overview of CIs.

1.4 Visual expertise

Visual expertise is the experience-dependent development of enhanced perception for domain-specific images. Example domains involving visual expertise include chess (e.g., Reingold et al., 2001; Reingold & Sheridan, 2012), medical imagery analysis (e.g., Fox & Faulkner-Jones, 2017; Gegenfurtner, Lehtinen & Säljö, 2011; Krupinski, 2010; Krupinski et al., 2006; Wolfe et al., 2016), pilot perception in aviation (e.g., Bellenkes, Wickens, & Kramer, 1997; Kasarkis et al., 2001; Schriver et al., 2008; Ziv, 2016), and remote sensing surveying (Lansdale, Underwood & Davies 2010; Šikl et al., 2019). An expert radiologist who is trained and experienced with thousands of hours of searching for and diagnosing spots in x-rays has developed visual expertise for this domain-specific task. While experts possess expertise within the domain, they typically do not show enhanced visual abilities in general (Nodine & Krupinski, 1998; Reingold et al., 2001). Expert chess players process more information about chess games in a glance than novices, but the chess players are not likely better at general visual tasks outside of chess (Reingold et al., 2001; Reingold & Sheridan, 2012). Performance measures tend to find expertise factors such as higher accuracies, shorter response times, and fewer gaze fixations on irrelevant items. This goes for tasks within the domain of expertise, but not necessarily outside it.

Semantic and visual expertise are interlinked in many fields of expertise (Harel, 2016). For example, an aircraft mechanic might attend to and describe attributes of an aircraft differently than novices (Rosch et al., 1976). The novice might not understand the seen information in the same way

as the expert, and therefore use less of it, or miss it completely (Seitz, 2017). Rosch et al. (1976) studied conceptual hierarchies and semantic labelling, finding that novices tend to use basic level labels ('bird') more so than superordinate ('animal') or subordinate ('robin') labels. Studying expert-novice differences in labelling, Tanaka and Taylor (1991) showed again that novices rely more on basic level concepts, but experts may differentiate subordinate categories as much as basic level categories. Experts in some fields have access to a larger lexicon of semantic labels within the domain, allowing conceptually more sensitive discrimination of information. Although this is the case in some fields of expertise (e.g., mechanics), it may not be the case in all fields. Many fields involve looking for features that are conceptually relatively easy to describe to novices, such as available moves in chess games, or field boundaries in remote sensing of aerial landscape images. For example, novices and experts may both conceptually understand that 'aerial image' means an image of a scene from the above perspective. However, aerial images are unfamiliar to human observers, and experience is key to enhancing visual processing skills in new domains (Chapter 2 and 4; Lansdale, Underwood & Davies, 2010; Lloyd, Hodgson & Stokes, 2002; Seitz, 2017, 2020; Šikl et al., 2019).

A major area of research on visual expertise is medical imagery analysis. Medical imagery sub-fields, such as radiology, provides images that are unfamiliar to human observers. Radiologists working with diagnostic tasks based on such images need years of training and experience with supervision to perform the tasks with expert-level performance (Bertram et al., 2013; Fox & Faulkner-Jones, 2017; Gegenfurtner, Lehtinen & Säljö, 2011; Krupinski, 2010; Krupinski et al., 2006; Wolfe et al., 2016). Expertise in this field is commonly studied in visual search paradigms, as radiologists make carefully considered diagnostic judgements based on searching images for anomalous spots. Eye-tracking methods have shown that experts tend to make fewer gaze fixations, and spend less time attending task-irrelevant information. Radiologists can also gain more information from briefly presented (<250ms) images within their domain (Drew et al., 2013; Evans et al., 2013; Kundel & Nodine, 1975). Overall, experts are more efficient than novices in processing global image structure for guiding initial eye movements towards task-relevant areas.

A few studies have explored expertise in photogrammetry of aerial images. Šikl et al. (2019) recruited expert image analysts in a study of expertise in visual recognition memory of 2D colour aerial images. The task involved recalling previously seen images of four different scene categories (historical centers, suburbs, parks and sport fields, and industrial and transportation buildings) and two different viewing heights (800m and 1,600m). During the recall task, target images were presented with an accompanying distractor item which was either between-category, within-category, or the target image but rotated. Participants were instructed to respond which image had

been previously shown during a learning phase. Accuracy measures showed that experts recognised more previously seen aerial scenes than novices. While Šikl et al. (2019) show a memory advantage for experts, the processes involved in the expertise remain largely unexplored. For example, we might wonder if the experts' superior recognition memory relies on spatial arrangements of local features in the scenes, or colour profile characteristics, or some other image cues that were considered particularly useful for memorisation by the experts but not the novices. Further, Šikl et al. (2019) did not employ a control condition with ground-view images, which could mean that their experts performed better at the aerial images only as a consequence of trying harder at the task in general. The study also did not otherwise control for any general memory advantage that the experts might have had. Finally, the authors used large-scale images depicting scenes spanning hundreds of square meters. This scale is useful to reveal general land-use categorical information such as 'residential' or 'woods'. However, most remote sensing tasks done at the OS would require a smaller-scale analysis ('closer zoom'). Lansdale, Underwood and Davies (2010) used remote sensing surveyors from the OS in a study of visual saliency and expertise in photogrammetry of aerial images. Participants inspected an aerial image stimulus for 12 seconds, then, after a 2 second delay, they searched a changed version for an added target, using a mouse click to identify the target's location. Eye movements were recorded, and experts and novices differed in eye movement patterns. Novices were consistently more drawn to visually salient features in the stimulus images, regardless of their relevance to the task, but experts were able to discount irrelevant but salient features. A small number of other studies have examined categorisation performance and search in aerial images using novice participants and geographers, with some authors recommending that future research include participants who are directly experienced with photogrammetric surveying tasks (e.g., Borders et al., 2020; Lloyd, Hodgson & Stokes, 2002; Pannasch et al., 2014; Rhodes et al., 2021).

1.5 Depth perception

In this thesis, depth perception is primarily studied with binocular disparity and luminance cues. Binocular disparity is a mechanism in binocular vision based on the two eyes receiving inputs of the same scene from slightly different angles. Binocular combination of these disparate angles can lead to a strong impression of 3D shape in the scene. Luminance cues also contribute to impressions of shape, where different light and dark image patterns can be interpreted as having been caused by 3D shapes in surfaces. Binocular disparity and luminance cues will be elaborated on below, and this section continues with a short discussion on other cues to depth perception.

Depth perception involves more than disparity and luminance cues. We use additional monocular cues that do not depend on binocular vision when perceiving the 3D world (Howard &

Rogers, 2002). Some of these cues involve prior assumptions about what 3D-world geometry is most likely to cause the appearance of a 2D image (Marr & Nishihara, 1978). In influential work on 3D shape perception, Pizlo and colleagues argue that 3D shape perception relies on prior assumptions about the nature of shapes in the world (Li & Pizlo, 2011; Li, Pizlo & Steinman, 2009; Pizlo, 2010). Perception of 3D shape is a difficult, 'inverse problem' which must be solved by the brain using 2D inputs from the retina. Inverse problems are difficult as they require inferring a model (3D shape) from data (2D retinal image), and there is always more than one model that can account for the data (Li & Pizlo, 2011). In visual perception, there are theoretically an infinite number of 3D shapes that can produce a given 2D retinal image. Pizlo and colleagues have argued that the visual system relies on prior assumptions of simplicity constraints (aligned with the Occam's razor principle) to limit the number of possible interpretations of 3D shape from the 2D data. 3D shapes can be recovered from a single viewpoint, without provision of exhaustive viewpoints, that is, without seeing all the sides of the object. The simplicity constraints include that: 1) 3D shapes are symmetrical, for example a mug or the human body, 2) the surfaces of 3D shapes are planar between their contours, 3) and 3D shapes are maximally compact and occupy minimal surface area (Li & Pizlo, 2011; Li, Pizlo & Steinman, 2009; Pizlo, 2010). A criticism of Pizlo and colleagues' shape models is that they tend be limited to explaining 'carpented' objects with flat surfaces and 'boxy' shapes.

In Figure 1.2, depth is readily perceived in this photograph, despite it being a 2D image with no involvement from binocular vision (binocular disparity is involved when viewing, but the whole image is presented on a flat page or screen, meaning that binocular disparity only provides a cue to flatness for 2D images). Features in the wall texture are assumed to be of similar size and shape across the image, and the wall itself is assumed as a planar (flat) surface. As the wall texture becomes denser as we look from the right-to-left in the image, the visual system interprets this texture gradient as a depth cue which helps perception of the 3D surface orientation. Seeing that the left-hand part of the wall is farther away than the right-hand part is a 3D reconstruction. Pictorial cues such as linear perspective and texture gradients (Figure 1.2) provide reliable depth perception in normal ground-view scenes. But such cues are often scarce, diminished, or missing in aerial-view scenes. The visual system can readily achieve depth perception in ground-view images without meaningful involvement from binocular disparity. But in aerial images, binocular disparity can provide a particularly valuable 3D cue when commonly available pictorial depth cues are diminished.

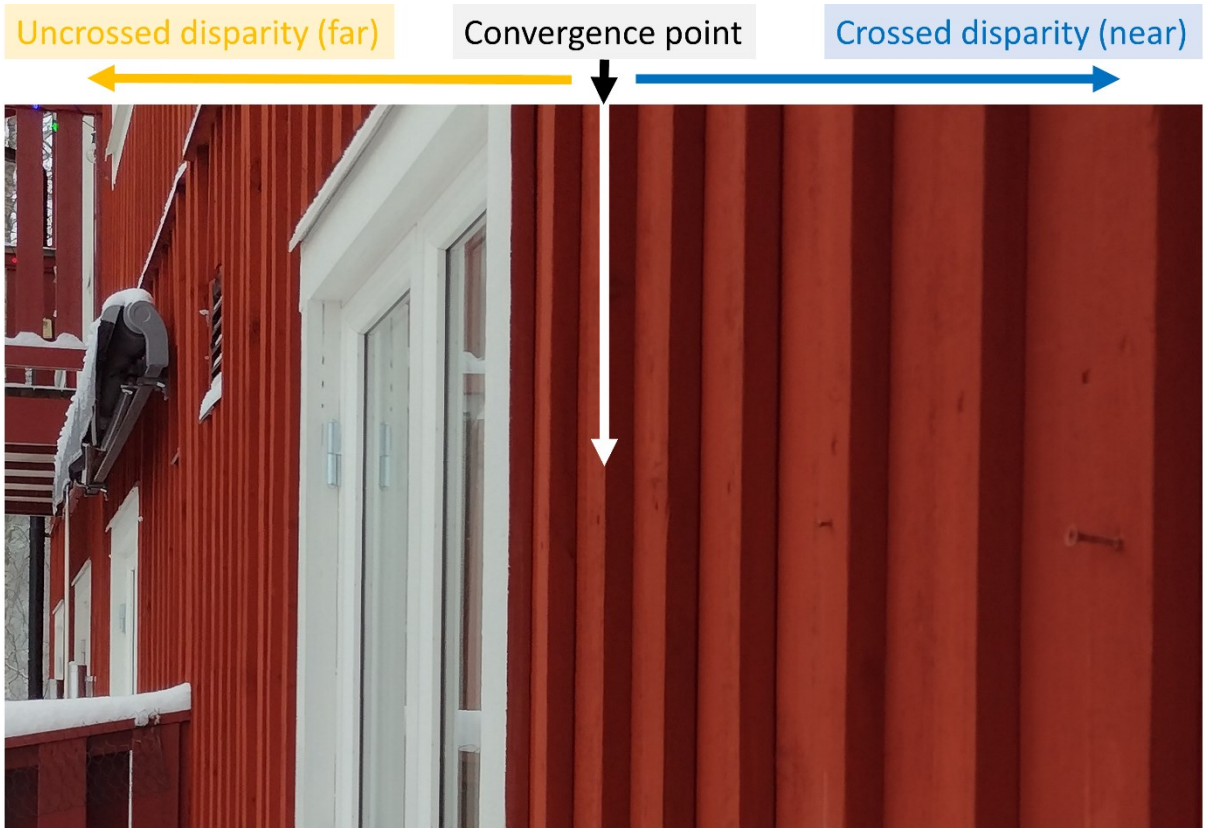

Figure 1.2: A 2D photograph of a slanted surface. Depth is readily perceived in this image with an assumption that the wall is flat and with the perception of a changing texture gradient. The top of the image contains markers that show the contribution of binocular disparity if this image is seen in the real world with binocular viewing. Crossed (near) and uncrossed (far) disparity provides a strong depth cue while viewing such surfaces binocularly.

1.6 Stereoscopic judgements

The view of objects in aerial images often provides less diagnostic information about landscape features compared to the usual ground viewpoint (Lloyd, Hodgson & Stokes, 2002; Loschky et al., 2015). Seeing aerial landscape images stereoscopically, in 3D, affords better depth perception which facilitates landscape feature classification. For example, this can be helpful in measuring changes in ground height, and in discrimination of mounds and holes, or hedges and ditches.

Binocular stereopsis, the perception of binocular disparity cues, is a primary mechanism for depth perception in binocular vision. The visual system can estimate the binocular disparity between the two eyes as they perceive the same scene from slightly different angles. Binocular processing of these disparities affords an estimate of depth. A previous section contains Figure 1.1, an example landscape stereogram: a stereoscopic pair of aerial images of two houses. Notice how the left- and right-hand images in Figure 1.1 are seen from slightly different angles. When fused, these disparities are binocularly combined to provide a strong impression of depth.

Two converging eyes gazing at the same point in space receive the same retinal input along a horizontal ring-like area defined by the eyes and the point of convergence called the horopter

(Howard, 2002). Surfaces deviating from the horopter cause disparities in the binocularly fused image so that the two eyes receive the input at different retinal locations. See Figure 1.3 for an illustration of the horopter including example disparities. If binocular deviations are too large, a surface or item may pass over the limits of binocular fusion and the observer may experience diplopia, or 'double vision'. Processing of disparities is most efficient horizontally, as the two eyes are horizontally arranged next to one another. The geometric horopter is a combination of the larger arc of a circle in the fixation plane and a line perpendicular to this (Howarth, 2011). The empirically estimated horopter is also different, as alluded to in Figure 1.3 (Howard, 2002). If disparity in the binocular image is within and without the horopter, the disparity is 'crossed' and 'uncrossed', respectively (Figure 1.3b). Crossed and uncrossed disparity lead to perceptions that things are near and far, respectively. Our visual system constantly processes binocular disparities in objects and surfaces, significantly contributing to depth perception. Figure 1.2 illustrates how binocular disparity operates to support depth perception of surfaces. Binocular disparity is not a cue to absolute depth, but rather relative depth, as it is a cue to relative distance from the horopter along the visual direction. An example of absolute depth is: 'this object is 80 cm away from me', and an example of relative depth is: 'a point in space is slightly farther away than an adjacent point'. The visual system must know the absolute distance to the horopter to achieve a full depth map from disparity. The common explanation of how estimations of absolute depth can be achieved is through oculo-motor cues to vergence – the angular rotations of the two eyes that allow fixation on a point in space (notice in Figure 1.3, the eyes are rotated inwards to fixate on a point in space). However, this explanation remains controversial, as Linton (2020) suggests that vergence cues might not be available for judgements of absolute depth.

In stereoscopic aerial images, an example of binocular disparity contributing to depth perception is a house appearing tall, standing out in 3D relief above the surrounding ground. The arrangement of aerial photographs taken from slightly different angles leads to tall and deep features having crossed and uncrossed disparities compared to a surrounding point of reference, respectively.

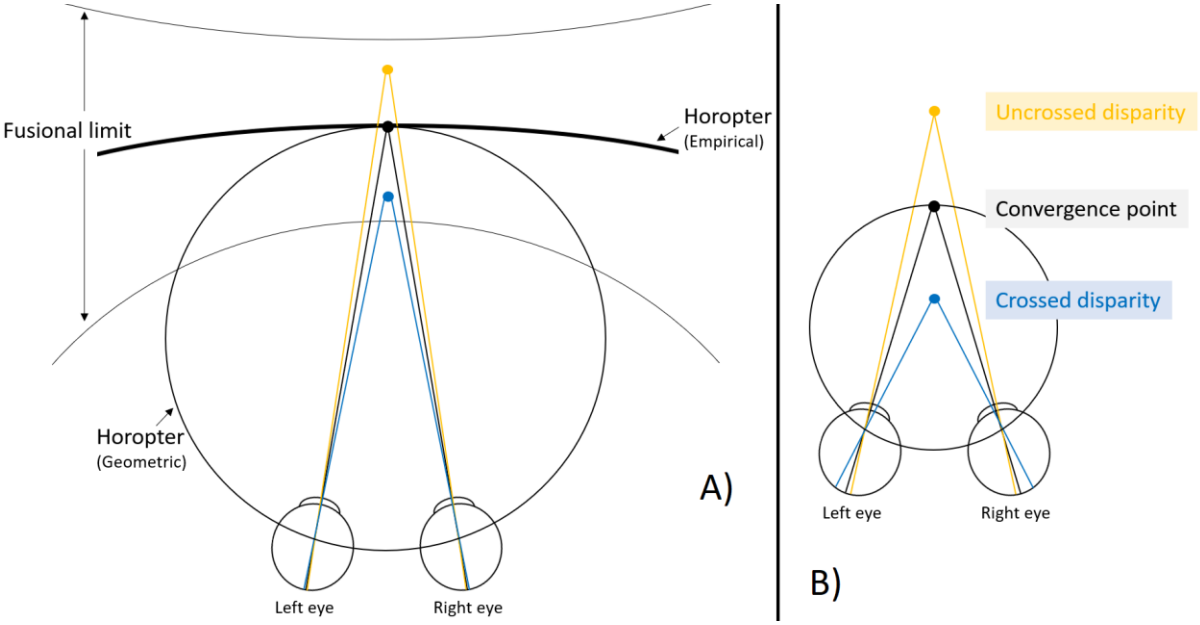

Figure 1.3: An illustration diagram of the horopter. A) The empirical horopter and the theoretical, geometric horopter. Binocular fusion does not occur beyond the fusional limit. B) Exaggerated example disparities for illustration. The convergence point in space is captured on the same retinal location in both eyes. Notice how different disparities cause different retinal locations to be involved.

Julesz (1971) introduced an influential technique for studying binocular stereopsis called random dot stereograms (RDS). RDSs rely on dichoptic fusion of two images, where one image is seen by each eye. In careful consideration of how disparities are produced relative to the point of convergence on the horopter in Figure 1.3, crossed and uncrossed disparities correspond to nasal and temporal displacements, respectively (Blakemore, 1970; Howard, 2002; Julesz, 1971). RDSs exploit this displacement pattern with dichoptic presentation of dot arrays, where dots within an array can be horizontally displaced nasally and temporally to create crossed and uncrossed disparity, respectively. RDSs can create impressions of depth from binocular disparity by displacing dots within the two images that are dichoptically fused. If some selected dots are displaced nasally in both images, they will be interpreted as being crossed (near), as the dots fall on different retinal locations (Figure 1.3). The dots in the images are interpreted as being in a different depth plane due to the involvement of different retinal locations compared to the rest of the fused image (Figure 1.3). Figure 1.4 shows an RDS-like image which contains a small square in uncrossed depth created from the above displacement principles. This image is not a classical RDS image, as it is a dense noise texture rather than a sparse dot array with empty spaces between dots. RDSs provide a way to study depth perception from binocular disparity cues in the absence of any monocular cues to depth. This is a powerful technique used to isolate stereopsis mechanisms and create carefully controlled experimental images. Howard and Rogers review much of the work using RDSs to probe binocular depth perception (Howard, 2002; Howard & Rogers, 2002).

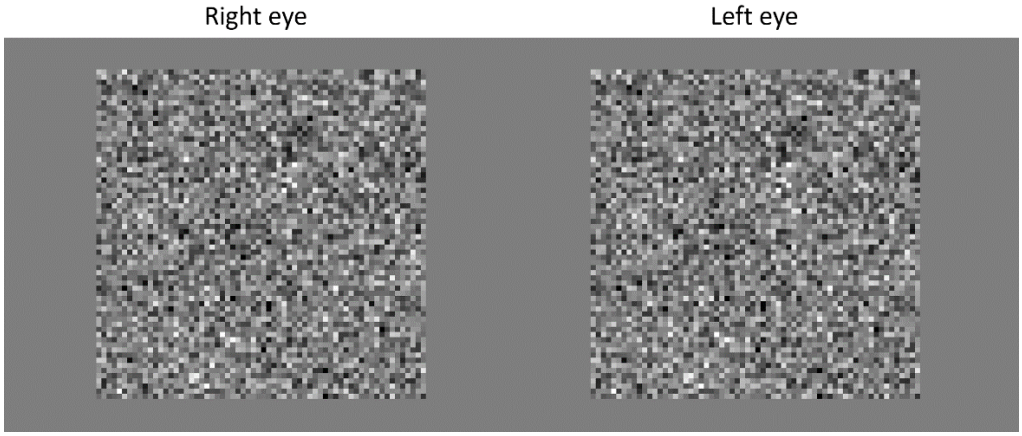

Figure 1.4: Crossed fusion reveals a small central square in uncrossed depth (i.e., a central square of dots hovering behind the surrounding region). Divergent fusion reverses disparities, making the square appear in crossed depth. The image pixels constituting the 3D square have been displaced horizontally by one 'pixel' in both textures.

1.7 Luminance cues to 3D shape

Luminance cues can be used to infer 3D shape. Mechanisms include lightness judgements, and prior assumptions regarding how lighting directions influence shape from shading. This sub-section highlights how luminance cues can support depth perception without involvement from binocular vision.

Structured variations in image intensity can lead to perception of 3D shape. Such variations are commonly light-dark interpreted as highlight-shading caused by 3D surface shape. The visual system may recover 3D shape interpretations from simple images merely containing a contoured luminance gradient, or images with varying contour intensities (Figure 1.5). By rotation alone, images can alternate in perceived relief between convex or concave interpretations (Berbaum, Bever & Chung, 1983; Gibson, 1950), as Figure 1.5 demonstrates (Andrews et al., 2013; Ramachandran, 1988). The visual system can use prior assumptions about the origin of the light source when recovering 3D shape from shading. In Figure 1.5, shape from shading is interpreted with the lighting-from-above prior, where we assume a single global illumination source coming from above; ecologically, the sun or ceiling lights being the source (Adams, Graf & Ernst, 2004; Berbaum, Bever & Chung, 1983; Brewster, 1826; Koenderink et al., 2003; Langer & Bülthoff, 2000; Pont, van Doorn & Koenderink, 2017; Ramachandran, 1988; Rittenhouse, 1786; Schofield, Rock & Georgeson, 2011; Sun & Perona, 1998). Ramachandran (1988) introduced the type of stimuli seen to the left in Figure 1.5, finding that multiple images with such ambiguities presented next to one another are interpreted with one assumed global illumination source. Sun and Perona (1998) showed that the lighting-from-above prior is not located exactly above on average, but slightly tilted to the left. Mamassian and Goutcher (2001) later replicated this. Langer and Bülthoff (2000) tested differences in convex-concave judgements under diffuse lighting ('cloudy day') and punctate lighting ('sunny day')

conditions. The punctate lighting condition, using collimated beams of light, was further divided into source lighting from above-left, above-right, below-left, and below-right. The authors found that observers made systematic errors in convex-concave judgements when lighting was from below. This suggests that the lighting-from-above prior can act as a dominant cue when resolving convex-concave ambiguities. This tendency for '180° flipped errors' congruent with lighting-from-above was also found by Koenderink et al. (2003) and Pont, van Doorn and Koenderink (2017) when shading cues dominated shape judgements in complex textures. As convex-concave judgements can be strongly influenced by lighting directions, a later study used such 180° flips with hedges (convex) and ditches (concave) to measure how expert surveyors interpret lighting direction cues. For a further elaboration on this topic, see Chapter 4.

Adams, Graf and Ernst (2004) showed that the lighting-from-above prior is malleable with experience in humans. The authors used adaptation sessions with multisensory haptic-visual feedback, where contoured-gradient visual stimuli like those in the left of Figure 1.5 were paired with haptic feedback ('feels convex or concave') that was shifted + or - 30°. Participants adapted to shift their lighting direction biases slightly based on the shift introduced by the haptic feedback, an effect which translated to a visual-only baseline task. In this seminal study, Adams, Graf and Ernst (2004) showed that human lighting direction biases are malleable and may be modified by new scene statistics. They also noted that their effect on the participants would likely not survive in the long term due to the natural environment resetting the baseline prior. Hershberger (1970) found that chickens raised in cages illuminated from below still infer shape from shading in a manner congruent with lighting-from-above, suggesting an innately encoded prior unaltered by experience in chickens.

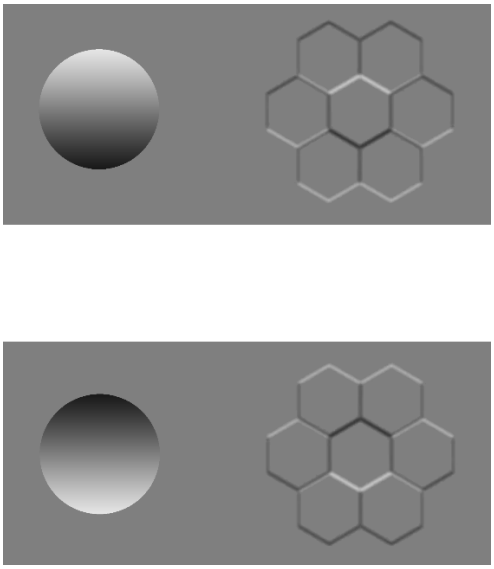

Figure 1.5: Images providing an interpretation of 3D shape. The lower row contains the same images as the upper row, but rotated 180°. Top-left circular luminance gradient is typically interpreted as convex ('bump') and bottom-left as concave ('dimple'). The reader may squint their eyes to aid 3D recovery by blur (Ramachandran, 1988). Right-hand hexagonal lattices ('honeycombs') with varying contour intensities lead to an interpretation

of highlighted and shaded contours (Andrews et al., 2013). The central hexagon in the top-right and bottom-right honeycomb are typically seen as a convex bump and a concave dimple, respectively.

With the diffuse lighting assumption, dark image regions appear deep, and light image regions appear in relief, or 'tall' (Chen & Tyler, 2015; Langer & Zucker, 1994; Langer and Bülthoff, 2000, Schofield, Rock & Georgeson, 2011; Sun & Schofield, 2012; Wright & Ledgeway, 2004). This can be an effective strategy for recovering shape from shading. Langer and Bülthoff (2000) found that shape from shading was accurately recovered during diffuse lighting ('cloudy day'). Judgements of brightness correlated with judgements of depth, indicating that a 'dark means deep' strategy is employed during diffuse lighting conditions (see also Langer & Zucker, 1994). An ecological explanation for this effect can be found in natural scene statistics, where the top of a 'hill' might receive illumination from many sources in the scene, but the bottom of a 'valley' might only receive diminished illumination due to surface depth occlusion (Langer & Zucker, 1994; Potetz & Lee, 2003). Schofield, Rock and Georgeson (2011) highlighted the problem that human vision needs to resolve shape from shading in natural environments that provide an ambiguous mixture of punctate light from a single-point light source, and diffuse light scattered from the surrounding surfaces. The authors found that observers have a strong tendency to expect diffuse lighting, in combination with a bias for lighting-from-above.

The diffuse lighting assumption is commonly exploited in painting, where light and dark can accentuate perceived depth. Leonardo da Vinci observed this phenomenon; "among bodies equal in size and distance, that which shines the more brightly seems to the eye nearer" (MacCurdy, 1938, p. 332). Another assumption is the convexity bias, which is a tendency to perceive an ambiguous relief as convex ('tall rather than deep') (Adams & Elder, 2014; Champion & Adams, 2007; Langer & Bülthoff, 2001; Perrett & Harries, 1988). Contrast differences between target and background are yet another monocular cue in depth perception, where both light and dark targets can appear closer depending on their relationship to the background contrast (Egusa, 1983; O'Shea, Blackburn & Ono, 1994).

Remote sensing surveyors at the OS have reported that luminance cues and cast shadows are important cues for depth perception in aerial images. Cast shadows in photographs on sunny days can be used to infer height differences between tall objects through shadow length. While measuring cast shadows might reflect a consciously available strategy, shape inferences from luminance cues such as priors are perceptual, and not always consciously available. Chapter 4 describes an experiment where the task was to discriminate between aerial images of hedges and ditches. In this experiment, luminance cues played a significant role in the participants' judgements, and the use of luminance cues varied on the individual level. Data from some participants suggest

that they employed a diffuse lighting rule ('dark-is-deep') for hedge-ditch discriminations, and data from others suggest a punctate lighting rule ('sunlight strikes from a specific orientation').

1.8 Cue combination of disparity and luminance

Multiple cues supporting depth perception tend to combine. Binocular disparity and luminance cues tend to coincide in the real world in structured ways. For example, tall 'hills' and deep 'valleys' are encoded with crossed and uncrossed disparity and tend to be lighter and darker, respectively. The combination of binocular disparity and luminance cues sometimes results in conflicts which the visual system must then resolve. Luminance cues of highlights and shading are also inherently ambiguous in many situations. Resolution of convex-concave ambiguities is often handled via biases in the system, such as the lighting-from-above bias, the diffuse lighting ('dark-is-deep') assumption, or the convexity bias, affording quick perceptual heuristics for interpreting 3D relief. Disparity and luminance cues typically interact in harmony in the real world, supporting veridical interpretations of 3D surfaces.

In studies of cue combination, experimenters may create conditions where cues are congruent and incongruent with one another, to measure how the cues combine. Doorschot, Kappers and Koenderink (2001) used a surface normal setting task and found that luminance and binocular disparity cues combined almost linearly (see also Landy et al., 1995). Lovell, Bloj and Harris (2012) independently varied disparity and luminance cues to examine how each cue contributes to depth judgements. The authors found that disparity cues were used more, but as disparity cues were manipulated to become less reliable, observers tended to shift to luminance cues. Hartle et al. (2022) studied convex-concave discriminations with stimuli varying in 3D relief from binocular disparity and shading intensity. When the shading cue was weak, binocular disparity was critical for reliable discrimination. But when the shading cue was strong, participants tended to show a convexity bias even when binocular disparity indicated concavities. Strong shading cues can thus dominate, and make observers ignore disparity cues (see also Chen & Tyler, 2015). Hartle et al. (2022) also noted individual differences in how the convexity bias influenced relief discriminations. Chen and Tyler (2015) showed that luminance cues can make disparity cues redundant in an experiment using sinusoidally corrugated luminance gratings. Despite a strong, competing, binocular disparity cue, surface relief was interpreted primarily based on the luminance cue, consistent with the diffuse lighting assumption where 'dark-is-deep'. Samonds, Potetz and Lee (2012) studied simultaneous excitation from luminance and disparity cues in neurons tuned to both cue modalities in macaque primary visual cortex. The authors found neurons that responded more to crossed and uncrossed disparities, that also responded relatively better to lighter and darker stimuli, respectively. Other

studies have also tested integration of disparity and texture cues to 3D shape, consistently showing an additive benefit of cue combination (Hillis et al., 2002, 2004; Johnston, Cumming & Parker, 1993; Knill & Saunders, 2003; Vuong, Domini & Caudek, 2006; see also Meese & Holmes, 2004 for a study on combining different pictorial cues).

1.9 Perceptual learning

Another topic of interest for this thesis is the mechanisms involved in how remote sensing surveyors learn and improve to better perform their tasks. Visual expertise does not develop at the moment of initial conceptual awareness, meaning that a brief lesson on aerial images is not sufficient to develop expertise for aerial images. Developing visual expertise requires long-term, experience-dependent learning and tuning of the visual system to enhanced processing of domain-specific images. Perceptual learning (PL) could account for some aspects of expertise, for example by enhancing certain visual processes (e.g., the ability to process binocular disparity) that are important for processing domain-specific images (elaborated below).

PL has often been studied for simpler visual processes using training on simple stimuli, such as discrimination of spatial frequencies and orientations (e.g., Fiorentini & Berardi, 1981), or hyperacuity in Vernier line discrimination (e.g., Poggio, Fahle & Edelman, 1992). The typical finding in such studies is that PL improves performance for the trained stimuli, but does not translate when relatively small changes are introduced, such as a 90° rotation. With stimuli that are highly homogeneous, the PL can also be expected to have limited transfer. Visual expertise, however, appears more general, superseding such highly specific training effects. Training to improve vision more generally might help the development of expertise, or benefit recovery for individuals with reduced vision, such as in amblyopia (Levi, Knill & Bavelier, 2015; Levi & Li, 2009). Researchers have suggested that PL can lead to more general visual learning (e.g., acuity or stereoacuity) in neurotypical or neurodivergent adults with sufficient training on a sufficiently diverse set of stimuli (Deveau, Ozer & Seitz, 2014; Godinez et al., 2021; Green & Bavelier, 2015; Levi, 2022; Levi, Knill & Bavelier, 2015; Levi & Li, 2009; Portela-Camino et al., 2018; Seitz, 2020; Vedamurthy et al., 2016). To support generalisable learning, stimuli might need to vary in image configurations such as surface orientation, spatial frequency content, and weighted combination with other supporting cues (e.g., other visual, haptic, or auditory cues), among other aspects.

PL is also consolidated by sleep (e.g., Karni et al., 1994; Karni & Sagi, 1993), and promoted by trial-by-trial performance feedback which can help observers to home in on diagnostic cues (e.g., Herzog & Fahle, 1997; Liu, Lu & Dosher, 2010). The discussion on PL will be elaborated further in Chapter 5, which describes a study that focuses on stereoscopic PL.

PL is a likely mechanism in developing visual expertise, thus PL interventions could potentially augment the development of expertise. In theories on feedforward hierarchical processing in vision, higher processes in the visual system, such as object recognition, rely on the output of lower visual processes (Serre, Oliva & Poggio, 2007). In such a hierarchical structure, the higher processes benefit if the supporting lower processes are enhanced through PL (Dosher & Lu, 1999; Sagi, 2011). Hypothesising along this line, remote sensing surveyors who are experts at stereoscopic photogrammetry might show enhanced processing of binocular disparity cues in landscape images, among other expertise factors. Such more basic mechanisms might support the broader expertise. From a research standpoint, studying visual expertise can be difficult, as it emerges from a natural environment that affords a relatively poor understanding of how isolated stimuli and tasks contribute to expertise development. Researchers will therefore commonly apply a reductionistic approach by limiting their scopes to singular expertise factors, such as evidence of PL for specific cues which are thought to contribute to the broader expertise.

1.10 Selecting a primary method

Remote sensing surveyors are tasked with creating and updating map data using stereoscopic aerial images. Surveyors rely on depth perception from binocular disparity and various luminance cues for processing of images seen from the aerial viewpoint. With experience, surveyors develop expertise in visual tasks of detection, discrimination, categorisation, and search in images that are unfamiliar to the average human observer.

A primary aim of this thesis was to uncover expertise involved in depth perception of stereoscopic aerial images. The work presented in later chapters thus required experimental methods which could uncover sampling of visual cues from images that had both 2D and 3D components. An example of a 2D component of an aerial landscape image is luminance cues to shape, and a 3D example is height/depth judgements based on binocular disparity. A psychophysical technique called classification images (CI) affords the possibility of measuring the visual cues that are sampled from stimulus images during behavioural tasks. This technique thus seemed well suited for satisfying a main aim of this thesis to uncover visual strategies for stereoscopic aerial images. The term CIs encompass a family of related methods, of which ‘reverse correlation’ is the method used in this thesis. Reverse correlation uses noise textures that provide a random modulation to a target, and for the sake of this thesis, the technique will be referred to as ‘classification images / CIs’. CIs have primarily been used with 2D luminance images, but could be extended with the addition of an RDS-like manipulation into a novel version that could simultaneously estimate 2D and 3D CIs. This novel version of CIs could be suited for studying how experts and novices might differ when using

disparity and luminance cues in visual tasks using stereoscopic aerial images. Such a method was developed for this thesis project, and its development is described in Chapter 3.

As a different potential method for this thesis, eye-tracking provides a measure of how images are foveated, correlating with how image areas are attended. Eye-tracking is particularly useful for studies on visual search. But for this project, information sampling mechanisms (including for 3D image aspects) were of primary interest. Eye-tracking seemed unsuitable as it is not directly able to measure the use of binocular disparity as a visual cue. Eye-tracking also does not provide a measure of how visual cues are used once gaze is directed to their location. For this reason, CIs were used as the primary method in this thesis. The CI technique is central to this thesis, and the following section will present a summary of the technique and how it has been used in vision research.

1.11 Classification images and the perceptual template

The CI technique is used to analyse perceptual strategies (Abbey & Eckstein, 2002; Ahumada, 1996; Beard & Ahumada, 1998; Murray, 2011). CIs have been used for a diverse range of investigations into topics in visual perception. CIs are typically gained through the addition of random noise textures to target images in visual tasks of detection, discrimination, or categorisation (Ahumada, 1996; Murray, 2011). Noise textures have a masking effect on targets, making detection more difficult. But they also have a modulating effect, changing the appearance of targets. Noise can sometimes promote and sometimes demote detection of targets.

Take the example of a small, low contrast, white square target overlaid with high contrast visual noise covering a larger region than the white square itself. If your task is to detect the white square, which is present on 50% of all trials and always located in the middle of the image, you must attempt to detect this low-contrast signal despite the high-contrast noise mask making detection difficult. Most of the variation in the stimulus images is due to random noise. As you decide whether the target was present or not on each trial, you might rely on a strategy of contrast judgement centred on the target location. If the target location appears ‘whiter’ with sufficient intensity, your response might be ‘yes – there was a white square present’, or if it appears darker, you might be inclined to respond ‘no – no white square’. This strategy attempts to detect the expected contrast change in the image facilitated by the target acting as a pedestal (Legge & Foley, 1980; Georgeson, Yates & Schofield, 2008; Meese, Georgeson & Baker, 2006; Murray, Bennett & Sekuler, 2002). Independently of the target, lighter patterns in the noise textures promote, and darker patterns demote detection, thus influencing your responses accordingly. CIs are generated from these types of noise patterns. By tagging noise textures with the participant’s response on every trial, we can sum up all noise textures that yield different responses to construct a CI.

CI studies tend to assume that the participants' stimulus-response relationship is approximately linear (Abbey & Eckstein, 2002, 2006; Murray, Bennett & Sekuler, 2002, 2005; Tjan & Nandy, 2006). In the above example of detecting the white square, an observer would resemble a linear observer with the above response pattern, where lighter and darker patterns increase probability of a positive and negative response, respectively.

Noise textures yielding different responses are typically grouped and summed according to four stimulus-response categories as follows: *hits* (signal present and 'present' response), *false alarms* (signal absent and 'present' response), *correct rejections* (signal absent and 'absent' response), and *misses* (signal present and 'absent' response) (Ahumada, 1996; Beard & Ahumada, 1998; Murray, 2011). Importantly, only the noise textures are used in the analysis, the signal is not included. Noise textures that yielded negative responses (*correct rejections* and *misses*) are typically subtracted from those that yielded positive responses (*hits* and *false alarms*) to form a CI (Equation 1[1]):

$$CI = \left(\sum Hits_{ij} + \sum False\ alarms_{ij}\right) - \left(\sum Correct\ rejections_{ij} + \sum Misses_{ij}\right). \qquad \text{Equation 1.}$$

The *false alarm* trials are illustrative for understanding CIs as, in these trials, the random noise texture has mimicked the target to yield a positive response but there was no target present (Abbey & Eckstein, 2002; Eckstein, Pham & Shimozaki, 2004; Gold, Sekuler & Bennett, 2004; Murray, Bennett & Sekuler, 2002).

A necessary condition for CIs to be captured is that the noise textures modulate the appearance of the target image and thus influence responses. If participants can detect the target regardless of any random configurations in the visual noise, the noise cannot influence the task, and no CIs will be captured. Experimenters who are preparing a CI study thus need to make sure that the signal-to-noise ratio (SNR) is set to a suitable level. Methods include 'manual staircasing' by testing different SNRs to approximate a desired detection threshold (for example 70% correct responses), or adaptive staircasing SNR to a fixed threshold throughout the experiment.

CIs can reveal the perceptual strategies, or 'templates' of the observers. Templates reveal the visual cues that observers use when performing a task. In studying the use of visual cues, this method can avoid potential biases from the experimenter when selecting stimuli that might occur with other designs. The technique is thus also referred to as a 'reverse correlation' approach, as the

[1] A CI is a 2D matrix where each element represents a grayscale value, and this matrix is commonly presented such that each element occupies multiple pixels on the physical monitor (e.g., 5x5 monitor pixels).

determination of what is diagnostic of the target is inferred from the observer's CI template rather than by a pre-determined correct response decided by the experimenter.

The CIs approach was originally borrowed from the study of electronic systems, control theory, and computer science; fields that have required studying the behaviours of 'black boxes'. A black box refers to a system with unknown internal processes that transforms inputs to outputs. Murray (2011) provided an influential review of CIs, where he describes the process of kernel analysis (Lee & Schetzen, 1965; Volterra, 1930; Wiener, 1958). A system, such as a retinal ganglion cell, can have an intricate internal relationship between inputs and outputs. Volterra (1930) and Weiner (1958) showed that, under certain conditions, such a system's kernels can be approximated. Kernels bear similarity to physiological receptive fields or psychophysical spatial filters (Neri & Levi, 2006; Ringach & Shapely, 2004). For example, a retinal ganglion cell might have a kernel/receptive field that resembles a 'Mexican hat' function (e.g., Difference of Gaussians or Gabor function), with a positive center and negative side-lobes. This is an example kernel with a spatial weighting function. Kernels can be characterised by noise inputs that feed through the kernel to produce a response. Different patterns in noise produce different responses, and with enough repetitions, certain noise patterns can systematically trigger certain responses (Lee & Schetzen, 1965; Murray, 2011; Neri & Levi, 2006; Ringach & Shapely, 2004). In analysing these noise patterns that yield different responses, kernels can be characterised. Kernel estimation – measuring outputs based on noise inputs – bears close analogy to CIs in visual perception that reveal perceptual templates from noise inputs (Neri & Levi, 2006; Ringach & Shapely, 2004). A black box system can be a single neuron transforming an input to an output, an artificial neural network classifying objects in images, or an observer viewing stimulus images on a screen and pressing buttons for responses (Pelli, 1990). CI approaches seek to estimate the behaviours of black boxes, to reveal the inner workings of the perceptual templates that generate response outputs from random, known, noise inputs. Note however, that physiological receptive fields from single neurons are not always sufficient to explain CIs estimated from an observer performing a behavioural task. An observer might combine the output of many receptive fields to build an overall detection mechanism, and CIs characterise this detection mechanism from button-press responses in behavioral tasks. For example, an observer tasked with detecting a white square might not necessarily possess and activate a single-neuron receptive field that resembles the white square. Rather, the visual system may combine the outputs of many receptive fields in the application of a perceptual template that fits the target.

1.12 Templates in neurophysiology

Analysing the behaviours of black boxes has immediate application to biological systems. CIs as a technique for kernel analysis inspired application to the study of cell communication and receptive field structure in neurophysiology. In a study of auditory neurons, de Boer and Kuyper (1968) analysed input-output cross-correlations using noise inputs, which characterised the linear filter of their current neuron model. Marmarelis and Naka (1972) studied a three-neuron chain in a catfish retina by injection of a random noise current. A horizontal cell stimulated a bipolar cell, which in turn stimulated a ganglion cell, and the ganglion cell's spike-triggered output was recorded. The authors were able to use noise to make predictions about the dynamic behaviour of the neuron chain (Marmarelis & Naka, 1972; Murray, 2011; Wiener, 1958). Application of noise for the study of neurophysiological systems has its own history and body of work which will not be covered in further detail in this thesis. See Marmarelis and Marmarelis (1978), and Pinter and Nabet (2018) for books on the topic, and Sakai (1992), and Wu, David and Gallant (2006) for review articles. See also Neri and Levi (2006), and Ringach and Shapely (2004) for discussions on the similarities between templates in neurophysiology and psychophysics.

1.13 Visual psychophysics with classification images

Ahumada and Beard first developed and popularised the CI technique (Equation 1) for use in visual psychophysics (Ahumada, 1996; Beard & Ahumada, 1997; Beard & Ahumada, 1998) many years after closely related methods had been used in auditory psychophysics (Ahumada & Lovell, 1971) and neurophysiology (Marmarelis & Naka, 1972). In visual psychophysics, the first CI studies showed extraction of relevant image features. Ahumada (1996) found differences in visual strategies between human and ideal observers for a Vernier acuity task. An ideal observer is a simulated observer that performs a given task as well as possible given the information in the stimulus. In a detection task, an ideal observer would tell of the optimal strategy for detection, and the ideal template is a perfect match to the target (Ahumada, 1996; Beard & Ahumada, 1998; Abbey & Eckstein, 2006). Watson and Rosenholtz (1997) showed that when observers discriminated a 'c' from an 'x', templates from 'c' responses to 'x' stimuli contained a 'c' with an inhibited 'x', and vice versa. Beard and Ahumada (1998) found templates revealing Gabor-like spatial filters involved in making 'aligned' and 'offset' judgements in a Vernier acuity task. Gold et al. (2000) used CIs to study illusory contour effects in Kanizsa squares, finding perceptual completion effects in human observers who 'filled in' illusory contours in Kanizsa squares. An ideal observer did not rely on illusory contour information, as no bottom-up signal is present there.

Building on the novel approach by Gold et al. (2000), Gosselin and Schyns (2003) further tested the idea that CIs reveal properties of internal representations in absence of *any* bottom-up signal by studying what they called 'superstitious perceptions'. The authors told their participants that the letter *S* was hidden in white noise images, but no signal was ever present. CIs, after 20,000 trials, showed a significant *S* shape that looked slightly different for each of the three participants. The authors correlated a smoothed version of these CIs with different fonts and found that the *S* for each participant correlated strongest with a different font. As no signal was ever presented, the authors concluded that the resulting CIs have revealed an approximation of the internal representations of the letter for each participant. Gosselin and Schyns (2003) also tested face stimuli, where the contours of a face were presented without the mouth area. When asked to detect a smiling mouth, CIs contained templates that resembled a smiling mouth. For templates that contain an *S*, a mouth, or any other more complex shapes, CIs reveal templates that are constructed out of multiple receptive fields. These fields combine to produce a perceptual template, which relates to the overall detection mechanism of the observer.

In Gosselin and Schyns (2003), participants showed large differences in response biases. Response bias, the tendency to give one response more often than another, can be avoided with a two-alternative forced-choice (2AFC) task, where participants are forced to select that the target was present in one out of two images. CI studies typically use either a 2AFC or a single-interval binary-response (SIBR) task where one image is presented, and a yes/no response (or similar) is given. This latter design can be faster to perform, as it only requires half the images. SIBR also affords a less constrained experience for the participant, as they are freely able to select yes/no for each stimulus image. An inherent problem with the SIBR design compared to the 2AFC design is that response biases can occur. In Gosselin and Schyns' (2003) second experiment the two participants responded 'yes' in 48% and 7% of the trials, respectively. CI templates were still reliably found for both observers, despite their very different response biases, lending support for the SIBR design. It might seem a surprising result that two observers with such differing response criteria should both produce robust CIs. The authors note that the more conservative observer had given a positive response only when they were certain to have seen the face as smiling, suggesting that their positive-response template reflected high-confidence responses. The other observer might have been more focused on equally distributing 'yes' and 'no' responses, as the observers were told that the smile was present in 50% of the trials. It is likely that less irrelevant noise was added to the positive response category due to the conservative observer's more strict response criterion. The topic of response biases and CIs is further explored in Chapter 3: Pilot 1.

We may consider the idea of a 'purely internal' template, for example an imagined shape or object, without any bottom-up influence prior or during the experiment. This template approximates the observer's top-down expectation of what the signal should look like, and can be captured with CIs (e.g., Gosselin & Schyns, 2003). The advantage of using no signal is that the stimuli do not constrain or lead the observer's expectation of what to 'see' when performing the task. This design is less advantageous when experimenters desire more control and constraint in their experiment, beyond the observers' expectations about the stimulus. Another disadvantage relates to the absence of any performance measures as there are no ground-truth signals.

Gosselin, Bacon and Mamassian (2004) extended this 'superstitious' or 'illusory' CI technique to 3D stimuli. CIs of a complex 3D pattern were be found when observers were tasked to look for a large '+' sign in RDS with no signal present. RDS are dichoptically presented images of a dot array where dots can be displaced to create binocular disparity (Julesz, 1971). In an earlier CI study using RDSs, Neri, Parker and Blakemore (1999) found that observers' 'yes' and 'no' responses in a detection task were supported by dots that contained the same, and opposite, sign of disparity as the target, respectively.

CIs have also been used to investigate PL (Dobres & Seitz, 2010; Gold, Sekuler & Bennett, 2004; Kuai, Levi & Kourtzi, 2013; Kurki & Eckstein, 2014; Li, Levi & Klein, 2004). A typical finding in such studies is that as PL progresses, participants are better able to make use of more relevant stimulus information – 'sampling efficiency' increases, as revealed by CIs. Gold, Sekuler and Bennett (2004) studied PL for discrimination of two faces and abstract textures. CIs showed that observers' templates had higher amplitude and greater spatial extent in the latter half of the experiment compared to the first half. The authors paired noise masking with the technique of double-pass response consistency, where the same stimulus is shown twice during a session to estimate internal noise from response consistency. Gold, Sekuler and Bennett (2004) found evidence that the PL improvement comes from increases in the ability to sample relevant visual cues rather than a reduction in internal noise. Dobres and Seitz (2010) found higher contrast CI templates post-learning compared to pre-learning in a study on PL of oriented gratings. Apart from these template improvements, accuracy measures also improved, indicating learning.

Other research groups have contributed with methodological advancements to CIs. Murray, Bennet and Sekuler (2002) discussed a commonality in CI studies that incorrect trials (*false alarms* and *misses*) tend to carry more of the CI template per trial than correct trials (*hits* and *correct rejections*). Equation 1 weights all stimulus-response categories equally, but the authors argue that optimal template estimation should consider SNR (if it varies throughout the experiment) and whether a response was correct or incorrect. For example, if SNR varies, incorrect trials where the

target is high in contrast likely contain more of a misleading noise pattern than incorrect trials where the target is low in contrast, as noise added to high contrast targets must provide a greater amount of misleading information to induce an incorrect response. In a later study, the same authors showed that all aspects of a linear observer's strategy for simple stimuli are reflected in a CI (Murray, Bennett & Sekuler, 2005). Observer performance can thus be predicted under limited circumstances. Tjan and Nandy (2006) emphasized the problem that if an observer is uncertain about the location of the signal, CI templates at different locations can cancel each other out, resulting in no measurable template (see also Beard & Ahumada, 1999; Murray, Bennett & Sekuler, 2002). The authors show benefits of using higher contrast signals to 'clamp' a template to a specific location. Certainty about the target location is typical in CI studies, but Abbey and Eckstein (2014, 2021) tasked observers to search larger images for small targets with unknown locations using a mouse cursor for localization responses. CIs were generated by saving the parts of the noise images that the observers clicked on. This alternative design merges the CI technique with visual search. Other studies with a similar approach have also combined eye-tracking with CIs in search tasks, studying visual saliency in search tasks (Kienzle et al., 2009) and finding differences in foveal and non-foveal processes where peripheral features resembling targets attract gaze (Rajashekar, Bovik & Cormack, 2006; Tavassoli et al., 2007).

1.14 Classification image analysis

As CIs come in the form of images, they readily afford direct presentation to the reader, as templates are often visible and convincing at first glance. Visual inspection of the raw CI is thus an essential part of modern CI research. While such casual inspection is valuable, it is also important to provide means of statistical testing for CIs. Chauvin et al. (2005) argued that available statistical tests for CIs were underdeveloped in the early years after Ahumada (1996). Studies had sometimes solely based their CI analysis on visual inspection by the reader (Abbey & Eckstein, 2002), increasing the risk of false positives and a bit of 'wishful thinking' from experimenters not relying on any statistical measures. A CI is a matrix containing measurable results in the form of pixel intensity values. An obvious starting point for analysing significant pixels in the image is by deciding on a luminance intensity threshold which filters out pixels that do not reach the threshold. This pixel-wise analysis was used on CIs in some early papers. Beard and Ahumada (1998) and Watson and Rosenholtz (1997) zeroed all pixels that were less than two standard deviations from the mean in their CIs. These threshold filters brought out the templates while zeroing much of the noise from non-template regions in the CIs. CI templates can thus be captured through a pixel-wise analysis procedure, but we may question the assumption that templates operate on a pixel-by-pixel level. Returning to the example task of

detecting a white square, the observer's template would not likely be operating on a pixel-by-pixel level. Receptive fields sum over the image space (over many pixels), which is exploited by the overall perceptual template when judging images. Many neighbouring pixels are pooled when determining if a region is considered light or dark. Thus, an analysis focusing on correlated neighbouring pixels in clusters could efficiently capture lower amplitude but significant template regions. Beard and Ahumada (1998) and Watson and Rosenholtz (1997) implicitly included the idea that neighbouring pixels were correlated by smoothing their CIs with a Gaussian filter, but they explicitly conducted a pixel-wise analysis. Smoothing attenuates high-spatial frequency components, which can help to bring out a low-spatial frequency template from white noise textures. By attenuating high-spatial frequency components, the authors assumed that their participants had applied templates that summed over a larger space, with several correlated neighbours.

The Kolmogorov-Smirnov test (Sheskin, 2020) can compare sample areas of CIs with a normal distribution to examine deviations from normal. A normal distribution in an image would be centred at mean luminance and contain an equal distribution of light and dark pixels. The test, used on CIs by Kontsevich and Tyler (2004), can detect templates that are dominated by a specific sign (such as a skewed distribution of mainly black pixels). The Kolmogorov-Smirnov test treats pixels in image regions as belonging to the same distribution, which can be useful for detecting templates that sum over multiple pixels.

Statistical testing of individual pixels can also prove challenging. Chauvin et al. (2005) argued that Bonferroni corrections to *t*-tests can become far too conservative during analyses of CIs, as a 256x256 pixel image contains 65,536 independent data points. A Bonferroni correction applied to this number of data points would filter out much of the template due to an extremely conservative estimation of significance. Even a dimensionally reduced image, e.g., 64x64 pixel textures, still have 4,096 independent data points. Chauvin et al. (2005) instead showed that random field theory (RFT) can be applied to CI analysis, drawing parallels to how RFT is used in brain imaging analysis. Data from a brain imaging technique such as fMRI share similarities to CI data, as voxel and pixel space are both noisy, producing random fluctuations which commonly reach statistical significance, even after smoothing. Chauvin et al. (2005) applied RFT to smoothed CIs to analyse the probability that a cluster of pixels should exceed a threshold compared to chance. This analysis considers correlations between neighbouring pixels, going beyond pixel-by-pixel analyses.

