

RESEARCH ARTICLE

Seeing our 3D world while only viewing contour-drawings

Maddex Farshchi, Alexandra Kiba, Tadamasa Sawada ^{*}

School of Psychology, National Research University Higher School of Economics, Moscow, Russia

^{*} tsawada@hse.ru, tada.masa.sawada@gmail.com

Abstract

Artists can represent a 3D object by using only contours in a 2D drawing. Prior studies have shown that people can use such drawings to perceive 3D shapes reliably, but it is not clear how useful this kind of contour information actually is in a real dynamical scene in which people interact with objects. To address this issue, we developed an Augmented Reality (AR) device that can show a participant a contour-drawing or a grayscale-image of a real dynamical scene in an immersive manner. We compared the performance of people in a variety of run-of-the-mill tasks with both contour-drawings and grayscale-images under natural viewing conditions in three behavioral experiments. The results of these experiments showed that the people could perform almost equally well with both types of images. This contour information may be sufficient to provide the basis for our visual system to obtain *much* of the 3D information needed for successful visuomotor interactions in our everyday life.

 OPEN ACCESS

Citation: Farshchi M, Kiba A, Sawada T (2021) Seeing our 3D world while only viewing contour-drawings. PLoS ONE 16(1): e0242581. <https://doi.org/10.1371/journal.pone.0242581>

Editor: Markus Lappe, University of Muenster, GERMANY

Received: June 26, 2020

Accepted: November 4, 2020

Published: January 22, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0242581>

Copyright: © 2021 Farshchi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All empirical data reported in this study are available from <https://osf.io/t5jgb/>.

Funding: This article was prepared within the framework of the Academic Fund Program at the National Research University Higher School of

Introduction

Artists can represent a 3D scene with 3D objects by using only contours in a 2D contour-drawing and people can recognize the scene and objects reliably from such drawings [1–9]. There are computer vision algorithms that try to emulate this artists' skill and can generate contour-drawings from 2D photographic-images of 3D scenes [10, 11] and from 3D information contained in the scene [12, 13]. These contours represent an abrupt change of the luminance, color, or texture in the image and characteristic features in the 3D information. These characteristic features include self-occluding boundaries on the surface of objects [14–16], ridges on the surface [17, 18], as well as sharp edges on surfaces (see [12, 13] for reviews). Neither the luminance-polarity nor luminance-gradients are present in a contour-drawing.

Human beings can see the shape and position of a 3D object veridically when given only 2D drawings of it, and they can also recognize such objects reliably [1, 3–6, 19–27]. These well-known facts present a problem because according to Inverse-Problem Theory, the recovery of the shape of a 3D object from a 2D drawing is an ill-posed inverse problem. There are infinitely many 3D interpretations of a 2D contour-drawing. Note that line-drawings lack most depth cues, including binocular disparity, shading, and cast-shadows. This inverse problem can be resolved by imposing *a priori* constraints on the family of possible 3D interpretations [3, 5, 28–30]. Now, consider ordinary objects we see and use in our everyday life. Such objects are

Economics (HSE University) in 2019 (grant № 19-04-006, awarded to TS) and by the Russian Academic Excellence Project «5-100». <https://www.hse.ru/science/scifund/nug/nug2019> The sponsors or funders play no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

not composed of a random scattering of points. They can be characterized by regularities of their shape, for example, their symmetry, volume, the planarity of their contours, and the presence of rectangular corners. These regularities, which introduce specific features into the drawing, also could be used to detect the presence and shape of an object [30–33]. Our visual system could make use these regularities as the *a priori* constraints needed to recover the shape of a 3D object from a 2D contour-drawing of the object [34–38].

Prior studies that tested 3D perception from contour drawings have shown that people can obtain 3D information from the contour drawing, but it is not clear just how useful such contour information actually is in real dynamical 3D scenes in which ordinary objects are recognized and utilized under natural viewing conditions. These studies generated contour-drawings of objects taking care to avoid using degenerate views [5, 30], but note that in the real 3D scenes, objects will often be seen with degenerate views. Also note that people often change their viewing positions providing them with different views of the objects. Put simply, people can interact with the objects in real dynamical scenes.

Visual perception in real dynamical 3D scenes can be studied using the XR (Augmented-, Mixed-, and Virtual-Reality) technology. This XR technology can provide immersive experiences of a 3D scene that can be controlled by a computer. It has been shown that people can see the 3D information in the scene and they can interact within the scene on the basis of the visual information provided by the XR technology even if the scene is not fully photorealistic [39–41].

We developed an Augmented Reality (AR) device that can show a participant both a contour-drawing and a grayscale-image of a real dynamical 3D scene in an immersive manner (see [42–44] for earlier studies using AR devices to test the human visual system). The grayscale images were used as a control. They provided a baseline for the performance of a participant conducting our kind of tasks while wearing this device in our experiments. Our AR device allowed us to determine how well the participant can interact dynamically with objects in a scene, by using only contours in the contour-drawing or by using luminance-polarity and luminance-gradients in the gray-scale image.

General methods

AR device

The AR device used in this study showed, in an immersive manner, a contour-drawing and a grayscale image that represented a scene "out there" (Fig 1). The device was composed primarily of a smart phone (Lenovo Phab 2 Pro) and a wearable stereoscope (VR head-set). The phone ran on the Google Android OS which was equipped with an LCD screen and a camera located on the back of the screen. The two halves of the screen were seen individually by the two eyes of a participant who looked through the lenses of the stereoscope. The distance between the eyes and the screen, when the stereoscope was worn, was 8.5 cm. The screen's resolution was 1248×2560 pixels, and its size was 6.9×14.2 cm.

The phone's camera captured a photographic image of the scene in front of the participant. A contour-drawing and grayscale-image representing the scene were generated from this photographic image. This image was first converted to a grayscale-image I_G . Then, I_G was passed through a set of image filters to generate the contour-drawing I_C :

$$I_C = 0.5|(I_G * B) * S_V| + 0.5|(I_G * B) * S_H|$$

where B was a Gaussian filter and S_V and S_H were Sobel filters [10, 11, 45] that emphasized

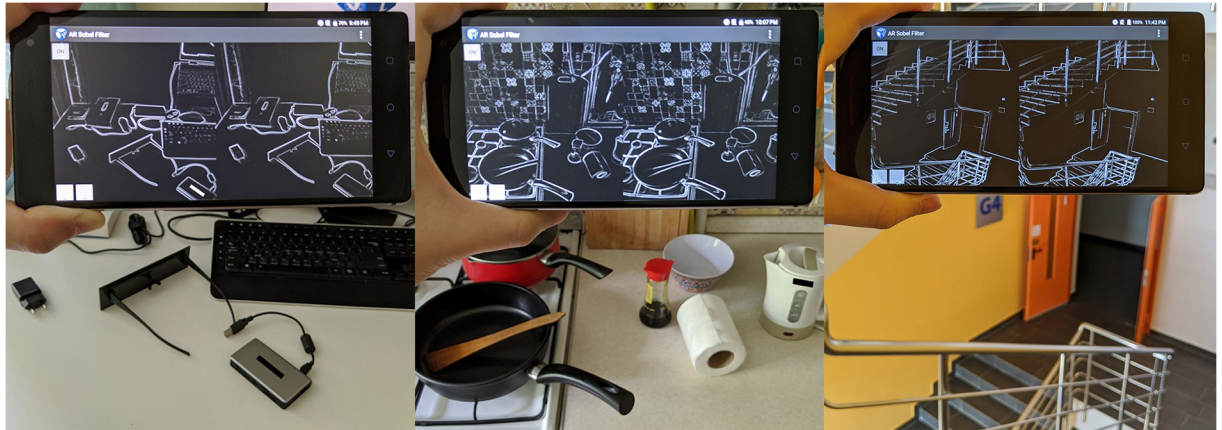


Fig 1. Contour-drawings of real scenes generated by the AR device used in this study. Copy-righted parts in the images have been blacked out.

<https://doi.org/10.1371/journal.pone.0242581.g001>

vertical and horizontal edges:

$$B = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

$$S_V = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}$$

$$S_H = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix}$$

Sobel filters were chosen because of their computational simplicity. This allowed our AR device to process the photographic images in near real-time. It is worth mentioning that an analogy of this algorithm with the visual system's process of edge detection in the primary visual cortex has been discussed [46]. This image process was implemented as an Android app using OpenCV library [47]. The resolution of the original photographic image and of the processed images was the same as the resolution of the screen (1248×2560 pixels).

