VOL. 13, NO. 20, OCTOBER 2018 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

DESIGN OF AN OBJECTIVE MODEL TO EVALUATE THE 3D ANAGLYPH VIDEO QUALITY

Jesús D. Quintero Polanco, Martin D. Bravo Obando and José de Jesús Salgado Patrón
Department of Electronic Engineering, Surcolombiana University, Neiva, Colombia
E-Mail: jdavid@usco.edu.co

ABSTRACT

This paper presents the design and implementation of an objective evaluation model for the quality of anaglyph 3D video based on standardized objective evaluation models for 2D video; each metric that composes the objective model is adapted so that it can identify the characteristic components of anaglyph 3D video. In order to verify the proposed model, a subjective evaluation is carried out, in which the evaluation of the viewers about the perception of video quality is carried out through a mobile device, for this purpose a web application is developed in charge of acquiring and store all the information in a database; the process is carried out following international recommendations. The results obtained from the objective model show that a good approximation is achieved compared to the results of the subjective evaluation.

Keywords: subjective evaluation, objective model, anaglyph, perception, quality, video.

INTRODUCTION

In recent years the consumption of multimedia services has increased massively, and every day new standards emerge for different services provided by cable television distributors, internet providers and telephone operators. For each of these scenarios it is essential to guarantee the quality of service, where the user is in charge of determining it based on his own criteria. Video quality is one of the important aspects that every provider must consider when providing multimedia services.

The processes of acquisition, compression, processing, storage, transmission and reproduction of information generate several kinds of distortion in the final anaglyph 3D video signal. Video compression techniques with losses reduce the needed capacity in the storage or transmission of information, in return they generate a degradation of the video quality that must be detected, evaluated and quantified in order to maintain it, control it, or possible to improve it (Sikora, 2005).

For a long time, the safest way to determine the quality of a video has been through the use of a subjective evaluation, where the end user is the one who determines how good the video signal is, the MOS (Mean Opinion Score) has been recognized as the most reliable method to assess quality (Joskowicz, *et al.*, 2012). However, to determine the MOS measure it is necessary to perform complex tests, where the disadvantage is the time and resources necessary to develop them (Huynh Thu, 2011).

For these reasons, the main objective is to design, implement and evaluate an objective evaluation model with a high degree of correlation with subjective evaluation, which can be used to establish the quality of anaglyph 3D video.

Some measures of distortion (mean square error, standard deviation or signal-to-noise ratio) formulated some time ago allow to quantify the error between the distorted signal and the original signal to assume a quality factor (Wang, *et al*., 2003). A great effort has been made to develop objective methods that integrate perceptual quality measures taking into account all the characteristics of the HVS (Human Visual System). In 1982, the first quantitative quality measure for video was proposed (Lukas and Budrikis, 1982). Subsequently, in October 1997, the VQEG group (Video Quality Experts Group) was created with the main objective of collecting reliable subjective ratings for a well-defined set of test sequences and standardizing the new objective evaluation methods in video quality.

The content of this paper is organized as follows. Section 2 describes the type of information collected, the processing of anaglyph 3D video sequences through the use of a free software tool, the correct implementation and implementation of a subjective evaluation, and finally the design and implementation is presented of the objective evaluation model (Seshadrinathan, 2010). In section 3 the results obtained are shown and an analysis is carried out to compare the evaluation models. The conclusions are presented in section 4.

METHODOLOGY

To make the objective model it was necessary to divide the project into stages with constant feedback between them. The stages that were developed were:

Collection of information

This stage consisted of consulting and investigating important aspects for the development of the project as sources of 3D video anaglyphs of reliable origin, free software that allowed to perform coding in H.265/HEVC (High Efficiency Video Coding), the different objective metrics for 2D video existing and the correct performance of a subjective evaluation (ITU-R BT 500-11, 2002).

Processing of video sequences

In this stage, the free software FFmpeg (Fast Forward Moving Pictures Expert Group) was used to perform the compression and decompression in H.265 of the three selected anaglyph 3D video sequences (Tomar, 2006). The purpose of the H.265/HEVC codec is to offer twice the compression efficiency and to deliver a significant improvement in visual quality compared to its

ARPN Journal of Engineering and Applied Sciences

predecessor, the H.264/AVC (Advanced Video Coding) standard; even so, the efficiency in the compression of the H.265 codec varies depending on the configuration of the codec and the content of the video.