CI experiments using simple stimuli tend to only require simple detection mechanisms. Templates from simple detection tasks can be described through modelling with common functions. A task such as detection of a central white gaussian bump in noise might generate a CI template containing a positive white bump with surrounding dark negative lobes (Abbey & Eckstein, 2002,

2006). Negative surrounds are commonly involved in contrast judgements and are often captured in CI studies. Cell receptive fields and spatial filters in early vision commonly have excitatory centres with inhibitory surrounds of different configurations (Barlow, 1953; Kuffler, 1953; Marr & Hildreth, 1980). Inhibitory surrounds enhance the contrast to the excitatory centre. For example, a white bump in noise will appear lighter if the noise adds darker pixels surrounding it, or a stereoscopically tall feature will appear taller with a deeper surround. Such filters are efficient at signalling sudden changes in contrast, such as in detection of edges and lines. Abbey and Eckstein (2002, 2006) showed that CI data for circularly symmetric templates can be reduced to a radial average cross-section, where the amplitude of the template decreases with increasing distance from the template centre. Abbey and Eckstein used this technique to measure template structure, including negative surround mechanisms, in different types of gaussian bump stimuli. Simplistic templates can resemble simple functions, such as Gaussian, Difference of Gaussians, or Gabor functions, etc. With cross-sectioned CIs, we can fit various functions to the cross-sections to explore how best to describe the data, and to examine the function parameters that provide the best fit. In our example of a template containing a central white bump with negative lobes, a Difference of Gaussians or Gabor function might accurately describe the template profile. By parameterising CI data, functions can describe e.g., the amplitude, spread, and peak location of the template. Cross-sections and function fitting provides reductionistic results about template structure which can be desired when rigorously examining subtle differences across conditions.

In a study on face expressiveness, Skog et al. (2023) found CI templates for the detection of 'surprise' expressiveness using neutral faces. The CIs appeared to contain a high-spatial frequency template of eyes (leftmost image in Figure 1.6). In analysis, the Kolmogorov-Smirnov test (used by Kontsevich & Tyler, 2004 in a similar face study) did not capture the template, likely because these templates were defined by both white and black pixels, and thus did not significantly deviate from a normal distribution. As CI templates appeared visible to human observers, Skog et al. (2023) devised an analysis method based on a filter bank including Gabor and isotropic filters with spatial frequency and orientation bands that resemble channels in human vision (Campbell and Robson, 1968; Blakemore & Campbell, 1969; Sachs, Nachmias & Robson, 1971). These decomposed CIs were rectified to produce energy maps that could highlight significant regions. See Figure 1.6 for an example part of the analysis procedure. Energy maps were analysed pixel-by-pixel with a Bonferroni correction. The main results showed that the eye region in multiple conditions (e.g., the example in Figure 1.6) contained a significant template. Subsequent confirmation of this result was achieved using the cluster analysis technique developed based on RFT by Chauvin et al. (2005). Filters with

varying spatial frequency and orientation bands can be applied to decompose CIs. This facilitates analysis of templates with Gabor-like spatial arrangements of both black and white pixel clusters.

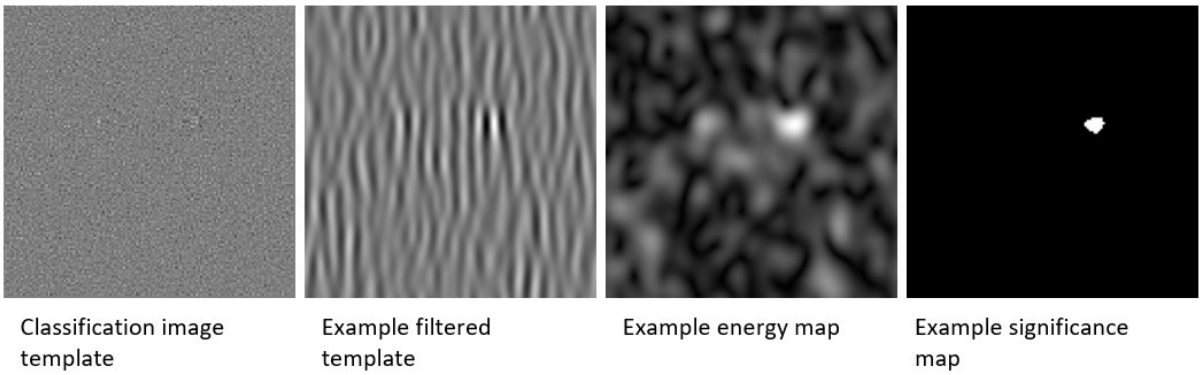

Figure 1.6: Analysis procedure showing example steps for reaching a significance measure, using the most pronounced eye-template out of nine conditions. Part of figure reproduced with permission from Skog et al. (2023).

In conclusion, several methods exist for analysing significant pixels and template regions in CIs. Different methods specialise in capturing significance in different template shapes. Low-spatial frequency templates can be smoothed to reduce high-spatial frequency noise components. Significance in such templates can be discovered through e.g., cluster analysis (Chauvin et al., 2005), reduction by cross-section and subsequent function fitting (Chapter 4 and 5), and testing distribution structure in sub-image areas (Sheskin, 2020). High-spatial frequency templates can be more difficult to prove statistically, as they cover fewer pixels. Cluster analysis (Chauvin et al., 2005) might capture small templates as a few neighbouring pixels of similar sign might significantly deviate from chance. Decomposing CIs via spatial filters can also serve to enhance template structure, increasing the sensitivity of statistical tests (Skog et al., 2023).

The analysis procedure for the CIs in Chapters 4 and 5 of this thesis included cross-sectional reduction and function fitting. Stimulus images contained horizontally arranged targets in the middle of the images, and vertical cross-sections of CIs showed template structures that had positive centres and negative surrounds above and below the middle, resembling Gabor or Difference of Gaussians functions. Fitting Gabor functions to the CI data facilitated analyses of function parameters. The templates were lower in spatial frequency, with Gabor wavelengths of ~2.6 degrees of visual angle. The Gabor fits were used to describe and analyse the use of disparity and luminance cues, group differences, lighting direction priors, and improvements from PL throughout the experiment.

1.15 Bubbles

Gosselin and Schyns (2001) developed another influential CIs technique called Bubbles. Bubbles are based on limiting viewing of the image to a few random locations through Gaussian windows. For

example, if the task is to detect a happy face expression, a window showing a hairline will not likely be diagnostic. A window showing the mouth might, however, be diagnostic. In making task judgements, certain window locations will promote positive responses and others will promote negative responses, similar to how the previously described CIs with noise textures are constructed out of positive- and negative-response templates. Bubbles also belongs to the family of related CI techniques that 'reverse correlates' image parts that promote and demote the task. But the term CIs is generally used throughout this thesis to refer to reverse correlation – the 'classical' CIs with noise textures. The result of combining all window locations for positive responses produces a Bubbles image, highlighting the image areas that the observer uses for the task. Experimentally, Bubbles require far fewer trials than the classical CIs with noise textures, as it essentially only requires the number of trials it takes to discover locations that reliably correlate with positive responses.

Gosselin and Schyns (2001) tasked observers with discriminating a face as expressive or not expressive, finding that observers relied on the mouth for this task. For discriminating male/female gender, observers relied on the mouth and eye regions. Schyns, Bonnar and Gosselin (2002) used Bubbles to examine recognition in identity, gender, and expressive (or not) tasks. They calculated 'attentional maps' which showed probabilities that different parts of faces would be diagnostic of different task conditions. Smith et al. (2005) used Bubbles to examine emotion categorisation of faces expressing six different basic emotions plus neutral. The authors further decomposed the face stimuli into different spatial frequency bandwidths to examine which spatial frequency bandwidths carry emotion expressive content in different parts of the faces. For example, Smith et al. (2005) found that surprise was classified via the mouth region (the face stimulus had an open mouth), but high-spatial frequency information in the eyes and eyebrows was also involved.

There are important differences between Bubbles and the classical CIs using noise textures. Bubbles reveal the areas of an image required for the task. CIs can, on the other hand, reveal the areas required but also the shape and nature of the template applied for the task. For example, in a task of discriminating a smiling mouth from a neutral mouth, CIs could reveal what parts of a mouth is used to make this assessment and what changes in the templates between smile and neutral. In the same task, Bubbles would only reveal that observers need to see the mouth area of the image, revealing less about the information that the observer is using for discriminating between smile and neutral. CIs can reveal template location and shape, while Bubbles are limited to template location. Bubbles also cannot reveal templates from illusory targets, such as detecting a 'superstitious' target or an expression in a neutral face (Gosselin & Schyns, 2003, 2004; Jack et al., 2012; Skog et al., 2023). This is not to say that the classical CIs with noise textures are a superior method. Bubbles can locate information sampling regions more quickly, and has had an impact in areas such as face perception

and brain imaging (not cited in this thesis). The original Bubbles paper by Gosselin and Schyns (2001) has been cited more than any other CI methodology paper, and has had widespread impact in discussions on how visual information is sampled. An example where Bubbles can bring benefit is studies correlating certain Bubbles windows with cognitive or neural processes to discover what image parts correlate with what processes.

Regarding the specific work supporting this thesis, a primary aim was to analyse visual strategies for stereoscopic aerial images. CIs are well suited to explore how visual cues are sampled, and how experts and novices might differ in internal templates. The options for a CIs technique stood between Bubbles and the traditional CIs from noise textures. After a literature overview, the traditional CIs from noise textures was concluded to provide a better suited technique for the current research questions. In order to study the use of binocular disparity cues, stereogram stimulus images were required. CIs with RDS-like noise images can be used to study 3D templates from binocular disparity noise (Gosselin, Bacon & Mamassian, 2004; Neri, Parker & Blakemore, 1999). Bubbles, providing a technique of analysing 2D Gaussian windows, lacks obvious relation to investigating cue sampling in stereograms. Stereograms aside, the project aimed to examine luminance cues and any potential cue combination between binocular disparity and luminance. Bubbles remain limited to information sampling location, while CIs based on noise textures can examine the types of cues that are sampled from locations. Thus, the traditional noise texture CIs were used to study visual strategies in stereograms (Chapter 4 and 5).

1.16 Aims of thesis summary

Visual expertise in remote sensing surveying is an underexplored research topic. This thesis aims to provide evidence that contribute to furthering our understanding of the mechanisms involved in interpreting aerial images, and how experience can change this ability. This thesis will describe expert performance within the domain, and the type of skillset that is associated with expertise in remote sensing surveyors. This thesis sets out to prove four broad hypotheses. First, the unfamiliar aerial viewpoint is more difficult to process, but expert surveyors are better at processing the aerial viewpoint. Next, surveyors are experienced with sampling binocular disparity cues in stereoscopic aerial images, and can thus make more use of this cue. Further, the surveyors also adapt to the aerial imagery, and this can alter perceptual priors for interpreting shape from shading. Finally, the surveyors develop expertise from experience, and this can in part be explained by PL. This thesis builds on these broad hypotheses with a set of specific studies elaborated on in four empirical chapters (2-5). For an elaboration on the specific research aims, see ‘Aims of thesis’ on page 8, or the later chapters which describes the studies in full detail.

To conclude this chapter, here follows an overview of the thesis structure. Chapter 2 begins with a study that characterises novices and expert surveyors' ability to interpret the aerial viewpoint. This study provides evidence that expert surveyors are better at recognising the configurations of features seen from the aerial viewpoint. Chapter 3 continues with the development of a more specific method in preparation for the work in Chapter 4 and 5. This method was a novel version of CIs that could simultaneously estimate templates from disparity and luminance cues. Chapter 4 uses this method to study expert-novice differences in the use of stereoscopic cues for discriminating aerial landscape features. As is later shown, Chapter 4 suggests that expert surveyors have learned to diminish the influence of the lighting-from-above prior in aerial images, likely as a result of their experience with lit-from-below images. This study also provides evidence that experts have a large advantage for sampling disparity cues in stereograms. Chapter 5 continues this theme by exploring the mechanisms of how such an advantage can be developed with training. This study trained novices with a PL intervention to improve their ability to sample disparity cues in stereograms. With CIs, the results characterise learning and show how training can change internal templates involved in stereopsis. Following this final empirical chapter, Chapter 6 provides a discussion to conclude this thesis.

Chapter 2

Expertise for aerial images: Evidence from scene gist and object matching across ground and aerial viewpoints.

2.1 Introduction

This study sought to explore and characterise expertise for classifying landscape features in aerial images. For this purpose, two specific experiments were conducted: the first explored categorisation of scenes in aerial images, and the second explored object matching across ground and aerial viewpoints. To characterise expertise in these tasks, remote sensing surveyors from the OS were compared to novices from the general population. A primary challenge with remote sensing surveying is the use of aerial images, as the aerial viewpoint provides an unusual view of landscapes. Surveyors must become familiar with the configurations of landscape features seen from the aerial viewpoint, and they undergo specific training for this, and learn from experience. This study sought to explore if, and characterise how, expert surveyors performed better in aerial viewpoints.

Remote sensing surveyors must resolve difficult classifications in aerial images, which are unfamiliar to human observers. The aerial viewpoint radically changes the appearances of landscape features compared to the ground viewpoint. Aerial scenes are more difficult to process and classify than ground-view scenes (Lloyd, Hodgson & Stokes, 2002; Loschky et al., 2015; Pannasch et al., 2014), in part because aerial images are more homogenous in spatial structure (Loschky et al., 2015; Oliva & Torralba, 2001). Objects can also be more difficult to recognise from such atypical viewpoints (e.g., Biederman & Gerhardstein, 1993; Lawson, 1999). Remote sensing surveyors undergo training to learn how to perform photogrammetric tasks with aerial images, and gain expertise over time which helps to enhance visual processing skills (Harel, 2016; Seitz, 2017, 2020; Chapter 4).

Expertise for remote sensing surveying of aerial images has previously been explored in a small number of studies. Šikl et al. (2019) studied expert aerial image analysts and compared their performance to novices from the general population in a task of visual recognition memory for aerial images. The authors found a memory advantage where the experts recalled aerial scenes more accurately. Lansdale, Underwood and Davies (2010) used remote sensing surveyors from the OS in a study of expertise and visual saliency in a visual search task. Novices were consistently drawn to salient features in aerial images, while experts were able to discount salient but irrelevant features. Lloyd, Hodgson and Stokes (2002) found that geographers who were familiar with aerial images were better than novices at categorising land-use in such images. Studies have also shown that training can improve novices' recognition in aerial images (Borders et al., 2020; Lloyd, Hodgson & Stokes, 2002).

Remote sensing surveying requires expertise with scenes seen from the aerial viewpoint. Much is previously known about perception of scenes in the literature, and the scene gist paradigm is a prominent area of research (Castelhano & Henderson, 2008; Fei-Fei et al., 2007; Malcolm, Groen & Baker, 2016; Oliva & Torralba, 2006; Rayner et al., 2009; Rousselet, Joubert & Fabre-Thorpe, 2005; Schyns & Oliva, 1994). The 'gist' of a scene refers to the basic encoding of scene information, which can be achieved even with very brief presentation durations. For example, a superordinate category discrimination between 'natural' and 'man-made' scenes is possible even if stimuli are viewed for only 8-20 ms (Furtak, Mudrik & Bola, 2022; Greene & Oliva, 2009; Joubert et al., 2007; Loschky & Larson, 2010). Global information regarding scene context and spatial structure may be encoded in very early processing stages of perception, even in images with added blur that makes objects indistinguishable (Oliva & Torralba, 2001, 2006; Schyns & Oliva, 1994). For example, observers realise that they are looking at a forest before processing individual trees. Gist processing has also been related to expertise with medical imagery. Radiologists achieve above-chance performance for detecting small targets in x-ray images (~70% correct responses at 200 – 250 ms), suggesting that local as well as global information can be processed in briefly presented images (Drew et al., 2013; Evans et al., 2013; Kundel & Nodine, 1975). Loschky et al. (2015) studied rapid visual categorisation for ground-view and aerial images using novice participants. Accuracy was consistently 15-20% lower for aerial images compared to ground-view images across a range of short durations (24 – 330 ms). The authors also tested 'natural' and 'man-made' scene categories, and analysed confusions within and between scene categories. Within category confusions dominated such that natural scenes tended to be confused with other natural scenes more so than with man-made scenes, and vice versa. Furtak, Mudrik & Bola (2022) found that background scene contexts were classified more accurately than foreground objects. The authors also added natural and man-made objects into natural and man-made scene contexts, showing that disrupting the congruency between scene context and objects can lead to lower accuracies (see also Davenport & Potter, 2004; Joubert et al., 2007). Gist processing can thus be influenced by both global context and local objects.

In Experiment 1 of this chapter, the scene gist paradigm provided a way to study rapid categorisation performance in novices and expert remote sensing surveyors using both ground-view and aerial images. As we are all experienced with ground-view images, no group differences were expected for this condition. However, novices were expected to have difficulty processing aerial images, while experts have overcome these difficulties. Experts should therefore make more accurate aerial-view scene categorisations than the novices. Although experts are experienced with aerial images, both groups were expected to have higher accuracies for ground-view than aerial images because aerial images are more homogenous in image structure (i.e., aerial images generally

have a diminished global contrast distribution across orientations and the spatial frequency spectrum) and thus provide less support for gist processing (Loschky et al., 2015; Oliva & Torralba, 2001). Experiment 1 used multiple natural and man-made scene categories, and the analysis included confusions across scene categories and groups. Here, the experts were expected to make fewer confusions than the novices, and be more consistent across viewpoints, owing to their familiarity with aerial scenes.

Beyond scene gist, this study also sought to explore identity judgements across ground and aerial viewpoints. Remote sensing surveyors classify landscape features in aerial viewpoints, but no study to date has examined their ability to match identities across ground and aerial viewpoints. Perspective switches across the ground and aerial viewpoints are unusual forms of 3D rotations. Processing of rotations and unusual viewpoints has commonly been studied in object perception (e.g., Biederman & Gerhardstein, 1993; Edelman & Bülthoff, 1992; Lawson, 1999; Newell et al., 2001; Tarr et al., 1998). Thus, a second experiment was designed to focus on the perception of objects. Here, observers were tasked to match the identities of houses seen from ground and aerial viewpoints. Achieving visual object constancy is often easy if all parts are visible from different viewpoints, e.g., with a 45° horizontal rotation that does not lead to any parts occlusions or changes (Biederman & Gerhardstein, 1993; Lawson, 1999). However, the aerial viewpoint can lead to parts changes, and differences in the relative emphasis of parts compared to ground views. The view of a house from the ground emphasises the façade features, while the roof may be partially occluded. Seeing the same house from aerial view can provide the opposite emphasis, with façade features occupying less space in the image while the roof is fully visible. Clearly, matching across these two viewpoints can be difficult, as parts can change and be emphasised differently. Further, the aerial viewpoint is unfamiliar to most, and atypical viewpoints can make recognition more difficult (Center et al., 2022; Edelman & Bülthoff, 1992; Newell et al., 2001; Tarr et al., 1998). Atypical viewpoints are often discussed in the context of 'canonical' views of objects (Palmer, Rosch & Chase, 1981). Recognition is most effective in canonical views, and 3D rotations can cause disruptions to recognition.

Ground-view images are canonically perceived within a 'gravitational frame', where images have a preferred orientation, or a 'perceptual upright', which is expected to appear congruent with gravity (ground down, sky up) (Asch & Witkin, 1948; Dyde, Jenkin & Harris, 2006; Loschky et al., 2015; Mittelstaedt, 1983). But for aerial images, the gravitational frame is frontoparallel to the observer, and house façades are oriented in all directions in the 2D image plane. The façades of houses are generally considered the 'canonical side', and are visible in all images used in Experiment 2. As the façades are often diagnostic in Experiment 2's matching task, the analysis will explore

whether participants mentally rotate aerial images to face downwards in the 2D image plane when matching with their ground-view counterparts. Mental rotation is a cognitive operation in which observers mentally rotate an object to a preferred orientation to better understand its shape (Shepherd & Metzler, 1971). Mental rotation tends to require longer response times (RT) with larger rotations. The aerial images of houses in Experiment 2 had façades facing in different directions, and façades facing downwards in the 2D image plane was hypothesised as a preferred orientation for the matching task. Loschky et al. (2015) has previously shown that aerial images are not mentally rotated when observers perform a scene categorisation task (as in Experiment 1 of the current study). Categorisation of aerial-view scenes can be accomplished in any image orientation, as identifying a parking lot from the aerial viewpoint can be done based on features such as cars and parking spaces. This feature identification strategy requires no mental rotation of images in the 2D plane. But observers might benefit by mentally rotating aerial images when forced to match a ground- and aerial-view house if the task is easier to perform when the façade is facing a preferred orientation (e.g., downwards) prior to matching.

In Experiment 2, participants were tasked with matching identities of houses across ground and aerial viewpoints. For novices, the ground view of houses is canonical, but the aerial view is atypical and unfamiliar. The surveyors, however, are experienced with aerial viewpoints, and they should be better able to perform this matching task compared to novices. Regarding observer strategies in Experiment 2, participants might employ one of two competing strategies: 1) a feature identification strategy ('feature ID') would match features in the house images without mentally rotating the aerial image, or 2) a mental rotation strategy would start with a mental rotation of the aerial image so that the house façade faces downwards, followed by feature matching. These strategies were investigated without explicit predictions about the strategies or any group differences.

Experiment 2 included two control experiments that reflect these two strategies, in the form of matching tasks using letters as stimuli. The 'feature ID experiment' did not require mental rotation as the task could be done on simple feature analysis across same or different letter pairs (Prather & Sathian, 2002). In contrast, the 'mental rotation experiment' required mental rotation of the letters prior to matching to avoid incorrectly matching mirror-asymmetric pairs. These two control experiments were used to anchor the results for the main experiment using house images. No effects of expertise were expected for these control experiments, as these tasks and images are outside the domain of the surveyors' expertise.

This chapter focuses on group differences between expert remote sensing surveyors and untrained novices from the general population in tasks involving aerial images. Experiment 1 was a

rapid scene categorisation task, and experts were expected to have processing advantages for aerial-view scenes, but not ground-view scenes, compared to novices. Furthermore, the experts were expected to show a greater consistency across viewpoints compared to novices in an analysis of confusions among scenes. Experiment 2 was an object matching task with images of houses seen from ground and aerial viewpoints, where experts were expected to be better at matching across viewpoints. This experiment also explored whether observers mentally rotate aerial images of houses prior to matching with a ground-view counterpart.

2.2 Experiment 1: Rapid scene categorisation

2.2.1 Method

2.2.1.1 Stimulus images

To ensure that a diverse set of scene categories were used, 14 categories were chosen spanning both 'natural' and 'man-made' superordinate categories. For each of the 14 categories, 10 ground-view and 10 aerial images were selected as stimuli, for a total of 280 images. See Figure 2.1 for examples from both viewpoints. The expert participants were recruited from OS, but scene categories were not selected and distinguished based on common OS task specifications. For example, the experts would be accustomed to categorising both 'crop field' and 'cattle field' as 'agricultural'. Ground-view images were sourced from the public domain on Flickr (www.flickr.com). Aerial images were sourced from OS. The aerial photographs originally covered land areas of approximately 2.5km x 1.5km (450 megapixels) but were cropped to smaller portions of land to isolate the relevant scene categories. Prior to the experiment, all images were cropped to a square, grayscaled, down sampled to 384 x 384-pixels using bicubic interpolation, and stored in Portable Network Graphics (PNG) format. This processing was conducted using MATLAB (The MathWorks Inc). During the experiment, images were scaled in PsychoPy (Peirce et al., 2019) using linear interpolation to a square 30% of the participant's monitor height in pixels. Two naïve participants and one experienced psychophysical observer were asked to judge all stimulus images based on whether they unambiguously corresponded to their supposed scene category. This led to the replacement of two ground-view images and nine aerial images (an example exclusion was a 'parking lot' right next to 'industrial buildings'). All 280 images used in the experiment were judged to be appropriate category exemplars.

A pre-experiment survey study was conducted online via Prolific (www.prolific.co) where 20 experimentally naïve, native English-speaking participants based in the United Kingdom rated the image categories from most-to-least natural. The survey provided only the 14 scene category names in text, and no images were provided for reference. The survey ranked the scenes most-to-least

natural in the following order: woods, beach, mountain/moor, river, lake, cattle field, crop field, park, train track, residential houses, urban city, parking lot, highway, and industrial buildings.

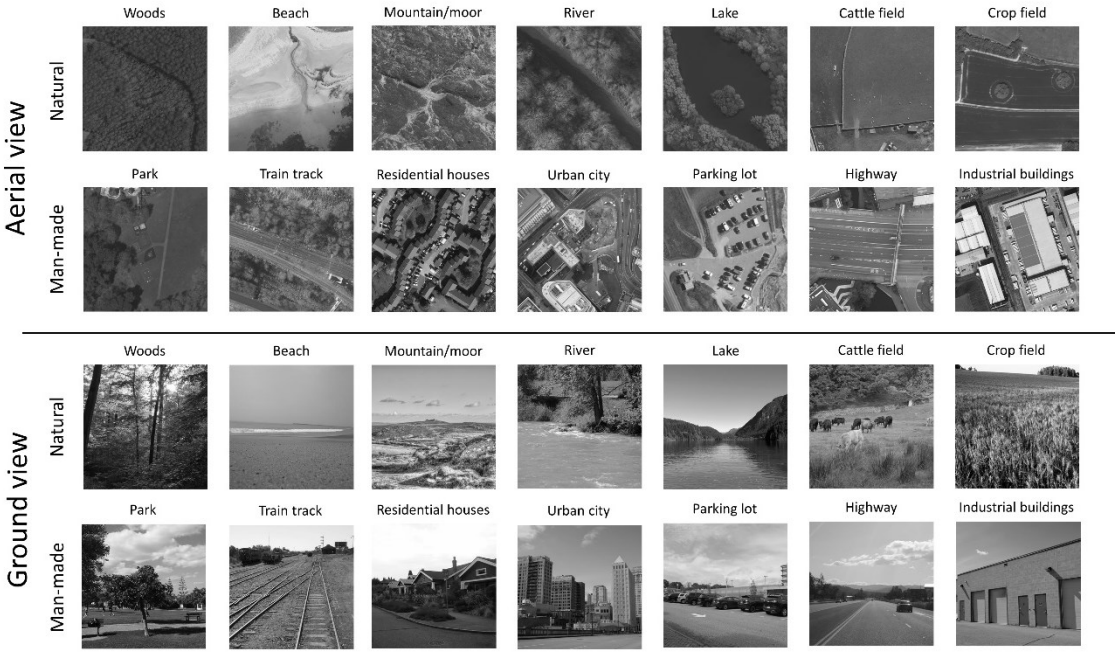

Figure 2.1: Example stimulus images from all 14 scene categories from both ground and aerial viewpoints. (Aerial views *© Crown copyright and database rights 2023 OS*, used with permission; ground views, public domain www.flickr.com).

2.2.1.2 Materials

The experiment was created using PsychoPy and JavaScript to run on the online experiment delivery platform Pavlovia (Pavlovia.org). As the experiment ran online, participants used their own desktop computers to access and run the experiment by clicking a hyperlink that would open the experiment in their web-browser. PsychoPy handled stimulus timings (PsychoJS, version 2021.2.0). The computer, monitor, mouse and keyboard, viewing distance, and testing environment were not otherwise controlled. The experiment was advertised via email or Prolific.

2.2.1.3 Participants

14 expert participants were recruited from OS (7 female; mean age 40 years (SD 12); mean experience of remote sensing surveying = 7 years (SD 6, range: 1-25 years)). 15 novice participants were recruited from Prolific, but one was excluded based on failing an attention check (7 female; mean age: 37 years (SD: 10)). The novices had an average of 737 (SD: 538) total approved participations in other studies and surveys on Prolific. All participants were fluent or native speakers of English and were based in the UK or Ireland. Participants were compensated at a rate of £10 an

hour. The number of recruited participants was limited by the number of experts who would volunteer their participation.

2.2.1.4 Procedure

All participants carried out the experiment during daytime hours, OS employees participated during work hours. At the start of the experiment, participants read a participant information sheet and gave informed consent by a button press when prompted to either agree and continue or exit the experiment. Participants were assured that their data would be confidential and anonymised. The project was reviewed by Aston University's College of Health and Life Sciences Ethical Review committee (approval number 1843). After agreeing to participate, participants indicated by button press if they had significant experience with aerial images. All experts indicated 'yes' and all novices indicated 'no'. The total time for completion was around 25 minutes.

The experiment started by informing the participants that their task was to identify scene categories in briefly presented scenes seen from both ground and aerial viewpoints. Example images of scene categories 'crop field' and 'residential houses' were shown with unlimited presentation time, from both viewpoints. This familiarised participants with the appearance of ground- and aerial-view scenes. This screen also showed the response options and informed participants how to give responses via clicking buttons on the screen with the mouse cursor. The buttons were ordered on two rows as follows: 1) Top row: Woods, Train track, Crop field, Residential houses, Lake, Highway, River, 2) Bottom row: Industrial buildings, Mountain/moor, Parking lot, Cattle field, Park, Beach, and Urban city. The instructions promoted careful attention to the task and highlighted that the images would be presented very briefly. Participants then continued to the practise trials.

The trial structure started with a blank screen / interstimulus interval (1,000 ms), followed by a small black fixation cross prompting attention (1,000 ms), followed by a target image (100 ms), immediately followed by a noise texture acting as a backwards mask (100ms; white noise with 32x32 elements). All trial components were presented in the centre of the screen. 100 ms presentation time corresponds to 6 frames on a 60 Hz monitor. After a short blank screen period (500 ms), the 14 response option buttons appeared below the middle of the screen. This task was a 14-alternative forced-choice task. Participants had unlimited time to respond[2], and each response started a new trial. The experiment started with 20 practise trials, 10 ground-view and 10 aerial images from different categories in the main experiment. These images were the same for all participants and were not used in the main experiment. The 280 stimulus images were presented in a fully random

[2] Response times were not analysed due to the requirement of mouse movements and clicks for responses.

order. An 'attention check' text stimulus was presented four times, after every 70th trial, for one second which read 'Red', 'Green', or 'Blue' in the middle of the screen. Following this, three buttons would appear with the corresponding response options, prompting a response. Participants were then shown a pause screen which afforded a self-timed break and an indication of progress.

2.2.2 Results

Figure 2.2 shows accuracy results for the two groups and viewpoint conditions. Experts and novices were 82.6% and 75.1% accurate for the ground-view images, respectively. For the aerial images, experts and novices were 64.9% and 46% accurate, respectively. A 2 x 2 repeated measures ANOVA (Group: expert, novice; Viewpoint: ground, aerial) was used to test the results statistically. Viewpoint produced a significant main effect ($F(1, 26) = 202.0$, $p < 0.001$), as did group ($F(1, 26) = 7.98$, $p = 0.009$). The interaction of viewpoint and group was also significant ($F(1, 26) = 11.9$, $p = 0.002$). Post-hoc t-tests with Tukey corrections showed that, compared to novices, experts gave significantly more correct responses for the aerial images ($t(26) = 3.18$, $p = 0.019$), but not for ground-view images ($t(26) = 2.02$, $p = 0.206$). Furthermore, and as expected, the within-group accuracy was significantly higher for the ground-view images than the aerial images for both the experts ($t(26) = 7.61$, $p < 0.001$) and the novices ($t(26) = 12.49$, $p < 0.001$).

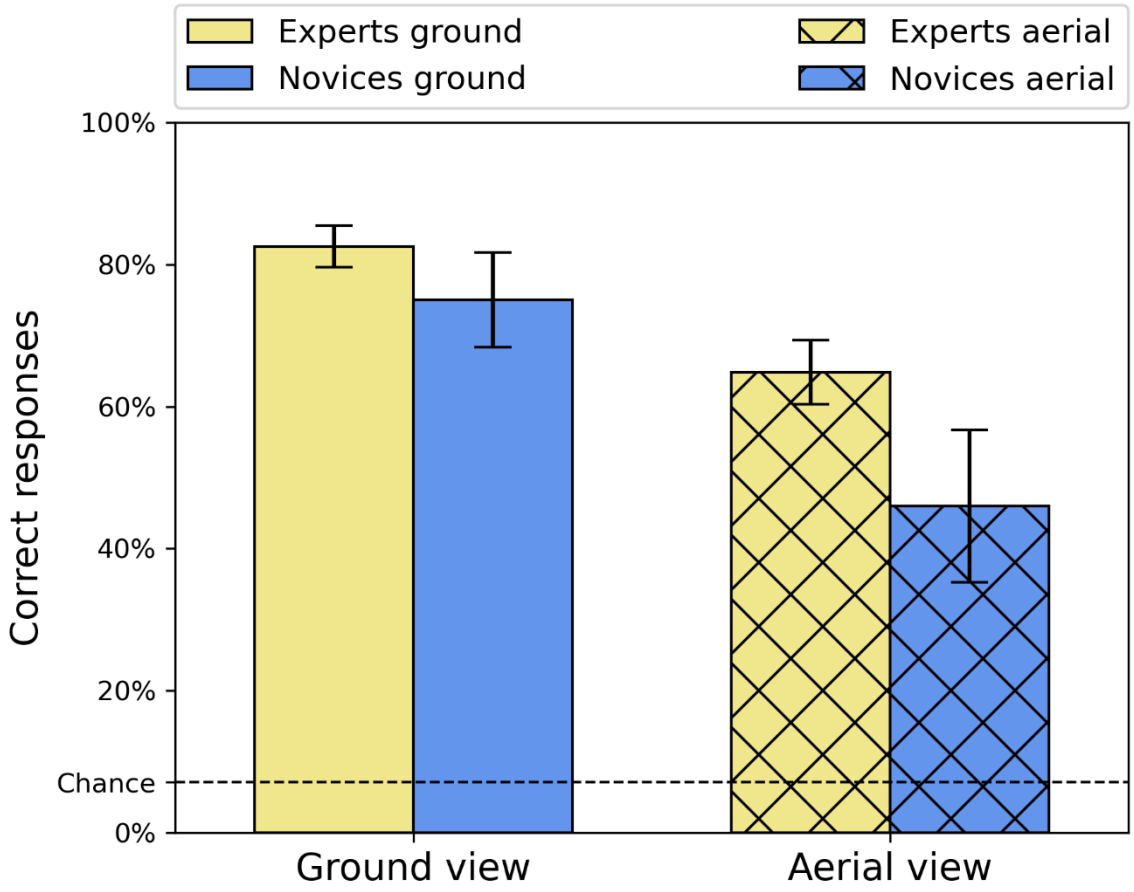

Figure 2.2: Group average percent correct responses for the two groups and viewpoint conditions. Error bars are 95% confidence intervals.

To provide a detailed description and analysis of scene categorisation performance, confusion matrices (CM) were constructed based on the responses given to each scene category. Figure 2.3 shows CMs for both groups and both viewpoint conditions. The main diagonal (top-left to bottom-right) in the figures show correct responses, and responses outside the main diagonal show confusions. An example which caused a relatively high number of confusions is rivers being confused with lakes when seen from ground-view (left side in Figure 2.3; row 4, column 5).

Inspection of the CMs suggests that novices make more confusions than experts in the aerial images, which relates to the lower total accuracy score (Figure 2.2). The distribution of responses appears to differ between groups and sections. The CMs were divided into four sections (four quarters of equal sizes) to examine if their distributions differed across conditions in some sections but not others. Cells on the main diagonal were excluded from this analysis, to focus on confusions. The sections were analysed using a repeated measures ANOVA (2 x 2 x 4; Group: expert, novice; Viewpoint: ground, aerial; Section: top left, top right, bottom left, bottom right) which was corrected with a Greenhouse-Geisser correction following a positive test for sphericity. The four sections correspond to different types of image-response relationships. See top of Figure 2.3 for an

illustration, where, for example, the top right section of the CMs corresponds to natural scenes and 'man-made' responses (Nat-Man). Results of the ANOVA showed significant main effects for viewpoint ($F(1, 41) = 16.22$, $p < 0.001$), section ($F(2.26, 92.52) = 13.60$, $p < 0.001$), and group ($F(1, 41) = 42.58$, $p < 0.001$). The interaction between viewpoint and group was significant ($F(1, 41) = 7.14$, $p = 0.011$). This interaction shows almost the same results as in Figure 2.2, but with the exclusion of data from the main diagonal. The interaction between section and group was near significant ($F(2.13, 87.16) = 2.96$, $p = 0.054$), but the interaction between viewpoint and section was not significant, nor was the interaction between all three factors.

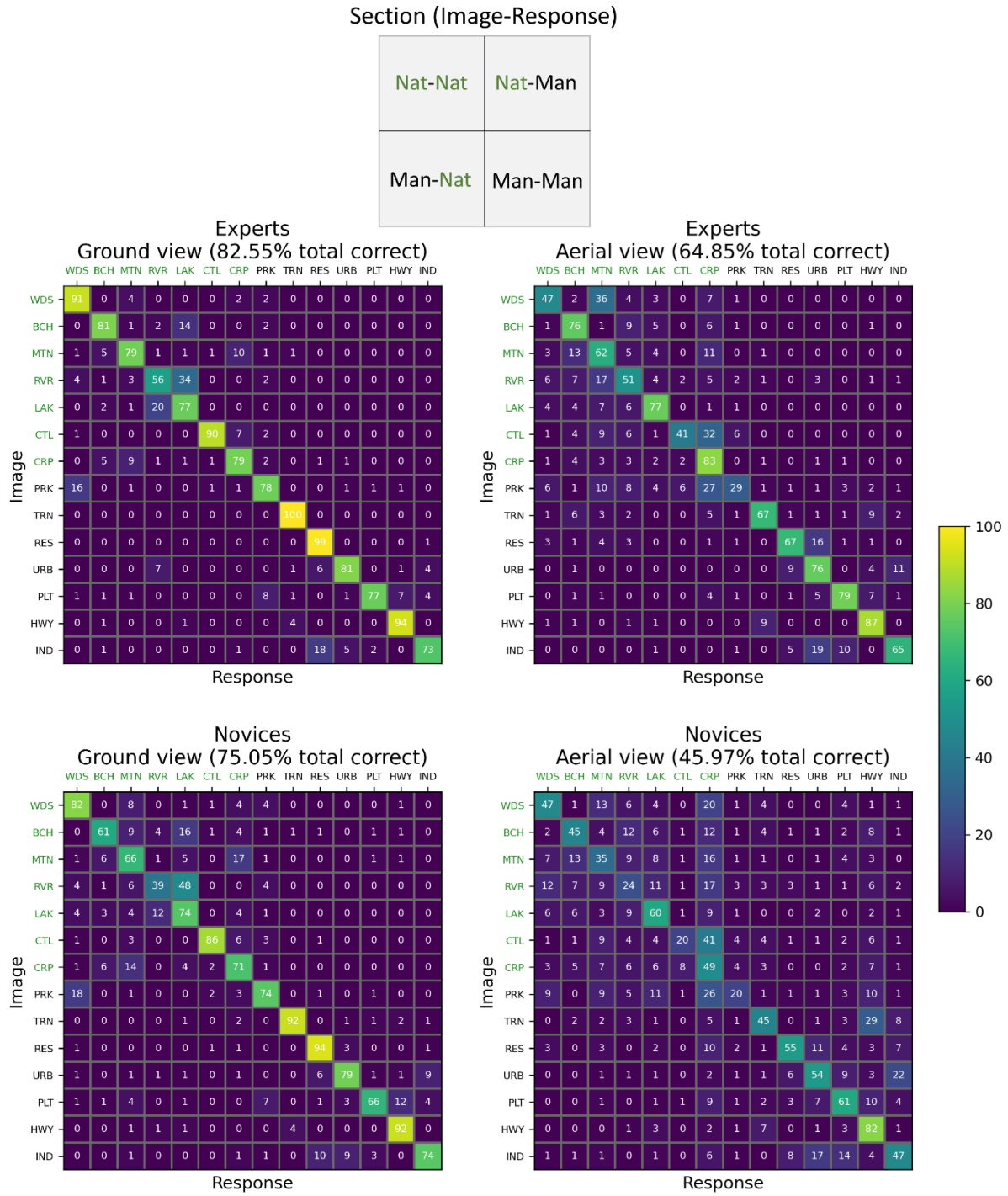

Figure 2.3: Group averaged confusion matrices for the two groups and viewpoint conditions. Cell values are represented in percent responses. Rows indicate the image category, and columns indicate the response given. The most natural half of the image category labels is shown in green, and the most man-made half is shown in black. A diagram at the top of the figure illustrates the division of sections (Nat = 'natural', Man = 'man-made'). Axis labels are coded as: WDS: woods; BCH: beach; MTN: mountain/moor; RVR: river; LAK: lake; CTL: cattle field; CRP: crop field; PRK: park; TRN: train track; RES: residential buildings; URB: urban city; PLT: parking lot; HWY: highway; IND: industrial buildings.

Motivated by the highly significant differences in distributions of cells across CM sections (Figure 2.3), post-hoc t-tests with Tukey corrections provided a piecemeal analysis by individual comparisons, while adopting a conservative estimate of significance. Table 2.1 shows selected

relevant comparisons, which omits the less meaningful comparisons across sections. Rows 1-4 in Table 2.1 show group comparisons for aerial images. One section, Nat-Man, significantly differs between the groups. This means that novices tended to give 'man-made' responses to natural scenes more so than experts. Rows 5-8 show group comparisons for ground-view images. No sections significantly differ after correction, suggesting an overall agreement between the groups for ground-view images. Rows 9-12 show viewpoint comparisons for experts. No sections significantly differ. Finally, rows 13-16 show viewpoint comparisons for novices. Two sections significantly differ, Man-Nat ('man-made' responses to natural scenes) and Man-Man (confusions among 'man-made' scenes). Rows 9-16 reveal that experts had good within-group agreement across ground and aerial viewpoints in all four sections of the CMs, but novices had two sections that differed and two that did not. Novices thus show worse within-group agreement across ground and aerial viewpoints compared to experts.

While the ANOVA tests for differences between the distributions of responses within sections, it does not consider the pattern of confusions within each section. Correlations across sections, groups, and viewpoints provide a complementary analysis of the patterns in the different sections (right of Table 2.1). Rows 1-8 provide a similar outcome as the above analysis, where experts and novices are more consistent with each other in the ground-view images (4/4 correlated sections) than in the aerial images (3/4 correlations). Rows 9-16 show that the ground and aerial viewpoints were prone to different confusion patterns. This was true for both groups, as both groups had two correlated and two not correlated sections each (Rows 9-16).

Different confusions across ground and aerial viewpoints were to be expected. For example, a lake and a river can be more discriminable from aerial view compared to ground view. From the ground view, both typically appear as an expanse of water in front of the observer, but from the aerial view, lakes may appear as an oval-shaped body of water, but rivers appear as curves or lines (Figure 2.1). Another example is cattle fields and crop fields, where cattle fields had a very similar appearance to crop fields from the aerial view, with tiny dots of cattle in a large field (Figure 2.1). From the ground view, however, the cattle were more noticeable, and individual bodies of sheep or cows occupied more space in the images (Figure 2.1). Examples such as these motivated an implicit expectation that confusions may differ across viewpoints, and thus be poorly correlated across viewpoints (Rows 9-16).

Row	Comparison						Post-hoc t-test					Correlation	
	Group	View-point	Section	Group	View-point	Section	Mean diff.	SE	*t*(41)	*p*	*p*(Tukey)	*r*	*p*
1	Experts	Aerial	Nat-Nat	Novices	Aerial	Nat-Nat	-1.89	0.86	-2.19	0.034 *	0.698	0.705	< 0.001 ***
2	Experts	Aerial	Nat-Man	Novices	Aerial	Nat-Man	-1.62	0.29	-5.59	< 0.001 ***	< 0.001 ***	0.132	0.367
3	Experts	Aerial	Man-Nat	Novices	Aerial	Man-Nat	-0.44	0.4	0.28	0.281	0.999	0.850	< 0.001 ***
4	Experts	Aerial	Man-Man	Novices	Aerial	Man-Man	-2.07	0.65	-3.18	0.003 **	0.148	0.731	< 0.001 ***
5	Experts	Ground	Nat-Nat	Novices	Ground	Nat-Nat	-1.63	0.48	-3.41	0.001 **	0.088	0.936	< 0.001 ***
6	Experts	Ground	Nat-Man	Novices	Ground	Nat-Man	-0.15	0.1	-1.55	0.13	0.969	0.728	< 0.001 ***
7	Experts	Ground	Man-Nat	Novices	Ground	Man-Nat	-0.24	0.2	-1.19	0.24	0.997	0.889	< 0.001 ***
8	Experts	Ground	Man-Man	Novices	Ground	Man-Man	-0.46	0.3	-1.53	0.135	0.972	0.823	< 0.001 ***
9	Experts	Aerial	Nat-Nat	Experts	Ground	Nat-Nat	2.62	1.38	1.89	0.065	0.864	0.160	0.311
10	Experts	Aerial	Nat-Man	Experts	Ground	Nat-Man	0.15	0.14	1.12	0.269	0.999	0.535	< 0.001 ***
11	Experts	Aerial	Man-Nat	Experts	Ground	Man-Nat	1.38	0.75	1.84	0.073	0.887	0.174	0.230
12	Experts	Aerial	Man-Man	Experts	Ground	Man-Man	1.46	0.7	2.1	0.042 *	0.757	0.366	0.017 *
13	Novices	Aerial	Nat-Nat	Novices	Ground	Nat-Nat	2.87	1.46	1.96	0.056	0.828	0.242	0.122
14	Novices	Aerial	Nat-Man	Novices	Ground	Nat-Man	1.62	0.29	5.61	< 0.001 ***	< 0.001 ***	0.137	0.347
15	Novices	Aerial	Man-Nat	Novices	Ground	Man-Nat	1.58	0.72	2.18	0.035 *	0.703	0.317	0.026 *
16	Novices	Aerial	Man-Man	Novices	Ground	Man-Man	3.08	0.81	3.79	< 0.001 ***	0.035 *	0.526	< 0.001 ***

Table 2.1: Results of statistical testing of sectioned confusion matrices. Post-hoc t-test results are shown for selected comparisons in the 2 x 2 x 4 repeated measures ANOVA. Correlations show Pearson's *r*. Sections are labelled according to image-response category (Nat = 'natural', Man = 'man-made'). SE = standard error. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.2.3 Discussion

In a rapid visual categorisation task with ground-view and aerial images of landscape scenes, expert surveyors trained on classification of aerial landscape images had higher accuracy than novices for aerial but not ground-view images (Figure 2.2). A within-group analysis showed that both experts and novices were better at ground-view than aerial images, likely due to ground-view images being richer in spatial structure (Loschky et al., 2015; Oliva & Torralba, 2001).

Confusion matrices showed that experts and novices made similar confusions in ground-view images, but differed in the aerial images such that novices more often confused natural scenes with man-made scenes. Within-group, experts were consistent across ground and aerial viewpoints, but novices differed in two out of four CM sections. Novices more often confused natural scenes with man-made scenes, and made more confusions among different man-made scenes. The novices thus showed a greater tendency to be confused with aerial images, and tended to give more man-made responses. These differences in confusions reveal group differences for aerial but not ground-view images, and suggest that experience with aerial images improves consistency across ground and aerial viewpoints. Previous studies have trained novice participants on aerial images, suggesting that experience can improve the consistency between ground and aerial images (Borders et al., 2020; Lloyd, Hodgson & Stokes, 2002). The current study supports these findings, as the experts show greater consistency across the viewpoints.

2.3 Experiment 2: Object matching

Continuing from gist processing of aerial scenes, Experiment 2 investigated identity matching across aerial and ground viewpoints using images of houses. Such perspective switches can lead to difficulties in recognising objects, but as experts are familiar with aerial viewpoints, they were

expected to perform better in this task. This experiment also included an investigation into whether aerial images of houses were mentally rotated in the 2D image plane prior to matching with ground-view counterparts. Furthermore, to anchor the results regarding such strategies, two control experiments were based on letter stimuli, where participants could perform the task based on feature identification ('feature ID experiment'), or where the task required mental rotation of the letters ('mental rotation experiment').

2.3.1 Method

2.3.1.1 Stimulus images

Ground-view images of houses[3] were collected through ground photography in a suburban district of Birmingham, UK, using a 48-megapixel camera. The street and location of each house image was recorded. Aerial images were sourced from the OS, and cropped to small portions focusing on individual houses in the same city. Ground-view and aerial photographs were captured in October 2022 and July 2022, respectively. Vegetation and movable objects such as cars could vary across viewpoint images, meaning that some 'non-house' features could be inconsistent across images.

100 houses were selected to create 'Same' pair stimulus images by pairing ground-view and aerial-view images of the same houses. 100 houses were selected to create 'Different' pair stimulus images by selecting a ground-view image of a house and finding an aerial image of a different house. The process of finding a different house from aerial-view followed some constraints. The different house had to be similar in one or two, and differ in one or two, of the following factors: 1) house shape outline, 2) roof shape (e.g., gabled or hipped roof), or 3) façade and roof features (e.g., bay windows, chimneys, or dormer windows). The different house was always of similar size and sourced from the same street, or a similar-looking, nearby street. Façades were visible in all ground and aerial images of houses. In some aerial images of houses, the back of the house was cropped out, but no diagnostic information would be lost due to this as the backs of the houses were never visible in the ground-view images.

Prior to the experiment, images were processed in MATLAB and were cropped to a square, grayscaled, resized to 300 x 300-pixels using bicubic interpolation, and stored in PNG format. Images in a stimulus pair were then arranged next to one another with a 300-pixel wide empty space between them to create stimulus images (300-pixel tall and 900-pixel wide). See Figure 2.4a for example house stimulus images. Ground- and aerial-view images were arranged left-right and right-left equally often. During the experiment, images were scaled in PsychoPy (Peirce et al., 2019) using

[3] Many of the houses were conjoined pairs of semi-detached homes that appear as one building. Such structures are also referred to as houses.

linear interpolation to a height of 30% and a width of 90% of the participant's monitor height in pixels. Ground-view images were higher than 300 pixels in resolution prior to resizing, but aerial images were lower in resolution (mean square pixel area: 122, SD: 25). The resolution of the aerial-view houses was limited by the resolution of the aerial photographs once cropped to the desired, small space depicting individual houses. This meant that aerial-view houses had a grainier appearance than the ground-view houses.

To create rotation conditions, aerial-view houses were categorised based on their orientation, with a definition that 0° orientation is houses with façades facing downwards in the 2D image plane (see Figure 2.4a for example rotations of the aerial images of houses). Houses were categorised into five rotation conditions: 0°, ±45°, ±90°, ±135°, and 180°. House images were counterbalanced for: rotation, street, general appearance, house size, shading, and sunlight direction. Within the 200 stimulus images, there were 40 images per rotation, divided evenly and counterbalanced across the same/different stimulus images. Only the aerial-view houses had varying orientations, and the ground-view houses were always in their original orientation (Figure 2.4a).

Control experiments were used to anchor the results of the main house experiment, reflecting two different strategies that observers might use: 1) a feature identification strategy ('feature ID'), or 2) a mental rotation strategy (see Introduction). In the control experiments, letters were paired to create stimulus images. Capital letters F, G, J, L, P, and R were presented in the 'Calibri' font. Letters were displayed with a height of 9.5% of the participant's monitor height in pixels, and were arranged and spaced apart similarly to the house images (Figure 2.4). The letter pairs were rotated similarly to the house images (0°, ±45°, ±90°, ±135°, and 180°). Both letters could be rotated, and the rotation condition of the stimulus was defined by the difference in rotation across the letter pair.

In the feature ID control experiment, a 'same' pair was defined as the same letter appearing in both locations (Figure 2.4b). A 'different' pair was defined as two different letters. The letters were rotated in a counterbalanced order, creating 12 stimulus images in each of the five rotation conditions, counterbalanced across the same/different trials, for a total of 60 trials.

The mental rotation control experiment introduced mirror-reversals (x-axis inversion; Figure 2.4c). In this experiment, a 'same' pair was defined as the same letter appearing twice with the same mirror-reversal status (neither mirrored or both mirrored). A 'different' pair was defined as the letter appearing with different mirror-reversal status (mirrored differently). Stimuli were counterbalanced across rotation conditions for a total of 60 trials.

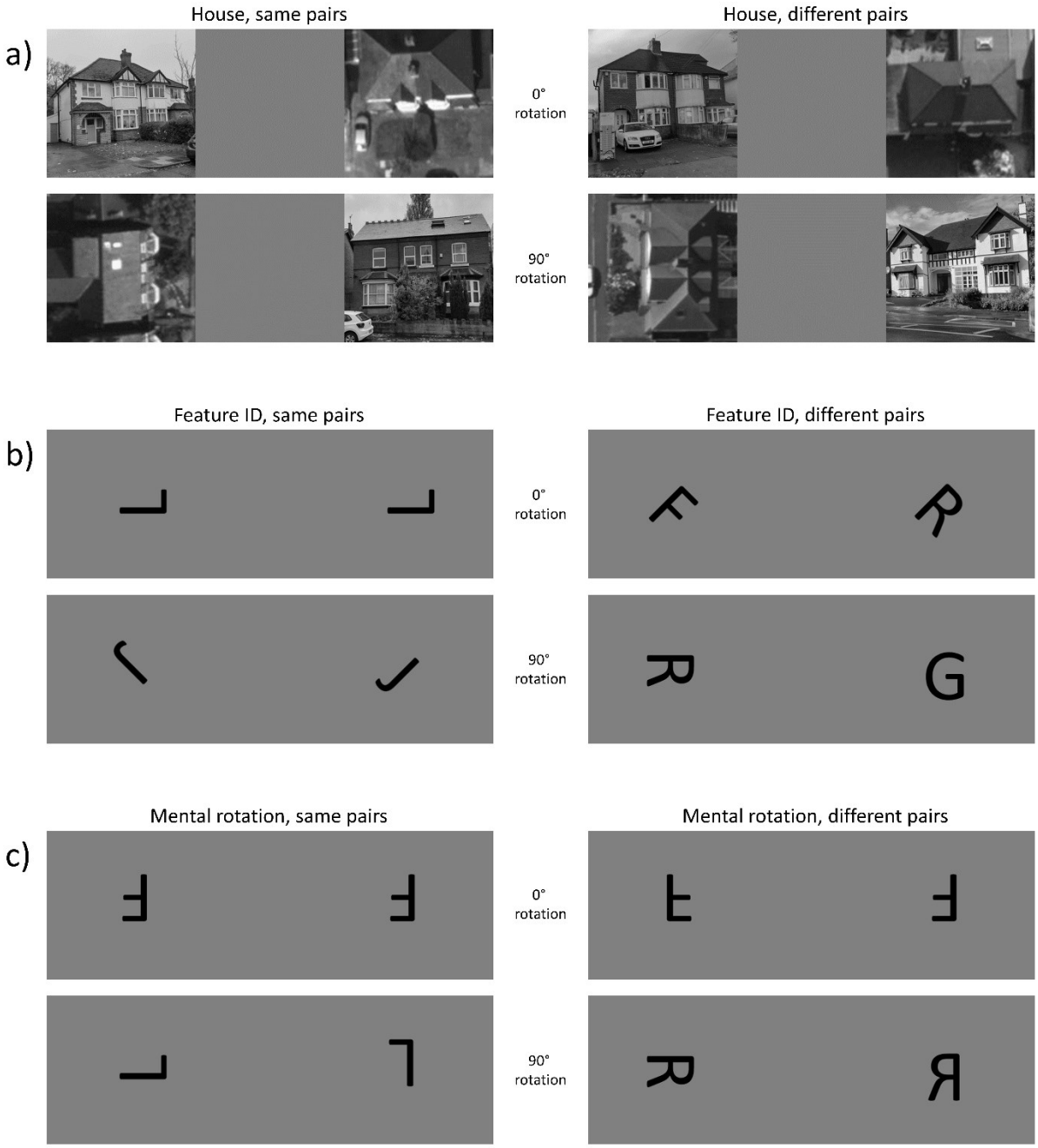

Figure 2.4: Example stimulus images for: a) the house experiment, b) the feature ID experiment, and c) the mental rotation experiment. Examples include 0° and 90° rotation conditions.

