INTERNATIONAL TELECOMMUNICATION UNION
TELECOMMUNICATION STANDARDIZATION SECTOR
STUDY PERIOD 1997 - 2000

COM 12- -E
December 1997
Original: English

Question: 11/12

STUDY GROUP 12

SOURCE* : BRAZIL

TITLE: VIDEO QUALITY ASSESSMENT USING OBJECTIVE PARAMETERS BASED ON IMAGE SEGMENTATION

Abstract

This contribution 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 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. 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 in this contribution show that the use of region-based objective measurements provides more accurate predictions compared to predictions based on global parameters.

* Contact Person: Antonio C. França Pessoa
CPqD/TELEBRÁS
Brazil
Tel.: +55 19 789 6746
Fax: +55 19 789 6833
Email: franca@cpqd.com.br

VIDEO QUALITY ASSESSMENT USING OBJECTIVE PARAMETERS BASED ON IMAGE 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 contribution presents a methodology 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 standard [1,2], 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 [3].

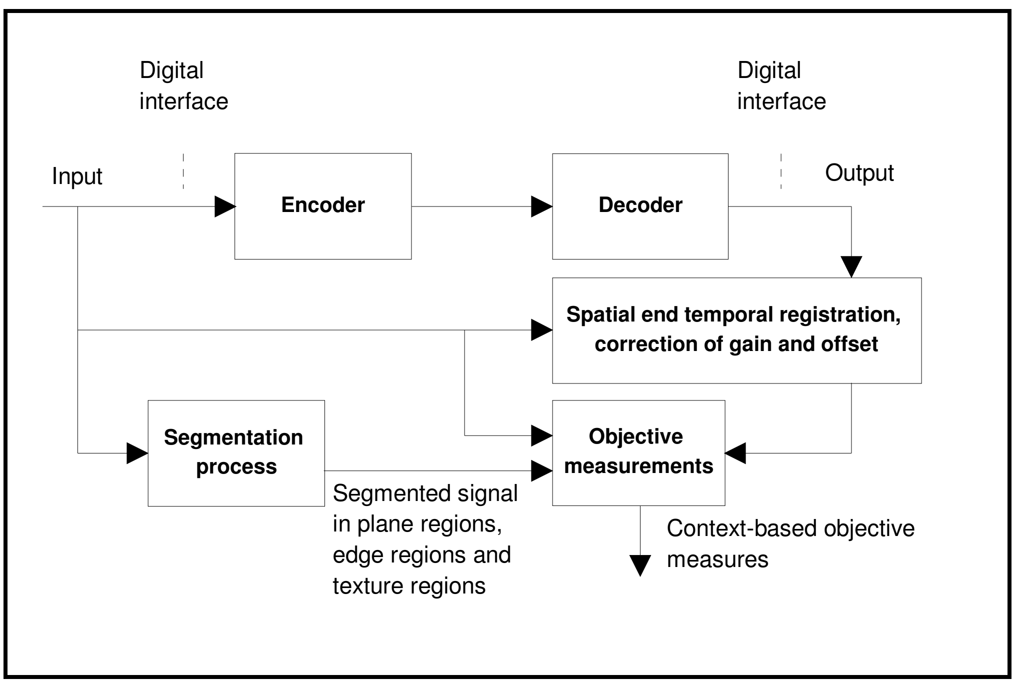

Figure 1 - Objective parameters computation

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 [4].

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 and reports the results that were obtained in this study. This section also 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 contribution are presented in Section 5.

2. Subjective assessment tests

The image processing laboratory of the CPqD/TELEBRÁS (Brazilian Research and Development Center for Telecommunications) has a special room for subjective evaluation trials, according to the Recommendation ITU-R BT.500-7 [5]. 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 [6]. The manufactured MPEG-2 codecs were provided by the TV Globo (a Brazilian TV company). The scenes were also processed by the MPEG-2 coding software available at CPqD/TELEBRÁS.