Two image segments (1248×1280 pixels) taken from regions in the grayscale-image I_G or in the contour-drawing I_C were shown on the left and right halves of the screen (Fig 2). These regions were horizontal translations of one another in I_G and I_C . The size of the translation Δ_S could be adjusted to allow the participant to fuse the retinal images of the halves of the screen when the screen was viewed with the stereoscope. Note that the two halves of the screen were seen binocularly but binocular depth cues (binocular disparity and vergence) could not be used to perceive the 3D scene. These cues simply represented a frontoparallel plane but its effect on the immersive experience with the AR device seemed to be small [48].

A small wide-angle lens was attached to the camera to widen the camera's field of view. The visual angle of each image segment that was displayed on half of the screen was $53 \times 54^\circ$ from the camera when this lens was attached. The visual angle of each screen half was $58 \times 59^\circ$ from

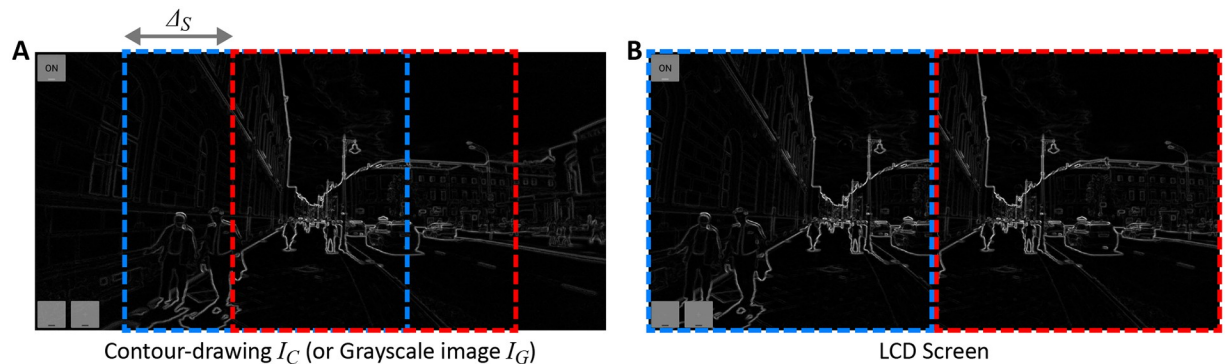


Fig 2. (A) Regions of two image segments (blue and red) taken from the contour-drawing I_C , or from the grayscale-image I_G , and (B) an image composed of these segments on the LCD screen. These regions are horizontal translations of one another for Δ_S in I_G and I_C .

<https://doi.org/10.1371/journal.pone.0242581.g002>

the eye of the participant. The refresh-rate of the screen was 10 Hz because of the time required to generate the grayscale-image and the contour-drawing. These processes also introduced a delay of 200–300 msec into the time required to refresh the screen. The refresh-rate (10 Hz) and the delay (200–300 msec) should be acceptable for an immersive experience when the AR device was used, but it could degrade a participants' interaction in the AR environment (see [49–51] for reviews; see the [Appendix](#)). Note that the contour-drawing was always generated regardless of whether the contour-drawing or the grayscale-image were shown on the screen. This made the refresh-rate and the delay constant with both types of images.

A shutter panel was also attached to the camera. This panel fully occluded the camera's view when it was closed. A clicking signal produced by pressing the middle button of a mouse was triggered when the shutter panel opened. This signal was used to signal the onset of a trial in the experiments.

Procedure

The experiments were conducted in a well-lit room. The participant's tasks were different from one another, but all of the tasks required interacting with objects on a desk in front of the participant. During these tasks, the participant sat on a chair in front of this desk and viewed the objects on it through our AR device. Note that the participant could only see objects that were relevant to the task at hand during each trial. All other objects were hidden from view.

The two conditions with different image filters were blocked within each experiment and the participant put on the AR device with one of the image filters before each block. The block started with a training phase during which the participant was asked to look at their hands and to examine the scene through the device for 1 minute. The participant got used to viewing through the filter in this adaptation phase. All of the objects relevant to the tasks used in the experiments were hidden from the participants during the adaptation phase.

Before each trial, the shutter panel of the AR device was closed to occlude the participant's view ([Fig 3](#)). The trial began by opening the panel. The participant was asked to finish a given task as soon as possible and to press a large green button on a wall in front of the participant when s/he was finished. The response time between opening the shutter panel and pressing the green button was recorded.

Participants were 36 undergraduate students (aged 18 or over) in the Department of Psychology at the National Research University Higher School of Economics. All had normal or corrected-to-normal vision. There were 12 participants in Experiment 1, 12 participants in Experiment 2, and 12 participants in Experiment 3 (see <https://osf.io/t5jgb/> for details). All

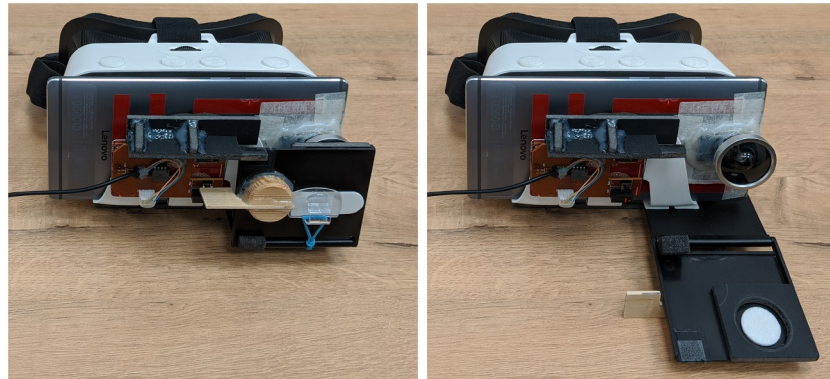


Fig 3. The AR device used in this study with its shutter panel closed (left) and opened (right).

<https://doi.org/10.1371/journal.pone.0242581.g003>

were naïve with respect to the purpose of the study. Written informed consents were obtained from all of the participants. They were compensated with 100 Rubles for their participation.

The experiments described were conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and approved by the institutional review board (the HSE Committee on Interuniversity Surveys and Ethical Assess of Empirical Research).

Experiment 1: Shape matching

In Experiment 1, we measured performance of a shape matching task with the image filters.

Procedures

A participant was given 12 prism-shaped objects that were randomly oriented on a tray and a box with 12 holes whose shapes corresponded to the 12 individual objects (Fig 4). The shapes of all of the holes were different from one another. They were the same as the shapes of the cross-sections of the individual objects, and the objects could go through only their corresponding holes. The participant had to insert all of the objects into the box by finding their unique holes. This task required matching the shapes of the objects with the shapes of the holes. Note that the participant could not see any of these objects before the first trial. The nature of this task was explained to the participant by using an analogous toy before the experiment. Put simply, this task required recognizing the objects' shapes and the shapes of their holes and then coordinating their relative positions and orientations.

The experiment had 2 blocks consisting of 3 trials. There were 2 groups of 6 participants. The first group ran the block with the contour-drawing filter first. This was followed by the block with the grayscale-image. The second group ran the blocks in the opposite order. The participants were asked to rest for 5 minutes between the blocks during which they did not wear the AR device.

