

Received September 20, 2021, accepted October 2, 2021, date of publication October 6, 2021, date of current version October 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3118438

# On the Validity of Virtual Reality Applications for Professional Use: A Case Study on Color Vision Research and Diagnosis

HALINA CWIERZ<sup>ID</sup>, FRANCISCO DÍAZ-BARRANCAS<sup>ID</sup>, JULIA GIL LLINÁS<sup>ID</sup>,  
AND PEDRO J. PARDO<sup>ID</sup>

Centro Universitario de Mérida, Department of Computer and Network Systems Engineering, University of Extremadura, 06800 Mérida, Spain

Corresponding author: Francisco Díaz-Barrancas (frdiaz@unex.es)

This work was supported in part by the Regional Government of the Junta de Extremadura under Grant GR18131, Grant IB16004, and Grant IB20094; and in part by the European Regional Development Fund.

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

**ABSTRACT** In recent years, there have been very important advances in graphic computing and technology related to the capture and representation of real objects in both 2 and 3 dimensions. One of these technologies is virtual reality, which can be incorporated into common tasks in research laboratories, especially in laboratories related to color vision and lighting research. To incorporate virtual reality devices into research tasks, newly developed applications must be validated with existing and known tests or techniques. The objective of this work was to study the validity of a commercial VR system for research and diagnosis in color vision. We carried out a comparative study on the behavior of these immersive systems for viewing 3D scenes in real time using a color vision test. In particular, we implemented a virtual version of the Farnsworth-Munsell 100 Hue test and compared the results obtained by 17 normal and 3 defective observers in both the physical and virtual tests. The results show that the functionality of both tests is very similar and that the diagnosis of both methods is equivalent. Detailed analysis of the results of both tests indicates that there is a slight difference in scale between the two tests. This difference in scale indicates a greater difficulty in the case of the virtual test but does not affect the final diagnosis. This could be due to the greater difficulty in using a head-mounted display (HMD).

**INDEX TERMS** Virtual reality, 3D scenes, computer vision, color research, color reproduction, color deficiency.

## I. INTRODUCTION

Technology is continuously evolving and offering new devices and/or techniques that professionals, researchers or end users try to incorporate into their daily lives, whether for professional, research or personal reasons, to obtain the possible benefits of the new technology. This integration process can be part of a long process in which the industry develops a new professional and specific device or service, and after several years of the refinement process, it finally reaches the end-user market. Another integration process is directly targeting new devices or techniques to the final end-user market, pursuing a generalization of their use, and after this generalization, several professional or specific applications

can be developed by third-party researchers. The last case corresponds to the current state of evolution and development of virtual reality (VR) devices or services. The fact that virtual reality systems have been developed with the consumer market in mind does not invalidate their use for research or professional applications. It is only necessary to check the validity of this new technology against a well-known traditional tool or application.

In relation to the specific fields of color vision and lighting research, the introduction of virtual reality systems is not the first time that new commercial devices have been introduced in research laboratories to replace specific scientific devices such as visual colorimeters or anomaloscopes: image devices based on cathode ray tubes were first applied [1], [2], with LCD [3], [4] and OLED screens being launched later [5]. In all these previous cases, the chromatic

The associate editor coordinating the review of this manuscript and approving it for publication was Charith Abhayaratne<sup>ID</sup>.

characterization of these new types of devices has been a key point [6]–[11].

From a visual point of view, VR technology creates the visual sensation of immersion in a three-dimensional world. This 3D visual sensation is created using specific hardware and software components. The hardware component is always based on head-mounted displays (HMDs). There are two different types of hardware: devices that do not have their own graphic hardware and need a personal computer for this task and devices that use a mobile phone or other specific graphic hardware without a personal computer. There are significant differences in performance between the two types of devices. In this work, we will refer exclusively to the first type: those devices that are associated with a personal computer with a dedicated graphics card.

In terms of the software component of a VR system, there are two main commercial software platforms for developing virtual reality content: Unreal Engine and Unity Game Engine. In both platforms, mathematical functions are used as basic rules of internal functioning that try to reflect to a greater or lesser extent the real world through physical laws [12].

It is possible to obtain rendered scenes with a high degree of visual appearance fidelity when treating the light–matter interaction simulating the physical laws governing this phenomenon. To handle the lighting and shading conditions, the graphic engine uses a physical bidirectional reflectance distribution function model (BRDF) with four main components: diffuse, specular, normal, and smoothness. The diffuse component corresponds to material color, the specular component corresponds to surface color, and the normal and smoothness components correspond to surface texture.

In this study we research whether it is possible to use virtual reality technology for research tasks related to color vision and lighting. There are other examples of the utility of this new technology in other research fields [13]–[15]. We can approach this topic from different perspectives. On the one hand, we have the paradigm of classical colorimetry which characterizes the image device through spectroradiometric measurements, from which we can distinguish between absolute colorimetry or relative colorimetry. In both cases, the only relationship with human visual perception is through the standard observer CIE XYZ 1931 defined by the International Commission on Illumination (CIE) [16]. On the other hand, we have the paradigm of psychometry, where we would measure the effect of a scene shown on a VR device on the perception of the observer without having to be linked to any physical property [17]. Finally, the paradigm of psychophysics tries to relate the physical world with the perceptual world but this case has an intermediate processing layer dedicated to the simulation of a physical world in a virtual world, in such a way that we could call it virtual psychophysics [18], [19].

The objective of this work is to check the validity of a commercial virtual reality system to be used in research and diagnosis tasks in color vision and lighting laboratories.

This validity arises from the point of view of what we have previously called virtual psychophysics, since it uses the simulation of the physical laws related to light–matter interaction. However, we have also studied the effect of this physical simulation on visual perception to check both aspects of psychophysics. Determining the validity of a specific use of this type of virtual reality system does not validate these systems for all uses in research tasks related to color vision research. However, if the result is positive, it opens the door to a wider future use in different fields provided each of its applications has been previously validated.

Specifically, in this study, a virtual version of the Farnsworth-Munsell 100 Hue test (FM 100) for color vision assessment has been implemented in a virtual reality environment. To verify its validity as a research and diagnosis tool, first, a relative colorimetric reproduction of the FM 100 test is necessary, and second, the results obtained with this virtual test are compared to those obtained with a real version of the mentioned test. This comparison is based on the analysis of the results obtained by both versions on the same population sample. These tasks try to answer the research questions posed previously: Is it possible to use virtual reality technology for research tasks related to color vision and illumination? What would be the validity of a commercial test implemented in a virtual reality environment?