A short description of the setup utilized for subjective evaluation of the aforementioned MPEG-2 video codecs is given as follows.

2.1. Sessions of subjective evaluation

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

Conditions for evaluation	According to the item 2.1 of the Recommendation ITU-R BT.500-7 [5]
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 [5].
Test sequences	5 scenes of conventional definition digital TV (see Section 2.2)
Presentation duration	10 seconds (original signal) + 3 seconds (gray signal) + 10 seconds (signal under evaluation) + 5 seconds for vote, as suggested by Figure 3.a of the Recommendation ITU-R BT.500-7 [5]
Assessors	14 experts and 34 non-experts
Assessors per session	5
Sessions per assessor	2
Presentations per session	48
Assessed items	See Section 2.3
Presentation of results	Mean and standard deviation of the impairment level regarding to the reference signal (original scene) Discarding of scores and assessors as suggested by the Recommendation ITU-R BT.500-7 [5]

Table 1 - Conditions of the subjective assessment tests

2.2. Natural scenes

The subjective evaluation sessions utilized a set of 5 natural scenes (see Table 2), which are defined as test sequences for conventional TV in the Recommendation ITU-R BT.802 [6]:

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

Table 2 - Natural scenes used for subjective evaluation

2.3. Systems under test

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

Group	Type	Characteristics	Assessed items
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 [2] in 4, 8, 16, 32 and 62	5
5	Composite signal conversion	NTSC and PAL-M	2

Table 3 - Systems under test

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 - Section 3.1), suggests three image segmentation methods that can be used to divide the video material into plane, edge and texture regions (Section 3.2), and presents the objective parameters that have been adopted in this study (Section 3.3).

3.1. Video material used for objective evaluation

The video material used for objective evaluation consists of a 17 seconds long video sequence, composed by 10 clips of natural scenes and 2 clips of artificial test signals.

Five clips of natural scenes, 2 seconds long each, were selected from the natural scenes presented in Section 2.2. 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:

- 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.
- 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.

The five other clips, 1 second long each, consist of scenes with low or no motion. These scenes have been used in the objective evaluation process, intercalating the previous 2 seconds long clips, in order to test the adaptive behavior of the MPEG-2 video codecs (i.e. the behavior regarding to rate and quality control, performance in regime and scene transition). They are also specified in the Recommendation ITU-R BT.802 [6]. Although this is not part of the scope of this contribution, it is important to say that the determination of the performance variation (dispersion of the signal-noise ratio) after each scene transition and in regime (difference of performance over I, P and B frames) has been used to characterize the dynamic behavior of manufactured MPEG-2 video codecs.

The artificial test signals are (1 second long each):

- Narrow-band noise [4] - static and trichromatic video signal defined by noise with resolution about 1/25 of the Nyquist's limit and with approximately uniform histogram for each of components Y, Cb and Cr.
- Circular zone-plate [4] - static and trichromatic video signal defined by a sinusoidal pattern for the components Y, Cb and Cr, with constant horizontal and vertical frequencies along the same column and along the same line of a given field of video, respectively, and outward crescent frequencies from the center of the image.

These artificial signals have been used to determine the following parameters:

- Displacement of active video;
- Active video area;
- Gain and offset;
- 2D frequency response;
- Displacement between chrominance and luminance (a vertical displacement between these components has been noticed very often in manufactured MP@ML MPEG-2 systems, due to the conversions YCbCr4:2:0 ⇔ YCbCr4:2:2, creating a halo of spurious chromaticity on the edges of the output signal).