The compressor was configured taking into account parameters such as compression speed, fps (Frames per second) and CRF (Costant Rate Factor). The latter is of great importance because it allows the encoder to achieve a certain output quality for the whole file when the video size is of less relevance, making each frame obtain the necessary bit rate to maintain the appropriate quality level.

The CRF can be set between 0 and 51, where lower values will result in better quality (at the expense of larger file sizes). The recommended value for a good size-quality ratio is 28, which was used as a reference point. The values chosen from the CRFs to degrade the sequences were 14, 28, 42 and 51; therefore, we worked with three original videos and four degradations for each one.

Implementation of the subjective evaluation

The most reliable way to measure the quality of a video is through the subjective evaluation carried out by a wide group of people who think about their perception. With these opinions you can calculate the MOS metric which is accepted as a measure of quality.

Before performing the subjective test, the spectators were asked to undergo a visual acuity test to determine if they had any kind of problem in their vision that would prevent the evaluation (Dahl, 2008). Afterwards, it was invited to enter the room where the visualization of the videos would take place. This room was designed in such a way that its interior luminance was less than 20 lux following the norms established in international recommendations (ITU-T P.910, 2008).

The DSIS (Double Stimulus Impairment Scale) method was chosen for the execution of this evaluation, because it is the only one that allows the degradation level between the original video and the gradient to be qualified. At the beginning of each session, the 21 observers were explained about the type of evaluation implemented, the scale of appreciation, the sequences and timing.

It was made clear that the worst quality observed does not necessarily correspond to the lowest subjective score and that their appreciation was based on the overall impression they perceive of the videos. Viewers were asked to watch the videos during periods T1 and T3, in period T2 it was established to show a gray image of short duration, finally the voting must be done only during T4, as shown in Figure-1.

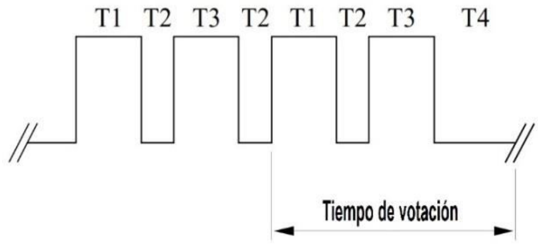

Figure-1. Presentation structure of the test material.

The DSIS method has a default category of values for MOS interpretation where viewers choose one of five options, as shown in Table-1.

Table-1. Grade scale of MOS/.

Degradation	Score
Imperceptible	5
Perceptible, but not annoying	4
Slightly annoying	3
Annoying	2
Very Annoying	1

A web application was designed to present the value categories of the DSIS method; the data taken by the application are stored in a database. In total, three videos were used, some with high mobility and others with low mobility, with the aim of creating a metric that works with different types of movement.

Four degradations were applied to each video, there were a total of three sessions and for each of them the viewers rated four MOS measures.

Design and implementation of the objective model

For the design of the objective model we chose the combination of measures based on error detection and techniques based on structural distortion, because both contribute to the perception of objective quality from different perspectives. The human visual system is adapted to extract structural information and detect changes in signals, which is why it is necessary to present an objective measure of structural information loss. On the other hand, an approach based on error detection is also important, since it estimates the perceived errors that the degraded signal has compared with the original signal.

Additionally, the red, green and blue colors characteristic of the anaglyph video was taken into account. For this, the RGB standard was used in order to analyze each plane of the test sequences.

Structure of the objective model

Four phases are established to determine the objective quality model in anaglyph 3D videos.

Phase I

The value of the peak signal to noise ratio is determined for the set of frames that constitute the original and degraded video in order to establish a quality value.

