Supplemental Data

Humans ignore motion and stereo cues in favour of a fictional stable world

Andrew Glennerster, Lili Tcheang, Stuart J. Gilson,
Andrew W. Fitzgibbon, Andrew J. Parker

The effect of a floor or ceiling on size matches

The height of the eyes above the ground plane may, under some circumstances, become a significant cue to viewing distance [*1, 2*]. The data in Figure S1 show that this cue is not necessary in order for subjects to show large biases in size judgements in an expanding room. The floor and ceiling were removed so that subjects appeared to be in an infinite shaft or well. In other respects the experimental conditions were the same as the expanding room experiment (figures 1 and 2). Thus, the size of the well and the bricks on the wall varied according to the subject's lateral position as before and subjects made a similar judgement of the relative size of the comparison and standard cubes. Figure S1 shows size matches at three comparison distances. Removing the floor and ceiling has some effect on size matches at the closer distances for subject LT but none at the furthest distance or at any distance for subject HB. Distortions in size matches clearly remain in the absence of a ground plane.

Computational analysis of images seen by a subject in a typical trial

In principle, motion parallax can be used to perform size constancy if a proprioceptive signal such as stride length provides a scale reference. We used commercially available computer vision software (*boujou* from 2d3, [*3, 4*]) to show that the motion parallax information in the presented images is sufficiently rich to allow near-veridical performance.

Boujou takes as input a sequence of images captured by a camera moving through a rigid 3D scene, and computes a maximum-likelihood estimate of the 3D scene structure and the camera trajectory. This estimate is obtained by tracking 2D points in the input images, and finding the set of 3D camera positions and 3D points that provide the closest prediction of the 2D point tracks. From monocular data, the estimate is always up to an unknown overall scale factor. Figure S2 shows an example frame from a processed sequence, and the recovered 3D scene and camera path.

In order to achieve size constancy, the unknown scale factor must be computed. This scale factor can be recovered using proprioceptive information, for example a known length of a pace. We used the original camera trajectory from which the images were generated (derived from the InterSense tracker data) in the place of proprioceptive information and estimated the scale relating the original trajectory to that recovered by *boujou*. Specifically, for a sequence of images I_1 to I_n, we denote by $\{\mathbf{x}_i\}_{i=1}^n$ the corresponding 3D camera positions recovered by *boujou*. The positions given by the InterSense tracker are $\{\mathbf{y}_i\}_{i=1}^n$. For a sequence in which the room is constant size, the arbitrary reference frame in which the vision-based reconstruction is computed is related to the InterSense coordinate system by a 3D rotation, translation, and scale. We need only the scale factor, s, to estimate the size of objects in the scene. This is determined as the ratio of lengths $\|\mathbf{x}_i - \mathbf{x}_j\|/\|\mathbf{y}_i - \mathbf{y}_j\|$ between two time instants, i, j. We chose i, j from

$i, j = \mathrm{argmax}_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|$. Other strategies for estimating s produced similar results, such as a least-squares estimate based on registration of the entire 3D camera tracks.

Thus, the experimental procedure is duplicated as follows. The images in which the left cube is visible are presented to *boujou*, and a 3D reconstruction and camera trajectory computed. The scale of the reconstruction is set using the known stride-length information as described above. Two points on the front face of the cube are manually identified in 2D (see Figure S2), their 3D positions computed using *boujou*, and the distance between them, d_l, is recorded. Repeating the process for the images in which the right-hand cube is visible produces an estimate d_r. The ratio d_r/d_l is an estimate of the relative size of the two cubes according to *boujou*. The open symbols in Figure 2 show the 'matched size' predicted from this size ratio, $4d_l/d_r$, for three image sequences in which the comparison cube was at three different distances.

References and Notes

1. J. J. Gibson. *The perception of the visual world*. Boston: Houghton Mifflin, 1950.

2. T.L. Ooi, B. Wu, and Z.J. He. Distance determined by the angular declination below the horizon. *Nature*, 414:197–200, 2001.

3. 2d3 Ltd. Boujou 2, 2003. http://www.2d3.com.

4. A. W. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *LNCS 1406: Computer Vision—ECCV '98*, pages 311–326. Springer, 1998.

