COMMITTEE T1
CONTRIBUTION

Document Number T1A1.5/92-134

****

STANDARDS PROJECT: Analog Interface Performance Specifications for Digital Video Teleconferencing/Video Telephony Service

****

TITLE: Spatial and Temporal Information Measures - Test Scene Evaluation II

****

ISSUE ADDRESSED: Video Quality

****

SOURCE: NTIA/ITS - Arthur Webster

****

DATE: July 13, 1992

****

DISTRIBUTION TO: T1A1.5

****

KEYWORDS: Video Teleconferencing, Video Telephony, Video Quality, Subjective Quality, Objective Quality

****

DISCLAIMER:

****

Spatial and Temporal Information Measures - Test Scene Evaluation II

1. Introduction

This contribution presents the results of VTC/VT test scene evaluation using the methods that are documented in References [1] and [2]. Results are given for test scene groups designated as: **ITS** (Institute of Telecommunication Sciences), **CBC** (Canadian Broadcasting Corporation), and **DIS** (Delta Information Systems). These test scenes are those which the T1A1.5 VTC/VT sub-working group chose to be considered for future tests. This contribution satisfies the action item accepted by ITS at the last T1A1.5 meeting.

We recommend that this contribution be inserted into Appendix A of the Draft Standard [4] in the section headed "Test Scene Selection." The method of evaluating test scenes presented here is applicable to video quality testing -- both now and in the future.

First, the definitions of the measures for spatial and temporal information will be repeated; then the results will be presented in plots and tables.

2. Spatial Information

The spatial information measure (SI) used is based on the Sobel filter [3]. Each video frame (luminance plane) is filtered with the Sobel operators. The standard deviation over the pixels in each Sobel-filtered frame is then computed. This operation is repeated for each frame in a video sequence and results in a time series of spatial information values. The maximum value in the time series is chosen to represent the spatial information content of the scene.

$$SI = MAX_{time}\ of\ STD_{space}\ of\ Sobel(VideoSequence)$$

3. Temporal Information

This class of temporal measures is based upon the motion difference feature, $M_n(i,j)$, which is the difference between the pixel values (of the luminance plane) at the same location in space but at successive times or frames. $M_n(i,j)$ as a function of time (n) is defined as,

$$M_n(i,j) = F_n(i,j) - F_{n-1}(i,j)$$

where $F_n(i,j)$ is the pixel at the ith row and jth column of the nth frame in time.

The measure of temporal information content, TI, is computed as the maximum over n (time) of the standard deviation of $M_n(i,j)$ over all i and j (space).

$$TI = MAX_{time}\ of\ STD_{space}\ [M_n(i,j)]$$

More motion in adjacent frames will result in higher values of TI. Note: For some of the scenes which contain scene cuts, two values are given: one using the scene cut as a measure of motion, and the other, using the video between scene cuts to represent the motion of the scene.

4. Results

Figures 1, 2, and 3 show how the three sets of test scenes analyzed can be placed on a spatial-temporal information matrix which has its origin at the upper left. The grid attempts to place scenes in boxes corresponding to the spatial-temporal matrix on page 18 of the current draft standard [4]. See Figure 4 for a reproduction of this matrix. Note that the scenes in Figures 1, 2, and 3 are divided into only 3 spatial levels instead of 4 as shown in Figure 4. This is due to the fact that Level 4 might be used for spatial performances beyond NTSC (such as HDTV).

The accompanying Tables 1, 2, and 3 provide data as well as naming references. Along the TI=0 axis (at the top edge of the plot) are found the still scenes and those with very limited motion. Along the SI=0 axis (at the left edge of the plot) are found scenes with minimal spatial detail. These values of SI and TI can be compared to other test scenes measured using the above equations which have been spatially sampled at 4 times sub-carrier ($4f_{sc}$) or approximately 756 x 486 pixels, 30 frames per second, and digitized with white set to 235 and black set to 16.

