Physical Measurements vs Visual Perception: Comparing colour appearance in reality to Virtual Reality

Beata Stahre and Monica Billger, Dep. of Architecture; Chalmers University of Technology; Gothenburg; Sweden

Abstract

Virtual Reality (VR) has great potential to become a usable design tool for the planning of light and colour in buildings. However, the complex interaction between light and objects makes the problem of lighting scenes a central conceptual and practical problem of computer graphics. This paper aims at discussing the problem of translating reality to its digital counterpart. We deal with methodological issues concerning how to compare visual results between different media, mixed adaptation and arbitrary parameter setting in the software. Our comparison between a 25 m² real room and different VR-simulations showed various problems related to the translation and comparison to reality.

Data were collected from video recorded interviews with 56 observers. Colour appearance was assessed with semantic descriptions, magnitude estimation and colour matching. Measurements of reflectance, luminance and spectral composition were made.

Introduction

The Virtual Reality¹ (VR) -technology of today enables the creation of very realistic models from a geometric point of view and scenes can be created that almost cannot be distinguished from reality. However, when it comes to the light distribution in a scene it is often not necessary with physical correctness. For many applications it is enough that the scene looks ‘believable’ [1]. The importance of reliable simulations of the lighting conditions of a building is pointed out for the fields of virtual prototyping and architecture and lighting design. [1] Today, the different actors in the design process have communication difficulties for visualizing and predicting how the not yet built environment is going to be experienced. To predict the colour appearance² in a room is difficult. A colour on a small sample appears differently than the same colour applied in a full scale room, since the visual appearance will be strongly affected by the light in the room. Realistic virtual environments could make it easier for architects, users and clients to participate in the planning process of material choices, illumination and colouring. The guiding principle ought to let the human perception direct the development of digital design tools.

This paper concerns comparisons of light and colour appearance between real and virtual rooms. It aims at contributing to the discussion on the problem of translating reality to its digital counterpart. The paper is based on studies conducted in the SCAVE²-project. In this project, observers performed sequential comparisons between colour, light and spatial appearance in real rooms and different VR-simulations.

We base our studies on the viewpoint of architectural research. An important standpoint was that the real as well as the virtual rooms should be studied from within, i.e. the observer had to be able to move around inside the room.

Relevant earlier studies

Research on colour vision has developed theories on how humans perceive and experience colours. These are mainly based on studies of colour patches and images, and in fewer cases on simple three-dimensional settings [2,3,4]. To create realistic scenes with the human visual system (HVS) as a base instead of physically correctness has been the aim for parts of the research concerning comparative studies between reality and digital simulations. Devlin et al [5] point out that new psychophysical research is needed. Changes of contrast and colour perception need to be considered in order to produce perceptually accurate images on a display. There are new psychophysically based visual models that aim to create solutions for these problems. However, the focus tends to lie on singular aspects for single purposes. The knowledge of HVS is still, according to Devlin, limited and needs developing. [5]

There has been much research on how to predict colour appearance under different viewing conditions. Colour appearance models are developed in order to enable colours to appear equal between different media. This is a complicated task and hitherto the datasets used in the psychophysical studies are predominately based on assessments of colour patches in light booths [6]. To predict the colour appearance in a room is difficult. The visual appearance of the room colour is depending on many different factors, such as the physical properties of the light; in which angle it strikes and its flux, and the texture and colour of the surface [7]. Some results on colour appearance in real interior spaces have been published in the latest years [8,9,10,11].

¹Virtual Reality (VR) has no absolute definition, but it is here referred to a computer-generated 3D-world that allows the user to feel present and interact with the world, i.e. move around in it, move objects etc.

²Colour appearance is in this paper referred to as a general concept for the perceived colour of a surface.