In Section II of this paper, we review the state-of-the-art color vision testing methods and explain in detail how the FM 100 test works. In Section III, we describe the methodology used in this work, including the chromatic characterization of the VR system, the spectroradiometric measurements of the FM 100 test and the development of the 3D virtual scene. Finally, in Sections IV and V, the experimental results are analyzed and discussed, and conclusions are drawn.

## II. TESTS FOR COLOR DEFICIENCY

Color perception depends on the light falling on the three types of cones of the retina. When a person suffers from color vision deficiency (CVD), it can limit his or her daily life, academic life or even work life.

Color vision tests play an important role in detecting visual deficiencies that may be related to pathologies such as optical neuritis, pituitary adenoma, glaucoma, and diabetes. In order to diagnose these pathologies very sensitive tests are required [20] which, on the one hand, allow the detection of small deviations from normal color vision, and, on the other hand, allow discrimination between different degrees of alterations in color vision.

The methods used in these tests for the detection of visual impairment can be grouped into three types:

- Pseudo-isochromatic tests: This type of test is based on the use of films composed of small colored surfaces (background) from which some colored areas stand out, forming a figure that will only be visible to normal observers on some occasions, while on others, it will only be visible to abnormal or defective observers.

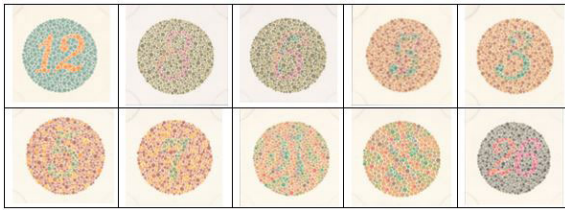


FIGURE 1. Different Ishihara test plates.

This method provides a quick result (Ishihara, HRR, SPP, Ulloa, etc.).

Among the pseudoisochromatic tests, the Ishihara test, the first version of which appeared in 1906 [21], is the most widespread and used for the study of chromatic defects. The main objective of this test is to detect red-green type deficiencies and in its current configuration, it consists of thirty-eight plates (Fig. 1), twenty-five with numbers for observers who can read and thirteen with sinuous lines for those who cannot. For this type of test, we simulated the visual aspect in virtual reality [22].

- Tests were carried out using special instruments such as an anomaloscope. This device is essentially a colorimeter that produces a metameric pair from a mixture of two pure spectral colors of variable proportions to equal a reference color on a bipartite field. This device can be adapted to each problem.
- Tests for sorting colored caps in the natural order, from blue to red and various shades in between. They allow the detection of color perception deficiencies (Farnsworth-Munsell test, FM 100 and its reduced versions D-15, B-20 and H-16, Roth 28-Hue test, Lanthony test, etc.).

Ordering tests are currently one of the main tools for the diagnosis of congenital and acquired anomalies. They provide sufficient information while maintaining relative simplicity in the task to be performed by the subject compared to methods using an anomaloscope.

Among the sorting tests, the most sensitive one is the 100-tone FM 100, although it actually consists of 93 (85 unfixed and 8 fixed) caps. This test is based on the notation system developed by Munsell [23]. It consists of a series of colored caps with constant saturation and clarity but different tones (Fig. 5). In this work, we develop a virtual scene with the FM 100 test applying color management techniques.

**A. FM 100 TEST SCORING METHOD**

The goal of this test is to place the color palettes in the correct order based on the color hue. Scores for the test are based on two factors:

- Frequency of the color caps are misplaced.
- The severity or distance of the misplacement.

The scoring tool provided with the test calculates the total error score (TES) and graphically represents the errors made by the observers (Figs. 2-4). A solid black line represents the observer's response to the test. The further it moves away

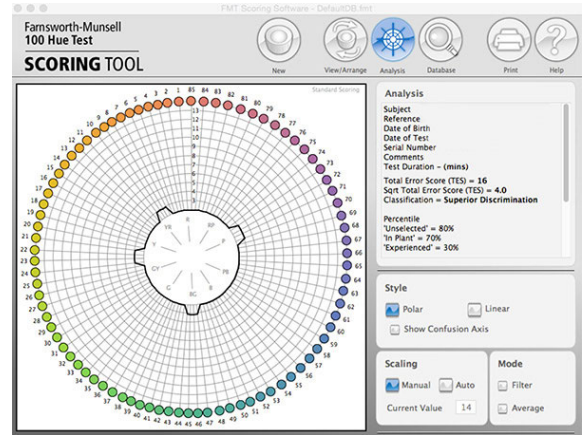


FIGURE 2. Superior (good) software score.

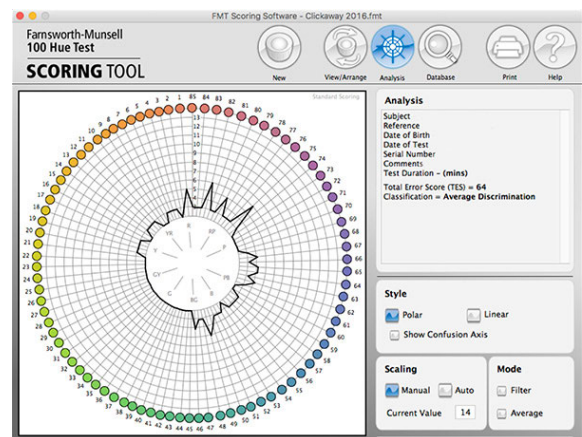


FIGURE 3. Average (normal) software score.

from the center circle, the larger the errors will be. The test results can be classified according to the following ranges:

**1) SUPERIOR (GOOD) SCORE**

Approximately 16% of the population makes 0 to 4 transpositions on the first test or has total error scores of zero to 16. This is a superior range of competence for color discrimination.

**2) AVERAGE (NORMAL) SCORE**

Approximately 68% of the population scores between 16 and 100 on the first tests. This is a normal range of competence for color discrimination.

**3) LOW (WEAK) SCORE**

Approximately 16% of the population has a total error score above 100. The first retest may show an improvement, but further retests do not significantly affect the score.