Note that the values are not normalized, although each plot is made to the same scale. Maxima for SI and TI are estimated to be, for our processing parameters, 406 and 219 respectively. These values will not be encountered in typical video, although special test scenes can be constructed to yield these values.

Figure 5 shows how all three sets of test scenes can be grouped into the spatial-temporal matrix. The asterisks mark the ITS scenes, the triangles mark the CBC scenes and the squares mark the DIS scenes. Presumably, for a video system to qualify at a given level, it must be tested with scenes that exercise that level of performance.

5. Conclusions

The location of the video scene within the spatial-temporal matrix is important because the quality of a transmitted video scene (especially after passing through a low bit rate codec) is often highly dependent on this location. The SI and TI measures can be used to assist the T1A1.5 VTC/VT sub-working group with scene selection for the upcoming VTC/VT tests. SI and TI can be used to assure uniform coverage of the spatial-temporal matrix, as well as assuring that each pertinent performance block in the spatial-temporal grid shown in Figure 4 is adequately stressed.

We recommend that this contribution be inserted into Appendix A of the Draft Standard [4] in the section headed "Test Scene Selection." For a video system to qualify at a given level, it must be tested with scenes that exercise that level of performance.

6. Acknowledgements

The work presented in this contribution is the result of the efforts of the Video Quality Project team (which is part of the System Performance Standards Group at the Institute for Telecommunication Sciences) consisting of Steve Wolf, Arthur Webster, Steve Voran, Coleen Jones, Margaret Pinson, and Paul King. The measures presented here complement work on other video quality measures that have been presented to the T1Q1.5 VTC/VT sub-working group by ITS and other participants.

7. References

[1] T1Q1.5/92-13, Webster, A. & Wolf, S., Spatial and Temporal Information Measures for Video Quality, January 22, 1992.

[2] T1A1.5/92-13, Webster, A., Spatial and Temporal Information Measures -- Test Scene Evaluation, March 31, 1992.

[3] Gonzalez, R. & Wintz,P., Digital Image Processing, 2nd Ed., Addison Wesley,1987.

[4] T1A1.5/92-107, Draft American National Standard, Digital Transport of Video Teleconferencing/Video Telephony Signals- System M-NTSC Analog Interface Specifications and Performance Parameters, March 4, 1992.

