VR 4334
25 April 2005 Disk Used
ARTICLE IN PRESS
No. of Pages 7, DTD = 5.0.1

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research xxx (2005) xxx–xxx
Vision Research
www.elsevier.com/locate/visres

The visual computation of 2-D area by human observers

M.J. Morgan *

Henry Wellcome Vision Research Laboratories, Department of Optometry and Visual Science, City University, Northampton Sq, London EC1V 0HB, United Kingdom

Received 14 October 2004; received in revised form 5 April 2005

Abstract

Normal human observers compared either the width, height or area of two simultaneously-presented shapes (the standard and the test), with a cue to indicate which decision had to be made. On 'area' trials, test width was a random variable, ensuring that neither shape (aspect ratio), width nor height by themselves was a reliable signal. Weber fractions for width and height of both ellipses and rectangles were in the range 5–10%, but for area they were higher (10–20%) than predicted from the combination of noisy width and height decisions. With ellipses, observers were more likely to overestimate width or height when the other dimension differed from the standard in the same direction (e.g. both greater). We conclude that observers have no access to high-precision codes for 2-D area, and that they base their decisions on a variety of heuristics derived from 1-D codes. A second experiment measured acuity for changes in aspect ratio. For ellipses, accuracy for aspect ratio was higher than predicted by the combination of noisy width and height signals; for rectangles it was worse, suggesting that 2-D curvature is a potent cue to shape.

Keywords: Psychophysics; Shape; Weber fraction

1. Introduction

We can see that an elephant is larger than a mouse even though the two animals have very different shapes. This could mean that we perceive the linear dimensions (such as the width and the height) of the elephant as larger; or we could have some special-purpose mechanism for computing the 2-D area, or even the 3-D volume, of an arbitrary shape. We consider here the simple case where the area is computable from two linear measurements, as is the case for conic sections like an ellipse. Observers might compute the area of an ellipse by multiplying their independent estimates of the major and minor axes, in which case the accuracy should be worse than either linear estimation taken separately, since they will be combining two sources of noise. If the two noise sources are independent, their joint variance will be the sum of their separate variances, according to the convolution theorem (Bracewell, 1965). The purpose of this paper is to see whether this is true. If accuracy for area is better than predicted by addition of variances, we shall be able to conclude that there is a specialised mechanism for 2-D area computation; if accuracy for area is worse than predicted, several conclusions are possible, as we shall see later in the paper.

The logic of the present approach is made clear by Heeley and Buchanan-Smith (1996), who applied it to the accuracy of angle estimation. They showed that angle estimation is more accurate than we should predict from the accuracy of estimating the angle of the component lines, and that, therefore, there is a specialised mechanism for angle computation. We are not aware of any analogous investigations of area. Laursen and Rasussen (1975) found that humans and monkeys can detect 2–6% changes in the aspect ratio of ellipses. Since

* Tel.: +44 207 040 0183; fax: +44 207 040 0182.
E-mail address: m.morgan@city.ac.uk

0042-6989/$ - see front matter © 2005 Published by Elsevier Ltd.
doi:10.1016/j.visres.2005.04.004

the Weber fractions for line length changes are similar, we may be tempted to conclude that there is a specialised mechanism for aspect ratio estimation, but without measurements of width and height accuracy in the same situation it is difficult to be certain. Using the method of single stimuli (a single rectangle elongated vertically vs elongated horizontally) Zanker and Quenzer (1999) found an accuracy of 1% when observers had to tell which of two ellipses was 'oriented more vertically', but this accuracy could have been based on height alone, since the area of the two shapes was the same. The same problem did not apply to Regan and Hamstra (1992) who randomly varied the relative area of the standard and comparison stimuli while requiring observers to compare aspect ratios. They found that accuracy was greatest when the reference shape was a perfect circle or square, falling off markedly with more extreme aspect ratios for the standard. They also measured independent width and height thresholds and found that these were 'comparable with, but not better than, corresponding aspect ratio near 1.4 and 0.7'. Although Regan and Hamstra did not make this comparison, the fact that thresholds for aspect were not worse implies that they were not derived from addition of independent noisy encoding of width and height. We return to this point in the discussion of our Experiment 2.