2.3.1.2 Participants

12 expert participants were recruited from OS (7 female; mean age: 40 years (SD: 10); mean experience with remote sensing surveying = 10 years (SD: 5), range: 1-25 years). All 12 experts had previously participated in Experiment 1. 13 novice participants were recruited from Prolific, but one was excluded based on failing an attention check criterion (8 female; mean age: 38 years, SD: 12). The novices had an average of 440 (SD: 439) total approved participations in other studies and surveys on Prolific. No novice had previously participated in Experiment 1. All participants were

fluent or native speakers of English and based in the UK or Ireland. Participants were compensated at a rate of £10 an hour.

2.3.1.3 Procedure

Experiment 2 was identical to Experiment 1 in terms of materials, access to the experiment, and ethical considerations. All experts, and none of the novices, reported having significant experience with aerial images.

For the sake of counterbalancing block orders, the main 'house experiment' and the control 'letter experiments' were treated as two different parts of the experiment, where half of the participants started with the house experiment followed by the letter experiments, and vice versa for the other participants. Furthermore, the letter experiments were counterbalanced in order such that half the participants did the feature ID experiment first and the mental rotation experiment second, and vice versa. Four different block orders were thus possible, and these were counterbalanced among the 12 participants in each group.

The instructions for all parts of the full experiment stated that the task would be a 'same or different' judgement where the 's' and 'd' keys would be used for 'same' and 'different' responses, respectively. Text was permanently present at the top of the screen throughout the full experiment reminding participants how to respond with these buttons. Participants were instructed to respond accurately and quickly. The instructions for the house experiment included examples of a 'same pair' and a 'different pair' of houses. Participants were instructed that two houses would appear next to one another from ground and aerial viewpoints, and that they should judge if the houses are the same house or different houses. After these instructions, participants did six practise trials before starting the house experiment. Throughout this experiment, text reminding participants how to respond read: "'S' key: Same house, 'D' key: Different houses". The instructions for the feature ID experiment stated that the 'same or different' judgement regarded letters, where a same pair was the same letter, and a different pair was different letters. Three same and three different pair examples were shown in the instructions. Participants continued with five practise trials. Throughout this experiment, text reminding participants how to respond read: "'S' key: Same letter, 'D' key: Different letters". The instructions for the mental rotation experiment stated that the 'same or different' judgement regarded mirror-reversal status. Three same pairs examples (with text stating: 'Both not mirrored' or 'Both mirrored') and three different pair examples (with text stating: 'Mirrored differently') were shown. Participants continued with 10 practise trials where the correct answer was given to them in text instructions, for the purpose of aiding task learning. Throughout this

experiment, text reminding participants how to respond read: "'S' key: Looks the same, 'D' key: Mirrored differently". No practise stimuli were repeated in the main section of any experiment.

Trials started with a 1,500 ms blank screen interstimulus interval, followed by unlimited presentation time of the stimulus. A response started the next trial. All images in all experiments were fully randomised in order. The house experiment included four attention check questions, after every 50th trial, which were similar to Experiment 1. The letter experiments included one attention check question at the end of each block. Following the attention check question, participants were shown a pause screen with either an indication of progress, or the next block would start with instructions for a new experiment.

2.3.2 Results

Individual trials where the RT was more than two standard deviations from the mean for the respective participant were removed from the data. This mainly had the effect of removing very slow trials.

The accuracy results (Figure 2.5) were converted to the sensitivity measure *d*' to avoid response bias between Same and Different responses. In this stimulus-response task, Same and Different images and responses were recorded as hits (Same-Same), misses (Same-Different), false alarms (Different-Same), and correct rejections (Different-Different). The Same and Different images are thus used in conjunction to derive the *d*' sensitivity measure displayed in Figure 2.5[4]. The RT results (Figure 2.6) are displayed in seconds, and split by whether the images were a Same image pair (solid lines) or a Different image pair (dashed lines). Only correct responses were used in the RT data. Both accuracy and RT results are split by rotation conditions in Figures 2.5 and 2.6. Repeated measures ANOVAs were used to test results statistically (2 x 5; Group: expert, novice; Rotation: 0°, 45°, 90°, 135°, 180°). The RT data further included another factor (Image type: same, different). Results of the ANOVAs are displayed in Table 2.2.

The results from accuracy (Figure 2.5) and RT (Figure 2.6) are shown in separate figures, and the details of the results are elaborated below.

[4] Some participants had a hit proportion of 1.0 and a false alarm proportion of 0.0, producing an infinitely high *d*'. This mostly occurred in the feature ID experiment. To avoid using infinite *d*'s prior to averaging across participants, proportions that were 1.0 or 0.0 were recalculated so that: proportion 1.0 = 1.0 – (1/*number of trials in condition*), and proportion 0.0 = 0.0 + (1/*number of trials in condition*). For example, in the house experiment (Same image pairs), the number of trials was 100, and the *d*' value corresponding to infinity was thus set at 4.65. In the letter experiments (e.g., Feature ID: Same pairs), the number of trials was 30, and the *d*' value corresponding to infinity was 3.67.

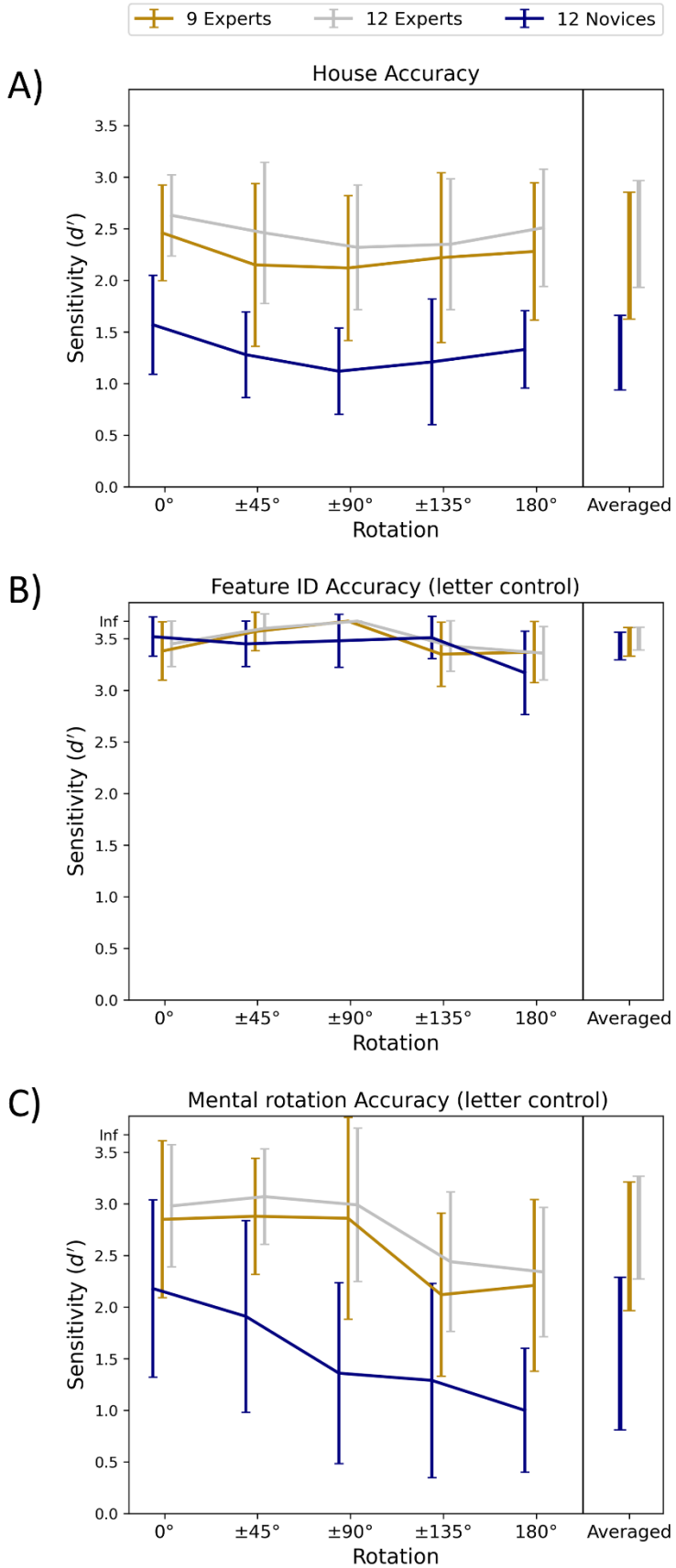

Figure 2.5: Accuracy data for the: a) House experiment, b) Feature ID control experiment, c) Mental rotation control experiment. Infinite sensitivities (corresponding to 100% hits and 0% false alarms) were capped at *d*' = 4.65 for the house experiment (not included in scale) and *d*' = 3.67 in both letter experiments ('Inf'). 'Averaged' is the average across the rotation conditions. Error bars are 95% confidence intervals.

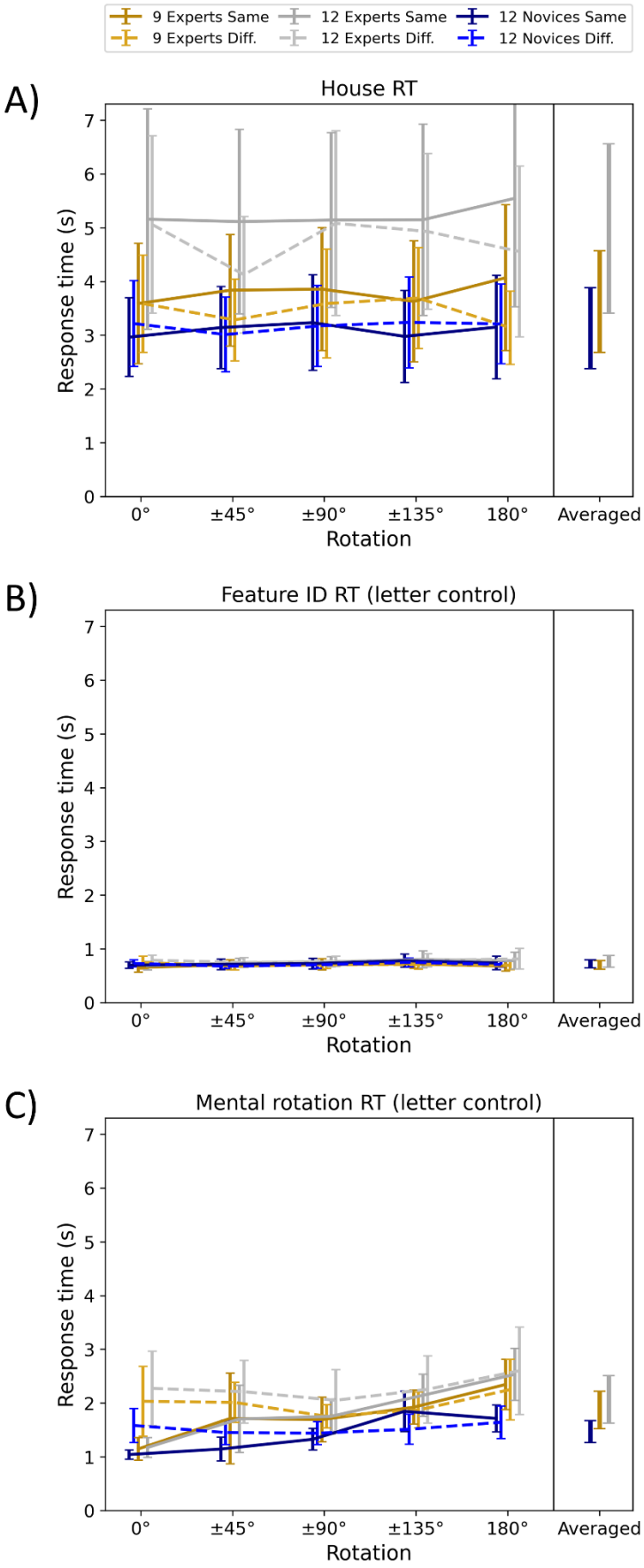

Figure 2.6: Response time (RT) data for the: a) House experiment, b) Feature ID experiment, c) Mental rotation experiment. 'Averaged' includes data from both Same and Different. Error bars are 95% confidence intervals.

Regarding accuracy in the house experiment (Figure 2.5a), experts and novices were 84.8% (*d*' = 2.45) and 71.1% (*d*' = 1.30) accurate, respectively. The results of the repeated measures ANOVA are displayed in Table 2.2: House accuracy. These results show that the experts were significantly more accurate, and that different rotations did not affect the accuracy results for either group.

For RT in the house experiment (Figure 2.6a), experts and novices took on average 4.99 and 3.13 seconds to respond, respectively (Table 2.2: House RT). These results show a low-powered significant main effect between groups, and no evidence of rotations affecting the RTs. Looking more closely at the RT results, the novices' RTs ranged from 1.47 to 6.27 seconds, and the experts ranged from 1.95 to 10.11 seconds. But three experts had notably slower RTs than the rest (these were: 10.11, 9.05, and 8.03 seconds), and the other nine experts' RTs ranged from 1.95 to 5.97. The relationship between speed (RT) and accuracy in the expert population was examined by removing these three experts and repeating the above analyses for both accuracy and RT. For accuracy, the outcome remained largely unchanged (Table 2.2: House accuracy), with a significant main effect between groups ($F(1, 19) = 7.47$, $p = 0.013$), with no other significant effect or interaction. But for RT, the main effect between groups changed, and was now not significant ($F(1, 19) = 0.66$, $p = 0.425$), while the other outcomes remained largely unchanged (Table 2.2: House RT). Overall, removing the three slowest experts produced only a small change in the accuracy results, but a large change in the RT results[5]. All results in Figures 2.5 and 2.6 thus include data from both the selected 9 experts (gold) and all 12 experts (silver). In comparing these gold and silver data, notice a relatively small change in Figure 2.5a: 'House Accuracy', but a relatively large change in Figure 2.6a: 'House RT'. This suggests that the speed-accuracy relationship in the expert population might not follow the 'standard model' of speed-accuracy trade-offs, where longer RTs lead to higher accuracies.

The letters control experiments (see Figure 2.4b, c for example stimuli) were analysed with the same ANOVAs as above. In both control experiments, removing the three slowest experts produced little change in the results. The results for both accuracy and RT thus regard all 12 experts compared to the 12 novices (Table 2.2). In the feature ID experiment, both groups completed the task with almost perfect accuracies (Figure 2.5b) and similar RTs of around 700-750 ms (Figure 2.6b). This shows no group differences in this baseline identification task using same or different letter pairs (Table 2.2; Figure 2.4b). Furthermore, as expected, no effect of rotation was found, as this task could be performed on features.

[5] The accuracy of 9 experts was 82.9% (*d*' = 2.24), and 12 experts was 84.8% (*d*' = 2.45). The RT of 9 experts was 3.63 seconds, and 12 experts was 4.99 seconds. Apart from a large change in mean RTs, the homogeneity of variances across groups for RT improved notably from using 12 experts (Levene's $F(1, 22) = 9.15$, $p = 0.006$) to using 9 experts (Levene's $F(1, 19) = 0.21$, $p = 0.652$).

In analysis of the mental rotation experiment, highly significant main effects of rotation were seen for both accuracy and RT (Figure 2.5c and 2.6c; Table 2.2), showing that mental rotation was engaged in this control experiment which used letters that could be mirror-asymmetric (Figure 2.4c). This outcome affords a contrast to the main house experiment, where no effect of rotation was observed for the aerial house images (Figure 2.5a and 2.6a; Table 2.2). Regarding group differences, the experts were slower (Figure 2.6c; Table 2.2) but more accurate (Figure 2.5c; Table 2.2) compared to novices in this experiment. Furthermore, the effect of mental rotation was significantly more present in the same compared to the different images for both groups (Figure 2.6c; Table 2.2).

Main effects and Interactions	House Accuracy (d')	House RT †	Feature ID Accuracy (d') †	Feature ID RT †	Mental rotation Accuracy (d') †	Mental rotation RT † ‡
Rotation	$F(4, 88) = 1.6$, $p = 0.182$	$F(2.89, 63.47) = 2.01$, $p = 0.124$	$F(2.61, 57.42) = 2.14$, $p = 0.113$	$F(1.45, 31.85) = 2.32$, $p = 0.128$	$F(3.04, 66.83) = 5.85$, $p = 0.001$ **	$F(2.53, 55.65) = 16.97$, $p < 0.001$ ***
Group	$F(1, 22) = 12.8$, $p = 0.002$ **	$F(1, 22) = 4.33$, $p = 0.049$ *	$F(1, 22) = 0.69$, $p = 0.416$	$F(1, 22) = 0.46$, $p = 0.503$	$F(1, 22) = 7.17$, $p = 0.014$ *	$F(1, 22) = 6.05$, $p = 0.022$ *
Same/Diff.	-	$F(1, 22) = 0.56$, $p = 0.461$	-	$F(1, 22) = 0.10$, $p = 0.753$	-	$F(1, 22) = 5.42$, $p = 0.030$ *
Rotation X Group	$F(4, 88) = 0.06$, $p = 0.993$	$F(2.89, 63.47) = 1.13$, $p = 0.344$	$F(2.61, 57.42) = 0.74$, $p = 0.517$	$F(1.45, 31.85) = 0.34$, $p = 0.645$	$F(3.04, 66.83) = 0.88$, $p = 0.458$	$F(2.53, 55.65) = 2.3$, $p = 0.097$
Same/Diff. X Group.	-	$F(1, 22) = 1.08$, $p = 0.309$	-	$F(1, 22) = 1.89$, $p = 0.183$	-	$F(1, 22) = 1.76$, $p = 0.199$
Rotation X Same/Diff.	-	$F(2.81, 61.88) = 2.77$, $p = 0.052$	-	$F(2.68, 59) = 3.00$, $p = 0.043$ *	-	$F(2.78, 61.22) = 13.09$, $p < 0.001$ ***
Rotation X Group X Same/Diff.	-	$F(2.81, 61.88) = 1.4$, $p = 0.253$	-	$F(2.68, 59) = 0.32$, $p = 0.787$	-	$F(2.78, 61.22) = 0.78$, $p = 0.501$

Table 2.2: Results of repeated measures ANOVAs to accuracy and response times (RT) in all three experiments. Significant cells are shaded in green. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. X = interactions. † Greenhouse-Geisser corrected column following a positive sphericity test (except for Group, Same/Diff., and Same/Diff. X Group). ‡ Analyses of RTs were always based on correct responses, but in the mental rotation experiment, some participants (one expert and five novices) provided no correct responses in at least one image category (e.g., Same image, 45° rotation). Such image categories consisted of six trials each. To avoid excluding data from participants on this basis, 13 out of 240 cells were filled in with the average of the other available participants in the same group and condition.

2.3.3 Discussion

The house experiment showed that experts were more accurate than novices for object matching across ground and aerial viewpoints. The experts further displayed a 'nonstandard' relationship between speed and accuracy, where three experts were notably slower than the rest of the participants but did not produce much higher accuracy than the other experts. After removing these three experts, accuracy results maintained the experts' advantage (Figure 2.5a), with RTs not significantly differing between groups (Figure 2.6a). The expert surveyors are trained to prioritise accuracy over speed in remote sensing surveying tasks at the OS. Due to this main occupational task where errors are potentially costly, some experts may spend time confirming an already correct first impression. As the results show a performance advantage for the experts, this could explain why the

experts were generally more accurate instead of faster. The study was followed up by asking one very experienced surveyor if this task was familiar. This surveyor reported that surveyors have experience with matching vertical aerial images from airplane photography with oblique-angled drone photography. These imaging techniques have been paired to register and update landscape changes. While the experts are not accustomed to matching ground-view images (from e.g., Google Street View) with aerial images, this experience could be directly relevant to the experts' advantage in the house experiment.

The results across the rotation conditions showed that participants do not mentally rotate the aerial images of houses in the 2D image plane prior to matching, and that this was true for both groups. These results suggest that participants start identifying and matching features immediately at the start of each trial, regardless of the aerial house image's orientation. This suggests that the expert surveyors had greater facility to identify and match features in the images. Studies that find strong effects of mental rotation commonly involve judgements of rotated abstract shapes, such as in the traditional cube figures of Shepherd and Metzler (1971), in which local features are not diagnostic but the spatial relationships between them are. In the current experiment, mental rotation was likely not necessary for the aerial-view houses because features provided enough diagnostic information (e.g., roof features). The task in the house experiment was also more complex a standard mental rotation experiment as some features were not visible in both images (e.g., a full view of the roof was only visible from the air).

Results from the feature ID control experiment showed that experts and novices performed this simple feature identification task with almost perfect accuracies (Figure 2.5b) and similarly short RTs (Figure 2.6b). As the participants used a feature identification strategy in the house experiment, these results suggest that group differences observed in the house experiment are not explained by the experts having a generalised performance advantage, outside of their area of expertise. Furthermore, these results afford an important indication that both groups, despite not doing the experiment with controlled equipment or environment, performed similarly in a baseline task. However, these results might be subject to a ceiling effect, masking any group differences.

The mental rotation control experiment shows that both groups did mental rotation of the letter stimuli (Figure 2.5c and 2.6c). As letters but not aerial images of houses were rotated, this further supports the conclusion that mental rotation was not utilized as a strategy in the house experiment. Regarding group differences, experts were more accurate but slower compared to novices. Experts and novices may have been motivated differently by their experiences and how they were recruited. Experts are trained to prioritize accuracy over speed, and thus may have wanted to perform well. Novices, who were recruited from an online platform and have participated in many

previous studies and surveys, may be experienced with performing studies quickly and thus prioritized speed over accuracy. Overall, the accuracy in this experiment was lower than what would be desired. Ideally, participants can be supervised and trained to perform with close to perfect accuracy, thus leaving only RT as a variable, which could have increased the clarity of the results for this group comparison. The lack of supervised training in this experiment could account for the highly variable accuracy scores across participants and groups in this experiment.

2.4 General discussion

Expertise in remote sensing of aerial landscape images is associated with higher accuracy in rapid categorisation of aerial-view scenes, and higher accuracy in an object matching task across ground and aerial viewpoints. The results further demonstrate that experts are more consistent across ground and aerial viewpoints, and that aerial images are not rotated in the 2D image plane prior to matching with a ground-view counterpart. These results provide novel evidence of expertise in remote sensing surveyors, who are professionally dedicated to photogrammetry of aerial images.

Lloyd, Hodgson and Stokes (2002) tasked participants with categorising land use in aerial images. The authors geographers as an 'experienced' group, but the expertise for photogrammetry of aerial images within this group is less clearly defined, likely more heterogenous, and was not made explicit by the authors. Šikl et al. (2019) used both psychology and geography students, defining them as novice and intermediate groups, respectively. Geography students, who had some anecdotal experience with aerial images, did not rival experienced remote sensing image analysts in their study's memory task. The authors studied visual recognition memory using a condition with aerial images, but did not provide any control conditions using e.g., ground-view images. The current study can complement Šikl et al. (2019), albeit in a different set of experiments, by finding effects of expertise with control conditions for both experiments. Previous studies have also shown that novices can be trained on aerial images and that this can reduce the gap between the ability to process ground and aerial viewpoints (Borders et al., 2020; Lloyd, Hodgson & Stokes, 2002). Experiment 1 supports these findings, as experts are better than novices at categorising aerial but not ground-view scenes, and the CMs show a greater consistency across viewpoints for the experts compared to the novices.

Humans are known to have processing difficulties with aerial images, which is clearly reflected in the current results obtained from novices. Initial fixations tend to be longer for aerial than ground-view images, suggesting a higher processing difficulty for the gist of the scene (Pannasch et al., 2014). Aerial images are also more homogenous in spatial structure, which is generally associated with worse processing in brief presentations (Loschky et al., 2015; Oliva &

Torralba, 2001). These are factors which make initial processing of aerial images difficult. The results of Experiment 1 suggest that experts have overcome some of this difficulty with experience, but that ground-view images are still easier to process. Studies of expertise in medical imagery could inform how expertise is related to the first fixations in aerial images, as medical images are, like aerial images, unusual to most humans. Radiologists show evidence of early-stage visual processing advantages compared to novices, with more efficient gist processing leading to earlier detection of targets using fewer fixations (Bertram et al., 2013; Drew et al., 2013; Evans et al., 2013; Fox & Faulkner-Jones, 2017; Krupinski et al., 2006; Kundel & Nodine, 1975). Using a visual search task in aerial images, Lansdale, Underwood and Davies (2010) showed that experts can ignore irrelevant but salient features, while novices were drawn to saliency. This further suggests that we have processing difficulties with aerial images which can be improved with experience. Experts can suppress items with higher visual saliency in favour of more meaningful items during search. Furthermore, the results of Experiment 2 in the current study suggest that novices struggled to process features when object matching across ground and aerial viewpoints. Large changes in object appearances due to 3D rotations are known to impair object recognition and constancy (Biederman & Gerhardstein, 1993; Center et al., 2022; Edelman & Bülthoff, 1992; Lawson, 1999; Newell et al., 2001; Tarr et al., 1998). The experts' improved accuracy suggests that experience with aerial viewpoints might improve the ability to maintain object constancy across unusual viewpoints, despite large changes in object appearances. In terms of future directions, a study could explore whether this expertise is robust to other 3D rotations beyond the ground-to-aerial viewpoints.

Remote sensing surveyors have had time and training to learn about the regular appearances of landscape objects and image-statistical regularities in aerial images. However, the development of this expertise in the surveyors' natural workplace environment remains largely unexplored. In terms of future directions, a further study could follow surveyors longitudinally from the beginning of their careers (perhaps testing throughout the first year of work), or define intermediate stages with surveyors who have different levels of experience. Studying remote sensing surveyors poses logistical challenges as there are only a relatively small number of surveyors that work in specialised organizations. This is likely why some previous studies have recommended the use of experts, but not included them in their experiments (e.g., Borders et al., 2020; Pannasch et al., 2014; Rhodes et al., 2021). In the current study, both experiments ran online to gain easier access to OS remote sensing surveyors. Running PsychoPy online via Pavlovia is known to provide reliable visual timing durations, with variability and lag varying by only a few milliseconds across operating systems and web browsers (Bridges et al., 2020). Furthermore, the results of the feature ID experiment in Experiment 2 suggest that, despite non-laboratory control over equipment and environment, both

accuracy and RT were similar in both groups in a simple baseline task. While these factors speak in favour of the current methods producing reliable results, the online procedure of this study remains a limitation.

2.4.1 Summary and conclusions

Aerial images are more difficult to process than ground-view images, but evidence from expert remote sensing surveyors show that some of this processing difficulty can be overcome with experience. In two experiments, experts were more accurate in tasks involving analysis of aerial images. In a rapid scene categorisation task (Experiment 1), experts and novices performed comparably with ground-view images, but experts were more accurate with aerial images. Experts also tended to be more consistent in scene categorisations across ground and aerial viewpoints. In an object matching task (Experiment 2), experts were more accurate when matching houses across ground and aerial viewpoints. This experiment also showed that neither experts nor novices mentally rotate aerial images prior to matching with ground-view images. This result suggests that experts are better at identifying features and their specific configurations in aerial images. The theme of feature identification in aerial images is investigated further in Chapter 4, which examines the features that experts and novices use when classifying stereoscopic aerial images. Overall, this study highlights the benefits of experience for processing aerial images, and provides novel and complementary evidence on expertise for human remote sensing surveying.

Chapter 3

Pilot studies

The previous chapter established that remote sensing surveyors have expertise for processing features seen from the aerial viewpoint. This chapter continues with the development of a more specific method that is later used to characterise expertise for stereoscopic features in aerial images.

This chapter describes preparatory work and pilot studies to extend the CI technique to allow simultaneous estimation of CIs for luminance and binocular disparity. This novel version of CIs was required for a later study which used CIs to estimate how expert surveyors use different 3D cues, such as luminance and disparity, in stereoscopic aerial images. This method was developed in three stages, described here in three experiments (Pilot 1-3). The first experiment in this series was an introductory CI study that used a detection task of a luminance target in luminance noise. This experiment further served to evaluate two different experimental designs that are commonly used in CI studies. The second experiment in the series continued with using a similar detection task but in 3D images defined by binocular disparity. Here, a novel version of CIs was developed that was based on RDS but used dense textures as stimuli rather than sparse dot arrays. Finally, the third experiment in the series used these novel stimulus images to explore a novel task where observers detected luminance and binocular disparity targets simultaneously to generate both types of CIs.

The data presented in this chapter are not used in other thesis chapters, and was acquired from compensated participants who signed informed consent. The projects were reviewed by Aston University's College of Health and Life Sciences Ethical Review committee.

3.1 Pilot 1: Evaluating experimental designs for classification image studies

3.1.1 Aims

As a first step for this project, an introductory experiment used luminance targets and noise to generate luminance CIs. This experiment was basic as it involved replication of previous similar CI studies that have found perceptual templates for targets with simple shapes (e.g., Beard & Ahumada, 1998; Watson & Rosenholtz, 1997).

This experiment also served to evaluate two experimental designs to measure potential differences in efficacy for generating CIs. The first design was a SIBR design, where one stimulus image was presented that contained a target on 50% of the trials, with 'yes or no' detection responses. The second design was a 2AFC design, where two stimulus images were presented in

temporal sequence on each trial, one with and one without the target. In the 2AFC condition, participants were forced to select one image that contained the target.

The two designs each have some strengths and weaknesses. The relative benefits of SIBR and 2AFC designs for CI studies were previously discussed in Chapter 1 (page 29), based on the results of Gosselin & Schyns (2003). The SIBR design is beneficial as it is quick to perform, with just one image and one response, and it affords a less constrained experience for the participants as they are freely allowed to respond 'yes or no' to all images. The SIBR design, however, allows response biases, where participants can provide uneven distributions of responses that could impact the quality of the results. For example, in extreme cases, CIs cannot be generated from participants who give only one type of response. The 2AFC design prevents response biases as, in each trial, one image is selected as containing the target and another image is selected as not. This benefit of avoiding response bias should be weighed against the fact that the 2AFC design requires participants to evaluate twice as many images, which requires longer time commitments for the experiment. The 2AFC design can also be experienced as more constrained for the participants, as they must select one image as 'yes' and another as 'no' on each trial, without having the options to respond 'both' or 'neither'.

3.1.2 Method

Three participants (including the author) were recruited for a CI experiment where the task was to detect a centrally located white square target pedestal (20x20 texture elements) in a white noise texture (64x64 texture elements), with a static SNR that was the same for all participants. The target pedestal added 10 (out of 256) grayscale values and the noise ranged from 0-246 on the grayscale, giving an average SNR of 0.081 when the target was present. The SNR was determined in preparatory work using two observers to approximate ~70% correct responses. The experiment was informally controlled, and performed from the participants' own computers outside of a laboratory environment[6]. Stimulus image sizes were scaled to a 30% ratio of the participants' monitor heights, but viewing distance was not controlled. Two participants were compensated at a rate of £8.33 an hour.

The two experimental designs described above were evaluated for potential differences in efficacy for generating CIs. Each new trial in the SIBR condition generated a random noise texture. The 2AFC condition used two independent noise textures that were presented in temporal sequence with a 500 ms interstimulus interval. Stimulus images were presented for 200 ms. The experiment was generated and run using PsychoPy (Peirce et al., 2019), and the participants had access to the

[6] During this stage, in-person testing was restricted as part of the response to covid-19.

full Python version of PsychoPy on their own computers. Each participant completed 20,000 trials, 10,000 in each condition (SIBR and 2AFC). CIs were generated according to the typical procedure of subtracting the sum of all noise textures which led to negative responses from the sum of all noise textures which led to positive responses (Ahumada, 1996; Beard & Ahumada, 1998; Murray, 2011). For the 2AFC condition, this procedure meant saving the selected noise texture as a positive response, and the other noise texture as a negative response.

Prior to the experiment, simulated ideal observers were used to evaluate the noise textures generated by PsychoPy (version 2020.2.10). PsychoPy's noise textures have a zero mean (at mid-grey luminance), meaning that a CI template consisting of lighter pixels must be accompanied by a surround consisting of darker pixels. Zero mean noise produces a confound to any estimate of inhibitory surround mechanisms that might be implied by a negative (dark) surround in CI templates. To prevent this confound, PsychoPy's noise generating component was modified to have a non-zero mean. Ideal observer simulations showed that the standard zero mean noise necessarily produced a negative surround when a positive centre template was applied. But with the modified non-zero mean noise, the ideal observer could apply a positive centre template without any negative surround. Thus, with non-zero mean noise, evidence of negative surrounds reflects the observer's template rather than inherent structure in the noise. For this reason, all CI experiments in this thesis used a modified version of PsychoPy's noise component that had a non-zero mean.

3.1.3 Results and discussion

Participants were between 67-75% accurate in this experiment. Results show that both the SIBR and 2AFC experimental design conditions produced CIs that contained obvious templates (Figure 3.1). The CI templates appear different in character between participants, where ES's template is offset to the top-left of centre, AP's is more centralised, and OS's is smaller, covering fewer pixels in total. Note that these are participant initials, and OS does not refer to Ordnance Survey here. Notably, these template characteristics were highly similar within-participant across the SIBR and 2AFC condition (Figure 3.1). Templates generally had the characteristics of a white bump, suggesting that lighter and darker noise patterns promoted and demoted detection of the white square target, respectively. Some templates also have subtle negative (dark) surrounds.

The SIBR condition was ~33% faster to complete, as the 2AFC condition required the participants to evaluate twice as many images. As the results clearly show similar CIs across the conditions, the SIBR condition was considered to be superior as it reduced experiment duration with little if any cost to the quality of the CIs. Reducing the time spent within the experiment is beneficial,

as CI studies usually require thousands of trials and several hours of commitment. Following these results, every following CI experiment in this thesis used a SIBR rather than a 2AFC design.

Response biases can typically be tolerated without much cost to CI quality (Gosselin & Schyns, 2003). But studies that use the SIBR design should carefully instruct participants of the presence of the target (e.g., 50% likelihood of target presence), and encourage participants to provide an even distribution of responses throughout the experiment.

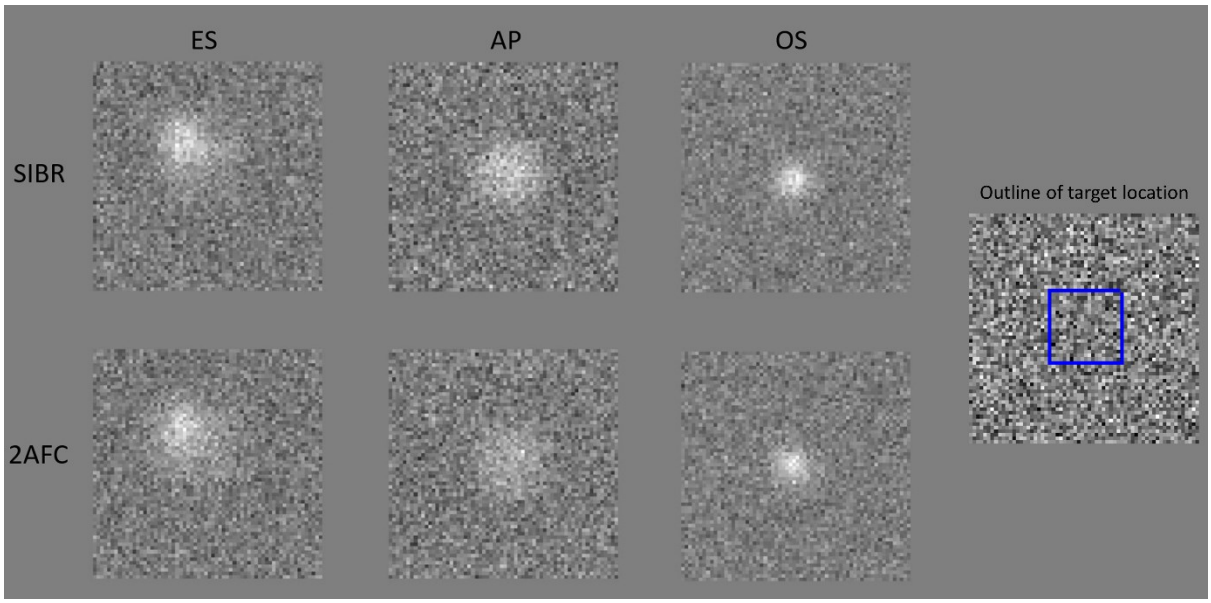

Figure 3.1: Classification images from three participants (initials in columns, note that ES is the author) and two conditions (rows): SIBR: single-interval binary-response design, 2AFC: two-alternative forced-choice design. To the right is an illustration of where the target was located (blue outline).

3.2 Pilot 2: Developing and evaluating a novel method for generating 3D classification images

3.2.1 Aims

Pilot 1 recorded 2D luminance CIs and evaluated experimental designs. Pilot 2 continues by developing a 3D CI from binocular disparity noise. CIs from disparity cues were required to meet the later aims of this thesis, which involved estimating the use of stereoscopic cues in classification of stereoscopic aerial images. The details of this later study are the subject of Chapter 4, which used the noise textures and subsequent CIs that were developed in the current pilot experiment.

Previous studies have demonstrated disparity CIs with RDSs (Gosselin, Bacon & Mamassian, 2004; Neri, Parker & Blakemore, 1999). But as later studies required masking of natural images, a sparse array of dots (Julesz, 1971) seemed unsuitable, as images would be visible through sparse noise. This project thus required dense noise to provide a sufficient masking effect, that could also produce CIs.

3.2.2 Method

RDS provide a simple method for generating horizontal shifts that cause binocular disparities in dichoptic images. With a sparse array of e.g., black dots on a white background, dots can simply be moved horizontally over the empty background, as the dots do not have to compete for space with other dots (Howard, 2002; Julesz, 1971). But with a dense noise texture, horizontal shifts of pixels will compete for space with their neighbouring pixels. By introducing sub-pixel shifts, this problem can be mitigated, as the 'centre of mass' of the pixels can be moved horizontally in steps that are smaller than the size of the pixels. These shifted pixels might still compete for sub-pixel space, but this problem can be mitigated with a competition rule where the pixel with the most crossed disparity occludes its competitor. That is, the 'nearer' pixel is shown. This algorithm for creating horizontal sub-pixel shifts in dense textures was initially inspired by Georgeson, Yates and Schofield (2009), who introduced sub-pixel disparity shifts in noise textures via phase shifted sinusoidal gratings. The algorithm in the current study is based on horizontally expanding the texture, and horizontally shifting pixel values across the image, followed by down sampling to the original image dimensions. This image now contains sub-pixel shifts that cause binocular disparities in dichoptic viewing. Figure 3.2a provides an example of how this algorithm works to introduce disparity noise by introducing horizontal sub-pixel shifts. Figure 3.2b shows an example stimulus image containing disparity noise but no target. The method for creating disparity noise is further expanded on in Chapter 4: Methods: Dual-noise classification images.

To determine how the shifts should be introduced, a 'disparity map' was created on each trial that mapped the application of disparity noise in both crossed and uncrossed directions across the carrier textures. In this experiment, the disparity map was lower in spatial frequency, to introduce smoother transitions between patches of crossed and uncrossed disparity (white noise texture with a Butterworth filter with a cut-off frequency of 9 cycles per image). See Chapter 4: Figure 4.2 for an illustration of a disparity map. The carrier textures were white noise textures that were subject to horizontal shifts as determined by the disparity map (Figure 3.2b). The disparity maps were saved on each trial, and tagged based on the participants' responses, to generate disparity CIs. Note that, as is always the case in CI studies, only the noise is used to generate CIs.

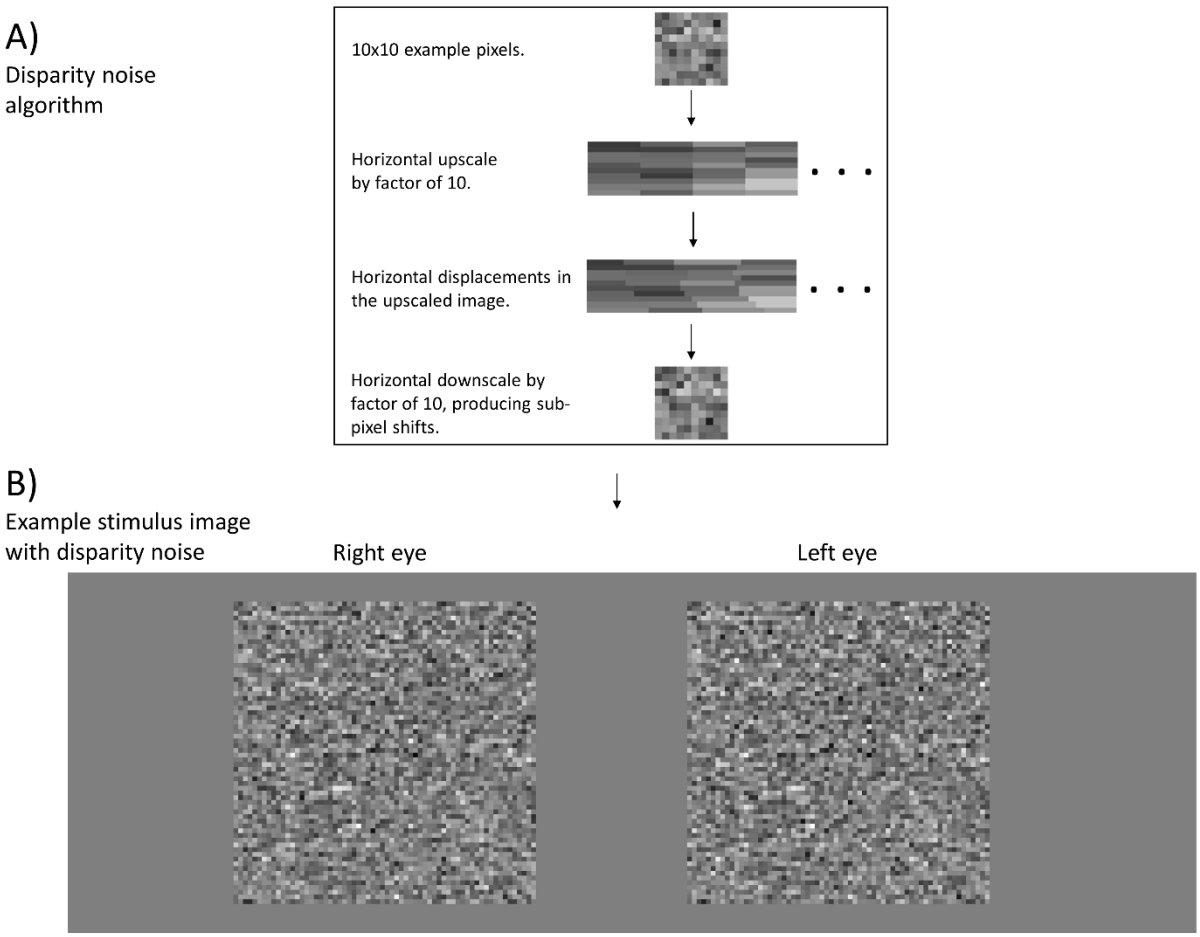

Figure 3.2: a) Disparity noise algorithm for introducing disparity noise by horizontal sub-pixel shifts. b) Stereogram pair of stimulus images containing disparity noise, without any target. Dichoptic fusion reveals disparity noise in both crossed and uncrossed directions, arranged for crossed fusion, and divergent fusion reverses disparities.

Viewing images in a mirror stereoscope, three participants (two female, mean age: 24.5) were tasked to detect a 'near' square (14x14 texture elements, 1.44 degrees of visual angle) defined as a centrally located static pedestal projecting 3.7 arcminutes of crossed disparity in a white noise texture (64x64 texture elements, 6.58 degrees of visual angle). This target was present on 50% of the trials. Stimulus images were presented for 750 ms. Participants were screened for the ability to discriminate the difference between crossed and uncrossed disparity with a similar target, in the absence of external noise. See Chapter 4 for further details on screening procedure, equipment, and how vergence control was supported in the stereoscope with a fixation cross and a surrounding border.

An adaptive staircase procedure was used to vary SNRs by, unconventionally, manipulating the disparity noise level rather than the signal. This could provide psychophysically more granular step sizes in the experimental software than if the signal level was manipulated, as the signal was defined by fewer sub-pixel steps than the total noise range. SNRs were varied in a 1-up, 2-down step procedure, designed to estimate a 70.7% threshold (Levitt, 1971). When the procedure determined

that the SNR should go up, the noise level was reduced, and vice versa. Two such staircases operated in parallel. Step sizes were linear, and would change the range of external disparity noise by increasing or decreasing its range by 37 arcseconds of disparity. The experiment followed the SIBR design and participants responded 'yes or no' with button presses on a keyboard. In this experiment, monitors were not corrected to linearised gamma for the carrier textures, but this was the case in all later experiments using CIs. The experiment consisted of 10,000 trials and participants completed it in two or three days. Participants were compensated at a rate of £10 an hour.

3.2.3 Results and discussion

Figure 3.3 shows CIs generated from the influence of disparity noise in detection of the crossed target ('near' square). All three participants produced a noticeable CI template in the target location which roughly corresponded with the target shape. Light and dark pixels represent crossed and uncrossed disparity, respectively (which is the presentation structure for disparity throughout this thesis). CIs generally contained templates with a central positive peak (light pixels) in the target location, with surrounding negative side-lobes (dark pixels). Differences in amplitude between participants were observed (Figure 3.3 and 3.4), suggesting that participants varied in their ability to sample disparity cues. To aid in visualisation of individual differences, Figure 3.4 shows a horizontal cross-section taken by averaging the 14 central rows, corresponding to the target's location at the full width of the image.

This pilot experiment validated the novel algorithm for generating disparity noise and CIs from dense textures. This technique is developed further in the following pilot experiment, where the method is extended to simultaneously include luminance targets and luminance CIs. The technique is also used in Chapter 4, in a study of stereoscopic judgements of aerial images, and in Chapter 5, in a study of PL for stereopsis in stereograms.

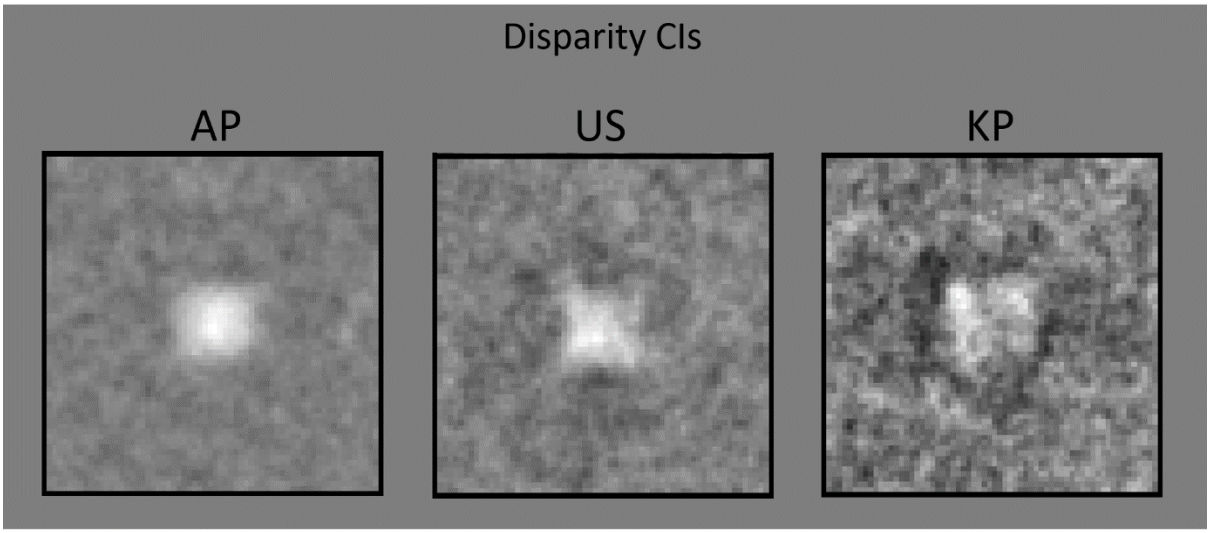

Figure 3.3 Classification images from 3D binocular disparity. Light and dark pixels represent crossed and uncrossed disparity, respectively. Images were scaled individually so that the lightest pixel is at full contrast.

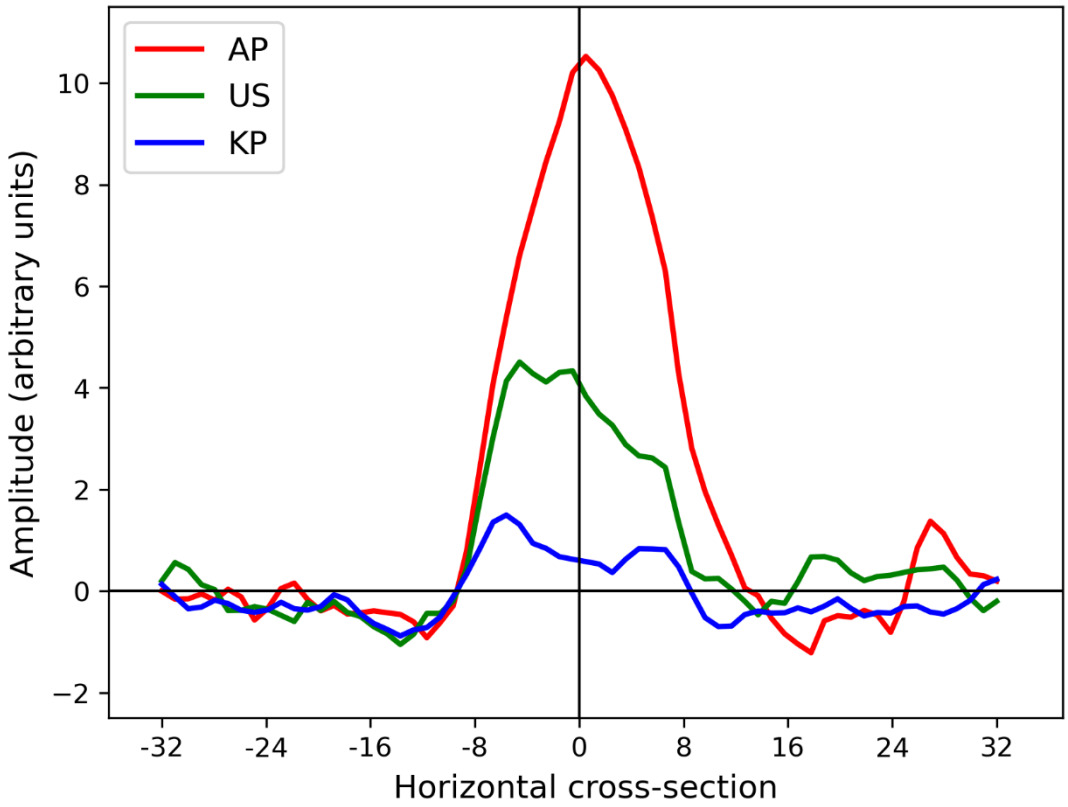

Figure 3.4: A horizontal cross-section of the disparity classification images. 0 on the x-axis indicates the image centre.

3.3 Pilot 3: Novel method for generating simultaneous 2D and 3D classification images

3.3.1 Aims

The previous experiment captured disparity CIs from dense stimulus images using a pedestal target in disparity noise. The current experiment sought to extend this novel technique with an experiment

consisting of three different conditions: 1) luminance, 2) disparity, and 3) both simultaneously. The novel contribution of this experiment was to show that CIs can be captured from both luminance and disparity cues simultaneously. This method is developed further in the next chapter, where this 'dual-noise' CI technique was applied to stereoscopic aerial images to discover the contribution of luminance and disparity cues for a discrimination task with aerial landscape features.

3.3.2 Method

Six undergraduate participants (five female, mean age: 23) were recruited and compensated at a rate of £10 per hour and/or course credits (at Aston University). Five of the participants were naïve and inexperienced with psychophysics experiments. Participant AP was informed about the study and experienced with the stimulus images, having participated in the previous experiment.

Before the start of the experiment, participants were told that they would be searching for a small central square in noise. The square would be defined either by binocular disparity, luminance, or both, varying between sessions but not within sessions. Text was always visible on the top of the screen reminding participants of which condition they were in. The text prompted a response, and read "Did you see a NEAR square?" for the disparity condition, "Did you see a WHITE square?" for the luminance condition, and "Did you see a WHITE and NEAR square?" for the compound condition. The participants completed 5,000 trials per condition over three or four days. The luminance and disparity conditions only contained the luminance and disparity targets, respectively. Luminance and disparity noise was always present in all conditions. The compound condition ('White and Near') presented both the luminance and the disparity targets together. The three conditions constituted a block and the orders inside the blocks were counterbalanced throughout the experiment. Monitors were corrected to linearised gamma.

This experiment used similar methods as the above experiment on 3D CIs (Pilot 2). Targets were present on half of the trials in all conditions. Two staircases were used in parallel to estimate a 75% threshold (1-up 2-down, estimating 70.7%, and 1-up 3-down, estimating 79.4%). The disparity target was a static pedestal (3.7 arcminutes of crossed disparity), and external noise was increased to make detection harder, and vice versa. The luminance target operated differently from the disparity target, by varying in intensity rather than being static. A pedestal of luminance was added in the target area which varied in luminance intensity depending on detection threshold. Detection was made harder by reducing the intensity of the target via the staircases, and vice versa. The luminance aspect was thus treated differently than the disparity aspect. As the luminance texture is a carrier for the disparity noise, varying the contrast of the luminance carrier would alter the carrier for the disparity noise, creating a complicated relationship between levels of external luminance and

disparity noise. The luminance target was thus staircased by altering the target contrast, but the disparity target was staircased by altering the external disparity noise. Despite these differences, both techniques serve to increase or decrease SNR depending on the staircase adjustments that track fixed thresholds.