Figure 1. ITS Test Scenes

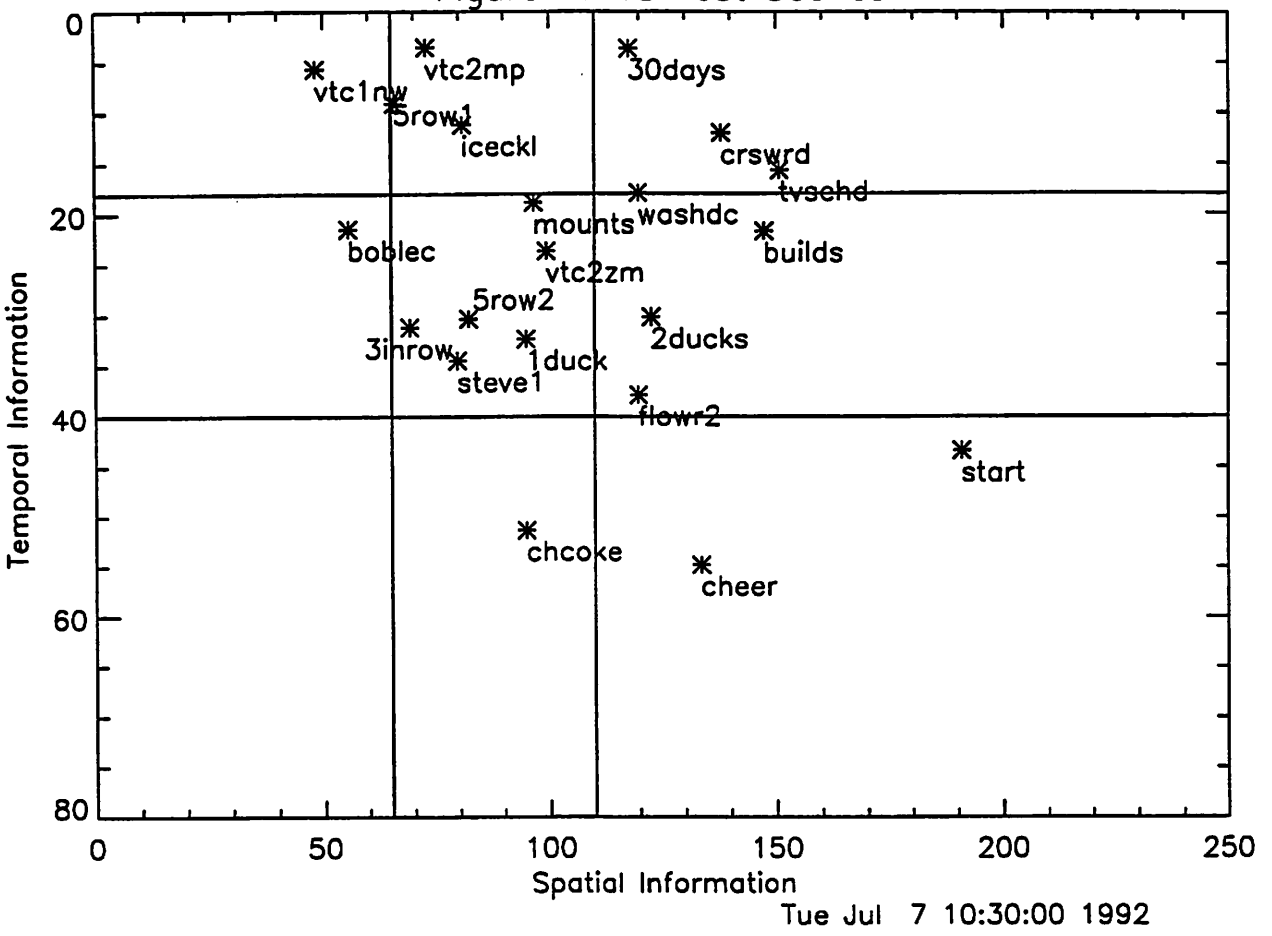

TABLE 1. ITS Test Scenes

Processing Name	Spatial Inform.	Temporal Inform.
vtc1nw	48.3	5.6
5row1	65.6	9.1
vtc2mp	72.7	3.5
iceckl	80.7	11.2
boblec	55.4	21.5
mounts	96.6	18.9
vtc2zm	99.4	23.6
3inrow	69.1	31.2
5row2	82.1	30.3
washdc	119.8	17.9
steve1	79.6	34.5
crswrd	138.0	12.0
1duck	94.9	32.3
2ducks	122.7	30.2
tvschd	150.8	15.7
builds	147.4	21.7
flowr2	119.8	37.9
chcoke	94.9	51.3
cheer	133.5	54.8
start	191.0	43.4
30days	117.5	2.7

