Practical Implementation of a Graphics Turing Test

Borg, Mathias; Johansen, Stine Schmieg; Thomsen, Dennis Lundgaard; Kraus, Martin

Published in:
Advances in Visual Computing

DOI (link to publication from Publisher):
10.1007/978-3-642-33191-6_30

Publication date:
2012

Document Version
Early version, also known as pre-print

Citation for published version (APA):
Borg, M., Johansen, S. S., Thomsen, D. L., & Kraus, M. (2012). Practical Implementation of a Graphics Turing Test. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, C. Fowlkes, S. Wang, M.-H. Choi, S. Mantler, J. Schulze, D. Acevedo, K. Mueller, & M. Papka (Eds.), *Advances in Visual Computing: 8th International Symposium, ISVC 2012, Revised Selected Papers, Part II* (Vol. 7432, pp. 305-313). Springer. https://doi.org/10.1007/978-3-642-33191-6_30

Practical Implementation of a Graphics Turing Test

M. Borg[1], S. S. Johansen[1], D. L. Thomsen[1] and M. Kraus[2]

[1]School of Information and Communication Technology, Aalborg University, Denmark
[2]Department of Architecture, Design and Media Technology, Aalborg University, Denmark

Abstract. We present a practical implementation of a variation of the Turing Test for realistic computer graphics. The test determines whether virtual representations of objects appear as real as genuine objects. Two experiments were conducted wherein a real object and a similar virtual object is presented to test subjects under specific restrictions. A criterion for passing the test is presented based on the probability for the subjects to be unable to recognise a computer generated object as virtual. The experiments show that the specific setup can be used to determine the quality of virtual reality graphics. Based on the results from these experiments, future versions of the Graphics Turing Test could ease the restrictions currently necessary in order to test object telepresence under more general conditions. Furthermore, the test could be used to determine the minimum requirements to achieve object telepresence.

1 Introduction

The purpose of the work presented in this paper is to evaluate the realism of computer graphics through a specific setup which restricts the human visual system and takes into account the limitations of today's monitors, for example insufficient black-levels and colour range. In this way, we can create a setup that allows us to test the realism of a virtual object compared to a genuine object. Instead of letting the test subjects decide which of two virtual objects looks more realistic, this setup forces the test subjects to compare a virtual object to a genuine real object in a similar environment. This results in an indirect measurement comparable to the just-noticeable-difference between a real and a virtual scene based on subjective assessments of the realism of the scenes. The aim is to present a setup which can assist in improving virtual reality environments.

Based on work by Slater and Usoh [1] and Steuer [2] we define telepresence as a sense of being physically present in a virtual environment. Presence has been defined as "*the subjective experience of being in one place or environment, even when one is physically situated in another*" [3]. Similarly, we define object telepresence as the subjective experience that a virtual object is situated in the real world.

In accordance with Mel Slater's statement "*Computer graphics is for ... people*" [4], results are obtained through a study of human vision akin the approach used in the Turing Test suggested by Alan Turing [5]. Rather than, for example, evaluating the pixel colour difference between a synthesized image and a photograph, test subjects assess object telepresence directly. A sufficient number of such subjective measurements leads to an evaluation of object telepresence. This conscious evaluation can complement previous results that measured unconscious reactions to virtual environments, e.g. Slater and Usoh [1]. Our main contribution is to implement a restricted Graphics Turing Test and to propose a practicable criterion for passing the test. We show that it is possible to pass the proposed test without excessive effort. While the setup has been tested with simple and static shapes, it is also applicable to more complex scenes.

2 Related Work

Several previous studies propose a virtual reality test in the spirit of Alan Turing's artificial intelligence test. Instead of testing artificial intelligence, these studies test how a real scene compares to a similar representation on a screen. Meyer et al. [6] tested a setup which required test subjects to view a real physical setup as well as the setup being displayed on a colour television (both through a view camera). The study showed that within the specific setup, the test subjects could not tell which scene was real. In this setup, however, the genuine scene was only indirectly compared to the virtual scene because the test subjects were watching it through a view camera. Similar setups have also been used to compare lightness in virtual and genuine scenes [7] and to identify specific parameters of realistic scenes [8].

