Article

3D22\textsuperscript{MX}: Performance Subjective Evaluation of 3D/Stereoscopic Image Processing and Analysis

Jesús Jaime Moreno Escobar *, Erika Yolanda Aguilar del Villar, Oswaldo Morales Matamoros and Liliana Chanona Hernández

Escuela Superior de Ingeniería Mecánica y Eléctrica, Unidad Zacatenco, Instituto Politécnico Nacional, Ciudad de México 07340, Mexico

Abstract: This work is divided into three parts: (i) a methodology developed for building a 3D/stereoscopic database, called 3D22\textsuperscript{MX}, (ii) a software tool designed for degradation of 3D/stereoscopic images, and (iii) a psychophysical experiment carried out for a specific type of noise. The novelty of this work is to integrate these three parts precisely to provide not only professionals who design algorithms to estimate three-dimensional image quality but also those who wish to generate new image databases. For the development of the 3D/stereoscopic database, 15 indoor images and 5 outdoor ones were spatial-calibrated and lighted for different types of scenarios. Criteria calibration is different for indoor images with respect to outdoor images. The software tool to degrade 3D/stereoscopic images is designed from MatLab programming language since images captured in the first part are calculated for achieving several image degradation. Our program has ten different types of noises for degradation, such as white Gauss impulse, localvar, spatial correlation, salt & pepper, speckle, blur, contrast, jpeg, and j2k. Due to each type of noise containing up to five levels of degradation, in this proposal, a database of 20 images is required to design a tool for degrading and generating three-dimensional images ranging from all types of noise to yield psychophysical. Finally, there are applied specific criteria to carry out some psychophysical experiments with 3D/stereoscopic images. Moreover, we analyzed the methodology used to qualify and apply images to the j2k noise, explaining every degradation level for this noise.

Keywords: performance evaluation; image analysis; 3D/stereoscopic; image databases; psychophysical experiments; image processing

MSC: 94A08

1. Introduction

The visual quality of images is an important factor in various digital processing applications. For researchers of 3D-image projections, it is not yet clear how the human eye captures the quality of a 3D/Stereo or three-dimensional image because it depends on the variability of the observer’s perception, and the place where this animation is projected has different subjective-quality parameters. 3D projections are not accepted as many critics and artists have labeled them as impossible to see. 3D-image induce nausea and headaches due to distortions and poor 3D-post-production quality. Although the visual quality of 3D images has been improved by designed algorithms, many of these algorithms have not been tested to treat 3D images because both researchers nor programmers have all the main tools for experimentation. Therefore, it is imperative to make a database containing original or source images along with the kind of distortion applied to original ones [1,2].

On the one hand, researchers focused on creating algorithms for processing 3D images do not have means for making psychophysical experiments applied to source images to test their three-dimensional image quality assessments or algorithms. On the other hand,
3D imaging has a wide area of research driven by the entertainment industry as well as television, cinema, video games, and an area of scientific and medical research. For instance, 3D medical imaging techniques are aimed at showing the internal structure of organs via many scanned images that are processed by computer algorithms to create volumetric three-dimensional models, Figure 1a.

![Figure 1a](image1.png)

**Figure 1a.** Medical imaging techniques for 3D visualization of internal structures.

Both academics and practitioners dedicate a great effort to developing objective metrics capable of quantitatively evaluating the amount of degradation suffered by a signal, an image, or a stereoscopic or volumetric video sequence. Despite several methods of quality assessment that have been proposed in imaging literature, there has been any effort to the quality assessment of 3D/stereo or three-dimensional images. With widespread 3D technology applied to many different, e.g., entertainment and applied medicine, 3D images and videos must be processed, Figure 1b.

![Figure 1b](image2.png)

**Figure 1b.** Usage of 3D-imaging in scientific and medical research. (a) Telesurgery and (b) volumetric inspection.

According to the above, it is important to develop and define subjective metrics, since, in a psychophysical experiment, the human-being response varies according to their preferences, culture, or age, and objective metrics to assess the 3D/Quality of the processed stereo images becomes a matter of extremely important [3–5].

With respect to 2D-images, the perception of stereo content involves several peculiar elements that are excluded when it comes to the evaluation of 2D content. Hence, this work aims to deal not only with the basic definition of quality but also with an extension of it called Quality of Experience (QoE), since 3D involves perception factors such as the feeling of immersion, presence, and reality tangible to human eye [6–8].

The quality of the image is no longer sufficient to represent QoE made by an observer immersed in a three-dimensional environment. Previous studies about Subjective evaluation of stereoscopic-image quality, realized by Moorthy et al. in [3] report distortion effects such as blocking or blurring. In addition, perception and constancy of quality in the field of computer vision are investigated as well as the impact of depth representation, their data formats and algorithms of compression used.

Thereby, it is essential to develop databases containing not only different images with added noises but also Most Opinion Scores (MOS) from viewers to assess the quality of each 3D/Stereoscopic-image subjectively. From this test, any Image Quality Assessment (IQA) performs a projection in a 3D/Stereoscopic-image in an appropriate way with its own methodology and algorithm. The problem complexity, together with the incipient understanding of 3D perception and the increasing commercial transfer towards 3D entertainment, makes an area of 3D/stereoscopic-image quality control interesting, formidable and relevant for a group of researchers. 3D-image processing and its understanding are very important topics as there is a lot of technical research behind them. The field of 3D-quality control is still extremely interesting, and it offers a huge opportunity for researchers.
Nonetheless, there are still important gaps to understand the perception of stereoscopic distortions and the appropriate statistical models for 3D-natural scenes for processing 3D scenarios. Databases for the verification of QoE Pure-3D metrics are designed rarely since many of these metrics are aimed at 2D-image quality assessments.

Extensive experimental results are the fundamental work for researchers who design both objective and subjective IQAs and, thus, improve viewer experience with respect to 3D/stereoscopic images. Our 3D22MX database has the following features: 3D-imaging, the year and the place where it was carried out, and the database contains 15 indoor and 5 outdoor images out of the 20 scenarios to be evaluated by human observers. The average results for each of the three-dimensional images or Stereo Pair (SP) processed can be considered as the 3D-MOS of each SP. In addition, in this proposal, we design a corpus for evaluating three-dimensional images in High-Definition (3D-HD) to support the creation of new algorithms for the treatment of the images in 3D/Stereoscopic. Moreover, this project includes the usage of advanced calibrating tools such as color checker passport, for instance, along with 10 types of noise like gaussian white noise, localvar, spatial correlation, salt, and pepper, peppered, blur, contrast, JPEG, and JPEG2000 with five-degradation levels for every noise.

This work is divided in Section 6 sections. In Section 2 we describe the main stereoscopic-image databases such as IRCCyN/IVC, NINGBO, and LIVE 3D IQA. In Section 3 Human Stereoscopic Vision, Noises for Image Distortion and Conditions of a 3D-Scenario are explained. In Section 4 methodology of this proposal is divided into three main phases: (i) Capture, (ii) Coding, and (iii) Representation. In Section 5 we show the experimental results in addition to the comparison with a state-of-the-art set of 3D-IQAs. Finally, in Section 6 our main findings, conclusions, and future work of our proposal are exposed.

2. Related Work

In this section, different current 3D/stereoscopic databases are described with their features, such as the number of images, types of distortions, levels of distortions, availability to the public, number of people who were used to perform the visual-experience experiments, as well as the methodology and images used. These three different current 3D/stereoscopic databases are

- IRCCyN/IVC [9],
- NINGBO [5], and
- LIVE 3D IQA [3].

IRCCyN/IVC Image-Database is developed at both University of Nantes and the University of Rome in Italy by Benoit et al. in [9].