Results

Fig 5 shows the averaged results observed in Experiment 1. The ordinate shows the response time. The abscissa shows the trials. The colors of the plots (blue and orange) represent the image filters (contour-drawing and grayscale image) and the styles of the plots represent the two groups of participants. The results were analyzed by using a three-way mixed-design ANOVA with repeated measures on two factors [52]: groups of participants, image filters, and

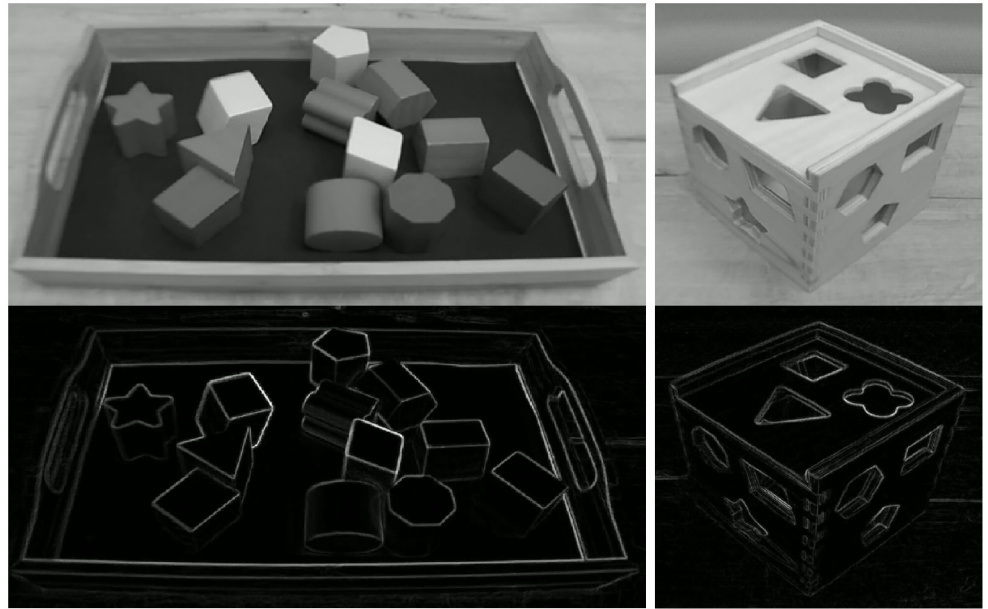


Fig 4. Gray-scale images and contour-drawings of 12 prism-shaped objects on a tray (left) and the box with 12 holes (right) used in Experiment 1. The shapes of the holes corresponded to the 12 individual objects.

<https://doi.org/10.1371/journal.pone.0242581.g004>

trial numbers in each block (1, 2, and 3). The effect of the trial numbers ($F_{2,50} = 9.0$, $p = 0.00046 \times 7$, where 7 is multiplied for a Bonferroni correction, see [53]) and the interaction between the filters and groups ($F_{1,50} = 27$, $p = 3.5 \times 10^{-6} \times 7$) were significant. The results of the other effects were not significant: the filters ($F_{1,50} = 0.31$, $p = 0.58 \times 7$), the groups ($F_{1,10} = 0.016$, $p = 0.97 \times 7$), the trial numbers \times filters ($F_{2,50} = 0.22$, $p = 0.80 \times 7$), the trial

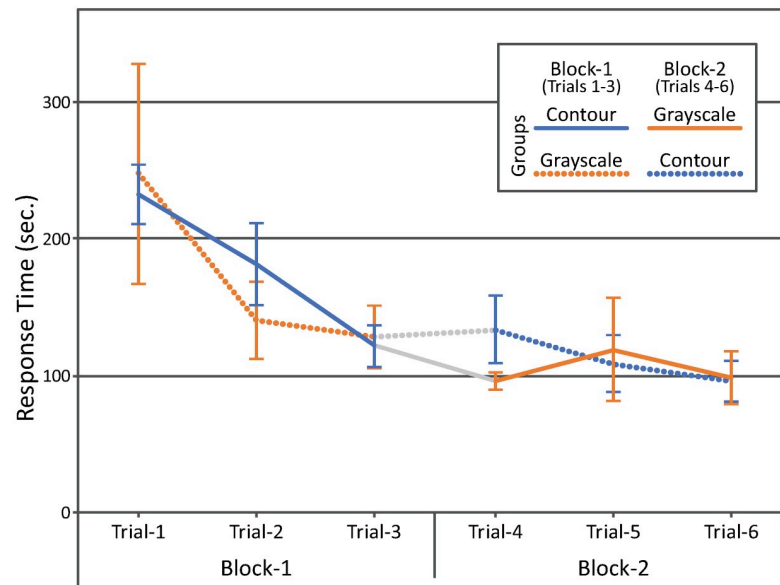


Fig 5. The results obtained in Experiment 1. The ordinate shows the response time and the abscissa shows the trials. The colors of the plots represent the image filters and the styles of the plots represent the groups of participants. The error bars show the standard errors across the participants. The 95 percent confidence intervals based on the t -distribution are 2.45 times of the standard errors ($CDF_t^{-1}(p = 0.975, n = 6) = 2.45$).

<https://doi.org/10.1371/journal.pone.0242581.g005>

numbers \times groups ($F_{2,50} = 1.4$, $p = 0.26 \times 7$), the trial numbers \times filters \times groups ($F_{2,50} = 5.4$, $p = 0.0075 \times 7$). (Note that p -values are adjusted to 1 if the p -values become larger than 1 after multiplying the Bonferroni factor.)

A posteriori test (Tukey) was performed to test the interaction between the filters and groups. The response time was shorter in the first block than in the second block: $p = 0.00091$ for the group that ran the block with the contour-drawing filter first and $p = 0.0095$ for the group that ran the block with the grayscale-image. The effect of the filter was significant neither in the first block ($p = 1.0$) nor in the second block ($p = 1.0$).

These results show that a human participant could conduct the shape matching task reliably with both a contour-drawing and a grayscale image that represented a scene "out there". We did not observe any difference in performance with both types of representation.

Experiment 2: Object recognition

In Experiment 2, we used our image filters to measure performance in an object recognition task.

Procedure

The participants sat in front of an open box and a collection of toys that were randomly oriented on a tray (Figs 6 and 7). These toys represented a variety of animals. The experimenter said the names of 3 target animals out loud, and the participant was asked to repeat these names while the shutter panel was kept closed to occlude the participant's view. When the panel was opened, the participant found the toys that represented 3 target animals, which had been chosen from the collection, and put them into the box. Put simply, this task only required that a participant could recognize an object s/he had not seen before.

This experiment had 2 blocks, each containing 4 trials. Four collections, consisting of a variable number of animals, were used (Figs 6 and 7). The order of the trials was determined by using the Latin-square method for each image filter. Ten animal toys were used in 3 of the 4 collections and 7 were used in the remaining collection. The target animals were chosen by using the following criteria: (i) no animal was used more than once throughout all blocks for each participant and (ii) individual animals in each collection were used at roughly equal frequencies throughout the experiment. (see <https://osf.io/t5jgb/> for a list of the target animals used in the experiment).

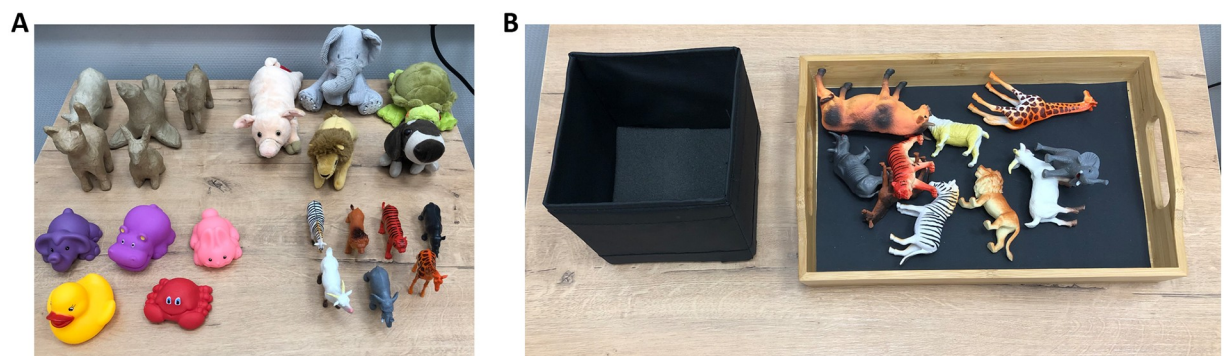


Fig 6. (A) 4 collections of animal toys used in Experiment 2: pulp-paper (left-top), stuffed (right-top), plastic-cartoon-like (left bottom), and plastic-realistic (right-bottom) animals. (B) The collection of toy animals on the tray, and the open box as they were arranged on the desk before each trial.