In the disparity and luminance conditions, the mean of the two staircases provided an empirical estimate of how much the external disparity noise or luminance target contrast needed to change from an estimated 70.7% threshold to an estimated 79.4% threshold. This was used to estimate two points on the participants' psychometric functions for the luminance and disparity elements when setting the SNR for the compound stimulus ('White and Near'). Equivalent step sizes were estimated so that both the luminance and disparity aspects of the compound stimulus remained approximately equally detectable when staircased together. The SNR for the compound stimulus was determined throughout the experiment for each participant based on their most recent disparity and luminance thresholds.

3.3.3 Results

Results replicate the previous experiments in showing that participants produced both luminance and binocular disparity CIs in separate conditions (Figure 3.5: 'Disparity: Near' and 'Luminance: White'). Extending these results, the compound condition shows that the two cue dimensions were combined and sampled simultaneously to produce CIs (Figure 3.5: 'White & Near' for both disparity and luminance). The luminance CIs were generated from the carrier textures, which were white noise textures (e.g., Figure 3.2b). But the disparity CIs were generated from smoother, low-pass filtered textures (see Pilot 2 for a description of the disparity maps). These differences account for the apparent spatial frequency differences between the luminance and disparity images seen in Figure 3.5.

Visualising the CIs generated from binocular disparity, Figure 3.6 shows cross-sectioned disparity CI data from the disparity and compound conditions. As in the previous experiment, participants varied in their ability to sample disparity cues, as shown by apparent individual differences in CI template amplitudes. Figure 3.7 shows cross-sectioned luminance CI data from the luminance and compound condition. Overall, most participants displayed at least marginal use of both cue dimensions in the compound condition (Figure 3.5, 3.6 and 3.7).

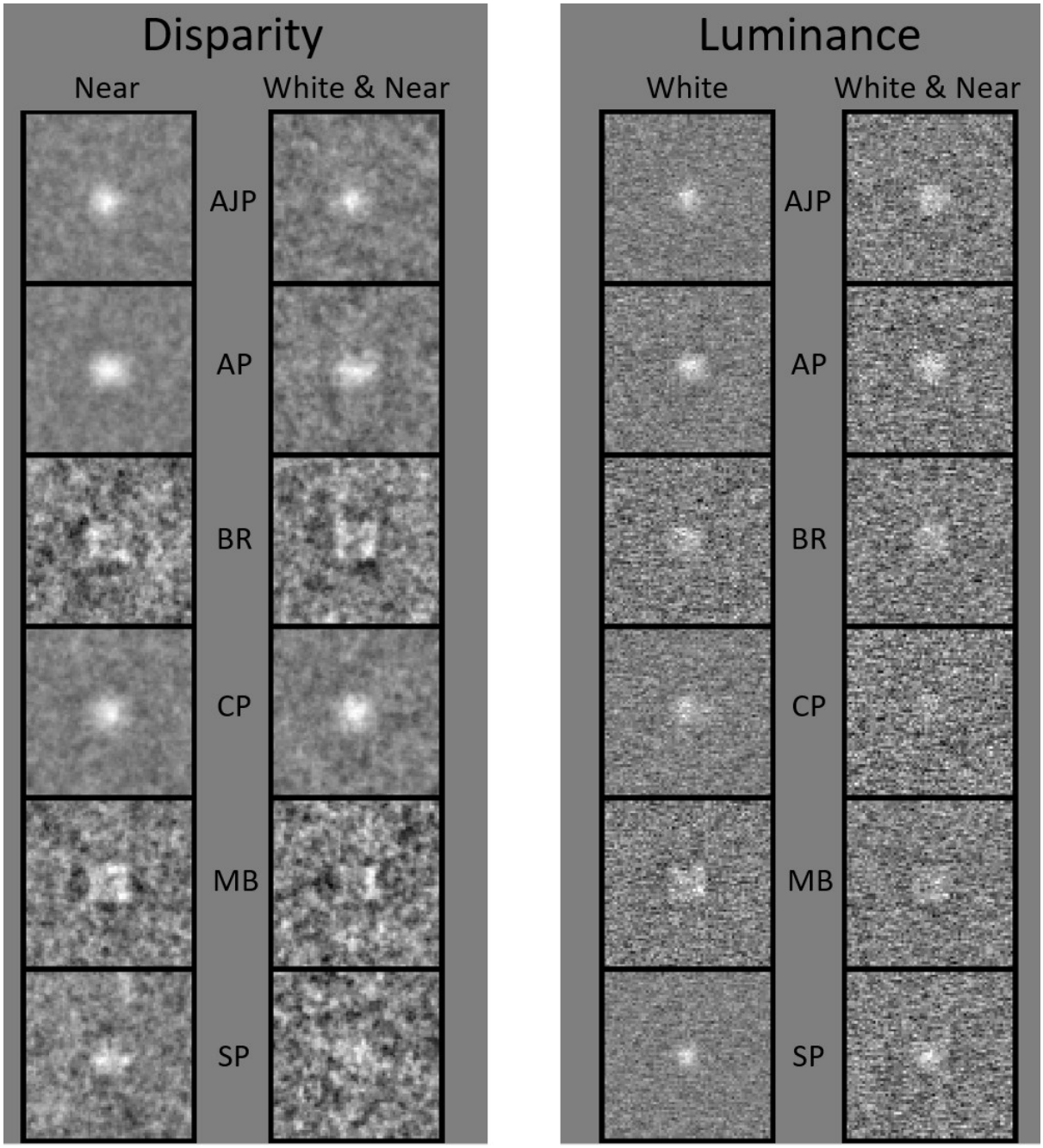

Figure 3.5: Classification images from all three conditions. Images are presented similar to Figure 3.3. Left side: classification images from binocular disparity, split by disparity ('Near') and compound ('White & Near') conditions. Right side: classification images from luminance, split by luminance ('White') and compound ('White & Near') conditions.

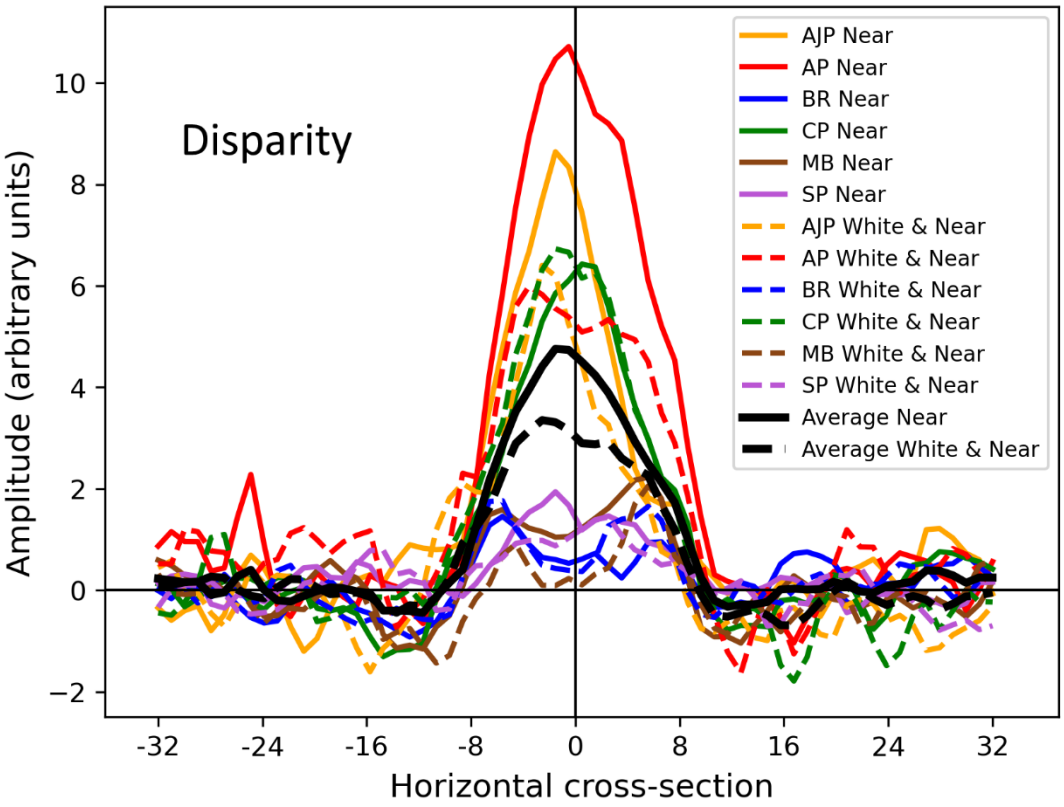

Figure 3.6: Horizontal cross-sections of disparity classification images as in Figure 3.4. Solid and dashed lines represent the disparity condition and the disparity aspect of the compound condition, respectively. Black curves indicate the average of the conditions.

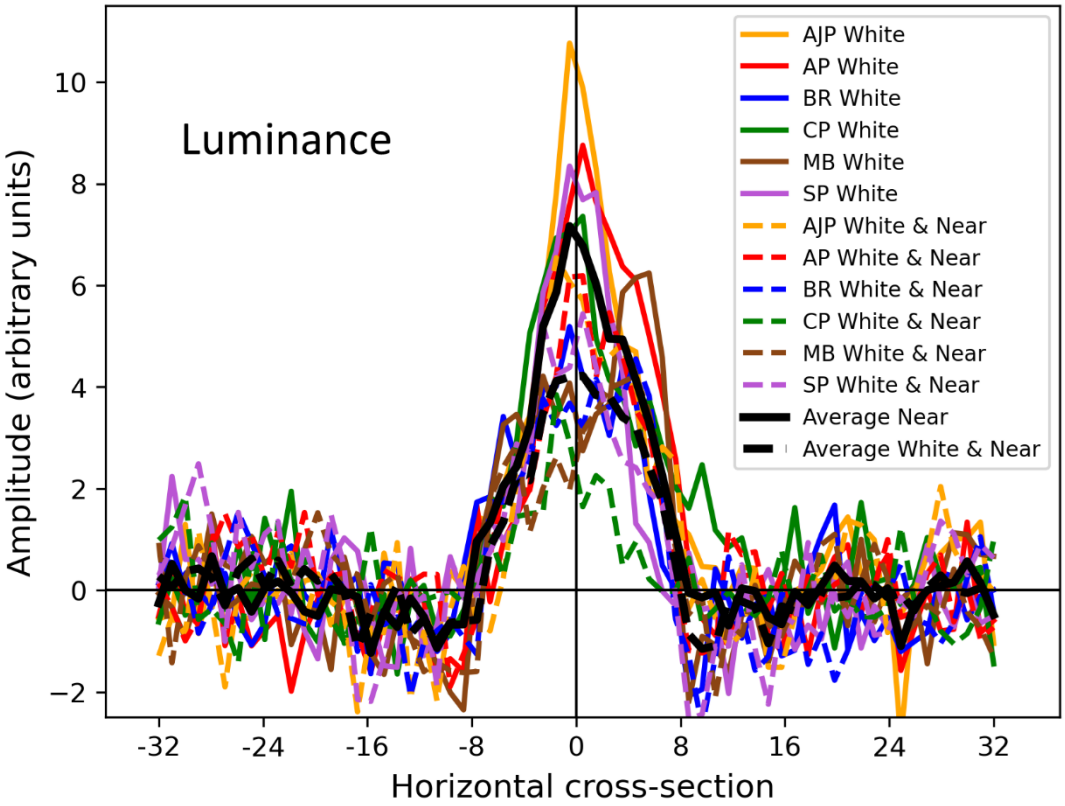

Figure 3.7: Horizontal cross-sections of luminance classification images similar to Figure 3.6.

3.3.4 Discussion

The results of this experiment satisfied the aims of producing simultaneous CIs from binocular disparity and luminance, as participants generally show the tendency of using both cues in the compound condition. The average result is that participants tended to sample each cue dimension less in the compound condition compared to the 'unidimensional' conditions (compare the solid and dashed black lines in Figure 3.6 and 3.7). This might seem to be an obvious effect, but the details regarding why and how this might have occurred can have two non-exclusive explanations: 1) differences in SNRs between the unidimensional and compound conditions, or 2) attention was solely focused on the one cue in the unidimensional condition, but shared among the two cues in the compound condition.

To expand on these different explanations: 1) The compound condition afforded two diagnostic cue dimensions rather than one, meaning easier detection consistent with some kind of probability summation, thus the staircase adjusted the SNR accordingly. Two cue sources afford lower detection thresholds: An example of summation is binocular summation where binocular contrast sensitivity is greater than monocular (Blake & Fox, 1973). In this condition, participants would start at a level estimated from the unidimensional conditions, but the staircase would typically reduce the SNRs of both cues to track a 75% threshold that would be suited for the compound stimulus. Such differences in SNRs could affect the CIs, as the noise had a different modulating effect on each cue dimension. 2) Differences in attentional focus between single-cue and two-cue conditions. Only the relevant cue was prioritized in the unidimensional conditions, but attentional weighting was shared between both cues in the compound condition, thus reducing priority of either one cue. This is another effect that could reduce the CIs in the compound condition, as cues must share priority. An extended study with further conditions, including controls for SNR differences between the compound and unidimensional conditions, could provide details for how cue sampling differed across the unidimensional and compound conditions.

The CIs show visual strategies to sample both cues simultaneously (Figure 3.5, 3.6 and 3.7), but some uncertainty remains regarding how this was achieved. Possible strategies that could explain these results include: 1) Cue combination within sessions, where 'White and Near' were weighted together on all trials, but one cue could dominate in some trials if sufficiently persuasive. 2) Cue disjunction within a session, where 'White' and 'Near' detection was prioritized separately on different trials, i.e., participants switched between cues but did usually not combine them within a session. A follow-up study might discover more details about such strategies in the compound stimulus condition with the addition of confidence ratings for each cue dimension, or with a three-

alternative response design where participants are afforded the response options: 'White', 'Near', or 'White and Near' to the compound stimuli.

Furthermore, this experiment suffered from an error in the staircasing of the luminance target signal for some participants. The staircase would sometimes reach a ceiling for some participants in some parts of sessions where they required a higher contrast luminance target than the software would allow. This would result in an underestimation of the participant's luminance target threshold, biasing the compound stimulus so that the disparity target was more detectable than the luminance target. For example, this occurred with participant CP, who used relatively more disparity than luminance cues in the compound condition (Figure 3.5), likely because the disparity target was more detectable than the luminance target owing to this error in the experimental software.

3.4 Conclusions on classification images

These three pilot experiments provided insights and practical guidance for directing the use of the CI technique in the two following chapters. Pilot 1 showed that the SIBR task design affords a faster, yet effective, method for capturing CIs compared to the 2AFC design. Pilot 2 validated a novel version of stereoscopic CIs which uses dense stereogram textures with binocular disparity noise. Pilot 3 showed a novel result that luminance and disparity CIs can be captured simultaneously with these stimulus images.

In discussion of Pilot 3, two additional experiments are proposed for uncovering more details about how participants were simultaneously able to sample both cue dimensions in the compound condition. These follow-up experiments were not conducted. Instead, priority was designated to applying these CIs to the main aims of this thesis – studying mechanisms involved in expertise for remote sensing surveying. The following chapter describes a study using this 2D and 3D CI method to study expert-novice differences in visual information sampling from stereoscopic aerial images.

In conclusion, this series of pilot experiments provide practical guidance for using the CI method in the later chapters of this thesis. These pilot experiments provided insights into the risks of experimental flaws to be avoided in the later studies. They also show that the method can provide the desired outcome, and the two following chapters develop this novel method further and apply it to meet the research aims of this thesis. These pilot experiments together provide a novel method that can simultaneously estimate CIs from binocular disparity and luminance cues. The benefits of this novel method are discussed in all the following chapters, where it is used to bring new insights into the mechanisms associated with visual information sampling.

3.5 Post-hoc analysis of perceptual learning for disparity targets

3.5.1 Aims

Remote sensing surveyors learn and improve from their experience with stereoscopic aerial images. An aim of this thesis sought to characterise stereoscopic PL, where experience with stereogram imagery might improve processing of disparity targets. Later in this thesis, Chapter 4 shows indirect evidence of PL for disparity targets in an expert-novice comparison. Following this study, Chapter 5 directly attempted to characterise learning for disparity targets. As two pilot experiments in the current chapter (Pilot 2 and 3) used stereograms with disparity targets, a supplementary analysis of these data was of interest to discover if the participants showed any evidence of PL throughout the pilot experiments. Here the level of external noise required to maintain fixed thresholds throughout the experiment is an indication of PL.

3.5.2 Results and discussion

In Pilot 2, which captured disparity CIs with three participants (Figure 3.3), participants required different levels of external noise to provide a sufficient masking effect to maintain the 70.7% threshold, as seen in Figure 3.8. This relates to the template amplitude differences seen in Figure 3.4, where the participant with the strongest CI template also required the most external noise (red curves), and vice versa with the participant who produced the lowest CI amplitude (blue curves). This can be interpreted as individual differences in the ability to sample disparity cues leading to differences in both CI and noise-threshold measures.

In this experiment (Pilot 2), two participants demonstrated evidence of PL in Figure 3.8. On the y-axis, disparity noise range indicates the average level of noise in each session. Note that, as the adaptive staircase adjusted the level of external noise, and not the target contrast, evidence of learning should be seen with increasing rather than decreasing thresholds in Figure 3.8. Increased noise implies a lower SNR and thus greater ability to detect the target signal. As the step sizes were linear in this experiment, the y-axis scale is also displayed in linear units. These data were fitted with linear regression models to examine significant slopes that could reveal improvements from PL. Two participants (US and KP) required increasing levels of external noise as the experiment progressed (Figure 3.8). Furthermore, fitting the linear regression model to the average of these three participants also produced a modest but highly significant slope (slope = 0.16, R^2 = 0.576, t = 4.94, p < 0.001). This effect is consistent with PL, where learning improves the ability to detect the target, thus requiring more external noise to maintain the threshold.

PL may be considered surprising given that no feedback was provided, and the participants who showed significant PL completed the experiment in only two and three days. Feedback and division of the experiment across multiple days (usually over four days), with sleep consolidating learning, is known to contribute to PL (Aberg & Herzog, 2012; Herzog & Fahle, 1997; Karni et al., 1994; Liu, Dosher & Lu, 2014; Sasaki, Nanez & Watanabe, 2010). But PL can occur without structured trial-by-trial feedback (Liu, Lu & Dosher, 2010, 2012; Lu et al., 2011), as seen in this experiment which was not specifically designed to induce PL.

In Pilot 3, PL in the disparity condition was analysed with a similar method, and the results are shown in Figure 3.9. This condition differed in three ways from Pilot 2: 1) The threshold tracked a 75% correct response rate rather than 70.7%, 2) it was half as long, with 5,000 instead of 10,000 trials, and 3) this disparity condition was interleaved with the luminance and compound conditions. The linear regression models show that the external noise required to maintain thresholds in the disparity condition increased across sessions for two participants (Figure 3.9). This is again consistent with PL for disparity cues in the absence of feedback. Fitting the average of these data with the linear regression model produced a positive but non-significant average slope in this experiment (slope = 0.12, R^2 = 0.178, t = 1.31, p = 0.225).

Regarding luminance thresholds in the luminance condition (Pilot 3), no significant slopes were observed across sessions, suggesting no PL for detecting the white square (data not shown). No PL for the luminance target was expected, and can be understood in terms of luminance contrast judgements being primitive and close to the sensory origins of the visual system. No PL was observed because most humans are already ‘experts’ at detecting intensity changes (Adini, Sagi & Tsodyks, 2002; Dorais & Sagi, 1997; Westheimer, 2001).

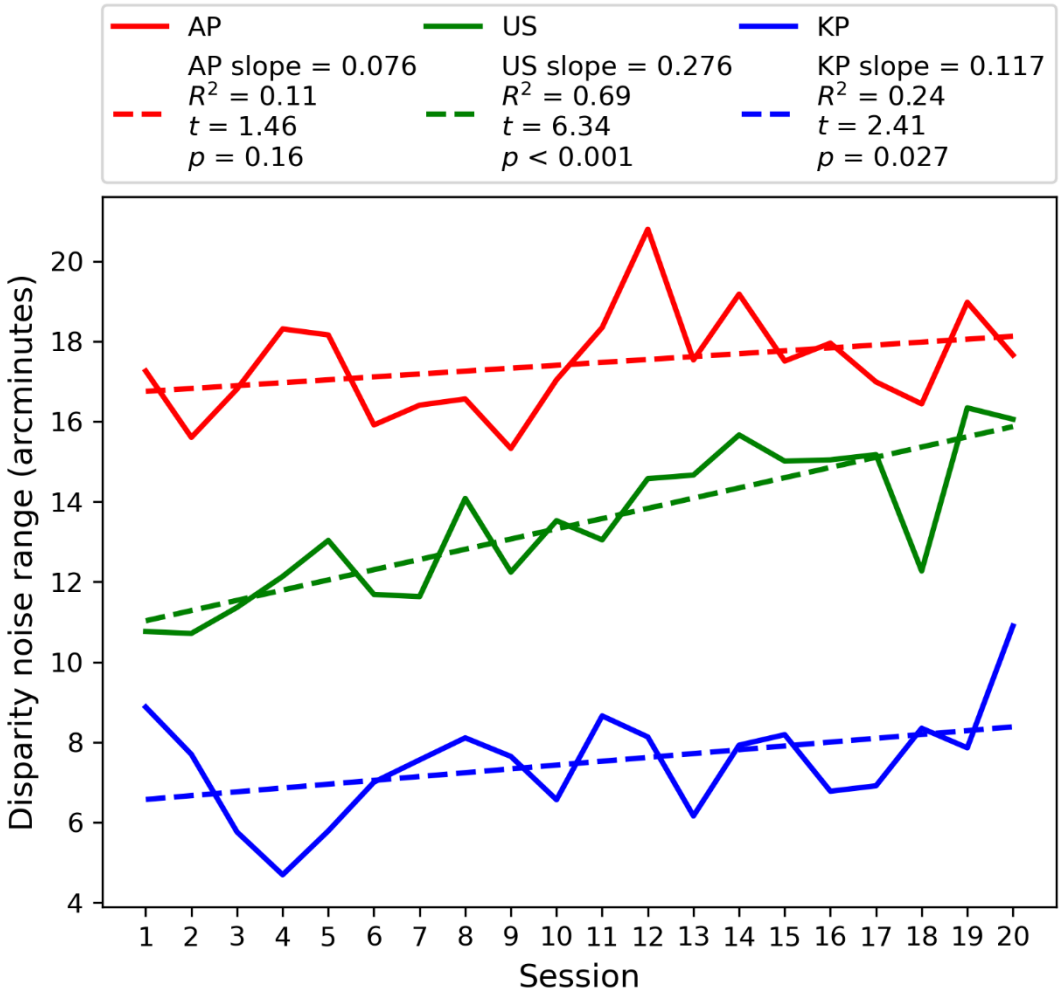

Figure 3.8: External noise required to maintain 70.7% thresholds for the three participants in Pilot 2. Solid curves show the mean level of added disparity noise across each session (500 trials). Dashed lines show a linear regression model fit to each participant's data (slopes and statistical outcomes indicated in legend). Disparity noise range indicates the highest possible crossed and uncrossed disparities in the noise, in arcminutes.

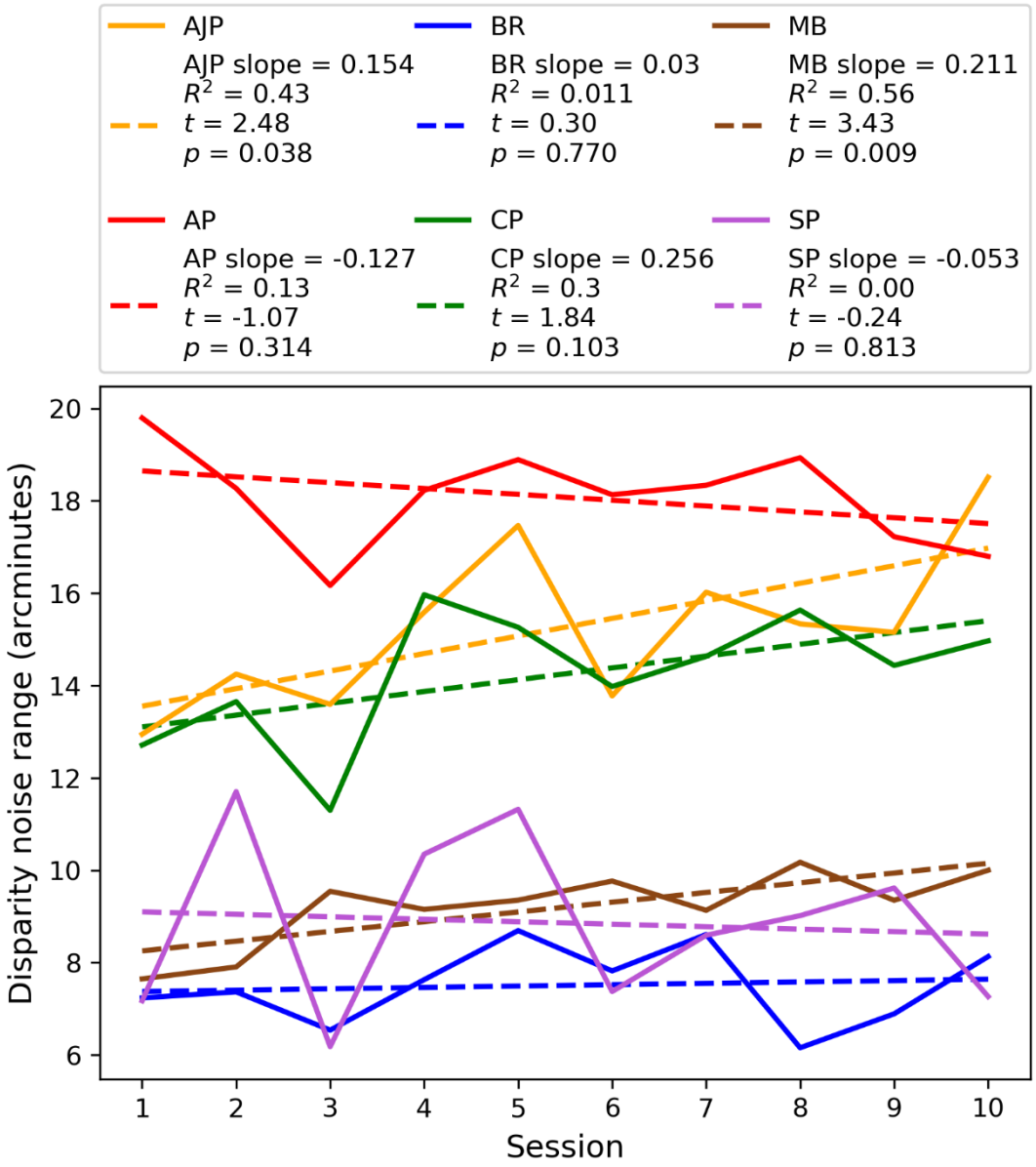

Figure 3.9: External disparity noise required to maintain 75% threshold in the disparity condition in Pilot 3, similar to Figure 3.9. Participant AP also participated in Pilot 2.

3.5.3 Conclusions

The results show stereoscopic PL for disparity targets in four out of eight participants (Figure 3.8 and 3.9) (AP participated in both experiments, and is thus not counted in the second). PL occurred despite no provision of feedback, and with a small number of participation days. These results suggest that experience can be important for improving the ability to process stereograms, consistent with well-established literature (e.g., Ramachandran, 1976; see also Chapter 5 for further discussion of relevant literature).

The results of this post-hoc analysis of PL strengthen the motivation for the study in Chapter 5, as these results suggest that PL can be present for this task. Chapter 5 directly aimed to induce PL

for disparity targets in stereograms. Further details regarding stereoscopic PL are discussed in Chapter 5. Furthermore, the potential future directions and implications of PL interventions for remote sensing surveying are discussed in Chapter 6.

Chapter 4

Classification images for aerial images capture visual expertise for binocular disparity and a prior for lighting from above

Collaboration acknowledgement

The author wishes to acknowledge that much of this chapter was written in collaboration with Andrew J. Schofield and Timothy S. Meese in preparation for a joint publication. The introduction, method and results for the main experiment described below were written in collaboration, whereas the general discussion and follow-up experiment were written more independently.

4.1 Introduction

Continuing with the CI method developed in Chapter 3, the current study aimed to compare visual strategies of novices and expert remote sensing surveyors. This was examined with the task of discriminating two landscape features that have similar visual textures but dissimilar 3D relief: hedges and ditches. In principle, these can be discriminated from luminance and/or binocular disparity cues. The details of these depth cues are further considered below.

The experimental approach for this investigation is novel and based on the pilot studies in Chapter 3. By imposing spatial noise made from luminance textures and random binocular disparities onto stereoscopic landscape images, simultaneous pairs of CIs were generated for each observer. By examining and quantifying these, the analysis established how observers used disparity and luminance cues when performing hedge/ditch classifications. The image treatments involved a 2x2 manipulation which flipped: 1) the disparity of half the images (to produce pseudoscopic viewing), so that hedges had ditch-like disparity profiles and vice-versa, and 2) the orientation of half the images (mirror-reversed around a horizontal axis) to change the lighting and shading cues (see below). The prediction was that experts would make more use of disparity cues than novices, and thus have more clearly defined disparity CIs, for two reasons. First, an informal preliminary report from a very experienced remote sensing instructor at the OS, advised that hedges and ditches are typically identified according to their perceived stereoscopic relief (i.e., their 3D quality from binocular disparity cues). Second, the expert surveyors who participated in the current study were more experienced than novices in making photogrammetric judgements involving disparity cues. Even so, participants were also expected to combine disparity and luminance cues instead of completely ignoring one of them because cue combination tends to support stronger stereoscopic perception (Doorschot, Kappers & Koenderink 2001; Hartle et al., 2022; Lovell, Bloj & Harris, 2012; but see Chen & Tyler, 2015 where luminance cues made disparity cues redundant).

Regarding the form of the CI templates, crossed and uncrossed disparities were expected to promote 'hedge' and 'ditch' responses, respectively, regardless of the ground truth of the image owing to the unambiguous 3D relief of tall hedges and deep ditches in the real world. Luminance was also expected to be an influential factor because luminance contrast is important for depth perception (Egusa, 1983; O'Shea, Blackburn & Ono, 1994) under two different assumptions about shape from shading. On the assumption of 'diffuse lighting', surface peaks and troughs align with light and dark image regions, leading to the perception that 'dark-is-deep' (Langer & Zucker, 1994; Langer and Bülthoff, 2000; Schofield, Rock & Georgeson, 2011; Sun & Schofield, 2012). On the assumption of 'punctate lighting', a single-point light source means luminance peaks are perceived as surfaces facing the light source, such as a hedge with a highlight on the side facing the sun (Adams, Graf & Ernst, 2004; Berbaum, Bever & Chung, 1983; Brewster, 1826; Koenderink et al., 2003; Pont, van Doorn & Koenderink, 2017; Ramachandran, 1988; Rittenhouse, 1786; Schofield, Rock & Georgeson, 2011; Sun & Perona, 1998; Sun & Schofield, 2012). These assumptions invoke subtly different relationships between luminance and shape. In the experiment, a diffuse lighting prior predicts a strategy of 'hedges are light, ditches are dark' (dark-is-deep), with luminance peaks (hedges) and troughs (ditches) aligned with the centre of the landscape feature. On the other hand, if the lighting is assumed to be punctate then this predicts luminance peaks that are offset from the centre of the landscape feature in the direction of the assumed light source. For example, consider an observer who assumes lighting-from-above, meaning light coming from the top of the 2D image plane. (Note that to avoid confusion of terminology with top/bottom and above/below in 3D, this direction will be referred to as 'north', meaning the top of the page regardless of what a compass would say). This observer would expect convex hedges to have a highlight towards the 'northern' part of the feature, with darker luminance in the 'southern' part, representing a shadow or internal shading. Similarly, this hypothetical observer would expect a concave ditch lit from the 'north' to be lighter towards the 'south' of the feature, as light would not reach the 'northern' concave region owing to surface depth occlusion. As will be seen, these asymmetries are important for the details of the luminance CIs.

The predicted outcome under the punctate lighting hypothesis is complicated further by the OS's practice of presenting aerial imagery with geographical north at the top of the image, consistent with most geographical maps. However, in the UK the sun shines predominantly from the south. This produces aerial images that are lit from the 'south', in this case meaning from the bottom of the page/screen. Expert remote sensing surveyors are thus accustomed to viewing aerial images as if lit from below their line of sight, which conflicts with a well-known bias in the population known as the lighting-from-above prior (Adams, Graf & Ernst, 2004; Ramachandran, 1988; Sun & Perona, 1998). As

the experts have spent many years working with aerial imagery lit this way, these natural lighting biases might have diminished, or even switched direction. No predictions were made on whether either of the two lighting assumptions (punctate or diffuse) would dominate in the experiment, accepting that both might be seen. Under the punctate lighting prior, the luminance peaks in the CI were expected to be offset 'north' of the perceived centre line of hedges for novices, as per the conventional prior, but 'south' of them, or with smaller offsets, for the experts. Similarly, under this hypothesis, bi-lobed luminance CIs were expected for the reasons to do with lighting and shading outlined above. More generally, because novices and experts have potentially different priors, the two groups were expected to have different sensitivities to image orientation (lighting direction in the hedge and ditch images) and to have qualitatively different luminance profiles in their classification images.

4.1.1 Overview and hypotheses

A novel variant of the CI technique was introduced, designed to provide simultaneous estimation of luminance and disparity templates (see Chapter 3 for the development of this method). This was employed for a feature identification task using aerial images to address the research questions developed above and summarised below as five specific hypotheses. Largely, these are about expected differences between experts and novices and are listed here to scaffold the results. However, observations and conclusions do extend beyond these *a priori* expectations.

Hypothesis 1 (stereo, CI): The stereo fidelity hypothesis: Compared to novices, experts will be better in sampling relevant information from stereoscopic aerial images. This will be shown by greater amplitudes and greater spatial extents of the disparity CIs for experts compared to novices.

Hypothesis 2 (stereo, CI): The cue strategy hypothesis: it is possible that experts and/or novices will prioritize one type of noise-cue (e.g. disparity cues) over the other type (e.g. luminance cues). This would be demonstrated by greater amplitudes and/or spreads across the two types of CI.

Hypothesis 3 (stereo, categorical): The stereo accuracy hypothesis: Compared to novices, experts will show greater sensitivity to the stereoscopic profiles of targets, as revealed by their accuracy in the categorical ratings.

Hypothesis 4 (lighting, categorical): The lighting sensitivity hypothesis: regardless of their CI structures, experts and novices will show different sensitivities to lighting direction, with novices having a greater tendency to respond according to an assumption of lighting-from-above.

Hypothesis 5 (lighting, CI): The lighting bias hypothesis: Compared to novices, experts will show different or diminished lighting direction biases in their CIs and by this token, novices will show

a greater tendency to lighting-from-above compared to experts. A relationship is expected between this and the lighting sensitivity hypothesis (H4) outlined above.

4.2 Methods

4.2.1 Visual stimuli and the image generation pipeline

High-resolution aerial-view landscape photographs covering land areas of approximately 2.5km x 1.5km were sourced from the OS. Stereogram pairs were created using two images that covered overlapping landscape areas, spaced apart along the aircraft's flight path. Six landscape features were isolated: three landscape features were hedges, found in Cambridgeshire, UK, and three features were ditches, found in Somerset, UK. Features were selected based on the following criteria: 1) The levels of shadow/sunlight were moderate. 2) The features were horizontally aligned within 15° of the aircraft's flight path to facilitate horizontal binocular disparity. 3) The features were of a similar vertical extent and spread across the width of the image segment selected. 4) The features were straight. 5) The features had usable stereoscopic information: shallow ditches were excluded[7].

The six image pairs were processed with MATLAB and Python to create landscape stimuli. Each image was: 1) rotated to horizontal alignment using bicubic interpolation (Mean rotation: 7.35°, range 0–14.5°); 2) resized so that features had the same vertical extent using bicubic interpolation (Mean scale factor: 0.85, range: 0.52–1.2); 3) linearized to undo a compressive nonlinearity applied in the OS image pipeline; 4) converted to grayscale using Equation 1:

$$L = 0.2125 * R + 0.7152 * G + 0.0722 * B, \qquad \text{(Equation 1)}$$

where *L* = luminance, *R* = red colour channel, *G* = green colour channel, *B* = blue colour channel; 5) cropped to 128x128 pixels; 6) standardized to have the same mean luminance and average root-mean-square contrast as the 12-image set. The images were processed and stored at 16-bit greyscale resolution throughout to prevent losses. These transformations were designed to produce horizontally oriented target features of similar sizes while removing colour and luminance variations in the original photographs that may have varied due to the feature types being sourced at different locations, time of year, and time of day. Figure 4.1 shows the final images used in the study.

[7] The exact quantity of binocular disparities in the features are unknown due to the unavailability of data on camera separation and height from the ground.

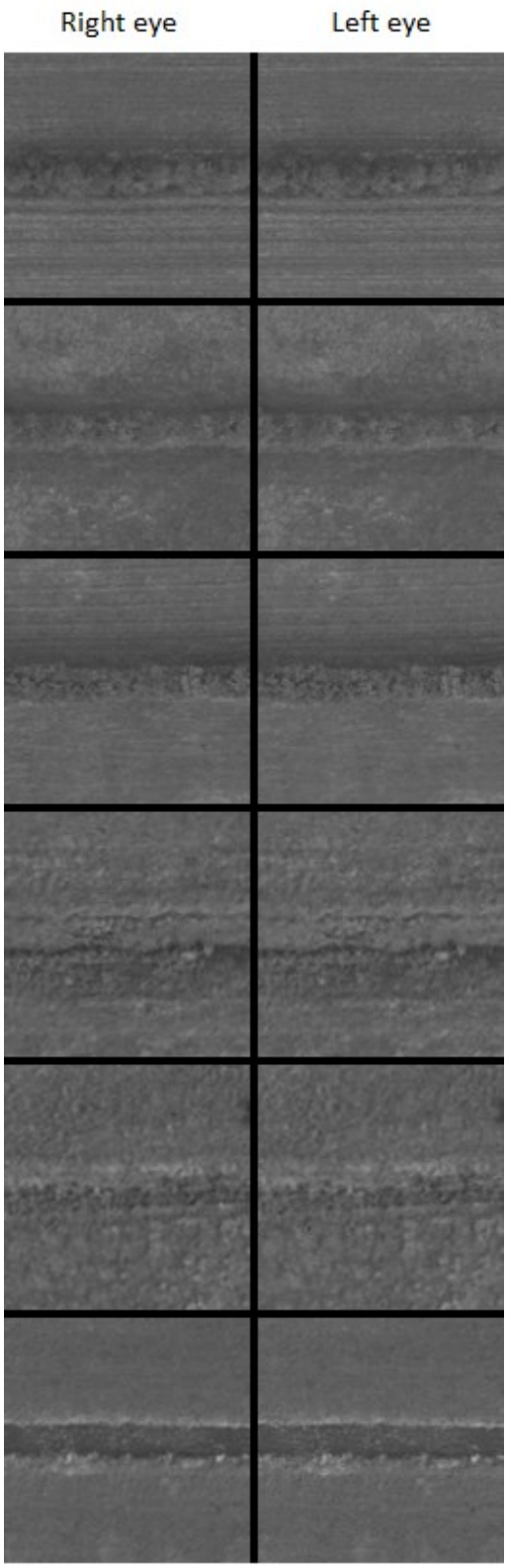

Figure 4.1: The landscape features used to make stimulus images (see Figure 4.3) after the minor rotation needed to achieve a visually horizontal feature. The top and bottom three pairs are hedges and ditches, respectively. The images are stereogram pairs arranged for crossed free-fusion, and divergent fusion reverses disparities. These images are shown with geographic north at the top of each image as per OS practice (see text for details). In these images, the terms 'north' and 'south' refer to the upper and lower halves of the images (and their parts), respectively, and also, to a first approximation, the compass. Notice subtle lighting cues owing to sunlight originating as if from the 'south' in these images. *© Crown copyright and database rights 2023 OS*, used with permission.

4.2.1.1 Dual-noise classification images

In a novel step, both luminance and disparity noise was imposed onto the test images allowing the simultaneous estimation of luminance and disparity CIs (see also Chapter 3). A unique white noise texture (range ±1.0, 128x128 pixels) with randomly varying, non-zero mean was generated on each trial. This texture was low pass filtered with a first-order Butterworth filter with a cut-off frequency of 9 cycles per image. The noise texture was then added to the two landscape images in each stereoscopic pair to create noise+feature images (noise and image contrasts were normalised to 35% and 65% of their original contrasts, respectively).

Another low pass filtered noise texture (with properties as above) was generated on each trial to create a random disparity map describing disparity offsets (see Figure 4.2 for an example image). Pixels in the two noise+feature images (one for each eye) were displaced horizontally by an amount determined by the random disparity map, thus adding disparity noise. Each image in the stereo pair bore half the required shift so that the images for the two eyes were transformed equally but in opposite directions. When presented in the stereoscope (described below) this produced a range of 0-296 arcseconds of random disparity (quantised to 8 levels) in the stimulus image pair and required sub-pixel shifts in the position of each pixel in the noise+feature images.

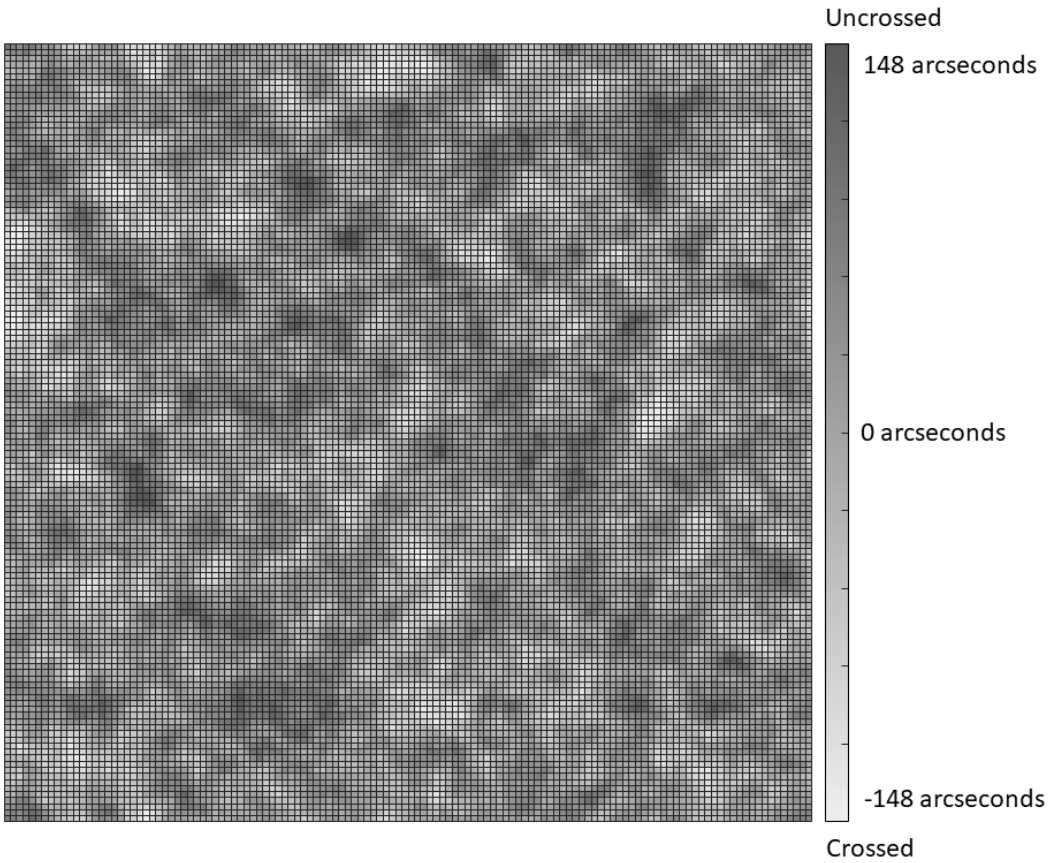

Figure 4.2: Example Z-coordinate texture used to map random disparities.

Figure 4.3 illustrates the procedure for adding noise to the stimulus images, including the process for producing sub-pixel shifts, which was similar to the one used by Georgeson, Yates & Schofield (2009). Each noise+feature image was first upsampled in the horizontal direction by a factor of 10 to produce a 128x1280-element image. The upsampled luminance elements were then displaced based on values taken from the equivalent location in the disparity map (Figure 4.2). The amount of displacement applied varied horizontally across the image meaning that two or more luminance elements in the original image could be displaced to the same location in the transformed image. To address this problem, the competing luminance element that was subject to the least crossed disparity (i.e., the one that would appear furthest from the observer) was discarded and only the element subject to the most crossed disparity (i.e., closest to the observer) was retained. The disparity shifts could also result in gaps where no luminance element was assigned to a location in the transformed image. These gaps were filled with random luminance values sampled from a white noise texture. The image array was then downsampled in the horizontal direction by averaging, thereby recreating the original image resolution. Where determined by the disparity map, this procedure resulted in sub-pixel disparity shifts by virtue of subtle variations of luminance between the two eyes such that the 'centre of mass' of the grey values comprising features in the stereo pair was subtly shifted in each eye. Note that the random/noisy disparity shifts were applied in addition to the existing disparities between features in the original landscape images. Thus, the original disparities were retained but were heavily distorted by the disparity map analogous to the distortion of the original luminance features by the luminance noise.

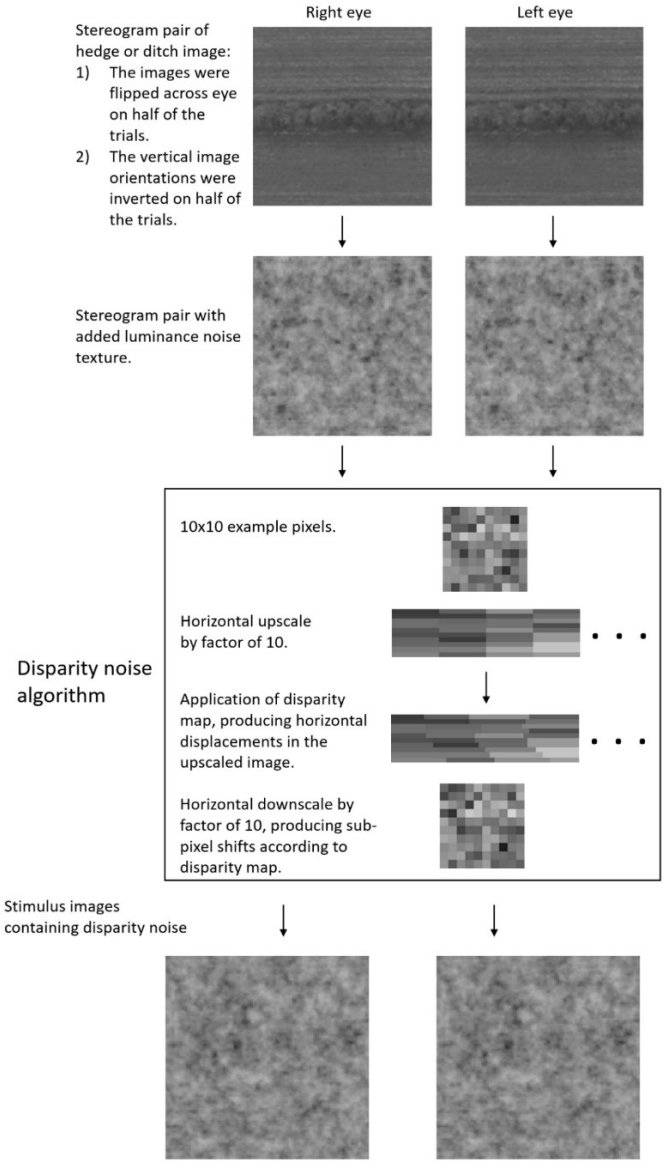

Figure 4.3: Dual-noise procedure for making stimulus images on each trial using the hedge and ditch images in Figure 4.1. Paired images are arranged for crossed fusion with the low signal-to-noise ratios used in the experiment. This means that the cross-fusing reader is unlikely to witness much meaningful signal. See text for further details. Top image pair, *© Crown copyright and database rights 2023 OS*, used with permission.

Finally, the inverse gamma functions of the monitors were applied to the stereo image pairs to ensure that luminance was linear for the displays. The bottom part of Figure 4.3 shows an example stimulus pair. Stimulus images were intentionally masked heavily with both luminance and disparity noise because the CI technique benefits from the strong influence of noise on behavioural responses.

4.2.1.2 Disparity and lighting direction

Before applying the luminance and disparity noise described above the stimuli were treated in each of two ways. In one manipulation, the landscape images were swapped between the two eyes, so that the disparity of the hedge or ditch was inverted and incongruent. A hedge image thus

changed from having substantially crossed disparity to substantially uncrossed disparity and appeared ditch-like, and vice versa for the ditch images. In a second manipulation, the landscape images were inverted about their horizontal axis, maintaining horizontal disparities but inverting the spatial relations of light and dark image features. In principle, these features can provide cues to 3D relief from highlights and shading. For example, most people report that the middle hexagon in the honeycomb stimulus in Figure 4.4a looks like a bump while the same image region in Figure 4.4b, rotated by 180°, looks like a dimple (Andrews et al., 2013). See also Chapter 1 for further details on lighting direction priors for interpreting shape from shading. Underpinning these perceptions is an assumption that lighting comes from above. Therefore, image orientation might also influence the perception of hedges (convex) and ditches (concave) in the same way as the honeycomb (Figure 4.4).

Each of the two image treatments was performed on a trial-by-trial basis with an independent probability of 50%. This created a stimulus set with an overall 2x2x2 design (hedge/ditch; correct/inverted disparity; original/inverted orientation).

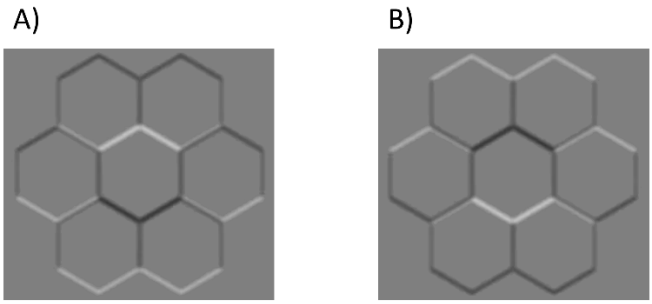

Figure 4.4: Honeycomb stimulus, as previously seen in Chapter 1, where the light and dark parts of the image can be interpreted as highlights and shading owing to 3D shape. (A) is the same as (B) but rotated by 180°. These images are included to demonstrate the point about lighting direction and the perception of 3D shape, and were not used in this experiment.

4.2.2 Equipment

Participants were seated in a dimly lit and secluded room with their chins on a chinrest in front of a mirror stereoscope. The monitors provided the primary light source in the room, apart from in the testing room in Southampton (see below), where window blinds allowed low levels of diffuse daylight to enter the room. This light applied equally to both viewpoints in the stereoscope. Two front-surface mirrors angled at 45° were mounted 6cm in front of the participant. These directed images to the observer from two ASUS ProArt PA329C monitors (3840 x 2160 pixel, 710x405 mm active screen region) placed on either side of the mirror mount with a total viewing distance of 990 mm. Each monitor pixel subtended ~0.01 degrees of arc. Images were scaled in PsychoPy (Version 2020.2.10; Peirce et al., 2019) so that a single element from a stimulus image occupied 5x5 pixels on the monitors. Thus, images subtended 6.58 degrees of visual angle. Apart from the pre-processing

noted above, stimuli were generated and presented using PsychoPy with a modified version of the noise component.

4.2.3 Participants and ethical considerations

Twelve participants (Mean age: 38.7, range: 23-62) were recruited by targeted email advertisement or direct communication. Participants were categorized as experts or novices depending on their level of experience with remote sensing surveying. An expert was defined as someone with two or more years of experience with remote sensing photogrammetric tasks. A novice was someone with no experience with remote sensing photogrammetry. Six participants were experts (Mean age: 43.8, range: 23-62) and employees at the OS with an average of 8 years of experience (range 2-20 years). The six novices (Mean age: 33.5, range: 25-45) comprised two non-surveying staff at the OS, one staff member at Aston University, and three PhD students. The eight OS employees were tested at their offices in Southampton, UK, and the other four participants were tested on the Aston University campus, Birmingham, UK. Both groups had an average of 4 years of completed university-level education. No participant was experienced in creating or participating in psychology or psychophysics studies. Participants gave informed consent and were compensated with payment at a rate of £10 an hour. All participants were assured that their data, including screening data, would be confidential and anonymised. The project was reviewed by Aston University's College of Health and Life Sciences Ethical Review committee (approval number 1843).

4.2.4 Screening and exclusion procedure

A screening procedure assessed the eyesight and binocular stereopsis of each potential participant for the main experiment. Participants wore their normal optical correction where appropriate. They were tested for standard visual acuity using a Snellen test and given a 'gold standard' (Garnham & Sloper, 2006) TNO test for stereoscopic vision, based on random-dot-stereograms that provide no monocular cues to the target.

The results of the TNO test are shown in Table 4.1 and are within normative bounds of a sample of 1058 participants who had a median TNO stereoacuity of 60 arcseconds (Bosten et al, 2015). No exclusion criterion was set for this test. Expert 5's relatively high TNO threshold will be discussed later in the chapter, but note that Bosten et al (2015) reported that 8.9% of their sample had a TNO stereoacuity measure of ≥480 arcseconds.

	TNO threshold	
Participant	Experts	Novices
1	15	120
2	30	60
3	30	15
4	30	30
5	450	15
6	30	60

Table 4.1: TNO thresholds for all 12 observers who took part in the main experiment. TNO threshold describes stereoacuity threshold in arcseconds from the TNO test.

Participants were familiarised with the stereoscope by observing ten images that contained either a 'flat' texture or a stereoscopic texture with a square target defined by crossed disparity. Participants then carried out a discrimination task (40 trials) where a central disparity-defined square (740 arcseconds of disparity, side length 1.44 degrees of visual angle) had either crossed or uncrossed disparities. The task was to report whether the square was in a 'near' or 'far' depth plane compared to the surround. Responses were made by pressing a button on a keyboard. Participants had to score above 90% correct to pass this test. Those who failed were thanked for their time and given £5. Two potential novice participants and no experts were excluded by this process.

4.2.5 Experimental procedure

4.2.5.1 Preliminary procedure: general familiarity

To familiarise all participants with the concept of aerial stereoscopic imagery, they were shown the same ground view photograph of two houses followed by an OS aerial stereogram pair of the same houses viewed through the stereoscope. They were told that these were different views of the same scene, the second one from above, and that they would be viewing aerial images containing stereoscopic depth like the houses but showing hedges and ditches. Participants were then shown ground-view images of a hedge and a ditch and told they would be looking for these features but from an aerial perspective. They were not shown any aerial-perspective images of hedges and ditches as part of the familiarisation procedure.

