Bernard Harper
bernard.harper@liverpool.ac.uk

Richard Latto
Department of Psychology
University of Liverpool
Liverpool, L69 7ZA
United Kingdom

Cyclopean Vision, Size Estimation, and Presence in Orthostereoscopic Images

Abstract

Stereo scene capture and generation is an important facet of presence research in that stereoscopic images have been linked to naturalness as a component of reported presence. Three-dimensional images can be captured and presented in many ways, but it is rare that the most simple and "natural" method is used: full orthostereoscopic image capture and projection. This technique mimics as closely as possible the geometry of the human visual system and uses convergent axis stereography with the cameras separated by the human interocular distance. It simulates human viewing angles, magnification, and convergences so that the point of zero disparity in the captured scene is reproduced without disparity in the display. In a series of experiments, we have used this technique to investigate body image distortion in photographic images. Three psychophysical experiments compared size, weight, or shape estimations (perceived waist-hip ratio) in 2-D and 3-D images for the human form and real or virtual abstract shapes. In all cases, there was a relative slimming effect of binocular disparity. A well-known photographic distortion is the perspective flattening effect of telephoto lenses. A fourth psychophysical experiment using photographic portraits taken at different distances found a fattening effect with telephoto lenses and a slimming effect with wide-angle lenses. We conclude that, where possible, photographic inputs to the visual system should allow it to generate the cyclopean point of view by which we normally see the world. This is best achieved by viewing images made with full orthostereoscopic capture and display geometry. The technique can result in more-accurate estimations of object shape or size and control of ocular suppression. These are assets that have particular utility in the generation of realistic virtual environments.

1 Introduction

Photographers are sometimes aware that the scenes they see with their normal direct vision will differ significantly from the 2-D representations produced when the scenes are imaged and transferred to photographic paper or a projection screen. Almost everything about the originally captured scene is conveyed in a modified or degraded form. The descriptions of classical image aberrations (for example, Langford (1989, ch. 2)) cover the effects of simple uncorrected lenses on only the shape or color of the imaged scene. However, there are many other changes in the transition from

Presence, *Vol. 10, No. 3, June 2001*, 312–330

the reality to the image. One of the best known (and most disconcerting to the subject) is the fattening effect of photography.[1]

The most obvious loss in conventional imaging is presence derived from stereo depth information (Freeman, Avons, Meddis, & Pearson, 2000; Freeman, Avons, Pearson, & IJsselsteijn, 1999; Hendrix & Barfield, 1996; IJsselsteijn, de Ridder, Hamberg, Bouwhuis, & Freeman, 1998). However, there are other more subtle effects of which we are often unaware that are worthy of note. Peripheral vision objects and scaling cues are usually excluded from photographic images. Photography almost always fails to reproduce scenes at their original (same-size) magnification. Even when this is achieved, photography cannot reproduce the detail that can be seen with normal vision from the original viewpoint while maintaining the angle of view. Natural brightness ranges are immensely difficult to reproduce because each image generation can add contrast and lose shadow or highlight detail. Accurate color reproduction is also almost impossible with conventional imaging, and subject color failure can be found in most types of image capture. These "fidelity failures" are often corrected for by trial and error or custom and practice techniques derived from professional knowledge (Langford, 1989, ch. 8).

The only thing that appears to be unchanged in a photograph is the point of view. However, the single-point perspective that makes a photographic image appear to be an accurate representation of the original scene can also convey inaccurate object information. Humans, too, perceive the world from a single-point perspective. By the process of cyclopean vision (Julesz, 1971), we see the world through a "cyclopean eye" that generates a single artificial viewpoint from a location midway between each real eye. In normal human vision, the processes of foveal convergence, accommodation, and stereo fusion allow the brain to construct a new perspective that differs from those seen by either eye individually. This cyclopean point of view appears to be similar to a 2-D photographic perspective. However, a single-lens system cannot reproduce the way in which we can focus/fuse on an object with two eyes and see both diverging and converging optical paths (figure 1) from the same position. With close-up objects, we have the ability to see the normal photographic perspective and also have "look-around" vision from a single head position. The result is that close-up objects viewed stereoscopically occlude less of the background than do their 2-D photographic equivalents. This paper investigates the possibility that the failure to reproduce this geometry in a display is a major cause of the fattening effects associated with conventional photographic images. A previous study (Yamanoue, 1997) found evidence of changes in size estimations in stereoscopic conditions. His experiments linked widening camera lens interaxial separations to smaller size perception and the "puppet theater" effect. He used direct observation of a mannequin and compared it with a same-size, parallel-imaged stereo video reproduction. In a later paper, Yamanoue, Okui, and Yuyama (2000) supported the use of lens separations and magnifications similar to those of the human visual system in order to reduce the appearance of an image artifact known as the "cardboard effect." In the stereo experiments reported here, only photographic images were viewed and only the stereoscopic disparity and convergences were changed.

In psychophysical experiments, monocular vision has consistently been linked to lower performance when compared to binocular vision, with the exception of the horizontal-vertical illusion (Prinzmetal & Gettleman, 1993). Tasks such as luminance increment detection, contrast sensitivity with sine wave gratings, color discrimination, vernier acuity, letter identification, and visual search (Banton & Levi, 1991; Blake, Sloane, & Fox, 1981; Jones & Lee, 1981) all show improved performance in the binocular condition. It is argued here that, whenever the visual system is presented with im-

1. It is commonly said in the fields of photography, film, and television that the camera "can put ten pounds on you." Yet we can find no academic reference for this effect, despite researching this phenomena with a number of institutions such as the British Journal of Photography, the Independent Television Commission, the Moving Image Society (BKSTS), the Royal Television Society, members of the American Society of Cinematographers, and more-conventional scientific resources. Although distortions are regularly mentioned anecdotally (Gunby, 2000; Kelly, 1998; Warner, 1995), until the present study, it appears that no one has examined the fattening effect of photography in a systematic way.

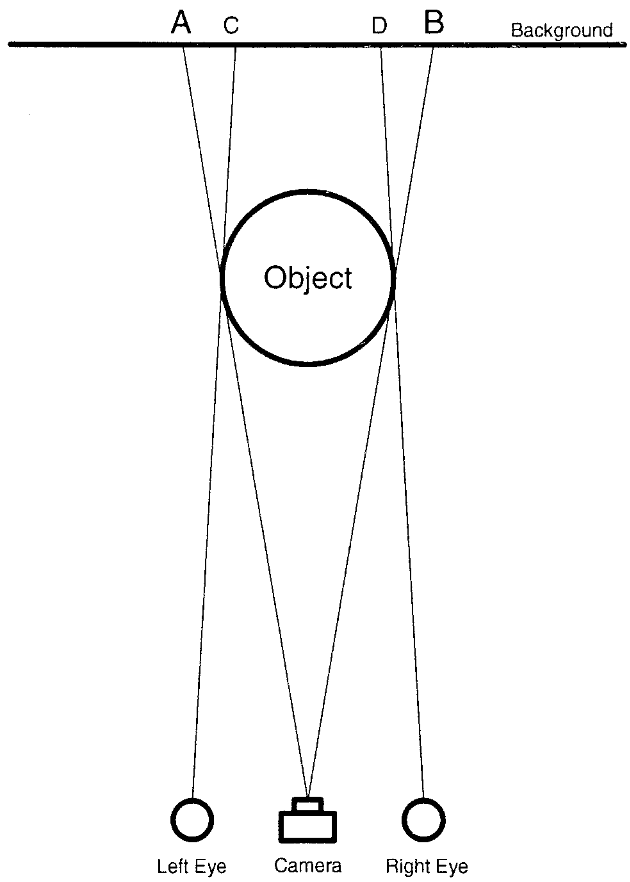


Figure 1. *The difference between a camera point of view and human stereo vision from the same position. The viewed object occludes more of the background in a 2-D photograph (AB) than in stereo vision (CD).*

ages that do not allow it to form a normal cyclopean view, predictable perceptual disturbances will occur: the display medium will be flawed in its ability to convey objects and people in their original proportions, size, and background occlusion characteristics. We propose that only a full orthostereoscopic capture and display system (Spottiswoode, Spottiswoode, & Smith, 1952)[2] can reproduce natural viewing geometries and provide a more lifelike visual experience.