<https://doi.org/10.1371/journal.pone.0242581.g006>

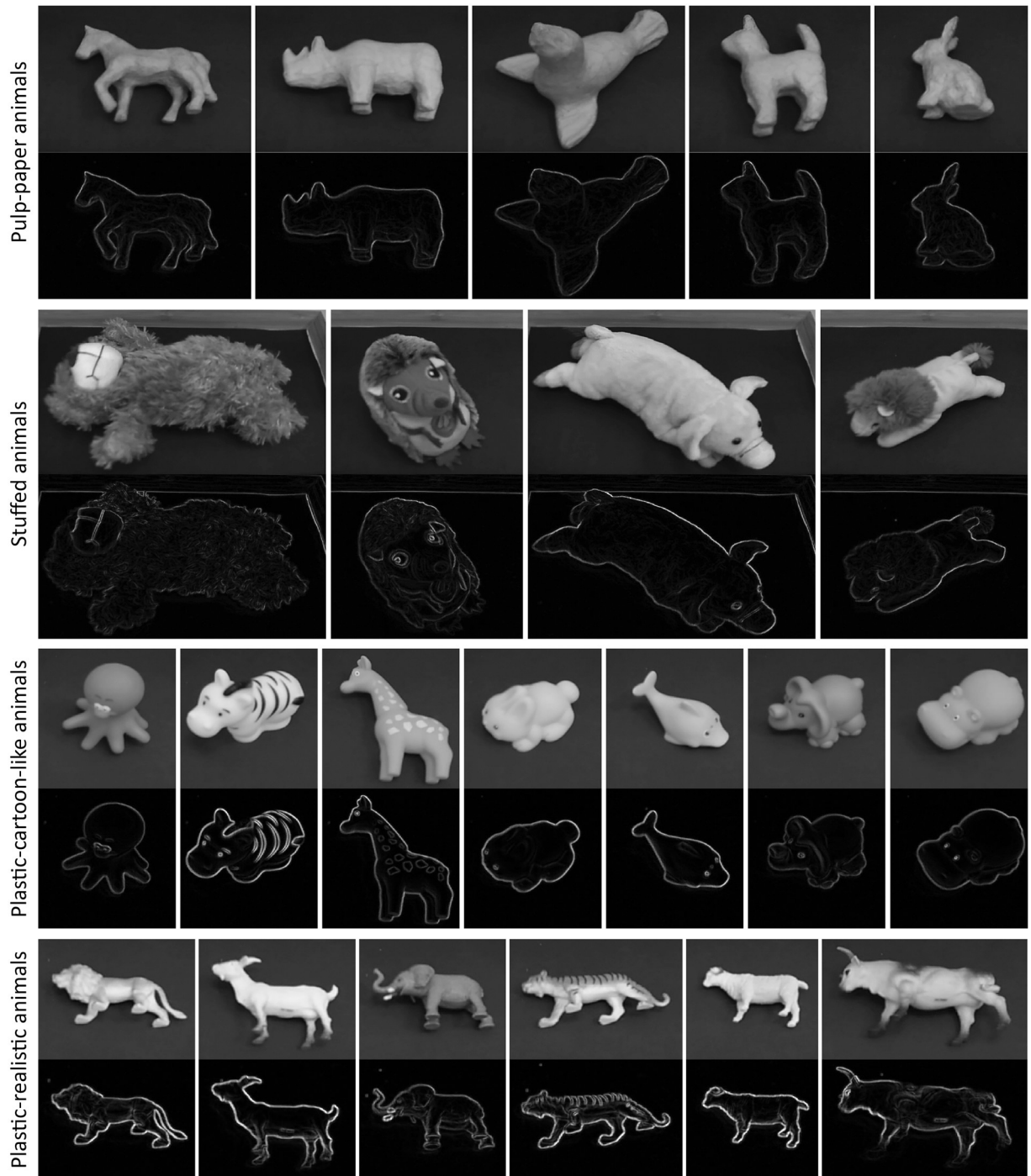


Fig 7. Gray-scale images and contour-drawings of several samples taken from the 4 collections of animal toys used in Experiment 2.

<https://doi.org/10.1371/journal.pone.0242581.g007>

There were 2 groups of 6 participants. The first group ran the block with the contour-drawing filter first. This was followed by the block with the grayscale-images. The second group ran the blocks in the opposite order. The participants rested for 5 minutes between blocks. They did not wear our AR device during these rest periods.

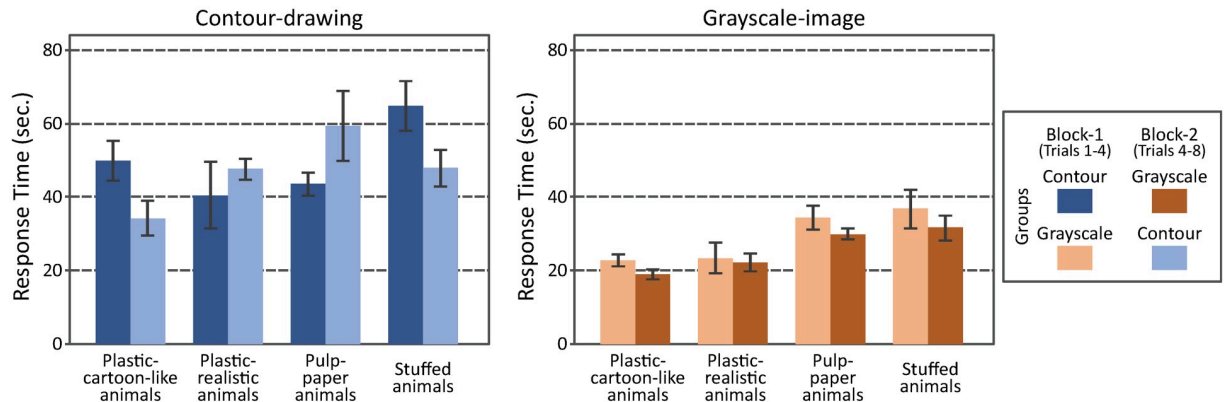


Fig 8. The two panels of bar-graphs show the results obtained with both kinds of filters, *i.e.*, contour-drawing and grayscale-image. The ordinate shows the response time. The abscissa shows the collections of toy animals. The brightness of the bars represents represent the two groups. The error bars show the standard errors across the participants. The 95 percent confidence intervals based on the t -distribution are 2.45 times of the standard errors ($CDF_t^{-1}(p = 0.975, n = 6) = 2.45$).

<https://doi.org/10.1371/journal.pone.0242581.g008>

Results

All participants recognized all of the objects. They made no errors in both image filter conditions, *i.e.*, contour-drawing and grayscale-image.

Fig 8 shows the averaged results observed in Experiment 2. The two panels of bar-graphs show the results obtained in both image filter conditions. The ordinate shows the response time. The abscissa shows the collections of animals. The brightness of the bars represents the two groups. The results were analyzed by using a three-way mixed-design ANOVA with repeated measures on two factors: groups of participants, image filters, and collections of toys. The two main factors were significant: the filters ($F_{1,70} = 75, p = 1.0 \times 10^{-12} \times 7$, where 7 is multiplied for a Bonferroni correction) and the toy collections ($F_{3,70} = 7.5, p = 0.00020 \times 7$). The results of the other effects were not significant: the groups ($F_{1,10} = 0.028, p = 0.87 \times 7$), the groups \times toy collections ($F_{3,70} = 2.7, p = 0.051 \times 7$), the groups \times filters ($F_{1,70} = 1.6, p = 0.21 \times 7$), the filters \times toy collections ($F_{3,70} = 0.062, p = 0.98 \times 7$), and the groups \times filters \times toy collections ($F_{3,70} = 3.2, p = 0.030 \times 7$).

A posteriori test (Tukey) was performed to test the effect of the toy collections. The response time was shorter with the plastic-cartoon-like animals than with the stuffed animals ($p = 0.00075$) and with the pulp-paper animals ($p = 0.018$). The response time was shorter with the plastic-realistic animals than with the stuffed animals ($p = 0.0047$). The other pair-wise comparisons were not significant: plastic-cartoon-like vs. plastic-realistic ($p = 0.94$), plastic-realistic vs. pulp-paper ($p = 0.074$), paper-realistic vs. stuffed ($p = 0.74$).

Experiment 3: Visuomotor coordination

In Experiment 3, we measured performance of a visuomotor interaction with objects in two tasks.

