

VIDEO QUALITY ASSESSMENT USING OBJECTIVE PARAMETERS BASED ON IMAGE SEGMENTATION

A. C. F. Pessoa¹, A. X. Falcão², A. E. Faria e Silva³, R. M. Nishihara¹ and R. A. Lotufo²

¹CPqD, ²UNICAMP, ³TV Globo
Rod. Campinas-Mogi Mirim, Km 118.5
13083-970 - Campinas - SP - Brazil
franca@cpqd.com.br

Resumo - Este trabalho apresenta uma metodologia para a estimação de qualidade de vídeo usando parâmetros objetivos e segmentação. Cenas naturais de vídeo são segmentadas em regiões planas, de bordas e de texturas, e emprega-se um conjunto de parâmetros objetivos definidos sobre cada um destes contextos. Um modelo baseado em percepção que estima notas subjetivas é definido computando-se a relação *entre medidas objetivas e resultados de testes de avaliação* subjetiva sobre um conjunto de cenas naturais processadas por codecs de vídeo MPEG-2. Neste modelo, a relação entre cada parâmetro objetivo e o nível de degradação subjetivo é aproximada por uma curva sigmoidal, resultando num nível de degradação estimado para cada parâmetro. O resultado final é obtido através de uma combinação linear dos níveis de degradação estimados, onde o peso de cada nível de degradação é proporcional à sua confiabilidade estatística. Os resultados apresentados no trabalho indicam que o uso de medidas objetivas baseadas em regiões proporcionam previsões mais precisas que as obtidas com parâmetros globais.

Abstract - This work presents a methodology for video quality assessment using objective parameters based on image segmentation. Natural scenes are segmented into plane, edge and texture regions, and a set of objective parameters are assigned to each of these contexts. A perception-based model that predicts subjective ratings is defined by computing the relationship between objective measures and results of subjective assessment tests, applied to a set of natural scenes and MPEG-2 video codecs. In this model, the relationship between each objective parameter and the subjective impairment level is approximated by a logistic curve, resulting an estimated impairment level for each parameter. The final result is achieved through a linear combination of estimated impairment levels, where the weight of each impairment level is proportional to its statistical reliability. The results presented show that the use of region-based objective measurements provides more accurate predictions compared to predictions based on global parameters.

Keywords: Video quality evaluation, video coding, MPEG, video segmentation.

1. INTRODUCTION

Video quality assessment has become a crucial issue with the increasing use of digital video compression systems and the subsequent video services, such as primary and secondary

distribution of digital TV, video on demand, videophone, videoconference, etc. Due to the flexibility of video coding standards, competing codecs don't provide the same picture quality. Therefore, methods for video quality assessment represent important tools to compare the video quality of competing codecs and to quantify their performance in a large number of applications.

The challenge in developing techniques to estimate the quality of video compression systems stems, in part, from the fact that compression algorithms introduce video impairments which are strongly dependent of the levels of details and motion in the scenes. Moreover, the visual perception of video impairments also depends on the details and motion of the scenes. Thus traditional evaluation methods, which are based on static test signals, are inadequate to quantify the performance of video compression systems.

This work presents a methodology [1] for video quality assessment, when the video is processed by unidirectional transmission systems that use digital interfaces and, ideally, digital transport facilities. The method has been applied to assess video compression systems according to MPEG standards [2] [3], but it could be also used to evaluate other types of systems, such as video codecs based on other analysis techniques (i.e. wavelets and prediction filters) and composite signal encoders/decoders.

Figure 1. shows the configuration of the objective parameters computation process used for video quality estimation. The file format of the input and the output digital video signals is YCbCr4:2:2, as determined by Recommendation ITU-R BT.601 [4].