2 General Method

The experiments reported here use orthostereoscopic imaging to investigate the distorting effects of photographic images. Two-dimensional images are less able to convey volumetric, contour, or shading information and can generate monocular optical illusions that fail with direct stereo vision (such as an Ames room). The hypothesis is that 2-D images distort because they do not present object information in the same way as a real object would under direct human observation. To minimize possible photographic distortions, the experiments use stereo image-capture geometry that is as close as possible to that of the human visual system. Conventional 3-D photography, which we are grouping under the term *parallel stereography,*[3] is inadequate because most stereo camera/display arrangements are not designed to match the geometry of human stereo vision.[4]

It was considered that viewing comfort should have a high priority in the presentations. There are limits (Panum's fusional area) to how far out of horizontal or vertical alignment binocular stimuli can be before there is loss of fusion and diplopia or suppression of one image (Howard & Rogers, 1995). We decided that the point

2. While recognizing the theoretical advantages of orthostereoscopic imaging and that this technique was "the condition of perfect image reproduction," Spottiswoode et al. (1952, p. 263) argued that this would constrain the artistic freedom of directors and cinematographers. So, their pragmatic solution was to reject these constraints for more-flexible and practical combinations of magnification, lens interaxial separations, and alignments. This often meant that images were captured using long telephoto lenses, wider-than-normal lens interaxials, "narrower than natural" convergences, and that the stereo window of reproduction was often placed behind the plane of focus/screen plane. They also considered that the primary orthostereoscopic conditions were 65 mm interaxial separation and same-size magnification.

3. Parallel stereography in this paper refers to stereo image-capture geometries that do not converge the lens axes on the center of focus and interest at the object plane and generate a double image at the plane of reproduction.

4. Almost all stereography uses different combinations of lens interaxial separations, magnifications, and convergences from those the human visual system would use when viewing the original scene. For instance, the average human interocular distance is approximately 65 mm, but stereo camera separations are often much wider than this. Also, they usually fail to reproduce the point of zero disparity from the original scene with zero disparity in the display. This means that they show a single point from the captured scene as two points on the screen and the viewers are required to "force fuse" these points to form a single stereo image.

of focus for each camera should coincide with the convergence point for each lens axis, and that this must be reproduced as a point of zero disparity in the display. This alignment was most likely to give comfortable viewing because, when the points of each camera's focus are horizontally aligned in the display, the center of interest (a face, for instance) appears as a single image. Zero separation in the display (no double image at the center of interest) means that relatively flat objects can be viewed successfully without polarizing spectacles. Typically, this condition has a high degree of 2-D compatibility as only the out-of-focus areas are not aligned at the screen. Polarizing spectacles allow the viewer to separate these areas into discrete channels by which they can then perceive the original scene depth. The principle that underlies all of the stereo experiments reported here is that orthostereoscopic images are presented to the participants for comparison with 2-D images from the same viewpoint and camera to subject distance. In practice, this means that, when participants are making size or shape judgments under experimental conditions, they are presented with images in which the only differences are of disparity.

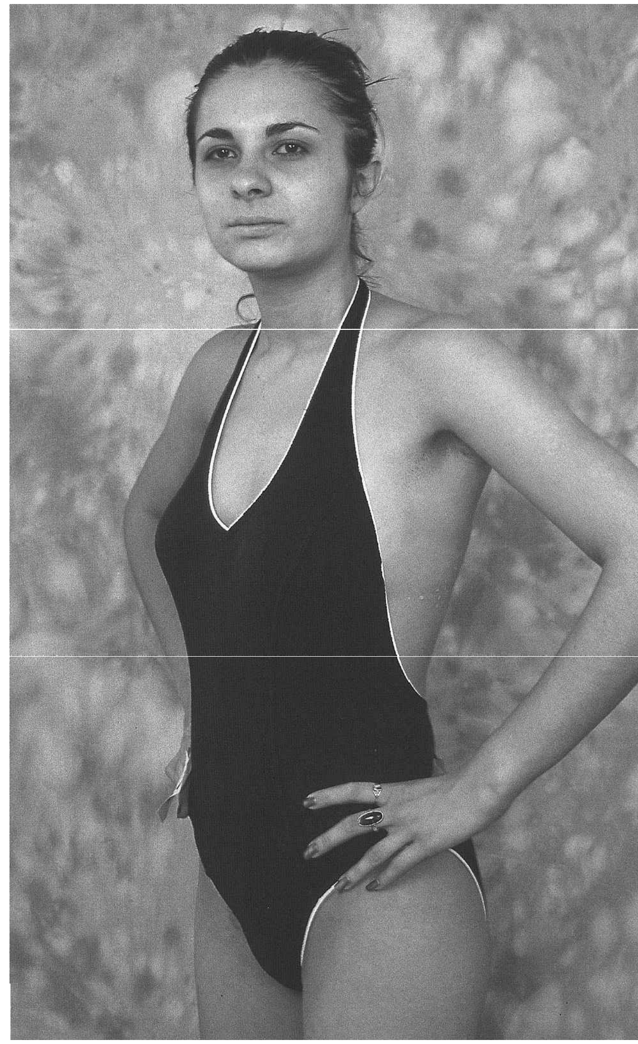


Figure 2. *Experiment 1: typical swimsuit image.*

2.1 The Stereo Camera

In experiment 1 and 2, a stereo camera was constructed using two Olympus OM1 cameras mounted vertically on a common baseplate and tripod mount. Standard 50 mm, f1.8 lenses were used to closely approximate the human eye's angle of view and magnification. The lens separation was 64 mm and the optical axis of each lens was converged on the point of focus 1.68 m away. The framing was for adults of normal height; the horizontal crop lines falling above the knees to just above head height (figure 2). Each shutter was triggered by a dual cable release staged to fire the flash lighting on the opening of the second curtain to ensure correct synchronization. This method allowed for bright, even illumination of the subject and for consistent exposures using small apertures (f16). It also ensured that the maximum depth of field and apparent sharpness would be recorded.

The images were recorded onto high-resolution Fuji 50 ASA transparency film. The transparencies were processed, selected for technical quality, and mounted into annotated 35 mm registration mounts. The left camera images were also copied to same-size magnification, and two color-matched copies were produced for synoptic[5] presentation. The exposures were carefully controlled, because the stereo images were intended for two-channel projection using cross-polarized filters and viewing

5. Following Koenderink, van Doorn, and Kappers (1994), we are using the term *synoptic* to describe the situation in which both eyes see exactly the same image with no binocular disparity, as in viewing a photograph, television screen, or a landscape at infinity.

through standard polarizing spectacles. This technique allows high-quality, full-color stereo images to be seen, but it causes a 50% loss of image brightness. Some of this brightness loss can be recovered because the technique requires the use of a polarization-maintaining (metalisized) projection screen. These are often used simply as high-brightness screens and, together with illumination by two projectors, this ensures a projected image of adequate brightness.

2.2 Stereo Projection

The transparencies were projected onto the metalisized screen using two carousel-type (Kodak Ektar) projectors with matched Kodak f2.8, 85 mm lenses. Because of their large size, the projectors could not be mounted side by side for correct orthostereoscopic projection, so a surface-silvered mirror was used to establish the correct optical path (figure 3). The right projector images were loaded normally, but the left projector images were laterally reversed to compensate for the mirror reversal in its optical path. Calibration images were then projected to same-size scale so that the projected model's interocular distance and height measured on the screen closely matched the measurements taken from the real person.

Side-by-side projection like this allows for the stereo window in which objects and scenes are reproduced to be easily moved towards or away from the viewer. For instance, it is possible by cross-converging the projectors (that is, moving one image horizontally) to place the background plane onto the projection screen and have the object appear to be reproduced in virtual space at the original camera-to-object distance. The projectors can also be diverged so as to move the object/stereo window behind the plane of reproduction. However, both of these alignments would require the images on the screen to be presented out of registration (figure 4). We speculated that this could cause the viewer to see objects as slimmer than they really are, because it might affect their perception of the true object boundary as it occludes the background. Incorrect vertical or rotational registration also might cause shape misperception (figure 5) for the same reason. So all of the images in

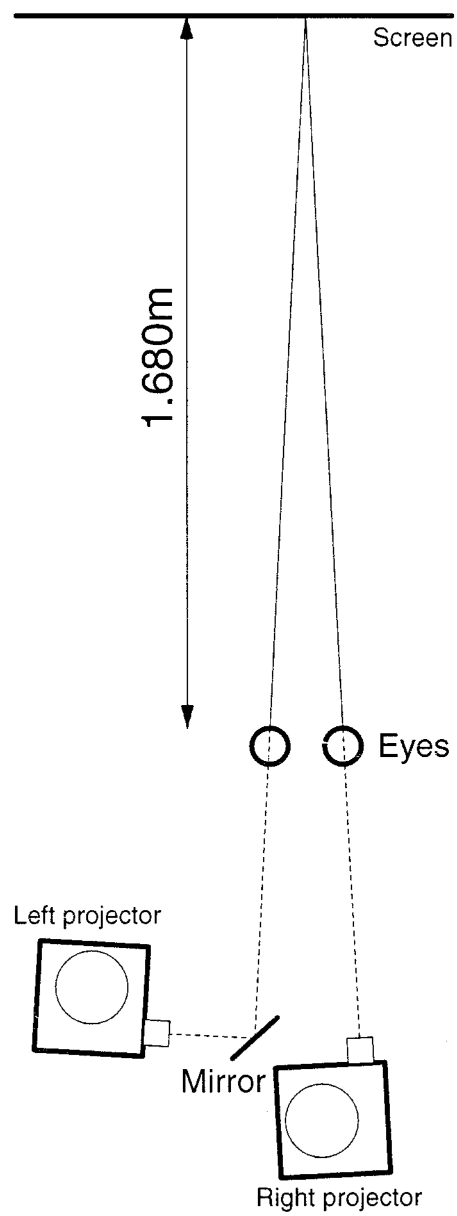


Figure 3. *Plan view of the projector alignment and viewing position for experiments 1 and 3. The viewers were positioned below the projectors' lenses to avoid occluding the image.*

these experiments were presented so that the vertical and horizontal registration of the point of interest/focus were of zero disparity at the plane of reproduction. Successful stereo projection also requires that image cross-talk (whereby one image channel can "leak" into another) be kept to a minimum. This can be achieved