Moving closer to virtual reality, Brack et al. [9] propose to test the presence of a virtual object when compared to a similar real object. The study concludes that the display of a virtual object would require fidelity reduction techniques to appear as present as the real object. However, it provides no recommendations as to which restrictions are necessary and to which extent. McGuigan [10] determines that computers are not able to create photo-realistic 3D real-time rendering yet. Similarly, Ventrella et al. [11] argue that computer-generated 3D movement is still not believable. In contrast to this earlier work, we show that under certain restrictions it is possible to render and display virtual, static 3D objects so convincingly that people cannot distinguish them from real objects.

3 Graphics Turing Test

Based on previous work on the Graphics Turing Test, we define a restricted version of the Graphics Turing Test, which can actually be put into practice. Thus, our approach is to define certain restrictions that are sufficient to make a virtual object appear as present as a similar real object. To determine the specific restrictions, we summarize the most important depth cues of the human vision [12]:

- Monocular cues: occlusion, size, position, ocular accommodation, linear perspective and motion parallax.
- Binocular cues: stereopsis and convergence.

The experiment is executed by restricting test subjects from using a number of the above cues. To follow the structure of the Turing Test, the test subjects are shown a real object and a virtual representation. They are told before the experiment that one is real and one is not. After viewing both objects, they are asked to pick the object that in their opinion is the real one.

4 Experimental Setup

To carry out the experiment described above, a rectangular box was built, which is illustrated in Fig. 1. One end of the box can be removed and replaced by a screen displaying a virtual representation of the scene inside the removable end. This removable end of the box has a hole in the top, which is 1.6 cm in diameter and allows for the effect of a spotlight on the object inside the box by placing a halogen lamp above the hole. The lamp was measured to illuminate with approximately 2000 lux just below the top hole. For the sphere scene (see below), the illumination was reduced to approximately 1500 lux.

To compensate for the expected limitations of the monitor, the view needs to be restricted. Since the acuity of the human eye is much higher than it is possible to show on a screen and many monitors suffer from insufficient black-levels, a setup was created which takes the capabilities of the human eye into account. A wooden plate with a 2.1 cm hole in the middle (large enough for the test subject to be able to see the entire object) covers the front of the box. The distance from the front of the box to the screen or the separate end of the box is determined by an equation for the visual acuity which specifies the minimum distance necessary for a human to be unable to separate the pixels [13]. Given a resolution of 1,680 × 1,050 pixels on a 22” display, the minimum distance is 163 cm. In order to accommodate variations of the human acuity (e.g. age of the test subjects), the removable part with the object is placed 200 cm from the front of the box.

The test subjects are asked to look through the small hole using only one eye. This restricts the depth perception by avoiding stereopsis, convergence and motion parallax. If the depth of the scene is small enough relative to the length of the box, the ocular accommodation will approximately be constant over the whole scene and therefore, it will not provide any depth cues [12]. It should be emphasized that — according to the measured illumination — the well lit parts of the objects inside the box were very well visible (even under the imposed viewing restrictions).

To ensure that it is not possible to recognise the screen from the box, the test subjects have to keep a distance of about 10 cm from the hole. This distance makes the eyes of the test subjects adapt to a bright environment because of the emittance of 50 lux from the well lit room. This makes them unable to adapt

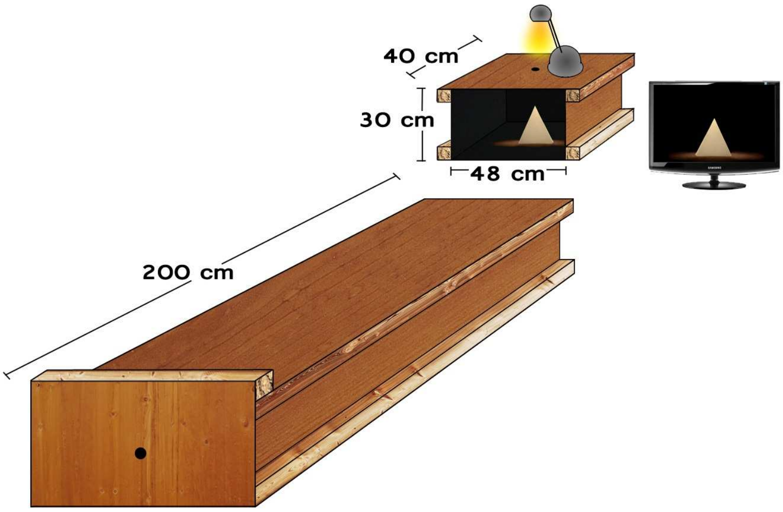

Fig. 1. A representation of the box used for the experiment.

to the dark interior of the box, which is covered with a black fabric. Thus, the ability of the test subjects to distinguish the not perfectly black colour of the screen from the also not perfectly black colour of the black fabric is inhibited. To prevent their eyes from adapting further to the dark environment inside of the box, the subjects are only given 10 seconds to look through the hole. [12]

The box is placed in the test environment so that the test subjects are only able to see the front of the box — the rest is hidden from the test subjects such that they cannot estimate the length of the box without looking inside it. Furthermore, this will keep them from knowing whether they are viewing the screen or the end of the box.