If we wish to measure how accurately observers can measure area or aspect ratio by eye, they must not be able to use either height or width separately, but must be forced to combine them. Therefore, like Regan and Hamstra (1992) we required observers to compare the areas of two shapes with randomly-differing widths and heights. One shape, the standard, always had the same area, but its width (and concomitantly its height) was a random variable over trials. The other (the test) had a different (randomly-chosen) width from the standard but in addition its area was different. The observer had to decide whether the test shape had a larger or smaller area than the standard. Randomly interleaved with 'area' trials were 'width' and 'height' trials, where the non-relevant dimension was randomly varied. An icon on the screen told the observer which kind of decision to make.

2. General methods

2.1. Apparatus and stimuli

The display (Mitsubishi DiamandPro colour monitor, with frame rate 100 Hz and background luminance halfway between the minimum and maximum of 0 and 75 cd/m^2) was programmed in MATLAB and stimuli were generated by a Cambridge Research Systems VSG 2/3 graphics card. Linear luminance look-up tables were constructed with the CRS Optical system. The viewing distance was 57 cm at which the total viewing area on the screen subtended $22.3° \times 22.3°$ and pixel size was 2.6 arcmin. The stimuli were white-filled figures (75 cd/m^2) on a grey background (37.5 cd/m^2), rather than outlines because it was felt that this was more like natural objects. No anti-aliasing was employed, but the minimum size change of 1 pixel (1% of standard width) was well below the thresholds we found.

2.2. Procedure

The shape of the fixation point on each trial indicated to the observer the decision (width, height or area) to be made. A horizontally-oriented rectangle (22×5.5 arcmin) indicated width, a vertically-oriented rectangle indicated height, and a square rectangle (25×25 arcmin) indicated area. One second after the fixation point appeared the stimulus array was presented, and remained visible for one second before being replaced by a mean luminance screen.

On each trial the width of the standard stimulus was randomly chosen from a uniform distribution in the range standard width ±50%. The standard width was 4.53°. The height of the standard stimulus was then chosen so that the area was constant and equivalent to the area of a rectangle of $4.53° \times 4.53°$.

The standard was always presented to the right of the fixation point and the test to the left, but the exact positions were jittered over trials to prevent the observer using alignment cues. The mean distance of the centres of the two stimuli were 5.7° from the fixation point, and the jitter was drawn from a rectangular distribution with limits ± standard width/4.

The procedure for generating the test stimulus depended on which condition was in operation on that trial:

Width trials. The width of the test stimulus was that of the standard, plus or minus a percentage change determined by the APE psychometric procedure (see below). For convenience, the percentage change will be referred to in future as 'the increment', although it is equally often a decrement. The height was randomly chosen in the range standard height ±50%.

Height trials. As for Width trials, but the increment was added to the height, and width was randomly varied.

Area trials. The width was randomly chosen from a uniform distribution in the range [standard width ±0.5 standard width], as the standard had been. The area was the same as the standard,, plus or minus a percentage change determined by the APE psychometric procedure (see below). The height was therefore determined as area/width. An example of the stimuli on an 'Area' trial is shown in Fig. 1.

Aspect ratio trials. In Experiment 2 observers compared the aspect ratio of the standard and comparison stimulus, instead of the area. The dimensions of the

VR 4334 ARTICLE IN PRESS No. of Pages 7, DTD = 5.0.1
25 April 2005 Disk Used

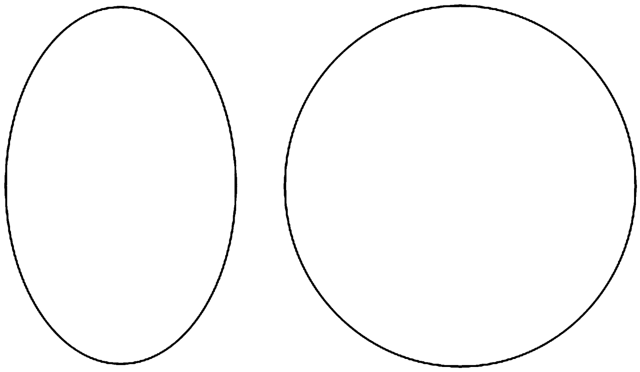

Fig. 1. Schematic representation of the stimuli on an 'Area' trial. In this case, the ellipse on the right has the greater area. If this were a 'Width' trial the correct choice would be the left-hand ellipse; if it were a 'Height' trial, the correct choice would be 'right'.