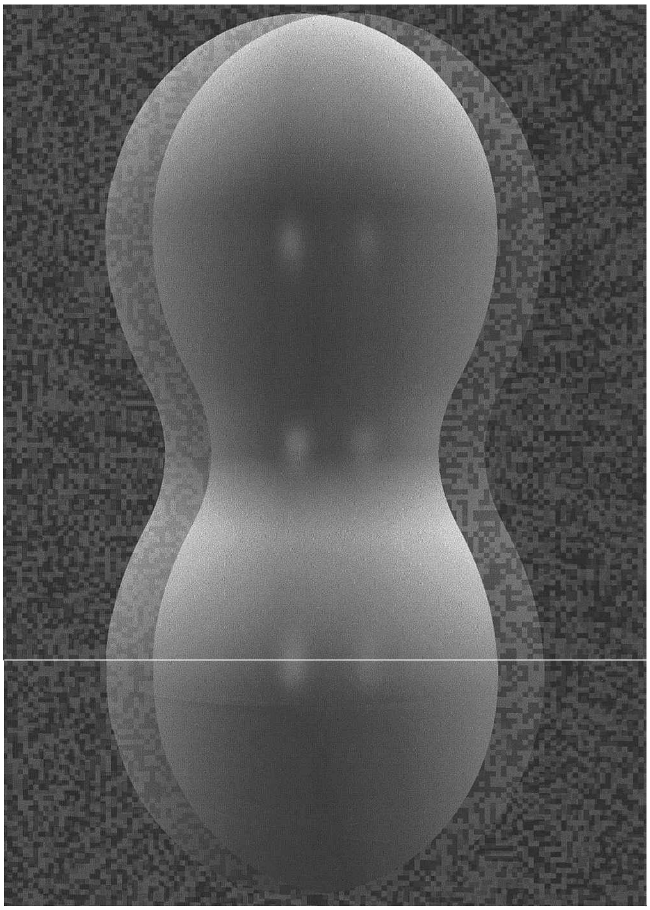


Figure 4. *Horizontal misalignment of the stereo window could cause a slimming effect by confusing viewers as to the true object boundary.*

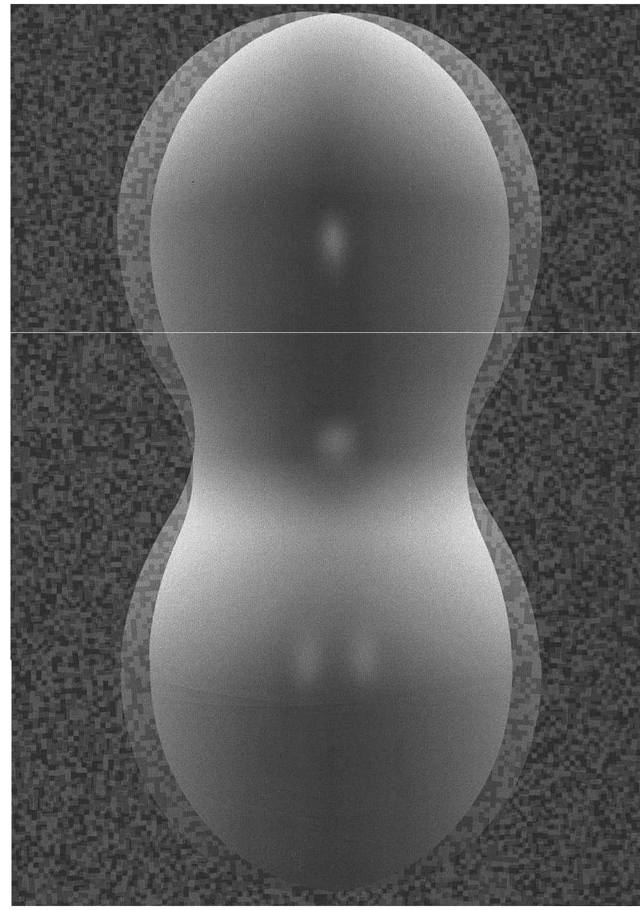


Figure 5. *Vertical or rotational misalignment of the projectors could cause a smaller waist to be seen in comparison with the hips and shoulder areas.*

by using professional-quality polarizing filters over each projector lens. These must be correctly aligned to 45 deg. (left and right) from the vertical to match the polarization angles of conventional 3-D movie spectacles. Image depolarization and cross-talk can still occur with these filters if the screen surface is not designed to maintain the polarization of the reflected image. In these experiments, image cross-talk was kept below 5% in each channel.

To test for the possibility that the slimming effect might be an artifact of projected stereo images, two Wheatstone viewers were used to present the transparencies in experiment 2. The advantage with this type of viewer (Pinsharp Viewer) is that it offers near same-size magnification, very high central resolution, zero cross-talk, and user control of the convergence for comfortable viewing. It also permits the presentation of a pair of conventionally mounted 35 mm stereo transparencies in one viewer and synoptic 2-D same-size copies in the other. When stereo pairs were shown to the participants, they could be asked to make comparisons between the 3-D and synoptic image while ensuring that the only difference between the conditions were the disparities presented.

2.3 The Virtual Stimulus

For experiment 3, a virtual "peanut-like" object was designed with the same imaging geometry as experiments 1 and 2 (see figures 11 and 12) using an architectural computer-aided design program (StrataVision 3D 4.0 from Strata, Inc.), with sophisticated rendering and lighting capabilities. The real-world image quality available with StrataVision is unlikely to generate the variable pixellation that could occur with simpler 3-D programs. It could also incorporate a random-dot background that was derived directly from stock Adobe PhotoShop files. When rendering stereo disparities using a computer-aided design package, it is important that the model is very accurately described, because small changes in topography or brightness due to aliasing can alter the stereoscopic detail within the image. The overriding design priority was that the virtual experiment could be repeated with a real object using stereo photography. It is therefore possible, should it be desired, for the virtual object and its background to be constructed and the camera/lighting simulation to be accurately reproduced.

3 The Fattening Effect of Zero-Disparity Images

A series of studies was performed to test the hypothesis that the absence of stereo depth information in 2-D images causes size and shape misperception of people and objects.

3.1 Experiment 1: Images of Female Models

3.1.1 Method. The stimuli, participants, and procedure of experiment 1 are as follows.

3.1.1.1 Stimuli. Ten female volunteers were photographed in stereo using the stereo camera described in section 2.1. The stereo photographs were taken with the models at three-quarter profile (figure 2). After being weighed and accurately measured, each model wore a dark swimsuit and was positioned in front of a flat photographic background over a floor mark. The left stereo photograph was copied to make a synoptic 2-D pair for the presentation.

3.1.1.2 Participants. Twenty-eight Liverpool University undergraduates were tested individually.

3.1.1.3 Procedure. Participants began by taking the TNO stereo acuity test (TNO, 1972) and viewing a series of projected 3-D slides to accustom them to stereo viewing. They were then shown life-size projected images of the ten models in alternating stereo and synoptic 2-D images, so that each model was never shown to the same participant in both 2-D and 3-D. Half the participants saw models 1 through 5 in stereo and models 6 through 10 in synoptic 2-D, and half saw 1 through 5 in synoptic 2-D and 6 through 10 in stereo. Trials were self-paced, and, during each presentation, participants rated the bodyweights of each model on a seven-point Likert scale: Very overweight, Overweight, Slightly overweight, Correct, Slightly underweight, Underweight, Very underweight.

3.1.2 Results. The mean perceived weight estimates of the ten models viewed either stereoscopically or synoptically are shown in figure 6. As the means and the very small standard errors indicate, there was a strong centralizing tendency in the participants' judgments, partly because the range of bodyweight in the models was not high but partly also probably because of a reluctance on the part of the participants to make negative judgments on the models. Nevertheless, a one-factor (viewing condition) ANOVA showed that the models were rated as significantly slimmer when viewed stereoscopically ($F(1,26) = 15.072$, $p = 0.001$).