5 Virtual Scene

A blue cardboard pyramid and a blue styrofoam sphere were chosen for the experiment because of the materials, which show only limited specular reflection and have simple textures. These surfaces can be displayed on the employed screen (Samsung SyncMaster 2233RZ), while a highly specular surface would be more difficult to display on the screen as it requires a wider range of colours. The pyramid has a height of 12 cm and a 16 cm base (measured diagonally). The sphere has a diameter of 12 cm.

5.1 Screen Usability and Reference Image

A short test was conducted with the purpose of determining the usability of the screen as well as making a reference image for creating the virtual scene.

A photograph of the pyramid inside of the box was captured and adjusted — giving the base more light and a higher contrast, the pyramid a slightly darker top and brighter bottom and the general image warmer colours (see Fig. 2). 24 test subjects were asked to view the photograph on the screen as well as the object in a randomized order. When asked, 19 of the subjects could not tell which object was real (3 of which chose to guess but did so incorrectly). These results indicate that virtual representations of the pyramid and the sphere can seem as present as the real objects within the scope of this experimental setup.

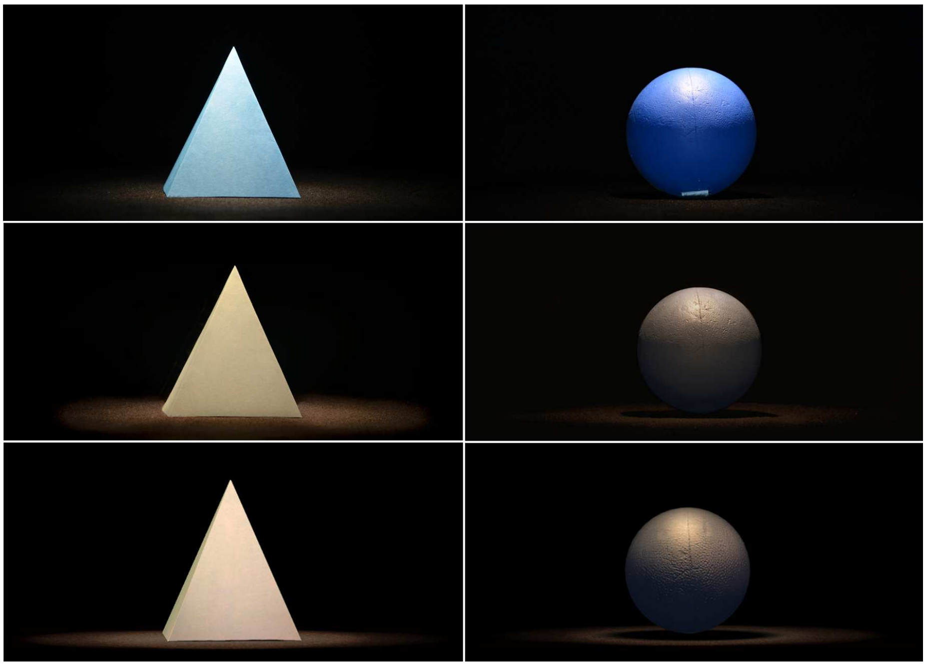

Fig. 2. Top row: The reference images captured. A slow shutter speed reveals the tape in the image of the sphere. This is not visible in the experiment. Middle row: The final reference images which have been adjusted and used for testing the screen. Bottom row: The final scene in Unity as displayed on the screen. Both the middle and the bottom images are displayed in darker colours due to a dynamic contrast function of the employed screen.

5.2 Modeling and Setup

The cardboard pyramid is modeled using a standard square pyramid shape. The edges are chamfered and additional vertices are added to make them crooked. The sphere is a standard sphere shape.

Both textures are created from adjusted photographs of the original object materials. They are applied by making UVW maps of the models. The texture for the planar base of the box is created using multiple layers of noise in yellow, orange and red colours on a black layer. All objects are imported to a game engine as FBX files. The reason for choosing a game engine is the possibility to extend the experiment (see Section 8).