4.2.5.2 Preliminary procedure: familiarity and instructions

At the start of their first experimental session, participants practised for ~20 trials under the supervision of the experimenter following the main procedure below. They were instructed to press the left and right arrow keys on a keyboard for 'ditch' and 'hedge' responses, respectively. While

these button press responses were not 'yes or no', they follow a SIBR design as the task design uses one image and a binary choice per trial. Participants were told that the task would often feel very difficult, but they should make their best estimates to find hedges and ditches. They were also told that various hedge and ditch stimuli would be presented, and that these were always in the same location and had the same size. The experimenter gestured with his hand to highlight the shape outline and the size of the hedges and ditches over the monitor.

To ensure appropriate vergence control, between each trial a black fixation cross was presented in the centre of the screen. The vertical bar of the cross was split across the two eyes. To achieve good convergence, participants were instructed to fuse the cross to make it appear 'complete', like a '+'. If the cross appeared to drift apart, participants were instructed to close their eyes or look away for a moment to 'reset' their convergence and on returning attention to the display, to wait until the cross appeared fused before making their response which would also start the next trial. To further aid fusion, a high contrast border featuring white rectangles on a black background surrounded the image presented to each eye.

4.2.5.3 Main procedure

Stimuli were presented for 750ms and participants were allowed unlimited response time. No feedback was provided. A response triggered a new trial after a 630ms delay. The high contrast border surround and the fixation cross were always present, except the fixation cross was removed when the stimuli were displayed.

The stimuli were presented with congruent (original) or incongruent (inverted) disparities and original or inverted orientation (see above) with equal probability on each trial. The third stimulus factor (original target feature = hedge or ditch) was blocked (i.e., a block of trials contained either only original hedges or only original ditches) but participants were not informed of this. This was done so that differences in visual textures across hedge and ditch images that were extraneous to 3D feature identification (e.g., the type of grassland in the scene) could not influence decisions within each block. The disparity reversals applied within each block ensured that the stimuli were presented as hedge-like and ditch-like with equal probability on each trial regardless of the block type. Sessions alternated between blocks of hedge and ditch targets, with the starting order counterbalanced across participants. Each session (block) contained 500 trials and there were 20 sessions lasting about 15 minutes each giving a total of 10,000 trails per participant. Breaks were permitted between sessions and sometimes this was overnight. Eleven participants completed the experiment over three days, for one it took four days. The total experimental time for each participant was about six hours.

Following their final session, participants were asked to describe what they had been looking for when deciding whether stimuli were hedges or ditches. Participants were also asked whether they were aware of sunlight and/or shading influencing their decisions.

4.3 Results and discussion

4.3.1 Debriefing

Expert 3 and Novices 2 and 4, reported a decisive luminance strategy supposing that hedges would look lighter and ditches darker (i.e., a 'dark is deep' strategy). Five out of six experts and four out of six novices reported using stereoscopic depth as a primary strategy. These five experts also stated that luminance cues (consistent with 'dark is deep') could be used mainly as a secondary strategy, but occasionally this would become primary. No participant mentioned that sunlight direction (from above or below) or shadow location influenced their decisions when quizzed about the direction of the light source. This suggests that participants were unaware of using a punctate lighting assumption.

4.3.2 Organisation of main results

Results will be described in the order of the hypotheses set out in the Introduction, which is to deal mainly with issues around disparity first followed by those around lighting and luminance. This involves starting with CIs, progressing to the details of the categorical data, then returning to CIs. However, the section begins with overall observations of the CIs, followed by a description of how CIs were quantified, before turning to interpretations and the five hypotheses.

4.3.3 Classification images and informal observations

Luminance noise textures were treated as the average across the stereo pair as presented to the participants on each trial. Disparity noise textures were the Z-coordinate maps for depth (see Figure 4.2), where light pixels represent crossed ('near') disparity, and dark pixels represent uncrossed ('far') disparity. (The convention here means that the CI grey levels relate to implied 3D relief in the same way for both types of CI.) For each participant, noise textures for luminance and disparity were tagged according to the 'hedge' or 'ditch' response on each trial and compound images were generated by summing the images for each tag. To generate a CI from all 10,000 trials, 'ditch' response compounds were subtracted from 'hedge' response compounds (Ahumada 1996; Murray, 2011). Figure 4.5 shows CIs for each of the twelve participants revealing individual decision templates for luminance and disparity.

The higher contrast CIs tend to be in the left and right columns for the experts and novices respectively (H1 & H2), suggesting that different classification strategies were being used across the two groups. The centre regions of the disparity CIs in Figure 4.5 are white, indicating that 'hedge' responses were promoted by crossed disparity and 'ditch' responses by uncrossed disparity, as to be expected. Individual differences in lighting assumptions are observed in the luminance CIs consistent with the lighting bias hypothesis (H5), as follows. For some participants, the template centres are white, indicating that 'hedges' were promoted by lighter patterns and 'ditches' by darker patterns, consistent with a diffuse lighting assumption in shape from shading (e.g., Expert 3 and Novices 2 and 4). For other participants, the luminance CIs show centrally offset positive and negative peaks, consistent with the influence of a punctate lighting assumption on the identification of hedges and ditches (e.g., Novices 1, 3 and 5). Further discussion of individual differences is provided below.

From casual inspection of the partial CIs from each of the six original hedge and ditch images (Figure 4.1), no systematic differences were observed across the different images (results not shown), confirming that participants were consistent in their use of visual strategies across the six images. This was also the case for the two image manipulations of disparity congruency and vertical image inversions. These CIs, representing partial CI data split by the different image manipulations, can be found for disparity CIs in Appendix A and luminance CIs in Appendix B. They show that participants applied similar templates throughout the experiment regardless of per-trial image manipulations.

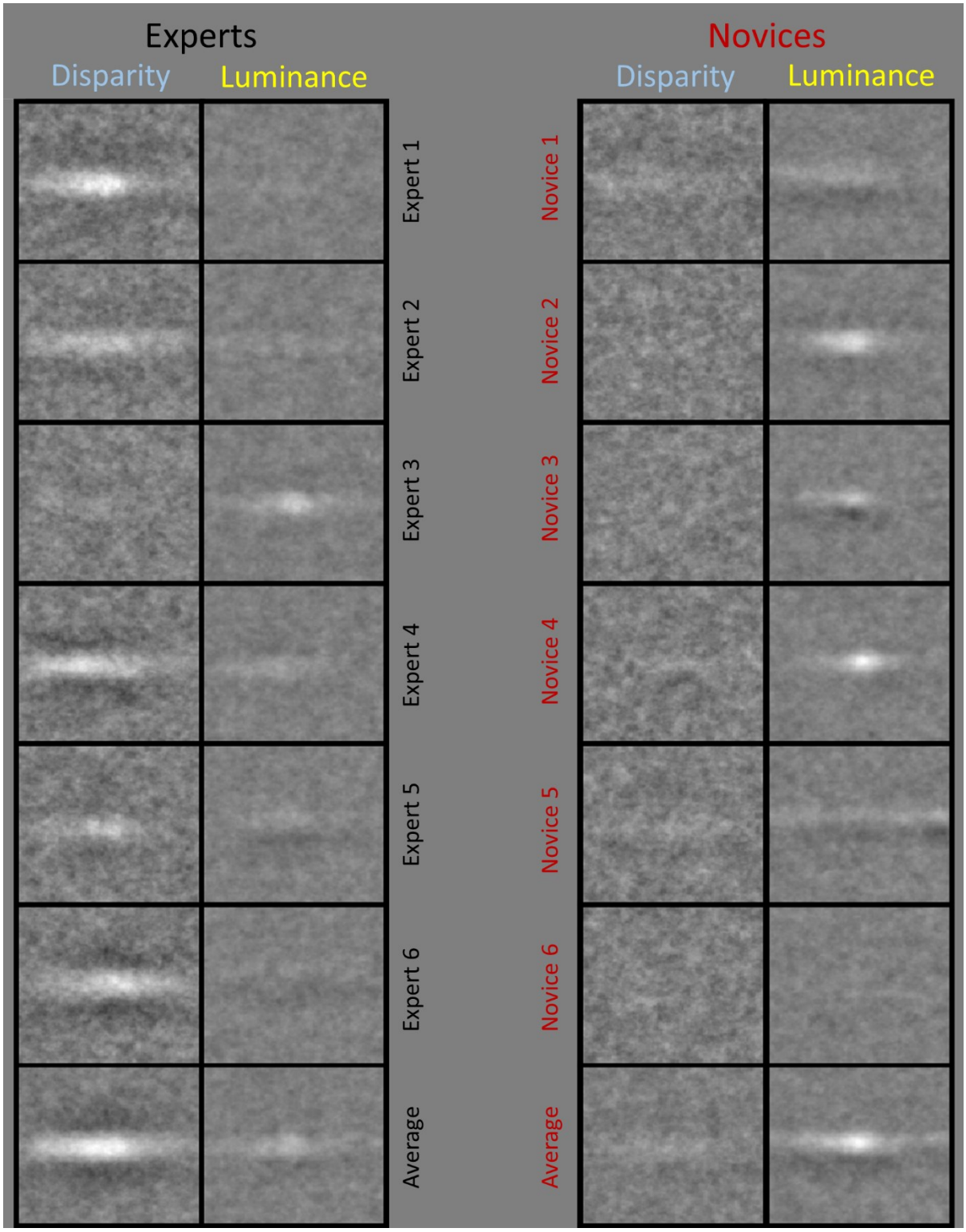

Figure 4.5: Classification images (CIs) for disparity and luminance where 'ditch' response textures were subtracted from 'hedge' response textures. For both types of classification image, lighter and darker pixels represent mounds and troughs, respectively. Thus, for the disparity CIs, the light regions derive from observers responding 'hedge' and 'ditch' when there were crossed and uncrossed disparities in those image regions, respectively. For the dark regions, the disparities were uncrossed and crossed, respectively. Similarly, for the luminance CIs, the light regions derive from observers responding 'hedge' and 'ditch' when there were light and dark pixels in those image regions, respectively. For the dark regions, the pixels were dark and light, respectively. The bottom of the figure shows group-average CIs.

4.3.4 Quantifying the CIs

To facilitate quantitative visualisation of the differences between the experts and novices, plots of cross-sections of the CIs for disparity (Figure 4.6) and luminance (Figure 4.7) are provided. Figure 4.6a shows the vertical cross-sections produced by averaging all the pixel columns in the disparity CIs (recall that the target features were horizontal) and has a profile like a Difference of Gaussians with a central positive lobe and two flanking negative lobes (sometimes called a 'Mexican hat'). Figure 4.6b shows the horizontal cross-sections produced by averaging the 20 central rows of pixels (±10 from the centre) corresponding with the target location. Only this central region was used for the horizontal cross-sections to avoid cancellations from the outer regions (Figure 4.6a) with opposite sign. Figure 4.6b reveals a curved profile with a peak in the centre of the CI. The solid red and black curves in Figure 4.6 are for the novices and experts, respectively, and illustrate large group differences for disparity.

The treatment of the luminance CIs in Figure 4.7 is similar to that for the disparity CIs in Figure 4.6, but the outcome is different. Figure 4.7a shows that all observers have distinct positive lobes and several also have negative lobes but often weighted more heavily on one side than the other. Furthermore, while several observers have central peaks, others have peaks offset from the centre, the most prominent of which are to the left. This corresponds with 'north' (up) in the stimuli, though in some cases the offset is in the other direction. These differences necessitated special treatment for the averaging in Figure 4.7b. As the negative lobes in Figure 4.7a were offset for some observers, they were prone to cancel the positive lobes when averaged across the central 20 rows. Therefore, to preserve amplitude, these rows were full-wave rectified before averaging.

The cross-sections defined above were characterised by fitting Gaussian (Equation 2) and Gabor (Equation 3) functions to the horizontal and vertical cross-sections, respectively:

$$f(x) = A_H \exp\left(-\frac{(x-\mu_H)^2}{2{\sigma_H}^2}\right), \qquad \text{(Equation 2)}$$

$$f(y) = A_V \exp\left(-\frac{(y-\mu_V)^2}{2{\sigma_V}^2}\right) \cos\left(2\pi\frac{y}{\lambda} - \psi\right), \qquad \text{(Equation 3)}$$

where *x*, *y* are column and row numbers (in pixels units, with 0 in the centre), *A* is amplitude, μ is spatial offset (in pixels), σ is spread (standard deviation in pixels), λ is wavelength (in pixels) and ψ is the absolute phase offset (in radians) of the co-sinusoidal component of the Gabor function. To separate the three shared parameters (*A*, μ, σ) in Equation 2 and 3, subscripts H and V indicate horizontal and vertical data, respectively. Absolute phase offset (ψ) was subtracted rather than

added (Equation 3) to harmonise the signs between spatial and absolute phase offsets. For convenience, absolute phase offsets were converted to pixel units, ψ_{pix} (Equation 4):

$$\psi_{pix} = \lambda\psi/2\pi, \qquad \text{(Equation 4)}$$

This makes the absolute phase offset parameter, ψ_{pix}, directly comparable to the spatial offset of the Gaussian, μ. Absolute phase (ψ) changes the peak position of the cosine component, and spatial offset (μ) changes the peak of the Gaussian envelope. The asymmetry of the Gabor function depends on the relative values of these two offsets. This was captured by a relative phase parameter (φ_{pix}), derived by subtracting the spatial offset (μ) from absolute phase (ψ_{pix}) and converted back to radians to give, φ. Thus, μ, expresses the lateral shift of the entire Gabor function and φ (and φ_{pix}) expresses the phase shift relative to this.

Equation 2 (the Gaussian) has three free parameters (A, μ, σ) and Equation 3 (the Gabor) has five (the same as the Gaussian plus λ and ψ (or ψ_{pix}) (or alternatively, λ and φ (or φ_{pix}))). In addition to these parameters, the location of the Gabor peak (which depends on μ, σ, and φ_{pix}) was determined using a MATLAB implementation of the Nelder-Mead simplex algorithm to find the lateral position, P, of the function maximum (Figure 4.8, Table 4.3).

The fits of Equations 2 and 3 to the group averages from Figures 4.6 and 4.7 are shown in Figure 4.8, and their parameter values are reported in Tables 4.2 and 4.3, respectively. The Gabor fits to individual observers are shown in Appendix C and D for the disparity and luminance CIs, respectively. The aim was to use whichever of the parameters above served us best in evaluating the three CI hypotheses (H1, H2 & H5). As will be illustrated, these proved to be: A, φ, P and σ. By comparison, λ and ψ, did less to distinguish between the factors of interest. They are included in Tables 4.2 and 4.3 for completeness but were not considered further. The spatial offset of the Gaussian, μ, was arguably more valuable, but for the vertical cross-sections it was subsumed by P. For the horizontal cross-sections, μ was always significantly negative meaning the fitted Gaussians were shifted a little to the left. This leftward shift remains unexplained, but given the variability of the data around the fitted profile (Figures 4.8c & d), the shift likely conveys little or nothing of value and thus μ is not considered further.

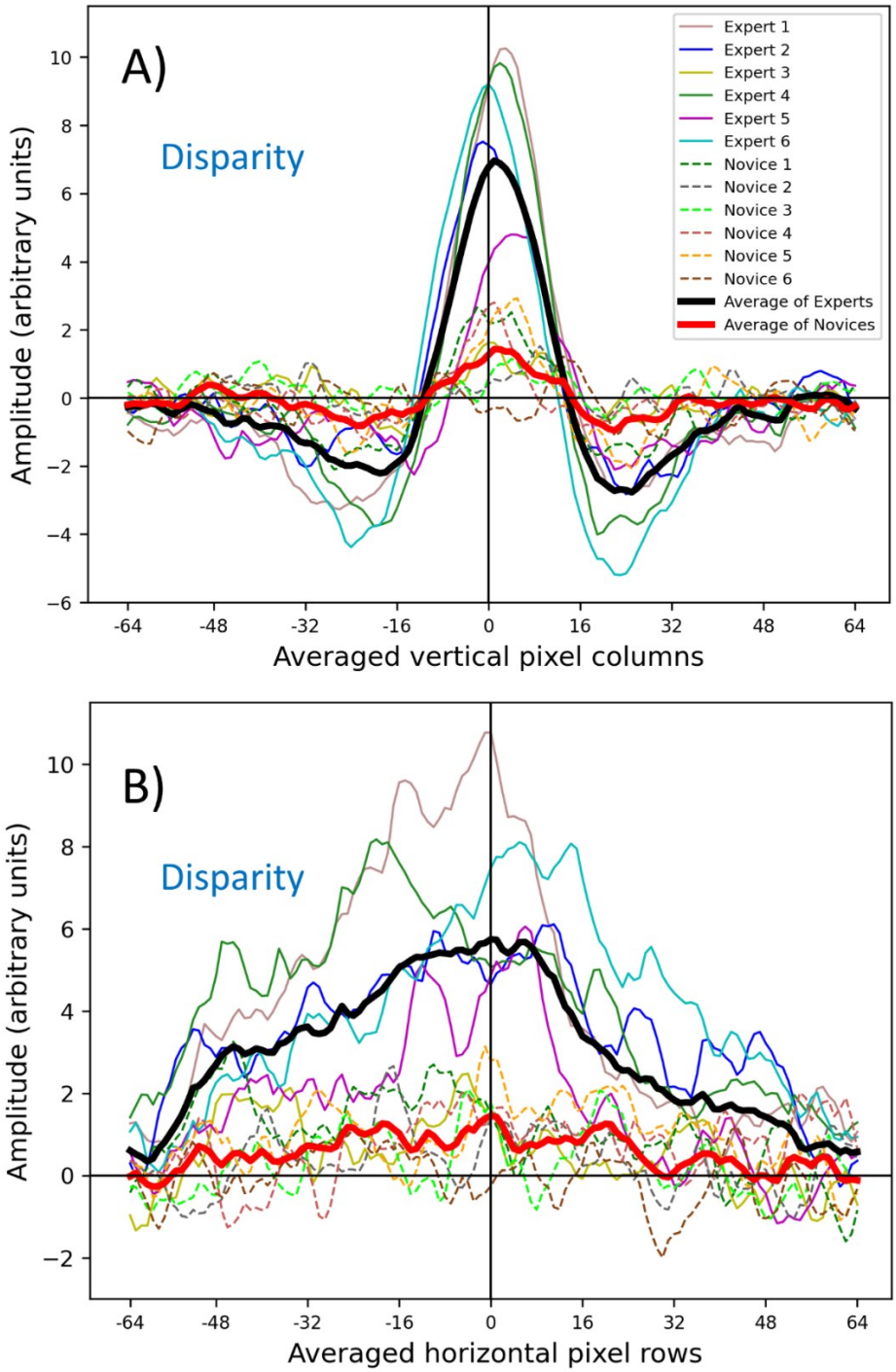

Figure 4.6: Cross-sections of disparity classification images. A) Vertical cross-sections. Left of centre corresponds to 'north' in the hedge and ditch images. B) Horizontal cross-sections of the central 20 rows. Experts and novices are shown by solid black and red curves, respectively. Different colours are for individual observers; the thick black curves are for group averages.

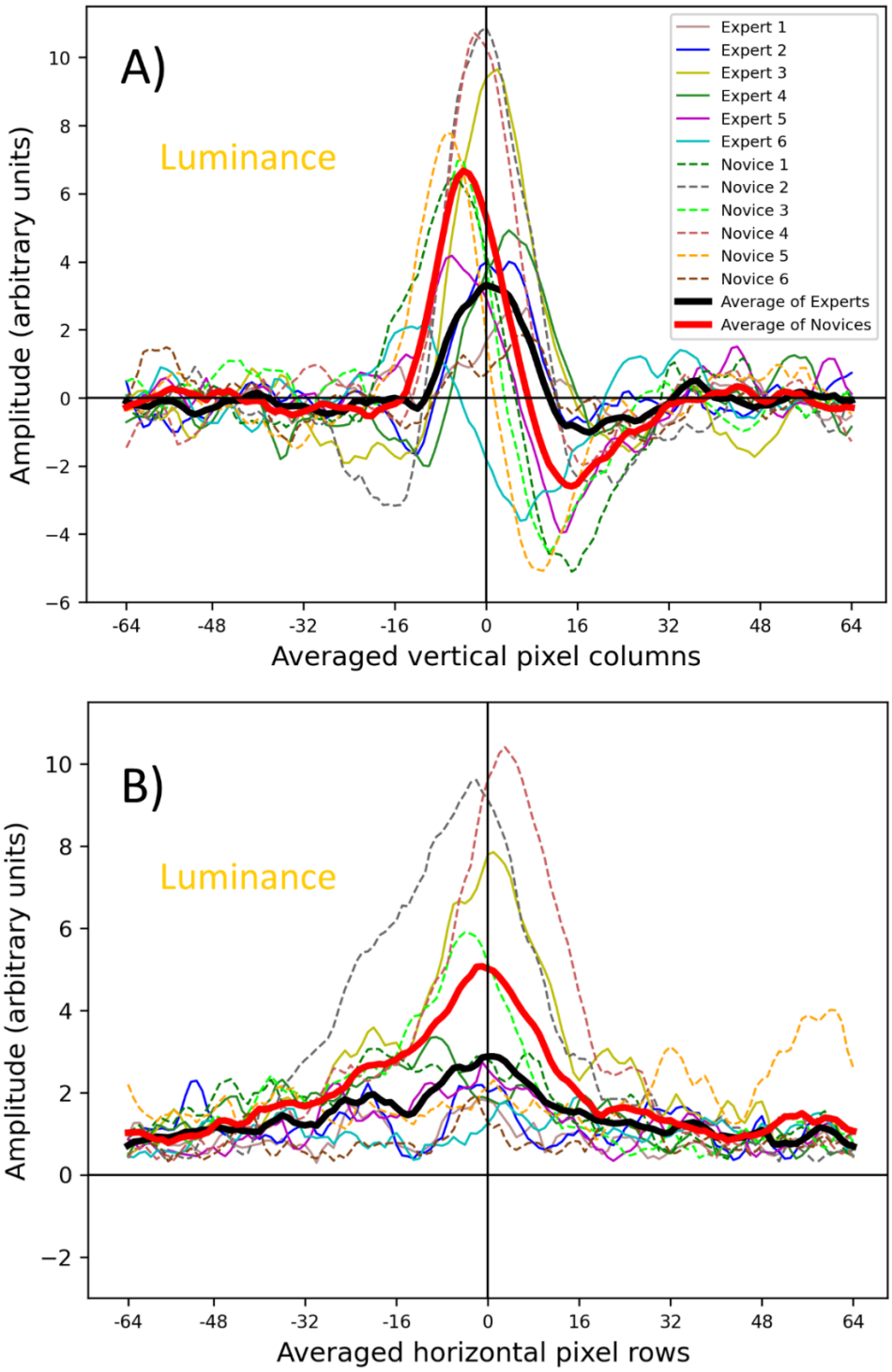

Figure 4.7: Cross-sections of luminance classification images. The details are as for Figure 4.6. A) Vertical cross-sections. B) Horizontal cross-sections of the central 20 rows after full-wave rectification (see text for further details).

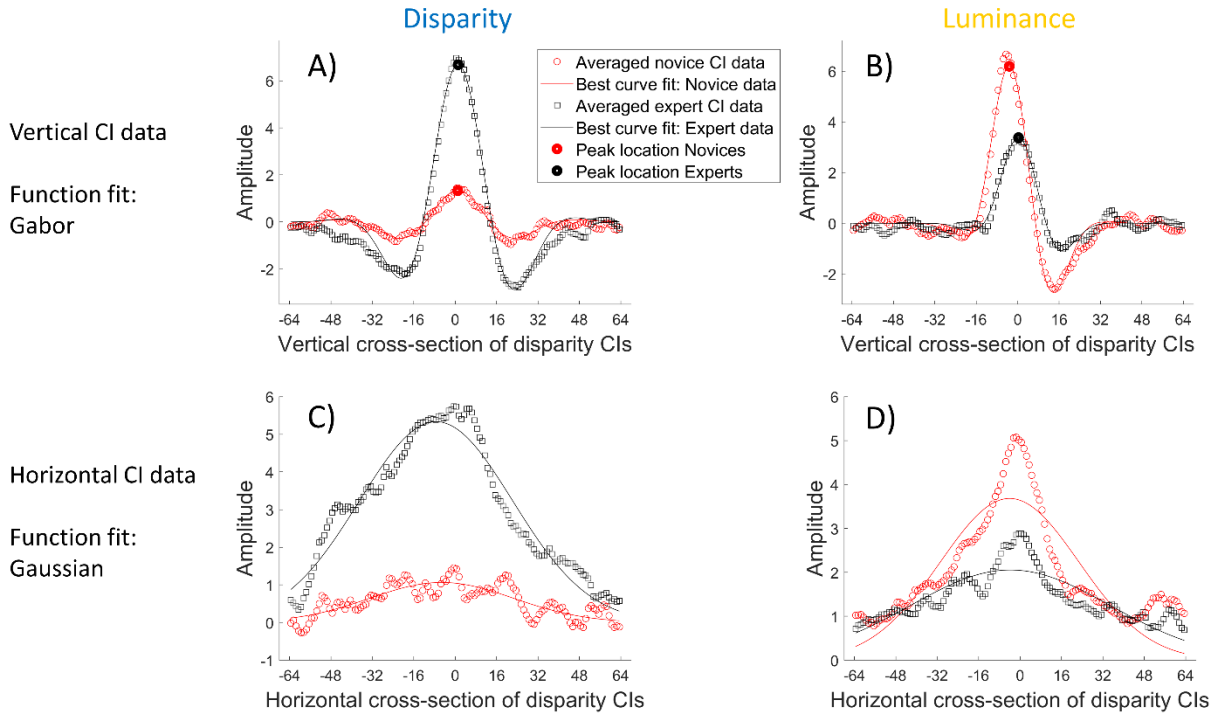

Figure 4.8: Fits of descriptive curves to the group averaged cross-sections of the CIs (Figures 4.6 and 4.7). A) Vertical disparity results fitted by Gabor functions. B) Vertical luminance results fitted by Gabor functions. C) Horizontal disparity results fitted by Gaussian functions. D) Full-wave rectified (see text for details) horizontal luminance results fitted by Gaussian functions. The origin of the x-axis is the stimulus centre. In the top row, left-to-right along the x-axis corresponds with top-to-bottom in the vertical CIs.

	Fitted Gaussian parameters to horizontal disparity CIs (with 95% confidence bounds)		Fitted Gaussian parameters to horizontal luminance CIs (with 95% confidence bounds)	
Group	Experts	Novices	Experts	Novices
Amplitude A	**5.33 (5.19, 5.48)**	**1.07 (0.98, 1.15)**	**2.06 (1.96, 2.15)**	**3.68 (3.46, 3.90)**
Spread σ	29.58 (28.61, 30.55)	27.78 (25.1, 30.47)	**38.63 (35.98, 41.27)**	**26.82 (24.9, 28.75)**
Spatial offset μ	-7.35 (-8.26, -6.43)	-5.21 (-7.79, -2.62)	-3.74 (-5.92, -1.56)	-4.02 (-5.89, -2.14)
Adjusted R^2	0.937	0.663	0.682	0.715

Table 4.2: Gaussian parameters (Equation 2) for fits to the group average horizontal CIs (Figure 4.8c, d). Non-overlapping confidence intervals between groups are shown in bold. (Confidence intervals were estimated by the Matlab fitting procedure). Goodness of fit is shown by adjusted R^2. Parameter values that belong to the x-axis in the figures (σ, μ) are in pixel units. Negative spatial offsets indicate leftward lateral shifts of the peaks in Figures 4.8c, d.

	Fitted Gabor parameters to vertical disparity CIs (with 95% confidence bounds)		Fitted Gabor parameters to vertical luminance CIs (with 95% confidence bounds)	
Group	Experts	Novices	Experts	Novices
Amplitude, *A*	**6.69 (6.44, 6.95)**	**1.33 (1.24, 1.42)**	**3.55 (3.37, 3.73)**	**6.65 (6.41, 6.9)**
Spread, σ	**17.59 (16.84, 18.33)**	**20.49 (18.85, 22.13)**	10.64 (9.95, 11.34)	11.56 (11.05, 12.07)
Spatial offset, μ	2.19 (1.32, 3.06)	0.92 (-0.8, 2.64)	**3.02 (2.17, 3.88)**	**0.44 (-0.12, 0.1)**
Wavelength, λ	53.1 (51.94, 54.26)	50.32 (48.7, 51.95)	48.08 (46.08, 50.09)	46.76 (45.52, 48)
Absolute phase, ψ_{pix}	0.89 (0.60, 1.18)	1.03 (0.56, 1.51)	**-1.07 (-1.51, -0.63)**	**-4.78 (-5.01, -4.55)**
Absolute phase, ψ (in radians)	0.03π (0.02π, 0.4π)	0.04π (0.02π, 0.6π)	**-0.04π (-0.06π, -0.03π)**	**-0.20π (-0.21π, -0.19π)**
Relative phase, φ_{pix}	**-1.3 (-1.59, -1.01)**	**0.11 (-0.36, 0.58)**	**-4.09 (-4.53, -3.65)**	**-5.22 (-5.44, -4.99)**
Relative phase, φ (in radians)	**-0.05π (-0.06π, -0.04π)**	**0.0π (-0.01π, 0.02π)**	**-0.17π (-0.19π, -0.15π)**	**-0.22π (-0.23π, -0.21π)**
Peak location, *P*	1.13	1.02	0.31	-3.26
Adjusted R^2	0.966	0.898	0.954	0.977

Table 4.3: Gabor parameters (Equation 3) for fits to the average vertical CIs (Figure 4.8a, b). Bold text shows non-overlapping confidence intervals between groups. Goodness of fit is shown by adjusted R^2. Several of the function parameters that relate to the x-axes in the figures (σ, μ, λ, ψ_{pix}, φ_{pix}) are in pixel units, as is the (lateral) peak location (*P*). Negative spatial offsets and phase indicate lateral shifts to the left.

4.3.5 Interpreting disparity CIs (including evaluation of H1 & H2)

In support of the stereo fidelity hypotheses (H1), the amplitude (*A*) of the average disparity CIs (Figure 4.8) was about five-times greater for the experts than for the novices (left of Tables 4.2 and 4.3; red and black curves in Figures 4.8a & c), confirmed by an independent samples t-test (Welch's $t(5.59) = 3.79$, $p = 0.005$; one-tailed). However, for the vertical cross-sections (Table 4.3), the Gabor spread (σ) for the novices was slightly greater than for the experts, though no reliable differences for the Gaussian spreads (σ) of the horizontal cross-sections were found (Table 4.2). Thus, the disparity CIs suggest that experts were better than novices at picking up disparity cues for depth (H1), but contrary to expectations (H1), did not sample this information over a greater spatial extent than novices.

The cue strategy hypothesis (H2) was also supported. Both groups reportedly attempted to use disparity cues (see *Debriefing*, above) but experts prioritized them over luminance cues, particularly by comparison to novices. Seeing this in the group fits (Table 4.3) is problematic since there is no measure of noise equivalence across noise types (disparity and luminance), and no signal amplitude-to-noise ratios can be derived to make the relevant comparisons. This was addressed by dividing the amplitudes from the fits to the *n*=12 individual observer results (Appendices A and B) by

their mean for each noise type. A 2 x 2 repeated measures ANOVA (Group: expert, novice; Cue type: disparity, luminance) on these normalised results revealed a significant interaction between participant group and cue type ($F(1, 5) = 15.61$, $p = 0.011$)[8]: on average, experts prioritised disparity over luminance but this was the other way around for novices (see Figure 4.9). Note also that the luminance amplitudes (*A*) for the novices were higher than for the experts (Tables 4.2 and 4.3). This means the expert superiority with disparity cannot be attributed simply to greater engagement with the task since under that account, novices would not be expected to outdo experts on luminance cues.

Notably, the CI disparity difference for experts and novices did not derive from stereoacuity because the two groups did not differ in their TNO test scores (Table 4.1) (One-Way ANOVA; Welch's $F(1, 5.53) = 0.430$, $p = 0.538$). When Expert 5 was removed as an outlier (Table 4.1), the difference remained non-significant (Welch's $F(1, 5.34) = 1.93$, $p = 0.220$). Furthermore, Expert 5 had a TNO threshold (450 arcseconds) fifteen times higher than the median and modal TNO threshold in the current sample (30 arcseconds) but produced a higher contrast disparity CI than any novice nonetheless (Figure 4.6a). A further observation is that the side-lobes in the disparity templates were more apparent for experts than for novices (Figures 4.6a & 4.8a). This shows that with experience, decisions about a feature's stereoscopic profile are influenced more by the surrounding context.

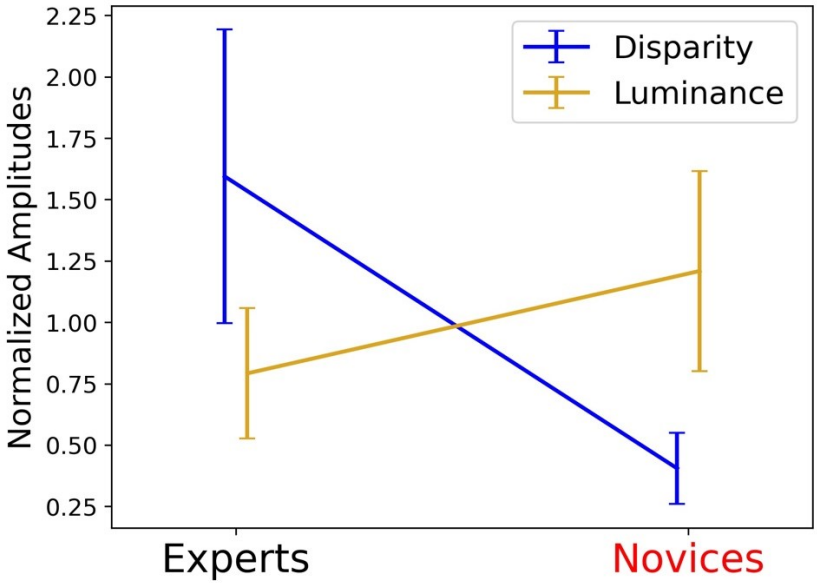

Figure 4.9: Interaction of CI amplitudes across cue type and groups. The amplitude, *A*, for each of the twelve participants (Appendices A & B) was normalised for each noise type (disparity and luminance) by dividing by their relevant mean. Error bars are ±95% confidence intervals.

[8] The fit to Expert 5's luminance CI was unusual, bringing the estimate of their amplitude (*A*) into question. The details and solution are outlined in Appendix D. It is also noted that when Expert 5 was removed as an anomaly, the conclusion from the ANOVA was unchanged ($F(1, 4) = 13.16$, $p = 0.022$).

4.3.6 Signal detection analysis (d' and bias)

Beyond CI measures, this experiment included three different manipulations to the images, such as inversions of disparity profiles, elaborated below. Thus, the observers' responses to these manipulated image factors (signals) can provide another performance measure and show how different factors influenced responses. The observer's sensitivity to signals can be estimated with the bias-free sensitivity measure *d*', based within signal detection theory (Green and Swets, 1966; Macmillan & Creelman, 2004). In a traditional yes/no procedure, the signal is the stimulus the observer is trying to detect, and the ground truth is whether the stimulus was presented. In this experiment, trials followed the SIBR design but always contained a stimulus feature, that was either a hedge or a ditch. Note that this bears similarity to the second experiment in Chapter 2 where 'Same' and 'Different' house images were assigned to target-distractor categories in generating *d*' measures reflecting accuracy results. In the current stimulus-response task, hedges were assigned to the target category and ditches to the distractor category. This was used to record stimulus-response categories: hits (hedge-hedge), misses (hedge-ditch), false alarms (ditch-hedge) and correct rejections (ditch-ditch), and *d*' was derived in the conventional way.

When performing the task, differences between hedges and ditches can be characterised by different image aspects, such as 3D relief. The sensitivity analysis focused on how three manipulated image factors influenced responses: 1) the original image (hedge or ditch), 2) binocular disparity (crossed or uncrossed) and 3) lighting direction cues (inversions of highlights and shading from vertical flips, associated with the lighting-from-above prior). The first image factor reflects the contents of the original photographs, such as texture features (Figure 4.1). But in the second and third image factors, the ground truth does not relate to whether the image was a photograph of a hedge or a ditch, but to binocular disparity and lighting direction cues for 3D relief interpretations, respectively. The sensitivity analyses were performed on each of the three assumptions about ground truth. The results are elaborated below, and they show how observers used the three image factors in their judgements.

4.3.6.1 Bias

The analysis of bias covers all three image factors as it simply relates to the two response categories 'hedge' and 'ditch'. Bias (Figure 4.10) has no relation to the current hypotheses, but the presence of bias in the data motivates the use of the bias-free sensitivity measure (*d'*), rather than percent correct responses. Biases are shown in Figure 4.10 for each observer where the bias measure is given by the number of 'ditch' responses subtracted from the number of 'hedge' responses. Biases were normalised by the total number of trials and expressed as percentages. All

observers responded ‘hedge’ more often than ‘ditch’ (all bars are above 0% in Figure 4.10). This direction of the bias is consistent with the well-known convexity bias (Adams & Elder, 2014; Champion & Adams, 2007; Langer & Bülthoff, 2001; Perrett & Harries, 1988). In some cases the bias was fairly strong (e.g. a bias of 30% indicates a 65:35 split for hedges:ditches).

These biases could be caused either by response bias, where observers tended to press one button more often, and/or a perceptual convexity bias. Liu and Todd (2004) suggest a perceptual origin to the convexity bias, and the current results are consistent with ambiguous 3D relief being more likely seen as a convex than a concave feature.

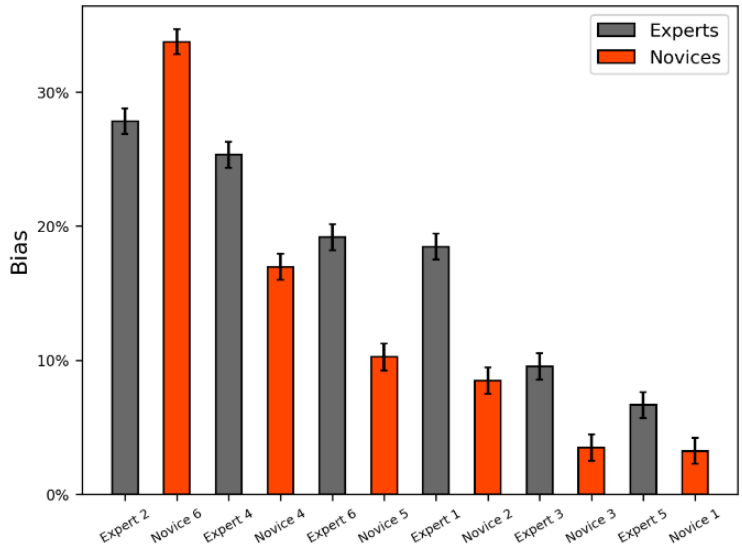

Figure 4.10: Individual biases (100 x (*n* ‘hedge’ responses - *n* ‘ditch’ responses)/(*n* ‘hedge’ responses + *n* ‘ditch’ responses)). All observers made more ‘hedge’ responses than 'ditch' responses (all bars are above 0%). Observers are rank ordered by bias within group (left to right). Error bars show 95% confidence intervals (Clopper-Pearson method).

4.3.6.2 Initial observations of d' analysis

Figure 4.11 shows individual sensitivities (*d'*) where the ground truth was defined by: a) the original image, b) the binocular disparity profiles (crossed or uncrossed), and c) the assumption of lighting-from-above. Although results from hedge and ditch images can be shown separately, they produced a highly similar outcome and were thus combined in Figure 4.11. Individual differences are seen across the three plots in Figure 4.11, where sensitivity measures varied as observers applied different assumptions of ground truth.

The original image content (Figure 4.11a) was the least valuable image factor (Figure 4.11). This shows that observers generally relied on disparity profiles and lighting direction cues more than other content (e.g., texture features) in the original images (Figure 4.1). This is a satisfactory result as judgements based on original image content were considered spurious, and the strong levels of noise were meant to completely mask such content in the images. Novices (red) used more of the original

image content than experts (dark grey), but sensitivities were overall low, and are not considered further. Novice 6's higher sensitivity to original image is notable. This participant might have detected some content in the original images, for example texture features, that other participants mostly ignored. Novice 6 is generally an outlier, who also had the strongest bias (Figure 4.10). Bias and sensitivity to original image might in part explain the absence of structure in their CIs (Figure 4.5) and the generally weak sensitivities to other ground truth aspects (Figures 4.11b, c).

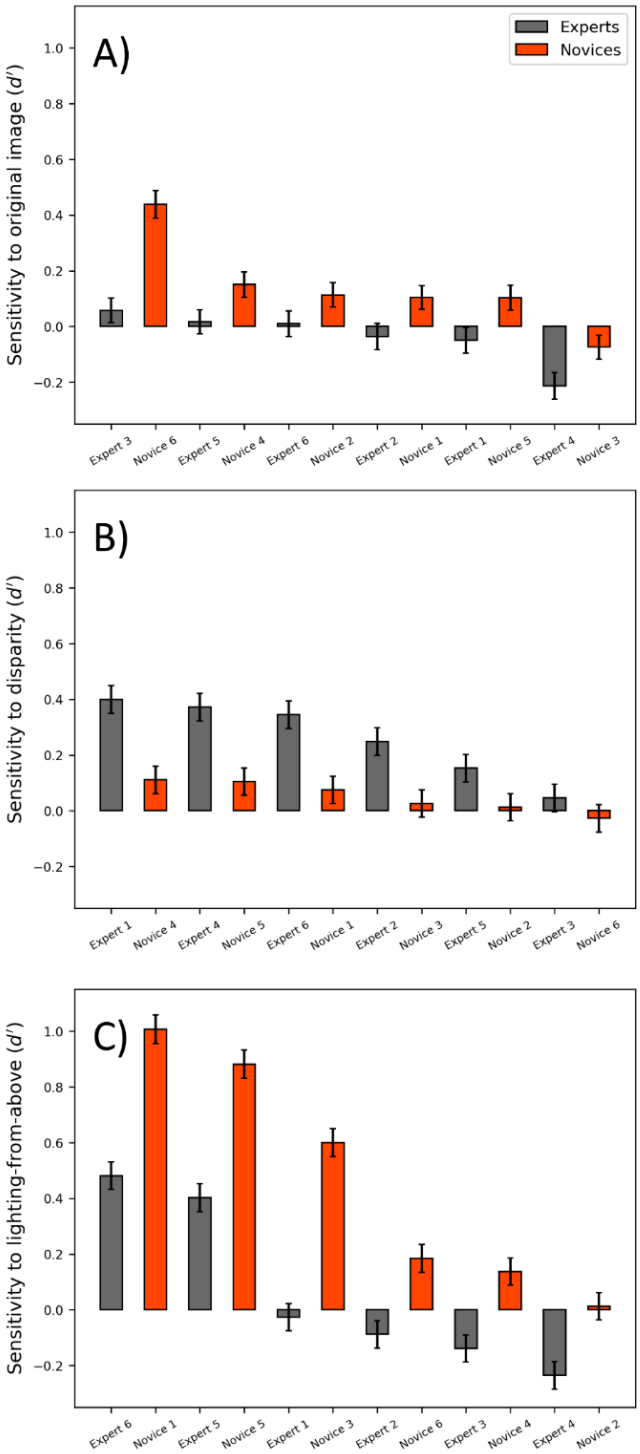

Figure 4.11: Individual sensitivities (*d*') to: A) original image, B) disparity profiles, and C) lighting-from-above. Observers are rank ordered within group (left to right) according to their sensitivity in each plot. Error bars show 95% confidence intervals (Macmillan & Creelman, 2004).

4.3.6.3 Sensitivity to disparity profiles (including evaluation of H3)

Figure 4.11b reveals a clear difference across groups for sensitivity to disparity profiles. The average expert sensitivity was *d*' = 0.264 for the three hedge images and *d*' = 0.277 for the three ditch images (not shown). Overall, novices were less sensitive than experts for both hedges (*d*' =

0.046) and ditches (d' = 0.051) (not shown), and within each ranked pair of observers across groups, the expert always had a greater sensitivity to disparity than the novice. These observations confirm the stereo accuracy hypothesis (H3) for categorical results ($t(10) = 3.51$, $p = 0.003$; one-tailed), where the experience of experts would expectedly allow for better use of disparity cues compared to novices. Furthermore, 1) the similar accuracies for the hedge and ditch images within group confirms that the stereoscopic profiles of the two image types were perceived equally well, and 2) the generally low performance levels indicate that the disparity noise was an effective mask. (For an unbiased observer, a d' of 0.3 corresponds with 56% correct in a single interval yes/no task.)

Figure 4.12 shows that across observers, d' sensitivity for disparity correlated strongly with the amplitude of the disparity CIs (A from the Gaussian fits to the individual cross-sections in Figure 4.6b) (Pearson's $r^2 = 0.919$, $p < 0.001$). This reaffirms the observations regarding H1 and H3: that overall, experts had higher scores for both measures compared to novices. This was to be expected since a high amplitude in a CI (in this case disparity) indicates a high global sensitivity for the relevant cue which is therefore identified/detected more reliably. On the other hand, d' for disparity did not correlate with TNO thresholds (Table 4.1) ($r^2 = 0.01$, $p = 0.755$). This is similar to the earlier, and presumably related, observation that TNO thresholds did not explain group differences in disparity CIs. This will be discussed further below.

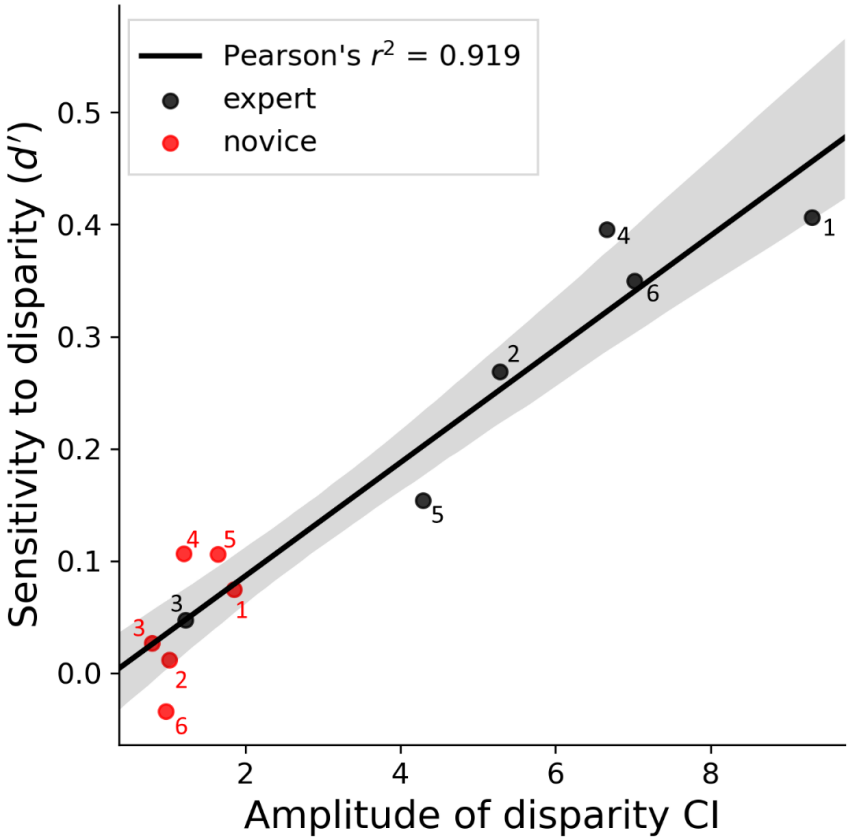

Figure 4.12: Relation between the amplitude of the disparity classification images (Figure 4.5 and 4.6b) and sensitivity (*d*') for classifying stimuli as a hedge or a ditch according to whether the landscape feature was presented with crossed or uncrossed disparity (Figure 4.11b). Individual participants are numbered (nominally) within the groups.

4.3.6.4 Sensitivity to lighting-from-above (including evaluation of H4)

Figure 4.11c shows sensitivities (*d*') to lighting direction, where values greater than and less than zero indicate perceptual priors for lighting-from-above and -below, respectively. This shows that: 1) novices were generally more prone to directional biases (were more distant from zero) than experts, and 2) novices had stronger biases towards lighting-from-above than experts, where some experts were biased towards lighting-from-below ($t(10) = -1.92$, $p = 0.042$; one-tailed). Within each ranked pair of observers across groups, the novice always had a greater sensitivity to lighting-from-above than the expert. This supports the lighting sensitivity hypothesis (H4), where the experts' experience with OS images lit from below the line of sight would expectedly diminish the conventional assumption for lighting-from-above.

Finally, Figure 4.11 shows that across the three different assumptions for ground truth, the greatest sensitivities were for lighting direction (Figure 4.11c), and this was for the novices. This point is expanded in the General Discussion.

4.3.7 Interpreting luminance CIs and individual differences (including evaluation of H5)

The differences between groups for the luminance CIs were smaller than for the disparity CIs (see Figure 4.5), but several comparisons are worthy of note. First, novices produced larger amplitudes (*A*) than experts (right of Tables 4.2 and 4.3; red and black curves in Figures 4.8b & d; Figure 4.9). No *a priori* prediction was made for this result, but it is consistent with the cue strategy hypothesis (H2) which stated that prioritisation of cues might take place differently across the groups. However, the experts had greater spreads (σ) in the horizontal direction than novices (Table 4.2). This shows that experts sampled luminance over a greater spatial range of the landscape feature despite giving it lower priority. The group level results from the luminance CIs show that this cue was used by both novices and experts in the task (Figure 4.5 and 4.7).

The lighting bias hypothesis (H5) predicted that patterns in the luminance CIs would relate to lighting direction biases and reveal group differences. The average luminance CIs for both groups (Figure 4.8b) show asymmetries, as revealed by the positional offset of the peak (*P*) and relative phase (φ) (Table 4.3), consistent with lighting-from-above. These effects were larger for the novices than the experts (Table 4.3), consistent with the expectation (H5). For the expert group the peak was located much more centrally than for the novice group (Figure 4.8b), suggesting that the assumed light source was more diffuse for the experts. However, the luminance CIs were less marked overall for this group (Figures 4.5 & 4.7), with less asymmetry (smaller relative phase shifts) and (as noted above) lower amplitude (Table 4.3). This suggests that the expert group was less prone to lighting priors and to luminance cues in general. This is consistent with the earlier observations of the categorisation results (Figure 4.11c) and provides the expected link between H4 and H5.

The analysis above is for group trends but as Figure 4.7a shows, there were marked individual differences in amplitudes, peak locations, and phase asymmetries within both the novice and expert groups. This suggests individual differences for assumptions about lighting in terms of both direction (above/below) and source (punctate/diffuse). To examine this, Gabor functions (Equation 3) were fitted to the individual luminance CIs from Figure 4.7a (see Appendix D for the fits). In all cases but one, the amplitude was positive, and the absolute value of the relative phase shift was less than 0.5 π radians, overall consistent with hedges being lighter and ditches darker. The exception was Expert 6, for whom the relative phase shift was -0.88 π radians, placing dark and light pixels more centrally for 'hedge' and 'ditch' responses, respectively (see Figure 4.5, bottom left). No explanation is provided for this participant's switch in polarity from the expectations, but they were one of only two experts who had a lighting-from-above prior (Figure 4.11c).

To visualise the individual differences and to show the relationship between the luminance CIs and the categorical results, the *d'* sensitivity for lighting direction (from Figure 4.11c) was plotted

against the two indices of asymmetry: 1) the lateral offset of the function peak in the vertical cross-sections (P), and 2) a metric related to relative phase (φ) which tells us about the asymmetry of the shape of the CI. These are shown in Figures 4.13a and 4.13b, respectively. (See figure caption for details of how the relative phase metric was derived to accommodate Expert 6). In both cases the correlations were good (Pearson's $r^2 = 0.537$, $p = 0.007$; Pearson's $r^2 = 0.516$, $p = 0.009$ in Figures 4.13a and 4.13b, respectively). These relationships further affirm the expected link between H4 and H5. The division across participant groups (different coloured symbols in Figure 4.13) illustrates the lighting bias hypothesis (H5), where conventional lighting cues were expected to be a more important factor for novices than for experts. These are most marked at the extremes (red symbols, top right; black symbols, bottom left). However, there is marked overlap in the central regions of the plots, showing that the two groups do not differ as strongly on this measure (see also Figure 4.5) as they did on disparity (Figure 4.12). Some of the details are highlighted below.

Observers with the strongest perceptual biases for lighting-from-above (e.g., Novices 1, 3 and 5; Figure 4.11c) also had an asymmetric CI with a negative side-lobe 'south' of the positive peak (Figure 4.5, 4.7a and 4.13b). This suggests that a shadow was inferred 'south' in hedge features, and/or a highlight 'south' in ditch features, consistent with a lighting-from-above bias and an assumption of punctate lighting. The opposite inferences of highlights and shadows is seen for observers with a bias for lighting-from-below (e.g., Experts 2 and 4; Figure 4.5, 4.7a and 4.13), but less strongly, presumably due to their weaker biases (Figure 4.11c). The observers with centralized luminance peaks (e.g., Expert 3 and Novices 2 and 4; Figure 4.5 and 4.7a) also showed weaker lighting direction biases (Figure 4.11c and 4.13). This is consistent with the diffuse lighting assumption where 'dark-is-deep' is the identification rule, where lighter and darker textures prompt 'hedge' and 'ditch' responses, respectively.