In Figure 1., each objective parameter is computed separately within the following contexts of the scenes: plane regions, edge regions and texture regions. This is one of the most important aspects of this methodology. A blocking distortion, for example, can be measured by an edge detector applied to the plane regions of the video scene, wherein the visual perception of this distortion is more noticeable. The computational complexity of the method is reduced by using low-complexity estimators and by constraining their computation within the correspondent contexts of the scenes. These contexts are defined by an image segmentation algorithm that is applied to the original natural scenes (i.e. the input test signal). This type of algorithm normally requires high computational complexity, however it is executed only once. Note that the spatial and temporal registration between the input and output video signals and the correction of

gain and offset are also required to compute the objective parameters correctly. The information about registration (or alignment) and calibration is addressed in [5].

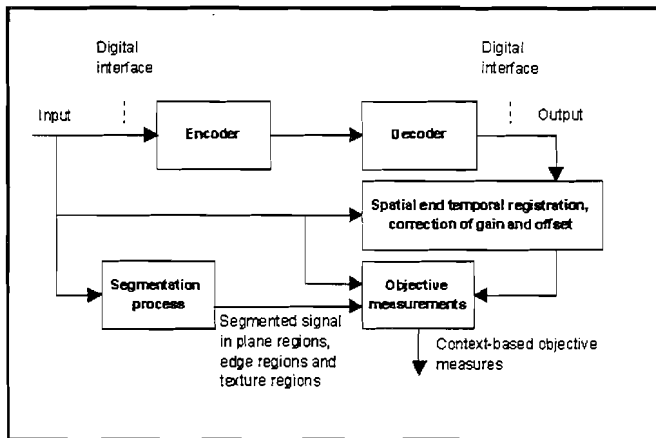


Figure 1. Objective parameters computation.

The objective parameters are computed by direct comparison between original and impaired scenes. All estimators are applied to fields rather than frames of video in order to ensure the statistical reliability of the measures in scenes with high level of motion.

A perceptual-based model that predicts subjective ratings is defined by computing the relationship between objective measures and results of subjective assessment tests applied to a set of natural scenes and MPEG-2 video codecs. These scene-dependent perceptual models are defined in two steps as follows:

I. The relationship between each parameter and the subjective impairment level is approximated by a logistic curve, resulting an estimated impairment level for each parameter;

II. The final result is achieved by linearly combining the estimated impairment levels, where the weight of each impairment level is proportional to its statistical reliability.

The details of the just outlined methodology are presented in the next sections. Section 2 gives a short description of the adopted setup for the subjective evaluation tests. The methods to determine the objective parameters and to segment the natural scenes are described in Section 3. Section 4 introduces the perceptual models for subjective quality estimation. Section 5 reports the results that were obtained in this study and points out the advantages of using objective parameters based on image segmentation for subjective quality estimation and the dependence of the perceptual models with the assessors category and with the viewing distance from the monitor (4H or 6H). The conclusions of this work are presented in Section 6.

2. SUBJECTIVE ASSESSMENT TESTS

The image processing laboratory of the CPqD (Research and Development Center) has a special room for subjective evaluation trials, according to the Recommendation ITU-R BT.500-7 [6]. This room was used to evaluate the performance of some manufactured and simulated MPEG-2

video codecs on a subset of the natural scenes suggested by the Recommendation ITU-R BT.802 [7]. The manufactured MPEG-2 codecs were provided by the TV Globo (a TV company). The scenes were also processed by the MPEG-2 coding software available at CPqD.

Table 1. presents a summary of the conditions used for the subjective assessment tests.

Conditions for evaluation	According to Recommendation ITU-R BT.500-7 [6]
Source of signals	D1 VTR
Monitor	20" studio monitor with digital interface
Viewing distances	4H and 6H
Assessment method	DSIS (Double-Stimulus Impairment Scale) method with nine points in the interval from 1 to 5
Test sequences	5 scenes
Presentation duration	10 sec. (original signal) + 3 sec. (gray signal) + 10 sec. (signal under evaluation) + 5 sec. for vote
Assessors	14 experts and 34 non-experts
Assessors per session	5
Sessions per assessor	2
Presentations per session	48
Source of signals	D1 VTR
Source of signals	D1 VTR
Processing conditions	26

Table 1. Conditions of the subjective assessment tests

The subjective evaluation sessions utilized a set of 5 natural scenes given in Table 2.), which are defined as test sequences for conventional TV in the Recommendation ITU-R BT.802 [7].