**III. METHODOLOGY**

The methodology applied in this work can be divided into six stages. The first stage describes the method applied for the chromatic characterization of the virtual reality device and the results obtained. The second stage shows how color

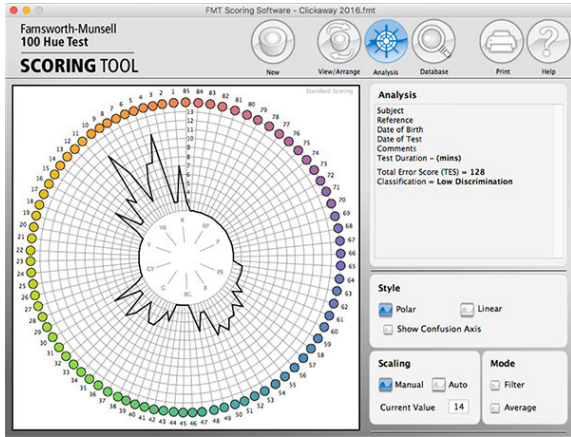


FIGURE 4. Low (weak) software score.



FIGURE 5. Original Farnsworth-Munsell 100 Hue test samples divided into four rows.

management has been carried out within the VR software platform. In the third stage, the results of the spectroradiometric measurements performed on the FM 100 test are shown. Finally, each of the following are described: the virtual scene design, the graphics settings employed, the population sample used and the procedure followed to perform the test.

**A. CHROMATIC CHARACTERIZATION OF HEAD-MOUNTED DISPLAY**

The first step to use a virtual reality system in tasks related to color vision research is the chromatic characterization of head-mounted displays (HMDs). Each device of this type has its own specific characteristics in terms of the chromaticity of its primary colors and the white point, as well as for the relationship between the digital values of the analog-to-digital converter (DAC) and the associated tristimulus values. However, the main commercial devices currently on the market use two LCD or OLED screens (one for each eye) and Fresnel-Aspherical hybrid lenses to adapt the observer’s accommodation distance to infinity, even when the screens are at a short distance from the user’s eyes. With these lenses, the field of view (FOV) covered is close to 110° per eye. The authors of this paper have used the same methodology with different HMDs in other works [22], [24] and have developed a simple color characterization model for VR devices. This model is based on a typical linear transformation between the RGB’ values and the XYZ tristimulus values using a 3 × 3 matrix. RGB’ values are obtained after a gamma correction

TABLE 1. Chromaticity coordinates of each primary channel and white point and gamma value of each channel.

Channel	Chromaticity		Luminance	Gamma	
	x	y	Y (Relative)	Value	R2
White	0.299 ± 0.002	0.315 ± 0.002	100.0		
Red	0.667 ± 0.004	0.332 ± 0.003	30.3 ± 1.1	2.34	0.999
Green	0.217 ± 0.007	0.710 ± 0.002	75.4 ± 2.4	2.30	0.999
Blue	0.139 ± 0.002	0.051 ± 0.002	7.8 ± 0.5	2.26	0.998
Black	0.335 ± 0.01	0.361 ± 0.004	0.4 ± 0.2		

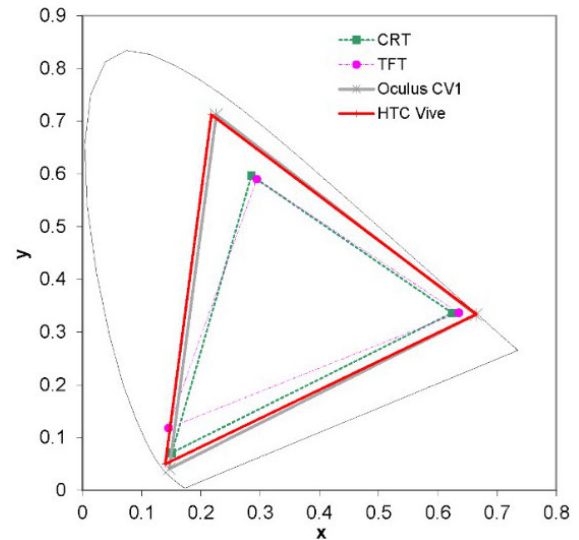


FIGURE 6. Color gamut of different displays represented in the CIE 1931 chromaticity diagram. The evolution of the color gamut with different display technologies can be seen.

of the original RGB values that guarantee the linearity of the system.

For the specific HMD model used in this work (HTC Vive), the chromaticity values of primary colors and the white point and the gamma value of each chromatic channel are shown in Table 1.

The color gamut associated with both HMD display types is similar to that of other displays with OLED technology, as shown in Fig. 6.

The measurements were performed in steps of 5 RGB values from 0 to 255 per channel. Then, we measured 50 random values to check the final error. Finally, we performed some systematic measurements. In total we have made 234 measurements. In the supplementary material, we have provided the RGB and XYZ values of these measurements. We obtained an average color error of 1.5 units in terms of CIEDE2000 with a standard deviation of 0.6.

**B. COLORIMETRIC CALCULATIONS MODULE FOR VR**

One of the main problems regarding color management in virtual reality devices is the high refresh frequency of images, between 90 and 120 Hz, due to the low latency in the interaction of the user with the environment. This high frequency reduces the time available for colorimetric calculations. For this reason, we have chosen a display characterization model

that does not require complex calculations but only seeks to relate, as simply and accurately as possible, the values of the DAC with chromatic values of the stimulus in any reference color space. In this case, the most appropriate color space will be the tristimulus space associated with the CIE 1931 standard observer; the chromatic characterization approach is a classical matrix model connecting RGB and XYZ spaces, including a previous linearizing gamma correction of RGB values. This simplified color characterization model is widely used in color management software. In this case, the most appropriate color space will be the tristimulus space associated with the CIE 1931 standard observer; the chromatic characterization approach is a classical matrix model connecting RGB and XYZ spaces, as shown in Eqn. 2, including a previous linearizing gamma correction from Eqn. 1.