Since we already have the measure of the mean square error for a couple of frames (original-degraded) in a plane (R or G or B), it is only necessary to perform the same calculation for the other pairs of frames and then averaging them. Later we do the same for the other two planes.

$$\mathrm{MSE_v} = \frac{1}{3\mathrm{F}}\left(\sum_{\mathrm{j=1}}^{\mathrm{F}} \mathrm{MSE_R(j)} + \sum_{\mathrm{j=1}}^{\mathrm{F}} \mathrm{MSE_G(j)} + \sum_{\mathrm{j=1}}^{\mathrm{F}} \mathrm{MSE_B(j)}\right) \quad (1)$$

$$PSNR = 10log_{10}\left(\frac{L^2}{MSE_v}\right) \quad (2)$$

Where F is the number of frames and $\mathrm{MSE_R(j)}$, $\mathrm{MSE_G(j)}$, $\mathrm{MSE_B(j)}$ are the value of the squared error calculated for the j-th frame in the R, G, B plane. Table-2 relates the PSNR with the MOS.

Table-2. Relationship between the values of PSNR and MOS.

PSNR	MOS
> 37	5
31 - 37	4
25 - 31	3
20 - 25	2
< 20	1

Phase II

The universal image quality index was calculated for the set of frames that make up the original and degraded video.

First, the size of the sliding window is chosen as 8x8 pixels, then the number of sampling windows (T) per frame is determined. Therefore, the universal quality index is calculated for each of the blocks.

$$UIQI = \frac{4\mu_x\mu_y\sigma_{xy}}{\left(\mu_x{}^2 + \mu_y{}^2\right)\left(\sigma_x{}^2 + \sigma_y{}^2\right)} \quad (3)$$

Where μ_x, σ_xcorrespond to the mean and variance of the original frame; μ_y, σ_y correspond to the mean and variance of the degraded frame and σ_{xy} is the covariance between the two frames.

The $UIQI$ index is applied to the components R, G and B independently, then they are combined in a local measure.

$$UIQI(i,j) = V_R \cdot Q_R(i,j) + V_G \cdot Q_G(i,j) + V_B \cdot Q_B(i,j) \quad (4)$$

Where $V_R = V_G = V_B = 1/3$are the weighting factors. Next all the indexes of the blocks of a frame are combined.

$$Q_i = \frac{1}{T}\sum_{i=1}^{T} UIQI(i,j) \quad (5)$$

Where $Q_i$equals the measurement of the quality index of the i-th frame in the set of frames representing the video sequence, and T represents the weight value given to the j-th sliding window in the i-th frame. Finally, the final universal quality index is given by equation 6, where F is the number of frames.

$$Q_{UIQI} = \frac{1}{F}\sum_{i=1}^{F} Q_i \quad (6)$$

Phase III

The index of structural similarity for the set of frames that constitute the original and degraded video is also found.

The size of the sliding window is chosen as 8x8 pixels, then the number of sampling windows T is determined per frame. Therefore, the structural similarity index is calculated for each of the blocks.

$$SSIM = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x{}^2 + \mu_y{}^2 + C_1\right)\left(\sigma_x{}^2 + \sigma_y{}^2 + C_2\right)} \quad (7)$$

Being $C_1 = (K_1L)2$ and $C_2 = (K_2L)2)$, where $L = 255$ and$K_1 = K_2 = 0.001$. Also,μ_x, σ_x correspond to the mean and variance of the original frame;μ_y, σ_y correspond to the mean and variance of the degraded frame and σ_{xy} is the covariance between the two frames.

The SSIM index is applied to the components R, G and B independently, then they are combined in a local measure.

$$SSIM(i,j) = W_R \cdot SSIM_R(i,j) + W_G \cdot SSIM_G(i,j) + W_B \cdot SSIM_B(i,j) \quad (8)$$

Where $W_R = W_G = W_B = 1/3$are the weighting factors.

Next, all the indexes of the blocks of a frame are combined:

$$Q_i = \frac{1}{T}\sum_{i=1}^{T} SSIM(i,j) \quad (9)$$

Finally, Table-3 is proposed that relates QUIQI and QSSIM with the MOS.

VOL. 13, NO. 20, OCTOBER 2018 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Table-3. Relationship between SSIM/UIQI and MOS values.