Subjective tests are performed using six-stereo images, Figure 2. Five levels of noise degradation are considered for each image. In this case, interesting results are shown when considering image distortions such as blur compression, JPEG and JPEG2000 applied symmetrically to stereoscopic images, with the following features:

- JPEG: Compression rate ranging from 0.244 bpp to 1.3 bpp.
- JPEG2000: Compression rate ranging from 0.16 bpp to 0.71 bpp.
- Sixty-degraded images in addition to six-original or source images.
- Average image size is 512 × 448 pixels viewed in standard resolution (without scaling, centered on the screen) at a resolution of 1024 × 768 on a 21” Samsung Sync Master 1100 MB TV.
- Test is carried out based on recommendations of the ITU BT 500-11 [10], similar to the ITU BT 500-10 standard, containing aspects such as:
  - Monitor resolution.
  - Monitor contrast.
  - Source of the signals.
  - Selection of materials for the subjective study exam.
  - Range of conditions.
• Observer:
  – Examination session.
  – Presentation of results.

Figure 2. IRCCyN/IVC Image Database.

For the pre-evaluation of the main assessment, the quality of the original stereoscopic image (reference image), hidden reference, and seven degraded versions were controlled on a continuous scale from 0 to 100. Each distorted image is collected in a random order. Each observer writes down the sixty-six images available in the test. Subjective experiments lead to ninety values of DMOS. Hence, the main objective of this database is to introduce an objective quality metric for the evaluation of the stereo-image quality. This proposed metric is based on both the use of 2D metrics and depth information. It is demonstrated that Structural Similarity Index Metrics (SSIM) are improved by adding the disparity distortion contribution. This fact is perfectly correlated with the evaluation of the luminance, contrast, and structure criterion of the original SSIM, but this is not enough to assess the quality from a perceptual point of view. The usage of disparity information provides an improvement on the original metric. The disparity estimation algorithm demonstrates that algorithms based on belief propagation are more efficient than methods based on other features. Indeed, sharper disparity maps from belief propagation are more correlated with subjective quality metrics than smoother disparity maps [11].

Working at the University of Ningbo in China, Wang et al. proposed in [5] a database containing 10 original images. In those images were applied the following distortions: JPEG, JPEG2000, Gaussian Blur, Gaussian Blur and white noise with 10 degradation-levels for having a total of 400 stereoscopic images distorted to be evaluated by 20 human observers, Figure 3.

The subjective experiment evaluated the quality of distorted stereoscopic-images with respect to the original ones. The Double Stimulus Continuous Quality Scale (DSCQS) method is followed for stereoscopic-image evaluation. In the cyclical-DSCQS method, evaluators viewed an image-pair about the same image, i.e., the original compressed and uncompressed ones, and were asked to evaluate the quality of both. After sessions, average scores for each test and image condition were calculated. For the subjective experiment, 20 subjects were recruited from the population of Ningbo University for each session, whose ages ranged from 20 to 25 years. All volunteers of the experiment met the minimum criteria for vision acuity of 20:30, stereo acuity at 40 arcseconds, and passed the color vision test. Neither all of them were directly related to image quality nor were experienced, image evaluators. Participants were unaware of the experiment’s purpose, or that one of the stereoscopic images was uncompressed. Before starting the experiment the subjects
received instructions and full-fill a practical test of the stereo screen. This stereo screen
contains four sets of stereo-images viewed and graded in the same way as for the essay’s
ones, but the latter are not included in the experimental analysis. Observers completed
all 100 experimental trials for each session with their best accuracy for judging the image
quality. To visualize the screen, tinted glasses are used to separate the left and right images
for each eye. The experiment was conducted in a dark room, with constant minimum-light
levels, and it lasted fifty minutes, including small pauses after each image type.

![Figure 3. Ningbo image database.](image3)

Lastly, at the University of Austin, Texas in, USA, Moorthy et al. in [3] developed
another database, incorporating symmetric distortions and covering a wider range of the
distortions. This database, called LIVE 3D IQA, is freely available for research purposes,
Figure 4, and it consists of 20 original images with 365 distorted images with JPEG200 or
J2K noise, JPEG, Gaussian Blur, Gaussian Blur with White Noise and Fast Fading.

![Figure 4. LIVE 3D IQA Image Database.](image4)

The creation of stereoscopic-content are based on (i) parallel and (ii) convergence
settings. In the parallel configuration, two-calibrated cameras separated by a fixed distance
are mounted on one piece of equipment, and the signal-pair acquired is known as a stereoscopic pair. The camera arrangement is a parallel baseline, i.e., the axes (of the lens) of the two cameras are parallel to each other. On contrary, in the convergence configuration, the axes (of the lens) intersect each other.

A continuous-assessment of the quality of a Single Stimulus Continuous Quality Evaluation (SSCQE) with a hidden reference study was carried out at the University of Texas at Austin, over the course of two weeks. The group of subjects consisted of 32 students mostly from the first semesters of this university. The subjects were a mixture of men and women, but the observers were mostly men. This study included two-viewing sessions lasting less than 30 min to minimize fatigue issues; the average test time was approximately 22 min. Informal feedback after the conducted study indicated that observers were able to perceive the stereoscopic signals well without discomfort or fatigue. Each image was displayed on a screen for 8 s. Each session started with a short training module to show observers six stereoscopic signals chosen to cover the range of distortions to see. The signals used for training are different from ours. Then, this study consisted in displaying the set of images in random order. Order was randomly assigned according to each subject and each session. A lot of care was taken to ensure that two consecutive sequences did not belong to the same reference, to minimize memory effects. Images were displayed on a 22-inch TV IZ3D passive stereoscopic display with a screen resolution of 800 × 600. The study was designed to warrant that each image received a score from 17 subjects, and scores given by the subject to distorted signal were subtracted from the score given by the same observer to reference signal to generate a Differential Most Opinion Score (DMOS) [12].

3. Materials and Methods

This section deals with the definition of Human Stereoscopic Vision and describes the noise types that cause discomfort to the human eye when images are perceived. Moreover, a scenario for 3D-image capture is discussed together with color checker passport tool.

3.1. Human Stereoscopic Vision

Stereoscopic vision allows observing of the 3D/Stereoscopic-images captured by the camera. A stereoscopic vision makes human-being able to integrate two images perceived into one to make it possible for the brain to analyze the data received from the two eyes and generates a unique three-dimensional image [13]. The basic principle is to have two images at different angles of separation to generate Stereoscopic Vision when processing images, Figure 5.

![Figure 5. 3D/Stereoscopic Image or Stereo Pair.](image)

A perspective image of the same object is formed on both retinas, differing from each other due to distinct positions taken by each point of view during the vision process, producing the relief effect. Distance between these two points of view, i.e., the separation between the receiving organs of the human being, has an average value of 65 mm and is called the interpupillary distance. In this vision process, each of the optical axes of the
lenses rotates inside their orbit until their directions intersect the point P in question; the generated function is called convergence, such that the images $P'$ and $P''$ are formed in the macula lutea existing in both retinas.

### 3.2. Noises for Image Distortion

We define noise as everything in a signal or information that lacks interest, degrading, distorting, and preventing the observation of the original information. There are several noise types that can be generated by default or by the programs, such as White Gaussian, Impulsive, Localvar, Spatial Correlation, Salt and Pepper, Speckle, Gaussian Blur, Contrast, JPEG, and JPEG2000 [14,15].

First, White Gaussian Noise (WGN) is defined as a stochastic process with time $t$ as an independent variable and $X_t$ as a random-dependent variable. Then a stochastic process can be interpreted as a succession of random variables whose characteristics can vary over time with mean zero, constant variance, covariance to zero, and a normal distribution [16]. In Equation (1), we define WGN as a stochastic process $X(t)$ for arbitrary coefficients $a_i$ $i$ times in a $N$ number of terms or signals as follows:

$$\sum_{i=1}^{N} a_i X(t_i) \sim \mathcal{N}(\mu, \sigma^2)$$  

(1)

In a similar manner, the function that describes Impulsive Noise $g(x, y)$ can be defined as a value taken by a pixel having no relation to its original value, but taking very high or very low values, i.e., almost white or black dots. Mathematically this noise is modeled with a non-Gaussian distribution function or step. Sometimes this noise is generated by digital or even analog transmissions [17]. Sometimes, the noise generated by digital (or even analog) transmission is impulsive. Hence, it can be modeled as

$$g(x, y) = (1 - p) \times f(x, y) + p \times i(x, y)$$  

(2)

where $i(x, y)$ is the impulse noise and $p$ belongs to $\{0, 1\}$.