³To simulate the experience of colour in virtual environments: a comparison of the experience of light and colour in real and virtual rooms (SCAVE). This project is interdisciplinary and aims to develop a virtual laboratory for trustworthy simulations of colour, light and spatial experience in enclosed spaces. It was initiated and planned by Monica Billger at the Dep. of Architecture at Chalmers University of Technology in 2001. Other institutions that are involved at Chalmers are the Centre for Digital Media and Higher Education, the Dep. of Computer Engineering and the Dep. of Technology Management and Economics. External collaborator is the Dep. of Information Technology at the University of Milan.
In the design process VR can be used as a tool which more correctly can reproduce the planned reality than traditional design tools are able to. While considerable work is being done on usability issues concerning technical interaction in Virtual Environments (VE) [12,13,14], there are few comparative studies focusing on the appearance of colour, light and space in reality compared to VR.

Horňáňský Dahlholm and Rydberg Mitchell report on the impact of colours and textures on size perception [15]. Colour appearance in itself is not in focus, it is used as a tool to affect the spatial appearance. Their results showed that the impact of colour and texture on size perception varied between different VR systems. A previous study [26] in the SCAPE-project discussed the potential of VE’s to become a usable design tool concerning light and colour highlighting the relation between the factors influencing the spatial experience. Benefits and disadvantages of the used models were discussed. The conclusions drawn was that better rendering quality led to higher task performance and had a positive effect on the spatial experience.

The rendering equation, developed by Kajiya [16], is one of the most central problems to solve in computer graphics. Attempts to solve this equation are made through all 3D display methods in computer graphics and there are various ways to do it. The local and global solution is two ways of classifying the rendering equation. In local solutions reflections between surfaces are not considered. Global solutions take into account the ways in which light is transferred between surfaces. When dealing with simulating colour appearance in rooms, only global solutions are relevant.

Mania et. al [17] aimed at photorealistic simulations and have made great efforts to control light and colours, using objects painted the same shade of blue. They did not visually assess or discuss colour appearance. Instead their focus lied on memory recalling of the objects in the room. Their study considered questions concerning cognition, through demonstrating the difference between using task performance based metrics and human evaluation of cognitive awareness states.

Meyer et al [18] introduced a method for comparing real environments with digital images. This method was built upon measurements of radiant energy flux densities in a simplified physical environment; i.e. a box with some blocks, compared to a simulation where radiosity was used. The simulation was then converted to a digital television image, which was compared with the physical environment. Both the image and the box were viewed simultaneously through a view camera. The matching between the simulation and reality turned out to be quite good. The observers found for example colours to match slightly better than good and the shadows slightly less than good.

Another study were simulations are compared to a real scene is done by McNamara et al. [19]. The study aims at investigating how much computation is enough in order to create a trustworthy image, based on the HVS rather than on physical correctness. For this the ray tracing based software Radiance is used. Objects are viewed in a real scene, i.e. a lighting booth and compared to rendered images. McNamara et al aims at study lightness, not colour. The observers’ impressions of the objects’ lightness have formed the base of the study, where levels of grey are matched. The scenes were viewed monocularly in order to eliminate depth cues.

Experimental design

Comparisons were made between a 25 m² real room and monographic VR-models on a desktop PC. The reference room, i.e. the real room, was designed to give examples of various colour phenomena (see Fig. 1). Only matte colours were analyzed. Different light situations were studied: incandescent, fluorescent 2700K and 3000K. In phase I of the studies, we compared the observer’s visual assessments of the real and the virtual rooms. In phase II, we elaborated with the software parameters. Finally, in phase III we conducted physical measurements. We have started to analyze them and in this paper we use a few examples to highlight the discussion.

A crucial criterion for choice of software was to simulate the chosen light fixtures correctly. The light fixtures were chosen on the basis of that the manufacturer could provide well produced digital models. These models support photometric values and in order to use them, a radiosity based software was needed. Thus, we chose the Windows based software Lightscape™ and 3Ds max®, commonly used by architects and designers in Sweden.