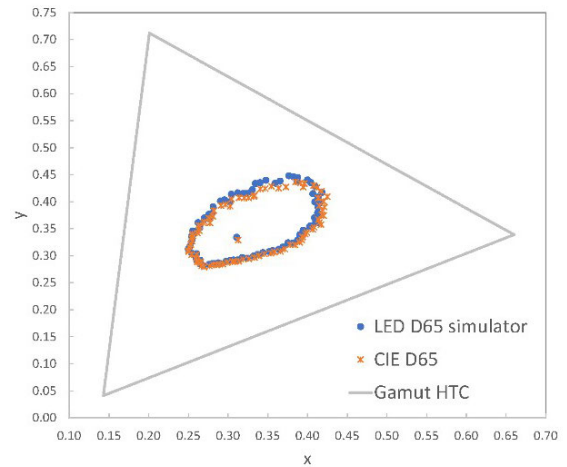
$$\begin{aligned} R' &= R^{\gamma_1} \\ G' &= G^{\gamma_2} \\ B' &= B^{\gamma_3} \end{aligned} \quad (1)$$

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} X_{R'max} & X_{G'max} & X_{B'max} \\ Y_{R'max} & Y_{G'max} & Y_{B'max} \\ Z_{R'max} & Z_{G'max} & Z_{B'max} \end{pmatrix} * \begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} \quad (2)$$

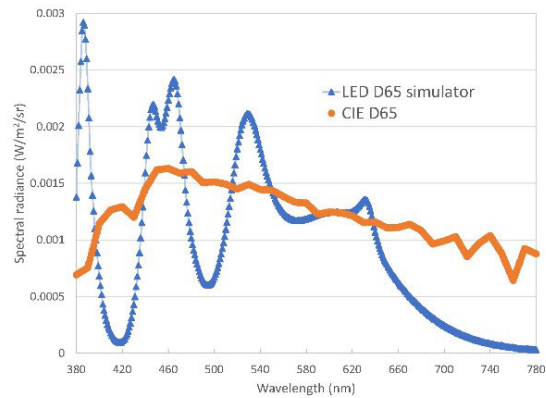
**C. FM 100 TEST: SPECTRORADIOMETRIC MEASUREMENTS**

In 1943, Deanne Farnsworth proposed the ‘Farnsworth-Munsell 100 - Hue’ test starting from 100 Munsell samples on matte paper, all of which had the same values of chroma and value (5/5) but had different tones (hues) varying in step constants from the perceptive point of view [23]. Subsequently, the test samples were reduced to 85, although the name of the test was maintained at FM 100. The FM 100 test is a sorting test with which two objectives can be achieved: distinguishing among people with normal color vision and those who have a high and low color discrimination ability and identifying different types of defective color vision. The FM 100 test has also been the subject of numerous specific scientific investigations over the past 50 years [25]–[30] and is a valid and well-known reference for any color vision laboratory.

To implement the virtual version of this test in a virtual reality environment, we performed 3 independent spectroradiometric measurements for each of the 85 samples in a completely new FM 100 test. Specifically, the spectral reflectance of each sample and the CIE 1931 XYZ tristimulus values of each sample were measured. A Konica-Minolta CS-2000 spectroradiometer was employed. The spectroradiometer was inclined 45° to the horizontal, and a Spectralon diffuse reflectance pattern (LabSphere, USA) was used as a reference to measure spectral reflectance. All measurements were made using a D65 simulator as the light source, with which the LED viewing light booth was equipped (Just Normlicht, Germany). Fig. 7 shows the CIE 1931 chromaticity coordinates of each of the 85 samples of the FM 100 test under the CIE D65 theoretical illuminator and under simulator D65 of



**FIGURE 7. Chromaticity of 85 FM 100 samples illuminated with CIE D65 illuminant (orange star) or LED D65 simulator (blue circle).**



**FIGURE 8. Spectral power distribution of the light sources employed in this work adjusted to a luminance value of 100 Cd/m2: CIE D65 (orange circles) and LED light booth D65 simulator (blue triangles).**

the light booth. Both power spectral distributions are shown in Fig. 8.

**D. VIRTUAL SCENE**

A virtual scene was created using the Unity Game Engine software platform in the software version 2019.1.5, which allows 3D scenes to be generated and displayed on virtual reality devices through their corresponding rendering. The scene simulates the Just NormLicht LED, available in our laboratory equipment (Fig. 9). This light booth is equipped with 12 LED spotlights and has been simulated both in physical and lighting geometry.

We simulated each of the 85 samples of the FM 100 test within this light booth. Each sample consists of a support part made of black Bakelite and another flat part, where the chromatic sample is located. The color of each sample was defined as a texture associated with a virtual material created in Unity, which allowed us to assign an RGB color to each sample.

To calculate the RGB color of each of the samples, a script that computes the tristimulus values of the FM 100 test

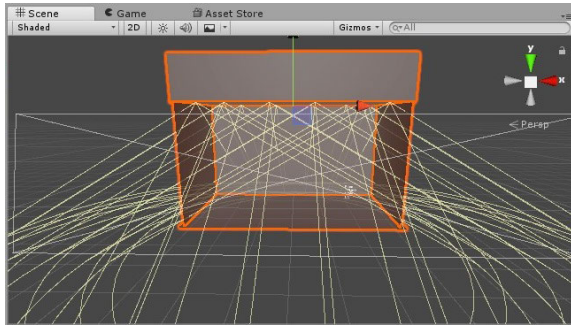


FIGURE 9. Virtual light booth designed in this work.

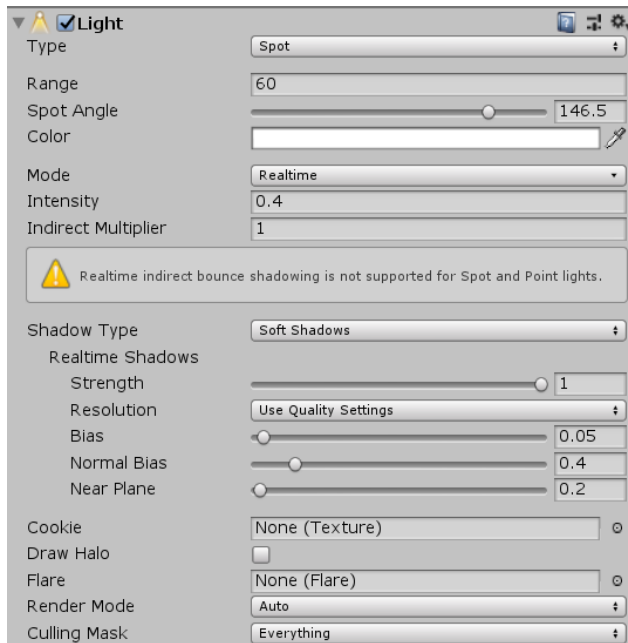


FIGURE 10. Spotlight settings.

samples under the SPD of the source used in the light booth was made. Subsequently, the corresponding RGB values were obtained through our colorimetric calculation module for VR.

We also measured every detail of the actual light booth. Details such as the angle of incidence of the lights or the separation between spotlights have been applied in the virtual scene. Next, in Fig. 10, an image of the lighting options is used to simulate the light booth.

Although the original recommendation was to use the C illuminant to perform the test, the CIE currently recommends using the D65 illuminant. However, it is not possible to use the D65 illuminant because of its theoretical nature. Many different D65 illuminant simulators exist on the market. A script has been introduced in the developed software that allows us to perform calculations on the RGB values for any light source so that it is possible to study the behavior of this test under different D65 simulators [31].