The contiguous test material of 17 seconds is described in Table 4:

Time Code (mm:ss:ff)	Scene	Short Name	Temporal Characteristic	Duration (seconds)
00:00:00	Narrow-Band Noise	Noise	static	1
00:01:00	Flower Garden	Garden	dynamic	2
00:03:00	Tree	Tree	static	1
00:04:00	Mobile & Calendar	Mobile	dynamic	2
00:06:00	Clown	Clown	static	1
00:07:00	Table Tennis	Tennis	dynamic	2
00:09:00	Balls of Wool	Balls	dynamic	1
00:10:00	Diva with Noise	Diva	dynamic	2
00:12:00	Boy with Toys	Boy	static	1
00:13:00	Kiel Harbour-4	Kiel	dynamic	2
00:15:00	Young Couple	Couple	static	1
00:16:00	Circular Zone-Plate	Zone Plate	static	1

Table 4 - Test material for objective evaluation

3.2. Spatial segmentation

There were developed three algorithms for image segmentation [7]. The first is an image segmentation algorithm based on edge detection using recursive filtering (Section 3.2.1), the second is a fuzzy image segmentation algorithm based on spatial features (Section 3.2.2) and the third is an image segmentation algorithm based on watershed (Section 3.2.3). The results of the objective evaluation using these algorithms are discussed in Section 4.3. 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:

3.2.1. 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 [7]. 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 & 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.

Figure 2 - Part of Mobile & Calendar

Figure 3 - Result of segmentation

3.2.2. 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 [8] applied to this function defines the membership function of the edge regions. The complement of the fuzzy union [9] 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.

3.2.3. 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 [10]. 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.

3.3. Objective parameters

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 and the underlying process to compute them are described as follows:

- MSE (Mean Square Error)
- PSD (Positive Sobel Difference)
- NSD (Negative Sobel Difference)
- ASD (Absolute Sobel Difference)

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 R (plane, edge or texture) is defined by the mean value of

$$SE(i,j) = [\ X(i,j) - Z(i,j)\ ]^2, \text{ where } (i,j) \in R.$$

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

$$PS(i,j) = \max [\ \text{sobel}(X_m(i,j)) - \text{sobel}(Z_m(i,j))\ ,\ 0\ ], \text{ where } (i,j) \in R.$$

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

$$NS(i,j) = -\max [\ \text{sobel}(Z_m(i,j)) - \text{sobel}(X_m(i,j))\ ,\ 0\ ], \text{ where } (i,j) \in R.$$

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

$$AS(i,j) = |\ \text{sobel}(X_m(i,j)) - \text{sobel}(Z_m(i,j))\ |, \text{ where } (i,j) \in R.$$

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 4 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 [1,2]. 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. Section 4.1 describes a perceptual model to estimate the subjective impairment level based on a single parameter. The results of this approximation for each objective parameter are linearly combined to estimate the final subjective impairment level. This linear prediction model is presented in Section 4.2. Section 4.3 presents and discusses the results of this study.

4.1. Subjective quality estimation based on a single parameter: logistic approximation

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 [5] as:

$$d = (U_{max} - U) / (U_{max} - U_{min}) \times 100\%.$$

The relationship between d and each objective parameter D is approximated by the following logistic function [5]:

$$\underline{d} = 1 / (1 + (D_M / D)^G) \times 100\%,$$

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

$$e = E [\{ d - \underline{d} \}^2]$$

for each scene and each objective parameter separately. The statistical reliability of $\underline{d}$ is defined as 1/e.

4.2. Subjective quality estimation: linear prediction in 3 steps

The estimation of the normalized impairment level d by a set of estimated impairment levels $\underline{d}$ (one per parameter as defined in Section 4.1) 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:

- $\underline{d}^{MSE}$
- $\underline{d}^{PSD}$ and $\underline{d}^{NSD}$
- $\underline{d}^{ASD}$
- $\underline{d}^{MSE}$, $\underline{d}^{PSD}$ and $\underline{d}^{NSD}$
- $\underline{d}^{MSE}$ and $\underline{d}^{ASD}$

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 combine the impairment levels of each selected set and outputs three estimation values (one per context) denoted by: $\underline{d}_{YP}$, $\underline{d}_{YE}$ and $\underline{d}_{YT}$.