Figure 2. CBC Test Scenes

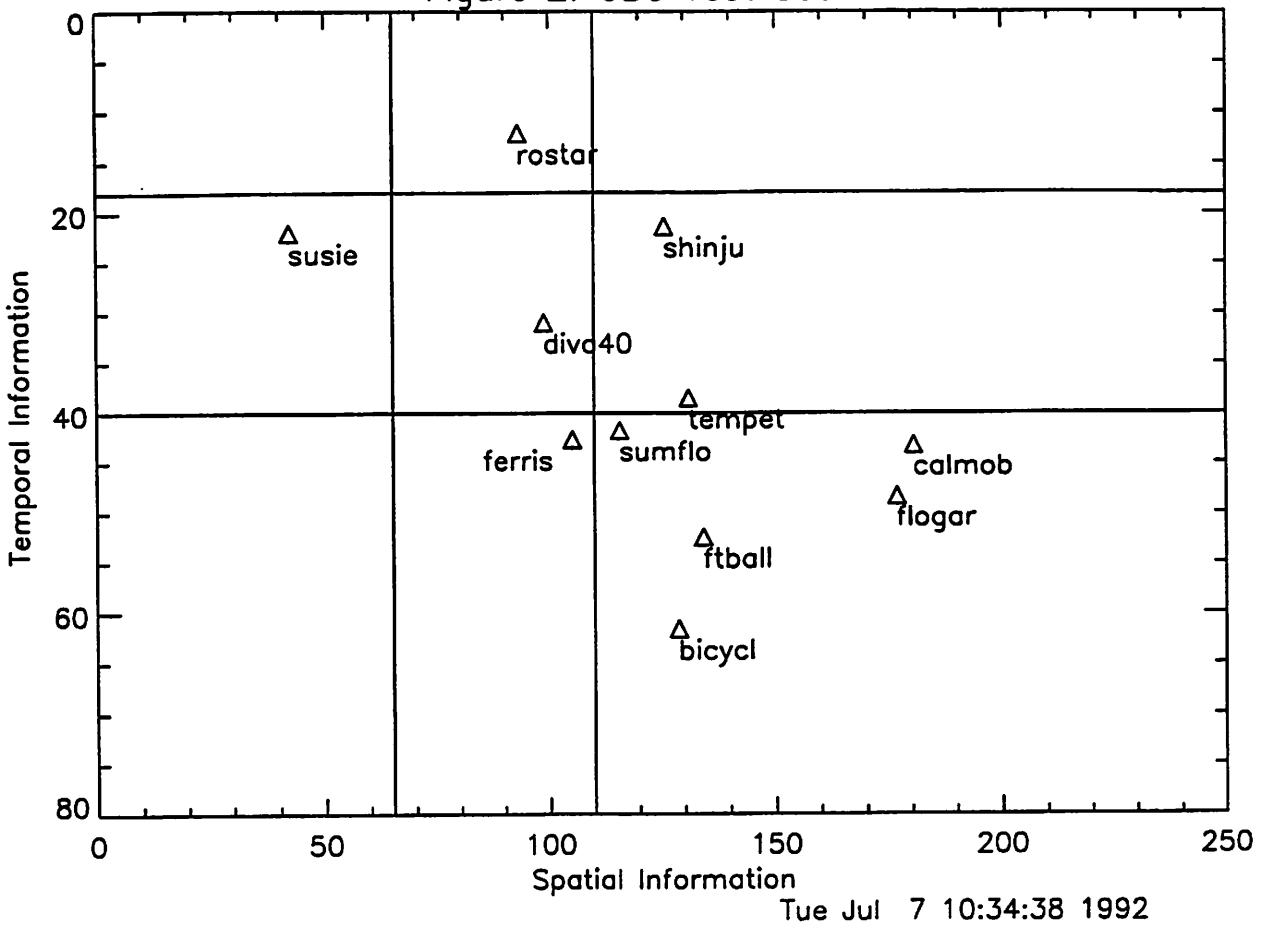

TABLE 2. CBC Test Scenes

Processing Name	Spatial Inform.	Temporal Inform.
bicycl	128.7	61.6
calmob	180.5	43.2
diva40	98.9	30.9
ferris	105.1	42.7
flogar	176.7	48.3
ftball	134.1	52.5
rostar	93.1	12.1
shinju	125.7	21.4
sumflo	115.6	41.8
susie	42.2	22.0
tempet	131.0	38.5

