

Methodology for Objective Evaluation of Video Broadcasting Quality using a Video Camera at the User's Home

Marcio L. Graciano
Dep. of Electrical Engineering
University of Brasilia
Campus Darcy Ribeiro,
Brasilia, Brazil

Alexandre R. S. Romariz
Dep. of Electrical Engineering
University of Brasilia
Campus Darcy Ribeiro,
Brasilia, Brazil

Jose Camargo da Costa
Dep. of Electrical Engineering
University of Brasilia
Campus Darcy Ribeiro,
Brasilia, Brazil

ABSTRACT

In this work, a methodology for objective evaluation of the quality of video programs, without reference, recording these programs in the users' residence using a video camera is presented. The methodology is based on the use of a digital watermark embedded in the original program. The watermark is invisible to the user, but capturable by the video camera. The recorded video is handled by specific software that evaluates the watermark degradation. The measure of degradation of this watermark is used to estimate the quality of the video broadcasting system. A case study is presented to validate the methodology. The results of video quality metrics using this methodology were compared to a standardized full reference metrics and the linear correlation between these metrics was superior to 93%, which indicates a high convergence. The result of video quality metrics were also compared to a pixel based difference metrics, PSNR (Peak Signal to Noise Ratio) and the linear correlation was superior to 99%.

Keywords

video quality, quality metrics, human visual system, modulation transfer function

1. INTRODUCTION

The use of digital video has increased in recent years. Although there have been great advances in compression and transmission techniques, impairments are often introduced along the several stages of a communication system. The visibility and annoyance of these impairments are directly related to the quality of the received/ processed video. For many applications, such as broadcasting, it is important to have a good estimate of the quality of the material being received [10][11].

There is an ongoing effort to develop video quality metrics that are able to detect impairments and estimate their annoyance as perceived by human viewers [1][5][14][16][18][21][23][24]. Most successful video quality metrics are Full Reference (FR) metrics. These metrics estimate the quality of a video by comparing original and impaired videos. Requiring the reference video becomes a serious impediment in many real-time applications. In these cases, it becomes essential to develop ways of blindly estimating the quality of a video using a No-Reference (NR) video quality metric. NR metrics, unfortunately, has a lower performance than FR what makes their use in real applications quite difficult [4].

One possible approach to estimate the quality of video signals without requiring the reference is to use a data hiding or watermarking system. In this approach, a digital mark is embedded into the original video frames before the compression and transmission stages. At the receiver, the mark is extracted and a measure of the degradation of the mark is used to estimate the quality of the test video. This type of metric has the advantage of being fast and not requiring the use of the original video [4]. In this work, our goal is to develop a methodology to implement an objective quality metric based on a watermarking technique [8].

Our approach is different than other approaches in the literature [1][5][14][15][16][18][21][23][24] in the following aspects: First, we acquire the watermarked video using a simple consumer electronic video camera placed in the room where the video is being played (i.e. the video screen is filmed with the video camera in order to capture the broadcasted program). Then, the captured video is processed in order to recover the watermark. Afterwards, a quality metrics function analyzes and determines the watermark level of degradation.

The video quality verification will be done as consumers evaluate quality at their homes using their human visual system (HVS). The methodology was developed in order to replace HVS by the camera and the video processing unit. The evaluation of the quality from this HVS-like system will be done using the contrast sensitivity function (CSF) of the human eye and the Modulation Transfer Function (MTF) of the optical system. Each consumer can verify quality in their environment without interfering with the broadcaster distribution equipment.

This approach has minimum interference in distribution systems and has several advantages when we take into account the heterogeneous environment of broadcasting networks. These networks have a wide range of equipment and technologies used in the distribution of video content.

In order for the metrics to accurately evaluate quality, it is necessary to compensate for different users' environment parameters using the MTF.

2. QUALITY METRICS DEFINITION

In this work, a digital mark is embedded into the reference video frames before the compression and transmission stages. At the receiver, the mark is extracted and a measure of the degradation of the mark is used to estimate the quality of the video received. These operations are done in a processing unit containing a set of programs that extract, decode and analyze

the mark in order to evaluate its quality. These programs were entirely developed in this work.