Similarly, the considered sets of estimated impairment levels for the chrominance components of the scene are:

- $\underline{d}^{MSE(Cb)}$ and $\underline{d}^{MSE(Cr)}$
- $\underline{d}^{ASD(Cb)}$ and $\underline{d}^{ASD(Cr)}$

and the three resulting estimation values (one per context) are denoted by: $\underline{d}_{CP}$, $\underline{d}_{CE}$ e $\underline{d}_{CT}$.

Step 2

The estimation values $\underline{d}_P$, $\underline{d}_E$ and $\underline{d}_T$ result from linear prediction based on the vectors ($\underline{d}_{YP}$, $\underline{d}_{CP}$), ($\underline{d}_{YE}$,$\underline{d}_{CE}$) and ($\underline{d}_{YT}$,$\underline{d}_{CT}$), respectively.

Step 3

The estimation values $\underline{d}_P$, $\underline{d}_E$ and $\underline{d}_T$ are combined by linear prediction to produce the estimated impairment level $\underline{d}$.

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

Let ($\underline{d}_1$, $\underline{d}_2$, ... , $\underline{d}_P$) be the input vector of the linear predictor. The output $\underline{d}_o$ is given by:

$$\underline{d}_o = \sum a_i \underline{d}_i,$$

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

$$E [\{ d - \underline{d}_o \}^2], \text{ such that}$$

$$\sum a_i = 1 \text{ and}$$

$$a_i / a_k = e_k / e_i,$$

where the statistical reliability of $\underline{d}_i$ is $1/e_i$, as defined in Section 4.1.

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 4.3.

4.3. Subjective quality estimation: presentation and discussion of results

This section is divided into three main topics. The results and prediction models obtained by the subjective quality estimation based on the Algorithm I (the image segmentation algorithm previously described in Section 3.2.1) are described in Section 4.3.1. Section 4.3.1 also presents 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). The variation of the estimation accuracy with the image segmentation algorithms is discussed in Section 4.3.2. Section 4.3.3 points out the advantages of the purposed subjective estimation method compared to other methods that are based on global measurements or optimal prediction.

4.3.1. Results: perceptual models and performance

Table 5 presents the results of the subjective estimation method based on the algorithm I (Section 3.2.1) for segmenting the following scenes: Garden, Mobile, Tennis, Diva and Kiel, separately. In Table5:

- The weights of the linear prediction described in Step 2 of Section 4.2 are equivalent to the relative subjective weights of luminance (Y) and chrominance (C) impairments in plane regions, edge regions and texture regions. The global mean value computed over all scenes is given at the last line of this table.
- The weights of the linear prediction described in Step 3 of Section 4.2 are equivalent to the relative subjective weights of the degradation in plane regions (P), edge regions (E) and texture regions (T). The global mean value computed over all scenes is given at the last line of this table.
- The mean square error (MSE) and the mean absolute error (MAE) between the normalized impairment level d and the estimated impairment level $\underline{d}$, taking into account a normalization scale from 0% to 100%, are shown at the two last columns of 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 of these columns.

The results presented in Table 5 refer to the perceptual models obtained from the subjective scores of the 34 non-expert assessors of the Table 1 and from the 26 assessed systems of the Table 3.

	Step 2: Plane		**Step 2: Edge**		**Step2: Texture**		**Step 3**			**Error**	
Scene	**Y(%)**	**C(%)**	**Y(%)**	**C(%)**	**Y(%)**	**C(%)**	**P(%)**	**B(%)**	**T(%)**	**MSE**	**MAE**
Garden	61	39	70	30	37	63	13	37	51	18.1	3.0
Mobile	74	26	75	25	63	37	83	7	9	24.2	3.6
Tennis	67	33	65	35	70	30	45	13	42	25.3	3.5
Diva	49	51	92	8	42	58	27	59	14	5.4	1.5
Kiel	62	38	66	34	40	60	32	39	29	22.7	3.6
Global	**63**	**37**	**73**	**27**	**50**	**50**	**40**	**31**	**29**	**6.2**	**1.8**

Table 5 - Perceptual models and results: non-expert assessors

Tables 6 and 7 show the dependence between perceptual models and results for:

- non-expert and expert assessors;
- viewing distance (4H and 6H) from the monitor (each case with 50% of the total number of assessors).