### E. GRAPHIC SETTINGS

As we have described in the introduction section, the main objective of this work is to study the validity of commercial

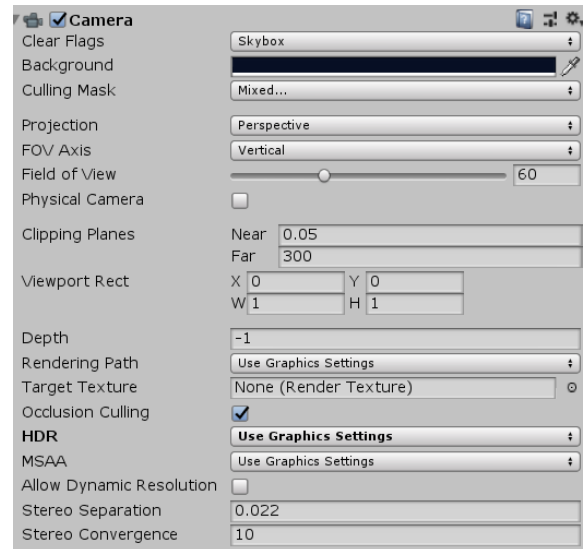


FIGURE 11. VR camera settings used.

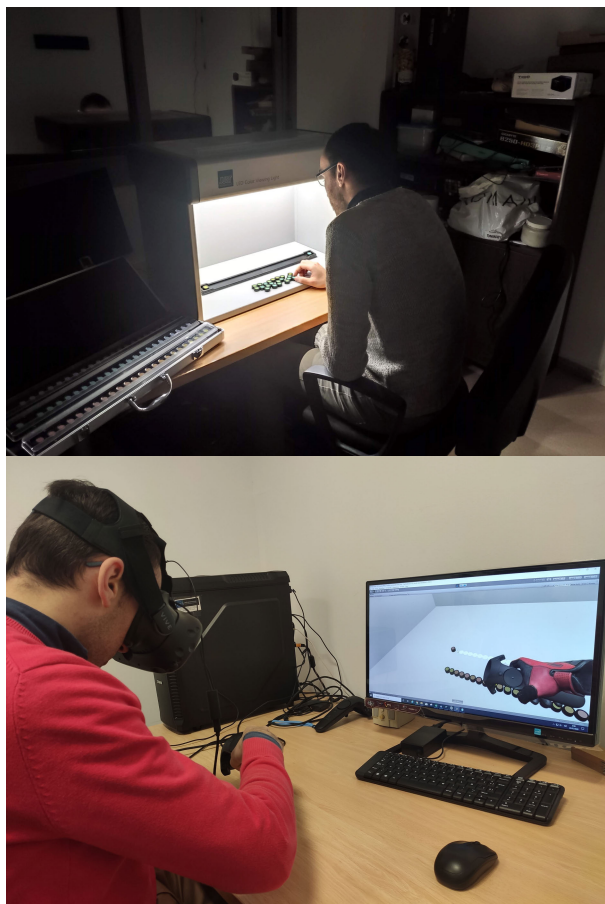
virtual reality systems to be used in professional tasks in color vision and lighting laboratories. From this perspective, our priority has been to carry out a visual impairment test with the greatest fidelity to reality, using hyperspectral textures as a new method for this purpose but without excessively penalizing the rendering and response time that virtual reality requires. For this reason, the standard render pipeline with deferred rendering has been used in our VR scenario. A previous work demonstrated the effectiveness of this configuration [12]. Light treatment is based on real-time lighting without baked global illumination light maps. Because of the simplicity of our scenario and the fact that there is no metallic material or transparency, we employ only the diffuse component of light reflected at the material surface (albedo color in Unity Terms). We did not include reflection probes or activate the HDR mode.

For the VR camera settings, we used a skybox with a black background. We have also performed the recommended VR default settings of field of view and stereo separation and convergence. The values used for the VR camera settings are shown in Fig. 11.

### F. OBSERVERS AND PROCEDURE

The population sample was composed of 20 observers (14 men and 6 women) aged between 20 and 56 years old. Three of them demonstrated color vision deficiencies by means of other color blindness validation tests, such as an Ishihara test. Those observers who need glasses in their daily life used the glasses for both the physical and virtual tests. Each observer answered the test five times randomly in different sessions. The test requires two parts: one corresponding to the physical test and another corresponding to the virtual version. All sessions were carried out on different days and by varying the order randomly.

The methodology employed with the physical test (Fig. 12 above) follows the recommendations of the original author of



**FIGURE 12.** (Top) A user’s procedure when taking the physical test. (Bottom) A user’s procedure when taking the virtual test.

the FM 100 test [23]. The methodology applied to the virtual test follows this sequence: first, all color samples correctly ordered are shown to the observer (Fig. 14 above). Subsequently, the test starts only showing the samples belonging to a single row. This row is chosen randomly, leaving the first and last caps at a fixed position in the same way that the physical test is applied (Fig. 14 bottom). The observer must place all the samples in the order he/she considers correct with a remote control for the VR device (Fig. 12 bottom). There is an equivalent remote control to the real one in the virtual scene, and the observers can see this remote control controlled by a virtual hand. Observers can modify the position assigned to each sample at all the times they consider necessary. After finishing each row, the next row is shown, and at the end of the test, the score obtained and the time spent are reported.

**IV. RESULTS**

The results of this work can be divided into two parts. The first part is related to the fidelity of the color reproduction of the virtual scene in relation to the original real scene. The second part deals with the comparison of the results obtained by means of both tests (physical and virtual tests). The results are described next.

**TABLE 2.** Color difference values.

Row- Tablet	$\Delta E^*$
1- 1	1.76
1- 10	2.42
1- 20	1.73
2- 23	0.99
2- 32	1.38
2- 41	1.44
3- 46	1.18
3- 55	1.43
3- 62	1.36
4- 64	1.36
4- 75	1.60
4- 83	1.42
Mean Error (Standar Desviation)	1.5 (0.4)

**A. MEASURED COLOR DIFFERENCE BETWEEN THE PHYSICAL AND VIRTUAL TESTS**

The International Commission on Illumination (CIE) defines the metric distance  $\Delta E^*_{ab}$  (also called  $\Delta E^*$ ) as a basic color difference formula associated with the CIELAB color space standardized in 1976 [32].