standard were determined as in Experiment 1. The dimensions of the comparison were then scaled by a random variable, before adding a percentage increment/decrement to the width. To help the observer to compare aspect ratios the mnemonic was suggested of thinking of the shapes as persons, and deciding which shape belonged to the 'thinner' person.

After each stimulus presentation the observer pressed the left of two buttons to indicate a 'smaller' decision and the right to indicate 'greater.'

2.3. Threshold determination

Thresholds (Weber fractions) were determined by the adaptive APE procedure (Watt & Andrews, 1981), which optimises the range and central values of the increment/decrement selected on each trial by a maximum likelihood fit of the data collected up to that trial to a cumulative Gaussian function having two parameters, σ representing the slope of the function and μ its central value or bias. The increment/decrement under control of APE was in the form of a Weber Fraction, that is, a proportion of the standard width/height/area on the trial in question.

Three independent APE-controlled sequences, each of 64 trials, were run in each block, corresponding to each of the three conditions (Width, Height, Area). The three sequences were randomly interleaved.

For the first 16 trials of each sequence, stimuli were selected in the range ±0.08 (Weber Fraction) as if the Method of Constant Stimuli were in operation, in order to give the observer clear examples of the stimuli. Following this initial run, stimuli were free to vary with a grain size of 0.01. Each block consisted of 192 (3×64) trials, and at least 4 blocks were run in each condition. Blocks with rectangles and ellipses were randomly interleaved.

When all the data had been collected for all conditions they were subjected to a single overall analysis to find the maximum likelihood fit to a Gaussian distribution with asymptotic probabilities for a 'larger' response of 0.01 and 0.99 to allow for the occasionally mistaken 'finger error'. Note that the full psychometric functions were fit, not just the thresholds. Two analyses were compared. In the first each condition width, height and area decisions had separate means and variances, giving rise to a six-parameter fit; in the second, the variance of the area decisions was calculated from the summed variances of the width and height estimates, giving rise to a five-parameter model. χ^2 for the difference between these two models was derived from twice the difference in their log likelihood, distributed with 1 (i.e. 6–5) degrees of freedom.

2.4. Observers

The observers in Experiment 1 were the author and one female undergraduate Optometry student (KS) naïve as to the purposes of the experiment. In Experiment 2 KS was replaced by DM, a male postdoctoral fellow in the Department of Optometry, who was an experienced psychophysical observer.

3. Experiment 1

Width, height and area trials were interleaved as described above. The object was to see whether area thresholds could be predicted by the addition of variances from width and heights thresholds.

3.1. Results

Weber fractions for the three kinds of decision (width, height and area) are shown in Fig. 2. Note that the error bars represent 95% confidence limits, not standard errors. Weber fractions for area were high, in the 10–20% region, much higher than those for width and height. The prediction for area is based on the notion that area is computed from independent measures of width and height with their own independent noise, in which case the threshold for area should be the square root of the sum of the squared thresholds for width and height (cf Ahumada & Watson, 1985; Burgess, Wagner, Jennings, & Barlow, 1981; Morgan, Hole, & Ward, 1990). Clearly, the actual thresholds for area are higher than those predicted by the addition of variances, so the independence model is rejected. We compared two models of the data by maximum likelihood as described in the Methods section. The χ^2 values for the difference in likelihood of the six- and five-parameter models were (for MM) 6.4 and 23.24 for ellipses and rectangles, respectively The equivalent values for KS were 6.64 and 14.42 All these values are significant (for df = 1) at the 0.05 level, leading to rejection of the