3.1.3 Discussion. Although there was a significant slimming effect of stereoscopic presentation, it was possible that this was an indirect effect of evoking increased presence in 3-D presentations. Informal reports from several participants suggested that they sometimes felt they were in the presence of real people. Perhaps increased presence may have led the participants to give judgments that were less harsh to models that they felt were more present in the laboratory. Although this seems unlikely, particularly as most viewers were unaware that the presentation mixed 2-D and 3-D images, it was decided in experiment 2 to test this finding using an inanimate object. The generalizability of the initial finding was also tested further by using

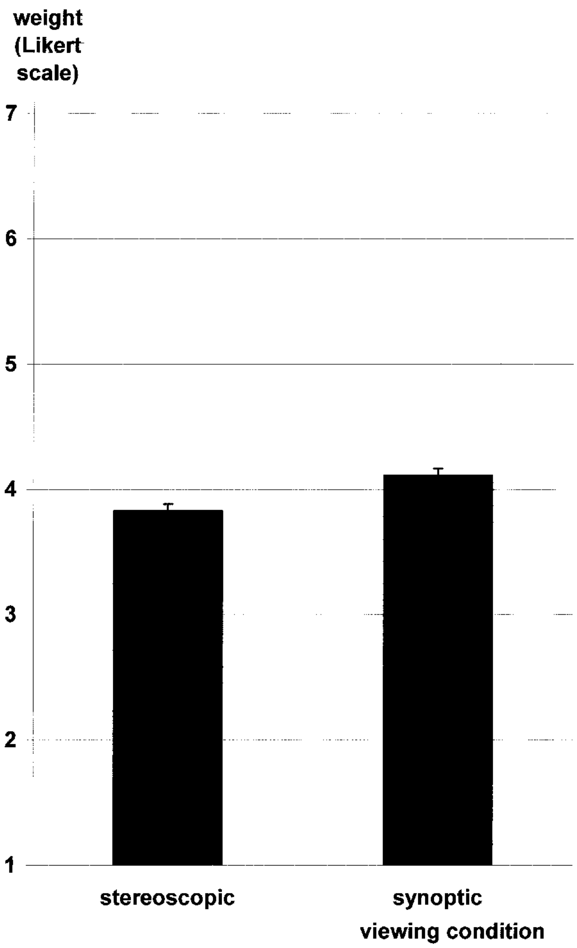


Figure 6. *Experiment 1: effect of viewing condition on mean perceived weight.*

Figure 7. *The stimulus used in experiment 2.*

Wheatstone viewers, rather than projected images, and a forced-choice rather than a scaling procedure for size estimation.

3.2 Experiment 2: Images without Human Presence

3.2.1 Method. The stimuli, participants, and procedure of experiment 2 are as follows.

3.2.1.1 Stimuli. Two large flower pots were arranged to form a waisted object (figure 7) which was then photographed using the same camera and image-capture geometry as used for the stimuli in experiment 1. The stereo transparencies were made using the method described in section 2, but the object was daylight-illuminated with the background plane imaged at infinity. The transparencies were mounted in a Wheatstone-type handheld stereo viewer. The horizontal/vertical field of view was 40 deg., and the viewer had user-variable vergence control. A second viewer held two same-size copies of one of the stereo transparencies, forming a synoptic pair.

3.2.1.2 Participants. Twenty Liverpool University undergraduate participants were tested individually.

3.2.1.3 Procedure. While viewing a series of pretest stereo images, each participant was shown how to use the two Wheatstone viewers.

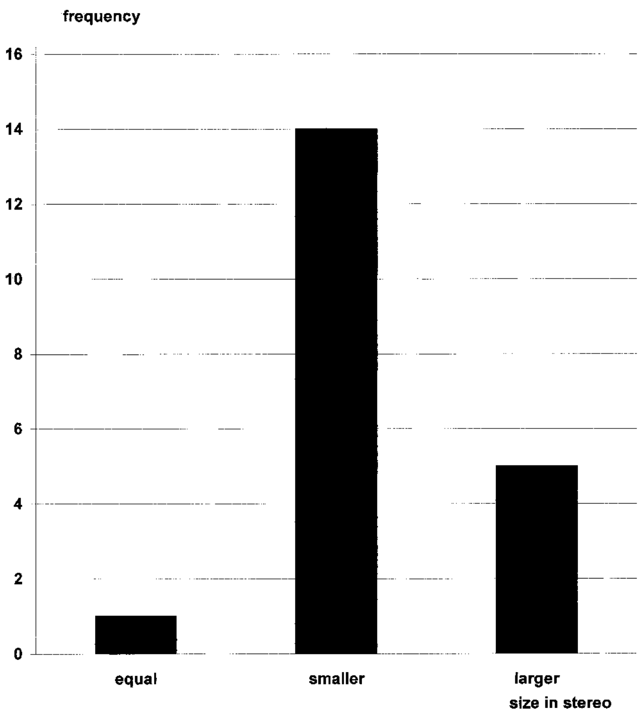


Figure 8. *Size comparisons of the stimuli in experiment 2 in synoptic and stereo conditions.*

Each viewer was then loaded with the stimuli, and the participants were asked to look carefully at the dimensions of the object in both viewers. They were asked if they could see any size difference between the images in each viewer. If they reported a difference, they were asked to choose which image was wider or larger than the other.

3.2.2 Results. The results shown in figure 8 confirm the prediction that the waisted object was viewed as slimmer or smaller in the stereo presentation ($\chi^2(2, N = 20) = 13.3, p < 0.001$). Almost three times as many viewers saw the object as slimmer or smaller when viewed binocularly compared to the synoptic image.

3.3 Experiment 3: Digital Variable-Waist Images

When directly comparing the synoptic and stereo images of female models in experiment 1, it seemed that not only were the models appearing to be slimmer but also that their proportions were subtly altered. Both necks and waists appeared to be disproportionately slimmer than their associated jaw and hip widths. The flowerpot stimuli used in experiment 2 also seemed to support this view, and simple trigonometry confirmed that this was possible (figure 9 and 10). A new shape-matching experiment was designed to test whether perceived waist-hip and jaw-neck ratios could be affected by changing between 2-D and stereo image presentation. Two additional conditions were also introduced. Two different disparities in the binocular condition were used to look at the relationship between the degree of size distortion and the magnitude of the disparity. A parallel-axis stereogram was also included to allow the direct comparison of the distortions in parallel and convergent stereo. All of the participants were also tested for stereo acuity using the TNO test to establish if this was a reliable predictor of performance.

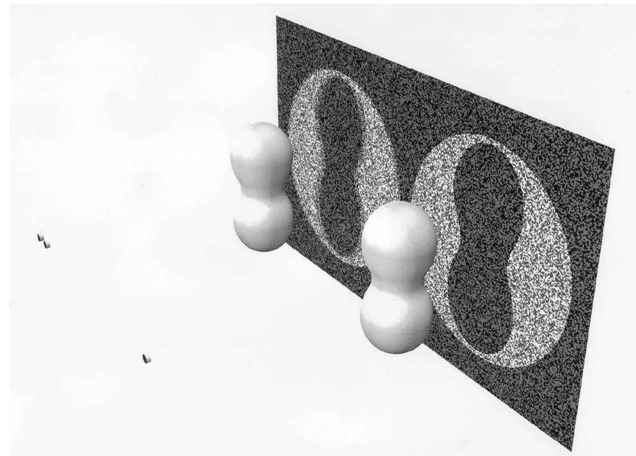


Figure 9. *The size and shape of the occluded area behind the object. The occluded area not only becomes smaller with disparity (left image), but the waist-hip ratio also changes; the wider the disparity, the lower this ratio becomes. (See also figure 10.)*

3.3.1 Method. The stimuli, participants, and procedure of experiment 3 are as follows.

3.3.1.1 Stimuli. A peanut-shaped 3-D model (see section 2.3) was designed. The widest part of the stimu-

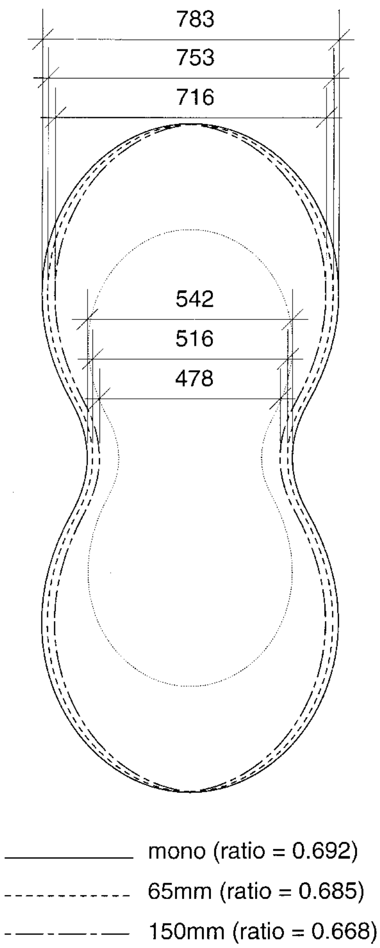


Figure 10. *Diagrammatic representation of the size and shape of the occluded area behind the object shown in figure 9 quantifying the way in which the occluded area becomes smaller with disparity and its waist-hip ratio lowers with increasing disparity. (It should be noted that the occluded area from the monocular position does not equal the 0.7 waist-hip ratio of the foreground object, because it was not imaged from a camera at infinity.)*

lus is described in these experiments as the "hips." The narrowest is the "waist." The waist circumference in figure 11 is 70% of the size of the hips, and this is described as a 0.7 waist-hip ratio. All of the stereo and synoptic images of the stimuli in experiment 3 are of this 0.7 ratio. Its surface was rendered without texture so that the only stereo information available to the viewer was from lighting-derived contour and shading and the trapezoidal distortion (perspective keystoning) of the background.