There are known perceptual nonuniformities in the CIELAB color space; during recent decades, several new color difference formulas have been proposed and employed to correct these nonuniformities (as recommended by the CIE in 1994 and 2000 [33]). These nonuniformities are important because the human eye is more sensitive to certain colors than to others. A good metric should take this into account to associate the notion of a “just noticeable difference” (JND) to a unity of its metric in the whole color space. Otherwise, a certain  $\Delta E^*$  may be insignificant between two colors in one part of the color space while being significant in some other part. With the concept of JND in mind, CIEDE2000 color difference values between 1 and 2 units are considered close to imperceptible for the human visual system. Values between 2 and 3 are considered an accurate representation of the color. Above these values, the result is considered to be inaccurate.

To obtain the average color difference between the real and virtual scenes of this work, we measured 12 FM 100 test caps (3 from each row, at both ends and at the middle) of the physical test, and measured the equivalent caps at the virtual scene for the virtual test. Table 2 shows the color difference between each pair of caps (virtual and real). The results showed an average color difference of  $CIEDE2000 = 1.5$  and a standard deviation of 0.4. This value matches the chromatic characterization mean error limit and reflects the accurate behavior of our system in a faithful color reproduction of a real scene in a virtual scene.

In addition, in Fig. 13, we see the values of chromaticity CIE 1931 (x, y) in each of the 12 physical and virtual tables. The results of both measurements denote a clear matching between the two systems.

Finally, in Fig. 14, we see the appearance of the virtual test, while if we look at Fig. 15, we see the appearance of the physical test. Although measuring the general appearance

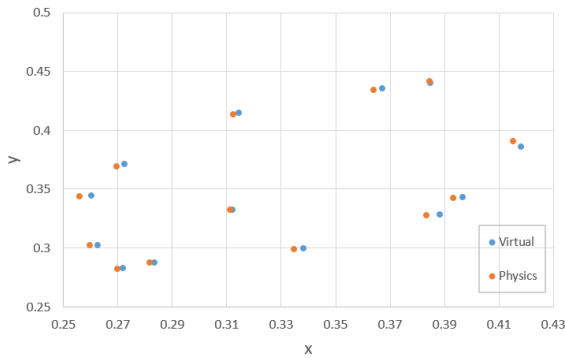


FIGURE 13. Chromaticity measurements (x,y) for the 12 FM 100 samples.



FIGURE 14. (Top) Initial aspect of the virtual reality application in which the 93 test caps are shown. (Bottom) Screen capture during the development of one of the tests in which the virtual interface is used to order the samples.

is not the main objective of this work, and we have only measured color values, we see that the appearance has many similarities.

**B. SCORES OBTAINED BY THE OBSERVERS**

As described in section II, the FM 100 test is a sorting test where a score is obtained according to the errors made in the sorting of color samples with equal saturation and clarity and different hues. With this score, it is possible to quantify how accurate the color discrimination ability of the person taking the test is.

Although it is possible to carry out a comparative study capsule by capsule, an accurate indicator of the correct functioning of the virtual test versus the physical test is the comparison of the errors obtained by the same observers in both tests. Since each of the tests was carried out 5 times, this comparison should be made with the average of the scores obtained in the 5 sessions.

Table 3 shows the results obtained by the 17 observers previously classified as normal observers and the three observers classified as defective. A first analysis of this data shows a high correlation between both scores: those corresponding to



FIGURE 15. (Top) Initial aspect of the real scene application in which the 93 test caps are shown. (Bottom) Screen capture during the development of one of the tests in which the virtual interface is used to order the samples.

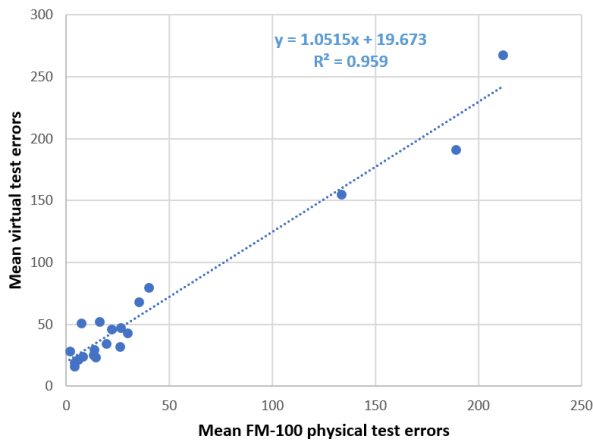
TABLE 3. Results obtained by the 17 observers previously classified as normal observers and the three observers classified as defective.

Observer	Age	Sex	Physical Test	Virtual Test
Normal 1	55	W	35.2	68
Normal 2	54	W	26.4	47.2
Normal 3	56	W	19.2	34.4
Normal 4	56	M	29.6	43.2
Normal 5	26	M	4	19.2
Normal 6	21	M	13	25.6
Normal 7	29	M	5.6	21.6
Normal 8	31	W	13.6	29.6
Normal 9	26	M	4	16
Normal 10	24	M	40	80
Normal 11	45	M	14	23.2
Normal 12	26	W	7.2	51.2
Normal 13	27	W	1.6	28.5
Normal 14	56	M	26	32
Normal 15	56	M	22	46
Normal 16	26	M	8	24
Normal 17	25	M	16	52
Defective 1	59	M	189	191
Defective 2	19	M	133.6	155
Defective3	25	M	212	268
Mean (defective excluded)			16.8	37.7
Standard deviation (defective Excluded)			11.6	17.8
Upper limit for normals			51.5	91.3
Pearson Coef.			0.98	

the physical test and those corresponding to the virtual test (Pearson = 0.98, Confidence level = 99%,  $p < 0.001$ ).

Moreover, the relationship between both scores is highly linear ( $R^2 = 0.96$ , Confidence Level = 99%,  $p$ -value  $< 0.001$ ) with a slope value very close to 1, and only an independent term of 19 points is noteworthy, which indicates that the virtual test, on average, obtained an error that was 19 points higher. Statistical data related with this linear model and its confidence interval are shown in Table 4. We can point out that the errors in the virtual reality tests are higher than those in the physical tests due to a higher complexity caused by the use of a technological device such as HMD. An influencing





**FIGURE 16.** The correlations between the errors of the physical and virtual test.

**TABLE 4.** Statistical data obtained from the linear regression applied between physical and virtual test results.

Linear model	Coefficients	Error	t	Sigma	Confidence Interval 99%	
					Lower	Upper
Constant	19.673	3.720	5.288	0.000	8.964	30.382
Slope	1.052	0.051	20.522	0.000	0.904	1.199

factor is the quality of the image, which is still not perfect due to several known effects, such as the screen door effect or the blurring effect. Regarding the differences in scale between normal and defective observers, we think that the lower quality of the virtual image compared to real scenes is the cause of the 19 points of difference, on average, at the independent term of the linear equation. This difference in score appears as a 4-fold difference at lower scores (better color observers) because it is relatively easy to change the order of four caps (16 points) instead of two caps (4 points) in the more difficult VR scenario. However, this factor is not related to the diagnostic criterion because this criterion depends, in our opinion, on the closeness to the slope factor of the linear equation to 1.