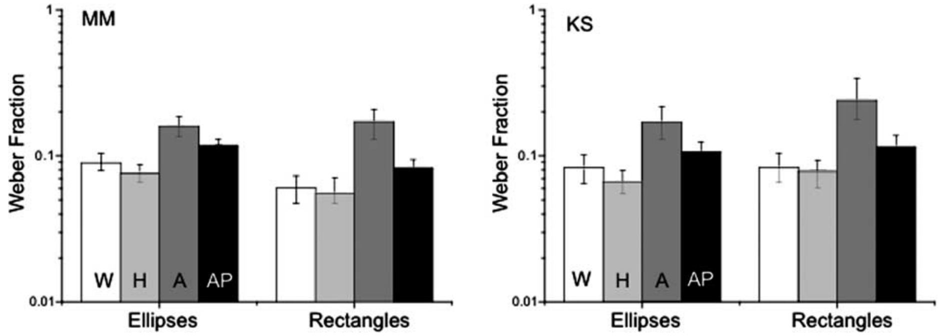

Fig. 2. Thresholds for width (W), height (H) and area (A) tasks for two observers and for two different reference shapes (ellipses and rectangles). Error bars represent 95% confidence limits, estimated from psychometric functions by a bootstrap procedure. The area prediction (AP) is based on the additivity of variances from independently-noisy width and height measurements. For details see the text.

addition of variances model. Note that the failure of the model is more extreme for rectangles than for ellipses, as Fig. 2 seems to indicate.

Some insight into the failure of additivity can be gained by a separate analysis of trials when the width and height vary in the same direction, as opposed to varying in opposite directions (Fig. 3). Accuracy tends to be higher in the 'same' direction (significant for both observers with ellipses, and for MM with rectangles). In other words, if both the height and width of the comparison stimulus are greater (or smaller) than that of the reference, observers tend to be more accurate in classifying it. The same tendency is seen for the width and height tasks themselves in the case of ellipses (significant for both width and height in MM; and for width in KS). These data suggest that observers cannot treat the variables of width and height independently. An analysis of biases (μ in the psychometric function, see Methods) supports this conclusion (Fig. 4). On 'width' trials when the test stimulus had a greater height than the standard, the observer was more likely to respond 'larger' than on trials when the height was smaller. Interpretation of these biases is complicated by an overall response bias towards 'smaller' responses (particularly in MM). However, the difference between the two kinds of trial is significant for both MM and KS in the case of the ellipse-width task and for the height task in MM.

To see if separate analysis of 'same' and 'different' trials significantly improved fit to the data, the likelihood of two-parameter fits to all the data were compared to those with four-parameter fits that had different values of μ, σ for the 'same' and 'different' trials. The results are shown in Table 1. In all cases, the values of χ^2 were significant at least the $p = 0.05$ level (df = 2).

We wondered whether our results might have been influenced by using solid rather than outline shapes. Author MM therefore repeated the Ellipse experiment described above, but with an outline rather than a filled figure. Results were very similar, and once again, thresholds for area were significantly higher than predicted from width and height by addition of independent noisy codes ($\chi^2 = 9.0$; df = 1; $p < 0.01$).

4. Experiment 2

As noted in the Introduction, several studies have claimed high accuracy for aspect-ratio discrimination, suggesting a high-precision mechanism for computing the ratio of width to length. However, only one of these

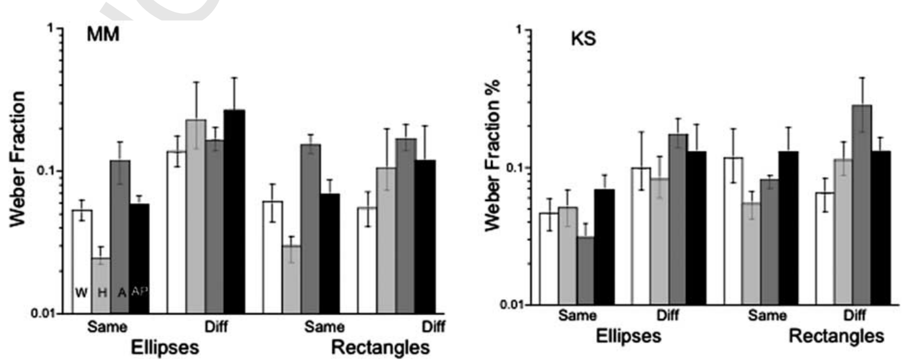

Fig. 3. Separate analysis of trials on which width and height co-varied in the same direction relative to the standard, vs. trials when they co-varied in opposite directions. For further explanation see the text.