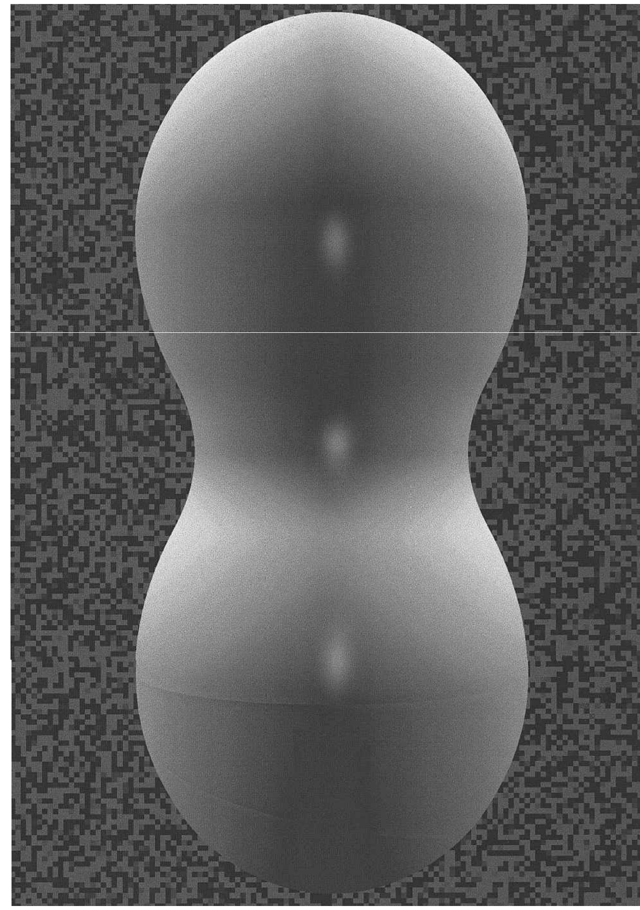


Figure 11. *An example of the peanut shape used in experiment 3 with a waist-hip ratio of 0.7. (This image is cropped for reproduction so the background is smaller than in the test stimulus.)*

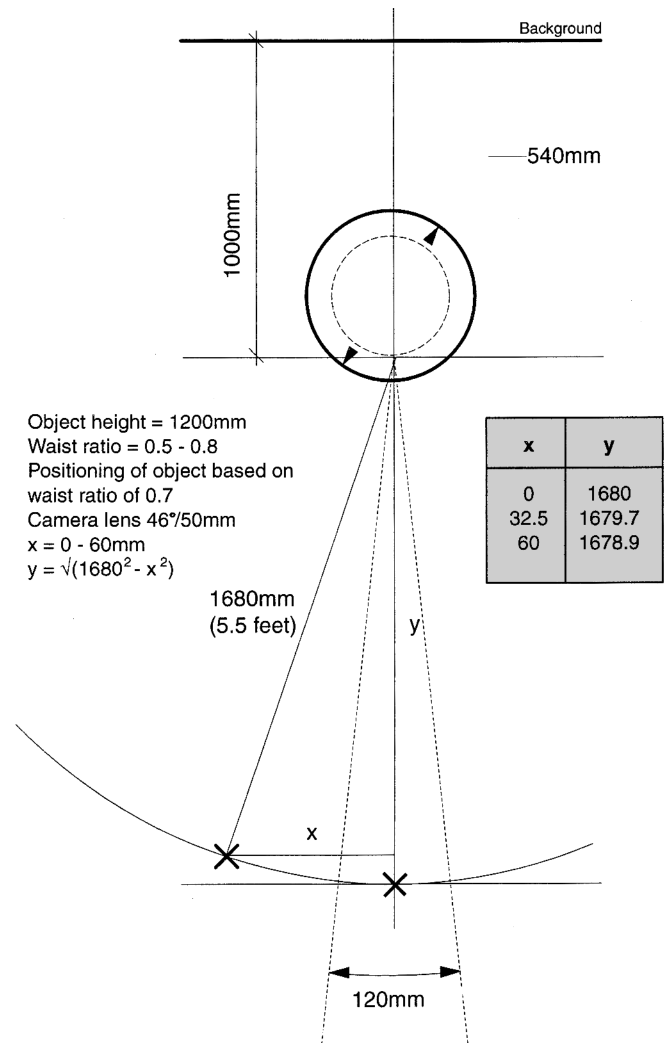
Four computer-generated stereogram pairs of the 0.7 peanut model were rendered for polarized projection to individual participants. These images were made in a series of widening disparities with 00 (synoptic, 2-D), 65P (65 mm, parallel axis), 65C (65 mm, convergent axis) and 120C (120 mm, convergent axis) interaxial equivalent separations. The peanut was constructed to approximate the "ideal" 0.7 waist-hip ratio of a healthy adult female (Singh, 1993) but with rotational symmetry in order to have the same shape from any horizontal angle. In the plan view (figure 12), the peanut and its relationship to the virtual cameras and the background are shown. These were designed to be identical to the arrangement used in experiment 1. The background was a random-dot wall of light-gray and dark-gray pixels. Stereoscopic and synoptic images were projected onto a screen using the same procedure as in experiment 1. These projected images were the equivalent of life-size, with the background subtending 31.6 deg. wide by 47.0 deg. high and the peanut subtending 18.6 deg. wide by 39.3 deg. high. Its waist subtended a visual angle of 13.0 deg.

The order of presentation of the four images was rotated round a Latin square to avoid order effects. A set of thirteen A4 comparison photographs was made of the peanut from the zero-disparity position. The image on each card was identical to the projected 3-D images except that their waist-hip ratios varied from 0.5 to 0.8 in 0.025 steps (figure 13).

3.3.1.2 Participants. Twenty Liverpool University undergraduates were tested individually.

3.3.1.3 Procedure. The thirteen comparison cards were randomized and the participants asked to place them in order from slimmest waist to fattest waist in order to familiarize themselves with the stimuli. They were then shown the first image of the sequence of varying-disparity images and asked to pick a card that matched the shape of the peanut as it appears on the screen. This was repeated with the remaining three images.

3.3.2 Results. Figure 14 shows the frequency distributions of participants' matches for the four different disparities. Figure 15 shows the overall group means for these choices. A one-factor (disparity) ANOVA found an overall effect of disparity on size judgement ($F(2,38) = 7.628$, $p = 0.002$). Post-hoc paired comparisons showed that the only significant differences were between the 0 deg. (synoptic) and the 65 deg. (stereo) ($t(19) = 3.367$, $p = 0.003$, two-tailed) and between the 0 deg. (synoptic) and the 120 deg. (stereo) $t(19) = 3.286$, $p = 0.004$, two-tailed).



x	y
0	1680
32.5	1679.7
60	1678.9

Figure 12. *Plan view of the dimensions of the peanut shape used in experiment 3, its relationship to the virtual camera positions, and the plane of the background. The virtual cameras generated views at each disparity in an arc to ensure that the magnification was constant in every image. The right-hand bold X shows the position of the camera when it was in the straight-ahead position (zero disparity). The left-hand bold X shows the position of the left-hand camera at a distance x* from the straight-ahead position. The disparity this generates is defined as 2X mm.