	Non-expert assessors						Expert assessors					
	Region			Component		Error	Region			Component		Error
Scene	**P(%)**	**B(%)**	**T(%)**	**Y(%)**	**C(%)**	**MSE**	**P(%)**	**B(%)**	**T(%)**	**Y(%)**	**C(%)**	**MSE**
Garden	13	37	51	52	48	18.1	12	53	34	51	49	23.5
Mobile	83	7	9	73	27	24.2	72	13	15	72	28	73.4
Tennis	45	13	42	68	32	25.3	47	12	41	70	30	48.1
Diva	27	59	14	73	27	5.4	22	42	36	55	45	21.2
Kiel	32	39	29	57	43	22.7	43	36	21	47	53	44.1
Global	**40**	**31**	**29**	**62**	**38**	**6.2**	**39**	**31**	**30**	**58**	**42**	**12.1**

Table 6 - Perceptual models and results: non-expert and expert assessors

	6H viewing distance						4H viewing distance					
	Region			Component		Error	Region			Component		Error
Scene	**P(%)**	**B(%)**	**T(%)**	**Y(%)**	**C(%)**	**MSE**	**P(%)**	**B(%)**	**T(%)**	**Y(%)**	**C(%)**	**MSE**
Garden	15	40	45	51	49	20.0	9	35	55	47	53	16.8
Mobile	83	8	9	77	23	24.7	71	13	16	62	38	59.4
Tennis	47	15	38	60	40	40.9	42	10	48	74	26	21.4
Diva	41	36	22	64	36	15.5	21	46	33	54	46	12.5
Kiel	34	40	26	54	46	18.3	31	46	23	57	43	26.0
Global	**44**	**28**	**28**	**61**	**39**	**7.9**	**35**	**30**	**35**	**59**	**41**	**9.6**

Table 7 - Perceptual models and results: 6H and 4H viewing distances

The results presented in Tables 5, 6 and 7 are commented below:

- The subjective quality estimation using objective parameters based on image segmentation, computed over the 26 systems described in Section 2.3, resulted a mean absolute error (MAE) less than 4% for each individual scene and a global MAE of 1.8% , considering 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 in the models based on expert assessors.
- Comparing the perceptual models based on 4H and 6H viewing distances, the weight of the impairments in edge and texture regions is significantly greater in the models based on 4H viewing distance, as expected.

4.3.2. The variation of the estimation accuracy with the image segmentation algorithm

The results of the subjective quality estimation, based on the Algorithms II and III (shortly described in Section 3.2) and obtained from the scores of the 34 non-expert assessors, are shown in Table 8. Comparing the results of this table with the results previously presented at the left side of Table 6 (for the Algorithm I), the estimation accuracy presented small variations for a 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 algorithms may provide satisfactory results.

	Algorithm II						Algorithm III					
	Region			Component		Error	Region			Component		Error
Scene	**P(%)**	**B(%)**	**T(%)**	**Y(%)**	**C(%)**	**MSE**	**P(%)**	**B(%)**	**T(%)**	**Y(%)**	**C(%)**	**MSE**
Garden	9	32	59	53	47	18.0	10	44	46	53	47	15.8
Mobile	65	26	9	59	41	20.7	82	11	6	60	40	18.7
Tennis	54	27	19	70	30	28.5	68	21	11	72	28	31.3
Diva	25	50	24	75	25	7.1	50	31	19	60	40	7.4
Kiel	23	31	46	64	36	25.9	28	33	38	59	41	22.4
Global	**35**	**33**	**31**	**66**	**34**	**7.4**	**48**	**28**	**24**	**63**	**37**	**6.5**