Sequence Name	Scene Number in the Rec. 802 [7]
Flower Garden	15
Mobile and Calendar	30
Table Tennis	29
Diva with Noise	17
Kiel Harbour-4	26

Table 2. Natural scenes used for subjective evaluation

In total, 26 systems were included in the sessions of subjective evaluation. These items are presented in Table 3..

3. OBJECTIVE MEASUREMENTS BASED ON CONTEXT

This section describes the video material used for the objective evaluation (i.e. the material used for objective parameters computation), suggests three image segmentation methods that can be used to divide the video material into plane, edge and texture regions, and presents the objective parameters that have been adopted in this study.

The video material used for objective evaluation consists of a 10 seconds long video sequence, composed by 5 clips

Group	Type	Characteristics	Processing conditions
1	Manufactured MP@ML MPEG-2 codec for CBR (constant bit rate) applications	Bit rates: 5, 10 and 15 Mbits/s; N=12 and M=2	6
2	Simulated MP@ML MPEG-2 codec for CBR applications	Bit rates: 2.5, 5, 7.5, 10, 12 and 15 Mbits/s; N=12 and M=1 and 2	12
3	Manufactured 422P@MPL MPEG-2 codec for CBR applications	Bit rate: 18 Mbits/s; N=2 and M=2	1
4	Simulated MP@ML MPEG-2 codec for VBR (variable bit rate) applications using intra-frame coding only	Fix quantizer scale [3] in 4, 8, 16, 32 and 62	5
5	Composite signal conversion	NTSC and PAL-M	2

Table 3. Processing conditions

of natural scenes. These five clips, 2 seconds long each, were selected from the natural scenes presented in Section 2. The purpose of using 2 seconds clips instead of 10 seconds clips, as in the subjective tests described in Section 2, was to reduce the computational complexity of the objective evaluation process. The choice of 2 seconds clips was based on the following criteria:

I. The clip of 2 seconds of a given scene represents a critical segment of its 10 seconds material compared to the mean criticality of the scene. This criticality was defined as the number of bits per frame resulting from the coding process of a MP@ML MPEG-2 codec (N=12 and M=2) with variable bit rate and quantizer scale equals 16.

II. The clip of 2 seconds of this scene also represents a critical segment from the subjective point of view, when the scene is processed by a MP@ML MPEG-2 codec (N=12 and M=2) at 5 Mbits/s.

There were developed three algorithms for image segmentation [8]. The first is an image segmentation algorithm based on edge detection using recursive filtering, the second is a fuzzy image segmentation algorithm based on spatial features and the third is an image segmentation algorithm based on watershed. The strategy in these segmentation algorithms is to classify the luminance component of each field of video into three mutually exclusive contexts: plane regions, edge regions and texture regions. These algorithms are shortly described as follows:

Algorithm I: image segmentation based on edge detection using recursive filtering

This algorithm initially classifies each pixel, based on the brightness variance computed within a neighborhood of the pixel, as belonging or not belonging to the plane regions of

the image. The resulting binary image is then smoothed by a median filter [8]. The algorithm also applies to the original image an edge detector based on recursive filtering. The edges on the boundary of the plane regions are classified as belonging to the edge regions. The texture regions are the remaining regions of the image.

As an example, Figure 2. shows part of the scene Mobile and Calendar. The result of segmentation by Algorithm I of this part can be seen in Figure 3.. Note that the plane regions are represented by white pixels, edge regions by gray pixels and texture regions by black pixels.

Algorithm II: fuzzy image segmentation based on spatial features

This algorithm is divided into two steps. In the first step, the algorithm assigns a membership function, defined in the interval [0,1], to each one of the three contexts under classification. In the membership function of the plane regions, the membership value of a pixel is defined inversely proportional to the brightness variance computed within a neighborhood of the pixel. The morphological gradient [9] applied to this function defines the membership function of the edge regions. The complement of the fuzzy union [10] between these two membership functions defines the membership function of the texture regions. In the second step, each pixel is classified as belonging to the context with highest value of membership among its three membership values determined in the previous step.