3.3.3 Discussion. It can be seen in figure 14 that the image-capture geometries (or disparities) used in this experiment reveal a previously unseen effect. The zero-disparity 2-D stimuli (card 9) was correctly matched to its projected equivalent (0.7 waist-hip ratio) by more than half of the participants (figure 14a). The average perceived waist-hip ratio of the group was 0.694 (figure 15). This is almost identical to the oc-

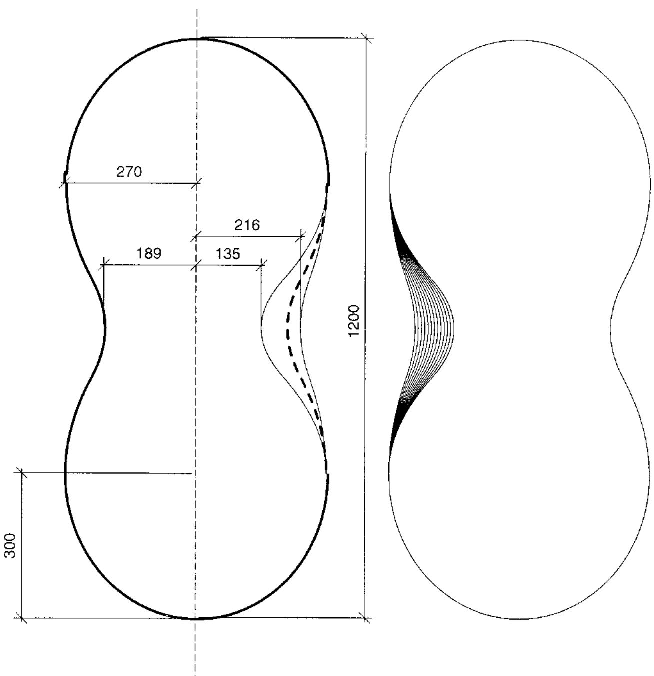


Figure 13. *The waist-hip ratios of the thirteen comparison stimuli used in experiment 3. Each stimulus was printed onto an A4 card (with a random-dot background, as in figure 11). Card 1 had the slimmest waist-hip ratio of 0.5. Each of the subsequent cards had a ratio that increased in 0.025 graduations. Card 9 (see also figure 11) was the same 0.7 ratio as the stereo and synoptic images. Card 13 was at a ratio of 0.8. The left diagram shows the largest and smallest physical dimensions of the varying waist sizes. The diagram on the right shows all of the intermediate ratios. The card images were scaled so that they were approximately the same size as the projected image when held at arms length.*

cluded area as shown in figure 10 of 0.692. However, when the viewers were shown the same shape but in stereo with 65 mm of convergence disparity (corresponding to the normal geometry of human stereo vision), a match with a significantly slimmer waist-hip ratio was selected. Conventional stereo cameras do not capture images with convergent lens axes but use parallel capture geometry. When this condition was simulated with a test image (65P), the mean perceived waist-hip ratio did not differ significantly from the synoptic condition. It can also be seen in figure 14c that there is much more variation in responses in this condition.

There was no correlation between the participants' stereo acuity, measured with the TNO test, and perceived waist-hip ratio in the 65C condition ($r = 0.29$, $n = 20$, $p = 0.904$) (figure 16). Subdividing the participants into those with high (15 sec. to 60 sec. of arc) and low (120 sec. to 480 sec. of arc) stereo acuity and using a mixed-design, two-factor (acuity and disparity) ANOVA showed that there was no effect of acuity on their performance in the size-judgment task ($F(1,18) = 0.46$, $p = 0.506$). Neither was there an interaction between the effect of disparity on size judgments and the performance in the stereo acuity test ($F(3,54) = 1.49$, $p = 0.228$).

3.4 Experiment 4: Varying Size Judgments in Zero-Disparity Images

In conventional photography, it is known that using lenses of different focal lengths can change the perceived size and shape of objects. Wide-angle lenses used in close proximity to scale models can make them look much larger than they really are. Telephoto lens compression can trick the viewer into misperceiving the spatial relationship between objects. For example, it can make the moon look oversized when it is framed with buildings or people. However, the perspective-flattening effect of telephoto lenses is rarely associated with the fattening effect that is so often mentioned in relation to photographic portraits, film, and television. Experiment 4 was designed to test the hypothesis that bodyweight appears higher in telephoto images and lower in wide-angle images. Of particular interest was the effect of different focal lengths of lens on the perceived width of the model's neck relative to the width of the jaw. Figure 17 shows how varying lens-to-subject distances can change the measured waist-hip ratio of the occluded area (as well as the expected size change) behind the peanut shape. It should be noted that quoting focal lengths in millimetres can be misleading. Lens calibrations can offer different image magnifications depending on the camera used. For instance, a 50 mm lens on a 35 mm SLR is considered to be a standard lens. On a 6 × 6 camera, it is a wide-angle lens. On a video camera, it would be a telephoto lens. For experiment 4, the inde-

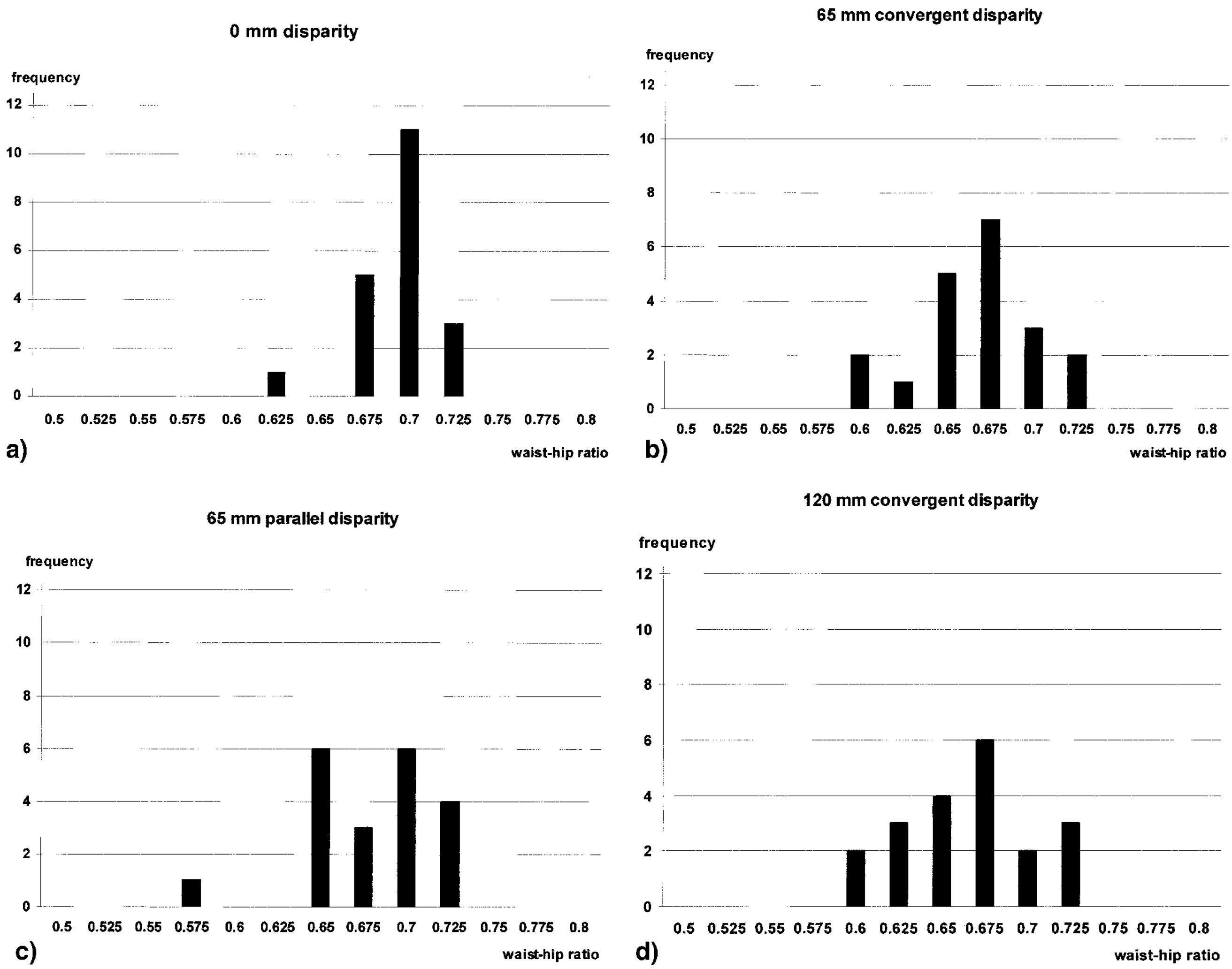


Figure 14. *Experiment 3. The matches that the participants made when shown the shape with a waist-hip ratio of 0.7: a) synoptically (0 mm disparity); b) stereoscopically with 65 mm, convergent disparity; c) stereoscopically with 65 mm, parallel disparity; d) stereoscopically with 120 mm, convergent disparity. Increasing the convergent stereo disparity to 120 mm results in a lower perceived waist-hip ratio.*

pendent variable reported is therefore camera-to-subject distance while maintaining a same-size image, because this is repeatable regardless of the camera system or lens design used.

3.4.1 Method. The stimuli, participants, and procedure of experiment 4 are as follows.

3.4.1.1 Stimuli. Two men and three women were photographed in identical poses using zoom lenses in a series of five focal lengths from wide-angle to telephoto. Using guides in the viewfinder, the lenses were zoomed very accurately for each of five camera-to-subject distances. This method made it possible to record the facial features of each model to the same magnification at the film plane from distances of 0.32 m, 0.45 m, 0.71 m, 1.32 m, and 2.70 m. Prints were made from the portraits and five sets were assembled, each containing one photograph of each model at one of the five focal lengths. Examples from the range are shown in figure 18.

