

Draft 1

Methods for the Production of Graded Impairment Video Test Materials for JND Measurement of Digital Video Quality

Author: John M. Libert
Submitted: 26 October, 1998

1 Purpose

Create a suite of video sequences exhibiting graded levels of compression artifacts (impairments) to support development of a Just-Noticeable-Difference (JND) scale of video impairment.

2 Objectives

- 2.1 **Develop or acquire a methodology by which to induce or introduce graded levels of MPEG-2 type video impairments into reference digital video sequences – Video impairments of interest include blocking, blurring, mosquito noise, color errors, and block errors.**
- 2.2 **Validate synthetic video impairments against actual MPEG-2 video impairments of each class with respect to perceptibility by human viewers and at least one objective measurement scheme.**
- 2.3 **Select public domain, i.e. unrestricted use, royalty-free, video sequences or create synthetic sequences as bases for the reference video calibration suite.**
- 2.4 **Apply the impairment-induction methodology to the reference video to create a set of graded test sequences suitable to measuring the threshold of human perception of just-noticeable-differences of impaired video against the reference sequence.**
- 2.5 **Identify subjective testing laboratories willing conduct JND measurements according to design developed as a separate task.**
- 2.6 **Disseminate standard reference video materials on suitable medium with appropriate calibration certification.**

3 Impaired Video Sequences: Requirements

3.1 Definitions¹

3.1.1 block: Group of pixels (see below). For example, a block of 8×8 pixels is the smallest coding block used in MPEG algorithms.

3.1.2 block distortion (or tiling): Distortion of the image characterized by the appearance of an underlying block encoding structure wherein entire or partial block boundaries are delineated by horizontal and vertical intensity gradients.

3.1.3

¹ Approximately as given in ITU-T Rec. P.930

- blurring:** Distortion of the entire image, characterized by reduced sharpness of edges and loss of spatial detail.
- 3.1.4 color errors:** Distortion of all or a portion of the final image characterized by the appearance of unnatural or unexpected hues or saturation levels. These hues or saturation levels were not present in the original image.
- 3.1.5 edge busyness:** Distortion concentrated at the edges of objects, and further characterized by its temporal and spatial characteristics.
- 3.1.6 error blocks:** A form of block distortion where one or more blocks in the image bear no resemblance to the current or previous scene and often contrast greatly with adjacent blocks.
- 3.1.7 jerkiness (or jerky motion) :** Motion that was originally smooth and continuous perceived as a series of distinct "snapshots".
- 3.1.8 mosquito noise:** Form of edge busyness distortion sometimes associated with movement, characterized by moving artifacts and/or blotchy noise patterns superimposed over the objects (resembling a mosquito flying around a person's head and shoulders).
- 3.1.9 pixel (or pel):** A picture element that describes the brightness or color of a discrete point in an image.
- 3.1.10 quantization noise:** A "snow" or "salt and pepper" effect similar to random noise, but not uniform over the image

3.2 Motivation

Conventional practice in the video compression development and testing community has required only rather coarse control over the impairment type and the level of distortion of video sequences used for testing. By selecting source material appropriately and by adjusting encoding parameters, it is possible to stress video codecs such that various impairments are generated to various degrees in the decoded output sequence. While these methods have been adequate to generate a variety of impairment types over a range of levels, they permit only coarse control over the actual composition and degree of video impairment.

Visual threshold measurements, e.g., where one seeks to measure the "just noticeable difference" between two stimuli, requires relatively finely graded values of the stimuli of interest, at least in the neighborhood of the threshold. Thus, in a study of JND's of video impairments due to compression, one would like to be able to generate a continuous variation in the distortion level. Moreover, in the case of video impairments, it may also be the case that different types of distortions vary with respect to their visibility and in their impact on subjective judgements of quality. Accordingly, one might also like to control the type of

impairment, at least such that an image sequence contains mostly a single type of distortion even where several are present.

It may be found, ultimately, that fine control over subjectively and objectively realistic video impairments is not possible or at least impractically difficult – hence the widespread use of non-threshold types of video quality measurement. Yet, the objective of the present task force is to identify the methodology to produce test materials suitable for JND measurements of video quality. Such a methodology should be able to produce test materials that are:

- 3.2.1 graded finely over a region including the JND.
- 3.2.2 relatively homogeneous with respect to impairment type (Interaction effects of combinations of impairments might be examined later.); and
- 3.2.3 close in visual appearance to those produced via normal MPEG-2 compression and are similar with respect to both subjective and objective estimation of magnitude.

4 Generation of Impaired Video Sequences: Several Alternative Methods

The following list does not presume to be exhaustive of possible approaches for generating test materials as described above. Moreover, any of the schemes suggested would need to be validated.

4.1 Direct Induction via MPEG-2 Codec

Select video sequences and adjust parameters such as target bit rate, motion estimation search region size, etc. so as to produce different types and degrees of impairment. It is not considered by the author to be likely that this method will provide the degree of control required for threshold measurements. But the method is included for completeness.