Radiosity is a widely used method for calculating diffuse light spread. As a rendering technique it was a good choice of method for the project because of: 1) the solution is view independent, meaning that it will look the same from every view point (only its projection changes), 2) due to the previous point, the solution can be precomputed and then displayed in real-time, which is one major reason to use it in this project, 3) it handles diffuse-diffuse interreflections, 4) and it is supported by many programs, including 3Ds max® and Lightscape™. However, there are also some disadvantages. Radiosity cannot handle specular reflections i.e. reflections from mirror-like surfaces, as opposed to diffuse reflections, gloss and shadows in 3D [24].

Another interesting light calculation technique to apply is photon mapping, which, however, does not support photometric light and is therefore not considered as choice here.

Studies show that software Radiance [20] gives good results [19]. This is a ray traced based light calculation programme, which “takes as input a three-dimensional geometric model of the physical environment, and produces a map of spectral radiance values in a colour image” [manual at www.radiance-online.org]. However, Radiance is image-generating only, and it is therefore not corresponding to our aims of analyzing 3D-environments.
A 3D-model of the room was made in 3dsmax 6.0®. This model was exported to VR [21] and showed on a desktop PC. The process to translate real paints into their digital counterparts is described in [22]. Important for the translation was the use of the colour reference box for transferring the visual matches from the real to the virtual room. Since the translation was complex it was important to find an acceptable level of correctness. The relations between differently coloured surfaces had to be as correct as possible. On the other hand, small translocations of the colour scale could be accepted, since they were results from the adaptation to the surrounding light and the light from the computer. The computer was calibrated to 5500K.

It is important to stress the simulations are based on the information given at the stage of planning, i.e. the physical conditions, such as materials and light fixtures, available before the room is built.

**Phase 1-3: Methods and Results**

**Phase 1: Observations in the real vs the virtual room**

**Methods**

The comparisons between the different situations were sequential. Data were collected from video recorded interviews and questionnaires. Six methods were used for evaluating light, colour and space: 1. Free description of the room, 2. Motivated semantic differential scaling, 3. Visual evaluation of light, 4. Verbal descriptions of the colours, 5. Magnitude estimation [30], 6. Colour matching with a light box [23] In total, 56 participants were involved. Half of them were male and half female. All of them had normal colour vision.

In each light situation in the real room, 20-30 observers were involved. In each VR-application, 10-20 observers were involved. Different methods required different numbers of observers. Of the methods mentioned above, the first four methods were used by all observers in all situations. Magnitude estimation was only used in the real room. Only 3-5 observers were chosen to use magnitude estimation and colour matching with the colour reference box. Earlier studies of colours in rooms showed that it is more difficult to assess whole walls than small colour patches [25]. It was concluded that a few well-trained observers were more reliable than a larger group with no training. Thus, the observers got specific training in estimating colour attributes according to the NCS system. This method requires long time; thus it was not possible to include the majority of observers. The previous study showed that colour matching was more precise than magnitude estimation [25]. However, the observer’s state of adaptation had to be taken into consideration. In the studies using this technique, the same illumination in the box as in the room was used.

Important for the studies was that the observers used both their eyes, in order to gain full depth cues.

**Results**

In the reference room, surfaces perpendicular to or opposite each other became more similar due to reflections. However, on each uniformly painted area, different colour variations were clearly visible. Moreover, between differently painted surfaces on the same level, effects of simultaneous contrast were evident. A brightness phenomenon appeared, i.e. a light surface in the darkest corner was perceived as whitest of all surfaces.

Our studies proved significant differences between real and virtual rooms. The VR-room had incorrect reflection effects between surfaces, too few colour variations and too achromatic shadows. Also it had incorrectly reproduced contrast effects for the lightest surfaces. The one surface that was perceived as the whitest one in the room was for example far too grey. Moreover, chromatic information on illumination is heavily simplified itself.

**Phase 2: Software parameters exploration**

**Methods**

In order to define the discrepancies, a direct comparison between the display and colour samples was made. The colour samples, picked out by the observers to match the surfaces in the real room, were placed and analyzed in the colour reference box. With this analysis as a ground, the parameters describing colour bleeding, reflectance and colour temperature (i.e. the colour of the light source) were manipulated to make the model resemble the real room as closely as possible. On the basis of these manipulations the changes that were needed in the parameter settings could be noted.