SSIM/UIQI	MOS
> 0.970	5
0.920 – 0.969	4
0.850 – 0.919	3
0.700 – 0.849	2
< 0.699	1

Phase IV

After obtaining the quality values for the different proposed techniques, these are converted to a MOS scale using Table-2 and Table-3. In this phase the MOS values equivalent to each one of the metrics are averaged to give a final result of objective quality, which will be used to compare the results of the subjective test. This final anaglyph 3D video quality target value is indicated in equation 11.

$$MOS_{goal} = \frac{MOS_{PSNR} + MOS_{QUIQI} + MOS_{QSSIM}}{3} \qquad (10)$$

RESULTS AND DISCUSSIONS

Subjective evaluation

After having an average MOS measurement for each type of degradation and for each sequence, we proceeded to graph the results as shown in Figure-2.

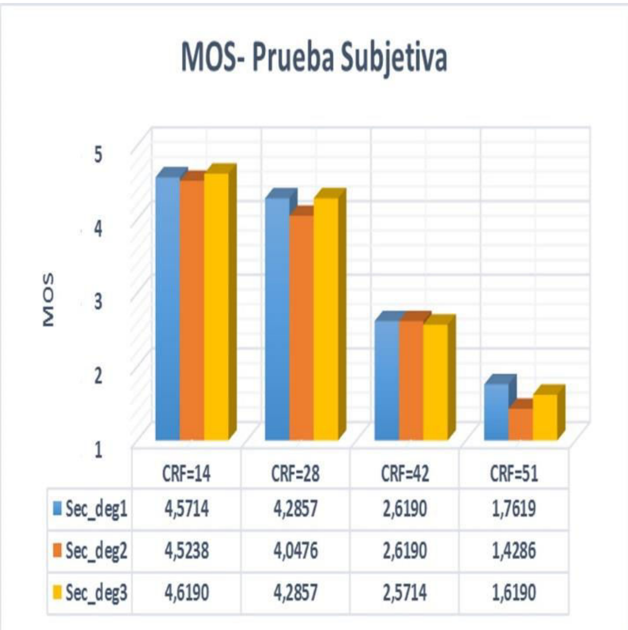

Figure-2. MOS rating of the subjective test.

From the results obtained in Figure-2 we can determine that:

- The ratings that had the highest score of the viewers were for the CRF=14 of each of the videos since they correspond to the parameter that allows having the highest possible quality at the time of coding. In the same way, it can be seen that the worst ratings were for the CRF=51 because this parameter degrades the video to such an extent that it is almost imperceptible.
- Observers rated similarly for each type of degradation regardless of the playback characteristics of the video.
- The recommendation of the ITU-T P.800 proposes to encode in H.265 with a constant rate factor of 28 since it maintains the quality of the original video and decreases the weight of the file maintaining the quality of the original video. We can verify this by looking at Figure-2, where for a CRF=28 the ratings were quite similar to those of CRF 14 which improves the quality much more, but increases the size of the video.
- In general, it can be seen that as the CRF degradation parameter increases, the perceived quality of the viewer decreases.

Objective model

In this section we analyze the results obtained from the objective model designed, which is constituted by the PSNR, SSIM and UIQI metrics adapted to evaluate the quality of anaglyph 3D videos.

The PSNR metric takes into account the importance of the colors (red and cyan) that stand out in this type of videos. The final value thrown by the metric is converted to the MOS scale to analyze and check the results using Table-2.

The results are detailed in Figure-3.

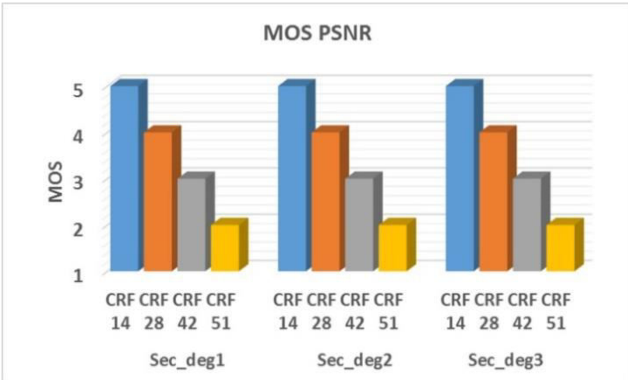

Figure-3. MOS measurement equivalent to the PSNR metric.