In this work, those programs run on a PC and the video camera is a model currently available in the consumer market. The processing tasks can be eventually made by a compatible platform (a mobile device like a smart phone or a dedicated device with processing capabilities and a video camera).

The following steps describe the process to extract and decode the digital watermark [13][19]:

—At the reception environment of the consumer, a video camera films the screen where the user watches the video and generates a RAW file.

—The frames of RAW file are separated, generating a frame sequence.

—Each frame is then normalized to a fixed size [13].

Some performance tests were done to choose the best normalization frame size to decode the mark. The performance parameters were: small processing time, low processing and ability to decode the mark using several frames resolutions. The best size for use in SD (standard definition) and HD (high definition) video was 512x512.

—The normalized frames were processed by edge detection. This process finds several areas where the marks could have been inserted [13].

—Each area is tested by a watermarking detection algorithm [13].

The video quality measure is performed by correlating the watermark retrieved at the user's environment and the watermark inserted in the content producer (Eq. (1) [13]).

$$N_c = \frac{\sum_{i=1}^n w(i)\hat{w}(i)}{\sqrt{\sum_{i=1}^n w^2(i)}\sqrt{\sum_{i=1}^n \hat{w}^2(i)}} \quad (1)$$

where N_c is the normalized correlation between marks, n is the mark length, w is the mark inserted during video production and \hat{w} is the mark recovered in the consumer environment. In this work 64 bit vectors were chosen as marks (after some preliminary tests with other vector lengths).

The correlation threshold ($N_{cT} = 0.6$) was chosen so that the mark was invisible to the human observer and identifiable by the processing algorithm. The choice was based on subjective tests [10]. A normalized correlation N_c above N_{cT} indicates the watermark presence.

—The highest N_c value from each area in a given frame becomes N_{cframe} .

—A fixed number of video frames (Num_{frames}) was chosen to calculate the overall N_{cvideo} .

The choice of a large number of frames increases the accuracy of the watermark retrieval. A large number of frames also cause the processing to become slower. An optimum value of Num_{frames} equal to 50 was chosen.

The N_{cframe} values were plotted against the corresponding Num_{frames} . A normal distribution was fitted to the obtained curve. This step removes the N_{cframe} outliers and the 95% percentile is then calculated resulting in the N_{cvideo} .

—This N_{cvideo} is then converted to the same scale of SQF (Subjective Quality Factor) [12]. Eq. (2) shows this scale conversion:

$$N_{cvideoQ} = 100 \times \frac{N_{cvideo} - 0.6}{(N_{cvideo} - 0.4) \times 0.6} \quad (2)$$

The SQF (subjective quality factor) Eq. (3) [12] is calculated using the MTF_T from Eqs. (5) and (6) presented in next section. The SQF is an objective value related with human eye CSF. The SQF was empirically verified in a well conducted observer study of perceived sharpness [9][12].

$$SQF = 100 \times \frac{\int_{v^-}^{v^+} MTF_T(v_d) d(\ln v_d)}{\int_{v^-}^{v^+} d(\ln v_d)} \quad (3)$$

where v_d is the spatial (position in space) frequency in cycles per degree at the retina and the limits of integration are $v^- = 3$ and $v^+ = 12$ cycles per degree [12].

—Using the $N_{cvideoQ}$ value calculated from the watermark recovery Eq. (2) and the SQF value calculated using the optical system MTF Eq. (3), the overall quality metrics FQ Eq. (4) can be evaluated.

$$FQ = a \times N_{cvideoQ} + b \times SQF \quad (4)$$

To find the coefficients a and b , it was performed a nonlinear least-squares data fitting of Eq. (4) to the differential mean opinion scores (DMOS) obtained from subjective tests conducted [10]. The values calculated are $a = 0.501$ and $b = 0.389$.

These values were calculated using the DMOS values according to quality metrics definition methodology steps [6].

3. OPTICAL SYSTEM MTF

The optical system MTF combines all components in the optical path of the system, among which we can highlight the image sensor, the lens, the distances involved between the components and the target image displayed on the video monitor [2]. Figure 1 shows a typical diagram for the optical system MTF. Eq. (5) shows the MTF calculation for the diagram in frequency domain.