**Results**

Through the elaboration, problems with the radiosity calculations were pointed out. It showed that in order to make true and realistic colour visualisations in virtual environments, the colour appearance of the real objects (rooms, buildings, etc) must be known. However, even if the real objects are known and available, correct simulation is not always possible. For example, high reflectance in the digital models turned out to brighten up the areas around the long red wall too much (see Fig 2). Low reflectance on the dark areas gave a better result, in combination with increased colour bleeding. For the light long wall high reflectance was needed for the light areas to get enough brightness. High reflectance for this wall resulted in its lightness being mirrored on the opposite red wall, which then became too light. In order to compensate for this, the luminance level needed to be decreased significantly. As with the colour bleed in both programmes, different colours need different levels of the parameter setting.

Through the conducted experiment a fairly good match was obtained between the real and the virtual room.
**Phase 3: Physical Measurements – Work in Progress**

**Methods**

In our study measured values were not used initially for describing the room, but instead applied as a retrospective reference and documentation. Measurements of reflectance, luminance and spectral composition were made.

40 local points of the real room and 27 corresponding points in the VR models were measured.

Table 1 gives examples of the collected data for four areas in the room, as well as the NCS-codes for the paints and results from magnitude estimation and verbal description. These areas are examples of surfaces that matched well and their matching values.

The reflectance of each painted area was measured with a BYK-Gardner Colour Guide spectrophotometer. The L*a*b-values are shown in column B. The display Lab-values, as defined in Photoshop, for the digital colours are shown in column G.

The spectral composition of the areas in the reference room and the matching colour patches in the light box were close. However, the corresponding areas in the VR-room had distinctly different L, x and y values.

In order to get comparable measurement values between the reference room and the display we had to re-calculate them to corresponding colours under the same illuminant (see Table 2). The white reference values for the reference room and the digital models were different. When bringing the L, x, y-values to a connection space (D65), using the Bradford transform [28] it showed that they were fairly close.