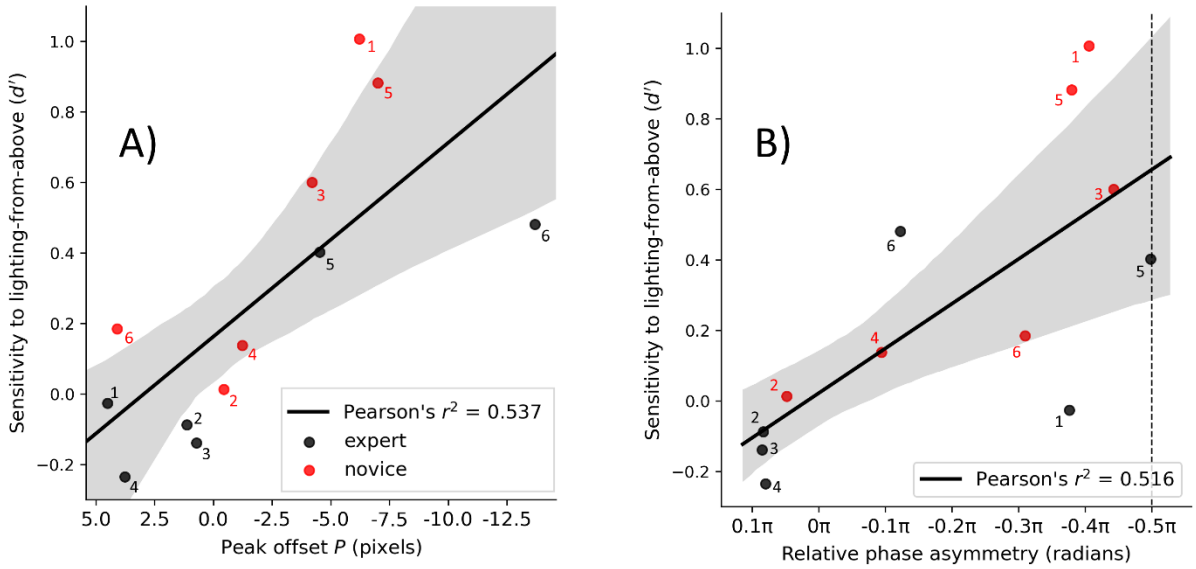

Figure 4.13: Relationships between sensitivities to lighting-from-above (Figure 4.11c) and peak location offsets (a) and relative phase (b) in individual luminance classification images. In (b), a phase of 0π radians indicates perfect cosine phase symmetry, and ±0.5π radians is maximally asymmetric sine phase (vertical dashed line). Individual participants are numbered (nominally) within the groups. In panel (b) Expert 6 was unusual in that $\varphi > |0.5\pi|$. Since the aim was for this figure to illustrate asymmetry in the CI, the result for Expert 6 was folded back across -0.5π. The result is a value of -0.12π under this metric but with opposite sign (light/dark; not depicted in this figure) compared with other participants.

In summary, lighting direction biases are implied by *d*' sensitivities to lighting-from-above (Figure 4.11c), and peak offsets and asymmetries in the luminance CIs (Figures 4.8b and 4.13). Novices showed a greater tendency for lighting-from-above, and lighting direction biases for experts tended to be diminished by comparison, or switched to lighting-from-below. Novices and experts thus differ in their assumptions about lighting and the influences these have on their luminance CIs, though there was much overlap between the two groups (Figure 4.13).

4.4 General discussion

The current study investigated judgements of stereoscopic aerial images and how novices and expert remote sensing surveyors might differ in their use of visual cues. This study applied the CI technique that was developed in the pilot studies of Chapter 3. New insights are brought with this novel CI technique that simultaneously estimates templates from luminance and disparity cues. The current study takes a novel methodological step in applying luminance and disparity noise to stereograms of natural images. Results show clear differences in the perceptual templates used by experts and novices when discriminating stereoscopic aerial images of hedges and ditches, under five specific hypotheses (H1-H5; see Introduction). The results showed that, compared to novices, experts made better use of binocular disparity cues (H1). Differences were also found in cue strategies, where

experts prioritised disparity cues over 'dark-is-deep' luminance cues[9] (H2). Conversely, novices prioritised luminance cues over disparity cues (H2), and had higher amplitudes in their luminance CIs. Sensitivity to stereoscopic profiles (*d'*) was greater for experts than novices (H3), but this did not relate to stereoacuity. The results from analyses of peak locations and asymmetries show individual differences in the interpretation of lighting direction cues, with experts less likely to adopt the conventional lighting-from-above prior (H4 & H5). This tendency of experts to have diminished lighting-from-above priors for aerial landscape features can be attributed to their experience with the counter-conventional lit-from-below imagery used by the OS.

4.4.1 Experts and novices

4.4.1.1 Stereoscopic judgements

The expert surveyors have years of experience with stereoscopic aerial landscape images. Confirming a primary hypothesis for this study, and this thesis, the current study shows that experts made better use of disparity cues than novices, and that this group difference was notably large. This result reveals a mechanism involved in interpreting stereoscopic aerial images that is strongly associated with expertise in remote sensing surveyors.

This expertise likely reflects learning from experience, and the CIs capture aspects of this expertise. Previous studies have used 2D CIs to study PL, finding learning effects for luminance CIs in detection of oriented gratings (Dobres & Seitz, 2010), face and texture identification (Gold, Sekuler & Bennett, 2004) and position discrimination (Kuai, Levi & Kourtzi, 2013; Kurki & Eckstein, 2014; Li, Levi & Klein, 2004). The current study is the first to demonstrate evidence of learning/expertise using disparity CIs. For an elaboration on stereoscopic PL following the results from disparity CIs in the current study, see Chapter 5.

Stereoacuities were measured using the TNO test. Perhaps surprisingly, disparity CI amplitudes were not correlated with the stereoacuity thresholds, and stereoacuities did not explain group differences. Expert 5's disparity CI was clearly defined and had greater amplitude than any novice, despite Expert 5 having a far higher stereoacuity threshold than all novices. Furthermore, TNO thresholds did not correlate with *d'* sensitivity for disparity profiles. Taken together, this shows that the ability to sample disparity cues in the CI task did not depend on stereoacuity. To elaborate on this disparity mechanism; stereoacuity concerns the smallest detectable difference (threshold) across binocular retinal images. But the current task required observers to extract a meaningful disparity signal in noise, where both the signal and noise might be detectable (above-threshold). Here, relevant disparity signal must be pooled against a background of disparity noise to contribute

[9] See Chapter 6 for a discussion on mechanisms associated with template shapes.

to the perception of a stereoscopic surface. Carrillo, Baldwin & Hess (2020) used disparity noise-masking and found no relation between stereoacuity thresholds and the level of external disparity noise that could be tolerated. Consistent with the current study, this suggests that detection of binocular disparities across the two eyes (stereoacuity threshold) operates at a different stage than spatial pooling of relevant disparity signal against a background of disparity noise. The current study shows that experts have a greater facility to extract a relevant disparity signal, but their stereoacuity thresholds are not better than novices.

4.4.1.2 Cue strategy

Experts prioritised and sampled disparity cues over 'dark-is-deep' luminance cues more so than novices. The results also show that, less expectedly, the novices had a greater luminance CI amplitude than the experts. This outcome was not reflected in the verbal debriefing, where most experts and novices reportedly used disparity cues as a primary strategy and 'dark-is-deep' luminance cues as a secondary strategy. Furthermore, the debriefing clearly showed that none of the observers were aware of using lighting direction cues (a punctate lighting assumption). That is, observers reported using luminance cues where 'dark-is-deep', but did not relate this to any asymmetries or shifts in peak locations consistent with punctate lighting-from-above or -below. This was rather surprising considering that the novices had relatively large *d'* sensitivities for the detection of lighting-from-above, and this sensitivity was greater for this cue/group combination than any other (Figure 4.11). This might suggest that some (novice) observers were aware of using lighting direction as a ground truth signal, but this was not the case.
Thus, CIs revealed group differences that the verbal reports did not: that the experts were better able to prioritise and use the disparity cues, and that the use of luminance cues varied across groups and individuals. This suggests that CIs can be a powerful technique for revealing visual strategies that observers are unaware of using, and this topic is elaborated further in Chapter 6.

A feature of the experimental design was that the sign of disparity and lighting direction were inconsistent in about half the trials (and consistent in the remainder). This means that for observers who detected both cues conventionally (e.g., Experts 5 and 6) these two cues would have been in conflict about 50% of the time, diminishing the performance that would otherwise be achieved. Note that in general, on removing the conflict trials from the analysis, *d'* equals the sum of those measured when each of the disparity and lighting from above cues were treated as ground truth. For Expert 6 this is quite a benefit; when hedge and ditch images are lit from above and have consistent disparity, this observer would benefit from both cue types. Note that this is not specific to observers who show a bias to lighting from above. Expert 4, for example, shows evidence for

detecting lighting from below (Figure 4.11c). Since the sign of *d'* in Figure 4.11c depends only on what was deemed to be the correct direction for lighting, Expert 4 would also benefit from the combined performance across cues (the sum of the absolute values of the *d'* measures) when hedge and ditch images were lit from below. Cue conflict arises when there is 1) inconsistency between the observer's lighting prior and the lighting direction in the image and 2) there is sensitivity to both cue types. The greatest *d'* sensitivities were found for an assumption of lighting from above for the novices. In similar tasks, but where images are presented without conflict, novices would benefit from lighting from above and would benefit further on being trained to use binocular disparity.

Furthermore, the separation of hedge and ditch images across blocks likely contributed to increasing the participants' reliance on the image manipulations as other image content was consistent within blocks.

4.4.1.3 Lighting direction priors

Lighting direction can be a strong cue to 3D shape from shading, as seen with the honeycomb stimulus (Figure 4.4 and 4.14a, b). But for hedges and ditches that are unmasked or have weaker noise masks, the lighting direction is likely less important for identification. For example, Figure 4.14c-f shows the effect of rotating an unmasked hedge and ditch. The reader is encouraged to try to ignore the labels in Figure 4.14 and decide which images are hedges and which are ditches. Further, the reader might directly reflect on the lighting direction structure, and decide which lighting direction makes hedges appear more 'hedge-like', and ditches appear more 'ditch-like'. Recall that the OS imagery are originally lit from the 'south', from below the line of sight. The impression shared by the author and colleagues is that the impact of inverting the hedges and ditches is small compared to the honeycomb (Figure 4.14a, b), but that there is some benefit for the impression of 3D relief when light comes from above (Figure 4.14c-f).

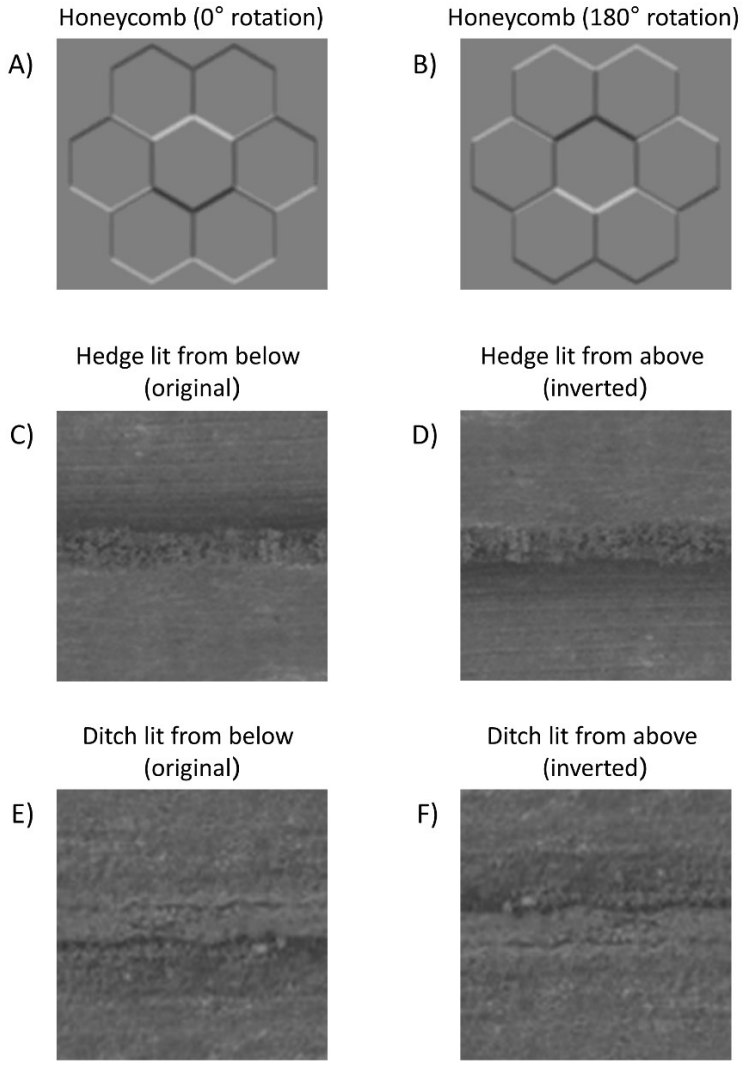

Figure 4.14: Images that might change in interpretations of 3D relief when inverted. A and B display the honeycomb stimulus with 0° or 180° orientations. C-F display hedges and ditches with either original or inverted orientations. Note that the hedges and ditches are in their original orientation when they are lit from below. Images c-f, *© Crown copyright and database rights 2023 OS*, used with permission.

The results, particularly from novices, show that such lighting direction cues provided an important cue to the hedge/ditch discrimination when images were heavily embedded in noise. With the approach of using extremely strong noise to mask the targets (Figure 4.3), the CIs reveal the perceptual strategies and/or the expectations that the observer brings to the task. See also the 'superstitious' method, where the stimulus is not presented, and a CI template reflects the observer's top-down expectations (Gosselin & Schyns, 2003). The current study reveals how perceptual strategies differed for the experts and novices, and the experts show evidence of having overcome the conventional prior for lighting-from-above (in 4 out of 6 cases; Figure 4.11c)[10].

[10] The likelihood of having no lighting-from-above bias was estimated based on previous literature, to explore whether the current sample could have occurred by chance. An estimate from Schofield et al. (2011) suggest that two out of nine participants do not have the conventional lighting-from-above bias (probability of 0.22). In

Previous studies have shown that experience can shape lighting direction priors. Adams, Graf and Ernst (2004) trained participants to shift their lighting-from-above priors by ~15° with cue-conflicting haptic training. However, the authors concluded that these shifts would most likely revert back to baselines after the experiment. It is unknown how the counter-conventional lighting assumption found with the experts in the current study would transfer to the real world. This research question is addressed below in a follow-up experiment which attempted to characterise this using the honeycomb image. The current study does suggest that experts' lighting direction priors are more permanently altered, as their experience with counter-conventionally lit aerial images impacted their strategies in a perceptual task (detecting signal in noise) that is different from their usual tasks. The current perceptual task was different from classifying features in natural images, but importantly, the observers brought the expectations of classifying domain-related aerial images.

4.4.2 *Summary and conclusions*

The current study tasked six expert remote sensing surveyors and six novices with discriminating expert-domain related stereoscopic aerial images of hedges and ditches with added luminance and binocular disparity noise. This novel method provided a new way to measure visual expertise by producing CIs for luminance and binocular disparity simultaneously. Compared to the novices, experts produced disparity CIs with five-fold greater amplitudes and detection sensitivities (*d*') to stereoscopic targets, revealing their advantage for sampling disparity cues. The study also shows that CIs can characterise lighting direction priors, and that experts were less likely to show the conventional lighting-from-above prior, which can be attributed to their counter-conventional experience with lit-from-below aerial imagery. These results are an important part of this thesis, and they suggest how visual mechanisms for interpreting aerial images can change with experience.

The methods and results of the current study have practical potential for directing visual training in remote sensing surveying, and for investigating basic perceptual mechanisms of human early vision. The CI technique can continue to bring insights into other domains of early vision, and into the development of visual expertise, in harmony with other approaches.

---

the current study, the probability of having four or more such observers in a sample of six is statistically unlikely (p = 0.0248). Furthermore, this is a conservative estimate, as Pickard-Jones, d'Avossa and Sapir (2020) found that all of fifty-eight children (ages 7-11) showed a bias for lighting-from-above with the honeycomb stimulus. Sun and Perona (1998) also found that twelve adults all had lighting-from-above biases with shaded bubble stimuli. It is thus concluded that, although the sample is small, this result is reliable and reflects the surveyors' lighting direction biases.

4.5 Follow-up experiment on lighting direction priors in expert surveyors

4.5.1 Aims

An online follow-up experiment was designed with two aims: 1) To explore if a different and shorter experiment can be used to capture lighting direction priors for aerial images of hedges and ditches, and 2) to explore how the experts use lighting direction priors outside of the domain of expertise. This experiment also used noise-masked hedges and ditches in different orientations, similar to the main experiment of this chapter, described above. See the above introduction to this chapter for an elaboration on how lighting directions can be used to discriminate hedges and ditches. This experiment used 2D images rather than stereograms. As the previous experiment in this chapter shows that binocular disparity can be a primary cue for discriminating hedges and ditches, the removal of this cue might increase the reliance on luminance cues such as lighting direction cues to shape from shading.

Owing to their unusual experience with lit-from-below imagery, expert surveyors were expected to show switched lighting direction priors for the hedges and ditches. The priors for these features will be compared to the prior for a well-established psychophysical stimulus – the honeycomb image, which is an image outside of the surveyors' domain of expertise. It is possible that surveyors might show different priors for the two types of images, which might suggest that they have context-specific priors for the domain-specific aerial images. The comparison across the aerial images and the honeycomb remains for open exploration.

As the below results and discussion will show, this follow-up experiment successfully captured lighting direction priors for the honeycomb image but failed to reliably capture lighting direction priors for the aerial landscape features. This was not the desired outcome in this experiment, and it is discussed in detail below. The current follow-up experiment was an online experiment, and further differences between it and the main experiment of this chapter are discussed below.

4.5.2 Method

The psychophysical stimulus was the honeycomb image (Andrews et al., 2013), constructed out of a hexagonal lattice with highlighted and shaded edges (Figure 4.14a, b). This image commonly elicits a strong impression of 3D shape, with the lighting-from-above prior leading to interpretations of 3D surface convexities and concavities from the highlighted and shaded edges (see also Chapter 1 for further details on this prior). With the lighting-from-above prior, the central hexagon in Figure 4.14a appears to contain a convex bump, and Figure 4.14b appears to contain a concave dimple.

The natural images from OS aerial landscape imagery were of hedges and ditches (Figure 4.14c-f). Hedges and ditches, being convex and concave, respectively, have opposing 3D profiles, causing directional lighting to create opposite shading patterns. For example, hedges lit from below (Figure 4.14c) are shaded on the top side in the image, but hedges lit from above are shaded on the bottom side in the image (Figure 4.14d). The opposite pattern is seen for ditches (Figure 4.14e, f). With the lighting-from-above prior, the hedges and ditches can look slightly more congruent in their 3D interpretation when they are lit from above (Figure 4.14d, f) than when they are lit from below (Figure 4.14c, e). Thus, lighting direction can be a primary cue to resolve the 3D profiles of these features. This experiment used image rotations to capture different response tendencies to different image orientations. Images were rotated in 15° steps, for a total of 24 rotations.

In the experiment, the honeycomb image was displayed as it is seen in Figure 4.14a and b, with no noise mask, for a total of 240 trials. But the hedges and ditches were masked with noise textures to prevent participants from basing their classifications on other features in the images, such as textures. The noise could serve to increase the reliance on directional lighting cues, leading the participants to judge the 3D profiles of the targets based on directional lighting and shading cues. The noise was weaker in this experiment compared to the main experiment, as pilot testing indicated that the noise mask had to be weaker for the observers to use lighting direction priors when the images could be presented in many orientations. See Figure 4.15 for example stimulus images containing two different hedges and two different ditches with added noise textures generated in PsychoPy. These four hedge and ditch images were used in the experiment to create 120 stimulus images per original image (480 total), all of which had different random noise textures. These hedge and ditch images were split across the 24 rotations, so that there were ten hedges and ten ditches per orientation.

The honeycomb and the aerial images (hedge and ditch) were separated into different conditions. Hedges and ditches were interleaved in the aerial image condition. There were five repeats of each condition, with 96 trials of the hedge/ditch and 48 trials of the honeycomb per condition. The conditions were randomly interleaved so that some participants always did the honeycomb condition first, followed by the hedge/ditch condition, and vice versa for other participants. In the honeycomb condition, participants responded ‘bump’ or ‘dimple’ to the central part of the honeycomb using buttons on their keyboard. In the hedge/ditch condition, hedges and ditches were interleaved, and participants responded ‘hedge’ or ‘ditch’ with button presses.

Prior to starting the experiment, participants were instructed that aerial images of hedges and ditches would be used, and they were shown images of a tall hedge and a deep ditch from the ground viewpoint. Instructions also showed a house from both the ground and aerial viewpoints,

illustrating the perspective switch that occurs from ground-to-aerial viewpoints. Participants were also shown example stimuli of hedges and ditches with noise masks, like the ones in Figure 4.15. No unmasked aerial images of hedges and ditches were shown. Before starting the experiment, participants were familiarized with 10 honeycomb and 20 hedge/ditch practise trials.

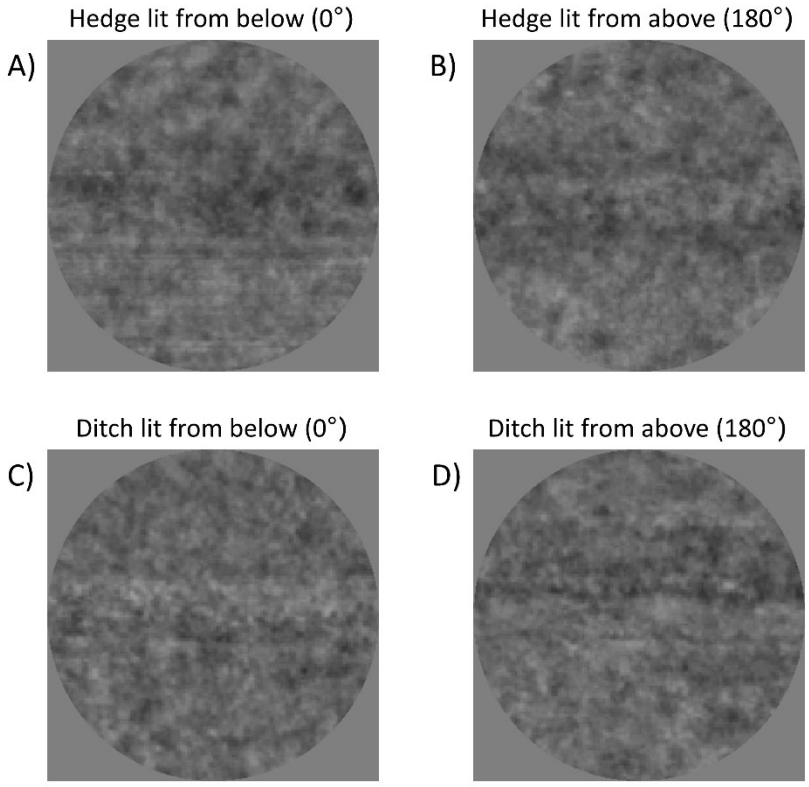

Figure 4.15: Example hedge and ditch stimulus images with two example orientations.

The experiment was created with JavaScript and PsychoJS and ran on the online platform Pavlovia (www.pavlovia.org). The experiment was advertised via email to remote sensing surveyors at the OS who had over one year of experience with surveying. Eleven surveyors participated and were compensated with £5. These were new participants compared to the previous experiment in this chapter, except one who volunteered in both experiments. The participants accessed and ran the experiment via a web browser on their own desktop computers in a quiet office environment during daytime hours. The computer, monitor, viewing distance, and testing environment were not otherwise controlled. Total time for completion was around 20 minutes. Participants provided informed consent by button press. The project was reviewed by Aston University's College of Health and Life Sciences Ethical Review committee (approval number 1843).

4.5.3 Results

The honeycomb is a well-established stimulus that is expected to capture lighting direction priors (Andrews et al., 2013; Pickard-Jones, d'Avossa & Sapir, 2020). In Figure 4.16, the black lines show the results of the honeycomb condition. Most participants classified the honeycomb as being convex when it was in orientations of around 0°, and concave when around 180° (Figure 4.14a, b). This is consistent with lighting-from-above, and the typical interpretation of Figure 4.14a and b. But one participant, Expert 3, show the opposite tendency, where the honeycomb was classified consistent with a lighting-from-below prior. In two other participants (Expert 5 and 8), the honeycomb failed to capture a lighting direction bias. Expert 5 showed an unexpected tendency to only respond 'concave' to the honeycomb stimulus in all orientations, and Expert 8 responded mainly 'convex' (Figure 4.16). Furthermore, the honeycomb data is slightly tilted to the left of upright (0°) in many participants, which is consistent with a known effect in lighting direction priors discovered by Sun and Perona (1998).

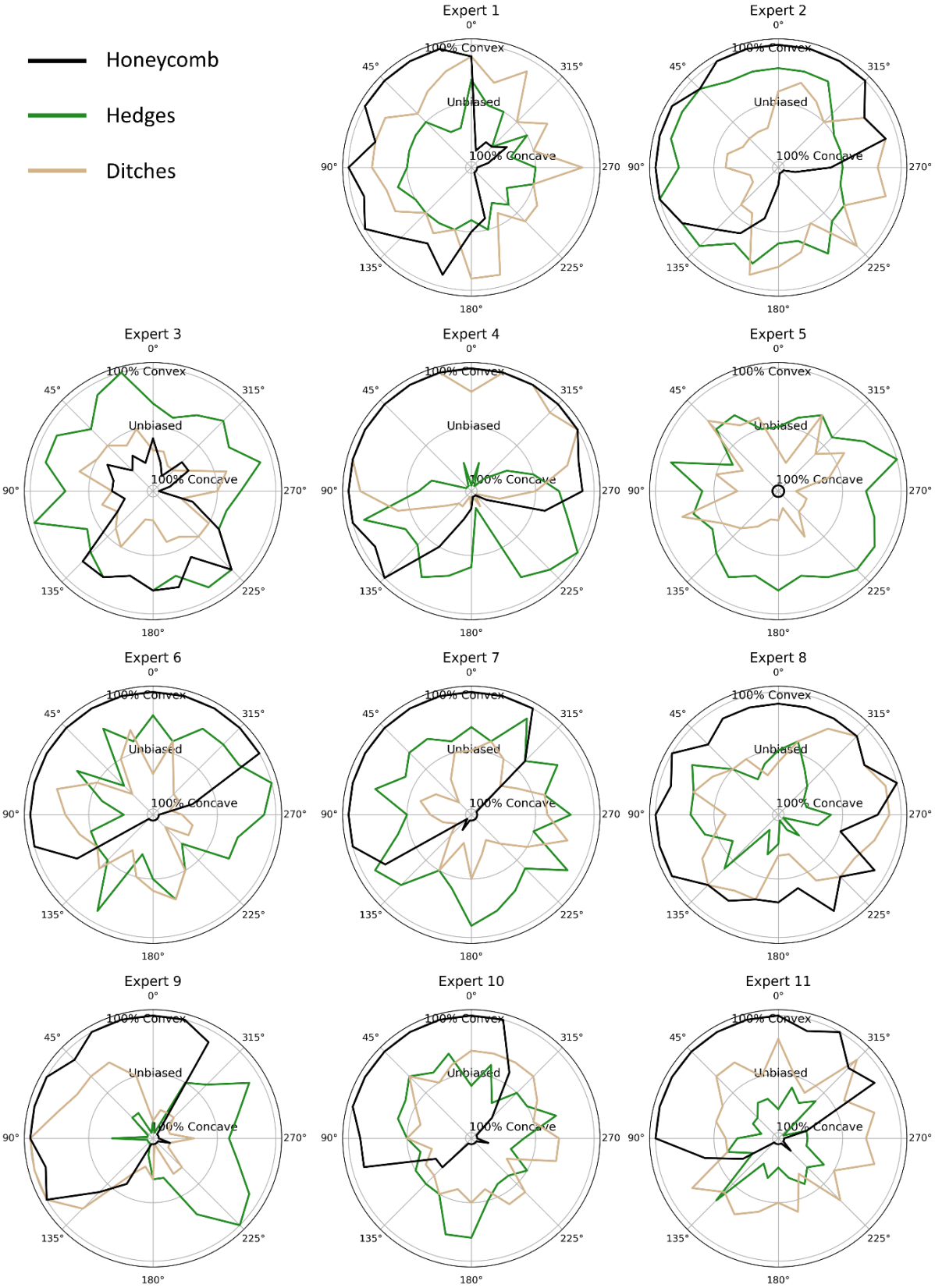

Figure 4.16: Polar plots for each individual participant showing the impact of image orientations on the three different image types. As indicated in the grids of each plot, the outer grid circle shows a rate of 100% convex responses, the middle grid circle shows an equal rate of convex and concave responses (i.e., unbiased), and the inner grid circle shows a rate of 100% concave responses. Note that, for hedges and ditches, a 0° orientation means that they were lit from below, and a 180° orientation means lit from above (Figure 4.14). Participants are ordered nominally, and the numbering is unrelated to the previous experiment.

Regarding the hedge/ditch condition, lighting direction priors were often not captured, as seen with the overall noisy data in Figure 4.16 for hedges (green lines) and ditches (tan lines). Note that the original orientation (0°) in the hedge and ditch images had light coming from below. Expert 4 provided data consistent with lighting-from-above for hedge/ditch as well as the honeycomb. To aid in interpretation of responses across orientations, the reader is encouraged to look at Expert 4 as a reference for lighting-from-above. Expert 2 shows an interesting tendency to classify the honeycomb according to lighting-from-above, but the hedge/ditch images more according to lighting-from-below, although these data are also noisy. This type of pattern was a priori expected in more participants, as it shows a context-specific lighting direction prior where stimuli outside the experts' domain, such as the honeycomb, is interpreted with lighting-from-above, but stimuli within the experts' domain are interpreted with lighting-from-below. Overall, this shorter online experiment failed to reliably capture lighting direction priors for the hedges and ditches in most participants.

4.5.4 Discussion

This follow-up experiment was exploratory and examined whether lighting direction priors could be captured for the honeycomb image and aerial images of hedges and ditches with a shorter online experiment. The honeycomb condition captured lighting direction priors in nine participants out of eleven. Out of these nine participants, all showed lighting-from-above priors except one who showed a lighting-from-below prior. The honeycomb is, however, not a domain-specific image to the expert surveyors, and the surveyors might apply a different bias for lighting directions in aerial images. But for the aerial images of hedges and ditches, the data are generally noisy and did not capture lighting direction priors, except in a few cases. These results suggest that directional lighting and shading cues mostly did not influence judgements in the aerial images, in contrast to the desired outcome in the current experiment.

This null result for the aerial images could be due to insufficient masking, where differences in e.g., textures between hedges and ditches could have driven responses, rather than directional lighting and shading cues. Ten images per orientation could also have been too small a number, and participants might have benefited from more task learning in each orientation in order for directional lighting cues to further influence judgements. Furthermore, this experiment ran online, without using a controlled lighting environment. This could have negatively impacted the ability to use subtle directional lighting cues in the hedges and ditches. But the honeycomb was able to capture lighting direction priors (Figure 4.16), likely because it provides a stronger directional lighting cue than the more subtle hedge/ditch stimulus images (Figure 4.14).

The current follow-up experiment differed from the main experiment in this chapter in some important respects. In the main experiment, participants classified aerial images of hedges and ditches that were either lit from above (180°) or below (0°) for 10,000 trials. This provided a large set of trials where participants had more chances to use subtle lighting direction cues that could vary between the images. Compared to the current follow-up experiment, the main experiment also used stronger noise (lower SNR), to decrease the influence of features such as textures in the original images. This, paired with many more trials on only two image orientations, could have served to increase the influence of directional lighting cues. The main experiment furthermore used a testing environment in a dark room, which is likely beneficial for discriminating subtle lighting direction cues.

To conclude, this follow-up experiment did not reliably capture lighting direction priors in aerial-view hedges and ditches, in contrast to the results of the main experiment where experts showed diminished or switched priors. However, this follow-up experiment does show that expert surveyors mostly interpret the honeycomb image with a lighting-from-above prior. This suggests that surveyors interpret the world outside of their domain according to lighting-from-above, but results from the main experiment suggest that this interpretation is diminished for aerial images. This might suggest that surveyors are combating the lighting-from-above prior when classifying aerial images, but use the lighting-from-above prior elsewhere. This topic is further discussed in Chapter 6.

Chapter 5

Characterising perceptual learning for stereopsis with stereoscopic classification images

5.1 Introduction

In Chapter 4, experts had a clear advantage over novices for sampling disparity cues in stereograms, showing visual expertise that likely developed from experience working with stereoscopic aerial images. The current study explored if such an advantage could be recreated in novices with a laboratory training intervention aimed to improve the ability to sample binocular disparity cues in stereograms. PL is the ability of perceptual systems to change depending on perceptual experiences, where experience with a particular array of stimuli improves processing of such stimuli (Gibson, 1963). In laboratory environments, PL interventions have been used to improve several aspects of early vision, and to examine how visual systems change with learning (Dosher & Lu, 2017; Goldstone, 1998; Lu & Dosher, 2022; Sagi, 2011; Seitz, 2017). Improvements in early visual processing have been found with, for example, visual acuity (Fahle, Edelman & Poggio, 1995), stereoacuity (Levi, Knill & Bavelier, 2015), orientation and spatial frequency (Fiorentini & Berardi, 1981), and motion (Ball & Sekuler, 1987). The current study used a PL approach, with a novel method to study PL in stereoscopic vision.

Some previous studies have examined PL with 2D CIs to examine how observers learn to use certain image cues. These studies have shown that PL is associated with improvements in internal templates, as observers become better at sampling relevant cues with learning (Gold, Sekuler & Bennett, 2004; Li, Levi & Klein, 2004). Such improvements have manifested as increases in template amplitude and/or spatial extent (Dobres & Seitz, 2010; Gold, Sekuler & Bennett, 2004; Kurki & Eckstein, 2014). The novel approach in the current study is based on examining stereoscopic PL with a 3D version of CIs based on stereograms (see Chapter 3 and 4).

Previous studies have investigated PL for stereoscopic vision, but not with CIs. Studies have investigated the efficacy of using PL interventions to recover stereopsis in participants with poor binocular vision (Birch, 2013; Ding & Levi, 2011; Godinez et al., 2021; Levi, 2022, 2023; Levi & Lee, 2009; Levi, Knill & Bavelier, 2015; McKee, Levi & Movshon, 2003; Rodán, Marroquín & García, 2022; Vedamurthy et al., 2016; Xi et al., 2014). In infancy, we learn to coordinate the inputs between the two eyes to develop binocular vision and a primary mechanism of depth perception – binocular stereopsis, which relies on the binocular disparity of images across the eyes (Howard, 2002). The ability to process binocular disparity is important for normal depth perception, but this can be impaired in those who suffer from amblyopia. Amblyopia is a neurodevelopmental abnormality associated with neural alterations in the visual pathways that usually originates from suppression of

one eye during a developmental period. Eye conditions such as strabismus, anisometropia, or cataracts can cause this suppression, as the developing brain prioritizes the development of neural pathways for one eye over the other. Amblyopia is associated with worse monocular vision (e.g., acuity in the affected eye), but also worse binocular vision and stereopsis as the two eyes can be poorly coordinated (Baker et al., 2007; Levi, 2020). Ding and Levi (2011) attempted to recover stereopsis in five human adults with subnormal binocular vision and stereopsis in a PL experiment. The authors combined monocular cues showing depth positions that were perfectly correlated with binocular disparity cues in gratings. The monocular cues helped to inform the participants of the disparity cues. These Stereodeficient participants significantly recovered stereopsis after many hours and thousands of training trials. Vedamurthy et al. (2016) also trained Stereodeficient participants using binocular disparity cues that could be in harmony or conflict with monocular texture cues. Their test required haptic interaction with a slanted surface seen through a virtual reality headset. After thousands of training trials across multiple weeks, most participants developed a greater reliance on binocular disparity cues relative to the texture cues. Individuals who learned to rely on disparity cues also tended to improve their stereoacuity. The results obtained by Vedamurthy et al. (2016) were stable at a 2-month follow-up after the training. In a review of different types of PL tasks, Levi, Knill and Bavelier (2015) found that stereoscopic tasks produce the highest recovery rate for stereopsis in Stereodeficient participants. Emerging evidence suggests that stereoscopic PL tasks can be based in immersive extended reality devices to stimulate recovery of stereopsis while providing a more tolerable task setting via video games or gamification of behavioural tasks (Coco-Martin et al., 2020; Foss, 2017; Godinez et al., 2021; Levi, 2023; Rodán, Marroquín & García, 2022; Vedamurthy et al., 2016). CIs, affording a method to study how visual cues are used, might provide another way of measuring improvements with PL in Stereodeficient observers.

The beneficial effect of stereoscopic PL has also been shown in neurotypical and stereo-normal observers (Fendick & Westheimer, 1983; Frisby & Clatworthy, 1975; Levi, 2022). Li et al. (2016) showed that stereoscopic PL with Gabor patch stimuli can transfer across the spatial frequency spectrum and across orthogonal orientations in the Gabor carriers (horizontal and vertical). These results suggest transfer of learning, but previous studies have highlighted transfer limitations in stereoscopic PL across orientations and retinal locations (Fahle, Edelman & Poggio, 1995; O'toole & Kersten, 1992; Ramachandran, 1976; Ramachandran & Braddick, 1973; Sowden et al., 1996). The issue of transfer and generalisability of stereoscopic PL for stereo-normal observers remains debated (Levi, 2022). Overall, evidence of PL from Stereodeficient and stereo-normal observers suggests that laboratory training can improve stereopsis, but we are yet to understand how generalised transfer of learning can be achieved.

The current study combined CIs with stereopsis training. The primary research question aimed to explore if and how stereo-normal participants improve their ability to sample binocular disparity cues, with the novel application of stereoscopic CIs. The stereoscopic CI technique developed in Chapter 3 was adapted for the current study. This technique relies on classification of targets defined by binocular disparity, embedded in random disparity noise. A primary goal of the experiment was to characterise CI template changes between the early and late parts of the experiment (Dobres & Seitz, 2010; Gold, Sekuler & Bennett, 2004). The CIs could capture details about how internal templates are associated with improvements from PL. There are two main possibilities regarding how templates might reveal improvements: 1) Increases in template amplitudes would suggest that more relevant binocular disparity cues are sampled from a location, or 2) Increases in spatial extent would suggest that disparity cues are sampled from a larger area. PL should lead to increases in one or both aspects of template shape.

In addition, the main experiment below used an adaptive staircase procedure to estimate a fixed rate of correct responses (Levitt, 1971). This technique maintains a constant rate of correct responses by adaptively manipulating the SNR throughout the experiment. Thresholds were estimated from response accuracies across different levels of SNRs. Thresholds were expected to change with learning throughout the experiment, where evidence of learning would be seen with increased tolerance to external noise (decreased SNRs). This would indicate learning as the staircase procedure must provide a more difficult task by adding more external noise to maintain the fixed rate of correct responses (see Procedure for further details). SNRs at threshold were thus analysed to provide an additional, and more conventional, measure of PL.

To promote learning, testing was divided over five different days, as sleep can consolidate PL (Karni et al., 1994; Karni & Sagi, 1993; Stickgold, James & Hobson, 2000). Participants were also provided with trial-by-trial feedback (auditory), which can increase learning by guiding participants to home in on diagnostic cues (Aberg & Herzog, 2012; Herzog & Fahle, 1997; Liu, Dosher & Lu, 2014; Liu, Lu & Dosher, 2010, 2012; Shibata et al., 2009).

5.2 Pilot experiment

This chapter includes two experiments on stereoscopic PL, the first of which is briefly described here under the label of a pilot experiment. This initial experiment produced a null result, but was informative to the design of the second experiment (Main experiment).

5.2.1 Method

A PL intervention was designed based on the hedge-ditch discrimination experiment in Chapter 4, with some additional manipulations aimed at promoting learning. As in Chapter 4, CIs could be generated from binocular disparity and luminance. The static SNR used for the hedge and ditch images with noise in Chapter 4 was increased by ~25%, increasing discriminability. In preparatory evaluation with this SNR, a very experienced psychophysical observer provided 60-65% correct responses. SNRs were constrained with these hedge and ditch images, as: 1) sufficient modulating disparity noise was required to provide disparity CIs, and 2) luminance noise had to sufficiently mask the original images to prevent participants from judging images based on features such as textures in the hedges and ditches. Auditory feedback was provided to half the participants, and correct responses were defined based on detecting the disparity profiles of the targets. The experimenter specifically instructed participants to try to judge disparity profiles, described to the participants as height and depth. The increased discriminability of the hedge and ditch images (compared to Chapter 4), provision of trial-by-trial response feedback, and division of the experiment across multiple days are factors which should help to promote PL (see Introduction). Apart from these factors, the stimulus images, screening procedure, experimental procedure, equipment, and ethical considerations were largely identical to those of Chapter 4.

Eight psychology undergraduate students enrolled at Aston University, Birmingham, UK, signed informed consent and were compensated at a rate of £10 an hour (three participants), or with credits in a research participation scheme (five participants). Participation required five visits on different days, within a 10-day period. Each visit took around 50 minutes. Participants completed 7,000 trials in total, which were split into 28 shorter sessions of 250 trials each.

The experiment further included an orthogonal transfer task, with new hedges and ditches that were vertically arranged in the images, to examine transfer of learning from the horizontal (Chapter 4) to the vertical targets. The experiment began and ended with two transfer sessions.

5.2.2 Results and discussion

The main results of this experiment are displayed in percent correct responses across sessions (Figure 5.1). The accuracy did not improve across sessions, as shown by a linear regression model fit to the averaged data, excluding the transfer sessions (Figure 5.1: Average; slope estimate = 0.044, $t = 1.36$, $p = 0.187$). This was also the case in comparison between the transfer sessions (before and after) ($t(1) = -9.00$, $p = 0.070$). This unexpected null result was likely due to the difficulty of discriminating the target images, as all participants were at, or close to, chance level accuracy throughout the experiment (Figure 5.1). As participants struggled to reliably discriminate the

disparity profiles in the hedge and ditch targets, their performance level was likely below a level required to promote learning. PL involves learning how to exploit diagnostic target cues, but in this experiment, targets were likely too corrupted by external noise for the participants to reliably learn from their cues, with and without feedback.

This null result was similarly found with CIs. Disparity CIs were constructed for comparison across the beginning and end parts of the experiment to examine effects of learning (Dobres & Seitz, 2010; Gold, Bennett & Sekuler, 2004). However, most participants produced CIs with templates that were too weak to characterize, and thus could not be used to estimate differences across different parts of the experiment.

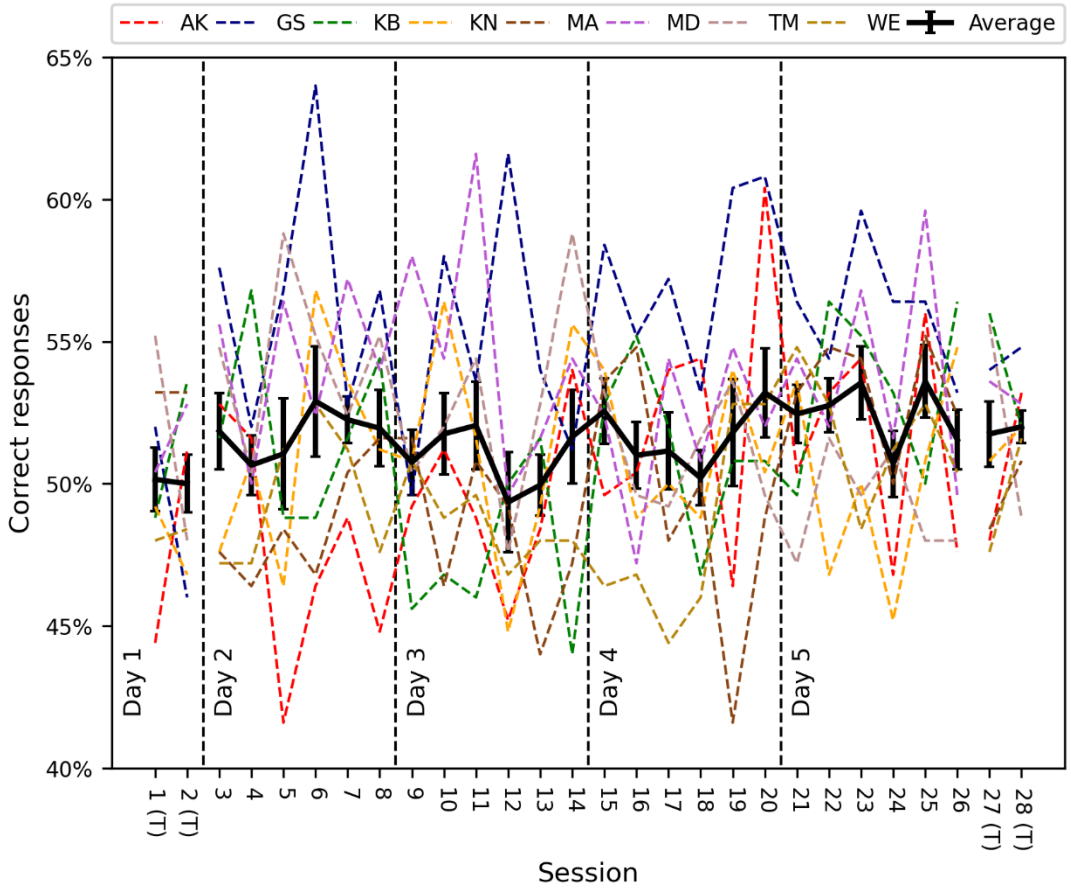

Figure 5.1: Correct responses across sessions for individual participants (coloured dashed lines), and their average (solid black line). A rate of 50% correct responses is chance. Sessions 1, 2, 27, and 28 are transfer sessions (T), which used vertical rather than horizontal targets. Error bars are SEM.

This null result inspired a further experiment, which is the main subject of this Chapter. The main experiment followed a similar procedure as the pilot experiment, but with some stimulus and procedural changes aimed at securing a higher correct response rate. A primary difference between the pilot experiment and the main experiment is the use of targets defined as pedestals in the noise textures themselves, rather than landscape images added to noise textures. This change facilitated

the use of stronger disparity signals, which were more discriminable than those of the hedge and ditch images with added luminance noise textures (Chapter 4). A lesson from the pilot experiment was that the targets were too difficult to discriminate (Figure 5.1). To ensure a higher and balanced rate of correct responses, an adaptive staircase procedure balanced the SNR to achieve a threshold of 70.7% correct responses, described below (Procedure). This staircase procedure ensured that a suitable SNR would be used throughout the experiment.

5.3 Main experiment

5.3.1 Method

5.3.1.1 Stimulus images

A unique white noise texture, with a non-zero mean, was generated on each trial (128x128 pixels). This texture was low pass filtered using a Butterworth filter with a cut-off frequency of 9 cycles per image. These textures were generated in the same way as those of Chapter 3 and 4. Targets and noise defined by binocular disparity were imposed onto the luminance textures. Figure 5.2 shows the disparity-defined targets and the process for adding disparity noise. Chapter 3 and 4 describes in detail the procedure for generating disparity noise by random horizontal displacements within carrier textures. The targets were disparity pedestals analogous to the disparity profiles of the stereoscopic images of hedges and ditches in Chapter 4. The 20 central rows of the noise images were manipulated to have 37.5 arcseconds of crossed or uncrossed disparity, and the rest of the image had 37.5 arcseconds of the opposite sign of disparity. A crossed ('tall') target with an uncrossed background (Figure 5.2a) therefore had 75 arcseconds more crossed disparity than its stimulus surround, and vice versa with uncrossed ('deep') targets (Figure 5.2b). These targets were horizontally oriented, with a height and width of 1.04 and 6.66 degrees of visual angle, respectively. The width corresponds to the full width of the stimulus images.

A transfer task was included to investigate transfer of learning to a new stimulus orientation. The effects of PL interventions are more useful for translational benefits if they can transfer to, for example, different retinal locations and orientations. This transfer task used vertical rather than horizontal targets to investigate transfer of learning to orthogonal targets. Apart from a changed orientation, the vertical targets were otherwise identical to the horizontal targets. This vertical transfer target was located in the 20 central columns (rather than the 20 central rows). This target had a height and width of 6.66 and 1.04 degrees of visual angle, respectively.

5.3.1.2 Equipment

Participants were seated in front of a mirror stereoscope with their chins on a chinrest. Two front-surface mirrors angled at 45° were mounted 6 cm in front of the participant. These directed images to the observer from two ASUS ProArt PA329C monitors (3840 x 2160 pixel, 710x405 mm active screen region) placed on either side of the mirror mount with a total viewing distance of 100 cm. These monitors provided the only light source in the room. Stimulus images subtended 6.66 degrees of visual angle and were scaled in PsychoPy (Version 2020.2.10; Peirce et al., 2019) so that a single element in the stimulus occupied 5x5 pixels on the monitors. Noise textures were generated and presented with PsychoPy with a modified version of the noise component. Stimulus images were linearised with inverse gamma functions for each display to ensure that luminance was linear in the displays.

5.3.1.3 Participants

Eight participants completed the experiment (4 female; Mean age: 22.25, SD: 3.5), but one was excluded from analysis (see below). Three other potential participants failed to pass a screening test prior to starting (see below), and one other withdrew from the study after starting. Participants were undergraduate optometry students enrolled at Aston University, Birmingham, UK. They were recruited via email advertisement and were compensated at a rate of £15 per visit (£75 for completing the five days). Participants gave informed consent, and the project was reviewed by Aston University's College of Health and Life Sciences Ethical Review committee (approval number 1843).

5.3.1.4 Screening and exclusion procedure

The participants self-reported having normal or corrected-to-normal eyesight, and wore their normal eyewear where applicable. No participant used bifocal or varifocal lenses. A three-stage screening procedure assessed the stereopsis of potential participants. The first part involved the TNO test for stereoscopic vision, which is based on random-dot stereograms that provide no monocular cues to the target. No exclusion criterion was set for TNO thresholds. TNO thresholds varied between 15 and 120 arcseconds across participants (Median: 60; Mean: 54, SD: 30). Participants then carried out a discrimination task using images in the mirror-stereoscope (40 trials) where a central disparity-defined square (750 arcseconds of disparity, side length 1.04 degrees of visual angle) had either crossed or uncrossed disparities. The task was to report whether the square was in a 'tall' or 'deep' depth plane compared to the surround. Responses were made by pressing a button on a keyboard. Auditory feedback was provided after each response via a headset where a beep or buzz would

indicate a correct or incorrect response, respectively. Participants had to score above 90% correct to pass this test and three potential participants were excluded based on this. The last part of the screening procedure involved familiarisation with the experiment's stimulus images which contained disparity targets without disparity noise (Figure 5.2a, b). Participants responded 'tall' and 'deep' with the right and left arrow keys, respectively, with the same auditory feedback. Participants had to give 10 correct responses in a row to pass. No participant was excluded based on this.

One participant who completed the experiment was excluded due to having failed to reliably reach a required rate of correct responses (~70%), despite the absence of external noise in most of their experiment sessions (see below regarding staircase procedure). Because the staircase set their noise level to zero, there was no noise that could contribute to a CI, thus data was missing for this participant.

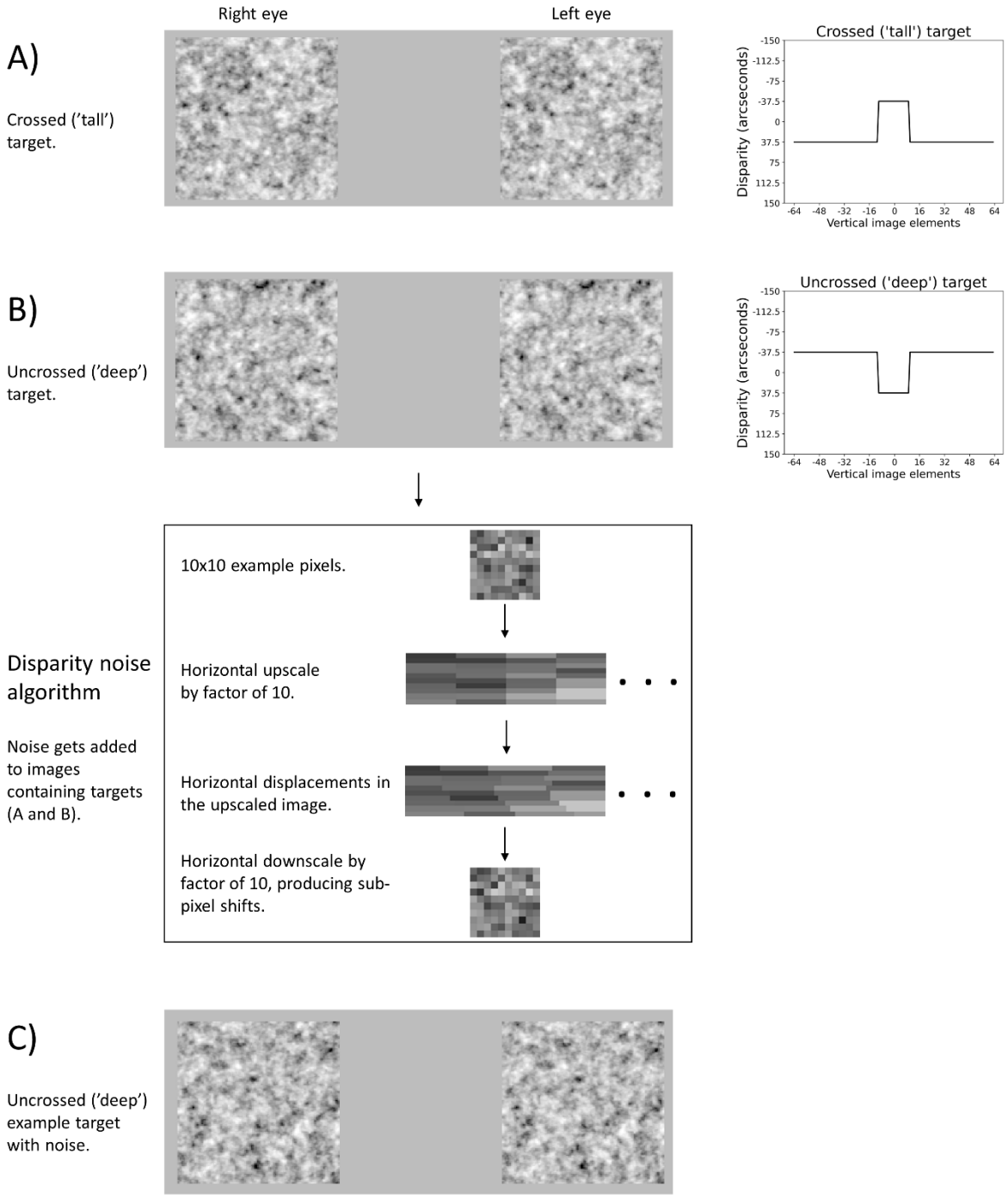

Figure 5.2: Procedure for generating stimulus images by introducing disparity noise. Images are arranged for crossed fusion: a) tall target without noise, b) deep target without noise, and c) an example deep target with added noise. Divergent fusion reverses the disparity profiles. The plots on the right of a) and b) display vertical image elements on the x-axis, where left-right corresponds to top-down in the image, with 0 as the centre. See text for further details.

5.3.1.5 Procedure

Participants were familiarised with the stereoscope and with responding to noise-free stimulus images during the screening procedure (Figure 5.2a, b). Familiarity with stereoscopic images can be important in inexperienced observers (Ramachandran, 1976). Prior to starting the

experiment, participants were informed that the experiment consisted of the same targets and button press responses, but that the task would be more difficult to perform. The experimenter explained that the stimulus images would appear noisier, and that these otherwise obvious targets would be masked by noise such that the participant should expect to give 70% correct responses. Trial-by-trial auditory feedback was provided to all participants throughout the experiment.