Figure 3. DIS Test Scenes

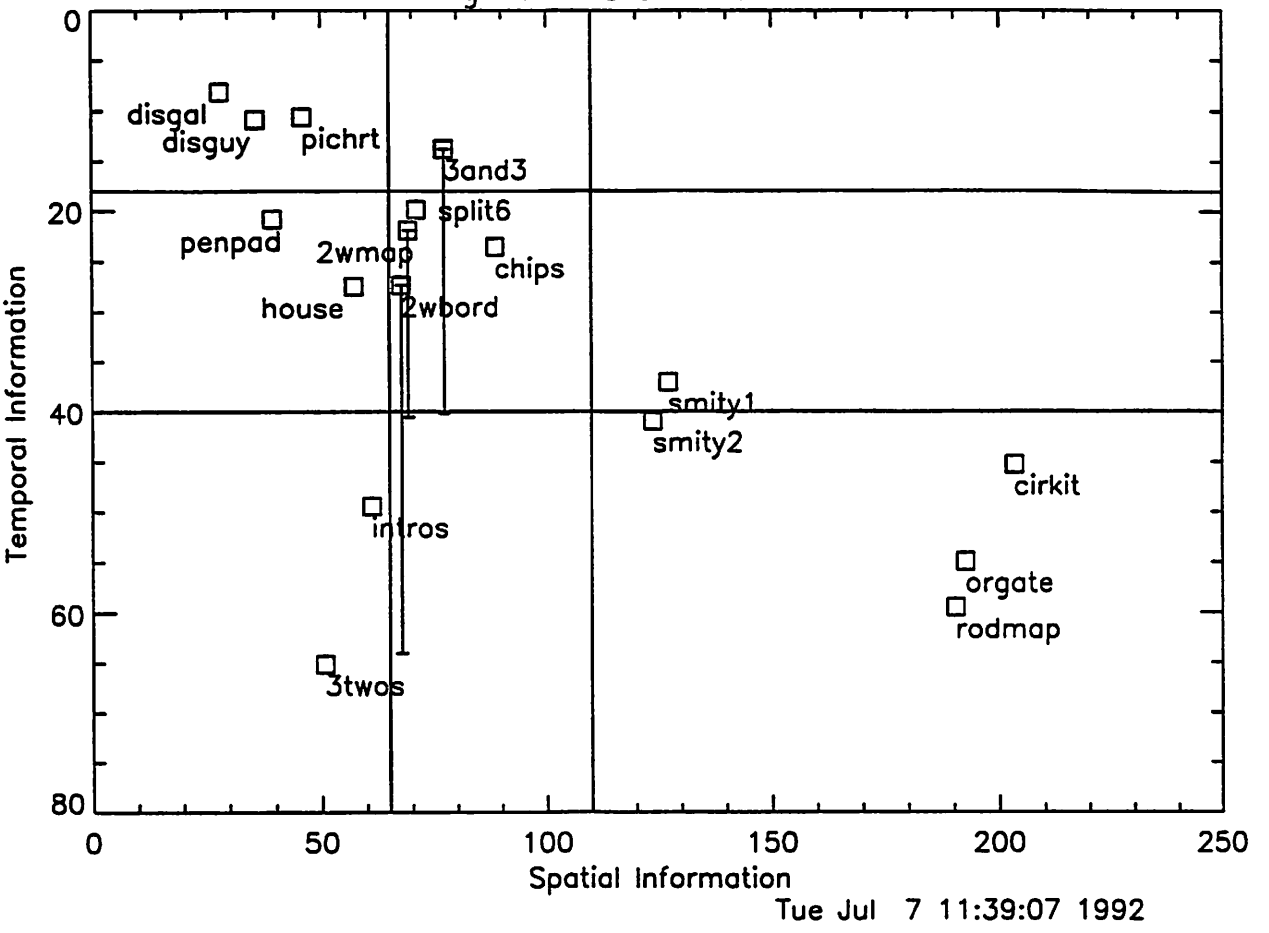

TABLE 3. DIS Test Scenes

DIS Name	Processing Name	Spatial Inform.	Temporal Inform. (w/o cuts)	Temporal Inform. (with cu
C5	2wbord	67.7	27.4	64.0
C4	2wmap	69.2	21.9	40.6
C2	3and3	77.1	13.8	40.2
C3	3twos	50.7	8.2	65.1
E1	chips	88.5	23.5	
D1	cirkit	203.3	45.3	
A2	disgal	28.1	8.1	
A1	disguy	35.8	10.9	
B2	house	57.2	27.5	
C1	intros	61.0	49.4	
B1	orgate	192.6	54.9	
B3	penpad	39.5	20.9	
C6	pichrt	46.0	10.6	
D2	rodmap	190.4	59.5	
A3	smity1	127.2	37.1	
A4	smity2	123.8	41.0	
C7	split6	71.0	19.9	