Algorithm III: image segmentation algorithm based on watershed

This algorithm first simplifies the luminance component by increasing its homogeneous regions through the application of an edge-preserving smoothing filter [11]. Secondly, a watershed algorithm is applied to the morphological gradient of the simplified image. The watershed detects homogeneous regions, denoted catching basins, with specified minimum relative contrast. The plane regions are the catching basins with area greater than a threshold. The texture regions are given by the erosion of the complement of the plane regions. The edge regions are the remaining regions of this process.

The objective parameters are obtained for each context (plane, edge and texture) and from the samples of luminance and chrominance of the input (Y_{ref} , Cb_{ref} and Cr_{ref}) and output (Y_{dec} , Cb_{dec} and Cr_{dec}) signals, after spatial and temporal registration and correction of gain and offset, as shown in Figure 1..

The measures MSE (Mean Square Error), PSD (Positive Sobel Difference), NSD (Negative Sobel Difference) and ASD (Absolute Sobel Difference), and the underlying process to compute them are described as follows.

Let $X(i, j)$ be the j -th pixel of the i -th line of the input signal, $Z(i, j)$ be the j -th pixel of the i -th line of the output signal and the elements $X_m(i, j)$ and $Z_m(i, j)$ be the pixels of the input and output signals, respectively, after a median filtering.

The computation of MSE in a context \mathcal{R} (plane, edge or



Figure 2. Part of the scene Mobile and Calendar.

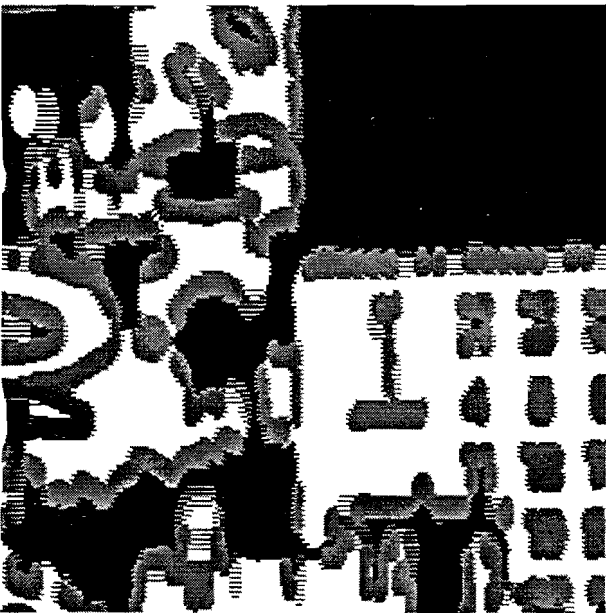


Figure 3. Result of segmentation.

texture) is defined by the mean value of

$$SE(i, j) = [X(i, j) - Z(i, j)]^2,$$

where $(i, j) \in \mathcal{R}$.

The computation of PSD in a context \mathcal{R} (plane, edge or texture) is defined by the mean value of

$$PS(i, j) = \max[S\{X_m(i, j)\} - S\{Z_m(i, j)\}, 0],$$

where $(i, j) \in \mathcal{R}$ and S is Sobel operator.

The computation of NSD in a context \mathcal{R} (plane, edge or texture) is defined by the mean value of

$$NS(i, j) = -\max[S\{Z_m(i, j)\} - S\{X_m(i, j)\}, 0],$$

where $(i, j) \in \mathcal{R}$ and S is Sobel operator.

The computation of ASD in a context \mathcal{R} (plane, edge or texture) is defined by the mean value of

$$AS(i, j) = |S\{X_m(i, j)\} - S\{Z_m(i, j)\}|,$$

where $(i, j) \in \mathcal{R}$ and S is Sobel operator. In other words, $ASD = PSD + NSD$.