In this way, Localvar Noise is generated by the Gaussian white noise with constant mean and variance, and it is dependent on the pixel-value intensity [18]. Namely, this kind of noise has a $\mu(x, y)$ for zero-mean Gaussian white noise with an intensity-dependent variance $\sigma^2(x, y) = \frac{\sum_{i=1}^{N} (p(x,y) - \mu(x,y))^2}{N}$, where $p(x,y)$ is the Intensity value of the $i$th pixel and $N$ the number of $p(x,y)$; along the image $\mathcal{I}$.

Furthermore, another well-known distortion is Spatial Correlation Noise (SCN), which is a digital image $\mathcal{I}$ represented by a function $\mathcal{I} : \Omega \rightarrow \Phi$, where both the domain of spatial coordinates, $\Omega$, and the set of values, $\Phi$, are finite and discrete. We will call each element of $\Omega$ as the position and each element of $\Omega$ as the Intensity channel, as long as it is one-dimensional, i.e., when referring to a gray-scale image. Otherwise, we will say that each element of $\Phi$ is a color image due to in a digital-color image, this value encodes the color in such a way that it can be interpreted and performed by some device, methodology, or algorithm. One way to encode color-image is to reference a certain color space so that the values of $\Phi$ represent coordinates within that space. So, it is possible to calculate a distance among colors in the color space where images are represented [19].

In addition, measuring SNC as a degree of association between colors based on their spatial distribution in the image is what is called spatial correlation between colors. Namely, two colors are said to be spatially correlated if they appear very close together in the image with respect to the other color pairs. Here we must highlight two features: (i) the measure of spatial correlation between colors is relative to a particular image, and (ii) it can be understood from a fully statistical point of view, and it is applicable to any digital signal or time series. Then, to calculate the spatial correlation between each pair of colors in an image, it is necessary to perform an implicit task: identify all possible color pairs. Many times it will not be possible to identify all the color pairs, because the number of ways in
which they can be combined is very large, and it implies a problem of processing capacity. For example, for a typical color image in RGB format, the Red, Green, and Blue channels are commonly digitized with 8 bits each, that is $256^3 = 16,777,216$ possible colors; therefore, we would have $\frac{256^6}{2}$ unordered pairs of colors.

In a similar way, Salt and Pepper noise consists of randomly generating black and white dots with a noise intensity captured as a parameter. It includes the spikes, glitches (spike pulses) and chirps (pulses with a certain structure) [20] and it is defined as follows:

$$X_{\text{abs}}(t) = X_{\text{in}}(t) + A_n \sum_{n=0}^{N_{\text{glitch}}} \delta(t - T_n)$$

where $\delta(t)$ is a Dirac Delta function and it is $\delta(t) = 1$ when $t = 0$, otherwise $\delta(t) = 0$ when $t \neq 0$. In addition, $T_n$ are the random values in the pixel domain, $A_n$ is the amplitude of the pixel, and $N_{\text{glitch}}$ is the number of impulses present and is a measure of the amount of impulse noise. If the original image has $N$ samples, it disappears completely when $N_{\text{glitch}} = N$. Each output value or $X_{\text{in}}(t) + A_n$ corresponds to either the maximum saturation or minimum value in the domain of the pixel, i.e., maximum intensity or salt noise, and minimum intensity or pepper noise.

Moreover, Speckle noise appears in images with coherent illumination, e.g., ultrasonic scanner, sonar, and Synthetic Aperture Radar (SAR). This noise is added to the image multiplication by means of an equation $J = I + (n \times I)$, where $n$ denotes the random noise of a uniform distribution with a variance $\nu = 0.04$ [21], as default value but this value can be adapted to the application of this kind of noise.

Gaussian Blur or Gaussian Smoothing Noise (GBN) is the result of blurring an image by a Gaussian function. It is a widely-used effect in graphic software, typically to reduce image noise and details. The visual effect of this blurring technique is a soft flash, like viewing the image through a translucent screen, distinctly different from the broken effect produced by an out-of-focus lens or the object shadow under usual lighting. Gaussian blur is also used as a pre-processing stage in computer-vision algorithms for improving the structures of an image at different scales. In addition, this kind of image noise has the effect of reducing the image’s high-frequency components; therefore, it can be considered as a low-pass filter [22]. GBN can be defined in a one-dimensional vector as a one-way Gaussian function; in the case of an image, it is important to realize that it is a two-dimensional matrix with $x$ columns, $y$ rows, and a variance $\sigma^2$, which is defined as follows:

$$\text{GBN}(x,y) = \frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Contrast Noise refers to the dispersion of gray levels in the image. Although other mathematical definitions of Contrast can be found, all of them are dispersion measures. An image with low contrast indicates a little variability of gray levels in the image, and its effect is shown in a very concentrated histogram with a small dynamic range. Dynamic range is the variation of gray levels in the image, pointing out the degree of gray dispersion in the image. If there were only one gray level, the energy would be a maximum and equal one. The higher the number of existing gray levels in the image, the lower the energy. Equations (5) and (6), respectively, demonstrate that the inverse function to $E(I)$ is $H(I)$ [23].

$$E(I) = \sum_{i=0}^{I-1} p(i)^2$$

$$H(I) = \sum_{i=0}^{I-1} p(i) \cdot \log(p(i))$$
An image is saturated when its histogram presents very high values at the extremes of the dynamic range; the histogram has a U shape. Images with low contrast or saturation have a loss of information in their acquisitions. The solution is in a new formation of the scene with different values of the capture parameters.

On the other hand, JPEG (Joint Photographic Experts Group) is an image-compression format developed in October 1996. This format is supported by most programs working with images, and it is a suitable format for storing photographic images because they take up little storage space when compressing images. Among other techniques, JPEG takes advantage of the deficiencies of human vision to eliminate information and produce a smaller storage space; hence it performs compression of lost. Then, the compression degree and the loss of quality are adjustable. The higher the compression level, the greater the quality loss, but the smaller the file size, and vice versa. Every time an image in JPEG format is opened with an image editor and saved again, a loss of quality occurs. Therefore, it is advisable to convert the photos to a lossless format before retouching them and return to JPEG when it is wanted to store the recovered image [24]. JPEG format stores an image using an algorithm that makes a trace of each pixel line and decides about it. When it finds two contiguous pixels on a line very similar in their colorimetry values, it decides what pixel will be saved and which one will be deleted. When the image is opened again, one pixel that was very similar to another will no longer be and will be in place only another recovered or equal from the original image.

Finally, JPEG2000 (J2K) is an image compression standard and encoding system. It compresses very better at low bitrates, less than 0.25 bits/pixel (bpp) for gray-scale images. The JP2K compression standard is used aided by the discrete-wavelet transform inside its frequency transformation instead of the Discrete Cosine one. J2K decomposes the image into a multi-resolution representation during the compression process. In addition, this algorithm provides big sets of data streams decompressed progressively both by pixel accuracy and image resolution or by image size. Therefore, viewers can see a lower-quality version of the final image, even before the full content of the file has been received. As the file is downloading progressively, the visual quality is also progressively improving. J2K code streams offer various mechanisms of random-spatial access, or to Regions of Interest (RoI), with several granularity degrees. Hence, it is possible to store distinct parts of the same image with different qualities. This entropy coder uses two types of file formats: (i) .jp2 and (ii) .jpx, allowing manipulation of color-space information and metadata as well as facilitating interoperability in network applications to have a full-embedded coding system [25].

3.3. Scenario for 3D Image Capture

Image Databases are made-up of different scenarios and objects necessary to take images for the creation of these databases. In the case of outdoor scenarios, it is avoided the presence of people in the place during their capture to make a framed shot without movement. Muslin is the name given to the green, blue, red, or white background fabric that is placed in the background for photo and video shoots. Figure 6 shows a Chroma-Key Green Muslin that is currently used more than any other color because image sensors in digital video cameras are more sensitive to green. Therefore, the green camera channel contains lower noise and needs less light for been illuminated.
To verify the homogeneity of the green channel in Muslin Fabric, a tool called Color Checker Passport (CCP) is used. CCP serves to achieve a color profile with the camera’s chromatic response [26], with a certain objective and under specific light conditions, obtaining the maximum color accuracy, taking advantage of the maximum reproduction range that our camera is capable of capturing. CCP contains the following three sections or charts:

1. Figure 7a depicts a classic CCP containing 24-color patches when combined with the camera-calibration software to produce DNG profiles, to respond to the lighting of the scene to achieve consistency, forecasting respectable results from image-to-image and camera-to-camera. This classic chart provides a visual color reference point, where each 24-color patch represents colors of natural objects, such as sky blue, skin tones, and leaf green; moreover, each patch reflects light as it is in its real equivalence. Moreover, each patch is individually colored using a solid shade to produce pure, flat, rich colors without smudges or mixing of dyes. In addition, this chart helps correct globally based on accurate information.