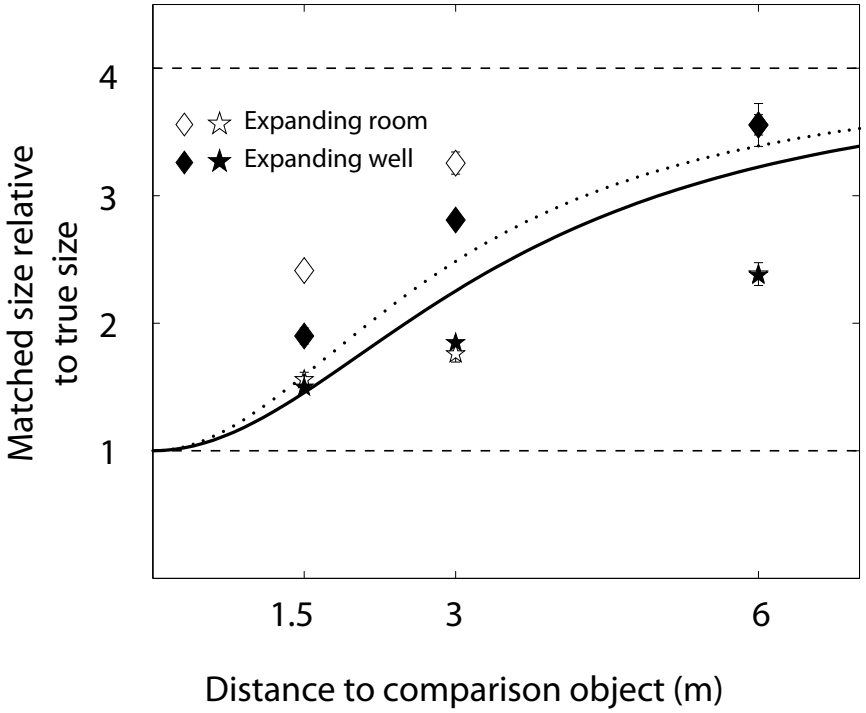

Figure S1: **Size matches without a ground plane.** Two subjects repeated the expanding room experiment without a floor or ceiling (symbol types for each subject as in Figure 1) but with the same texture on the walls. Subjects appeared to be suspended in an infinite shaft or well. The size of the well varied as the subject moved, as in the expanding room. Size matches were distorted by the expanding well (solid symbols) in a similar way to that found in the expanding room (open symbols, re-plotted from figure 2). The curves show the best fit of equation 5 to the mean data in each condition (expanding well, solid line; expanding room, dotted line). At a viewing distance of 6m the difference in performance was particularly small and the symbols overlie one another.

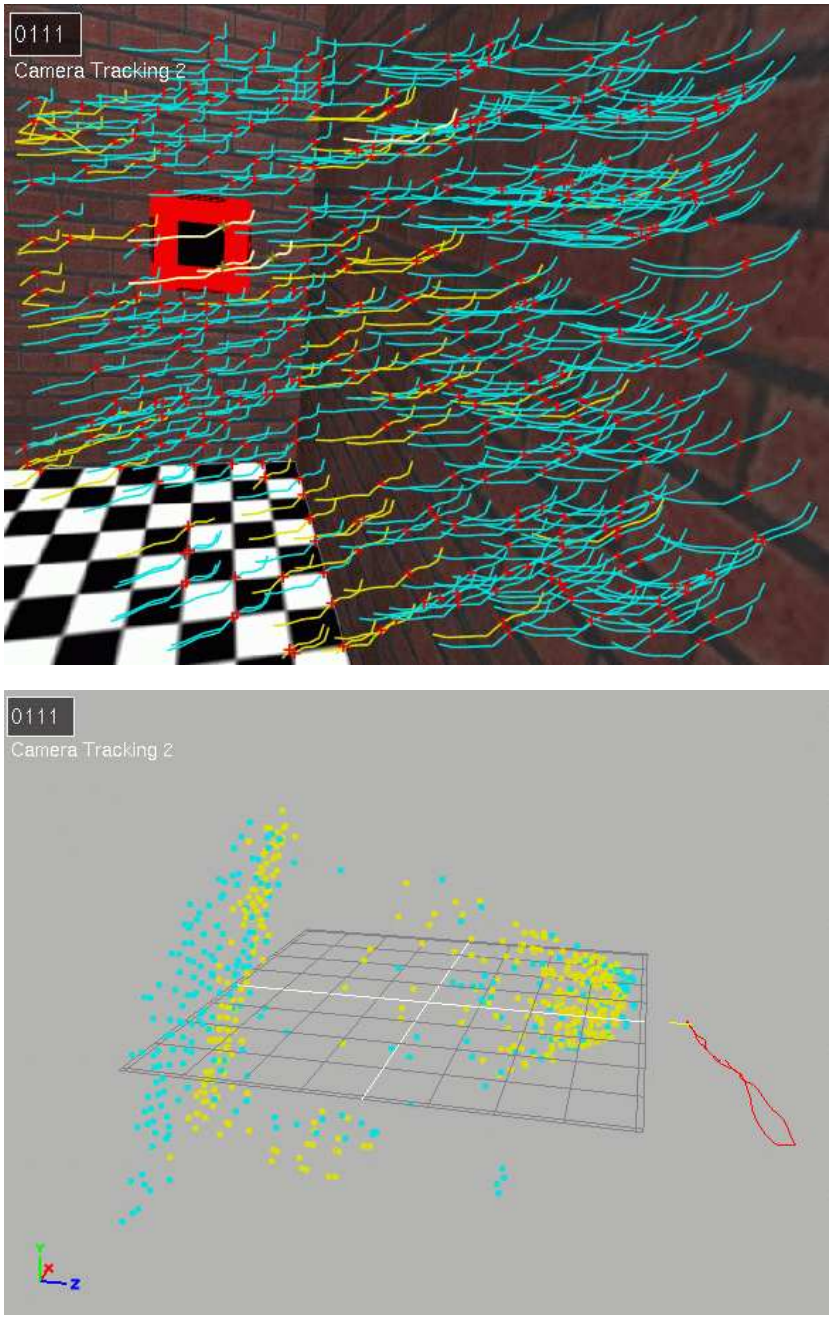

Figure S2: **Camera path and 3D structure recovered from a typical image sequence for one trial.** The top image illustrates one frame from the sequence, showing the virtual room and the comparison cube (with a superimposed black square for the purpose of feature tracking). The red crosses indicate the locations of tracked image features. The blue and yellow lines show the path of these features across sequential frames before and after the current frame. The image below shows a 3D view of the reconstructed location of the tracked features (blue and yellow points) and the computed path of the camera (red). The images used for this reconstruction were from the part of the trial in which the comparison cube was visible. The scene remained static throughout this period.