The light in the virtual pyramid scene was set up to simulate the employed halogen lamp. The final result can be seen in Fig. 2.

6 Results

In the experiment using the pyramid scene, four male and four female subjects with an age ranging from 21 to 47 years provided five answers each — giving a total of 40 answers. Of these answers, 22 were correct and 18 were incorrect. For the experiment using the sphere scene, six male and two female subjects with an age ranging from 21 to 28 years were used. They provided 25 correct and 15 incorrect answers.

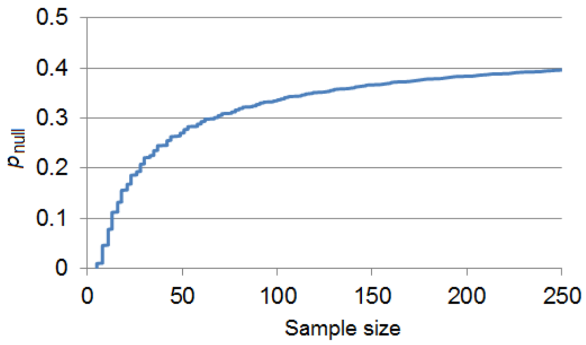

Fig. 3. Graph of the relation between p_{null} and the required sample size that results in a probability of 95 % that the null hypothesis is rejected in the case of complete object telepresence (i.e. subjects are just guessing).

We propose that the Graphics Turing Test is successfully passed if the probability for test subjects to incorrectly identify the virtual object as the real one is estimated as greater than 19 % with a confidence level of 95 %. In other words, we suggest that the null hypothesis of a probability $p_{\text{null}} = 0.19$ (or less) for identifying the objects incorrectly, has to be rejected with a significance level of 5 % by the results. This respects true hypothesis testing, where the null hypothesis is either rejected or not. The threshold corresponds to the commonly used threshold of subjects guessing incorrectly at least 25 % of at least 100 trials [14]. With 40 trials and $\text{p}_{null} = 0.19$, the corresponding threshold is slightly higher to compensate for lower samples sizes. This threshold may also be understood in terms of Alan Turing's statement "*a considerable proportion of a jury, who should not be expert about machines, must be taken in by the pretence*" [15] and we define "a considerable proportion" to be more than 19 %. Note that significantly higher values of p_{null} would require impractically large sample sizes as illustrated in Fig. 3.

The probability mass function for the number i of incorrectly identified objects under the null hypothesis is:

$$f(i|n,p_{null}) = \frac{n!}{i!\,(n-i)!}\,(p_{null})^{i}\,(1-p_{null})^{(n-i)} \tag{1}$$

where n is the number of trials, i is the number of incorrect answers and p_{null} is the probability for the null hypothesis. With this function, a critical number i_{c} of incorrectly identified objects can be computed such that the probability of incorrectly rejecting the null hypothesis is less than 5 %:

$$i_c(n,p_{null}) = \min\left\{ i \;\middle|\; \sum_{j=i}^{n} f(j|n,p_{null}) < 0.05 \right\} . \tag{2}$$

With 40 trials, the critical number of incorrect answers is 13 (see Fig. 4). Since 18 and 15 incorrect answers have been observed, both tests have been passed.

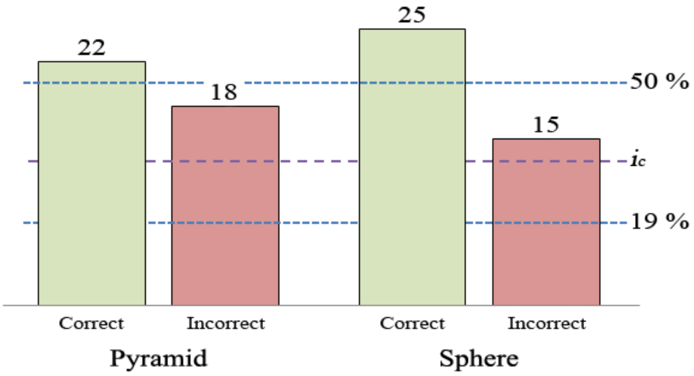

Fig. 4. Results from the two tests with the 50% line representing complete object telepresence (i.e. subjects are just guessing) and the 19% line representing the null hypothesis. i_{c} represents the required critical number of incorrect answers to reject the null hypothesis.