4.2 Synthetic Impairment Injection

The ITU-T Recommendation P.930 describes a Reference Impairment System for Video (RISV). In many respects, the RISV resembles the sort of product we seek in the present effort. The appendix of the P.930 details one implementation of the recommendation, i.e., a software system named VIRIS (Video Reference Impairment System) developed by Bellcore. Other implementations may have been developed, such as by KPN.

Applied to a source video sequence, the VIRIS was able to simulate five types of impairment including block distortion, blurring, edge busyness, jerkiness, and noise, as defined above. The Appendix to P.930 describes, in some detail, the means by which these impairments were simulated by VIRIS. Should the original software be lost to the "netherland of abandoned projects," it may be possible to reproduce the essential functionality of the VIRIS software with reference to the P.930 document. Alternately, KPN or other ITU members might have implemented similar simulations that we could acquire.

This approach would have to be validated as to the degree to which the synthetic impairments correlate with actual impairments, with respect to both subjective and objective measurements.

4.3 Merging of Impaired and Reference Video

For JND measurements suitable graded test material might be generated by either global or local mixing of reference video with corresponding frames of impaired video. In this approach, the video impairment can be graded as finely as needed with respect to its perceptibility by forming the weighted average of each reference video frame with its corresponding impaired frame. Thus, for a given impairment level, a weight, w_R , of the reference is set to a value between 0 and 1, and the weight of the impaired video, w_I , set to $1-w_R$. At one extreme, the sequence would match the reference. At the other extreme, the impairments, of course, would match those of the processed sequence. The progression of impairments thus produced could be as fine as desired, within the limits of the bit depth of display and processing systems.

This method appears to work reasonably well in producing a gradation of blocking impairments between the two "end members." Of course, while the visibility of the impairments gradually increases via this method, the spatial distribution of the impairments remains the same though the entire range.

5 Validation of Impairment Induction Method

Whichever method is chosen to produce the JND test materials, some level of validation of the methodology will be needed. Ideally, the artificially induced impairments at various levels will be similar to actual impairments both visually and with respect to some objective measurement value. While it might be reassuring if artificial impairments were visually indistinguishable from the natural distortions, it is more important that they have similar visibility and would be judged to have similar impact on picture quality.

5.1 Subjective Validation

One might test for similarity via a series of two alternative forced choice presentations of artificially and naturally impaired sequences. Here the point at which preference is divided evenly between the alternatives would become the point of subjective equivalence. If all is well, an objective quality model or distortion measure should yield similar values for the sequences found to be subjectively similar.

5.2 Objective Validation

Some objective measure of distortion will be used to quantify the quality scale. While one measure should be used in developing the graded test sequences, it would be useful to apply several different models to the sequences. As long as each responded to the graded distortion in some reasonably well behaved fashion, each might also be calibrated with respect to subjective JND's. Candidates would include a DCT error scheme similar to that described by Watson, Peterson, and Ahumada [2-4], and PSNR. Nominations will be accepted for candidate computational vision models.

6 Source Video Sequences

Whatever else their characteristics, it should be important to select source imagery that can be freely distributed without royalty or other restriction.² Any sequences that can be secured by G-2.1.6 for this exercise should be considered for inclusion. Members are encouraged to make recommendations.

7 Test Sequence Building

NIST would be willing to work on construction of the data production. Depending upon the anticipated size of the job, other laboratories will be enlisted.

8 Subjective Measurement

Proposal of the G-2.1.6 effort to subjective labs should proceed as soon as design of the experiment nears completion.

9 Dissemination of Reference Materials

Mechanisms for this and media are to be determined.

10 References

1. International Telecommunications Union. (1996). Principals of a Reference Impairment System for Video, ITU-T Recommendation P.930, SERIES P: TELEPHONE TRANSMISSION QUALITY Audiovisual quality in multimedia services.
2. A. J. Ahumada, Jr. and H. A. Peterson. "A visual detection model for DCT coefficient quantization." *Computing in Aerospace* 9, 314-318, 1983.
3. H. A. Peterson, A. J. Ahumada, Jr., and A. B. Watson. "An Improved Detection Model for DCT Coefficient Quantization." *Human Vision, Visual Processing, and Display IV*. Rogowitz and Allebach, eds., 1993 SPIE, Bellingham, WA.
4. A. B. Watson. "DCT matrices visually optimized for individual images." *Human Vision, Visual Processing, and Display IV*. Rogowitz and Allebach, eds., 1993 SPIE, Bellingham, WA.
- 5.

² Though details remain in the planning, NIST plans to assemble or build and distribute such a suite of test materials during the 1999 fiscal year. According the developing plan. NIST will provide sequences with impairment measurements at several levels and other calibration information and release these materials according to its standard practice for Standard Reference Materials (SRM's).