2. Figure 7b shows the White-Balance chart, ensuring that the color captured is real and provides a reference point for post-capture editing. This chart is spectrally flat, facilitating a neutral reference through different lighting conditions encountered during the photographs capture. Light is reflected equally through the visible spectrum, creating a customized white balance in the camera to compensate correctly for the lighting variation. The White Balance chart allows one to eliminate the color tones and improve color preview on any screen for getting a more reliable histogram, producing faster color editing in post-production.

3. Figure 7c shows the CCP Creative Improvement chart generated for high-level color creativity and workflow control. The enhancement chart includes four lines of colored patches designed for image editing. Whether shooting in a studio or in colorful nature, or in multiple scenes of photography events, CCP Creative Improvement is able to expand the powerful-photo editing software into a virtual Raw processing software. When cropping is needed, the improvement card highlights working in Raw. A cropped-patches line from the beginning to the end serves as a visual reference for judging, controlling, and editing images, to highlight shadow or crop details. Despite shadow or highlight details have been lost because the processing software has cropped them, they are still available in Raw file, and with a few adjustments, CCP Creative Improvement can recover them again. Moreover, trimmed patches are separated into two groups: (i) lighter and (ii) darker. The former is ordered with a third part of an F-Stop difference among them, while the latter are ordered the same, with an exception of the last patch due to it representing the blackest patch on the
color-checker passport card. The exposure difference between the darkest and the next darkest patch is out one-tenth part of a Stop, and the chart’s dynamic range is 32:1 (5 Stop).

![Color Checker Passport Charts](image)

Figure 7. Color Checker Passport Charts: (a) classic, (b) white balance and (c) creative improvement.

4. 3D22MX Image Database

Figure 8 shows the methodology to assess the quality of stereoscopic images, which is divided into three parts:

1. **Subjective Assessment.** Observers will give their opinion on the quality of the stereoscopic images (the input). This part is considered subjective since the same stereoscopic pair can have a different evaluation by two observers due to their visual perception;

2. **Objective Assessment of a 3D Coding process by a 3D Image Quality Assessment (3DIQA).** This part is considered objective since the same input stereoscopic pair will be given the same output evaluation; and

3. **Strength of the relationship** between the objective and subjective assessments to estimate the correlation between the observer’s opinion and a 3DIQA. The more these evaluations are correlated, the greater the relationship they will have with the average opinion of a human being.

Moreover, to highlight that to assess the quality of stereoscopic images, all three parts are essential, and the absence of any of them would cause the correlation between a digital quality metric and the average opinion of an observer to be incorrectly estimated.

That is why this section described first the three phases performed to develop our 3D22MX Image Database, i.e., a **Subjective Assessment**, while in Section 5 both **Objective Assessment** and **Strength of the Relationship** are discussed, as follows:

1. Capture,
2. Coding, and
3. Representation.

In **Capture Phase** are exposed the considerations taken into account to create a suitable enclosure for collecting the samples used for the Image Database generation, as well as the elements making up this enclosure and the camera and Space-Calibration techniques used. In addition, properties and considerations of outdoor shooting and the image formats used are described. In **Coding Phase** are explained images contained in this database, different noise types applied, disparity maps, and the software developed to add different noise types to each image. In **Representation Phase** are described the tools used for the visualization of the database.
4.1. Capture Phase

This phase breaks down the necessary procedure for capturing 3D-Images contained in 3D22\textsuperscript{MX} Image Database. On the one hand, capturing indoor images has motivated the design of a dark room with the following characteristics: (1) The dark room must be completely dark to avoid the entry light; (2) lighting applied to the stage must be controlled to capture source images. On the other hand, capturing outdoor images requires the use of sunlight to capture scenarios, since lighting cannot be controlled. For the case of capturing the images outdoors, since lighting cannot be controlled, a different procedure is followed where sunlight is used to be able to capture scenarios. It is important to highlight that both the interior and exterior images were taken at the Zacatenco Campus of the National Polytechnic Institute of Mexico (IPN). Figure 9 shows the dark-room designed to develop our 3D22\textsuperscript{MX} Image Database, divided into two-fundamental parts: Stage and Camera.

1. Stage: This part consists of a wooden table -whose dimensions are 56 × 33 cm- to put the objects to be captured, this table is placed behind the green muslin so that the captured objects are placed on top, Figure 9a. For lighting this stage, we use an arrangement of 75-Watt halogen lamps with a light dimmer, Figure 9b. This arrangement is placed above the stage to control the amount of Luminous Intensity or \( I_v \) in luxes (\( I_v = \frac{l_m}{m^2} \)), where \( l_m \) in lumens is the amount of light in the stage, and \( m \) in meters is the distance from this arrangement to the stage. Moreover, for covering the stage, we use a muslin, i.e., a piece of chromatically green, blue, or white fabric; the main objective of this background is to help researchers develop better segmentation algorithms.

Figure 9. The design of the dark room: (a) Scenario, (b) stage with background and (c) laser calibration of LOREO lens.
2. Camera: We used a Sony α230 camera with a LOREO lens, Figure 9c. When LOREO Lens is used, it is not necessary to calibrate the captured stereoscopic images because it is a device based on a three-dimensional single camera model, but the lens reduces the resolution by half since a single 3D-image has left and right stereoscopic pair in the same arrangement. Moreover, we use a tripod, placed at 1.30 m from the stage, to be the measure recommended by the LOREO’s manufacturer to achieve a 3D effect. In fact, a laser was calibrated in the three axes to have an accurate capture of the stage with the same conditions at all times.

Figure 10 shows the final conditioning of the stage. $I_0$ is measured with X-Rite Color Munki Display used as luxmeter, Figure 10a. The measurement must be equal to 0 luxes to guarantee that Dark-Room is conditioned to control the light conditions inside it. Figure 10b depicts how $I_0$ is read from the X-Rite Color Munki Display device. Figure 10c shows an example of 3D-Image captured in this proposal.

On the other hand, for capturing 3D-Image, Figure 11 shows how a characteristic curve of $I_0$ captured outside is made in this case instead of using the focus as is done in the shots of the scenarios inside, Figure 12a. Measurement is made for obtaining the same $I_0$ that were used in the captures of the scenarios indoors once this was done with the data obtained, it followed that the outdoor shots were taken at 6:50 AM because the $I_0$ obtained at that time on average were $I_0 = 715 \text{ lux}$, which would be half of $I_0$, and from 7:00 AM approximately $I_0 = 1240 \text{ lux}$ was obtained for full light captures, Figure 12b.

Source Outdoors Images were performed with the Sony α230 camera and LOREO Lens. Figure 11a shows a tripod that holds α230 camera, and it was placed at a distance for taking a 3D image of all objects in this scenario, according to implementation of indoor captures, described in Section 4.1. For calibration of outdoor 3D images to take pictures, first, we used a level calibrator in the center of the tripod to ensure the straightness of it and then placed a Color Checker Passport on it, Figure 11b. Then a 2D-Image of Color Checker Passport was taken. 3D-Images were made with the color checker passport at the same time, date, and place, Figure 11c. Finally, an Image with the Color Checker Passport superimposed on 3D-Image of the original stage, Figure 11d, the same day, time, and location with the program Adobe Photoshop, Figure 11e. The distance at which 3D-Images of each scenario varied from 6.5 m to 38.7 m.
Figure 11. Source Outdoors Images: (a) Sony a230 camera with LOREO Lens, (b) Color Checker Passport, (c) Capture of Color Checker Passport in a 2D-Image, (d) original scenario, and (e) scenario with Color Checker Passport superimposed.