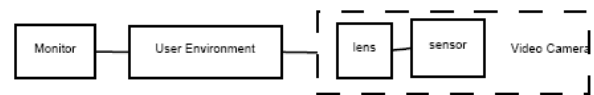


Fig. 1: Optical system MTF diagram

$$MTF_T = MTF_{monitor} \times MTF_{env} \times MTF_{len} \times MTF_{sensor} \quad (5)$$

where MTF_T is the total optical system MTF, $MTF_{monitor}$, MTF_{env} , MTF_{len} and MTF_{sensor} are the MTFs due to monitor, environment, camera lens and camera sensor respectively.

Table 1. Camera and monitor parameters

	Parameter	Value
Camera	sensor size	6.17x4.55mm
Canon	pixels quantity	4000x3000
D10	pixels size	1.54x1.52µm
Monitor	diagonal size	22"(55.88cm)
Samsung	pixels quantity	1920x1080
P2270HN	pixels size	0.248x0.248mm
Environment	visualization distance	100cm

The frequency domain is chosen, because the transfer functions' theory for optical systems can be used, which

makes the MTF calculation simpler and faster [2][7]. In the frequency domain (Fourier transform), the product of the MTFs of the various systems' components gives the optical system MTF.

From Eq. (5) we derive a simplified equation taking in account a term due to MTF_{env} and other due to MTF_{elect} relating to all electronic equipment used, Eq. (6).

$$MTF_T = MTF_{elect} \times MTF_{env} \quad (6)$$

where $MTF_{elect} = MTF_{monitor} \times MTF_{len} \times MTF_{sensors}$ from Eq. (5).

Using the MTF_T obtained with the slanted-edge method [3], it is possible to obtain the values of the SQF from Eq.(3) and the FQ from Eq. (4). The MTF_T allows the calculation of MTF_{elect} values or MTF_{env} values if we fixed one of them while calculating the other.

The $MTF_{elect} = MTF_{elect1}$ fixed value can be calculated, because all electronic equipment do not change. Estimation of MTF_{elect1} could be done averaging measurements of MTF_T while maintaining the value MTF_{env} constant for the test environment.

MTF_T and FQ measurements were done with equipment setup presented in Table 1, where the video camera was mounted fixed on a tripod at distance of 100 cm from the video monitor.

4. RESULTS AND DISCUSSION

In this section we present a case study for validation of the methodology.

In this work we are keeping the camera and monitor parameters fixed as shown in Table 1. Those camera's and monitor's parameters results in MTF_{elect1} to be used to calculate the MTF_{env} .

Three different video sequences were selected (*cactus*, *crowdrun* and *BQTerrace*). These sequences are commonly used in video quality tests [20]. The watermarks were inserted and then, from each watermarked sequence, other sequences were generated applying degradations indicated in column 3 of Table 2. After that step, seventeen (17) test video sequences were available.

Using these sequences and the same equipment of Table 1 the new MTF_{env} can be calculated from the constant MTF_{elect1} previously calculated.

The evaluation of the video quality metrics was done comparing the results obtained by our metrics with another standardized FR metrics known as VQM (Video Quality Metric) [17] available in the literature. The graphic in figure 2 and Table 2 show the correlation between our metrics FQ and the VQM metrics.

Each numbered point in the graphic of Figure 2 represents a pair relating VQM quality to FQ quality for the same video sequence from Table 2. The straight line represents the correlation least-squares line. The metrics VQM and FQ were normalized to show results from 0 to 1, where values presented near 0 means bad quality and values near 1 means excellent quality for both metrics. The distribution of points on the graph is compatible with the types of degradation presented in Table 2. The videos encoded with higher bit rates of 30Mbps and 10Mbps have better quality while others encoded with lower bit rates or with some kind of packet loss have lower quality. These quality values are in agreement with the subjective tests carried out according to [10]. The

results of VQM metrics for videos which present the same type of degradation have values very close with little excursion on the horizontal axis of Figure 2.