The objective parameters utilized for subjective quality estimation refer to the mean value of the aforementioned measures computed over a set of $2N$ fields of the final portion (i.e., in regime condition) of each one of the 5 clips indicated in Table 2. and that belong to the scenes submitted to subjective evaluation. The value N is a multiple of the interval between intra-frame coded images (type I), that is, it is a multiple of the GOP length [2] [3]. In order to satisfy this condition to all systems defined in Table 3., it is chosen $N = 12$.

4. SUBJECTIVE QUALITY ESTIMATION

This section describes how the subjective quality estimation models are defined for each scene. Initially, a perceptual model to estimate the subjective impairment level based on a single parameter is described. The results of this approximation for each objective parameter are linearly combined to estimate the final subjective impairment level.

For each scene, the relationship between each objective parameter D and the subjective result U is initially defined as follows.

A normalized impairment level between 0% and 100% is defined by [6] as:

$$d = \frac{U_{max} - U}{U_{max} - U_{min}} 100\%.$$

The relationship between d and each objective parameter D is approximated by the following logistic function [6]:

$$\lambda = \frac{1}{1 + (D_M/D)^G} 100\%,$$

where the values D_M and G are computed in order to minimize the mean square error

$$e = E[(d - \lambda)^2]$$

for each scene and each objective parameter separately. The statistical reliability of λ is defined as $1/e$.

The estimation of the normalized impairment level d by a set of estimated impairment levels $\lambda_1, \lambda_2, \dots$ (one per parameter) is implemented in three steps of linear prediction as described below.

Step 1

First consider the following sets of estimated impairment levels selected for the luminance component: $\{\lambda_{MSE}\}$, $\{\lambda_{PSD}, \lambda_{NSD}\}$, $\{\lambda_{ASD}\}$, $\{\lambda_{MSE}, \lambda_{PSD}, \lambda_{NSD}\}$ and $\{\lambda_{MSE}, \lambda_{ASD}\}$, where λ_X is the impairment level based on the objective parameter X .

For a given scene and context of this scene (plane, edge or texture), the best set is the one with the least

prediction error. By using this criterion to choose a set of estimated impairment levels for each context, this step linearly combines the impairment levels of each selected set and outputs three estimation values (one per context) denoted by: λ_{YP} , λ_{YE} and λ_{YT} .

Similarly, the considered sets of estimated impairment levels for the chrominance components of the scene are $\{\lambda_{MSE(Cb)}, \lambda_{MSE(Cr)}\}$ and $\{\lambda_{ASD(Cb)}, \lambda_{ASD(Cr)}\}$, and the three resulting estimation values (one per context) are denoted by: λ_{CP} , λ_{CE} and λ_{CT} .

Step 2

The estimation values λ_P , λ_E and λ_T result from linear prediction based on the pairs $\{\lambda_{YP}, \lambda_{CP}\}$, $\{\lambda_{YE}, \lambda_{CE}\}$ and $\{\lambda_{YT}, \lambda_{CT}\}$, respectively.

Step 3

The estimation values λ_P , λ_E and λ_T are combined by linear prediction to produce the estimated impairment level λ .

In all steps above, the predictors satisfy the following restrictions:

Let $\{\lambda_1, \lambda_2, \dots, \lambda_P\}$ be the input vector of the linear predictor. The output λ is given by:

$$\lambda = \sum_i \alpha_i \lambda_i,$$

where the weights $\{\alpha_i\}$ are computed in order to minimize the mean square error

$$E[(d - \lambda)^2]$$

with the conditions

$$\sum_i \alpha_i = 1$$

$$\frac{\alpha_i}{\alpha_k} = \frac{e_k}{e_i}$$

where $1/e_i$ is the statistical reliability of λ_i .

It has been observed that this type of prediction is more robust than the one obtained by optimum predictors, because it is less dependent of the training data base. It reaches better results when applied to test data bases, as exemplified in Section 5.

5. PRESENTATION AND DISCUSSION OF RESULTS

This section is divided into three topics. In Topic 1, the results and prediction models obtained by the subjective quality estimation based on the Algorithm I (the image segmentation algorithm previously described in Section 3) are described. The dependence between the perceptual models and the assessors category (experts and non-experts) and between the perceptual models and the viewing distance from the monitor (4H and 6H) are presented. The variation of the

estimation accuracy with the image segmentation algorithms is discussed in Topic 2. The Topic 3 points out the advantages of the purposed subjective estimation method compared to other methods that are based on global measurements or optimal prediction.