The results show that a "considerable proportion" of the test subjects are unable to recognise which object is real and which is not. A possible source of error is that each test subject had five trials, i.e. not all trials are independent.

Another possible source of error is the test setup which required the test participants to exit the room and wait approximately 20 seconds before being allowed to have a new look. Also, the visual differences between the real and the virtual object in both experiments can be a source of error as the test subjects have a tendency to either guess correctly or incorrectly rather consistently. However, the experiment is not about comparing the real object to a 3D model but to make the test subjects believe that the 3D object is real.

7 Conclusion

In this work, a restricted Graphics Turing Test has been defined and implemented by the approach of avoiding certain cues for depth perception. A practicable criterion for passing the test has been presented. Our experiments show that within the specific scope, the test can be passed without an excessive number of test subjects. A positive outcome of the proposed test strongly suggests that a limited form of object telepresence has been achieved.

8 Future Work

We are currently planning to extend the test setup by including stereoscopic vision and motion parallax. This would allow us to test less restricted forms of object telepresence. Furthermore, the proposed test could be applied to test the quality of many display technologies and rendering techniques. For example, it is also possible to compare animated renderings with real movements to test visual effects such as motion blur.

References

1. Slater, M., Usoh, M.: Body Centred Interaction in Immersive Virtual Environments. In: Artificial Life and Virtual Reality. Volume 1. (1994)
2. Steuer, J.: Defining virtual reality: dimensions determining telepresence. Journal of Communication **42** (1992) 73 – 93
3. Witmer, B., M.J.Singer: Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence: Teleoperators and Virtual Environments **7** (1998) 225 – 240
4. Slater, M.: The Influence of Rendering Styles on Participant Responses in Immersive Virtual Environments (2007) Invited Keynote Talk: Second International Conference on Computer Graphics Theory and Applications http://www.cs.ucl.ac.uk/research/vr/Projects/VLF/Media/CGTA_keynote_2007/vlf%20pit.pdf Last accessed February 24, 2012.
5. Turing, A.M.: Computer Machinery and Intelligence. Mind **LIX** (1950) 422 – 460
6. Meyer, G.W., Rushmeier, H.E., Cohen, M.F., Greenberg, D.P., Torrance, K.E.: An Experimental Evaluation of Computer Graphics Imagery. ACM Transactions on Graphics **5** (1986) 30 – 50

7. McNamara, A., Chalmers, A., Troscianko, T., Gilchrist, I.: Comparing real & synthetic scenes using human judgements of lightness. In: Proceedings of the Eurographics Workshop on Rendering Techniques 2000, Springer-Verlag (2000) 207–218
8. Rademacher, P., Lengyel, J., Cutrell, E., Whitted, T.: Measuring the Perception of Visual Realism in Images. In: Proc. of the 12th Eurographics Workshop on Rendering Techniques. (2001) 235 – 248
9. Brack, C.D., Clewlow, J.C., Kessel, I.: Human Factors Issues in the Design of Stereo-rendered Photorealistic Objects: A Stereoscopic Turing Test. In: Proc. SPIE '10. Volume 7524. (2010)
10. McGuigan, M.D.: Graphics Turing Test (2006) http://arxiv.org/abs/cs/0603132.
11. Ventrella, J., El-Nasr, M.S., Aghabeigi, B., Overington, R.: Gestural Turing Test: A Motion-Capture Experience for Exploring Nonverbal Communication. In: Proc. AAMAS International Workshop on Interacting with ECAs as Virtual Characters. (2010)
12. Wolfe, J.M., Kluender, K.R., Levi, D.M., Bartoshuk, L.M., Herz, R.S., Klatzky, R.L., Lederman, S.J., Merfeld, D.M.: Sensation & Perception. 2nd edn. Sinauer (2009)
13. Clark, R.: Notes on the Resolution and Other Details of the Human Eye (2005) Last updated 2009 http://clarkvision.com/imagedetail/eye-resolution.html Last accessed December 9, 2011.
14. McKee, S.P., Klein, S.A., Teller, D.Y.: Statistical Properties of Forced-Choice Psychometric Functions: Implications of Probit Analysis. Perception & Psychophysics **37** (1985) 786–298
15. Turing, A.M.: Can automatic calculating machines be said to think? (1952) Typescrip of broadcast discussion transmitted on BBC Third Programme, 61th sheet http://www.turingarchive.org/viewer/?id=460&title=5 Last accessed December 9, 2011.