The points 3 and 4 plotted by VQM do not match quality by degradation as the others points in Figure 2. The quality of video coded with low bit rate should be plotted to the left. The FQ metrics plots those points in low quality region at the bottom, in accordance with another videos coded with low bit rate.

Our FQ metrics provides a little variation compared to VQM (more excursion on the vertical axis of Figure 2). This range of values follows SQF quality values [9]. The FQ results obtained are consistent with the different types of degradation shown in Table 2. Through our experiment it was demonstrated that our proposed metrics is closely related to VQM in terms of prediction accuracy. A high correlation exceeding 93% (linear correlation, $r=0.9388$) was observed between these two metrics.

The graphic of Figure 3 represents a pair relating PSNR to FQ quality for the same video sequence from Table 2. The linear correlation of our metrics FQ with PSNR is equal to $r=0.9908$.

5. CONCLUSIONS

In this work, a new methodology using optical characteristics of the user environment to evaluate video quality based on a watermarking technique was proposed.

The methodology usage was checked presenting a case study for NR video quality metrics. Our proposed methodology using optical parameters allows utilization in various user environments and technologies of video broadcasting transmission and distribution. The performance of the objective video quality metrics was evaluated comparing with other standardized metrics [17] and the bit difference metrics PSNR.

A correlation of 99% between our metrics and PSNR and a correlation of 93% between our metrics and VQM were attained. As a result one can see that our no-reference metrics was successfully used in place of a full-reference one, which implies that a straight-ahead, almost real-time, low cost, high quality video evaluation methodology was developed and is now available.

A further refinement to improve robustness can consider using quaternion Fourier domain for calculation of watermarks [22].

6. ACKNOWLEDGEMENTS

The authors acknowledge the support of CNPq (Brazilian Council for Science and Technology Development) and INCT-NAMITEC.

7. REFERENCES

- [1] S. Daly. The visible differences predictor: an algorithm for the assessment of image fidelity. In Andrew B. Watson, editor, *Digital Images and Human Vision*, pages 179–206, Cambridge, Massachusetts, 1993. MIT Press.
- [2] J. B. DeVelis and G. B. Parrent. Transfer function for cascaded optical systems. *J. Opt. Soc. Am.*, 57:1486–1490, 1967.
- [3] M. Estrideau and P. Magnan. Fast mtf measurement of cmos imagers using iso 12233 slanted edge methodology. In *SPIE Detectors and Associated Signal Processing*, volume 5251, pages 243–251, 2004.