Procedure

The participants performed two visuomotor interactions, one with tongs (Fig 9); the second with a brick (Fig 10). In the tongs task, the participant picked up and moved 7 objects from a tray to an open box. The tongs were used to minimize haptic information being used to

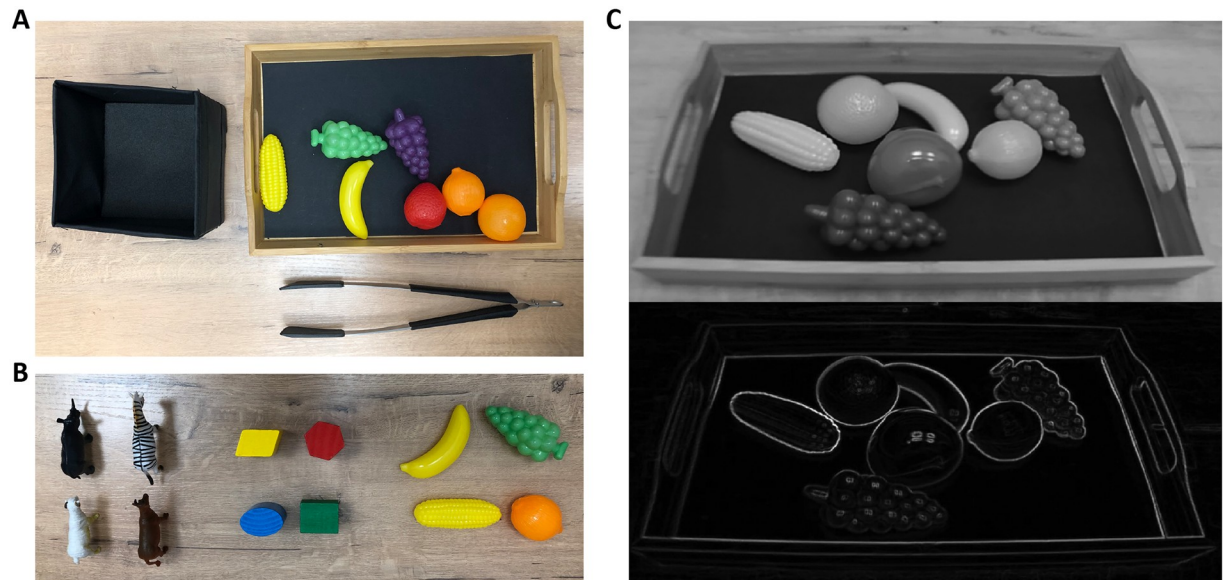


Fig 9. (A) A collection of 7 objects on a tray, and the open box as it was arranged on the desk before each trial of the tongs task in Experiment 3. (B) The 3 collections of objects used for the tongs task: plastic animals (left), wooden geometrical objects (middle), and plastic fruits-and-vegetables (right). (C) A gray-scale image and contour-drawing of the objects on the tray.

<https://doi.org/10.1371/journal.pone.0242581.g009>

perform the task. The participants had to control the tongs, using only dynamical visual information provided by the tongs and by the objects on the tray. In the brick task, the participant was given an open box containing 12 rectangular bricks and used his/her hands to build a stack, with 4 layers of 3 bricks on the top of a stand.

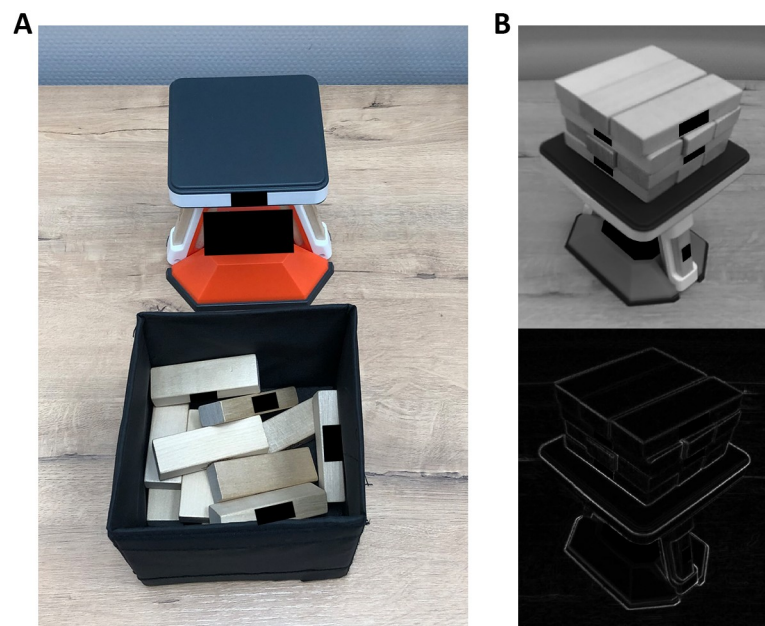


Fig 10. (A) The bricks and the stand as they were arranged on the desk before each trial of the brick task in Experiment 3. (B) A gray-scale image and contour-drawing of the 4 layers of 3 bricks built from the 12 bricks on the stand. Copy-righted parts in the images have been blacked out.

<https://doi.org/10.1371/journal.pone.0242581.g010>

The experiment had 4 blocks: 2 image filters (contour-drawing and grayscale image) \times 2 tasks (tongs and brick). Each block consisted of 3 trials. Three collections consisting 7 objects were used in the tongs task. These trials were repetitions of the blocks used in the brick task. The order of blocks and the order of trials within each block of the tongs task were randomized by using the Latin-square method with the following restriction: blocks with the contour-drawing and with the grayscale-image filters were conducted alternatively (see <https://osf.io/t5jgb/>). The participants rested for 5 minutes between blocks. They did not wear our AR device during these rest periods.

Results

Fig 11 shows the averaged results for the tongs and brick tasks observed in Experiment 3. The ordinate shows the response time. The abscissa in Fig 11A shows the trials within each block. The abscissa in Fig 11B shows the collections of objects. The colors of the plots (blue and orange) represent the image filters (contour-drawing and grayscale image).

The results of the tongs task were analyzed by using a two-way between-subject-design ANOVA with repeated measures: image filters, and collections of objects. The main factor, object collections, was significant ($F_{2,55} = 49, p = 6.4 \times 10^{-13} \times 3$, where 3 is multiplied for a Bonferroni correction). The effect of the filters and the interaction between the filters and object collections were not significant: the filters ($F_{1,55} = 0.52, p = 0.48 \times 3$), the filters \times object collections ($F_{2,55} = 0.18, p = 0.83 \times 3$). A posteriori test (Tukey) was performed to test the effect of the object collections. The response time was shorter with the plastic fruits-and-vegetables than it was with the plastic animals ($p = 1.2 \times 10^{-7}$) and with the wooden geometrical objects ($p = 4.1 \times 10^{-12}$). The response time was shorter with the plastic animals than with the wooden geometrical objects ($p = 0.0042$).

The results of the brick task were analyzed by using a two-way between-subject-design ANOVA with repeated measures: image filters, and trials. No effect was significant: the filters ($F_{1,55} = 0.0046, p = 0.95 \times 3$), the trials ($F_{2,55} = 3.2, p = 0.048 \times 3$), and their interaction ($F_{2,55} = 0.29, p = 0.75 \times 3$).