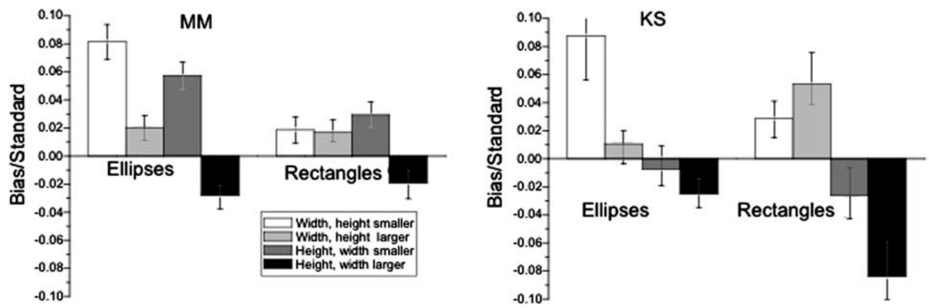

Fig. 4. Biases (μ of the psychometric function) in sub-sets of the trials in the experiment. Positive bias indicate a tendency to respond 'smaller' when the reference and test stimuli are physically equal. For further explanation see the text.

Table 1
Likelihood for two- and four-parameter fits to the data, and χ^2 values for the likelihood ratios

	Ellipses			Rectangles		
	Width	Height	Area	Width	Height	Area
MM 2 par	−394.32	−409.47	−431.59	−220.32	−211.78	−240.37
MM 4par	−346.39	−343.47	−426.05	−189.72	−194.78	−234.37
Chi-squared	−95.86	−132	−11.08	−61.2	−34	−12

For further explanation see text.

studies (Regan & Hamstra, 1992) required observers to compare figures of different overall size, so the decision could have been based on one dimension (width or height) alone. No study, as far as we are aware, has measured width, height and aspect-ratio thresholds concurrently in the same block of trials, to see if thresholds for the latter are predicted from separate noisy estimates of width and height. We therefore adapted the methods of Experiment 1 to measure thresholds for aspect ratio instead of area. The standard stimulus was generated as before and the tasks of width and height were the same. On 'aspect trials' the comparison stimulus was generated by scaling the overall size of the standard, to produce an invariant shape, and by then adding a percentage increment/decrement to the width. Thus, the best performance could be obtained only by examining the relative shape of the standard and comparison.

5. Results

The results are shown in Fig. 5. Qualitatively they were similar for the two observers. For ellipses, aspect ratio thresholds were better than those predicted from the summed variances of noisy width and height tasks; for rectangles they were worse. To assess significance of these differences, likelihood of independent (six-parameter) vs. dependent (five-parameter) fits to all the data were computed (see Experiment 1). For MM the χ^2 values derived from the likelihood ratios were 5.46 and 8.74 respectively, leading to rejection of the null hypothesis at the $p = 0.01$ level in both cases (df = 1). However, for observers DM the null hypothesis could not be rejected ($\chi^2 = 0.65$ in both cases). This means that DM could have been combining independent noisy estimates of width and height, although qualitatively his results go in the same direction as that of MM