- [4] M. C. Q. Farias, M. Carli and S. K. Mitra. Objective video quality metric based on data hiding. *IEEE Transactions on Consumer Electronics*, 51:983–992, 2005.
- [5] M. C. Q. Farias and S. K. Mitra. No-reference video quality metric based on artifact measurements. In *IEEE International Conference on Image Processing*, volume 3, pages 141–144, 2005.
- [6] M. C. Q. Farias and S. K. Mitra. A methodology for designing no-reference video quality metrics. In *Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, pages 1–6, 2009.
- [7] J.W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill Physical and Quantum Electronics Series, 1968.
- [8] M. L. Graciano, A. R. S. Romariz and J. C. Costa. Cmos image sensor device for objective evaluation of video quality in mass distribution networks. In *IEEE 7th Consumer Communications and Networking Conference (CCNC)*, pages 1–2, 2010.
- [9] E. M. Grainger and K. N. Cupery. An optical merit function (sqf) which correlates with subjective image judgments. *Photographic Science and Engineering*, 16:221–230, 1972.
- [10] ITU-R. Recommendation BT.500-13, chapter Methodology for the subjective assessment of the quality of television pictures. Recommendations of the ITU, Radiocommunication Sector, 2012.
- [11] ITU-T. Final report from the video quality experts group (VQEG) on the validation of objective models of video quality assessment, volume 4, chapter COM 9-80-E. approved for release at VQEG meeting, 2000.
- [12] B. W. Keelan. Objective and subjective measurement and modeling of image quality: a case study. In *SPIE Applications of Digital Image Processing XXXIII*, volume 7798, pages 779–815, 2010.
- [13] L. Li, B. Guo and L. Guo. Rotation, scaling and translation invariant image watermarking using feature points. *The Journal of China Universities of Posts and Telecommunications*, 15:82–87, 2008.
- [14] W. Lin and C.C. Jay Kuo. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 22(4):297–312, 2011.
- [15] H. Loukil, M. H. Kacem and M. S. BouhleL. A new image quality metric using system visual human characteristics. *International Journal of Computer Applications*, 60(6):32–36, 2012.
- [16] A. K. Moorthy and A.C. Bovik. Visual quality assessment algorithms : What does the future hold? *International Journal of Multimedia Tools and Applications*, Special Issue on Survey Papers in Multimedia by World Experts, 51(2):675–696, 2011.
- [17] M. Pinson and S. Wolf. *Video Quality Measurement Users Manual*. NTIA Handbook HB-02-01, 2002.
- [18] M.H. Pinson and S. Wolf. A new standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, 50:312–322, 2004.
- [19] S. Poongodi and B. Kalaavathi. Comparative study of various transformations in robust watermarking algorithms. *International Journal of Computer Applications*, 58(11), 2012.
- [20] F. De Simone, L. Goldmann, J.S. Lee and T. Ebrahimi. Towards high efficiency video coding: Subjective evaluation of potential coding technologies. *Journal of Visual Communication and Image Representation*, 22(8):734–748, 2011.
- [21] VQEG. Final report from the video quality experts group on the validation of objective models of video quality assessment - Phase II. Tech. Report, 2003.
- [22] X. Wang, C. Wang, H. Yang, and P. Niu. A robust blind color image watermarking in quaternion fourier transform domain. *Journal of Systems and Software*, 86(2):255–277, 2013.
- [23] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13:600–612, 2004.
- [24] S. Winkler and P. Mohandas. The evolution of video quality measurement: From psnr to hybrid metrics. *IEEE Transactions on Broadcasting*, 54:660–668, 2008.

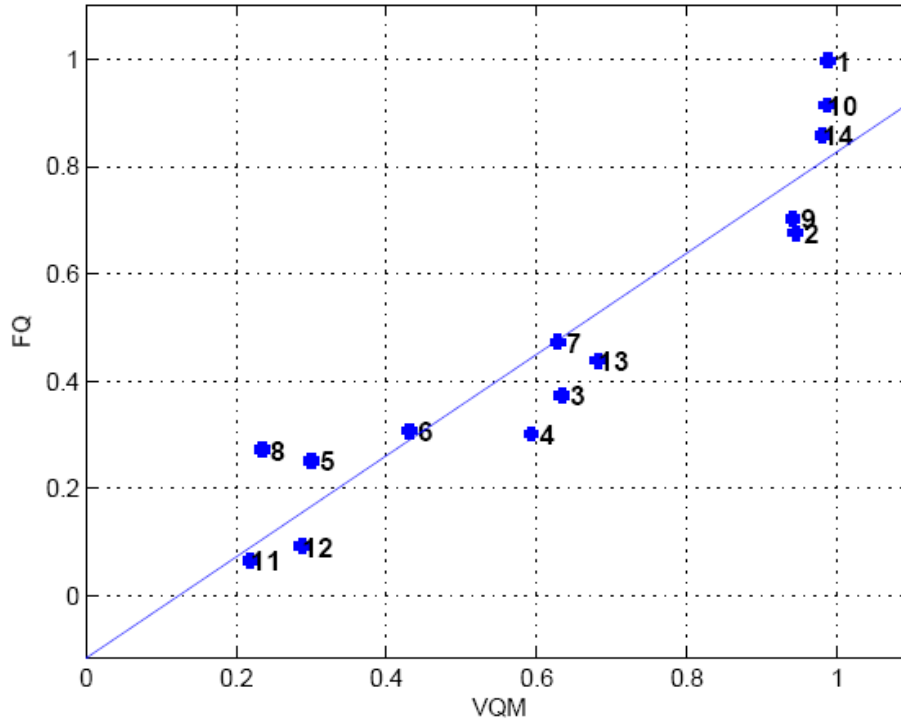


Fig 2: FQ Metrics results compared with VQM [17]