To ensure appropriate vergence control in the mirror-stereoscope, the fixation cross and fusion border described in Chapter 4 were used. A small black fixation cross was presented in the centre of the screen between each trial. The vertical bar of the cross was split across the two eyes. To achieve good convergence, participants were instructed to fuse the cross to make it appear 'complete', like a '+'. Participants were instructed to wait with responses and to attempt to 'reset' their convergence if the cross appeared malformed due to imperfect fusion in the stereoscope. The fixation cross was removed when stimuli were displayed. To further support fusion, a zero-disparity high contrast border comprising white rectangles on a black background surrounded the stimulus images.

An adaptive staircase procedure was used to vary SNRs by, unconventionally, manipulating the disparity noise level rather than the signal. This could provide psychophysically more granular step sizes in the experimental software than if the signal level was manipulated, as the signal was defined by fewer sub-pixel steps than the total noise range. SNRs were varied in a 1-up, 2-down step procedure, designed to estimate a 70.7% threshold (Levitt, 1971). When the procedure determined that the SNR should go up, the noise level was reduced, and vice versa. Two such staircases operated in parallel. Step sizes were logarithmic, and the smallest step size was 1 decibel (dB). At the beginning of each session, the staircases used larger steps to help track thresholds more quickly. These steps started with three iterations of three 1dB steps, followed by three iterations of two 1dB steps, and the rest of the session used 1dB steps. Staircase reversals would trigger the change in step sizes. In the experiment software, steps were defined as a range of disparity noise in arcseconds. Prior to generating a stimulus image on each trial (Figure 5.2), steps were quantised to units of 37.5 arcseconds of disparity. The quantised and logarithmic nature of the step sizes meant that, when external noise was low, the SNR might be unaltered by a single step in the staircase. Conversely, when external noise was high, the SNR could move across multiple quantised steps in one staircase step. The threshold estimate from each session was carried over to the next session, except that at the start of each day the noise level was approximately halved to help participants see the targets again.

The main part of the experiment was structured into shorter sessions of 256 trials, and participants did up to six of these in one day, which took around 50 minutes. Across five days of

testing, each participant completed 26 sessions for a total of 6,656 trials and a total completion time of around 4 hours. In the transfer task, participants did 256 trials of the same task with the same feedback but with the vertical targets. To allow for initial learning (familiarity and learning how to do the task), only the second half of this data was used in analysis, and the first 128 trials of the transfer task were treated as practise trials, both before and after the main experiment. The first day of participation consisted of signing informed consent, the screening procedure, the transfer task, and two sessions of the main experiment. The second, third, and fourth day each consisted of six sessions of the main experiment. The fifth and final day consisted of four sessions of the main experiment, followed by the transfer task, and finishing with a debrief from the experimenter. Participants had to complete the experiment within ten days of starting.

5.3.2 Results

5.3.2.1 Thresholds

Each 256-trial session employed different levels of external noise depending on the staircase adjustments of SNRs. Thresholds were estimated from the participant's percent correct responses at the six different noise levels where the most responses occurred in each session. These noise levels were the quantised noise levels that correspond to what was displayed on the screens. For each session, an inverse Weibull function was fitted to these data:

$$x = \alpha\left(-\log\left(\frac{1-y}{0.5}\right)\right)^{\frac{1}{\beta}}, \qquad \text{(Equation 1)}$$

and the noise level corresponding to 70.7% correct responses was estimated from the fitted curve. These noise levels at thresholds were recorded for each session and participant, and are shown in Figure 5.3. Note that, as the adaptive staircase adjusted the level of external noise, and not the target contrast, evidence of learning should be seen with increasing rather than decreasing thresholds in Figure 5.3. Increased noise implies a lower SNR and thus greater ability to detect the target signal. Some participants failed to achieve responses at the 70% correct level in the first transfer session, and the staircase thus tended to add no external noise. The fits to the staircase data thus tended to estimate their thresholds at close to zero noise in this session (Figure 5.3: "session 1 (T)").

On average, threshold noise levels increased between the first and last sessions (Figure 5.3). Increases in threshold noise levels suggest improving performance, as more noise is required to maintain the threshold. These data (excluding the transfer sessions) were fitted with linear regression models to examine slopes, where positive slopes show a performance increase across

sessions. Table 5.1 shows the individual slopes for each participant. On the individual level, these varied in statistical outcomes, and were significant for some and non-significant for other participants. These individual slopes were further tested as a distribution in a one-sample t-test, which was significantly above zero ($t(6) = 2.90$, $p = 0.027$). The average of all participants (Figure 5.3; Table 5.1) produced a highly significant positive slope ($t = 6.01$, $p < 0.001$). Overall, threshold estimates across sessions show improvements, indicating learning.

To examine transfer of learning to the vertical targets, results suggest that the threshold improvements throughout the experiment transferred to the orthogonal targets. In Figure 5.3, the transfer sessions (1 and 26) significantly differed ($t(6) = -3.48$, $p = 0.039$). The last session of the main experiment (session 25) was also compared to the transfer sessions, showing a significant difference from the first transfer session ($t(6) = -5.25$, $p = 0.006$), but not the last transfer session ($t(6) = 1.76$, $p = 0.384$). These paired samples t-tests were Bonferroni corrected for three comparisons.

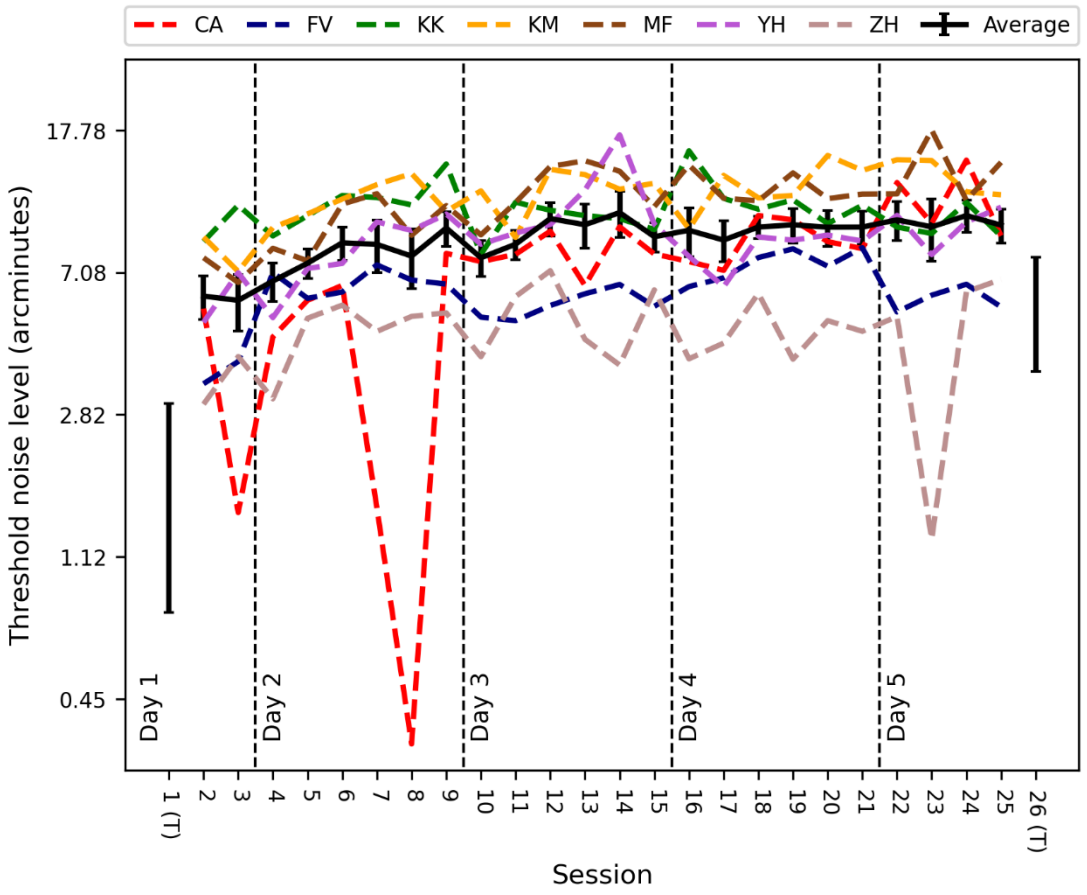

Figure 5.3: Thresholds across sessions for all seven participants. Sessions are ordered in temporal order. Sessions 1 and 26 were transfer task sessions (T), which used vertical rather than horizontal targets. Transfer task sessions only show the average for display purposes, as individual data were sometimes close to zero thus falling far outside the y-axis ($\log_{1dB}$). Error bars are SEM.

Participant	Slope estimate	R^2	*t*	*p*
CA	0.350	0.568	5.38	< .001
FV	0.0669	0.163	2.07	0.051
KK	-0.00795	0.001	-0.157	0.877
KM	0.174	0.359	3.51	0.002
MF	0.253	0.471	4.42	< .001
YH	0.114	0.112	1.66	0.111
ZH	0.0303	0.026	0.763	0.454
Average	0.140	0.621	6.01	< .001

Table 5.1: Individual slopes estimated from linear regression model fits to the threshold data in Figure 5.3, excluding data from the transfer sessions.

5.3.2.2 Classification images

CIs were generated from disparity noise textures. Noise textures were saved based on the response given on each trial, and these were combined into compound images for each response category. 'Deep' response compounds were subtracted from 'tall' response compounds to generate a disparity CI for each participant (Ahumada 1996; Murray, 2011). Partial disparity CIs are shown in Figure 5.4, where CIs from the first eight sessions of the main experiment can be seen next to the CIs from the last eight sessions. These batches of eight sessions correspond to the first and last thirds of the experiment, excluding transfer sessions. Light and dark pixels represent crossed and uncrossed disparity, respectively. CIs generally contain templates with a central positive peak (light pixels) in the target location, with negative side-lobes above and below the peak (dark pixels). Although difficult to discern in the raw CIs, templates appear generally stronger with the last sessions' CIs compared to the first sessions' CIs (Figure 5.4).

To quantify such differences in template shapes across the first and last sessions, these partial CIs were decomposed into vertical cross-sections by averaging the columns in the CIs. These vertical cross-sections were then characterised by fitting with Gabor functions:

$$f(y) = A\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)\cos\left(2\pi\frac{y}{\lambda}-\psi\right), \qquad \text{(Equation 2)}$$

where *y* are vertical image elements (in pixels units, with 0 in the centre). *A* is amplitude, μ is spatial offset (in pixels), σ is spread (standard deviation in pixels), λ is wavelength (in pixels) and ψ is the absolute phase offset (in radians).

Figure 5.5 shows cross section and function fitting results. The data and curves generally contain central peaks with negative side-lobes. Templates significantly increased in amplitudes from the first to the last sessions ($t(6) = -3.09$, $p = 0.021$). Amplitude parameter, *A*, values are thus included in the plots in Figure 5.5. Amplitudes were overall larger in the last compared to the first

sessions, but participants FV and YH did not show this tendency. This increase suggests that participants learned to better sample disparity cues in the stimulus images.

Templates remained more uniform from the first to the last sessions in terms of parameters related to spatial extent: wavelength λ ($t(6) = 0.06$, $p = 0.954$) and spread σ ($t(6) = 1.85$, $p = 0.114$). Increases in such parameters could otherwise have suggested that participants develop an ability to sample a larger area in the images, but this was not seen. The other two parameters, spatial offset (μ) and absolute phase offset (ψ), relate to template offsets captured by the Gaussian and cosine components of the Gabor function, and carry no direct relevance to the current research questions. Note, however, that the positive peaks in Figure 5.5 coincided with the middle of the images (0 on the x-axis). This shows that all participants sampled disparity cues within the bounds of the target location (±10 on the x-axis).

In examining the relationship between the two different measures of learning, CI amplitude improvements (Figure 5.5; $A_{Last} - A_{First}$) did not correlate with estimated slopes from thresholds across sessions (Table 5.1) ($r = 0.277$, $p = 0.548$). This outcome was surprising, as we might expect two different measures that both indicate PL to be correlated. Three participants (CA, KM, and MF) showed both significant improvements in thresholds (positive slopes) across sessions and increased CI amplitudes. But two participants (KK and ZH) showed no threshold improvements yet demonstrated increased CI amplitudes. Curiously, these two participants showed a tendency to better sample binocular disparity cues without a decrease in their threshold SNR.

CIs from luminance textures were also generated, and are presented in Appendix E. Although luminance cues carried no diagnostic information for the task of discriminating tall and deep disparity profiles, three participants (CA, KM, and ZH) used luminance cues to some extent, producing CI templates that were observable with the above cross-section and function fitting procedure (not shown). This can be understood in terms of cue combination of disparity and luminance. As light and dark textures appear stereoscopically near and far, respectively, these cues can be combined with crossed and uncrossed disparity cues to support a stronger impression of depth (Chen & Tyler, 2015; Doorschot, Kappers & Koenderink 2001; Egusa, 1983; Hartle et al., 2022; Langer & Zucker, 1994; Langer and Bülthoff, 2000; Lovell, Bloj & Harris, 2012; O'Shea, Blackburn & Ono, 1994; Schofield, Rock & Georgeson, 2011; Sun & Schofield, 2012). This experiment required discrimination of disparity profiles, and all participants used disparity cues, but three participants also used luminance cues to some extent. The tendency to use luminance cues for some participants bears similarity to the results of Chapter 4, although less so, as the hedges and ditches in Chapter 4 carried usable luminance cues to support depth perception. In this experiment, the targets carried

no usable luminance cues although three participants still used luminance cues in combination with disparity cues.

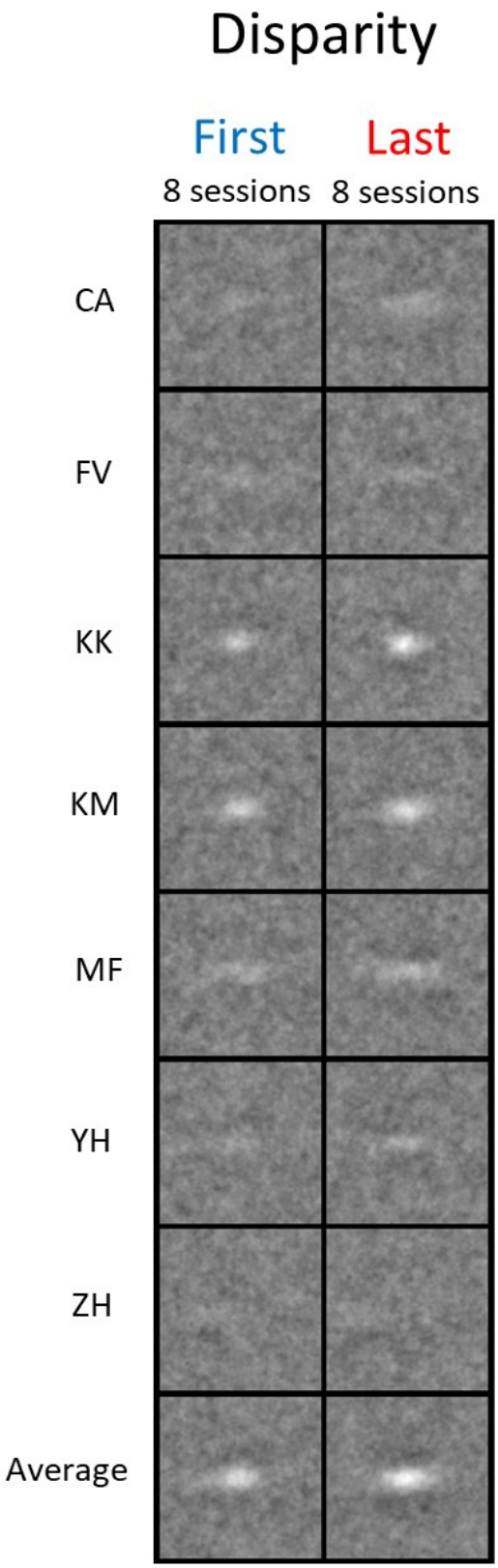

Figure 5.4: Partial disparity classification images for each participant, generated from the first and last thirds of the main part of the experiment. See text for details.

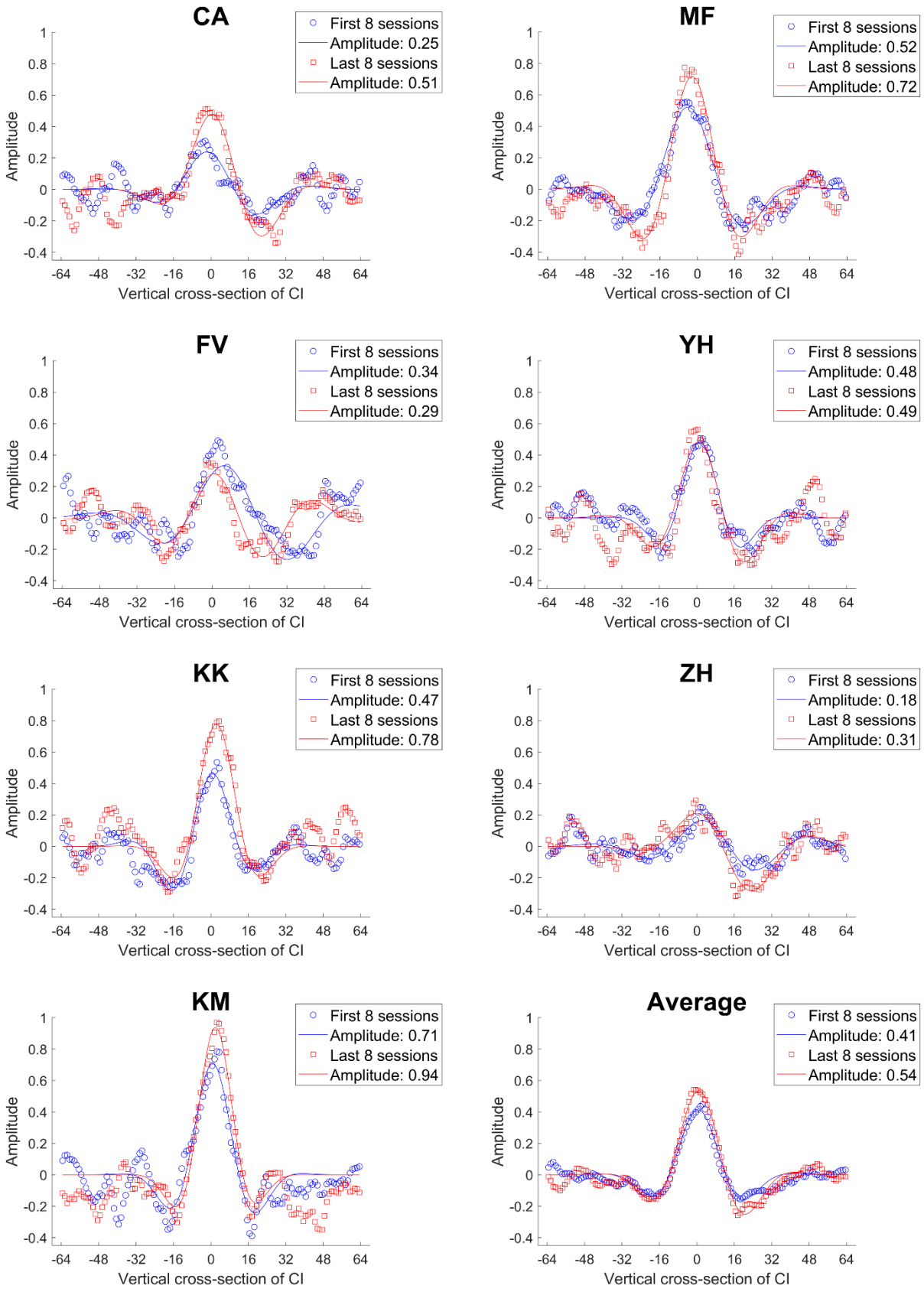

Figure 5.5: Vertical cross-sections of partial disparity classification images, from the first and last eight sessions of the main part of the experiment. Curves show the fitted Gabor function, and amplitude parameters are given in each plot legend.

5.4 Discussion

The primary aim of the current study was to explore if and how stereo-normal participants improve their ability to sample binocular disparity cues. The stereoscopic PL intervention led to improvements in both main outcome measures. Firstly, thresholds estimated with an adaptive staircase procedure showed that, as the experiment progressed, more external noise was generally required to maintain a fixed rate of correct responses (Figure 5.3). That is, signal detection performance improved. Secondly, CIs generated from binocular disparity cues showed improved templates from the first to the last thirds of the experiment (Figure 5.4 and 5.5). Templates increased in amplitudes but not spatial extent[11], suggesting that participants learned to better sample binocular disparity cues from a select location.

The stereoscopic CIs showed novel evidence that links stereoscopic PL to improvements in internal templates. This link has previously been demonstrated in different 2D tasks (Dobres & Seitz, 2010; Gold, Sekuler & Bennett, 2004). The combination of threshold estimation and CIs affords two routes to studying learning. Taken together, the results from these two measures suggest strong evidence of PL for stereopsis. CIs provide a rather unique measure of observer performance in revealing internal templates, which are difficult to estimate with other psychophysical techniques. This technique could be applied in further investigation of the mechanisms of PL. For example, CIs could characterise the ability to sample binocular disparity cues in Stereodeficient observers, and reveal potential improvements from PL. Both CIs and PL benefit from long experiments with thousands of trials. Thus, future studies may add CIs to their PL experiments to gain an additional route to studying learning effects with little extra cost to participation time. Adding CIs requires a target-modulating random component to the stimulus images, such as added noise textures (Dobres & Seitz, 2010; Gold, Sekuler & Bennett, 2004) or random modulations to the target itself in e.g., position (Kuai, Levi & Kourtzi, 2013; Kurki & Eckstein, 2014; Li, Levi & Klein, 2004).

The results regarding the transfer task suggest that learning in the main part of the experiment transferred across orthogonal orientations (horizontal to vertical). Studies have shown that transfer across retinal locations and orientations can be limited (O'toole & Kersten, 1992; Ramachandran, 1976; Ramachandran & Braddick, 1973), but Li et al. (2016) found evidence of transfer across orthogonally orientated Gabor patch carriers (in the same retinal location). The current study controlled for initial learning in the transfer task by providing 128 practise trials with the vertical target. Vergence control was also controlled for prior to starting any experimental sessions. A limitation in the design in the current study is that the horizontal and vertical targets

[11] See Chapter 6 for a discussion on mechanisms associated with template shapes.

shared a central overlapping area (a square 1.04 degrees of visual angle), thus no transfer was required to discriminate crossed and uncrossed disparities in this small central area[12].

The pilot and main experiments shared many similarities, but also differences in some important respects. The two experiments differed in choice of targets and SNRs, and in participant details. Regarding choice of targets and SNRs, the pilot experiment was instructive in showing that the hedge and ditch targets were too difficult to discriminate, thus not likely to facilitate PL (see above: Pilot experiment). The main experiment thus utilized new targets, defined as pedestals of crossed and uncrossed disparity. These new targets facilitated the use of stronger disparity signals compared to the noise-masked hedges and ditches of Chapter 5. Furthermore, to achieve consistent and standardised SNRs, the main experiment used an adaptive staircase procedure, moving away from the static SNR in the pilot experiment. Regarding participant details, the pilot and main experiment used psychology and optometry undergraduate students, respectively. The training experience of these groups differ: optometry students are trained to administer perceptual tasks with forced choice judgments and are familiar with this paradigm; Psychology students are more familiar with questionnaire studies and reaction time experiments. The participants were also compensated differently, where the psychology students received either credits (five participants) or £10 per visit (three participants), but all optometry students received £15 per visit. Such differences in training and compensation could impact the participants' attention throughout the experiment in favour of the optometry students.

Regarding the broader context of this thesis, in Chapter 4, expert remote sensing surveyors showed a large increase in sampling rate of disparity cues compared to novices. The current study and previous literature suggest that PL is a likely mechanism that contributes to the development of this expertise. Future studies could aim to expand these results through e.g., stereopsis training involving multiple and more diverse targets that might promote more general learning. The issue of whether stereopsis can be generally improved with laboratory training in stereo-normal adults is still debated, and more research is required (Levi, 2022; Lu, Lin & Dosher, 2016). If future work can identify if and how stereopsis can be improved, laboratory PL interventions could be used in workplaces such as in remote sensing surveying as supplementary training for inexperienced surveyors (see also Future directions in Chapter 6).

[12] This central area was foveated following the instructions to look at the central fixation cross. Although this central area constituted only 15.6% of the target areas, the classification images from the horizontal target clearly show that much cue sampling occurred in this area (Figure 5.4). The results of the transfer task would thus have been better controlled with the addition of a condition where the target is defined only in the small central area (like the square target in Chapter 3). A comparison between this and the vertical target could afford a measure of how learning on the horizontal target might transfer to the 'tails' of the vertical target, beyond the central area.

Chapter 6

Discussion

6.1 Meeting the aims of the thesis

This thesis set out to test four broad hypotheses regarding the mechanisms associated with expertise in remote sensing surveying of stereoscopic aerial landscape images. First, the unfamiliar aerial viewpoint is more difficult to process, but expert surveyors are better at processing the aerial viewpoint. Next, surveyors are experienced with using binocular disparity cues in stereoscopic aerial images, and they are better able to process this cue. The surveyors also adapt to the aerial imagery, and this can alter perceptual priors for interpreting shape from shading. Finally, the surveyors develop their expertise from experience, and the expertise can in part be explained by perceptual learning (PL).

Discussions with Ordnance Survey (OS) remote sensing surveyors helped to develop these broad hypotheses into elaborated and more specific research questions. This section provides an overview of the specific aims of this thesis, and the following section provides a more detailed discussion of the results of the studies.

The first study aimed to explore the effects of surveyor experience with processing aerial viewpoints (Chapter 2). This aim was met with two experiments that examined scene gist, and object matching, with images of both scenes and objects seen from both the ground and aerial viewpoints. This study provided evidence that surveyors have a superior ability to process both scenes and objects from the aerial viewpoint. The results also suggest that the surveyors have an advantage with expertise for processing the featural configurations typical of aerial images.

Having established that experts show evidence of expertise for detecting diagnostic features in 2D aerial images, later work aimed to explore classification of stereoscopic features that support perception of 3D shape in aerial images. To provide a suitable method for this, Chapter 3 aimed to develop a novel version of classification images (CIs) that could simultaneously estimate CIs from binocular disparity and luminance cues. This CI technique was developed in three stages of pilot experiments, and was applied in the last two empirical chapters of this thesis.

The aims of Chapter 4 sought to discover how experts and novices use different stereoscopic cues when discriminating aerial images of hedges and ditches. The CI technique served to capture the use of binocular disparity cues, diffuse luminance (‘dark-is-deep’), and lighting direction priors. The use of these different image cues revealed group differences related to visual expertise in remote sensing surveyors, discussed in detail below.

A result of Chapter 4 showed that better sampling of binocular disparity cues is a hallmark of expertise in expert surveyors. Chapter 5 followed up on these results with a stereoscopic PL intervention aimed to characterise learning to better sample disparity cues in stereograms. This study used CIs to characterise changes to internal templates that occurred with learning.

6.2 Overview of results

The main empirical chapters in this thesis paint a picture of expertise in remote sensing surveying. The results show that expertise is associated with more accurate scene categorisation in briefly presented aerial images, more accurate object identification across aerial and ground viewpoints, and an improved ability to sample binocular disparity cues in stereograms. The results also suggest that surveying is associated with a modified interpretation of the lighting-from-above prior, and that PL is a likely contributor to developing expertise for stereograms.

6.2.1 Chapter 2: Expertise in the aerial viewpoint

Chapter 2 shows evidence of expertise in tasks related to higher levels of visual perception, using natural images of scenes and objects seen from the aerial and ground viewpoints. A primary challenge with remote sensing surveying is the reliance on aerial images, as aerial viewpoints provide unusual views of the world. The results suggest that surveyors have overcome some of this difficulty with experience.

Expert surveyors were more accurate than novices at categorising briefly presented (100ms) scenes from aerial but not ground viewpoints. This suggests improved gist processing of aerial scenes, where experience brings processing benefits in the first moments of stimulus exposure. Expert surveyors were also better able to judge the identity of objects across viewpoints. This second experiment suggests that experience with featural configurations in aerial viewpoints allow greater recognition across viewpoints. These results together show that expert surveyors have developed expertise for aerial viewpoints, similar to how we develop expertise for ground viewpoints with our everyday perceptual experiences. This study also provides evidence that objects from aerial viewpoints are not mentally rotated in the 2D image plane prior to identity matching with a ground-view counterpart, extending previous findings which have shown that aerial images are not rotated during scene categorisation (Loschky et al., 2015).

6.2.2 Chapter 3: Contributions to the CI technique and pilot results

A new version of CIs was developed for this project, outlined in detail in Chapter 3 (Pilot 1-3). This CI technique facilitates simultaneous estimation of templates from 2D luminance and 3D binocular disparity cues from RDS-like noise textures.

Two previous studies have developed stereoscopic CIs from disparity cues in RDS (Gosselin, Bacon & Mamassian, 2004; Neri, Parker & Blakemore, 1999), and many studies have utilized CIs from luminance textures (see Chapter 1 for these details). The previously demonstrated stereoscopic CIs were based on RDSs, which are a sparse array of dots on a uniform background (e.g., black dots on a grey background) that can carry a stereogram (Julesz, 1971). The aims of the thesis included masking a natural image (for example, a hedge) with an RDS-like image to provide stereoscopic CIs. However, classical RDSs are unsuitable to add as a mask to another image, as the sparse dots do not provide the desired masking effect. Thus, Chapter 3 described the development of a novel version of stereoscopic CIs based on dense noise textures, which provided the desired masking effect. The ability to mask another image extends the potential applications of the technique. Furthermore, as this manipulation is based in a carrier noise texture, analysis of this carrier texture simultaneously affords luminance CIs. Simultaneous CIs from luminance and disparity could be useful for studies that seek to examine e.g., 1) stereoscopic judgements with both cue modalities, 2) cue combination of luminance and disparity, or 3) cue sampling in stereograms where a dense mask is required.

Furthermore, a post-hoc interest in stereoscopic PL motivated further analysis of two pilot experiments (Pilot 2 and 3) that used disparity targets with disparity noise. In these experiments, half the participants showed evidence of improvement throughout the experiment. That is, an adaptive staircase procedure increased the level of external noise (lowered the signal-to-noise ratios; SNR) throughout the experiment. This indicates that the participants became better able to detect the targets in noise. This evidence of PL was seen despite no provision of feedback, and with a smaller number of participation days, suggesting that experience is important for stereoscopic tasks. This supplementary analysis of the pilot experiments was instructive to the development of the study in Chapter 5, described below.

6.2.3 Chapter 4: Stereoscopic cues with experts and novices

The first study in Chapter 2 involved higher levels of visual perception with more general tasks of recognition in natural images. The study in Chapter 4, however, focused on more specific visual mechanisms related to depth perception in judgements of stereoscopic aerial landscape images.

This study implemented the novel approach to CIs described above, and provides novel evidence of expertise in remote sensing surveyors. Compared to novices, expert surveyors were five times better at sampling binocular disparity cues in both CI and sensitivity (*d*') measures (Figure 4.12). In the study, participants had access to binocular disparity cues that could define targets as stereoscopically tall or deep features. The task of discriminating tall vs. deep in a foveated region is a simple perceptual task which does not require much attentional resources or instrumental learning. Further, the procedure ensured that all participants had appropriate vergence control in the stereoscope, that is, stimulus images were seen with good dichoptic fusion. The simple nature of the task and the insurance of vergence control strongly suggest that these group differences reflect visual expertise, where the experts have a greater facility in sampling binocular disparity cues.

Chapter 4 also suggests that surveyors show evidence of adaptation for the lighting-from-above prior. The surveyors are accustomed to working with aerial landscape imagery where the sun provides light from below the line of sight. Application of a lighting-from-above prior to lit-from-below images would systematically provide the wrong interpretation in shape from shading. That is, interpretations of convex and concave shapes from directional lighting and shading cues would systematically be inverted and thus be counterproductive. The surveyors' visual systems might thus have adjusted to this environment by diminishing the typical lighting-from-above prior. A ranked pairing of expert surveyors against novices (Figure 4.11c) shows that novices consistently had the stronger priors for lighting-from-above, and some experts had a lighting-from-below prior, which would generally not be expected in novices (e.g., Pickard-Jones, d'Avossa & Sapir, 2020). These results suggest a striking shift in this prior. Adams, Graf and Ernst (2004) showed that the lighting-from-above prior can be malleable within smaller angular changes, but no study to date has found evidence of systematically diminished priors or even full inversions to lighting-from-below in a natural population. While these biases were very robust within-participants and were statistically significant between groups, they are limited by the smaller sample size of six participants per group, and should be interpreted with caution.

Chapter 4 also contained an online follow-up experiment which sought to estimate lighting direction priors from the honeycomb stimulus and aerial-view hedges and ditches. Expert surveyors tended to interpret the honeycomb with a lighting-from-above bias. Note, however, that this follow-up experiment used different participants than the main experiment in Chapter 4. The honeycomb is an image outside of the experts' domain, and the results from this image could suggest that surveyors interpret the world outside of their domain according to lighting-from-above. This would, of course, be an appropriate interpretation as the world is generally lit-from-above. The results of the main experiment in Chapter 4 do not show strong lighting-from-below biases in expert surveyors, but

rather generally weaker biases that sometimes switch to smaller lighting-from-below biases (Figure 4.11c). Taken together with the results from the honeycomb image in the follow-up experiment, this could suggest that surveyors are combating the lighting-from-above prior when classifying aerial images, but use the lighting-from-above prior elsewhere. Experienced surveyors might have developed a context-specific exception for aerial images, but they still interpret the real world with an assumption that light comes from above. The implications of this are discussed below.

Regarding observer strategies, most participants in both groups verbally reported using binocular disparity cues as a primary strategy, with a secondary strategy of diffuse lighting judgements of luminance cues ('dark-is-deep'). Despite this, experts were better able to use disparity cues while novices tended to use luminance cues (Figure 4.9). Furthermore, the use of lighting direction priors did not reflect a consciously available strategy in the participants. Thus, CIs revealed group differences that the verbal reports did not: that the experts were better able to use the disparity cues, and that the use of luminance cues varied across groups and individuals. This suggests that CIs can be a powerful technique for revealing visual strategies that observers are unaware of using, and the implications of this are elaborated below.

6.2.4 Chapter 5: Perceptual learning for disparity cues

Expertise (in general) can originate from two non-exclusive factors: 1) self-selection, where inherently talented individuals are drawn to specific domains where they excel, or 2) development, where abilities improve with experience. In general, people are likely more attracted to jobs or activities where they feel like they can perform well, thus self-selection can likely account for some expertise as a natural part of human individual differences. But a large part of expertise likely develops from experience. In the previous chapter, expert surveyors showed a large advantage for sampling binocular disparity cues in stereograms. In a supplementary analysis of earlier pilot data, it was shown that PL occurred, with half the participants improving their ability to detect a disparity target in stereograms. Following this, the study in Chapter 5 was designed to directly induce stereoscopic PL, to capture and characterise details regarding how the experts might have gained their advantage for disparity cues.

The results of Chapter 5 were split across two main outcome measures: 1) threshold estimates across sessions, and 2) comparison across CIs from the first and the last thirds of the sessions. For both measures, the results show converging evidence of stereoscopic PL. Thresholds improved, where SNRs reduced (more external noise was tolerated) as the experiment progressed. Disparity CIs increased in amplitudes but not spatial extent, showing that participants learned to

better sample disparity cues from a focused area. See below for further discussion of template shapes. Note, however, that the threshold and amplitude improvements did not correlate.

Previous literature commonly discusses PL as a mechanism for developing perceptual expertise (e.g., Lu & Dosher, 2022; Seitz, 2017; see also Chapter 1 for an elaboration). PL can be defined in laboratory settings where a controlled intervention is applied to improve visual processing of some stimuli. But PL also occurs in natural environments where experience improves stimulus processing in expertise domains, such as in remote sensing surveying or radiology. In such natural environments, the development of expertise does not occur with formal PL interventions, but rather through experience-dependent improvements from processing a diverse set of domain-specific stimuli.

Chapter 5 shows that PL can improve the processing of disparity cues in stereograms, which was the primary expertise factor in Chapter 4. Taken together, these results suggest that experience is important for processing stereograms, and that remote sensing surveyors develop expertise through long-term work with stereoscopic aerial images. Furthermore, the results of Chapter 5 raise the possibility of increasing the rate of expertise development through formal PL interventions. This topic serves as a suggestion for future directions, and is discussed further below.

6.2.5 Mechanisms of template shapes in Chapter 4 and 5

This section elaborates on the template structure in the CIs from Chapter 4 and 5. The disparity CIs in these two studies are discussed together, as they used similar targets in terms of their disparity profiles. Disparity templates often contained negative side-lobes, above and below the central positive peaks. These side-lobes show that cues were integrated from the surrounding area where, for example, a central target could appear ‘taller’ if an immediate surrounding point of reference was ‘deeper’. The templates thus reveal how the participants integrated relative disparities by using non-target areas as reference points when judging target relief.

The shape of the CI templates appeared centralised, with amplitudes that peaked at the foveated region of the images. The central template peaks decayed with horizontal eccentricity from the centre, as seen along the horizontal widths of the signal locations. For example, in Chapter 4, the fitted Gaussian functions had spreads that were smaller than the width of the target images (Figure 4.8; Table 3.2). The CIs in Chapter 5 similarly showed highly concentrated templates which clearly did not spread to the full width of the target signals (Figure 5.4). This suggests that relevant target signals were ignored in eccentric areas. Furthermore, learning in Chapter 5 was characterised by increased processing (template amplitudes) in the foveated area, without any increases in template spatial extent to sample a larger signal area. A similar effect is seen in Chapter 4 for disparity CIs, where

experts did not have wider templates than novices. But sampling a larger area could also have served to improve task performance as it would incorporate more relevant signal. For example, an ideal template is a perfect match to the target, and an ideal observer would utilise the full signal area (e.g., Gold et al., 2000). Overall, neither within-participant learning nor between-group expertise were associated with sampling disparity cues over more of the signal area.

This limitation in processing could be explained with visual mechanisms. As the visual system generally loses sensitivity and resolution with eccentricity (Baldwin, Meese & Baker, 2012; Strasburger, Rentschler & Jüttner, 2011), eccentric image regions were of less value than central ones, and this remained the case with PL. Attentional mechanisms could also provide a non-exclusive explanation. Perhaps observers had to limit their attentional window (Downing & Pinker, 1985; Posner, 1980) to a smaller location due to the high task demands of discriminating stereoscopic surfaces with strong external noise. Observers can reduce the size of their attentional window to increase processing efficiency, and efficiency decreases gradually with distance from the attentional focus (Castiello & Umiltà, 1990).

In conclusion on disparity CIs, the disparity templates were sub-optimal as they did not extend the full width of the signal area. The spatial extent of the templates remained unchanged with PL (Chapter 5), and experts did not prioritise larger spatial extents compared to novices (Chapter 4). Instead, learning and expertise were associated with higher amplitudes, indicating that experience enhances the ability to process disparity cues in a focused area.

Finally, the luminance CIs in Chapter 4 contained peaks with variable vertical offsets. These offsets were correlated with the sensitivity measure (*d*') for lighting-from-above (Figure 4.13). The offsets thus reveal that CIs captured the use of lighting direction priors, where participants who applied stronger priors for lighting-from-above tended to have asymmetric luminance peaks that were located above-centre. Conversely, participants with weaker priors tended to have symmetric peaks located closer to the centre. These results characterised group differences in lighting direction priors, elaborated above. Further, these differences in asymmetry and peak location show whether participants interpreted lighting cues as coming from a diffuse light source ('dark-is-deep') or a punctate light source ('sunlight strikes from a specific location').

6.3 Future directions

6.3.1 Future directions: Perceptual learning for remote sensing surveying

The OS employs a small number of new surveyors every year who need training and experience to develop expertise. These newly recruited surveyors could be ideal participants in studies of how PL

interventions might serve to speed up the development of expertise. Explicit training on photogrammetric tasks familiarises newly recruited surveyors with landscape features in aerial images. Developing generalised expertise in the workplace likely requires a very diverse set of experiences with stereoscopic aerial images over several months or years. Potentially, as a future addition to the natural development of expertise, PL interventions could help to improve the use of binocular disparity cues in stereograms by direct training. Having good stereopsis for stereograms is considered important by senior OS surveyors, and Chapter 4 clearly shows that surveyors are very adept at using binocular disparity cues in judgements of stereoscopic surfaces. Chapter 5 further shows that stereopsis in stereograms can be improved with a formal PL intervention. Direct training with stereoscopic PL could potentially afford a shortcut to developing expertise for stereopsis in stereograms. However, this proposition faces a few challenges. The main challenge is that the scientific literature does not provide a clear path to generalisable stereopsis training in stereo-normal participants (Levi, 2022; Lu, Lin & Dosher, 2016). An initial challenge is therefore to discover stereoscopic PL that – at least – transfers within stereograms. Learning must generalise within the domain of stereoscopic aerial images to be useful. This includes transfers across e.g., retinal locations, orientations, and the spatial frequency spectrum.

Following the potential validation of a generalisable stereoscopic PL intervention, this could be applied to see if the training brings benefit to inexperienced surveyors during surveying tasks. Inexperienced surveyors could be tested for their abilities with stereograms, and with a test/control group design, a stereoscopic PL intervention could be used to examine the efficacy of PL on improving the surveyors' abilities to process binocular disparity cues. Performance measures could be gathered multiple times in the first year of the surveyors' employment. This could afford measures of the natural development of expertise in the control group, and if any relative improvements are brought with the PL training in the test group. This proposed study is logistically challenging, as only a small number of new surveyors could be recruited as participants each year from a company such as OS. Likely, multiple years of participant recruitment and testing would be required, which could also afford multi-year longitudinal measures of expertise development. Although difficult to conduct, such a study could provide major insights into the real-world applications of PL and how organizations might want to utilise PL in their employee training regimens.

6.3.2 Future directions: Replicating the adaptation to the lighting-from-above prior

The lighting-from-above prior can be malleable within smaller angular shifts (Adams, Graf & Ernst, 2004), but the results of Chapter 4 suggests that the prior might be more malleable than

previously shown. Although limited by a smaller sample size, the results suggest that the prior can be diminished after years of engagement with lit-from-below images that are incongruent with the visual system's typical bias for lighting-from-above. In terms of future directions, another follow-up study could attempt to replicate this finding with a larger number of surveyors to produce a more reliable result. If replicated, this result would show that the lighting-from-above prior can be diminished and sometimes inverted as a result of adulthood adaptation. Furthermore, a new follow-up study could also seek to estimate if the surveyors interpret lighting direction cues differently when they classify aerial images compared to images outside of this domain (e.g., the honeycomb). This aim bears similarity to the online follow-up study that was conducted at the end of Chapter 4, but a new follow-up study would be more likely to succeed with a closer replication of the main study in Chapter 4. A new study on expert surveyors is thus warranted to explore and replicate the diminished lighting direction priors, and whether they are context-specific for aerial images. This context-specificity could be explored by comparison with e.g., the honeycomb image.

6.3.3 Future directions: Expertise for 3D rotations

Perspective rotations from ground to aerial viewpoints are an unusual form of rotation. More commonly, an observer might pass by an object horizontally, for example, if you look at a house as you walk past it. It might be easy to recognize the house across such a rotation, but a rotation into the aerial perspective would create an unusual viewpoint where novices might struggle to maintain object constancy and have difficulties with recognition and identification. Expert surveyors are familiar with this aerial viewpoint, and show evidence of improved processing in Chapter 2.

In terms of future directions, a study could explore expert-novice differences in object recognition with different viewpoints/rotations. These rotations could cover multiple common and uncommon viewpoints, including the above (aerial) perspective. Two competing hypotheses are proposed for open exploration: 1) The experts will be better at recognition in above perspectives, but not in other translated viewpoints such as different side-views. This would suggest that experience plays a specific role to developing the surveyors' expertise with aerial images, as recognition is specifically enhanced for this viewpoint but not others. 2) The experts show improvements across all viewpoint rotations. This outcome would be more surprising and more difficult to interpret, as it would be less clear what might be causing the experts' advantage. Potentially, the expert surveyors are inherently talented with mental rotations across all viewpoints, or their experience with processing one unusual viewpoint translates to other viewpoints. Testing with less experienced surveyors could define an 'intermediate' group which might help to control any effect of 'self-

selection', where individuals who are inherently talented with mental rotations decide to work with surveying tasks.

6.4 Implications

6.4.1 Implications: The skillset of expert surveyors

The work described in this thesis provides novel and complementary evidence that can be used to define expertise in remote sensing surveyors. Expertise is associated with strategic shifts in visual cue sampling, with a greater facility in early visual processing of disparity cues in stereograms (Chapter 4). Beyond early vision, expertise is associated with improved performance in: 1) categorisation of briefly presented aerial scenes, 2) consistency across ground and aerial viewpoints, and 3) object identity judgements across these viewpoints (Chapter 2). These results can inform training requirements for newly recruited surveyors as they develop expertise. At the OS, much emphasis in the surveyors' training is placed on familiarisation with classifying aerial images and the configural features specific to this viewpoint. The results of Chapter 2 reinforce the utility of using direct training to gain experience with the aerial viewpoint of landscapes, as a part of the expert skillset is increased accuracy with aerial images.

The results of Chapter 4 show advantageous processing in early vision, highlighting the importance of long-term PL to achieve visual expertise. Knowing that fully-developed visual expertise might take years to develop can constrain expectations on newly recruited surveyors, but also inspire new research on PL for increasing the rate of expertise development and thus workplace performance, elaborated above (Future directions).

We rely on remote sensing to gather geospatial data more than ever, and the demand is likely to increase in the future. Remote sensing surveyors have a unique skillset for analysing aerial images, and increased knowledge about this skillset can provide new suggestions for improving remote sensing analysis. The characteristic skillset of expert surveyors could be incorporated into training regimens for newly recruited surveyors, or in machine vision models, with the aim of improving performance. For example, Chapter 4 shows that excellent stereopsis in stereograms is a hallmark of expert performance in remote sensing. As the experts show a strong reliance on disparity cues, machine vision implementations could potentially benefit from training on images that provide depth coordinates in the landscape. While a discussion on the importance of 3D cues in machine vision is outside the scope of this thesis, remote sensing models could benefit from studies that benchmark expert human performance.

6.4.2 Implications: The utility of classification images

Chapter 4 shows that CIs can capture expert-novice group differences. This could have direct applications in other domains of expertise. For example, radiologists search for and classify spots in x-rays, and CIs could be used in experiments to investigate the radiologists' expectations on the characteristics of different types of targets. Such an experiment could investigate discrimination between benign and malignant spots, and characterise expert visual strategies. CIs might capture template differences across target types and groups in such a task.

CIs reveal visual strategies by showing the cues that observers use and prioritise. Chapter 4 found that observers can employ visual strategies that are not consciously available, as seen with the use of lighting direction priors that were not remarked upon despite direct questioning in the post-experiment debrief. This suggests that CIs can be a powerful technique for revealing visual strategies that observers are unaware of using. In many expertise domains, it is likely that experts have different visual strategies compared to novices, but the groups might not be able to articulate such differences. In Chapter 4, both groups reported that they primarily relied on binocular disparity cues to depth, with a secondary strategy of diffuse lighting judgements ('dark-is-deep'). CIs revealed group differences that the verbal reports did not: that the experts were better able to use the disparity cues, and that the use of luminance cues varied across groups and individuals. In other expertise domains, CIs could provide insights into visual strategies that might not be possible to capture with other methods or with verbal reports.

6.4.3 Implications: Lighting directions in the Ordnance Survey imagery

The OS mostly arranges aerial landscape images so that they face north-up. This convention is congruent with the traditional orientation of maps where the bottom of the page represents south, and the top of the page represents north. When surveyors encounter this convention in aerial landscape images, lighting comes from below as the UK is in the northern hemisphere. This lighting structure conflicts with the well-known prior for lighting-from-above (e.g., Ramachandran, 1988; see also Chapter 1 for further details). This means that the imagery is systematically incongruent with this natural bias in the visual system for recovering shape from shading. As Chapter 4 suggests, surveyors have adapted to this to diminish or switch the lighting-from-above prior in aerial landscape images. Surveyors have had to develop an exception to the natural bias, which we all encounter everywhere else in the real world. Surveyors also show evidence of having lighting-from-above priors in the honeycomb stimulus in the follow-up experiment in Chapter 4 (Figure 4.16), suggesting that surveyors' biases outside of the expertise domain are more typical.

An implication from this thesis for the OS to consider is how to arrange the imagery to best help their surveyors perform their tasks. If the imagery was oriented differently, so that the sunlight mostly comes from above, the surveyors would not have to combat the natural lighting-from-above prior. This reorientation would be easy to implement with rotations in the image presentation software. Currently, with lit-from-below imagery, the natural bias provides an incongruent interpretation which must be ignored or switched to not incur a cost to recovery of shape from shading. Instead, with lit-from-above imagery, the natural bias could serve to improve recovery of shape from shading. Such a change would likely aid photogrammetric performance in both beginner and experienced surveyors, as recovery of shape from shading in the imagery would be congruent with the real world.

6.5 Conclusions

The interpretation of aerial images is key for gathering geospatial information about the world. Remote sensing surveyors are experienced with interpreting aerial images that provide an unfamiliar view of the landscape. Despite considerable interest in remote sensing, only a small number of previous studies have explored the visual mechanisms associated with expertise in remote sensing of aerial images. This thesis presents three primary studies that aimed to extend our understanding of such expertise and the mechanisms that expert surveyors rely on to interpret aerial images.

The work presented in Chapter 2 established that experienced remote sensing surveyors have a superior ability to process both scenes and objects from the aerial viewpoint. In the first experiment of this study, compared to novices, experts were more accurate in categorising aerial-view scenes but not ground-view scenes. The second experiment shows that experts were better at recognising the identity of objects in a matching task involving the aerial viewpoint. This experiment also showed that mental rotation is not required for aerial images in this matching task. This study suggests that expert surveyors have an advantage for processing the featural configurations that are typical to aerial images. While aerial images provide an unusual view of landscapes that can be difficult to process, remote sensing surveyors have overcome some of this difficulty with experience.

Remote sensing surveyors at the OS view stereograms of aerial images, where binocular disparity cues significantly contribute to depth perception. To capture how experts and novices use different stereoscopic cues when judging 3D profiles in aerial features, a novel version of the CI technique was developed (Chapter 3). This method allows simultaneous estimation of CIs from binocular disparity and luminance cues. Chapter 4 tested experts and novices with this method, finding that the groups used stereoscopic cues in different ways when classifying aerial images. Compared to novices, experts had a greater facility to sample binocular disparity cues, likely due to

their experience with judgements in stereograms. This group difference was notably large, and revealed a mechanism that is strongly associated with visual expertise in the surveyors. Furthermore, the experts and novices interpreted lighting direction cues differently, where experts were less likely to adopt the conventional lighting-from-above prior. This prior was diminished or even inverted to lighting-from-below in the experts, likely due to the experts' experience with aerial images that are unconventionally lit from below the line of sight. Finally, both groups reportedly relied on disparity cues as a primary strategy, however, the experts were better able to use disparity cues while novices relied more on luminance cues. CIs thus revealed details about visual strategies that the verbal reports did not. This study in Chapter 4 reveal some of the impact of experience with aerial images. Experience with surveying aerial images is associated with a better ability to sample and prioritise relevant stereoscopic cues. Experience can also modify the interpretation of lighting direction cues in recovery of shape from shading.

Following the results regarding expertise for sampling disparity cues in stereograms, the study in Chapter 5 sought to characterise how novices might learn to improve this ability. With a stereoscopic PL intervention, participants generally improved their ability to detect disparity targets in noise. CIs revealed that learning was also associated with an improved ability to sample disparity cues. This improvement was concentrated to a focused area which did not expand with learning, and a similar effect was also seen in the previous study where experts and novices had comparable CI template extents. The results of this study help to characterise stereoscopic PL. Remote sensing surveyors develop expertise from long-term experience with stereoscopic aerial images, and this study provides a link between expertise and PL. This link can inspire suggestions for future research which could explore the efficacy of using PL interventions to augment the development of expertise.

In conclusion, this thesis contributes to furthering our understanding of the mechanisms associated with human vision in aerial images, and how these mechanisms can change with experience. The results paint a picture of how experience is associated with improvements and changes to vision, highlighting the key role of experience for interpreting stereoscopic aerial images. These novel and complementary results are useful for future research, and the results may also apply to provide an improved understanding of remote sensing surveying in the workplace.

References

Aberg, K. C., & Herzog, M. H. (2012). Different types of feedback change decision criterion and sensitivity differently in perceptual learning. *Journal of Vision*, *12*(3), 1–11. https://doi.org/10.1167/12.3.3

Abbey, C. K., & Eckstein, M. P. (2002). Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments. *Journal of Vision*, *2*(1), 66–78. https://doi.org/10.1167/2.1.5

Abbey, C. K., & Eckstein, M. P. (2006). Classification images for detection, contrast discrimination, and identification tasks with a common ideal observer. *Journal of Vision*, *6*(4), 335–355. https://doi.org/10.1167/6.4.4

Abbey, C. K., & Eckstein, M. P. (2014). Observer efficiency in free-localization tasks with correlated noise. *Frontiers in Psychology*, *5*(MAY), 1–13. https://doi.org/10.3389/fpsyg.2014.00345

Abbey, C. K., Lago, M. A., & Eckstein, M. P. (2021). Comparative observer effects in 2D and 3D localization tasks. *Journal of Medical Imaging*, *8*(04), 1–17. https://doi.org/10.1117/1.jmi.8.4.041206

Adams, W. J., & Elder, J. H. (2014). *Effects of Specular Highlights on Perceived Surface Convexity*. *10*(5). https://doi.org/10.1371/journal.pcbi.1003576

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

Adini, Y., Sagi, D., & Tsodyks, M. (2002). Context-enabled learning in the human visual system. *Nature*, *415*(6873), 790–793. https://doi.org/10.1038/415790a

Ahumada Jr, A. J. (1996). Perceptual classification images from Vernier acuity masked by noise. *Perception, 25(1_suppl),* 2-2. https://doi.org/10.1068/v96l0501

Ahumada, A. J., Jr., & Lovell, J. (1971). Stimulus features in signal detection. Journal of the Acoustical Society of America, 49, 1751–1756. https://doi.org/10.1121/1.1912577

Andrews, B., Aisenberg, D., D'Avossa, G., & Sapir, A. (2013). Cross-cultural effects on the assumed light source direction: Evidence from English and Hebrew readers. *Journal of Vision*, *13*(13), 1–7. https://doi.org/10.1167/13.13.2

Asch, S. E., & Witkin, H. A. (1948). Studies in space orientation: I. Perception of the upright with displaced visual fields. *Journal of Experimental Psychology*, *38*(3), 325–337. https://doi.org/10.1037/h0057855

Baldwin, A. S., Meese, T. S., & Baker, D. H. (2012). The attenuation surface for contrast sensitivity has the form of a witch's hat within the central visual field. *Journal of Vision*, *12*(11), 1–17. https://doi.org/10.1167/12.11.23

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research, 27*(6), 953–965. https://doi.org/10.1016/0042-6989(87)90011-3

Baker, D. H., & Meese, T. S. (2014). Measuring the spatial extent of texture pooling using reverse correlation. *Vision Research*, *97*, 52–58. https://doi.org/10.1016/j.visres.2014.02.004

Baker, D. H., Meese, T. S., Mansouri, B., & Hess, R. F. (2007). Binocular summation of contrast remains intact in strabismic amblyopia. *Investigative Ophthalmology and Visual Science*, *48*(11), 5332–5338. https://doi.org/10.1167/iovs.07-0194

Barlow, H. B. (1953). Summation and inhibition in the frog's retina. *The Journal of physiology*, *119*(1), 69-88. https://doi.org/10.1113%2Fjphysiol.1953.sp004829

Beard, B. L., & Ahumada, A. J., Jr. (1997). Relevant image features for Vernier acuity [Abstract]. Perception, 26, ECVP Abstract Supplement

Beard, B. L., & Ahumada Jr, A. J. (1998). Technique to extract relevant image features for visual tasks. *Human vision and electronic imaging III*, *3299*. 79-85. https://doi.org/10.1117/12.320099

Beard, B. L., & Ahumada, A. J. (1999). Detection in fixed and random noise in foveal and parafoveal vision explained by template learning. *JOSA A, 16*(3), 755-763. https://doi.org/10.1364/JOSAA.16.000755

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine, 68*(7), 569–579.