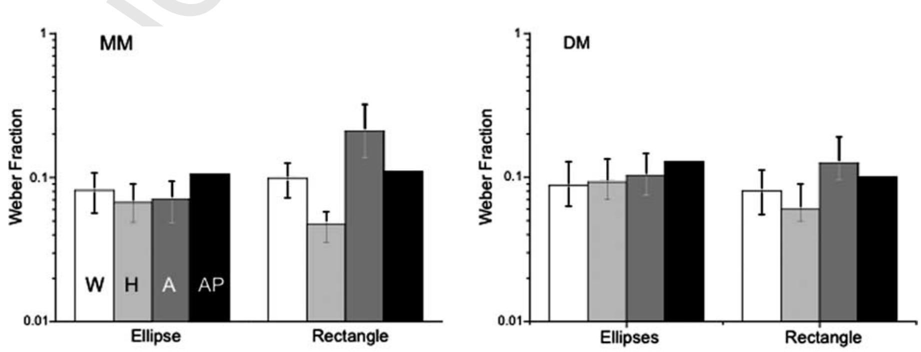

Fig. 5. Thresholds for width, height and aspect ration tasks in Experiment 2. Apart from the label 'A' referring to aspect ratio rather than area, conventions are the same as in Fig. 2

6. General discussion

6.1. Aspect ratios

We found no evidence in the case of rectangles that observers did any better at comparing aspect ratios than they would have by combining independent noisy codes for width and height. One observer (MM) did even worse. Only in the case of ellipses, and only for one observer (MM) was there evidence for a higher-precision mechanism The most likely candidate for the special-purpose mechanism is curvature discrimination, which has been shown to be a high-precision mechanism (Dobbins, Zucker, & Cynader, 1987; Fahle & Braitenberg, 1983; Koenderink & Richards, 1988; Watt & Andrews, 1982; Wilson, 1985). Of course, curvature discrimination is no use for rectangular shapes.

Our results for rectangles appear to contradict Regan and Hamstra (1992) who argued for a special-purpose mechanism. The discrepancy probably arises because we randomly varied the aspect ratio of our standard as well as its area. For extreme aspect ratios of the standard Regan and Hamstra found Weber fractions approaching 0.1, comparable to ours. It should also be noted that our Thresholds are based on the 18–82% points on the psychometric function rather than on the 25–75% points that were used by Regan and Hamstra. We have no reason to doubt Regan and Hamstra's conclusion that aspect-ratio thresholds for near-perfect ellipses and rectangles are better than predicted from separate width and height accuracy. As they point out, for squares observers could use the angle of intersection of imaginary lines joining opposite corners; for circles they could use curvature.

6.2. Area

First, we note that keeping the standard always on the same side, as we did in this experiment, means that this was the 'Method of Constant Stimuli' in the correct historical sense of that term (Morgan, Watamaniuk, & McKee, 2000) rather than two-alternative forced choice (2AFC). A potential problem with the Method of Constant Stimuli is that the observer can internalize the standard and need only look at the variable stimulus to make a decision, making the task equivalent to the Method of Single Stimuli (MSS). However, this was not true in this experiment, because width and height were random variables for the standard, so the observer was forced to compare the standard and comparison stimuli. Only on 'area' trials could the observer have ignored the standard without loss of accuracy, since the area of the standard was constant. However, the prediction of Signal Detection Theory is that ignoring the standard should decrease threshold by $\sqrt{2}$ (Morgan et al., 2000). The fact that the area of the standard was constant is thus unlikely to explain our finding that area thresholds are *higher* than those predicted from width and height.

We found that the accuracy of measuring the area of a shape by eye is not limited only by the accuracy of measuring its width and its height. The Weber fractions for area are amongst the highest ever reported, and it is likely that they would have been even higher for irregular polygons with unequal numbers of sides. If observers estimate the area of an ellipse by multiplying their estimates of length and width, there must be considerable noise following the multiplication stage itself. Our findings (Fig. 3) suggest that observers use a variety of heuristics for combining width and height estimates into an estimate of area. One such might be to classify the test shape as larger than the standard if both its width and height are greater. This condition produces greater accuracy than the cases where the two dimensions vary in different directions. If the width is greater but the height smaller, a decision could be reached by deciding which difference from the standard is greater. This amount to estimating the quantity:

$$\Phi = (\omega_1 - \omega_2) - (\eta_1 - \eta_2) \quad (1)$$

where ω and η are random variates corresponding to the widths and heights of the two stimuli. This makes identical predictions to the multiplication model we have already rejected. The implication is that there is further noise in the representation of the height difference and the width difference. For example, the observer might have to attend to the width and height dimension sequentially and store the intermediate results in a noisy memory. What is certain is that early in the encoding of width and height is not the only source of noise limiting area computation.