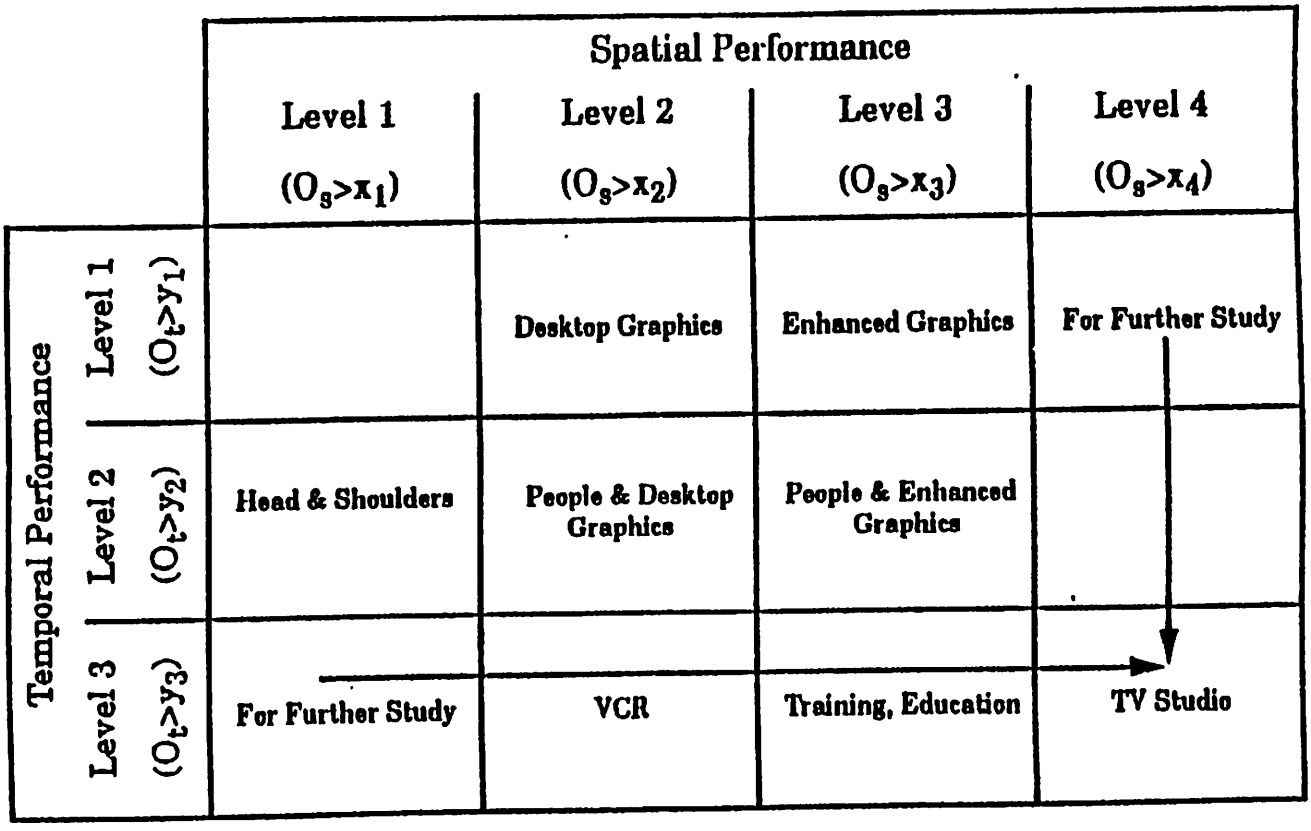

Figure 4. Video Performance Specifications

Figure 5. VTC/VT Test Scenes: ITS & CBC & DIS