Table 8 - Perceptual models and results: algorithms II and III

4.3.3. The advantages of the adopted subjective quality estimation method

The example illustrated in Table 9 focuses two very important properties of any method for subjective quality estimation based on objective parameters: accuracy and robustness [11-14]. This example compares the adopted subjective quality estimation method, which relies on context-based objective measurements and the linear prediction method described in Section 4.2, with the following methods:

- a method that relies on the same context-based objective measurements, but uses optimal prediction;
- a method that uses the linear prediction described in Section 4.2, 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 3.3. The results were obtained from the scores of the non-expert assessors. The input values of the table are mean square prediction errors. The last line of the table shows the mean value of this parameter computed over the set of scenes.

	Adopted method		Optimum predictor		Global measurements	
Scene	**Training**	**Test**	**Training**	**Test**	**Training**	**Test**
Garden	3.9	87.6	2.8	71.8	3.9	62.3
Mobile	30.1	48.6	10.5	82.1	179.1	162.5
Tennis	10.8	91.3	7.7	335.0	108.9	221.2
Diva	1.4	8.9	0.8	17.7	1.8	34.3
Kiel	22.4	9.3	20.5	13.4	30.6	27.7
Mean	**13.7**	**49.1**	**8.5**	**104.0**	**64.9**	**101.6**

Table 9 - Comparison: robustness and accuracy

The advantage of computing objective parameters based on context becomes clear when the procedure described in Sections 4.1 and 4.2 is also applied to global measurements. Note that the use of context-based measurements can significantly improve the estimation results in all scenes (with exception of Flower Garden). Possibly, this indicates that the image segmentation process for Flower Garden needs to be refined.

The example also shows that the prediction process described in Section 4.2 is more robust (i.e. it is less dependent of the training data base) when it is compared to the optimum predictor, improving the prediction results on the test data base.

5. Conclusions

This contribution 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 in this contribution 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 4.2. 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);
- 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.

Therefore, we suggest the inclusion of the linear prediction method presented in this contribution and of context-based objective measurements in new ITU Recommendations, which are related to objective evaluation of video quality.

References

[1] ISO/IEC 11172, “Information technology - Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s”, 1992.

[2] ITU-T Recommendation H.262, “Generic Coding of Moving Pictures and Associated Audio - Information: Video”, 1995.

[3] ITU-R Recommendation BT.601-5, “Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios”, 1995.

[4] ANSI T1.801.03-1996, “Telecommunications - Digital transport of one-way video signals - Parameters for objective performance assessment”, 1996.

[5] ITU-R Recommendation BT.500-7, “Methodology for the subjective assessment of the quality of television pictures”, 1995.

[6] ITU-R Recommendation BT.802-1, “Test pictures and sequences for subjective assessment of digital codecs conveying signals produced according to Recommendation 601”, 1994.

[7] Gonzalez and P. Wintz, “Digital Image Processing”, Addison Wesley, 1987.

[8] Dougherty, “An Introduction to Morphological Image Processing”, SPIE Optical Engineering Press, Bellingham, WA, vol. TT9, 1992.

[9] A. Kaufmann, “Introduction to The Theory of Fuzzy Subsets”, Academic Press, New York, NY, vol. 1, 1975.

[10] Barrera J., Banon J. F., and Lotufo R.A., “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, 24-29, July, 1994.

[11] ITU-T Contribution COM 12-66, “Selections from the draft American National Standard: - Digital transport of one-way signals - parameters for objective performance assessment”, USA.

[12] ITU-T Delayed Contribution D.021, “Objective and subjective measures of MPEG video quality: summary of experimental results”, USA.

[13] ITU-T Delayed Contribution D.101, “A Two-Stage Objective Model for Video Quality Evaluation”, Bellcore.

[14] ANSI T1A1 Contribution Number T1A1.5/96-121, “Objective and subjective measures of MPEG video quality”, GTE Labs., NTIA/ITS, October, 1996.