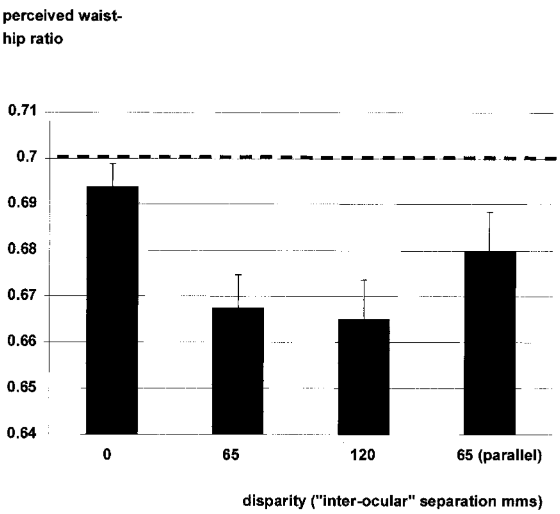


Figure 15. *Perceived waist size when the object was projected in the four different disparities used in experiment 3. The dashed line shows the actual waist-hip ratio of the stimulus.*

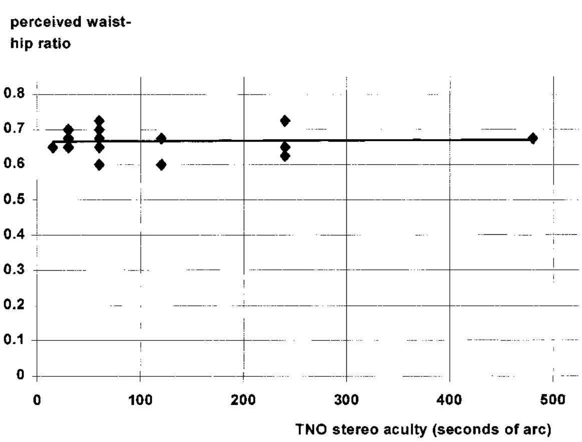


Figure 16. *The relationship between perceived waist-hip ratio in the 65 mm, convergent disparity condition, and the stereo acuity of the individual participants in experiment 3.*

3.4.1.2 Participants. Twenty Liverpool University undergraduates were tested individually.

3.4.1.3 Procedure. One set of photographs was shown to each of four groups of five participants. They were never shown the same model photographed at more than one focal length (unlike the examples in figure 18). Each participant was asked to place the five different model portraits in a rising order of apparent bodyweight using the same seven-point Likert scale as in experiment 1 (see section 3.1.1.3), and to apply a number from 1 to 7 to each image. A number greater than 4 was given to people who appeared to be overweight and numbers less than 4 to people who appeared to be underweight. The most overweight would be given a score of 7, the most underweight a score of 1.

3.4.2 Results. Figure 19 shows that, as camera-to-subject distance (and focal length) increases, a higher score was given on the Likert scale ($r = 0.824$, $N = 5$, $p < 0.05$, one-tailed). A one-factor (camera-to-subject distance) ANOVA found an overall effect of distance on size judgment ($F(4,76) = 8.858$, $p < 0.001$). Planned comparisons using two-tailed t-tests, showed that the wide-angle, close-proximity images (0.32 m) showed underweight estimations ($t(19) = 4.073$, $p = 0.001$). The standard lens image (0.71 m) showed a slight but not significant overweight estimation ($t(19) = 1.097$, $p = 0.287$). The telephoto distance images (1.32 m and 2.7 m) showed overweight estimations ($t(19) = 2.101$ and 5.101, $p = 0.049$ and < 0.001).

3.4.3 Discussion. Because of the limitations of the photographic location and lenses available, it was not possible to test if extending the range of focal lengths would show a continuing positive relationship between focal length and perceived bodyweight. It is likely, however, that the focal lengths used in this experiment cover the range in which the strongest effects could be demonstrated. Extreme wide-angle distortions at one end of the scale and proportionally smaller changes in the depth-compression effect of telephoto lenses at the other would probably act to curtail the effect.

4 General Discussion and Conclusions

These experiments support the theory that conventional imaging methods can convey misleading object information. Images of people seem to carry the

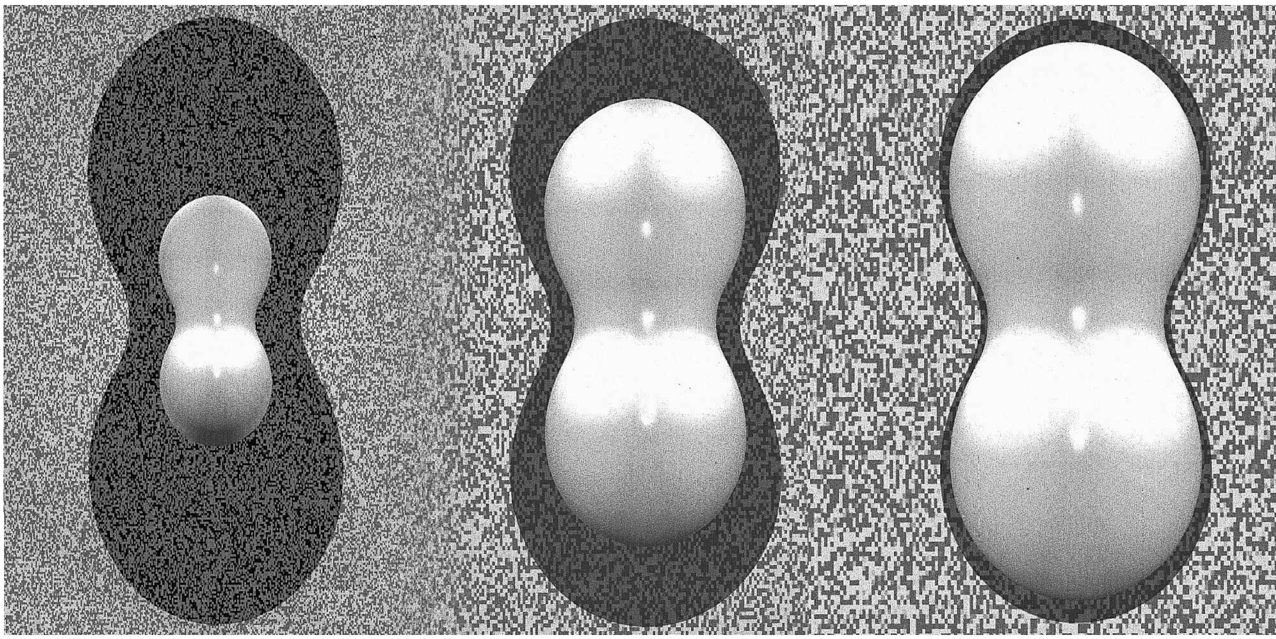


Figure 17. *Only objects viewed or illuminated from optical infinity can generate an occluded area that is the same size as the object. In this illustration, a light source is moved closer to the object in three stages (from right to left). The waist-hip ratio of the occluded area becomes lower as the source becomes closer.*

strongest effect as the tendency to use long-focal-length lenses combined with 2-D reproduction produces a significant flattening and fattening effect. It may be that we have specific mechanisms for shape recognition of the human body (Perrett, Harries, Mistlin, & Chitty, 1990) which are particularly sensitive to interference by different methods of imaging. Experiment 1 supported the theory that people look slimmer when viewed stereoscopically. Experiments 2 and 3 showed that the slimming effect of binocular disparity is seen with inanimate objects as well as human participants. Experiment 4 indicated that 2-D photography, which is usually considered to be a veridical method of record, can cause inaccurate size judgments under certain common conditions. In portraiture, it is likely that a model's directly seen jaw-neck ratio will be perceived as slimmer than in a conventional 2-D photographic image taken from the same viewpoint. The body-image distortion described here could be reduced by comparatively simple changes in 2-D imaging techniques. Some correction of the most common fattening effects can be achieved by using wide-angle lenses with carefully controlled subject proximity. However, only a well-designed stereoscopic or volumetric display can properly solve all of these problems.