Figure 12. Characteristic curve of $I_c$ captured outdoors: (a) Daylight-Hour Graph, sample every 30 min and (b) One-Hour Graph, sample taken every 10 min, during the sunrise.

Finally, all 3D-Images of 3D22$^{\text{MX}}$ database must have a color calibration. Hence, it is necessary to use the Color Checker Passport software. For achieving a color profile with the chromatic response of our camera, CCP is placed in all shots in this project, with a certain objective and specific light conditions, obtaining the maximum color accuracy, and taking advantage of the maximum reproduction range that our camera is capable of capturing.
For this, developer tools such as Adobe DNG Converter, Color Checker Passport Software, and Adobe Photoshop CS6 are employed. With Adobe DNG Converter, ARW image (extracted from our Sony α230 camera directly) is converted into a DNG format, since CCP software only detects this kind of image format, Figure 13a. Then, Color Checker Passport Software helps us generate a Sony α230 camera profile; when this program is opened, Figure 13b, 3D-Image is loaded mainly in a window. CCP Software then detects CCP in 3D-Image by placing the green boxes. Thus, when these green boxes are placed, CCO can be created a Color-Profile (CP), which is saved in a format .dcp.

![Adobe DNG Converter](image1.png) ![Color Checker Passport Software](image2.png) ![Adobe Photoshop CS6](image3.png)

(a) (b) (c)

Figure 13. Color calibration: (a) Adobe DNG Converter, (b) Color Checker Passport Software, and (c) Adobe Photoshop CS6.

Once CP is estimated, inside Adobe Photoshop CS6, there is a tool called Camera Calibration, when the profile of Sony α230 camera can be loaded, Figure 13c. This tool modifies the color of the 3D-Image. Figure 14a shows an Original Stereo pair, while Figure 14b shows a Recovered Stereo pair, i.e., re-painted, with the Sony α230 camera profile, specifically. Once the recovered stereo pair is re-painted, PhotoSteroMaker is used for spatial calibration by means of auto alignment tool. Finally, Recovered Stereo pairs are stored in BMP format, and all processes are repeated for each Original Stereo pair in the 3D22MX Image database.

![Adobe Photoshop CS6](image4.png)

(a) (b)

Figure 14. Adobe Photoshop CS6 Color calibration by means of Sony α230 camera profile: (a) Original Stereo pair and (b) Recovered Stereo pair.

4.2. Coding Phase

This section mainly covers three aspects:

1. Properties of images taken for the 3D22MX Image Database,
2. Amount of images that were taken for the full light intensity and average light intensity, and
3. Usage of Color Checker Passport (CCP) for color calibration in each element of $3D22^{\text{MX}}$ image database.

On the one hand, $3D22^{\text{MX}}$ image database contains 20 images whose dimensions are $1980 \times 1080$ with eight variants of the same scenario, which are:

1. Shot of the stage indoors and outdoors in full light.
2. Indoor and outdoor scene shot in full light with CCP Classic.
3. Indoor and outdoor scene shot in full light with CCP white balance.
4. Indoor and outdoor scene shot in full light with CCP creative improvement.
5. Shot of the scene indoors and outdoors in medium light.
6. Indoor and outdoor scene shot in medium light with CCP classic.
7. Indoor and outdoor scene shot in medium light with CCP white balance.
8. Indoor and outdoor scene shot in medium light with CCP creative improvement.

4.3. Representation Phase

This section mainly covers five aspects of the software tool developed in this work called $3D22^{\text{MX}}$ Toolbox, see Figure 15, for image degradation to manage:

1. Amount of noises,
2. Noise levels,
3. Degradation of a specific image,
4. Degradation an entire folder of images, and
5. Naming for saving images and along with the store’s format.

![Figure 15. $3D22^{\text{MX}}$ Toolbox main screen.](image)

Finally, the entire $3D22^{\text{MX}}$ Image Database, $3D22^{\text{MX}}$ Toolbox and obtained results can be downloaded at the following hyperlink: https://drive.google.com/file/d/14gX4IGlZuu9wInPigu3QZK5gfne_KS5/view?usp=sharing $3D22^{\text{MX}}$ (accessed on 1 November 2022). Moreover, this compendium of images and Matlab toolboxes can also be accessed through the QR code, shown in Figure 16.
5. Experimental Results

5.1. Experimental Methodology

In this section, the methodology used to carry out the psychophysical experiment is described. This methodology consists of three main features: (i) Visualization conditions used for the experiment; (ii) Properties of Stereoscopic Images from 3D22\textsuperscript{MX} Image Database that are projected to the viewer; and (iii) Visualization tests that are performed on the viewers to find out if they are candidates to evaluate stereoscopic images of database. From Figure 9, in a dark room, a SAMSUNG 3D UN40EH6030 television was placed with the following specifications: Screen resolution: 1920 × 1080, Screen size: 40 inches, Weight without stand: 8.9 kg, and Dimensions (W × H × D): 36.5 × 21.5 × 3.6 inches (or 927.5 × 548 × 93.1 mm). 3D TV contains glasses for visualizing the 3D effect; these glasses are the SSG-4100GB model to communicate with the Samsung 3D TV through the 2.4 GHz RF band. For communicating 3D-Glasses with the television, where necessary, perform four steps: (i) Glasses had to be brought closer to the TV, the maximum distance between them should be 50 cm; (ii) 3D TV must be in 3D mode; (iii) Glasses were turned on, a green led light up for 3 s; and (iv) A 3D Glasses connected to the TV message was displayed on the screen.

Moreover, the Optical Specifications of 3D-Glasses are: (i) shutters are made with Liquid Crystal, (ii) transmission of data has 37 ± 2% of Error, (iii) recommended viewing distance varies from 2 to 6 meters, and (iv) frequency field is 120 hertz at 7 s.

3D-Glasses are active, which means that they have an infrared sensor synchronizing images and alternating them on the screen, so that left and right eyes only see left and right perspectives, respectively. Namely, the infrared sensor simply synchronizes images that should be displayed for each eye. LCD crystals contain an active shutter that quickly alternates images on the screen. Blinking and changing the image from one eye to the other occur at such a speed that the brain does not manage to perceive any change and interprets it as a single three-dimensional scenario. A meter and a half of the television is positioned chair where the viewer takes a seat to evaluate 3D-Images when images are projected on 3D TV. The dark room has an illumination of approximately 0 lux, having as its only source of lighting the screen SAMSUNG 3D model UN40EH6030. Image projection is performed with the option of 3D-TV to show images sequentially, and the approximate time from one 3D-Stereo Pair to another is 15 s; in this time, viewers must evaluate 3D Image giving their Opinion Score (OS), and evaluation range used in this experiment is from 1 to 5, considering 1 as a very good experience and 5 as a very bad one.

3D22\textsuperscript{MX} has 20 scenarios with a J2K noise, but in the 5-levels of degradation, we get 100 images. These images were divided into two blocks, considering 3D-Stereo Pairs from
To order each Set of 3D-Stereo Pairs containing 10 subsets with 5 3D-Stereo Pair, in these subsets are placed different distorted 3D-scenarios with the following order: 1, 3, 5, 2, and finally 4 degradation levels. This order for the degradation level is used with the purpose that the viewer does not recognize any sequence or pattern of distortion applied and only evaluates each 3D-Stereo Pair projected on 3D-TV. Evidently, the first group of 30 viewers was projected in the first subset, and then the second group of 30 viewers was projected in the 3D-Stereo Pairs from 51 to 100, i.e., the second subset. The final subset of 60 viewers was obtained from a bigger set of 180 viewers, because 120 viewers were not surpassed Ishihara and Randot Stereo tests to determine, respectively, if they were color blind and determining their visual acuity. Hence, the first criterion to rule out observers as if they were colorblind, obviously, and after their visual acuity exceeded level 6, that is, having 40 s of arc or greater. If the viewer observes at least 40 s of arc, the following tests were performed: Measurement of interpupillary distance and dominant eye of the viewer.

