COMMITTEE T1

CONTRIBUTION

Document Number: T1A1.5/93-153

****

STANDARDS PROJECT: VTC/VT Performance Standards Projects

****

TITLE: Methods of Measurement for Two Objective Video Quality Parameters Based on the Fourier Transform

****

ISSUE ADDRESSED: Objective Video Performance Testing

****

SOURCE: NTIA/ITS, Arthur Webster

****

DATE: November 8, 1993

****

DISTRIBUTION TO: T1A1.5

****

KEYWORDS: Objective Video Performance Testing, Video Quality Parameters

****

Methods of Measurement for Two Objective Video Quality Parameters Based on the Fourier Transform

1 Introduction

This document provides the method of measurement for two video quality distortion measurements which are based upon the spatial frequency content of the source and the destination video. The distortion measurements are calculated from vector information extracted from the source and corresponding destination video. The measurements presented here, $PD(t_n)$ and $ND(t_n)$, are improved versions of the measures $SD3$ and $SD4$ which were introduced in contribution T1Q1.5/92-113, January 22, 1992. Objective video quality parameters P12 and P13 are obtained from the measurements $PD(t_n)$ and $ND(t_n)$. The scalar parameters P12 and P13 extend the ITS objective video quality parameter list given in T1A1.5/93-152.

The measurements $PD(t_n)$ and $ND(t_n)$ are obtained from the two dimensional FFT (Fast Fourier Transform) of subregions of each video frame. These measurements detect spatial impairments in the destination video. $PD(t_n)$ measures the positive distortion (loss of edge energy, as in blurring) and $ND(t_n)$ measures negative distortion (the addition of spurious edge energy, such as added noise, blocking, or other digital artifacts). These measurements are made on the video sequence 5 times per second (i.e., on frames 1, 7, 13, 19, 25, etc.). The temporal subsampling is performed to reduce overall computation time and data storage requirements.

ITS is currently in the process of performing these objective measurements on the hypothetical reference circuit (HRC) video (see contribution T1A1.5/93-014R1) and will be reporting the results to T1A1.5. Questions regarding the methods of measurement in this contribution should be directed to:

Arthur Webster
U.S. Department of Commerce
NTIA/ITS.N3
325 Broadway
Boulder, CO 80303

Phone: (303) 497-3567
Fax: (303) 497-5323
E-mail: webster@its.bldrdoc.gov

2 Methods of Measurement

Given a time series of luminance images (or frames) digitized as described in T1A1.5-93/152, let

$Y_S(t_n)$ = the image (or frame) sampled at time t_n of the *source* video sequence, and

$Y_D(t_m)$ = the image (or frame) sampled at time t_m of the *destination* video sequence.

Assume that the source and destination video sequences have been time-aligned such that $t_m = t_n + d_v$, where d_v is the video delay. That is, each destination image, $Y_D(t_m)$, corresponds to the source video image, $Y_S(t_n)$. Methods for time-alignment are discussed in T1A1.5/93-152.

$Y_S(i,j,t_n)$ and $Y_D(i,j,t_m)$ specify the pixels in the i^{th} row (video line) and j^{th} column (samples along a video line) of the n^{th} and m^{th} images respectively.

The *viewable region* of $Y_S(t_n)$ and $Y_D(t_m)$ (see Figure 3 of T1A1.5/93-152 for a definition of *viewable region*) is divided into 6 overlapping square subregions, each of dimensions 256 x 256 pixels (see Figure 1).

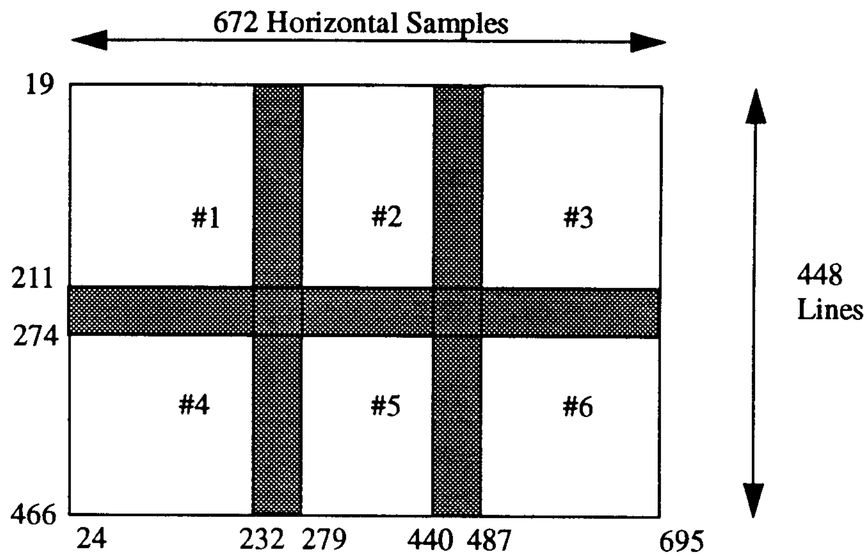

Figure 1 Subregion Boundaries

There are 486 (0 through 485) lines in the digitized image and 720 (0 through 719) samples (pixels) per line. The overlapping processing subregions are defined by the following (inclusive) coordinates:

Subregion #1: rows 19 through 274 and columns 24 through 279

Subregion #2: rows 19 through 274 and columns 232 through 487

Subregion #3: rows 19 through 274 and columns 440 through 695

Subregion #4: rows 211 through 466 and columns 24 through 279

Subregion #5: rows 211 through 466 and columns 232 through 487

Subregion #6: rows 211 through 466 and columns 440 through 695

Let $Y_{Sk}(t_n)$ denote the k^{th} subregion of the source image at time t_n.