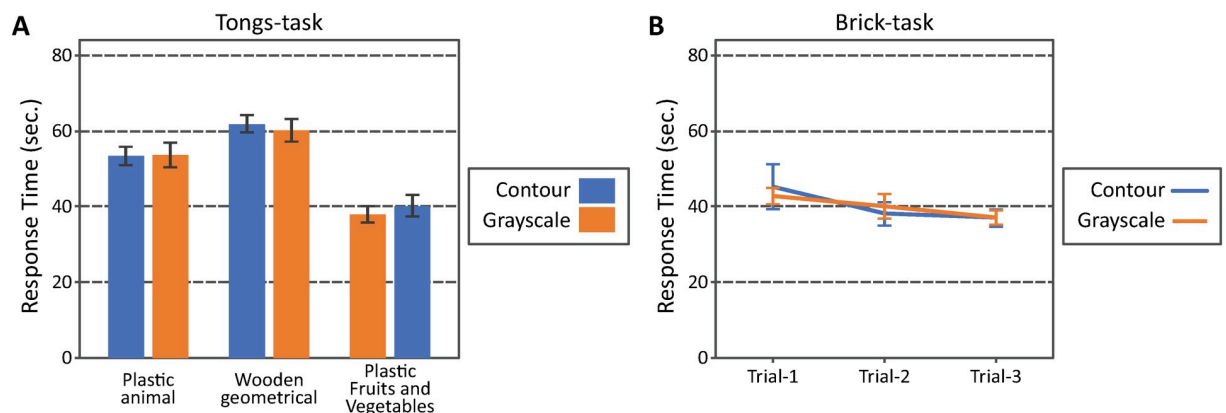


Fig 11. (A) The results of the tongs task obtained in Experiment 1. The ordinate shows the response time and the abscissa shows the collections of objects. The colors of the plots represent the image filters. The error bars show the standard errors across the participants. The 95 percent confidence intervals based on the t -distribution are 2.18 times of the standard errors ($CDF_t^{-1}(p = 0.975, n = 12) = 2.18$). (B) The results of the brick task obtained in Experiment 3. The ordinate shows the response time and the abscissa shows the trials. The colors of the plots represent the image filters. The error bars show the standard errors across the participants. The 95 percent confidence intervals based on the t -distribution are 2.18 times of the standard errors ($CDF_t^{-1}(p = 0.975, n = 12) = 2.18$).

<https://doi.org/10.1371/journal.pone.0242581.g011>

General discussion

We conducted three behavioral experiments that tested human performance when people interacted dynamically with objects in a real 3D scene, solely on the basis of a contour-drawing, or a grayscale-image that represented the scene. Contour-drawings were generated by applying an image filter to grayscale-images. The filter emphasized edges and eliminated luminance-gradients in the grayscale-image. The gray-scale images were used as a control. They provided a baseline for a participant's performance in our kind of tasks while wearing our AR device. The effect of the image filter was observed in the response time in an object recognition task (Experiment 2). Responses were slower with contour-drawings than with the grayscale-images, but note that participants could also recognize objects reliably from both contour-drawings and grayscale images. This difference in response time with these two types of images was not observed in our shape matching task (Experiment 1) or in our visuomotor coordination tasks (Experiment 3).

The tasks in Experiments 1 and 3 were designed so that the tasks required a dynamical visuomotor-coordination. The participant had to interact with multiple objects and to control their positions and orientations by using her/his hands. This kind of dynamical visuomotor-coordination is required in many run-of-the-mill tasks in which we use our hands in our everyday life. Note, however, that the tasks in this study did not require precise control of timing or quick reactions to unexpected events that could happen in a real-life scene. This kind of highly dynamical task could not be tested in this study because of technical limitations of our AR device. Note that such rapid processing of visual information is required in sports, and that it has been shown that the human visual system can process static contour-drawings very quickly [3, 6, 9]. So, it is possible that contour information is essential for visual processing in highly-dynamical tasks.

The results of Experiment 2 showed that a participant can recognize an object quite well when given a contour-drawing but recognition was even better when given its grayscale-image. This difference in performance can be attributed: (i) to the image filter used to generate the contour-drawing, and (ii) to the luminance-polarity and the luminance-gradient that are present in the grayscale-image but not in the contour-drawing. The image filter used a very simple algorithm that only emphasized the luminance edges on the basis of local information in the photographic image. This filter missed information in 2D images of a scene, such as edges between two isoluminant regions and between two regions that had different textures [54]. The filter could also miss contours that represent important features of the 3D information in a scene, for example, ridges on the surface of objects [17, 18]. This filter could also detect edges that are usually not drawn as contours in drawings made by an artist, for example, the edge of a shadow [55], and details of texture. These missed and redundant edges can degrade performance in an object recognition task [32, 56–59]. Note that the algorithm of the filter can be analogous to the visual system's process of edge detection in the primary visual cortex [46]. The human visual system must organize the edges in a retinal image in such a way that makes it possible for an observer to perceive the 3D information in the scene veridically [32, 58].

Also note that the luminance-polarity and the luminance-gradient present in the grayscale-image, but not in the contour-drawing, could also explain the difference in performance observed in Experiment 2. These two types of luminance information helped the visual system organize the luminance distributions and edges in the grayscale-images [42, 60–64]. Also, the luminance-gradient called "shading" could help the visual system to perceive the shape of the object's surface [18, 25, 27, 65–72] and to decompose the object on the basis of the surface's shape [73]. Note that this kind of luminance information did not improve performance in

Experiments 1 and 3, which suggests that the luminance-gradient and the luminance-polarity are not as important as the contours produced by luminance-edges in the retinal image are for perceiving 3D information in a real scene.

It is well-known that people can obtain 3D information from a contour drawing, but, until now, it was not clear how useful contour information actually is in a real dynamical scene. Our study shows that contour information, alone, is sufficient for ordinary people to perform a variety of run-of-the-mill tasks. We believe that our demonstration suggests that contour information, alone, may be sufficient to provide the basis for our visual system to obtain much of the 3D information needed for successful visuomotor interactions in our everyday life.

Appendix

The AR device, used in this study, introduced some technical limitations in what the participants could do in Experiments 1, 2, and 3: specifically, the refresh-rate (10 Hz), the delay (200–300 msec), and the visual angles [49–51]. Furthermore, the Images shown on the screen of the device were achromatic, and binocular depth cues (binocular disparity and vergence) could not be used to perceive the 3D scenes. Also, the participants could move their heads less freely when they wore the AR device on their heads. All of these factors could degrade the participants' performance even in the baseline condition with the grayscale-images. Two of the authors (MF, TS) performed these tasks without wearing the AR device to get an idea about how difficult these tasks were under more natural conditions. Specifically, MF and TS ran sessions in Experiments 1, 2, and 3 without wearing the AR device. Their response times in these sessions are summarized in Table 1 (See Figs 5, 8 and 11 for comparison).

The response times of MF and TS were analogous to one another in the individual sessions and they were substantially shorter than the response times of the naïve participants even with the grayscale-image filter in Experiments 1, 2, and 3. It suggests that the participants'

Table 1. Response times (sec) of MF and TS, under a more natural viewing condition, measured in sessions run when the AR device was not worn. The response time with the two image filters in Experiments 1, 2, and 3 are also shown for comparison (average \pm standard error).

Exp. 1	Trial-1	Trial-2	Trial-3	Trial-4	Trial-5	Trial-6
MF	56.3	46.2	42.8	38.4	37.7	53.1
TS	56.0	39.8	39.1	40.1	37.5	36.0
Contour	232.4 \pm 22.7	181.6 \pm 30.7	121.7 \pm 16.3	133.7 \pm 25.6	108.7 \pm 21.6	96.3 \pm 15.8
Grayscale	248.9 \pm 81.5	140.9 \pm 29.1	128.8 \pm 23.8	95.7 \pm 7.5	118.8 \pm 38.6	98.4 \pm 19.8
Exp. 2	Plastic-cartoon-like animals	Plastic-realistic animals	Pulp-paper animals	Stuffed animals		
MF	6.7	8.6	7.6	7.5		
TS	7.2	5.9	11.0	7.8		
Contour	42.1 \pm 4.4	44.0 \pm 4.8	51.5 \pm 5.5	56.4 \pm 4.9		
Grayscale	20.9 \pm 1.3	22.7 \pm 2.5	32.2 \pm 2.0	34.2 \pm 3.3		
Exp. 3: Tongs-task	Plastic animal	Wooden geometrical objects	Plastic fruits and vegetables			
MF	12.5	13.6	12.6			
TS	13.0	19.5	14.8			
Contour	53.3 \pm 2.7	61.9 \pm 2.4	40.2 \pm 3.1			
Grayscale	53.6 \pm 3.5	60.1 \pm 3.3	37.9 \pm 2.4			
Exp. 3: Brick-task	Trial-1	Trial-2	Trial-3			
MF	21.1	16.5	19.3			
TS	16.1	16.3	16.4			
Contour	45.2 \pm 6.3	38.1 \pm 3.4	37.1 \pm 2.5			
Grayscale	42.8 \pm 2.5	40.1 \pm 3.6	37.1 \pm 2.2			

<https://doi.org/10.1371/journal.pone.0242581.t001>

performance with both types of image filters (contour-drawing and grayscale image) was suppressed by the loss of color information and as well as by technical factors in the AR device used in the experiments. These technical factors should be minimized to make the viewing condition of the experiments more natural. This will be addressed in a future study.