Berbaum, K., Bever, T., & Chung, C. S. (1983). Light source position in the perception of object shape. *Perception*, *12*(4), 411–416. https://doi.org/10.1068/p120411

Bertram, R., Helle, L., Kaakinen, J. K., & Svedström, E. (2013). The Effect of Expertise on Eye Movement Behaviour in Medical Image Perception. *PLoS ONE*, *8*(6). https://doi.org/10.1371/journal.pone.0066169

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing Depth-Rotated Objects: Evidence and Conditions for Three-Dimensional Viewpoint Invariance. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(6), 1162–1182. https://doi.org/10.1037/0096-1523.19.6.1162

Birch, E. E. (2013). Amblyopia and binocular vision. *Progress in Retinal and Eye Research*, *33*(1), 67–84. https://doi.org/10.1016/j.preteyeres.2012.11.001

Blake, R., & Fox, R. (1973). The psychophysical inquiry into binocular summation. *Perception & Psychophysics*, *14*(1), 161–185. https://doi.org/10.3758/BF03198631

Blakemore, C. (1970). The range and scope of binocular depth discrimination in man. *The Journal of Physiology*, *211*(3), 599-622. https://doi.org/10.1113/jphysiol.1970.sp009296

Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. The Journal of physiology, 203(1), 237-260. https://doi.org/10.1113/jphysiol.1969.sp008862

Borders, J. D., Dennis, B., Noesen, B., & Harel, A. (2020). Using fMRI to Predict Training Effectiveness in Visual Scene Analysis. In *Augmented Cognition. Human Cognition and Behavior: 14th International Conference, AC 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22* (pp. 14-26). Springer International Publishing.

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.

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, 1–29. https://doi.org/10.7717/peerj.9414

Castelhano, M. S., & Henderson, J. M. (2008). The Influence of Color on the Perception of Scene Gist. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 660–675. https://doi.org/10.1037/0096-1523.34.3.660

Castiello, U., & Umiltá, C. (1990). Size of the attentional focus and efficiency of performance. *Acta Psychologia*, *73*, 195–209. https://doi.org/10.1016/0001-6918(90)90022-8

Center, E. G., Gephart, A. M., Yang, P. L., & Beck, D. M. (2022). Typical viewpoints of objects are better detected than atypical ones. *Journal of Vision*, *22*(12), 1. https://doi.org/10.1167/jov.22.12.1

Champion, R. A., & Adams, W. J. (2007). Modification of the convexity prior but not the light-from-above prior in visual search with shaded objects. *Journal of Vision*, *7*(13), 1–10. https://doi.org/10.1167/7.13.10

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, *5*(9), 659–667. https://doi.org/10.1167/5.9.1

Chen, C. C., & Tyler, C. W. (2015). Shading beats binocular disparity in depth from luminance gradients: Evidence against a maximum likelihood principle for cue combination. *PLoS ONE*, *10*(8), 1–17. https://doi.org/10.1371/journal.pone.0132658

Coco-Martin, M. B., Piñero, D. P., Leal-Vega, L., Hernández-Rodríguez, C. J., Adiego, J., Molina-Martín, A., De Fez, D., & Arenillas, J. F. (2020). The Potential of Virtual Reality for Inducing

Neuroplasticity in Children with Amblyopia. *Journal of Ophthalmology*, *2020*. https://doi.org/10.1155/2020/7067846
Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564. https://doi.org/10.1111/j.0956-7976.2004.00719.x
de Boer, E., & Kuyper, P. (1968). Triggered correlation. IEEE Transactions on Biomedical Engineering, 15, 169–179.
Deveau, J., Ozer, D. J., & Seitz, A. R. (2014). Improved vision and on-field performance in baseball through perceptual learning. *Current Biology*, *24*(4), R146–R147. https://doi.org/10.1016/j.cub.2014.01.004
Ding, J., & Levi, D. M. (2011). Recovery of stereopsis through perceptual learning in human adults with abnormal binocular vision. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(37), 733–741. https://doi.org/10.1073/pnas.1105183108
Dobres, J., & Seitz, A. R. (2010). Perceptual learning of oriented gratings as revealed by classification images. *Journal of Vision*, *10*(13), 1–11. https://doi.org/10.1167/10.13.8
Doorschot, P. C. A., Kappers, A. M. L., & Koenderink, J. J. (2001). The combined influence of binocular disparity and shading on pictorial shape. *Perception and Psychophysics*, *63*(6), 1038–1047. https://doi.org/10.3758/BF03194522
Dorais, A., & Sagi, D. (1997). Contrast masking effects change with practice. *Vision Research*, *37*(13), 1725–1733. https://doi.org/10.1016/S0042-6989(96)00329-X
Dosher, B. A., & Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision research*, *39*(19), 3197-3221. https://doi.org/10.1016/S0042-6989(99)00059-0
Dosher, B., & Lu, Z. L. (2017). Visual Perceptual Learning and Models. *Annual Review of Vision Science*, *3*, 343–363. https://doi.org/10.1146/annurev-vision-102016-061249
Downing, C. J., & Pinker, S. (1985). The spatial structure of visual attention. In M. I. Posner & O. S. M. Marin (Eds.), Attention and performance XI: Mechanisms of attention (pp. 171–187). Hillsdale, NJ: Erlbaum.
Drew, T., Evans, K., Võ, M. L. H., Jacobson, F. L., & Wolfe, J. M. (2013). Informatics in radiology: What can you see in a single glance and how might this guide visual search in medical images? *Radiographics*, *33*(1), 263–274. https://doi.org/10.1148/rg.331125023
Dyde, R. T., Jenkin, M. R., & Harris, L. R. (2006). The subjective visual vertical and the perceptual upright. *Experimental Brain Research*, *173*(4), 612–622. https://doi.org/10.1007/s00221-006-0405-y
Eckstein, M. P., & Ahumada, A. J. (2002). Classification images: A tool to analyze visual strategies. *Journal of Vision, 2(1),* i. https://doi.org/10.1167/2.1.i
Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of threedimensional objects. *Vision Research*, *32*(12), 2385–2400. https://doi.org/10.1016/0042-6989(92)90102-O
Egusa, H. (1983). Effects of brightness, hue, and saturation on perceived depth between adjacent regions in the visual field. *Perception*, *12*(2), 167-175. https://doi.org/10.1068/p120167
Evans, K. K., Georgian-Smith, D., Tambouret, R., Birdwell, R. L., & Wolfe, J. M. (2013). The gist of the abnormal: Above-chance medical decision making in the blink of an eye. *Psychonomic Bulletin and Review*, *20*(6), 1170–1175. https://doi.org/10.3758/s13423-013-0459-3
Fahle, M., & Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research*, *35*(21), 3003-3013. https://doi.org/10.1016/0042-6989(95)00044-Z
Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*(1), 1–29. https://doi.org/10.1167/7.1.10
Fendick, M., & Westheimer, G. (1983). Effects of practice and the separation of test targets on foveal and peripheral stereoacuity. *Vision Science*, *23,* 145–150. https://doi.org/10.1016/0042-6989(83)90137-2

Fiorentini, A., & Berardi, N. (1981). Learning in grating waveform discrimination: Specificity for orientation and spatial frequency. *Vision research*, *21*(7), 1149-1158. https://doi.org/10.1016/0042-6989(81)90017-1

Foss, A. J. (2017). Use of video games for the treatment of amblyopia. Current Opinion in Ophthalmology, 28(3), 276–281, https://doi.org/10.1097/ICU.0000000000000358

Fox, S. E., & Faulkner-Jones, B. E. (2017). Eye-tracking in the study of visual expertise: Methodology and approaches in medicine. *Frontline Learning Research*, *5*(3 Special Issue), 43–54. https://doi.org/10.14786/flr.v5i3.258

Frisby, J. P., & Clatworthy, J. L. (1975). Learning to see complex random-dot stereograms. *Perception*, *4*(2), 173-178. https://doi.org/10.1068/p040173

Furtak, M., Mudrik, L., & Bola, M. (2022). The forest, the trees, or both? Hierarchy and interactions between gist and object processing during perception of real-world scenes. *Cognition*, *221*(November 2021). https://doi.org/10.1016/j.cognition.2021.104983

Garnham, L., & Sloper, J. (2006). Effect of age on adult stereoacuity as measured by different types of stereotest. *British Journal of Ophthalmology*, *90*(1), 91–95. https://doi.org/10.1136/bjo.2005.077719

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise Differences in the Comprehension of Visualizations: A Meta-Analysis of Eye-Tracking Research in Professional Domains. *Educational Psychology Review*, *23*(4), 523–552. https://doi.org/10.1007/s10648-011-9174-7

Georgeson, M. A., Yates, T. A., & Schofield, A. J. (2008). Discriminating depth in corrugated stereo surfaces: Facilitation by a pedestal is explained by removal of uncertainty. *Vision Research*, *48*(21), 2321–2328. https://doi.org/10.1016/j.visres.2008.07.009

Georgeson, M. A., Yates, T. A., & Schofield, A. J. (2009). Depth propagation and surface construction in 3-D vision. *Vision Research*, *49*(1), 84–95. https://doi.org/10.1016/j.visres.2008.09.030

Gibson, E. J. (1963). Perceptual learning. *Annual review of psychology*, *14*(1), 29-56. https://doi.org/10.1146/annurev.ps.14.020163.000333

Gibson, J. (1950) The Perception of the Visual World (Cambridge, MA: Houghton Mifflin, The Riverside Press)

Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology*, *10*(11), 663–666. https://doi.org/10.1016/S0960-9822(00)00523-6

Gold, J. M., Sekuler, A. B., & Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, *28*(2), 167–207. https://doi.org/10.1016/j.cogsci.2003.10.005

Godinez, A., Martín-González, S., Ibarrondo, O., & Levi, D. M. (2021). Scaffolding depth cues and perceptual learning in VR to train stereovision: a proof of concept pilot study. *Scientific Reports*, *11*(1), 1–16. https://doi.org/10.1038/s41598-021-89064-z

Gold, J. M., Sekuler, A. B., & Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, *28*(2), 167–207. https://doi.org/10.1016/j.cogsci.2003.10.005

Goldstone, R. L. (1998). Perceptual learning. *Annual review of psychology*, *49*(1), 585-612. https://doi.org/10.1146/annurev.psych.49.1.585

Gosselin, F., Bacon, B. A., & Mamassian, P. (2004). Internal surface representations approximated by reverse correlation. *Vision Research*, *44*(21), 2515–2520. https://doi.org/10.1016/j.visres.2004.05.016

Gosselin, F., & Schyns, P. G. (2003). Superstitious perceptions reveal properties of internal representations. *Psychological Science*, *14*(5), 505–509. https://doi.org/10.1111/1467-9280.03452

Gosselin, F., & Schyns, P. G. (2004). No troubles with bubbles: A reply to Murray and Gold. *Vision Research*, *44*(5), 471–477. https://doi.org/10.1016/j.visres.2003.10.007

Green, C. S., & Bavelier, D. (2015). Action video game training for cognitive enhancement. *Current Opinion in Behavioral Sciences*, *4*, 103–108. https://doi.org/10.1016/j.cobeha.2015.04.012

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1, pp. 1969-2012). New York: Wiley.
Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137–176. https://doi.org/10.1016/j.cogpsych.2008.06.001
Harel, A. (2016). What is special about expertise? Visual expertise reveals the interactive nature of real-world object recognition. *Neuropsychologia*, *83*, 88–99. https://doi.org/10.1016/j.neuropsychologia.2015.06.004
Hartle, B., Irving, E. L., Allison, R. S., Glaholt, M. G., & Wilcox, L. M. (2022). Shape judgments in natural scenes: Convexity biases versus stereopsis. *Journal of Vision 22(6),* 1–13. https://doi.org/10.1167/jov.22.8.6
Hershberger, W. (1970). Attached-shadow orientation perceived as depth by chickens reared in an environment illuminated from below. *Journal of Comparative and Physiological Psychology, 73*(3), 407–411. https://doi.org/10.1037/h0030223
Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research*, *37*(15), 2133–2141. https://doi.org/10.1016/S0042-6989(97)00043-6
Hillis, J. H., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science*, *298*(5598), 1627–1630. https://doi.org/10.1126/science.1075396
Hillis, 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*(12), 967–992. https://doi.org/10.1167/4.12.1
Howard, I. P. (2002). *Seeing in depth, Vol. 1: Basic mechanisms*. University of Toronto Press.
Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth, Vol. 2: Depth perception*. University of Toronto Press.
Howarth, P. A. (2011). The geometric horopter. *Vision Research*, *51*(4), 397–399. https://doi.org/10.1016/j.visres.2010.12.018
Johnston, E. B., Cumming, B. G., & Parker, A. J. (1993). Integration of depth modules: Stereopsis and texture. *Vision Research*, *33*(5–6), 813–826. https://doi.org/10.1016/0042-6989(93)90200-G
Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*(26), 3286–3297. https://doi.org/10.1016/j.visres.2007.09.013
Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature,* 365, 250–252 https://doi.org/10.1038/365250a0
Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. M., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science*, *265*(5172), 679–682. https://doi.org/10.1126/science.8036518
Kasarskis, P., Stehwien, J., Hickox, J., Aretz, A., & Wickens, C. (2001). Comparison of expert and novice scan behaviors during VFR flight. *Proceedings of the 11th international symposium on aviation psychology* (6).
Kienzle, W., Franz, M. O., Schölkopf, B., & Wichmann, F. A. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of vision*, *9*(5), 1-15. https://doi.org/10.1167/9.5.7
Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*(24), 2539–2558. https://doi.org/10.1016/S0042-6989(03)00458-9
Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., te Pas, S. F., & Pont, S. C. (2003). Illumination direction from texture shading. *Journal of the Optical Society of America A*, *20*(6), 987. https://doi.org/10.1364/josaa.20.000987
Kontsevich, L. L., & Tyler, C. W. (2004). What makes Mona Lisa smile? *Vision Research*, *44*(13), 1493–1498. https://doi.org/10.1016/j.visres.2003.11.027

Krupinski, E. A. (2010). Current perspectives in medical image perception. *Attention, Perception, & Psychophysics*, *72*(5), 1205-1217. https://doi.org/10.3758/APP.72.5.1205

Krupinski, E. A., Tillack, A. A., Richter, L., Henderson, J. T., Bhattacharyya, A. K., Scott, K. M., Graham, A. R., Descour, M. R., Davis, J. R., & Weinstein, R. S. (2006). Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience. *Human Pathology*, *37*(12), 1543–1556. https://doi.org/10.1016/j.humpath.2006.08.024

Kuai, S. G., Levi, D., & Kourtzi, Z. (2013). Learning optimizes decision templates in the human visual cortex. *Current Biology*, *23*(18), 1799–1804. https://doi.org/10.1016/j.cub.2013.07.052

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of neurophysiology*, *16*(1), 37-68. https://doi.org/10.1152/jn.1953.16.1.37

Kundel, H. L., & Nodine, C. F. (1975). Interpreting chest radiographs without visual search. *Radiology*, *116*(3), 527-532. https://doi.org/10.1148/116.3.527

Kurki, I., & Eckstein, M. P. (2014). Template changes with perceptual learning are driven by feature informativeness. *Journal of Vision*, *14*(11), 1–18. https://doi.org/10.1167/14.11.6

Langer, M. S., & Bülthoff, H. H. (2000). Depth discrimination from shading under diffuse lighting. *Perception*, *29*(6), 649–660. https://doi.org/10.1068/p3060

Langer, M. S., & Bülthoff, H. H. (2001). A prior for global convexity in local shape-from-shading. *Perception*, *30*(4), 403-410. https://doi.org/10.1068/p3178

Langer, M. S., & Zucker, S. W. (1994). Shape-from-shading on a cloudy day. *Journal of the Optical Society of America A*, *11*(2), 467-478. https://doi.org/10.1364/josaa.11.000467

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, *35*(3), 389–412. https://doi.org/10.1016/0042-6989(94)00176-M

Lansdale, M., Underwood, G., & Davies, C. (2010). Something overlooked? How experts in change detection use visual saliency. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *24*(2), 213-225. https://doi.org/10.1002/acp.1552

Lawson, R. (1999). Achieving visual object constancy across plane rotation and depth rotation. *Acta Psychologica*, *102*(2–3), 221–245. https://doi.org/10.1016/s0001-6918(98)00052-3

Lee, Y. W., & Schetzen, M. (1965). Measurement of the Wiener kernels of a non-linear system by cross- correlation. International Journal of Control, 2, 237–254. https://doi.org/10.1080/00207176508905543

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Josa*, *70*(12), 1458-1471. https://doi.org/10.1364/JOSA.70.001458

Levi, D. M. (2020). Rethinking amblyopia 2020. *Vision Research*, *176*(August), 118–129. https://doi.org/10.1016/j.visres.2020.07.014

Levi, D. M. (2022). Learning to see in depth. *Vision Research*, *200*(April), 108082. 1–13. https://doi.org/10.1016/j.visres.2022.108082

Levi, D. M. (2023). Applications and implications for extended reality to improve binocular vision and stereopsis. Journal of Vision, 23(1):14, 1–14. https://doi.org/10.1167/jov.23.1.14

Levi, D. M., Knill, D. C., & Bavelier, D. (2015). Stereopsis and amblyopia: A mini-review. *Vision Research*, *114*(2015), 17–30. https://doi.org/10.1016/j.visres.2015.01.002

Levi, D. M., & Li, R. W. (2009). Perceptual learning as a potential treatment for amblyopia: A mini-review. *Vision Research*, *49*(21), 2535–2549. https://doi.org/10.1016/j.visres.2009.02.010

Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467–477. https://doi.org/10.1121/1.1912375

Li, R. W., Levi, D. M., & Klein, S. A. (2004). Perceptual learning improves efficiency by re-tuning the decision "template" for position discrimination. *Nature Neuroscience*, *7*(2), 178–183. https://doi.org/10.1038/nn1183

Li, R. W., Tran, T. T., Craven, A. P., Leung, T. W., Chat, S. W., & Levi, D. M. (2016). Sharpening coarse-to-fine stereo vision by perceptual learning: Asymmetric transfer across the spatial frequency spectrum. *Royal Society Open Science*, *3*(1). https://doi.org/10.1098/rsos.150523

Li, Y., & Pizlo, Z. (2011). Depth cues versus the simplicity principle in 3D shape perception. *Topics in cognitive science*, *3*(4), 667-685. https://doi.org/10.1111/j.1756-8765.2011.01155.x

Li, Y., Pizlo, Z., & Steinman, R. M. (2009). A computational model that recovers the 3D shape of an object from a single 2D retinal representation. *Vision research*, *49*(9), 979-991. https://doi.org/10.1016/j.visres.2008.05.013

Linton, P. (2020). Does vision extract absolute distance from vergence? *Attention, Perception, and Psychophysics*, *82*(6), 3176–3195. https://doi.org/10.3758/s13414-020-02006-1

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

Liu, J., Dosher, B., & Lu, Z. L. (2014). Modeling trial by trial and block feedback in perceptual learning. *Vision Research*, *99*, 46–56. https://doi.org/10.1016/j.visres.2014.01.001

Liu, J., Lu, Z. L., & Dosher, B. A. (2010). Augmented Hebbian reweighting: Interactions between feedback and training accuracy in perceptual learning. *Journal of Vision*, *10*(10), 1–14. https://doi.org/10.1167/10.10.29

Liu, J., Lu, Z. L., & Dosher, B. A. (2012). Mixed training at high and low accuracy levels leads to perceptual learning without feedback. *Vision Research*, *61*, 15–24. https://doi.org/10.1016/j.visres.2011.12.002

Lloyd, R., Hodgson, M. E., & Stokes, A. (2002). Visual categorization with aerial photographs. *Annals of the Association of American Geographers*, *92*(2), 241–266. https://doi.org/10.1111/1467-8306.00289

Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made before basic-level distinctions in scene gist processing. *Visual Cognition*, *18*(4), 513-536. https://doi.org/10.1080/13506280902937606

Loschky, L. C., Ringer, R. V., Ellis, K., & Hansen, B. C. (2015). Comparing rapid scene categorization of aerial and terrestrial views: A new perspective on scene gist. *Journal of Vision*, *15*(6), 1–29. https://doi.org/10.1167/15.6.11

Lovell, P. G., Bloj, M., & Harris, J. M. (2012). Optimal integration of shading and binocular disparity for depth perception. *Journal of Vision*, *12*(1), 1–18. https://doi.org/10.1167/12.1.1

Lu, Z. L., & Dosher, B. A. (2022). Current directions in visual perceptual learning. *Nature Reviews Psychology*, *1*(11), 654–668. https://doi.org/10.1038/s44159-022-00107-2

Lu, Z. L., Lin, Z., & Dosher, B. A. (2016). Translating perceptual learning from the laboratory to applications. *Trends in Cognitive Sciences*, *20*(8), 561-563. https://doi.org/10.1016/j.tics.2016.05.007

Lu, Z. L., Hua, T., Huang, C. B., Zhou, Y., & Dosher, B. A. (2011). Visual perceptual learning. *Neurobiology of Learning and Memory*, *95*(2), 145–151. https://doi.org/10.1016/j.nlm.2010.09.010

MacCurdy E (Ed.), (1938). *The Notebooks of Leonardo da Vinci* (London: Jonathan Cape), p. 332.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press. https://doi.org/10.4324/9781410611147

Malcolm, G. L., Groen, I. I., & Baker, C. I. (2016). Making sense of real-world scenes. *Trends in cognitive sciences*, *20*(11), 843-856. https://doi.org/10.1016/j.tics.2016.09.003

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition*, *81*(1), B1. https://doi.org/10.1016/S0010-0277(01)00116-0

Marmarelis, V. (2012). *Analysis of physiological systems: The white-noise approach*. Springer Science & Business Media.

Marmarelis, P. Z., & Marmarelis, V. Z. (1978). Analysis of physiological systems: The white-noise approach. New York: Plenum Press.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *207*(1167), 187-217. https://doi.org/10.1098/rspb.1980.0020
Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *200*(1140), 269-294. https://doi.org/10.1098/rspb.1978.0020
McKee, S. P., Levi, D. M., & Movshon, J. A. (2003). The pattern of visual deficits in amblyopia. *Journal of Vision*, *3*(5), 380–405. https://doi.org/10.1167/3.5.5
Meese, T. S., Georgeson, M. A., & Baker, D. H. (2006). Binocular contrast vision at and above threshold. *Journal of Vision*, *6*(11), 1224–1243. https://doi.org/10.1167/6.11.7
Meese, T. S., & Holmes, D. J. (2004). Performance data indicate summation for pictorial depth-cues in slanted surfaces. *Spatial Vision*, *17*(1–2), 127–151. https://doi.org/10.1163/156856804322778305
Mittelstaedt, H. A new solution to the problem of the subjective vertical. *Naturwissenschaften* 70, 272–281 (1983). https://doi.org/10.1007/BF00404833
Murray, R. F. (2011). Classification images: A review. *Journal of vision, 11*(5), 1-25. https://doi.org/10.1167/11.5.2.
Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2005). Classification images predict absolute efficiency. *Journal of Vision*, *5*(2), 139–149. https://doi.org/10.1167/5.2.5
Neri, P., & Levi, D. M. (2006). Receptive versus perceptive fields from the reverse-correlation viewpoint. *Vision Research*, *46*(16), 2465–2474. https://doi.org/10.1016/j.visres.2006.02.002
Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature*, *401*(6754), 695–698. https://doi.org/10.1038/44409
Newell, F. N., Ernst, M. O., Tjan, B. S., & Bülthoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological science*, *12*(1), 37-42. https://doi.org/10.1111/1467-9280.00307
Nodine, C. F., & Krupinski, E. A. (1998). Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO. *Academic Radiology*, *5*(9), 603–612. https://doi.org/10.1016/S1076-6332(98)80295-X
Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3), 145–175. https://doi.org/10.1023/A:1011139631724
Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research* (Vol. 155, pp. 23–36). https://www.sciencedirect.com/science/article/pii/S0079612306550022
O'Shea, R. P., Blackburn, S. G., & Ono, H. (1994). Contrast as a depth cue. *Vision Research*, *34*(12), 1595–1604. https://doi.org/10.1016/0042-6989(94)90116-3
O'toole, A. J., & Kersten, D. J. (1992). Learning to see random-dot stereograms. *Perception*, *21*(2), 227-243. https://doi.org/10.1068/p210227
Palmer, S., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley, Attention and Performance IX (pp. 135-151). Hillsdale, New Jersey: Lawrence Erlbaum.
Pannasch, S., Helmert, J. R., Hansen, B. C., Larson, A. M., & Loschky, L. C. (2014). Commonalities and differences in eye movement behavior when exploring aerial and terrestrial scenes. In *Cartography from Pole to Pole: Selected Contributions to the XXVIth International Conference of the ICA, Dresden 2013* (pp. 421-430). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-32618-9_30
Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y
Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), Vision: Coding and efficiency (pp. 3–24). Cambridge, MA: Cambridge University Press.

Perrett, D. I., & Harries, M. H. (1988). Characteristic views and the visual inspection of simple faceted and smooth objects:'tetrahedra and potatoes'. *Perception*, *17*(6), 703-720. https://doi.org/10.1068/p170703

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London - Biological Sciences*, *223*(1232), 293–317. https://doi.org/10.1098/rspb.1985.0003

Pinter, R. B., & Nabet, B. (2018). *Nonlinear vision: Determination of neural receptive fields, function, and networks*. CRC Press.

Pizlo, Z. (2010). *3D shape: Its unique place in visual perception*. Mit Press.

Pont, S. C., van Doorn, A. J., & Koenderink, J. J. (2017). Estimating the illumination direction from three- dimensional texture of Brownian surfaces. *I-Perception*, *8*(2), 1–18. https://doi.org/10.1177/2041669517701947

Posner, M. I. (1980). Orienting or attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Potetz, B., & Lee, T. S. (2003). Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes. *JOSA A*, *20*(7), 1292-1303. https://doi.org/10.1364/JOSAA.20.001292

Portela-Camino, J. A., Martín-González, S., Ruiz-Alcocer, J., Illarramendi-Mendicute, I., & Garrido-Mercado, R. (2018). A random dot computer video game improves stereopsis. *Optometry and Vision Science*, *95*(6), 523-535. https://doi.org/10.1097/OPX.0000000000001222

Prather, S. C., & Sathian, K. (2002). *Mental rotation of tactile stimuli*. *14*, 91–98. https://doi.org/10.1016/S0926-6410(02)00063-0

Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2006). Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis. *Journal of Vision*, *6*(4). https://doi.org/10.1167/6.4.7

Ramachandran, V. S. (1976). Learning-like phenomena in stereopsis. *Nature*, *262*(5567), 382-384. https://doi.org/10.1038/262382a0

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature*, *331*(6152), 163-166. https://doi.org/10.1016/0002-9394(88)90349-2

Ramachandran, V. S., & Braddick, O. (1973). Orientation-specific learning in stereopsis. *Perception*, *2*(3), 371-376. https://doi.org/10.1068/p020371

Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science*, *20*(1), 6–10. https://doi.org/10.1111/j.1467-9280.2008.02243.x

Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. *Psychological science*, *12*(1), 48-55. https://doi.org/10.1111/1467-9280.00309

Reingold, E. M., & Sheridan, H. (2012). Eye movements and visual expertise in chess and medicine. *The Oxford Handbook of Eye Movements*, *May 2014*, 524-550. https://doi.org/10.1093/oxfordhb/9780199539789.013.0029

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53*(1), 245–277. https://doi.org/10.1146/annurev.psych.53.100901.135125

Rhodes, R. E., Cowley, H. P., Huang, J. G., Gray-Roncal, W., Wester, B. A., & Drenkow, N. (2021). Benchmarking Human Performance for Visual Search of Aerial Images. *Frontiers in Psychology*, *12*(December). https://doi.org/10.3389/fpsyg.2021.733021

Ringach, D., & Shapley, R. (2004). Reverse correlation in neurophysiology. *Cognitive Science*, *28*(2), 147–166. https://doi.org/10.1016/j.cogsci.2003.11.003

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

Rodán, A., Candela Marroquín, E., & Jara García, L. C. (2022). An updated review about perceptual learning as a treatment for amblyopia. *Journal of Optometry*, *15*(1), 3–34. https://doi.org/10.1016/j.optom.2020.08.002

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, *8*(3), 382-439. https://doi.org/10.1016/0010-0285(76)90013-X

Rousselet, G. A., Joubert, O. R., & Fabre-Thorpe, M. (2005). How long to get to the "gist" of real-world natural scenes? *Visual Cognition, 12*(6), 852–877. https://doi.org/10.1080/13506280444000553

Sachs, M.B., Jacob Nachmias, J., and John G. Robson, J.G., (1971) Spatial-Frequency Channels in Human Vision" JOSA. 61, 1176-1186

Sagi, D. (2011). Perceptual learning in Vision Research. *Vision Research*, *51*(13), 1552–1566. https://doi.org/10.1016/j.visres.2010.10.019

Samonds, J. M., Potetz, B. R., & Lee, T. S. (2012). Relative luminance and binocular disparity preferences are correlated in macaque primary visual cortex, matching natural scene statistics. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(16), 6313–6318. https://doi.org/10.1073/pnas.1200125109

Sasaki, Y., Nanez, J. E., & Watanabe, T. (2010). Advances in visual perceptual learning and plasticity. *Nature Reviews Neuroscience*, *11*(1), 53-60. https://doi.org/10.1038/nrn2737

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

Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise differences in attentional strategies related to pilot decision making. *Human Factors*, *50*(6), 864–878. https://doi.org/10.1518/001872008X374974

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*(5), 402–409. https://doi.org/10.1111/1467-9280.00472

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological science*, *5*(4), 195-200. https://doi.org/10.1111/j.1467-9280.1994.tb00500.x

Seitz, A. R. (2017). Perceptual learning. *Current Biology*, *27*(13), R631–R636. https://doi.org/10.1016/j.cub.2017.05.053

Seitz, A. R. (2020). Perceptual Expertise: How Is It Achieved? *Current Biology*, *30*(15), R875–R878. https://doi.org/10.1016/j.cub.2020.06.013

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(15), 6424–6429. https://doi.org/10.1073/pnas.0700622104

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701-703. https://doi.org/10.1126/science.171.3972.701

Sheskin, D. J. (2020). *Handbook of parametric and nonparametric statistical procedures*. crc Press.

Shibata, K., Yamagishi, N., Ishii, S., & Kawato, M. (2009). Boosting perceptual learning by fake feedback. *Vision Research*, *49*(21), 2574–2585. https://doi.org/10.1016/j.visres.2009.06.009

Šikl, R., Svatoňová, H., Děchtěrenko, F., & Urbánek, T. (2019). Visual recognition memory for scenes in aerial photographs: Exploring the role of expertise. *Acta Psychologica*, *197*(April), 23–31. https://doi.org/10.1016/j.actpsy.2019.04.019

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in cognitive sciences*, *1*(7), 261-267. https://doi.org/10.1016/S1364-6613(97)01080-2

Skog, E., Qian, C. S., Parmar, A., & Schofield, A. J. (2023). What surprises the Mona Lisa ? The relative importance of the eyes and eyebrows for detecting surprise in briefly presented face stimuli.

Vision Research, *211*(September 2022), 108275. https://doi.org/10.1016/j.visres.2023.108275
Sowden, P., Davies, I., Rose, D., & Kaye, M. (1996). Perceptual learning of stereoacuity. *Perception*, *25*(9), 1043-1052. https://doi.org/10.1068/p251043
Stickgold, R., James, L., & Hobson, J. A. (2000). Visual discrimination learning requires sleep after training. *Nature Neuroscience*, *3*(12), 1237–1238. https://doi.org/10.1038/81756
Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5), 1–82. https://doi.org/10.1167/11.5.13
Sun, J., & Perona, P. (1998). Where is the sun?. *Nature neuroscience*, *1*(3), 183-184. https://doi.org/10.1038/630
Sun, P., & Schofield, A. J. (2012). Two operational modes in the perception of shape from shading revealed by the effects of edge information in slant settings. *Journal of Vision*, *12*(1), 1–21. https://doi.org/10.1167/12.1.12
Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*(3), 457–482. https://doi.org/10.1016/0010-0285(91)90016-H
Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience*, *1*(4), 275–277. https://doi.org/10.1038/1089
Tavassoli, A., Van der Linde, I., Bovik, A. C., & Cormack, L. K. (2007). An efficient technique for revealing visual search strategies with classification images. *Perception & Psychophysics*, *69*(1), 103-112. http://dx.doi.org/10.3758/BF03194457
Vedamurthy, I., Knill, D. C., Huang, S. J., Yung, A., Ding, J., Kwon, O. S., Bavelier, D., & Levi, D. M. (2016). Recovering stereo vision by squashing virtual bugs in a virtual reality environment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1697). https://doi.org/10.1098/rstb.2015.0264
Volterra, V. (1930). Theory of functionals and of integral and integrodifferential equations. London: Blakie.
Vuong, Q. C., Domini, F., & Caudek, C. (2006). Disparity and shading cues cooperate for surface interpolation. *Perception*, *35*(2), 145–155. https://doi.org/10.1068/p5315
Watson, A. B., & Rosenholtz, R. (1997). A Rorschach test for visual classification strategies. *Investigative Ophthalmology and Visual Science*, *38*, S1.
Westheimer, G. (2001). Is peripheral visual acuity susceptible to perceptual learning in the adult? *Vision Research*, *41*(1), 47–52. https://doi.org/10.1016/S0042-6989(00)00245-5
Wiener, N. (1958). Nonlinear problems in random theory. New York: John Wiley & Sons.
Wolfe, J. M., Evans, K. K., Drew, T., Aizenman, A., & Josephs, E. (2016). How do radiologists use the human search engine? *Radiation Protection Dosimetry*, *169*(1), 24–31. https://doi.org/10.1093/rpd/ncv501
Wright, M., & Ledgeway, T. (2004). Interaction between luminance gratings and disparity gratings. Spatial Vision, 17, 51–74. https://doi.org/10.1163/156856804322778260
Xi, J., Jia, W. L., Feng, L. X., Lu, Z. L., & Huang, C. B. (2014). Perceptual learning improves stereoacuity in amblyopia. *Investigative Ophthalmology and Visual Science*, *55*(4), 2384–2391. https://doi.org/10.1167/iovs.13-12627
Ziv, G. (2016). Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation. *International Journal of Aviation Psychology*, *26*(3–4), 75–104. https://doi.org/10.1080/10508414.2017.1313096

Appendix A: Disparity CIs, subdivided by different image manipulation conditions.

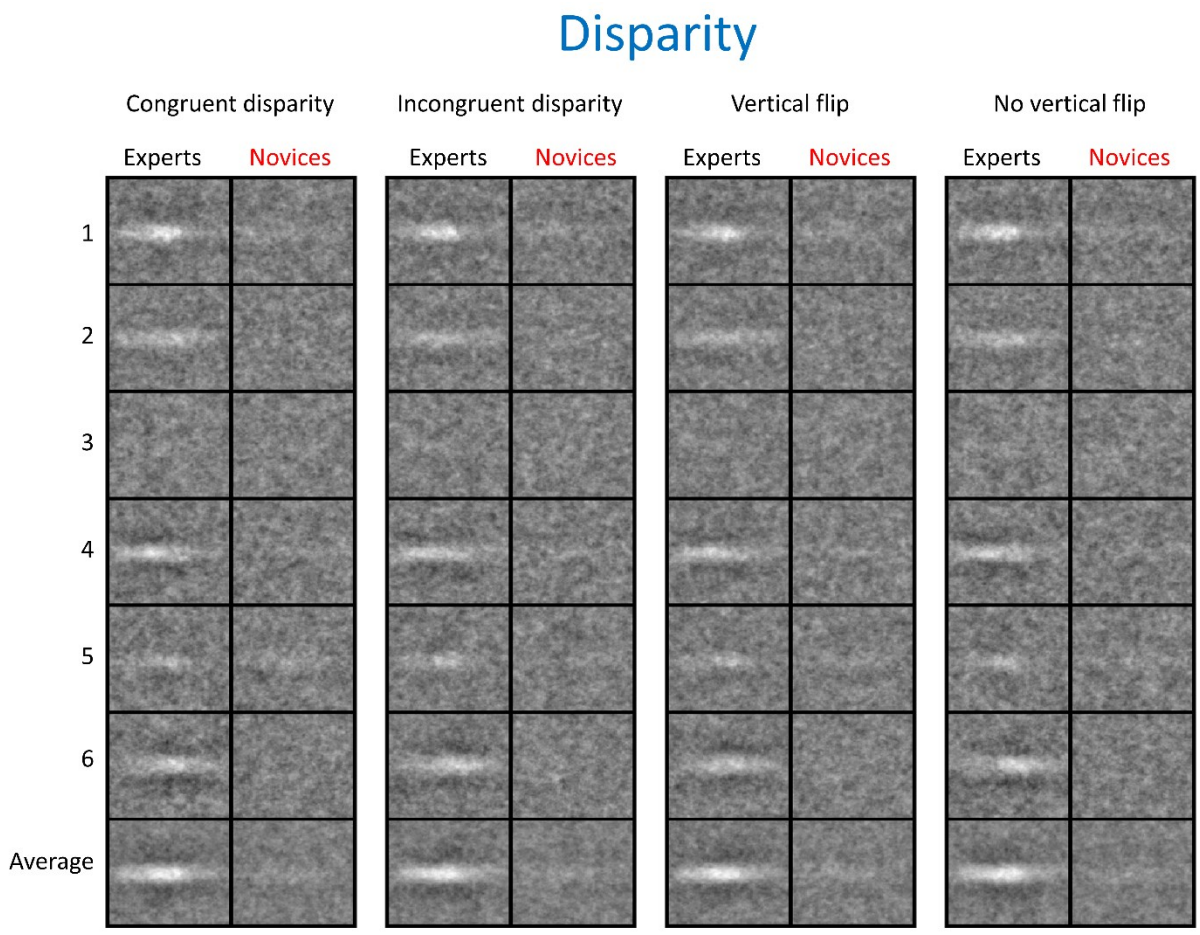

Figure A1: Disparity CIs from different conditions of image manipulations. Participants applied similar templates in all conditions. Numbers on the left-hand side represent individual participants at each row. 'Congruent disparity' represents all trials where disparity was congruent. 'Incongruent disparity' shows all trials where disparity was incongruent. 'Vertical flip' shows all trials where the light source originated from above the line of sight. 'No vertical flip' shows all trials where the light came from below.

Appendix B: Luminance CIs, subdivided by different image manipulation conditions similar to Appendix A.

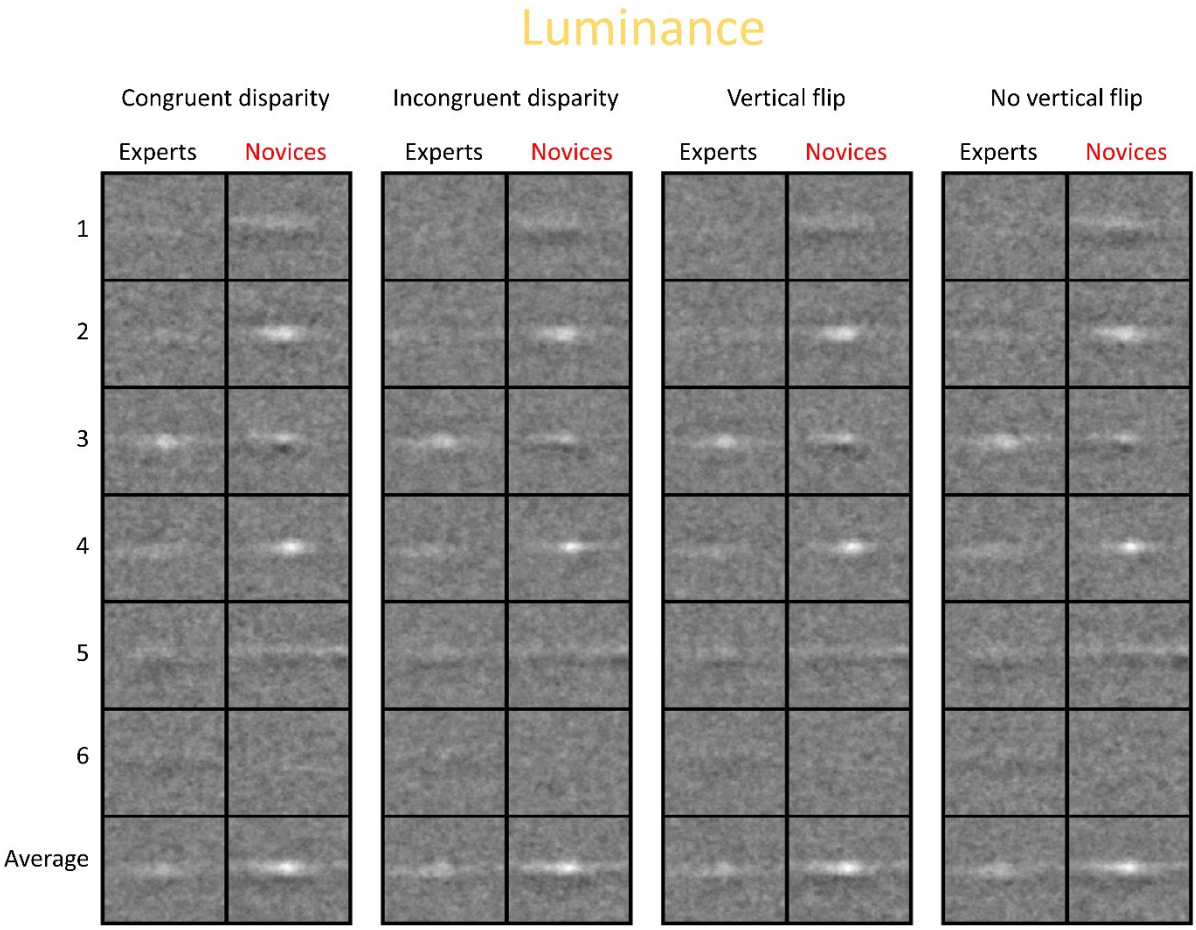

Figure B1: Luminance CIs from different image manipulations conditions similar to Appendix A. Participants applied similar templates in all conditions.

Appendix C: Fits to vertical cross-section of the disparity CIs for individual participants from Chapter 4

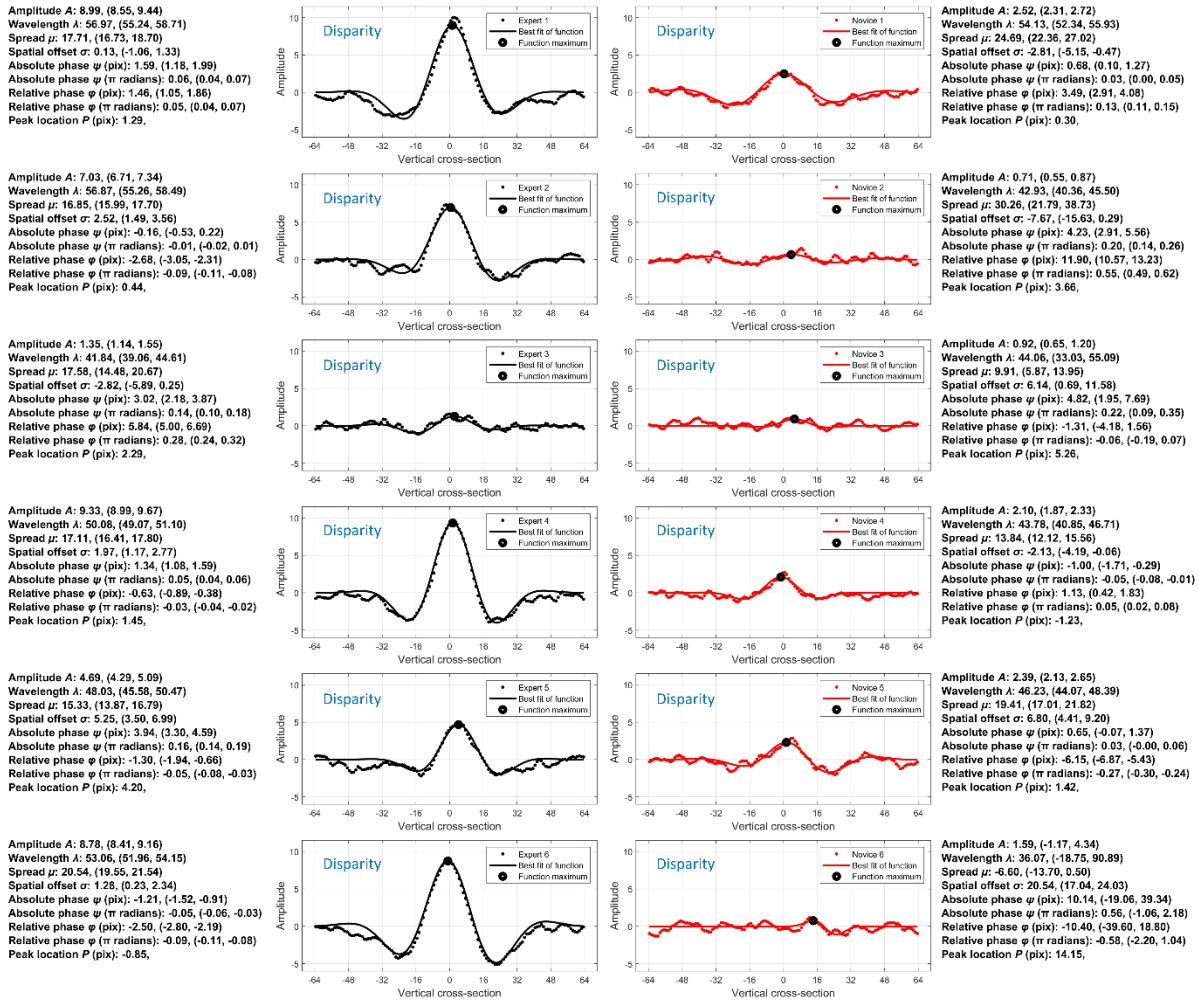

Figure C1. Vertical cross-sections of disparity classification images fitted with a Gabor function (Equation 3) for each participant. Gabor parameter values are listed above each participant's plot (with 95% confidence intervals shown in parentheses).

Appendix D: Fits to vertical cross-section of the luminance CIs for individual participants from Chapter 4

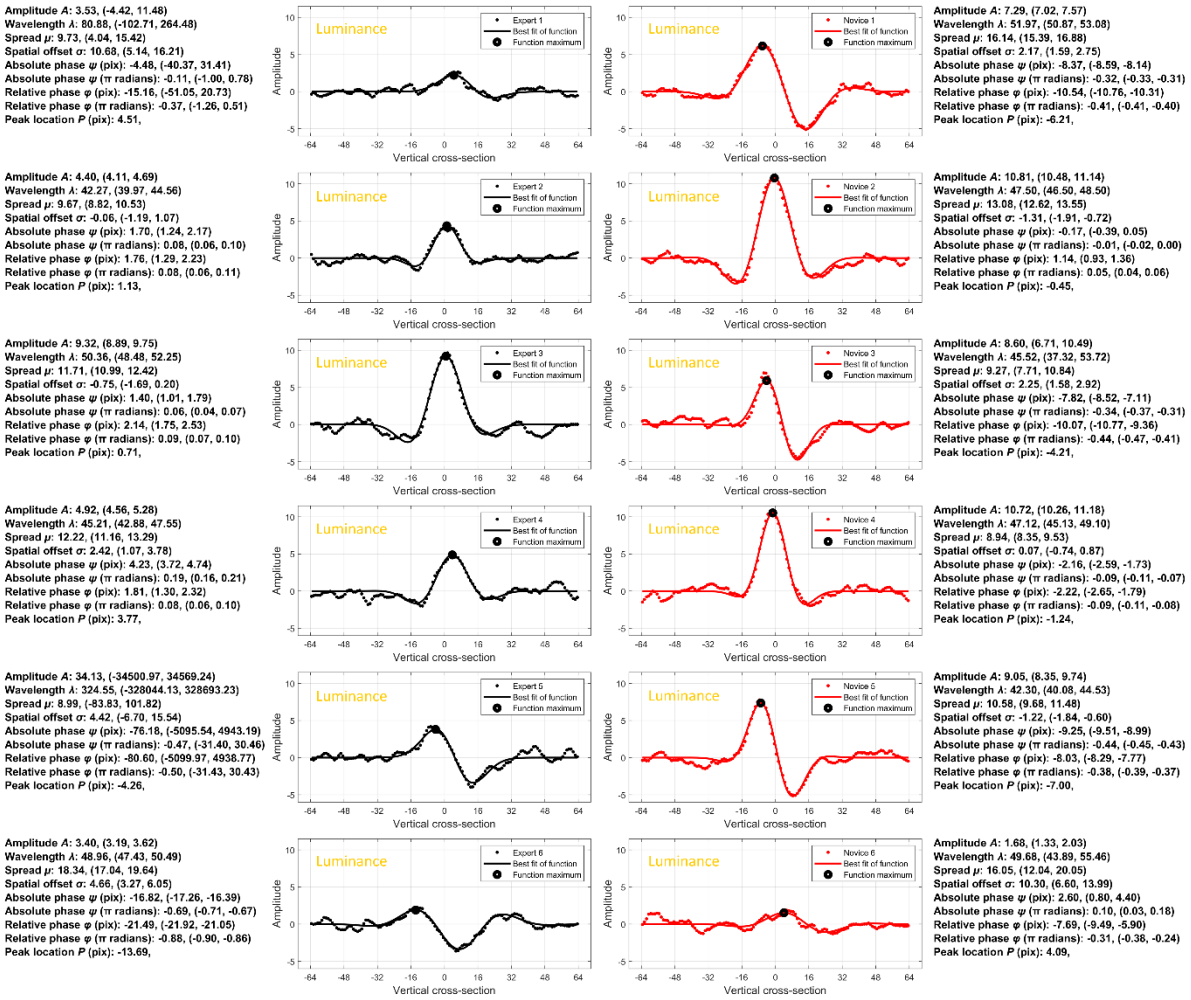

Figure D1. Vertical cross-sections of luminance classification images fitted with a Gabor function for each participant. Details are as for Figure C1. The fit to Expert 5's data is unusual. The Matlab optimisation routine was drawn towards an unusually high Gabor amplitude (much greater than the amplitude of the data) and a long wavelength. This produced a very shallow sine-wave component to the Gabor function around the zero-crossing that was amplified to reach the data by the high Gaussian amplitude. This unsual nuance had little or no influence on out estimates of relative phase (in π radians) and peak location, which are each reliable indicators of asymmetry, but posed a problem for the group statistical analysis of the amplitudes (*A*) from the individual fits. To address this, we tried constraining A, but found the fits were always drawn to the value of the constraint. We then tried constraining wavelength (λ), but found an interdependence between the value of the constraint and the estimate of amplitude. In a second approach (used in the main body of the report), we calculated the mean ratio between Gabor amplitude and the maximum and minumum values in the data for the other five experts and used this to estimate the Gabor amplitude for Expert 5 from their maximum and minumum values in their data.

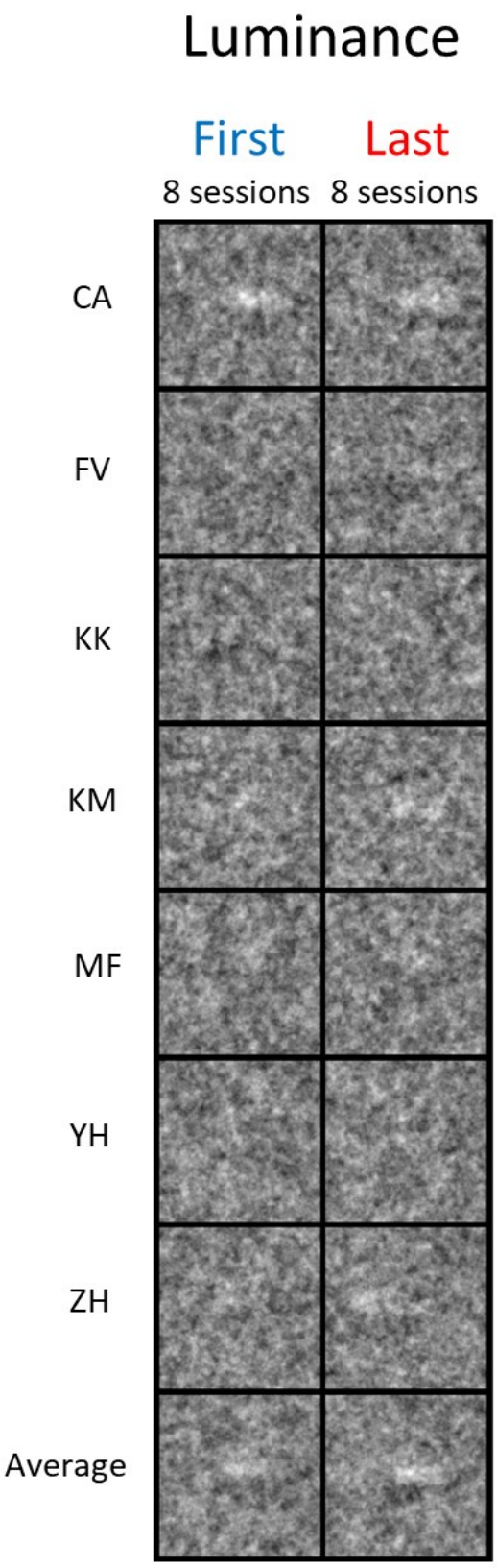

Figure E1: Partial luminance classification images.