The MOS rating corresponding to the PSNR is high when a small CRF is used in the coding; to the extent that the latter parameter increases, the quality of the video will decrease, and the first metric will make it notice through its MOS rating.

For this case, the video encoded with CRF14, the metric qualifies it with five, that is, the video is of excellent quality. However, for a CRF 51 the video will have a rating of two that represents poor quality.

As it can be seen, the MOS rating of the PSNR for the different videos is the same for each type of

VOL. 13, NO. 20, OCTOBER 2018 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

distortion, providing results very similar to those included in the subjective test.

The SSIM metric defines luminance, contrast and structural comparison measures, and like the first, the importance of colors is maintained.

The measurement of this metric is converted to the MOS scale using Table-3 to compare it with what viewers think. The results are illustrated in Figure-4.

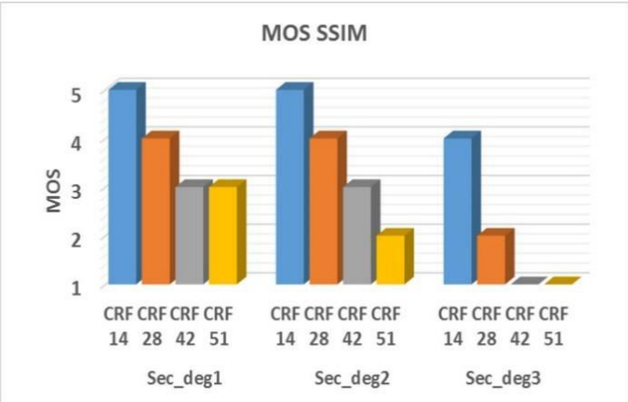

Figure-4. MOS measurement equivalent to the SSIM metric.

It can be inferred that unlike the PSNR, this metric presents variation in terms of the MOS ratings of the degraded sequence Sec_deg3, which has a high mobility compared to the other videos.

The designed metric makes that when correlating the original video sequence and the degraded sequence based on the luminosity, the contrast and the structural information, distortions can be seen that the viewer does not perceive, giving a much more objective qualification of the video.

The UIQI metric also defines the luminance and contrast measurements focused on measuring mainly the loss of correlation instead of the structural similarities between the original and degraded video sequences. Table-3 relates the values of this metric to the MOS. The results are presented in Figure-5.

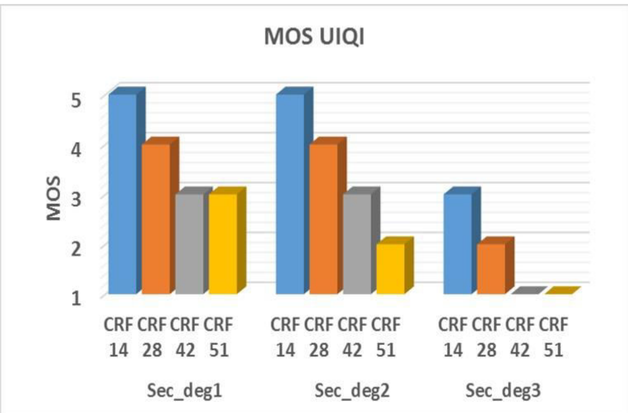

Figure-5. MOS measurement equivalent to the UIQI metric.

Like the SSIM, the objective metric, UIQI presents variations in the degraded sequence Sec_deg3 due to the high mobility it presents; it is also due to the correlation that the metric makes to detect changes and loss of information.

After obtaining all the MOS measurements obtained in each designed objective metric, a single measurement is calculated by finding the arithmetic mean of the three, which is called the objective MOS.

Subsequently, the results of the relationship between the objective MOS designed and the subjective MOS performed are presented in Figure-6.

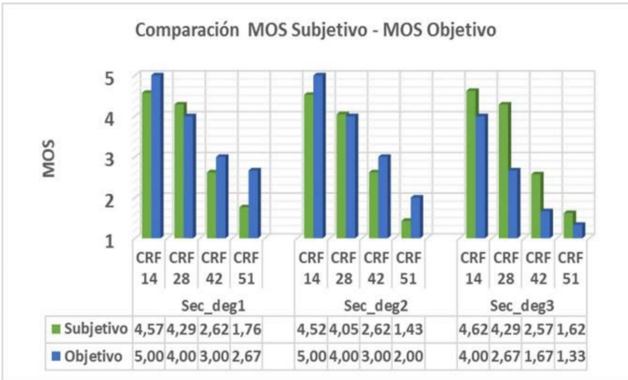

Figure-6. Objective MOS and subjective MOS comparison.