Acknowledgments

We thank Svetlana V. Salomasova for helping to run the experiments reported in this study.

Author Contributions

Conceptualization: Alexandra Kiba, Tadamasawa Sawada.

Data curation: Maddex Farshchi.

Formal analysis: Maddex Farshchi, Tadamasawa Sawada.

Funding acquisition: Tadamasawa Sawada.

Investigation: Maddex Farshchi, Tadamasawa Sawada.

Methodology: Maddex Farshchi, Alexandra Kiba, Tadamasawa Sawada.

Project administration: Tadamasawa Sawada.

Resources: Tadamasawa Sawada.

Software: Alexandra Kiba.

Supervision: Tadamasawa Sawada.

Validation: Tadamasawa Sawada.

Visualization: Maddex Farshchi, Alexandra Kiba, Tadamasawa Sawada.

Writing – original draft: Maddex Farshchi, Tadamasawa Sawada.

Writing – review & editing: Maddex Farshchi, Alexandra Kiba, Tadamasawa Sawada.

References

1. Cole F, Sanik K, DeCarlo D, Finkelstein A, Funkhouser T, Rusinkiewicz S, et al. How well do line drawings depict shape? *ACM Trans. Graph.* 2009; 28(3), 28.
2. Hertzmann A. Why do line drawings work? A realism hypothesis. *Percept.* 2020; 49(4): 439–451. <https://doi.org/10.1177/0301006620908207>
3. Kroll JF, Potter MC. Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *J Verbal Learning Verbal Behav.* 1984 Feb 1; 23(1):39–66. [https://doi.org/10.1016/S0022-5371\(84\)90499-7](https://doi.org/10.1016/S0022-5371(84)90499-7).
4. Pizlo Z. 3D shape: Its unique place in visual perception. Cambridge, MA: MIT Press; 2008.
5. Pizlo Z, Li Y, Sawada T, Steinman RM. Making a machine that sees like us. New York, NY: Oxford University Press; 2014.
6. Potter MC, Faulconer BA. Time to understand pictures and words. *Nature.* 1975 Feb 6; 253(5491):437–8. <https://doi.org/10.1038/253437a0>. PMID: 1110787
7. Sayim B, Cavanagh P. What line drawings reveal about the visual brain. *Front Hum Neurosci.* 2011; 5: 118. <https://doi.org/10.3389/fnhum.2011.00118> PMID: 22065509
8. Walther DB, Chai B, Caddigan E, Beck DM, Fei-Fei L. Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proc Natl Acad Sci.* 2011 Jun 7; 108(23):9661–6. <https://doi.org/10.1073/pnas.1015666108>. PMID: 21593417
9. Walther DB, Shen D. Nonaccidental properties underlie human categorization of complex natural scenes. *Psychol Sci.* 2014 Apr; 25(4):851–60. <https://doi.org/10.1177/0956797613512662>. PMID: 24474725

10. Bhardwaj S, Mittal A. A survey on various edge detector techniques. *Procedia Technology*. 2012 Jan 1; 4:220–226.
11. Spontón H, Cardelino J. A review of classic edge detectors. *IPOLE*. 2015 Jun; 5:90–123. <https://doi.org/10.5201/ipol.2015.35>.
12. DeCarlo D. Depicting 3d shape using lines. *Proc SPIE*. 2012; 8291 (Hum Vis Electron Imag XVII): 361–376.
13. Cole F, Golovinskiy A, Limpaecher A, Barros HS, Finkelstein A, Funkhouser T, et al. Where do people draw lines? *ACM Trans. Graph*. 2003; 27(3): 88. <https://doi.org/10.1145/1360612.1360687>
14. DeCarlo D, Finkelstein A, Rusinkiewicz S, Santella A. Suggestive Contours for Conveying Shape. *ACM Trans. Graph*. 2003; 22(3): 848–855.
15. Koenderink JJ, Van Doorn AJ. The shape of smooth objects and the way contours end. *Percept*. 1982; 11: 129–137. <https://doi.org/10.1068/p110129> PMID: 7155766
16. Koenderink JJ. What does the occluding contour tell us about solid shape. *Percept*. 1984; 13: 321–330. <https://doi.org/10.1068/p130321> PMID: 6514517
17. Judd T, Durand F, Adelson E. Apparent Ridges for Line Drawing. *ACM Trans. Graph*. 2007; 26(3): 19. <http://doi.acm.org/10.1145/1276377.1276401>
18. Todd JT. The visual perception of 3D shape. *Trends Cog Sci*. 2004; 8(3): 115–121. <https://doi.org/10.1016/j.tics.2004.01.006> PMID: 15301751
19. Attneave F. Some informational aspects of visual perception. *Psychol Rev*. 1954; 61: 183–193. <https://doi.org/10.1037/h0054663> PMID: 13167245
20. Biederman I. Recognition-by-components: A theory of human image understanding. *Psychol Rev*. 1987; 94: 115–147. <https://doi.org/10.1037/0033-295X.94.2.115> PMID: 3575582
21. Biederman I. Recognizing depth-rotated objects: A review of recent research and theory. *Spat Vis*. 2000; 13(2–3):241–53. <https://doi.org/10.1163/156856800741063>. PMID: 11198235
22. Hochberg J, Brooks V. Pictorial recognition as an unlearned ability: A study of one child's performance. *Am J Psychol*. 1962; 75(4): 624–628. PMID: 13963734
23. Jahoda G, Deregowski JB, Ampene E, Williams N. Pictorial recognition as an unlearned ability: A replication with children from pictorially deprived environments. In Butterworth G, editor. *The Child's Representation of the World*. New York, NY: Plenum Press, 1977. p. 203–217.
24. Kennedy JM, Ross AS. Outline picture perception by the Songe of Papua. *Percept*. 1975; 4: 391–406.
25. Koenderink JJ, van Doorn AJ, Christou C, Lappin JS. Shape constancy in pictorial relief. *Percept*. 1996; 25: 155–164. <https://doi.org/10.1068/p250155> PMID: 8733144
26. Li Y, Pizlo Z. Depth cues versus the simplicity principle in 3D shape perception. *Top Cogn Sci*. 2011 Oct; 3(4):667–85. <https://doi.org/10.1111/j.1756-8765.2011.01155.x>. PMID: 25164504
27. Tsuruhara A, Sawada T, Kanazawa S, Yamaguchi MK, Corrow S, Yonas A. The development of the ability of infants to utilize static cues to create and access representations of object shape. *J Vis*. 2010; 10(12): 2. <https://doi.org/10.1167/10.12.2> PMID: 21047734
28. Pizlo Z. Perception viewed as an inverse problem. *Vis Res*. 2001; 41: 3145–3161. [https://doi.org/10.1016/S0042-6989\(01\)00173-0](https://doi.org/10.1016/S0042-6989(01)00173-0) PMID: 11711140
29. Poggio T, Torre V, Koch C. Computational vision and regularization theory. *Nature*. 1985; 317: 314–319. <https://doi.org/10.1038/317314a0> PMID: 2413361
30. Sawada T, Li Y, Pizlo Z. Shape Perception. In Busemeyer J, Townsend J, Wang ZJ, Eidels A, editors. *Oxford Handbook of Computational and Mathematical Psychology*. New York, NY: Oxford University Press, 2015. p.255–276.
31. Leeuwenberg E, van der Helm PA. *Structural Information Theory: The Simplicity of Visual Form*. New York, NY: Cambridge University Press; 2013.
32. Michaux V, Jayadevan V, Delp E, Pizlo Z. Figure-ground organization based on 3D symmetry. *J. Electron. Imaging*. 2016; 25(6): 061606. <https://doi.org/10.1117/1.JEI.25.6.061606>
33. Perkins DN. (1972). Visual discrimination between rectangular and nonrectangular parallelepipeds. *Percept Psychophys*. 1972; 12(5): 396–400
34. Jayadevan V, Sawada T, Delp E, Pizlo Z. Perception of 3D symmetrical and nearly symmetrical shapes. *Symmetry*. 2018; 10(8): 344. <https://doi.org/10.3390/sym10080344>
35. Li Y. Perception of Parallelepipeds: Perkins's Law. *Percept*. 2009; 38: 1767–1781. <https://doi.org/10.1068/p6397> PMID: 20192127
36. Li Y, Pizlo Z, Steinman RM. A computational model that recovers the 3D shape of an object from a single 2D retinal representation. *Vis. Res*. 2009; 49: 979–91. <https://doi.org/10.1016/j.visres.2008.05.013> PMID: 18621410