Topic 1: perceptual models and performance

This topic presents the results of the subjective estimation method based on the algorithm I for segmenting the following scenes: Garden (Flower Garden), Mobile (Mobile and Calendar), Tennis (Table Tennis), Diva (Diva with Noise) and Kiel (Kiel Harbour-4), separately.

Tables 4. and 5. show the relative subjective weights of impairments and the dependence between perceptual models and results for non-expert and expert assessors.

The relative subjective weights of impairments in plane regions (w_P), edge regions (w_E) and texture regions (w_T) are equivalent to the weights of the linear prediction described in Step 3 of Section 4:

$$w_P = \alpha_P$$

$$w_E = \alpha_E$$

$$w_T = \alpha_T$$

The relative subjective weights of impairments in luminance (w_Y) and chrominance (w_C) components are given by:

$$w_Y = \alpha_P \alpha_{YP} + \alpha_E \alpha_{YE} + \alpha_T \alpha_{YT}$$

$$w_C = \alpha_P \alpha_{CP} + \alpha_E \alpha_{CE} + \alpha_T \alpha_{CT}$$

using the weights of the linear predictions described in Steps 2 and 3 of Section 4.

The mean value computed over all scenes is given at the last line of these tables.

Scene	w_P	w_E	w_T	w_Y	w_C
Garden	.13	.37	.51	.52	.48
Mobile	.83	.07	.09	.73	.27
Tennis	.45	.13	.42	.68	.32
Diva	.27	.59	.14	.73	.27
Kiel	.32	.39	.29	.57	.43
Mean	.40	.31	.29	.62	.38

Table 4. Perceptual models: non-expert assessors

Comparing the perceptual models based on the opinion of expert and non-expert assessors, the weight of the impairments in chrominance is slightly greater for the models based on expert assessors.

Tables 6. and 7. show the dependence between perceptual models and results for 6H and 4H viewing distances from the monitor (each case with 50% of the total number of assessors).

Comparing the perceptual models based on 4H and 6H viewing distances, the weight of the impairments in edge and

Scene	w_P	w_E	w_T	w_Y	w_C
Garden	.12	.53	.34	.51	.49
Mobile	.72	.13	.15	.72	.28
Tennis	.47	.12	.41	.70	.30
Diva	.22	.42	.36	.55	.45
Kiel	.43	.36	.21	.47	.53
Mean	.39	.31	.30	.58	.42

Table 5. Perceptual models: expert assessors

Scene	w_P	w_E	w_T	w_Y	w_C
Garden	.15	.40	.45	.51	.49
Mobile	.83	.08	.09	.77	.23
Tennis	.47	.15	.38	.60	.40
Diva	.41	.36	.22	.64	.36
Kiel	.34	.40	.26	.54	.46
Mean	.44	.28	.28	.61	.39

Table 6. Perceptual models: 6H viewing distance

Scene	w_P	w_E	w_T	w_Y	w_C
Garden	.09	.35	.55	.47	.53
Mobile	.71	.13	.16	.62	.38
Tennis	.42	.10	.48	.74	.26
Diva	.21	.46	.33	.54	.46
Kiel	.31	.46	.23	.57	.43
Mean	.35	.30	.35	.59	.41

Table 7. Perceptual models: 4H viewing distance

texture regions is significantly greater for the models based on 4H viewing distance, as expected.

In Table 8., the following parameters are presented: I. MSE (mean square error) II. MAE (mean absolute error) III. CORR (Pearson linear correlation coefficient) IV. RANK (Spearman rank order correlation coefficient) computed between the normalized impairment level d and the estimated impairment level λ , taking into account a normalization scale from 0% to 100%.

The results presented in Table 8. refer to the subjective scores of the 34 non-expert assessors over the set of 130 test conditions (5 scenes and 26 processing conditions) of the Table 1..