From the Figure it can be concluded that:

- The MOS measure that represents the objective model designed has a high degree of relationship with respect to the MOS rating representative of what the viewers think. The graph shows that the trends are similar for each type of distortion depending on the analyzed sequence.
- The objective model presents a strict analysis in the video sequence Sec_deg3 because its mobility content is high, and in comparison with the evaluation of the viewer, this model does perceive the structural differences overlooked.
- The best grades were for the CRF 14 in all the videos for both the subjective test and the objective model. Likewise, the worst rating for all the videos was obtained by the CRF 51 degradation parameter.
- The recommended degradation parameter for the coding in HEVC/H265 is CRF 28, in the graph it is not the best, but it is of good and constant quality, apart from that it compresses very well the size of the file.

Confidence interval

Finally, calculating the confidence interval in order to find the probability of success that the objective model has, at the moment of qualifying any type of distortion present, is fundamental and its results are observed in Table-4.

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Table-4. Confidence intervals (95%).

CRF	Mean	Confidence interval
14	4.65	0.65
28	3.65	0.65
42	2.65	0.65
51	2.00	1.13

In the case of CRF 14, the arithmetic mean of the measurements that the objective model yields, together with the confidence interval (4.65 ± 0.65) indicates that the metric presents a MOS measurement between 4 and 5 for this distortion. With regard to CRF 51, the confidence interval is very large because the characteristics of the degraded videos show great differences.

Verification test

A verification of the objective metric designed to verify the efficiency of the same was made. Two new anaglyph 3D videos were used, and thirteen people were summoned to carry out the subjective test.

Once this test was completed, the MOS measure and its respective confidence interval were calculated. The results are shown in Figure-7.

Figure-7. MOS verification - subjective test.

It is observed that for the two sequences the viewers rated very similar, although the videos were of different speed of reproduction; as evidenced in the first three sequences used.

Subsequently, the results obtained from the objective model applied to the videos used in the verification were analyzed. These data are shown in Figure-8.

Figure-8. MOS verification - Objective test.

It can be seen that the verification sequence Sec_ver1, which has a slow mobility, had a much higher rating than the verification sequence Sec_ver2 with fast mobility; this trend was explained in previous sections.

The results of the designed metric and the subjective test that are shown in Figure-9.

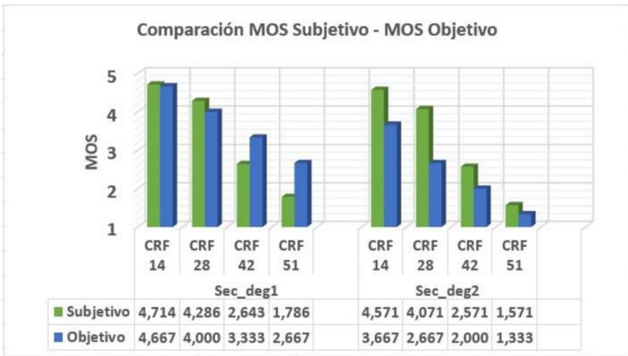

Figure-9. Subjective and objective MOS comparison.

Both the MOS measure of the objective metric designed as the MOS measure of the subjective evaluation had the same tendency as the results of the three videos used during the realization of the project; from here it is concluded that for videos with low mobility, the objective model designed has a high correlation index with the subjective MOS, except in the case where the degradation is too great.

For videos with high mobility, the designed objective model presents a stricter measure than the subjective MOS because it detects such degradations that the user can not perceive.

The MOS rating of the designed model tends to be below the average opinion in the subjective test; however, the difference is not much.

Graphical interface of the objective model

A graphic environment of the objective model (Figure-10) was designed to evaluate the quality in anaglyph 3D videos in order to facilitate its execution.