37. Li Y, Sawada T, Shi Y, Kwon TK, Pizlo Z. A Bayesian model of binocular perception of 3D mirror symmetrical polyhedra. *J. Vis.* 2011; 11(4): 11. <https://doi.org/10.1167/11.4.11> PMID: 21505106
38. Perkins DN. How good a bet is good form? *Percept.* 1976; 5(4): 393–406.
39. de Gelder B, Kätsyri J, de Borst AW. Virtual reality and the new psychophysics. *Br J Psychol.* 2018 Aug; 109(3):421–426. <https://doi.org/10.1111/bjop.12308>. PMID: 29806694
40. Scarfe P, Glennerster A. Using high-fidelity virtual reality to study perception in freely moving observers. *J Vis.* 2015; 15(9):3. <https://doi.org/10.1167/15.9.3>. PMID: 26161632
41. Triesch J, Ballard DH, Hayhoe MM, Sullivan BT. What you see is what you need. *J Vis.* 2003; 3(1):86–94. <https://doi.org/10.1167/3.1.9>. PMID: 12678628
42. Anstis S. Visual adaptation to a negative, brightness-reversed world: Some preliminary observations. In Carpenter GA, Grossberg S, editors. *Neural Networks for Vision and Image Processing*. Cambridge, MA: MIT Press, 1992. p. 1–14.
43. Grush R, Jaswal L, Knoepfler J, Brovold A. Visual adaptation to a remapped spectrum. In Metzinger T, Windt JM, editors. *Open MIND: 16(T)*. Frankfurt am Main, Germany: MIND Group, 2015. <https://doi.org/10.15502/9783958570283> PMID: 26664821
44. Cohen MA, Botch TL, Robertson CE. The limits of color awareness during active, real-world vision. *Proc Natl Acad Sci USA.* 2020; 117(24): 13821–13827. <https://doi.org/10.1073/pnas.1922294117> PMID: 32513698
45. Sobel I. History and Definition of the so-called "Sobel Operator", more appropriately named the Sobel-Feldman Operator. Research gate. 2014 February 2 [Cited 2020 June 20] https://www.researchgate.net/publication/239398674_An_Isotropic_3x3_Image_Gradient_Operator
46. Wu Q, McGinnity M, Maguire L, Belatreche A, Glackin B. Edge detection based on spiking neural network model. *ICIC 2007, LNAI 4682*. 2007 Aug 21; 26–34.
47. Bradski G. *The OpenCV Library*. Dr Dobb's J. Software Tools. 2000.
48. Wijntjes M, Füzü A, Verheij MES, Deetman T, Pont SC. The synoptic art experience. *Art Percept.* 2016; 4(1–2): 73–105. <https://doi.org/10.1163/22134913-00002046>
49. Chen JY, Thropp JE. Review of low frame rate effects on human performance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans.* 2007 Oct 29; 37(6):1063–76.
50. Cummings JJ, Bailenson JN. How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology.* 2016 Apr 2; 19(2):272–309.
51. Thropp JE, Chen JY. The effects of slow frame rates on human performance. Aberdeen Proving Ground, MD: Army Research Laboratory; 2006 Jun.
52. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. *Applied Linear Statistical Models*. 4th ed, Boston, MA: McGraw-Hill; 1996. PMID: 8807197
53. Cramer AOJ, van Ravenzwaaij D, Matzke D, Steingroever H, Wetzels R, Grasman RPPP, et al. Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychon Bull Rev.* 2016; 23: 640–647. <https://doi.org/10.3758/s13423-015-0913-5> PMID: 26374437
54. Rosenholtz R. Texture perception. In Wagemans J, editor. *Oxford Handbook of Perceptual Organization*. New York, NY: Oxford University Press, 2015. p.167–186. <https://doi.org/10.1167/15.3.9> PMID: 25780063
55. Metzger W. *Laws of Seeing* (Spillman L, Lehar S, Stromeyer M, Wertheimer M, translators). Cambridge, MA: the MIT press; 2006.
56. Harrison SJ, Feldman J. The influence of shape and skeletal axis structure on texture perception. *J Vis.* 2009; 9(6): 13. <https://doi.org/10.1167/9.6.13> PMID: 19761304
57. Kwon TK, Agrawal K, Li Y, Pizlo Z. Spatially-global integration of closed, fragmented contours by finding the shortest-path in a log-polar representation. *Vis Res.* 2015; 125: 143–163.
58. Li Y, Sawada T, Latecki LJ, Steinman RM, Pizlo Z. A tutorial explaining a machine vision model that emulates human performance when it recovers natural 3D scenes from 2D images. *J Math Psychol.* 2012; 56: 217–231.
59. Sassi M, Vancleef K, Machilsen B, Panis S, Wagemans J. Identification of everyday objects on the basis of Gaborized outline versions. *i-Percept.* 2010; 1(3): 121–142. <https://doi.org/10.1068/i0384> PMID: 23145218
60. Elder JH, Trithart S, Pintilie G, MacLean D. Rapid processing of cast and attached shadows. *Percept.* 2004; 33(11): 1319–1338. <https://doi.org/10.1068/p5323> PMID: 15693674
61. Ghose T, Palmer SE. Extremal edges versus other principles of figure-ground organization. *J Vis.* 2010; 10(8): 3. <https://doi.org/10.1167/10.8.3> PMID: 20884578

62. Kersten D, Knill DC, Mamassian P, Bülthoff I. Illusory motion from shadows. *Nature*. 1996; 379: 31. <https://doi.org/10.1038/379031a0> PMID: 8538738
63. Lauenstein L. Über räumliche wirkungen von licht und schatten. *Psychol Forsch*. 1938; 22: 267–319.
64. Ramachandran VS. Perceiving shape from shading. *Sci Am*. 1977; 256(8): 76–83.
65. Kunsberg B, Holtmann-Rice D, Alexander E, Cholewiak S, Fleming R, Zucker SW. Colour, contours, shading and shape: flow interactions reveal anchor neighbourhoods. *Interface Focus*. 2018; 820180019. <http://doi.org/10.1098/rsfs.2018.0019> PMID: 29951196
66. Langer MS, Bülthoff HH. Depth discrimination from shading under diffuse lighting. *Percept*. 2000; 29: 649–660. <https://doi.org/10.1068/p3060> PMID: 11040949
67. Liu B, Todd JT. Perceptual biases in the interpretation of 3D shape from shading. *Vis Res*. 2004; 44: 2135–2145. <https://doi.org/10.1016/j.visres.2004.03.024> PMID: 15183680
68. Mingolla E, Todd JT. Perception of solid shape from shading. *Biol Cybern*. 1986; 53: 137–151. <https://doi.org/10.1007/BF00342882> PMID: 3947683
69. Nasanen R. Spatial frequenc bandwidth used in the recognition of facial images. *Vis Res*. 1999; 39: 3824–3833. [https://doi.org/10.1016/s0042-6989\(99\)00096-6](https://doi.org/10.1016/s0042-6989(99)00096-6) PMID: 10748918
70. Norman JF, Todd JT, Phillips F. The perception of surface orientation from multiple sources of optical information. 1995; 57: 629–636.
71. Pentland A. Shape information from shading: A theory about human perception. *Spat Vis*. 1989; 4: 165–182. <https://doi.org/10.1163/156856889x00103> PMID: 2487162
72. Todd JT, Reichel FD. Ordinal Structure in the Visual Perception and Cognition of Smoothly Curved Surfaces. *Psychol Rev*. 1989; 96: 643–657. <https://doi.org/10.1037/0033-295x.96.4.643> PMID: 2798652
73. Nefs HT, Koenderink JJ, Kappers AML. The influence of illumination direction on the pictorial reliefs of Lambertial surfaces. *Percept*. 2005; 34: 275–287. <https://doi.org/10.1068/p5179> PMID: 15895627