Table 2. Metrics Results used to generate Figure 2 and Figure 3

Sequence	video name	degradation	PSNR	VQM	FQ
1	<i>BQTerrace</i>	video coded H.264 bitrate 30Mbps	48.947	0.987	0.9977
2	<i>BQTerrace</i>	video coded MPEG2 bitrate 10Mbps	46.6079	0.9447	0.6756
3	<i>BQTerrace</i>	video coded MPEG2 bitrate 1Mbps	32.4961	0.6336	0.3734
4	<i>BQTerrace</i>	video coded H.264 bitrate 300kbps	31.6951	0.5927	0.301
5	<i>cactus</i>	video coded H.264 packet loss rate 10%	30.5056	0.3004	0.251
6	<i>cactus</i>	video coded MPEG bitrate 1Mbps	33.6848	0.4307	0.3071
7	<i>cactus</i>	video coded H.264 packet loss rate 1%	37.1515	0.6281	0.4729
8	<i>cactus</i>	video coded H.264 bitrate 300kbps	30.8936	0.2345	0.2718
9	<i>cactus</i>	video coded MPEG bitrate 10Mbps	43.3981	0.9408	0.7022
10	<i>cactus</i>	video coded H.264 bitrate 30Mbps	49.4828	0.9858	0.9142
11	<i>crowdrun</i>	video coded H.264 bitrate 300kbps	25.7758	0.2186	0.065
12	<i>crowdrun</i>	video coded H.264 packet loss rate 10%	27.0725	0.2876	0.0918
13	<i>crowdrun</i>	video coded H.264 packet loss rate 1%	34.1506	0.6813	0.4378
14	<i>crowdrun</i>	video coded H.264 bitrate 30Mbps	45.1649	0.9797	0.8584

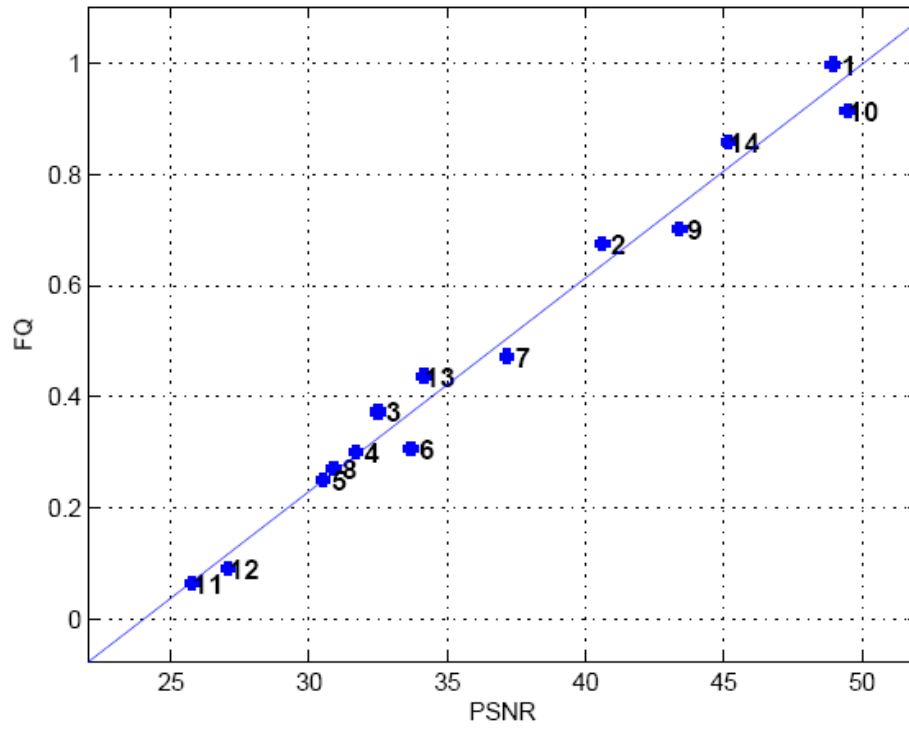


Fig 3: FQ Metrics results compared with PSNR