MSE	18.8		
MAE	3.1		
CORR	0.98		
RANK	0.92		

Table 8. Performance

Topic 2: the variation of the estimation accuracy with the image segmentation algorithm

The results of the subjective quality estimation, based on the Algorithms I, II and III (shortly described in Section 3) and obtained from the scores of the 34 non-expert assessors, are shown in Table 9..

The mean square error (MSE) between the normalized

impairment level d and the estimated impairment level λ , taking into account a normalization scale from 0% to 100%, are shown in this table. The error between the mean normalized impairment level and the mean estimated impairment level, computed over all scenes, is shown at the last line.

Scene	Algorithm I	Algorithm II	Algorithm III
Garden	18.1	18.0	15.8
Mobile	24.2	20.7	18.7
Tennis	25.3	28.5	31.3
Diva	5.4	7.1	7.4
Kiel	22.7	25.9	22.4
Global	6.2	7.4	6.5

Table 9. Results: algorithms I, II and III

Comparing the results given in this table, the estimation accuracy presented small variations for each given scene depending on the image segmentation algorithm. On the other hand, there was no relevant variation in the global estimation accuracy considering the three image segmentation algorithms. This suggests that even simpler image segmentation

Topic 3: the advantages of the adopted subjective quality estimation method

The example illustrated in Table 10. focuses two very important properties of any method for subjective quality estimation based on objective parameters: accuracy and robustness [12]. This example compares the adopted subjective quality estimation method, which relies on context-based objective measurements and the linear prediction method described in Section 4, with the following methods:

- I. a method that relies on the same context-based objective measurements, but uses optimal prediction;
- II. a method that uses the linear prediction described in Section 4, but adopts global measurements.

The example used the systems of the Group 2 and the NTSC system of the Group 5 as training data base, and the systems of the Group 1 and the PAL-M system of the Group 5 as test data base (see Table 3.). The objective parameters used in this example were the ones based on MSE and described in Section 3. The results were obtained from the scores of the non-expert assessors. The input values of the table are the mean square error (MSE), the mean absolute error (MAE), the Pearson linear correlation coefficient (CORR) and the Spearman rank order correlation coefficient (RANK) over the set of 5 scenes and all processing conditions used in each case (training or test).

The advantage of computing objective parameters based on context becomes clear when the procedure described in Section 4 is also applied to global measurements. Note that the use of context-based measurements can significantly improve the estimation results.

The example also shows that the prediction process described in Section 4 is more robust (i.e. it is less dependent of the training data base) when it is compared to the optimum

	Adopted method	
	Training	Test
MSE	11.8	54.9
MAE	2.4	4.8
CORR	0.98	0.77
RANK	0.92	0.77
	Optimum predictor	
	Training	Test
MSE	7.0	103.9
MAE	2.1	6.8
CORR	0.99	0.59
RANK	0.95	0.53
	Global measurements	
	Training	Test
MSE	61.8	109.7
MAE	4.2	6.5
CORR	0.88	0.69
RANK	0.84	0.81

Table 10. Comparison: robustness and accuracy

predictor, improving the prediction results on the test data base.

6. CONCLUSIONS

This work presents a methodology for subjective quality estimation using objective parameters based on image segmentation. The objective parameters are computed within plane regions, edge regions and texture regions resulting from the image segmentation process.

The results presented show that the use of context-based objective parameters compared to global parameters leads to more accurate predictions. This aspect is reinforced by the use of the perceptual model based on the linear prediction method described in Section 4. This method has led to more robust prediction results when it is compared to the optimal prediction.

These results can be further improved if the temporal information is included in the image segmentation process (e.g. edge regions could be further classified into edge regions with low motion and edge regions with high motion) or if the plane, edge and texture regions of chrominance are also considered in the image segmentation process, since the Algorithms I, II and III were used to segment the luminance component only.

REFERENCES

- [1] Contribution Number COM12-39, "Video quality assessment using objective parameters based on image segmentation," *ITU-T Study Group 12*, Brazil, December 1997.
- [2] IS 11172, "Information technology - Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s," *ISO/IEC*, 1992.