The finding that width decisions are influenced by height and vice versa could be explained by a cognitive confusion between the two dimensions (tending to say that a stimulus is less wide when it is in fact less high) or by geometrical effects. If an ellipse is flattened in height, its ends begin to resemble the ingoing arrows of a Muller–Lyer figure, possibly leading to underestimation of its width. Individual differences and differences between ellipses and rectangles in the bias data (Fig. 4) suggest that both cognitive and geometrical factors may be operating.

It may initially seem puzzling that there is not a high-precision, special-purpose mechanism for computing 2-D area. After all, the visual system has no difficulty in carrying out signal multiplication when it is useful, as the example of the Reichardt detector shows (Morgan & Chubb, 1999). The ecological answer may be the rarity of occasions where it is valuable to compare different shapes by their area. To see that an elephant is larger than a mouse is easy and useful because it is larger in all its dimensions; to decide whether an elephant has a larger 2-D area than a giraffe is more difficult and of more doubtful utility. We suggest that the encoding of 3-D shape attributes, such as volume, is derived directly from 1-D measurements using a variety of heuristics, without involving an explicit 2-D intermediate.

7. Uncited reference

Regan, Hadjur, and Hong (1996).

References

Ahumada, A. J., Jr., & Watson, A. B. (1985). Equivalent noise model for contrast detection and discrimination. *Journal of the Optical Society of America A, 2*(7), 1133–1139.

Bracewell, R. (1965). *The Fourier transform and its applications*. NY: McGraw-Hill.

Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science, 214*, 93–94.

Dobbins, A., Zucker, S. W., & Cynader, M. S. (1987). Endstopped neurons in the cortex as a substrate for calculating curvature. *Nature, 329*, 438–441.

Fahle, M., & Braitenberg, V. (1983). Curvature detection in the central and peripheral visual field of human subjects. *Neuroscience Letters Supplement, 14*, 108.

Heeley, D. W., & Buchanan-Smith, H. M. (1996). Mechanisms specialized for the perception of image geometry. *Vision Research, 36*(22), 3607–3627.

Koenderink, J. J., & Richards, W. (1988). Two-dimensional curvature operators. *Journal of Optical Society of America, 5*, 1136–1141.

Laursen, A. M., & Rasussen, J. B. (1975). Circle–ellipse discrimination in man and monkey. *Vision Research, 15*, 173.

Morgan, M., & Chubb, C. (1999). Contrast facilitation in motion detection: Evidence for a Reichardt detector in human vision. *Vision Research, 39*, 4217–4231.

Morgan, M. J., Hole, G. J., & Ward, R. M. (1990). Evidence for positional coding in hyperacuity. *Journal of Optical Society of America A, 7*, 297–304.

Morgan, M. J., Watamaniuk, S. N. J., & McKee, S. P. (2000). The use of an implicit standard for measuring discrimination thresholds. *Vision Research, 40*, 2341–2349.

Regan, D. M., Hadjur, L. V., & Hong, X. H. (1996). Two-dimensional aspect ratio discrimination for shape defined by orientation texture. *Vision Research, 36*, 3695–3701.

Regan, D. M., & Hamstra, S. (1992). Shape Discrimination and the judgement of perfect symmetry: Dissociation of shape from size. *Vision Research, 32*, 1845–1864.

Watt, R. J., & Andrews, D. P. (1981). APE: Adaptive probit estimation of a psychometric function. *Current Psychological Review, 1*, 205–214.

Watt, R., & Andrews, D. (1982). Contour curvature analysis: Hyperacuities in the discrimination of detailed shape. *Vision Research, 22*, 449–460.

Wilson, H. R. (1985). Discrimination of contour curvature: Data and theory. *Journal of Optical Society of America A, 2*(7), 1191–1199.

Zanker, J., & Quenzer, T. (1999). How to tell circles from ellipses: Perceiving the regularity of simple shapes. *Naturwissensshafter, 86*, 492–495.