ARPN Journal of Engineering and Applied Sciences

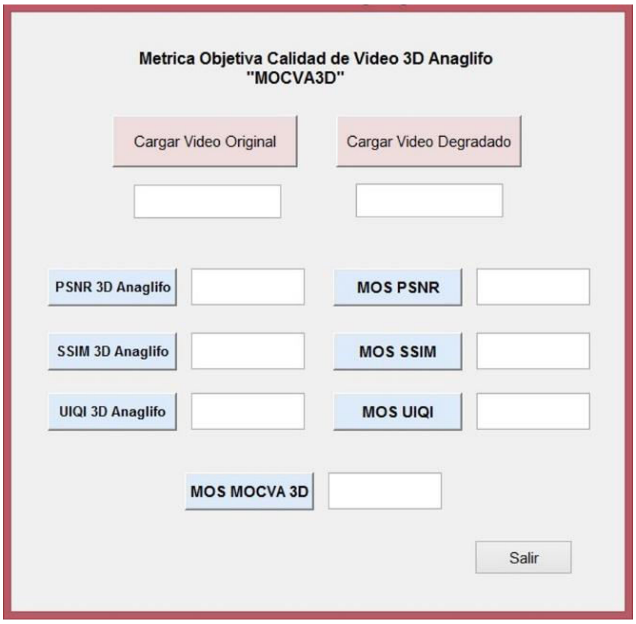

Figure-10. Graphic interface of the designed objective model.

CONCLUSIONS

A contribution was made to the video quality estimation models, proposing the design of an objective model in anaglyphic 3D videos. This purpose was achieved by making a considerable investigation of the metrics used for 2D videos, identifying their characteristics and adapting them to the proposed model. On the other hand, a subjective evaluation test was carried out whose results were used to compare and verify the designed model.

The characteristics of the current coding systems were analyzed with an emphasis on the aspects that are related to the quality of the obtained final video. In particular, the effect of the variation of the constant rate factor (CRF) on the used anaglyph 3D videos was examined. The best ratings obtained correspond to the CRF 14 of each video because this parameter enters more information to the videos at the time of compression, resulting in a higher quality; however, this parameter is dispensable if the aim is to reduce the size of the file since it considerably increases its size in exchange for a better quality.

The objective evaluation model achieves a low computational load and achieves a correlation of more than 80% with the subjective evaluation.

With a confidence level of 95%, it was determined that videos with a high movement index have wide confidence intervals; the opposite occurs when the video has low mobility.

Through the graphic environment developed, a very valuable tool is provided for easy and quick evaluation of the perceptual quality of anaglyph 3D videos.

Through a web application, a database is generated that includes subjective tests of multiple anaglyph 3D videos, which may be used for future research.

REFERENCES

Huynh-Thu Q., Garcia M. N., Speranz, F., Corriveau P., Raake A. 2011. Study of Rating Scales for Subjective Quality Assessment of High-definition Video. Broadcasting, IEEE Transactions. pp. 1-14.

Joskowicz J., Sotelo R., Arado J. 2012. Comparison of Parametric Models for Video Quality Estimation: Towards a General Model. Broadband Multimedia Systems and Broadcasting (BMSB), 2012 IEEE International Symposium. pp. 1-7.

Lukas F. X., Budrikis Z. L. 1982. Picture Quality Prediction Based on a Visual Model. Communications, IEEE Transactions. pp. 1679-1692.

Rec I. T. U. 2008. P. 910: Subjective Video Quality Assessment Methods for Multimedia Applications. International Telecommunication Union, Geneva.

Recommendation I. T. U. R. B. T. 2002. 500-11: Methodology for the Subjective Assessment of the Quality of Television Pictures. 6.

Seshadrinathan K., Soundararajan R., Bovik A.C., Cormack L.K. 2010. Study of Subjective and Objective Quality Assessment of Video. Image Processing, IEEE Transactions. pp. 1427-1441.

Sikor T. 2005. Trends and Perspectives in Image and Video Coding. Proceedings of the IEEE. pp. 6-17.

Tomar S. 2006. Converting video formats with FFmpeg. Linux Journal. 10.

Wang Z., Sheikh H.R., Bovik A.C. 2003. Objective Video Quality Assessment. The Handbook of Video Databases: Design and Applications, CRC Press Ed. pp. 1041-1078.