### Table 1

<table>
<thead>
<tr>
<th>Area:</th>
<th>Reference room:</th>
<th>Visual assessments:</th>
<th>VR-room:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A = Matching NCS-code, B = Reflectance (spectrophotometer measurements), C = Spectral composition (Spectroradiometer measurements), D = Light box 3000K (measurements of matching NCS-sample), E = Magnitude estimation (assessed in NCS), F = Verbal description, G = Display Lab values, as defined in Photoshop, H = Display spectral composition. Note that D to F is visual assessments results using 3 different techniques.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reference room:</td>
<td>Visual assessments:</td>
<td>VR-room:</td>
</tr>
<tr>
<td>-------</td>
<td>----------------</td>
<td>---------------------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>A 1030-G90Y</td>
<td>B 84,24</td>
<td>C 4907</td>
</tr>
<tr>
<td></td>
<td>1030-G90Y</td>
<td>96,8</td>
<td>491</td>
</tr>
<tr>
<td>1</td>
<td>A = Matching NCS-code, B = Reflectance (spectrophotometer measurements), C = Spectral composition (Spectroradiometer measurements), D = Light box 3000K (measurements of matching NCS-sample), E = Magnitude estimation (assessed in NCS), F = Verbal description, G = Display Lab values, as defined in Photoshop, H = Display spectral composition. Note that D to F is visual assessments results using 3 different techniques.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reference room:</td>
<td>Visual assessments:</td>
<td>VR-room:</td>
</tr>
<tr>
<td>-------</td>
<td>----------------</td>
<td>---------------------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>A 1030-G90Y</td>
<td>B 84,24</td>
<td>C 4907</td>
</tr>
<tr>
<td></td>
<td>1030-G90Y</td>
<td>96,8</td>
<td>491</td>
</tr>
<tr>
<td>2</td>
<td>A = Matching NCS-code, B = Reflectance (spectrophotometer measurements), C = Spectral composition (Spectroradiometer measurements), D = Light box 3000K (measurements of matching NCS-sample), E = Magnitude estimation (assessed in NCS), F = Verbal description, G = Display Lab values, as defined in Photoshop, H = Display spectral composition. Note that D to F is visual assessments results using 3 different techniques.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reference room:</td>
<td>Visual assessments:</td>
<td>VR-room:</td>
</tr>
<tr>
<td>-------</td>
<td>----------------</td>
<td>---------------------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>A 1030-G90Y</td>
<td>B 84,24</td>
<td>C 4907</td>
</tr>
<tr>
<td></td>
<td>1030-G90Y</td>
<td>96,8</td>
<td>491</td>
</tr>
<tr>
<td>3</td>
<td>A = Matching NCS-code, B = Reflectance (spectrophotometer measurements), C = Spectral composition (Spectroradiometer measurements), D = Light box 3000K (measurements of matching NCS-sample), E = Magnitude estimation (assessed in NCS), F = Verbal description, G = Display Lab values, as defined in Photoshop, H = Display spectral composition. Note that D to F is visual assessments results using 3 different techniques.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reference room:</td>
<td>Visual assessments:</td>
<td>VR-room:</td>
</tr>
<tr>
<td>-------</td>
<td>----------------</td>
<td>---------------------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>A 1030-G90Y</td>
<td>B 84,24</td>
<td>C 4907</td>
</tr>
<tr>
<td></td>
<td>1030-G90Y</td>
<td>96,8</td>
<td>491</td>
</tr>
<tr>
<td>4</td>
<td>A = Matching NCS-code, B = Reflectance (spectrophotometer measurements), C = Spectral composition (Spectroradiometer measurements), D = Light box 3000K (measurements of matching NCS-sample), E = Magnitude estimation (assessed in NCS), F = Verbal description, G = Display Lab values, as defined in Photoshop, H = Display spectral composition. Note that D to F is visual assessments results using 3 different techniques.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reference room:</td>
<td>Visual assessments:</td>
<td>VR-room:</td>
</tr>
<tr>
<td>-------</td>
<td>----------------</td>
<td>---------------------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>A 1030-G90Y</td>
<td>B 84,24</td>
<td>C 4907</td>
</tr>
<tr>
<td></td>
<td>1030-G90Y</td>
<td>96,8</td>
<td>491</td>
</tr>
</tbody>
</table>

The L, x and y-values were measured for the closest matching colour patch placed in the light box (column D).

**Data analysis and Preliminary Results**

L*a*b*-values measured in reality could not be directly applied as digital Lab-values. They needed to be slightly adjusted. Corresponding Lab-values on the display were too brownish for our four different yellow paints. The strong red and green needed to be adjusted in the opposite way. Otherwise the simulation became too brilliant and whitish.

The luminance was naturally very different between the reference room and VR. However, the light level in the VR-room was assessed almost the same and the colours agreed fairly well. The lightest areas were not simulated well; they became too grey.

Spectral composition of the areas in the reference room and the matching colour patches in the light box were close. However, the corresponding areas in the VR-room had distinctly different L, x and y values.

In order to get comparable measurement values between the reference room and the display we had to re-calculate them to corresponding colours under the same illuminant (see Table 2). The white reference values for the reference room and the digital models were different. When bringing the L, x, y-values to a connection space (D65), using the Bradford transform [28] it showed that they were fairly close.
Concluding discussion

Our study deals with how to compare visual results between different media, mixed adaptation and arbitrary parameter setting in the software. It showed various problems related to the translation and comparison of reality to VR.

For most surfaces, the colour appearance in reality and in VR proved to be visually similar. When bringing the x and y-values to a connection space (D65), it showed that they were fairly close. Overall, the methods used here provide a reasonable accuracy to describe the colour appearance of the room.