5.2. Results and Analysis

To perform the experiments in this work, sixty people were chosen randomly. The main results can be summarized as follows:

- 19 people were women and the other 41 men;
- 38 people had 40 s of arc, 8 people with 30 s of arc, 5 people with 25 s of arc, and 9 people for 20 s of arc;
- 25 people used glasses while the other 35 did not use glasses;
- 33 people had a right dominant eye and the other 27 had a left dominant eye; and
- Only 10 out of the 60 people presented some discomfort, such as dizziness, nausea, or headache.

For the evaluation of Stereo Pairs, degradation ranged from 1 to 5 considering 1 very good and 5 very bad; the viewers can choose numbers they perceive within that range to give their evaluation. MOS of all Stereo Pairs in 3D22MX Image Database is shown in Table 1.

To analyze the behavior of each of the MOS obtained from viewers, a new order is performed where Degradation Levels (DL) were placed with each of Stereo Pairs, namely five subsets of Degradation Levels containing twenty 3D-Images. After analyzing this new order, it was found that observers felt uncomfortable when they viewed outdoor scenarios because these kinds of 3D images were displayed on full screen without causing a 3D effect, so discomfort is more important to increase the level of degradation. Therefore, small objects were evaluated with a lower level of degradation to yield the best quality. Thus, Stereo Pairs captured indoors were more pleasing to the eye of view when objects were not together; on the contrary, when objects were together, viewers evaluated that the Stereo Pairs had a greater degradation. In addition, it was found that small objects’ ratings were in accordance with the level of degradation of the pair; for example, if a small object had a Level 1 of degradation, the average evaluation was MOS = 1.2, but if this small object had a level 5 of degradation, the average evaluation was MOS = 4.4.

Figure 17 shows the scatter plot of the expected score versus MOS obtained from all Degradation Levels. When analyzing Stereo Pairs with degradation level 1, scores ranged from 1 to 2 with the exception of outdoor scenarios, i.e., Stereo Pairs 16 to 20. Results of Stereo Pairs with degradation level 2 show that scores ranged from 2 to 3. When analyzing the images of degradation level 3, scores ranged from 3 to 4. Finally, Stereo Pairs, with degradation levels 4 and 5, scores ranged from 4 to 5 or tended to 5, respectively. According to scores obtained for the indoor scenarios, it could be claimed that spectators are pleased to see small objects separated among them due to the 3D-effect differences. When small objects were very close together, it was very difficult for viewers to observe the 3D effect. In this figure dotted line indicates the average of the data, while the solid lines represent the samples that are in a 95% confidence interval.
Table 1. Most Opinion Score obtained for each Stereo Pair.

<table>
<thead>
<tr>
<th>Name</th>
<th>Degradation Level</th>
<th>MOS</th>
<th>Name</th>
<th>Degradation Level</th>
<th>MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>01_01</td>
<td>1</td>
<td>2.00</td>
<td>11_01</td>
<td>1</td>
<td>1.70</td>
</tr>
<tr>
<td>01_02</td>
<td>2</td>
<td>2.67</td>
<td>11_02</td>
<td>2</td>
<td>2.60</td>
</tr>
<tr>
<td>01_03</td>
<td>3</td>
<td>3.47</td>
<td>11_03</td>
<td>3</td>
<td>3.27</td>
</tr>
<tr>
<td>01_04</td>
<td>4</td>
<td>3.97</td>
<td>11_04</td>
<td>4</td>
<td>3.53</td>
</tr>
<tr>
<td>01_05</td>
<td>5</td>
<td>4.33</td>
<td>11_05</td>
<td>5</td>
<td>4.07</td>
</tr>
<tr>
<td>02_01</td>
<td>1</td>
<td>2.97</td>
<td>12_01</td>
<td>1</td>
<td>1.40</td>
</tr>
<tr>
<td>02_02</td>
<td>2</td>
<td>2.50</td>
<td>12_02</td>
<td>2</td>
<td>1.37</td>
</tr>
<tr>
<td>02_03</td>
<td>3</td>
<td>3.23</td>
<td>12_03</td>
<td>3</td>
<td>2.17</td>
</tr>
<tr>
<td>02_04</td>
<td>4</td>
<td>3.40</td>
<td>12_04</td>
<td>4</td>
<td>2.50</td>
</tr>
<tr>
<td>02_05</td>
<td>5</td>
<td>4.23</td>
<td>12_05</td>
<td>5</td>
<td>3.33</td>
</tr>
<tr>
<td>03_01</td>
<td>1</td>
<td>1.07</td>
<td>13_01</td>
<td>1</td>
<td>1.83</td>
</tr>
<tr>
<td>03_02</td>
<td>2</td>
<td>1.47</td>
<td>13_02</td>
<td>2</td>
<td>2.40</td>
</tr>
<tr>
<td>03_03</td>
<td>3</td>
<td>2.27</td>
<td>13_03</td>
<td>3</td>
<td>3.33</td>
</tr>
<tr>
<td>03_04</td>
<td>4</td>
<td>3.30</td>
<td>13_04</td>
<td>4</td>
<td>4.13</td>
</tr>
<tr>
<td>03_05</td>
<td>5</td>
<td>3.37</td>
<td>13_05</td>
<td>5</td>
<td>4.27</td>
</tr>
<tr>
<td>04_01</td>
<td>1</td>
<td>1.30</td>
<td>14_01</td>
<td>1</td>
<td>1.73</td>
</tr>
<tr>
<td>04_02</td>
<td>2</td>
<td>1.80</td>
<td>14_02</td>
<td>2</td>
<td>2.07</td>
</tr>
<tr>
<td>04_03</td>
<td>3</td>
<td>2.90</td>
<td>14_03</td>
<td>3</td>
<td>2.87</td>
</tr>
<tr>
<td>04_04</td>
<td>4</td>
<td>3.20</td>
<td>14_04</td>
<td>4</td>
<td>3.43</td>
</tr>
<tr>
<td>04_05</td>
<td>5</td>
<td>3.77</td>
<td>14_05</td>
<td>5</td>
<td>4.33</td>
</tr>
<tr>
<td>05_01</td>
<td>1</td>
<td>1.33</td>
<td>15_01</td>
<td>1</td>
<td>1.60</td>
</tr>
<tr>
<td>05_02</td>
<td>2</td>
<td>1.37</td>
<td>15_02</td>
<td>2</td>
<td>2.27</td>
</tr>
<tr>
<td>05_03</td>
<td>3</td>
<td>2.33</td>
<td>15_03</td>
<td>3</td>
<td>3.77</td>
</tr>
<tr>
<td>05_04</td>
<td>4</td>
<td>2.67</td>
<td>15_04</td>
<td>4</td>
<td>3.53</td>
</tr>
<tr>
<td>05_05</td>
<td>5</td>
<td>3.07</td>
<td>15_05</td>
<td>5</td>
<td>3.87</td>
</tr>
<tr>
<td>06_01</td>
<td>1</td>
<td>2.83</td>
<td>16_01</td>
<td>1</td>
<td>4.57</td>
</tr>
<tr>
<td>06_02</td>
<td>2</td>
<td>1.60</td>
<td>16_02</td>
<td>2</td>
<td>4.40</td>
</tr>
<tr>
<td>06_03</td>
<td>3</td>
<td>2.80</td>
<td>16_03</td>
<td>3</td>
<td>4.87</td>
</tr>
<tr>
<td>06_04</td>
<td>4</td>
<td>3.70</td>
<td>16_04</td>
<td>4</td>
<td>4.80</td>
</tr>
<tr>
<td>06_05</td>
<td>5</td>
<td>3.70</td>
<td>16_05</td>
<td>5</td>
<td>4.83</td>
</tr>
<tr>
<td>07_01</td>
<td>1</td>
<td>2.47</td>
<td>17_01</td>
<td>1</td>
<td>2.30</td>
</tr>
<tr>
<td>07_02</td>
<td>2</td>
<td>3.43</td>
<td>17_02</td>
<td>2</td>
<td>2.93</td>
</tr>
<tr>
<td>07_03</td>
<td>3</td>
<td>3.53</td>
<td>17_03</td>
<td>3</td>
<td>3.53</td>
</tr>
<tr>
<td>07_04</td>
<td>4</td>
<td>3.90</td>
<td>17_04</td>
<td>4</td>
<td>4.37</td>
</tr>
<tr>
<td>07_05</td>
<td>5</td>
<td>4.60</td>
<td>17_05</td>
<td>5</td>
<td>4.03</td>
</tr>
<tr>
<td>08_01</td>
<td>1</td>
<td>1.30</td>
<td>18_01</td>
<td>1</td>
<td>3.30</td>
</tr>
<tr>
<td>08_02</td>
<td>2</td>
<td>2.13</td>
<td>18_02</td>
<td>2</td>
<td>3.70</td>
</tr>
<tr>
<td>08_03</td>
<td>3</td>
<td>3.97</td>
<td>18_03</td>
<td>3</td>
<td>3.73</td>
</tr>
<tr>
<td>08_04</td>
<td>4</td>
<td>3.57</td>
<td>18_04</td>
<td>4</td>
<td>4.40</td>
</tr>
<tr>
<td>08_05</td>
<td>5</td>
<td>4.47</td>
<td>18_05</td>
<td>5</td>
<td>4.47</td>
</tr>
<tr>
<td>09_01</td>
<td>1</td>
<td>1.47</td>
<td>19_01</td>
<td>1</td>
<td>2.07</td>
</tr>
<tr>
<td>09_02</td>
<td>2</td>
<td>2.03</td>
<td>19_02</td>
<td>2</td>
<td>3.40</td>
</tr>
<tr>
<td>09_03</td>
<td>3</td>
<td>3.13</td>
<td>19_03</td>
<td>3</td>
<td>4.27</td>
</tr>
<tr>
<td>09_04</td>
<td>4</td>
<td>3.77</td>
<td>19_04</td>
<td>4</td>
<td>4.50</td>
</tr>
<tr>
<td>09_05</td>
<td>5</td>
<td>4.33</td>
<td>19_05</td>
<td>5</td>
<td>4.70</td>
</tr>
<tr>
<td>10_01</td>
<td>1</td>
<td>1.70</td>
<td>20_01</td>
<td>1</td>
<td>3.47</td>
</tr>
<tr>
<td>10_02</td>
<td>2</td>
<td>2.83</td>
<td>20_02</td>
<td>2</td>
<td>4.00</td>
</tr>
<tr>
<td>10_03</td>
<td>3</td>
<td>3.63</td>
<td>20_03</td>
<td>3</td>
<td>4.43</td>
</tr>
<tr>
<td>10_04</td>
<td>4</td>
<td>4.47</td>
<td>20_04</td>
<td>4</td>
<td>4.57</td>
</tr>
<tr>
<td>10_05</td>
<td>5</td>
<td>4.73</td>
<td>20_05</td>
<td>5</td>
<td>4.83</td>
</tr>
</tbody>
</table>
5.3. Comparison of 3D-Image Quality Assessments