Figure 16 shows the graphical representation of the results obtained and the expression of the linear relationship model between both sets of results. This figure clearly shows the difference between the observers classified as defective and normal. If we look at purely statistical criteria, a non-normal observer is one that is outside the normal probability distribution, which is usually defined as an observer who obtains a result that is more than 3 standard deviations away from the mean value of the population when calculating that mean value. This is what is known as the 99.9% confidence level. If we perform these calculations for the results obtained with normal observers, we can see how this upper limit would be marked by a score of 52 points for the physical test and 92 for the virtual test. In both cases, observers previously classified as defective are those who are outside these limits and would again be classified as defective.

No statistically significant differences were obtained after analyzing the data obtained according to the age and sex of the participants.

## V. CONCLUSION

Virtual reality devices have been in constant evolution in recent years, allowing more diverse use. The general objective of this work was to contribute to the use of commercial virtual reality systems in professional tasks. Specifically, we have implemented a virtual version of a known arrangement test for color vision assessment, the FM 100 test, as an example of the use of this technology in research and diagnosis tasks at lighting and vision research labs. Our main objective was that the functionality of both tests (virtual and physical) would have the same behavior. The results obtained show that the functionality of both tests, the virtual and physical tests, are very similar. Furthermore, the diagnosis of both methods is equivalent. The scores obtained by real observers in both tests are slightly different in scale but have a clear linear relation. The small difference in scale does not affect the classification made of defective or normal observers and could be related to the greater difficulty in using an HMD as a display. The main limitation of this work lies in the dependence of a previous chromatic characterization of the virtual reality system (hardware and software). This limitation is identical for other types of electronic media used in research tasks (CRT, TFT, OLED). However, it is an advantage to use this technological support to perform color vision tests because there is no deterioration of the samples since the observers do not manipulate the samples. Another advantage is that we can control the illumination both in intensity and spatial distribution.

In view of the results, we conclude that this technology based on HMDs and virtual reality contents is valid for this research task. A future line of work could be to extend this type of generic tests based on virtual reality to other more specific types of tests such as visual evaluations carried out to obtain automobile driver or airline pilot licenses. This future work should be also oriented to deepen the improvement of the visual appearance of virtual 3D scenes. Despite the high quality of virtual reality systems, there is still room for progress both in the development of display devices and the improvement of the management of real-time rendering software. These improvements could expand the potential use of this virtual reality-based technology.

## ACKNOWLEDGMENT

The authors would like to thank the Centro Universitario de Mérida, University of Extremadura for the cession of its areas and materials used during this work.

## REFERENCES

- [1] P. Cavanagh, D. Maurer, T. Lewis, D. MacLoad, and G. Mather, "Computer-generated screening test for color blindness," *Color Res.*, vol. 11, pp. 63–66, Oct. 1986.
- [2] A. Hoffmann and M. Menozzi, "Applying the ishihara test to a PC-based screening system," *Displays*, vol. 20, no. 1, pp. 39–47, Feb. 1999.
- [3] O. K. Ozgur, T. S. Emborgo, M. B. Vieyra, R. F. Huselid, and R. Banik, "Validity and acceptance of color vision testing on smartphones," *J. Neuro-Ophthalmol.*, vol. 38, no. 1, pp. 13–16, 2018.
- [4] P. J. Pardo, A. L. Pérez, and M. I. Suero, "Validity of TFT-LCD displays for colour vision deficiency research and diagnosis," *Displays*, vol. 25, no. 4, pp. 159–163, Nov. 2004.