However, when brought to the connection space, a small difference between the measured values was visible. To a certain degree this could either be the result of measurement errors, since different measuring instrument have been used, or problems with the formula used for bringing the values to the connection space, e.g. the normalisation of the white point. Other problems include the assessment methods and the sequential comparisons. Sequentiality is necessary, because we need to study the rooms from within and adapted ot the illumination. In order to do the comparison we have encircled the colour appearance by verbal descriptions conducted by all observers. They were asked to compare the different coloured surfaces in the room (e.g. which one was the whitest) in order to describe the colour attributes. This gave us a good description of the relation between the surfaces of the room. The observers agreement was unanimous. The magnitude estimations gave further precision to the description of the relation between the equally painted but differently perceived surfaces in the real room. The colour reference box enabled a match between the chosen samples from the reference room and the display. From what we can see, this provides us with a well-functioning method. Xiao et al [9] have developed an interesting alternative method, where the matching is conducted directly on the screen.

However, we still have the problem of mixed adaptation [29] when looking at a display in a nearly dark room and compare it with samples in the lighting box.

Small differences in colour appearance can be significant for the experience. It is desirable to obtain estimations of at least 5 NCS-steps (not in the B-G area were the differences are small indeed). Differences between illuminations are often experienced to have especially great impact on light colours or when the colour category changes. The colour category of yellow only covers a small colour area and it does not take much for the colour to be experienced as beige, white or green. This was obvious in our study, which involved four different yellow colours. These colours showed very different appearance in the different light situations.

Concerning the radiosity calculations in the used software, some problems were pointed out. To know the colour appearance of the real objects (rooms, buildings, etc) proved to be essential for the making of true and realistic colour visualisations in virtual environments. Still, to obtain correct simulations is not always possible, even if the real objects are known and available.

Finally, the architectural perspective opens up and demands other aspects to be taken into account than a technical or (natural-) scientific one. For a meaningful result concerning colour, qualitative aspects are more important than quantitative ones, which also involve other methods.

Guiding principles for our project has been the assumption that to gain knowledge about colour appearance in rooms, we must study enclosed spaces, and consider the complexity of the context as an essential feature of the studied situation. Spatial conditions such as the volume of the room, materials, colours, and light cannot be separated, and should all be taken into account.

It is assumed that methods for meaningful investigation can be found and, although it may not be possible to make predictions for every coloured material in all situations, they are means to deepen the awareness and understanding of how spatial and lighting factors affect coloured material.

Acknowledgement

Many thanks to Dr Rejean Barbier at the National Research Council in Canada for assisting us with calculations and analysis of the measurement data. Any mistakes in interpretation are our own. Many thanks also to Professor Ronnier Luo at the University of Leeds for his reflections and comments on our study. Both have been of great assistance in pointing out important issues to proceed with.

References


Table 2. X and y differences in the original data compared to the connection space.

<table>
<thead>
<tr>
<th>Area:</th>
<th>Difference between Reference room and Vr-room:</th>
<th>Difference between Reference room and Vr-room:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original data (table I)</td>
<td>Connection space</td>
</tr>
<tr>
<td></td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>1</td>
<td>-0.0690</td>
<td>-0.0200</td>
</tr>
<tr>
<td>2</td>
<td>-0.0750</td>
<td>-0.0220</td>
</tr>
<tr>
<td>3</td>
<td>-0.0640</td>
<td>-0.0330</td>
</tr>
<tr>
<td>4</td>
<td>-0.0730</td>
<td>-0.0330</td>
</tr>
</tbody>
</table>

Table 2. X and y differences in the original data compared to the connection space.
Author Biography

Beata Stahre and Monica Billger work at the Dep. of Architecture at Chalmers in Sweden. Stahre is an architect currently studying for a Ph.D. Her licentiate thesis is focusing on spatial colour appearance in real and virtual rooms.

Billger, PhD, is an interior architect who works as a research scientist and teacher. She presented her dissertation "Colour in enclosed space" in 1999. Since then she has run projects focused on comparing real rooms to VR-rooms.