- [3] Recommendation H.262, "Generic Coding of Moving Pictures and Associated Audio - Information: Video," *ITU-T*, 1995.
- [4] Recommendation BT.601-5, "Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios," *ITU-R*, 1995.
- [5] T1.801.03-1996, "Telecommunications - Digital transport of one-way video signals - Parameters for objective performance assessment," *ANSI*, 1996.
- [6] Recommendation BT.500-7, "Methodology for the subjective assessment of the quality of television pictures," *ITU-R*, 1995.
- [7] Recommendation BT.802-1, "Test pictures and sequences for subjective assessment of digital codecs conveying signals produced according to Recommendation 601," *ITU-R*, 1994.
- [8] Gonzalez and P. Wintz, *Digital Image Processing*, Addison Wesley, 1987.
- [9] Dougherty, *An Introduction to Morphological Image Processing*, SPIE Optical Engineering Press, vol. TT9, 1992.
- [10] A. Kaufmann, *Introduction to The Theory of Fuzzy Subsets*, Academic Press, vol. 1, 1975.
- [11] J. Barrera, J. F. Banon, and R. A. Lotufo, "Mathematical Morphology Toolbox for the Khoros System," Conference on Image Algebra and Morphological Image Processing, V International Symposium on Optics, Imaging, and Instrumentation, SPIE's Annual Meeting, San Diego, USA, pp. 24-29, July 1994.
- [12] Contribution Number T1A1.5/96-121, "Objective and subjective measures of MPEG video quality *ANSI*, GTE Labs., NTIA/TTS, October 1996.

Antonio Claudio França Pessoa was born in Rio de Janeiro, Brazil in 1963. He received the B.S. degree from the Federal University of Paraíba, Brazil in 1985 and the M.S. degree from the State University of Campinas, Brazil in 1990, all in electrical engineering. Since 1987, he has been at CPqD (Research and Development Center), where has performed research in video and audio coding, video quality evaluation and channel coding. His main interests include TV/HDTV digital transmission, voice recognition and synthesis.

Alexandre Xavier Falcão received a B.S. in Electrical Engineering in 1988 from the Federal University of Pernambuco, Brazil. He has worked in image processing since 1991. In 1993, he received a the M.S. degree in Electrical Engineering from the State University of Campinas, Brazil. During 1994-1996, he worked on user-steered image segmentation methods for his PhD. In 1996, he got his doctorate in Electrical Engineering from the State University of Campinas, Brazil. He worked on video quality assessment methods at CPqD from 1998-1999. Currently, he

is an Assistant Professor at the State University of Campinas, Brazil.

Ana Eliza Faria e Silva graduated from the State University of Campinas, Brazil, where she also received her MSc degree in Electrical Engineering in 1997, with a thesis on subjective evaluation of digital compressed images. After graduating, she joined the Digital Signal Processing Division at CPqD (Research and Development Center), where she worked for two years. She is working at Globo TV Network since 1998. Her current interests include the evaluation of new broadcast systems and standards, in particular, high definition television.

Ricardo Massahiro Nishihara was born in Campinas, Brazil, on November 27, 1968. He received the B. E. E. and M. S. degrees from UNICAMP (State University of Campinas), Brazil, in 1992 and 1995. Since 1994 he has been employed by CPqD (Research and Development Center), Campinas, Brazil, where he is a member of the Digital Signal Processing Division. His technical experience has been in areas of digital image and video coding, image quality assesment, digital television (HDTV), and digital signal processing.

Roberto Alencar Lotufo received the Electronic Engineering Diploma from Instituto Tecnológico de Aeronáutica, Brazil, in 1978, the M.S. degree from State University of Campinas, UNICAMP, Brazil, in 1981, and the Ph.D. degree from the University of Bristol, UK, in 1990, all in electrical engineering. Since 1981, he is within the Department of Computer Engineering and Industrial Automation, at State University of Campinas (UNICAMP), Brazil. His principal interests are in the areas of image processing applications, multidimensional image processing and analysis, mathematical morphology, video processing, and medical imaging.