Let $Y_{Dk}(t_m)$ denote the k^{th} subregion of the destination image at time t_m.

$F_{Sk}(t_n) = \mathcal{F}\{Y_{Sk}(t_n)\}$ and $F_{Dk}(t_m) = \mathcal{F}\{Y_{Dk}(t_m)\}$,

where $\mathcal{F}$ denotes the two dimensional Fourier transform operator. The resulting transform images (each of size 256 x 256 complex-valued pixels) have been quadrant shifted so that the DC (or zero spatial frequency) term is at the center of the image (i.e. the pixel with address [129, 129]). This is accomplished by exchanging the 1st and 3rd quadrants and the 2nd and 4th quadrants.

The magnitude of the two dimensional Fourier spectrum is averaged radially to produce the basic feature, $R_{Sk}(f,t_n)$, from which the distortion measurements are obtained. This radial averaging produces one real-valued vector, $R_{Sk}(f,t_n)$, for each subregion of each processed frame. Each element of this vector is the average of the pixel values which lie within a circular band of inner radius f_{a-1} and outer radius f_a $(0 \le f_a \le 127)$ as measured from the center of the array. See Figure 2 for a diagram of the radial averaging process.

We define $R_{Sk}(f,t_n)$ to be the radial average of the Fourier magnitude, $|F_{Sk}(i,j,t_n)|$, calculated as

$$R_{Sk}(f_a, t_n) = \sum_{f > f_{a-1}}^{f \le f_a} \frac{|F_{Sk}(i,j,t_n)|}{m_f}, \text{ where } f = \left[i^2 + j^2\right]^{1/2} \tag{1}$$

and m_f is the number of pixels in the annular region between radii = f_{a-1} and f_a (see Figure 2). The variable 'f_a' is an integer and is used to index the spatial frequency bins 0 through 127. Bin 0 is the DC component and bin 127 is about 25 cycles per degree for the subjective viewing configuration given in the subjective test plan, T1A1.5/93-014R5. The other variables range as follows: i=0,255; j=0,255; t_n=1,7,13,...265, 271 (for a nine second scene); and k= 1,6.

The Fourier magnitude is defined as $|F| = (F F^*)^{1/2}$ where * denotes complex conjugate.

Similarly, $R_{Dk}(f,t_m)$, is calculated from the destination video sequence.

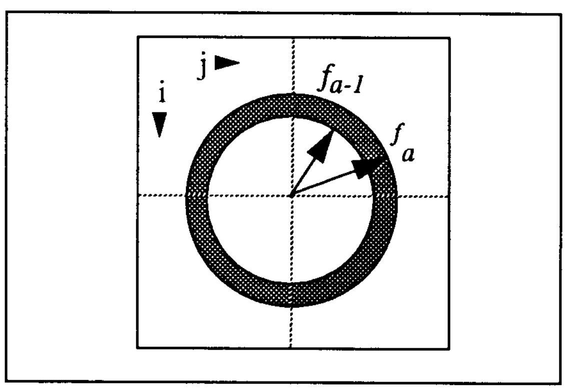

Figure 2 Radial Averaging Over the Fourier Magnitude

These radial averages of the source and destination video form the basis for the distortion measurements given below.

The positive distortion measurement, $PD(t_n)$, is given by

$$PD(t_n) = \sum_{k=1}^{6} \sum_{f=6}^{80} \left(\frac{R_{Sk}(f, t_n) - R_{Dk}(f, t_n + d_v)}{R_{Sk}(f, t_n)} \right) \Bigg|_{\text{positive values only}} \quad (2)$$

and the negative distortion measurement, $ND(t_n)$, is given by

$$ND(t_n) = \sum_{k=1}^{6} \sum_{f=6}^{80} \left(\frac{R_{Sk}(f, t_n) - R_{Dk}(f, t_n + d_v)}{R_{Sk}(f, t_n)} \right) \Bigg|_{\text{negative values only}} \quad (3)$$

To obtain a *single value* for the entire video scene (a video scene is nine seconds for the T1A1.5 VTC/VT test), the maximum absolute value over time is used. The positive distortion parameter, P12, is given by

$$P12 = MAX_{time}\left[\left| PD(t_n) \right| \right] \quad (4)$$

The negative distortion parameter, P13, is given by

$$P13 = MAX_{time}\left[\left| ND(t_n) \right| \right] \quad (5)$$

3 Conclusion

The two spatial distortion parameters presented in this contribution, P12 and P13, will be evaluated at NTIA/ITS for all 625 HRC - scene combinations. The results will be presented to T1A1.5 for use in further data analysis.