In this work, 3D22MX database has been tested with 23 Image Quality Assessments (IQA). These IQAs have been designed for image quality models of RGB two-dimensional images or 2DIQA. Equation (7) shows the procedure used, i.e., how the individual qualities of the left (2DIQAₗ) and right (2DIQAᵣ) image of a 3D-stereoscopic pair were averaged to obtain a three-dimensional image quality value, 3DIQA. The set of 2DIQA comprised the following 23 metrics: Mean-Squared Error (MSE) [27], Peak Signal-to-Noise Ratio (PSNR) [28], Structural Similarity Index (SSIM) [29,30], Multiscale Structural Similarity Index (MSSIM) [30,31], Visual Signal-to-Noise Ratio (VSNR) [32], Pixel-Based Visual Information Fidelity (VIF) [11,33], Visual Information Fidelity (VIFP) [34,35], Universal Quality Index (UQI) [36], Image Fidelity Criterion (IFC) [37,38], Noise Quality Measure (NQM) [39], Weighted Signal-to-Noise Ratio (WSNR) [39,40], Signal-to-Noise Ratio (SNR), Average Difference (AD), Maximum Difference (MD), Normalized Absolute Error (NAE), Normalized Cross Correlation (NCC), Structural Content (SC), Blind Image Quality Index (BIQI) [41], Blind/Referenceless Image Spatial Quality Evaluator Index (BRISQUE) [42], Naturalness Image Quality Evaluator (NIQE) [43], Feature-Similarity (FSIM) [44], Riesz-Transform Feature-Similarity (RFSIM) [45], and Peak Signal-to-Noise Ratio with Contrast Sensitivity Function (PSNRHVSM) [46].

$$3DIQA = \frac{2DIQAₗ + 2DIQAᵣ}{2}$$ (7)

Then, each 3DIQA was evaluated by predicting each 3D22MX Image or Predicted DMOS, and the correlations of these results with those evaluations given by human observers were calculated, i.e., DMOS vs. Predicted DMOS. It is important to say that Pearson’s Linear Correlation Coefficient (LCC) was used to measure the linear behavior of the observers’ response, while Rank Ordered Correlation Coefficients, especially Kendall’s (KROCC) and Spearman’s (SROCC), were used to measure their non-parametric response since not all 3DIQA’s predict with evaluations as a human being does. Moreover, the Root-Mean-Squared Error (RMSE) was used to measure the error between evaluations. Thus, Table 2 shows the results in terms of performance that were obtained across 3D22MX, NQM, and RFSIM, to predict the best results in both linear correlation and non-linear correlation coefficients. The results were highlighted in bold and italics, respectively. In the case of the RMSE, MD, and NIQE metrics, results were yielded from the least differences with respect to evaluations of DMOS made by a human being.
### Table 2. Performance across 3D-IQA Set in predicting perceived stereoscopic image quality: Linear-Linear Correlation Coefficient (LCC), Spearman’s Rank Ordered Correlation Coefficient (SROCC), Kendall’s Rank Ordered Correlation Coefficient (KROCC), and Root Mean Squared Error (RMSE). **Bold** indicates the best metric, while *italics* indicate the second best.

<table>
<thead>
<tr>
<th>3D-IQA</th>
<th>LCC</th>
<th>SROCC</th>
<th>KROCC</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.5664</td>
<td>0.7697</td>
<td>0.5792</td>
<td>0.4235</td>
</tr>
<tr>
<td>PSNR</td>
<td>−0.7584</td>
<td>−0.7700</td>
<td>−0.5796</td>
<td>0.3886</td>
</tr>
<tr>
<td>SSIM</td>
<td>−0.4961</td>
<td>−0.6080</td>
<td>−0.4409</td>
<td>0.3722</td>
</tr>
<tr>
<td>MSSIM</td>
<td>−0.5864</td>
<td>−0.7657</td>
<td>−0.5772</td>
<td>0.4075</td>
</tr>
<tr>
<td>VSNR</td>
<td>−0.7549</td>
<td>−0.7876</td>
<td>−0.5983</td>
<td>0.3851</td>
</tr>
<tr>
<td>VIF</td>
<td>−0.5809</td>
<td>−0.6536</td>
<td>−0.4681</td>
<td>0.4476</td>
</tr>
<tr>
<td>VIFP</td>
<td>−0.5501</td>
<td>−0.6041</td>
<td>−0.4360</td>
<td>0.4200</td>
</tr>
<tr>
<td>UQI</td>
<td>−0.1721</td>
<td>−0.2778</td>
<td>−0.1967</td>
<td>0.4161</td>
</tr>
<tr>
<td>IFC</td>
<td>−0.4443</td>
<td>−0.5223</td>
<td>−0.3541</td>
<td>0.4349</td>
</tr>
<tr>
<td>NQM</td>
<td><strong>0.8710</strong></td>
<td><strong>−0.8773</strong></td>
<td><strong>−0.6863</strong></td>
<td><strong>0.4129</strong></td>
</tr>
<tr>
<td>WSNR</td>
<td>−0.7575</td>
<td>−0.7677</td>
<td>−0.5824</td>
<td>0.4169</td>
</tr>
<tr>
<td>SNR</td>
<td>−0.6116</td>
<td>−0.6648</td>
<td>−0.4843</td>
<td>0.4043</td>
</tr>
<tr>
<td>AD</td>
<td>−0.0641</td>
<td>−0.1230</td>
<td>−0.0973</td>
<td>0.2599</td>
</tr>
<tr>
<td>MD</td>
<td>0.5951</td>
<td>0.5931</td>
<td>0.4198</td>
<td><strong>0.1811</strong></td>
</tr>
<tr>
<td>NAE</td>
<td>0.5245</td>
<td>0.5725</td>
<td>0.4093</td>
<td>0.2210</td>
</tr>
<tr>
<td>NCC</td>
<td>−0.6187</td>
<td>−0.6217</td>
<td>−0.4518</td>
<td>0.3929</td>
</tr>
<tr>
<td>SC</td>
<td>0.5960</td>
<td>0.5936</td>
<td>0.4303</td>
<td>0.3159</td>
</tr>
<tr>
<td>BIQI</td>
<td>−0.3212</td>
<td>−0.2415</td>
<td>−0.1371</td>
<td>0.3616</td>
</tr>
<tr>
<td>BRISQUE</td>
<td>−0.1516</td>
<td>−0.1115</td>
<td>−0.0556</td>
<td>0.3610</td>
</tr>
<tr>
<td>NIQE</td>
<td>0.5603</td>
<td>0.5726</td>
<td>0.3991</td>
<td>0.1968</td>
</tr>
<tr>
<td>FSIMC</td>
<td>−0.6652</td>
<td>−0.7458</td>
<td>−0.5512</td>
<td>0.3524</td>
</tr>
<tr>
<td>RFSIM</td>
<td>−0.8607</td>
<td>−0.8659</td>
<td>−0.6814</td>
<td>0.4262</td>
</tr>
<tr>
<td>PSNRHVS</td>
<td>−0.7885</td>
<td>−0.7950</td>
<td>−0.6125</td>
<td>0.4026</td>
</tr>
</tbody>
</table>