We have also demonstrated that orthostereoscopic images can affect object ratio judgments in shape perception. In experiment 3, the circumference of the waist of the peanut shape was seen as 5.4% slimmer (when averaged across all participants) in the 65C condition than the waist of the synoptically viewed object. It should be noted, however, that the participant's view of the orthostereoscopic display geometry used in experiments 1 and 3 was not as well corrected as it could have been. Firstly, although the stereo transparency pairs were converged at the object's waist, it was not possible to actively adjust the vergence angles so that all of the other gaze points on the stimuli were seen as having zero disparity as they were viewed. In this respect, the display could not perfectly simulate direct viewing of a real object as a small amount of nonveridical vertical disparity was fused as the observers moved their gaze away from the center of interest. However, it was at the waist (the area of zero disparity) that the object ap-

Figure 18. *Three images (from a sequence of five) of two of the five photographic models used in experiment 4. The images on the left are extreme wide-angle photographs with a camera-to-subject distance of 0.32 m. The central images use a standard lens at 0.71 m. The images on the right were taken with a telephoto lens and a camera-to-subject distance of 2.7 m.*

peared to change shape. The background was perceived as flat throughout, even though the vertical disparity increased towards the image periphery. Incorrect vertical disparities generate pincushion (concave) or barrel (convex) distortion that would affect the perceived flatness of the background plane. As the background in

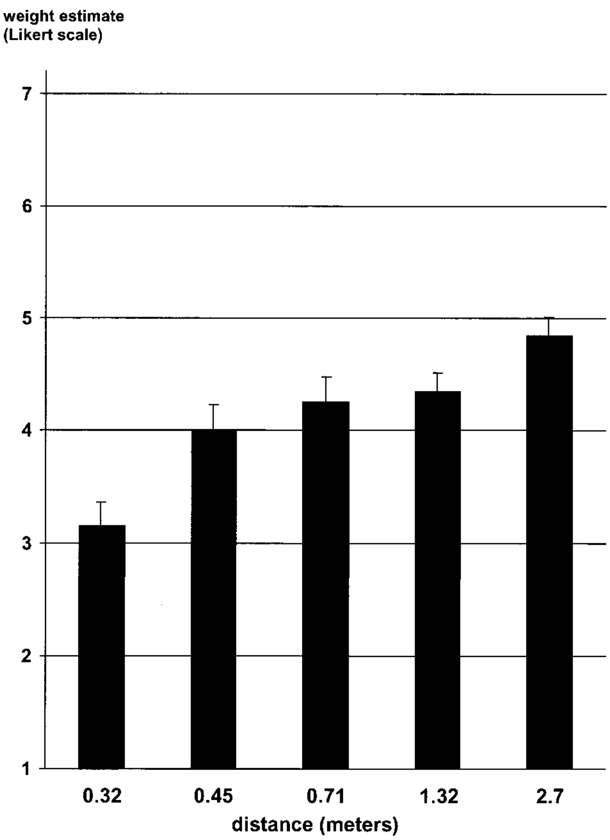


Figure 19. *The mean perceived bodyweight for the five different camera-to-subject distances (in meters), and therefore five different lens focal lengths, used to photograph the models in experiment 4.*

experiment 3 was perceived as flat, it can be inferred that the nonveridical vertical disparities did not generate obvious image artifacts. Secondly, any projected/reflected image system is likely to be compromised by the fact that the ideal viewing position (Koenderink, 1998) will occlude the projectors' optical paths. In these experiments, this was partially addressed by placing the viewing position between, but slightly below, each projector lens. The Wheatstone viewer used in experiment 2 resolves the occluded projection problem (and provides high-brightness images with zero cross-talk) but introduces others. The simple optics in this viewer are likely to induce slight curvature of field and resolution fall-off towards the edge of the image. A back-projected stereo display could, in theory, solve these problems and give a very high-brightness image. However, back projection tends to depolarize light, and, as yet, the materials required to manufacture a low cross-talk screen are currently not available. Despite these limitations, the results reported indicate that the orthostereoscopic technique used in these experiments appears to offer some advantages in veridical perception over 2-D representations of the same scene. Two-dimensional compatibility is another useful feature demonstrated by the orthostereoscopic display used in these experiments. Aligning the convergence to the point of zero disparity allows a viewer to see a single image at the center of interest in a scene without the need for polarizing glasses. This is especially true of scenes captured with low disparities.

We had expected, based on previous experience of stereoscopic displays, that some participants or experimenters would experience a degree of viewing discomfort during our experiments. However, in debriefing, no participant reported viewing discomfort in any of the experiments reported here and no experimenter experienced viewing discomfort despite very long exposure to the images. We therefore speculate that the polarized orthostereoscopic image could probably be viewed continuously for extended periods. Orthostereoscopic imaging may allow the muscles of the eyes to converge each optical axis in a natural and unstrained way. This is difficult with conventional stereography where the image separations at the screen plane require the eyes to "force fuse" two images, as if an object is at a closer position than would be the case with direct vision. Also, it can be seen in the "peanut experiment" (experiment 3, section 3.3) that shape perception may be more difficult in 65 mm parallel stereo image and causes more variation in shape matching than was found with the convergent orthostereoscopic images (figures 14b and c).

The analysis of the TNO stereo-acuity data also supports the view that the convergent images were easier to fuse for all the participants than were conventional parallel stereo images. The TNO stereo-acuity test uses a parallel stereo image-capture technique for its random-dot anaglyph plates. These anaglyph disparities are rendered to indicate the limit of a subject's ability to fuse red-green "double images." We had predicted that those participants who had above-average measured stereo acuity would perform consistently better in the size-matching task than would those with below-average stereo acuity. No such correlation was found. Participants who scored poorly on the TNO stereo test were able to easily fuse the stereo stimuli used in our experiments. Because the stimuli we used did not contain large disparities, this result suggests that the stereoscopic stimuli used in the TNO test differ in some important respects from orthostereoscopic images.

It is likely that most users of photography are unaware that it can produce distorted images in its normal modes of operation. Two-dimensional photography purports to be a truly representational medium. Yet, in common conditions, such as the imaging of people and close-up objects, it can be very misleading. It is reasonable to speculate that the "peanut" stimuli in experiment 3 correlates not only to the human female waist-hip ratio that it was designed to simulate, but also to the perceived jaw-neck ratio of both genders. This is because its waist design is similar to the way the human neck separates the head from the shoulders in males and females. It seems clear that a 2-D image of this geometry cannot accurately reproduce the information gathered with direct stereo vision from the same position. Thus, it can be inferred that the 2-D condition is almost always likely to distort when compared with an otherwise identical stereoscopic image. As parallel stereoscopic imaging seems to convey object information that causes more variation in the size-matching task, it appears that only an orthostereoscopic image can convey truly lifelike information (and therefore the presence) of objects and people.

Acknowledgments

The stimuli in Experiment 3 and all of the technical illustrations were designed and rendered by Philip Berridge of Liverpool University Department of Architecture and Building Engineering.

References

Banton, T., & Levi, D. M. (1991). Binocular summation in vernier acuity. *Journal of the Optical Society of America A, 8*(4), 376–380.

Blake, R., Sloane, M., & Fox, R. (1981). Further developments in binocular summation. *Perception & Psychophysics, 30*(3), 266–276.

Freeman, J., Avons, S. E., Meddis, R., & Pearson, D. E. (2000). Behavioral realism: Using postural responses to estimate presence. *Presence: Teleoperators and Virtual Environments, 9*(2), 149–164.

Freeman, J., Avons, S. E., Pearson, D., & IJsselsteijn, W. (1999). Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence: Teleoperators and Virtual Environments, 8*(1), 1–13.

Gunby, L. (2000, September 14). Notes & queries. *The Guardian, G2,* p. 16.

Hendrix, C., & Barfield, W. (1996). Presence within virtual environments as a function of visual display parameters. *Presence: Teleoperators and Virtual Environments, 5*(3), 274–289.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis.* Oxford: Oxford University Press.

IJsselsteijn, W., de Ridder, H., Hamberg, R., Bouwhuis, D., & Freeman, J. (1998). Perceived depth and the feeling of presence in 3DTV. *Displays, 18*(4), 207–214.

Jones, R. K., & Lee, D. N. (1981). Why two eyes are better than one. *Journal of Experimental Psychology: Human Perception & Performance, 7*(1), 30–40.

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago: University of Chicago Press.

Kelly, L. (1998, September 20). Q and A. *Sunday Mirror Personal Magazine,* p. 3.

Koenderink, J. J. (1998). Pictorial relief. *Philosophical Transactions of The Royal Society of London, 356*(1740), 1071–1086.

Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. L. (1994). On so-called paradoxical monocular stereoscopy. *Perception, 23*(5), 583–594.

Langford, M. (1989). *Advanced photography*. London: Focal Press.

Perrett, D., Harries, M., Mistlin, A. J., & Chitty, A. J. (1990). Three stages in the classification of body movements by visual neurons. In H. Barlow, C. Blakemore, & M. Weston-Smith (Eds.), *Images and understanding* (pp. 94–107). Cambridge, U.K.: Cambridge University Press.

Prinzmetal, W., & Gettleman, L. (1993). Vertical-horizontal illusion: One eye is better than two. *Perception & Psychophysics, 53*(1), 81–88.

Singh, D. (1993). Bodyshape and womens attractiveness: The critical role of waist-hip ratio. *Human Nature, 1*(3), 297–321.

Spottiswoode, R., Spottiswoode, N. L., & Smith, C. (1952, October). Basic principles of the three-dimensional film. *Journal of the Society of Motion Picture & Television Engineers, 59,* 249–285.

TNO Test for Stereoscopic Vision (10th ed.) (1972). Veenendaal: Laméris Ootech B.V.

Warner, M. (1995, Summer). Stealing souls and catching shadows. *tate: the art magazine, 6,* 40–46.

Yamanoue, H. (1997, April). The relationship between size distortion and shooting conditions for stereoscopic images. *Journal of the Society of Motion Picture and Television Engineers, 106,* 225–232.

Yamanoue, H., Okui, M., & Yuyama, I. (2000). A study on the relationship between shooting conditions and cardboard effect of stereoscopic images. *IEEE Transactions on Circuits and Systems for Video Technology, 10*(3), 411–416.