- [5] E. A. Cooper, H. Jiang, V. Vildavski, J. E. Farrell, and A. M. Norcia, "Assessment of oled displays for vision research," *J. Vis.*, vol. 13, no. 12, p. 16, 2013.
- [6] R. S. Berns, R. J. Motta, and M. E. Gorzynski, "CRT colorimetry. Part I: Theory and practice," *Color Res. Appl.*, vol. 18, no. 5, pp. 299–314, Oct. 1993.
- [7] J. A. Díaz, J. R. Jiménez, E. Hita, and L. J. del Barco, "Optimizing the constant-channel chromaticity and color gamut of CRT color displays by control of brightness and contrast levels," *Appl. Opt.*, vol. 35, no. 10, pp. 1711–1718, 1996.
- [8] K. Kam, Y. Tiantong, K. Behrman, Y. Caroline, and I. Kymissis, "57-3: Characterizing image retention for HDR OLED displays," in *SID Symp. Dig. Tech. Papers*, 2020, vol. 51, no. 1, pp. 850–853.
- [9] D. Tian, L. Xu, and M. Luo, "The characterization of HDR OLED display," *Electron. Imag.*, vol. 2019, pp. 326-1–326-6, 2019.
- [10] A. Almustanyir, J. Hovis, and M. G. Glaholt, "Predicting the Farnsworth–Munsell D15 and Holmes–Wright lantern outcomes with computer-based color vision tests," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 37, no. 4, p. A1, Apr. 2020. [Online]. Available: <http://josaa.osa.org/abstract.cfm?URI=josaa-37-4-A1>
- [11] J. S. Werner, B. Marsh-Armstrong, and K. Knoblauch, "Adaptive changes in color vision from long-term filter usage in anomalous but not normal trichromacy," *Current Biol.*, vol. 30, no. 15, pp. 3011–3015, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0960982220307417>
- [12] P. J. Pardo, M. I. Suero, and Á. L. Pérez, "Correlation between perception of color, shadows, and surface textures and the realism of a scene in virtual reality," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 35, no. 4, p. B130, 2018.
- [13] M. A. Cohen, T. L. Botch, and C. E. Robertson, "The limits of color awareness during active, real-world vision," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 24, pp. 13821–13827, Jun. 2020. [Online]. Available: <https://www.pnas.org/content/117/24/13821>
- [14] D. Kamińska, T. Sapiński, M. Kucharczyk-Pośpiech, and M. Wilczyński, "Ishihara color test using a mobile virtual reality headset for immobilised patients," *Roczniki Kolegium Analiz Ekonomicznych/Szkola Główna Handlowa*, pp. 297–304, 2017.
- [15] K. Szczerowski and M. Smith, "Emulating perceptual experience of color vision deficiency with virtual reality," *Stud. Health Technol. Informat.*, vol. 256, pp. 378–389, Jan. 2018.
- [16] J. Schanda, *Colorimetry: Understanding the CIE System*. Hoboken, NJ, USA: Wiley, 2007.
- [17] W. G. Backhaus, R. Kliegl, and J. S. Werner, *Color Vision: Perspectives From Different Disciplines*. Berlin, Germany: Walter de Gruyter, 2011.
- [18] J. J. Koenderink, "Virtual psychophysics," *Perception*, vol. 28, no. 6, pp. 669–674, Jun. 1999.
- [19] D. A. Gusev, D. M. Whittinghill, and J. Yong, "A simulator to study the effects of color and color blindness on motion sickness in virtual reality using head-mounted displays," in *Mobile and Wireless Technologies (Lecture Notes in Electrical Engineering)*, vol. 391. Singapore: Springer, 2016, pp. 197–204. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-981-10-1409-3\\_22](https://link.springer.com/chapter/10.1007/978-981-10-1409-3_22), doi: 10.1007/978-981-10-1409-3\_22.
- [20] A. F. Zarazaga, J. G. Vázquez, and V. P. Royo, "Review of the main colour vision clinical assessment tests," *Arch. Sociedad*, vol. 94, no. 1, pp. 25–32, Jan. 2019.
- [21] S. J. Dain, "Clinical colour vision tests," *Clin. Exp. Optometry*, vol. 87, nos. 4–5, pp. 276–293, Jul. 2004.
- [22] F. Díaz-Barrancas, H. Cwierz, P. J. Pardo, Á. L. Pérez, and M. I. Suero, "Spectral color management in virtual reality scenes," *Sensors*, vol. 20, no. 19, p. 5658, Oct. 2020.
- [23] D. Farnsworth, "The Farnsworth–Munsell 100-Hue and dichotomous tests for color vision," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 33, no. 10, pp. 568–578, Oct. 1943. [Online]. Available: <http://www.osapublishing.org/abstract.cfm?URI=josa-33-10-568>
- [24] F. Díaz-Barrancas, H. C. Cwierz, P. J. Pardo, A. L. Perez, and M. I. Suero, "Visual fidelity improvement in virtual reality through spectral textures applied to lighting simulations," *Electron. Imag.*, vol. 2020, no. 15, pp. 1–259, 2020.
- [25] P. R. Kinnear, "Proposals for scoring and assessing the 100-HUE test," *Vis. Res.*, vol. 10, no. 5, pp. 423–433, May 1970.
- [26] K. Knoblauch, F. Saunders, M. Kusuda, R. Hynes, M. Podgor, K. E. Higgins, and M. F. De Monasterio, "Age and illuminance effects in the Farnsworth–Munsell 100-Hue test," *Appl. Opt.*, vol. 26, no. 8, pp. 1441–1448, 1987.
- [27] M. E. Breton, D. E. Fletcher, and T. Krupin, "Influence of serial practice on Farnsworth–Munsell 100-Hue scores: The learning effect," *Appl. Opt.*, vol. 27, no. 6, pp. 1038–1044, 1988.
- [28] S. Ghose, D. Shrey, P. Venkatesh, T. Parmar, and S. Sharma, "A simple modification of the Farnsworth–Munsell 100-Hue test for much faster assessment of color vision," *Indian J. Ophthalmol.*, vol. 62, no. 6, p. 721, 2014.
- [29] M. García-Romera, M. Grosman, M. Sánchez-Marañín, L. Gómez-Robledo, and M. Melgosa, "On the uniformity of the 'Farnsworth–Munsell 100-Hue' test," *Optica Pura Aplicada*, vol. 44, pp. 139–148, Jan. 2011.
- [30] K. G. Foote, M. Neitz, and J. Neitz, "Comparison of the Richmond HRR 4th edition and Farnsworth–Munsell 100 Hue Test for quantitative assessment of tritan color deficiencies," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 31, no. 4, p. A186, 2014.
- [31] H. C. Cwierz, F. Díaz-Barrancas, P. J. Pardo, A. L. Perez, and M. I. Suero, "Application of spectral computing technics for color vision testing using virtual reality devices," *Electron. Imag.*, vol. 2020, no. 15, pp. 1–260, 2020.
- [32] *Colorimetry*, vol. 1, CIE5, Vienna, Austria, 2018. [Online]. Available: <https://cie.co.at/publications/colorimetry-4th-edition>
- [33] M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Res. Appl.*, vol. 26, no. 5, pp. 340–350, 2001.



**HALINA CWIERZ** received the B.S. degree in computer engineering and the M.S. degree in computer technologies research from the University of Extremadura (UEX), Mérida, Spain, in 1987 and 2016, respectively. She is currently pursuing the Ph.D. degree in the application of virtual reality devices to the screening of color vision deficiencies.

She is currently an Associate Professor at UEX. She has been involved in several projects related to the application of augmented reality to education. Her research interests include color vision and augmented and virtual reality devices.



**FRANCISCO DÍAZ-BARRANCAS** received the B.S. degree in computer engineering and the M.S. degree in computer technologies research from the University of Extremadura (UEX), Mérida, Spain, in 2017 and 2018, respectively. He is currently pursuing the Ph.D. degree in the visual appearance in virtual reality devices.

He is currently a Research Assistant at UEX. He has been involved in several projects related to the application of hyperspectral and virtual reality techniques to measure the color performance of LED light sources.



**JULIA GIL LLINÁS** is currently an Associate Professor at the University Centre of Mérida, University of Extremadura, Spain. Her current research interests include active and collaborative learning strategies as well as learner-centered instructional and learning strategies in the context of higher education in science.



**PEDRO J. PARDO** received the B.S. degree in physical sciences and the M.Sc. and Ph.D. degrees in physics from the University of Extremadura (UEX), Mérida, Spain, in 1998, 2000, and 2004, respectively.

He is currently an Associate Professor at UEX. He has been involved in several projects related to the detection and influence of visual anomalies in students. His research interests include color vision, optical instrumentation, neural networks, computer networks, and physics education.

• • •