Finally, Figure 18 shows the scatter plots of the performance of the best NQM (Figure 18a) and RFSIM (Figure 18b) metrics, where a tendency to the ideal result is observed, i.e., the scattered point cloud is concentrated in a straight line. While Figure 18c,d show, respectively, the performance of UQI and BRISQUE, which were the worst metrics to predict human evaluation within 3D22^{MX}, since they show an almost random performance as a whole.

![Figure 18](image_url)
Figure 18. Scatter plots of linear performance of the best two metrics (a) NQM, and (b) RFSIM and the two worst assessments (c) UQI, and (d) BRISQUE.

6. Conclusions

The environment of the 3D-image databases was compared in this proposal from a standardized database that contained more images and noises, unlike its predecessors, such as NINGBO, IRCCyN/IVC, and LIVE 3D IQA. NINGBO Image Database contains ten Original Stereo Pairs, including 400 Distorted Stereo Pairs, where authors performed psychophysical experiments with 20 viewers. IRCCyN/IVC Image Database contains ten Original Stereo Pairs, including 90 Distorted Stereo Pairs, where authors performed psychophysical experiments with 17 viewers. LIVE 3D IQA Image Database has ten Original Stereo Pairs, including 365 Distorted Stereo Pairs, where authors performed psychophysical experiments with 30 viewers. This project, 3D22\textsuperscript{MX} Image Database, contains more images than NINGBO, IRCCyN/IVC, and LIVE 3D IQA Image Database, and our psychophysical experiment was more exhaustive than in the previous ones. To accomplish the requirements for the realization of the psychophysical experiment, such as seconds of arc, the number of spectators were analyzed, taking into account the development and application of the tools for image projection and projection time.

Therefore, Human Stereoscopic Vision features, Noise for Image Distortion, and Scenario for 3D-Image Capture were defined for describing the noise types that cause discomfort to the human eye when images are perceived along with the scenario for 3D-Image captured with color checker passport tool. In addition, 3D22\textsuperscript{MX} Image Database has been described as the three-phases algorithm, namely Capture, Coding, and Representation Phases. In Capture Phase it was created a suitable enclosure for collecting samples used for the Image Database generation, as well as elements making up this enclosure and the camera and Space-Calibration techniques used. Moreover, properties and considerations of outdoor shooting and image formats used were described and analyzed. In Coding Phase were described the images contained in this database, different noise types applied, disparity maps, and the software developed to add different noise types to each image. In Representation Phase were described the tools used for visualization of the database.

Furthermore, a psychophysical experiment was prepared for recording and evaluating the experimented quality by 180 viewers with respect to the Distorted Stereo Pairs. However, 120 viewers were discarded due to the visual acuity of the 3D-effect was not enough according to the test Randot Stereotest. Results yielded from the psychophysical experiment carried out were the support to validate both the 3D22\textsuperscript{MX} Image Database and proposed software tool. Our results were the following: (i) 19 out of 60 viewers were women, and the other 41 viewers were men, (ii) average arc seconds was 40, (iii) counting 38 viewers with 40 sec.arc., (iv) 8 viewers with 30 sec.arc., (v) 5 viewers with 25 sec.arc., (vi) 9 viewers with 20 sec.arc, and (vii) an amount of 25 viewers used glasses while 35 did not use them. We found and defined Dominant Eye (DE) of any viewer of the sample, determining that 33 viewers had right DE and 27 viewers had left DE. Finally, at the end of the experiment,
we asked each one of the participants if they presented any discomfort such as dizziness, nausea, or headaches; 50 viewers answered negatively and the remaining 10 viewers positively. Thus, according to the central limit theorem, by performing these random repetitions, it is possible to normalize the data from a larger population, which makes it possible to calculate reasonably well the probabilities of random variables associated with this experiment and even if the sample size is large enough, the distribution of the sample means will approximately follow a normal distribution.

Based on our findings, we conclude that our database of images 3D/stereoscopic high-definition or 3D22\textsuperscript{MX} enables to perform psychophysical experiments to evaluate the quality of the user experience, and 3D22\textsuperscript{MX} is a useful tool in the research field of 3D/stereoscopic color naming algorithm, color constancy algorithm, high dynamic range algorithm, the capture of macroscopic images in 3D/stereoscopic and realization of other types of psychophysical experiments.

Contributions of this work induce us that as a future work, it will be possible to verify the operation of different algorithms, calibrate color images, calibrate contrast images, calibrate images in lighting, perform other types of psychophysical experiments, degrade images with more noises than this proposal, validate noise levels, recalculate images among others. Therefore, it is expected that 3D22\textsuperscript{MX} Image Database will be used by other researchers of stereoscopic images and their branches, such as High Dynamic Range Algorithms or in the Capture of macroscopic images in 3D/stereoscopic projected, for instance.

Finally, these results show that obtained performance across 3D22\textsuperscript{MX}, points out that NQM and RFSIM obtain the best results in terms of both linear correlation and non-linear correlation coefficients. In the case of the RMSE, MD, and NIQE metrics, results were yielded from the least differences with respect to evaluations of DMOS made by a human being.

Author Contributions: Conceptualization, J.J.M.E.; formal analysis, O.M.M. and L.C.H.; investigation and resources, O.M.M. and L.C.H.; data acquisition, J.J.M.E. and O.M.M.; writing—original draft preparation, E.Y.A.d.V., J.J.M.E. and O.M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This article is supported by the National Polytechnic Institute (Instituto Politécnico Nacional) of Mexico by means of projects No. 20220385, and 20220962 granted by the Secretariat of Research and Postgraduate (Secretaría de Investigación y Posgrado), National Council of Science and Technology of Mexico (CONACyT).

Data Availability Statement: Not applicable.

Acknowledgments: The research described in this work was carried out at the Superior School of Mechanical and Electrical Engineering (Escuela Superior de Ingeniería Mecánica y Eléctrica) of the Instituto Politécnico Nacional, Campus Zacatenco. It should be noted that this research is part of a Bachelor’s Degree thesis entitled Prototipo para la evaluación de Imágenes Tridimensionales en Alta Definición supported by Jessica Andrea González Sánchez, and Denisse Ariane García Lara, work directed by Jesús Jaime Moreno Escobar.

Conflicts of Interest: The authors declare no conflict of interest.

References


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.