Contribution to:

Turin ITU Experts Meeting on subjective and objective audiovisual quality assessment methods

October 14, 1997

Subjective and Objective Measures of Scene Criticality

Stephen Wolf and Arthur Webster
Institute for Telecommunication Sciences (NTIA/ITS)
325 Broadway, Boulder, CO 80303

I. Abstract

The difficulty of coding a video scene to achieve a constant perceived quality level increases with the amount of spatial detail and motion present in the scene. The performance of a given digital video system must therefore be expressed as a function of scene criticality, or coding difficulty. This paper describes an objective model of scene criticality that has been derived using subjective video quality judgments of test scenes from a recent large scale subjective experiment that involved 25 different scenes injected through 24 different digital video systems.[1] This objective model has several potential uses, including use as a tool for systematically selecting an appropriate range of test material without unnecessary duplication, and as a method for performing dynamic bit-rate allocation in a "constant quality, variable bit rate" statistically multiplexed transmission channel.

II. Introduction

This paper describes preliminary results of an investigation to derive a combined spatial-temporal metric for estimating scene criticality, or coding difficulty. The emphasis for this investigation was to determine if scene criticality could at least be coarsely estimated from the recent set of low bandwidth spatial information (SI) and temporal information (TI) features given in [1, 2, 3]. The advantage of using these particular image features is that they are simple to compute in real time and they can be readily transmitted or stored as digital side information due to their extremely low bandwidth and data storage requirements. They thus have use for automatic controlling and monitoring the behavior of digital video transmission systems. The SI feature that was examined here is given by

$$SI(t_n) = \mathrm{rms}_{space}[\mathrm{Sobel}(F(t_n))],$$

while the TI feature that was examined is given by

$$TI(t_n) = \mathrm{rms}_{space}[F(t_n) - F(t_{n-1})],$$

where $F(t_n)$ is the luminance video frame at time n and rms is the root mean square function. The rms operator is used to provide a measure of the total "energy".

Preliminary results on 25 scenes indicate that a coarse model of scene criticality can be derived using these image features. Further results on independent test scene data are required to determine the robustness of the scene criticality model presented in this paper.

---

[1] The digital video systems that were used included the effects of the coder, the digital transmission channel, and the decoder.

III. Subjective measure of scene criticality

In 1995, ANSI accredited committee T1A1 undertook an extensive subjective experiment that involved the subjective evaluation of 25 test scenes injected through 24 different digital video systems for a total of 600 scene-system combinations. The test scenes are described in detail in [4]. Most of the digital video systems were video teleconferencing systems that included a range of bit rates from 64 kbits/sec to 1.5 Mbits/sec. VHS and 45 Mbit/sec were the only exceptions exceptions to this. To obtain a subjective estimate of the scene criticality, we have averaged the subjective scores for each scene across all viewers and digital video systems that were used in the test. This computed average is referred to as the scene main effect by statisticians and provides a measure of the portion of the mean opinion score that is due solely to the test scene. Since a wide range of digital video systems were used in this test, the scene main effect should also provide an estimate of the scene criticality. Scenes that are the most difficult to code will have a lower scene main effect (or average mean opinion score—MOS) while scenes that are easy to code will have a higher scene main effect. Table 1 presents a summary of the subjective estimate of scene criticality (s) for the 25 test scenes. Since the subjective scores were impairment scores that ranged from 1 to 5 (i.e., 5 = "imperceptible", 4 = "perceptible but not annoying", 3 = "slightly annoying", 2 = "annoying", and 1 = "very annoying"), we see from the table that the scene main effect varied from "annoying" to somewhere between "slightly annoying" and "perceptible but not annoying." As expected, the football scene (ftball) was the most difficult to code while a head and shoulders scene (disguy) was the easiest to code. The 25 points in Table 1 were used to develop the objective model of scene criticality that is presented in this paper.

Table 1 Subjective measure of scene criticality for 25 test scenes

Scene Abbreviation	Scene Description	s (subjective measure)
ftball	Football game	2.05
cirkit	Circuit diagram, camera pan	2.16
2wbord	Two people at white board, scene cuts	2.33
rodmap	Road map with hand and pen motion, camera pan	2.56
smity2	Salesman at desk with magazine	2.56
smity1	Salesman at deck with box	2.58
flogar	Flower garden with windmill, camera pan	2.62
washdc	Washington DC map with hand and pointer	2.63
ysmite	Yosemite map with hand motion (slowly varying intensity fluctuations)	2.73
fredas	Fred Astaire tap dancing (black and white)	2.73
split6	Split screen, 6 people	2.77
intros	Introductions of people sitting at table, camera pans	2.8
boblec	Bob's lecture at chalkboard	2.86
3inrow	Men at table, camera pan	3.02
vowels	Woman at whiteboard teaching vowels	3.1
vtc2zm	Woman standing next to map with pointer, map zoom and pan	3.14
inspec	Woman at document camera	3.14
3twos	2 pairs of people, scene cuts	3.17
susie	Susie on telephone	3.28
5row1	Five people in a row sitting at a table	3.37
filter	Filter diagram on yellow pad with hand motion	3.51
disgal	Female announcer	3.65
vtc1nw	Woman sitting reading news story	3.66
vtc2mp	Woman standing next to map	3.67

disguy	Male announcer	3.68

IV. Objective measure of scene criticality

The objective model of scene criticality (o) that was developed is given by the model

$$o = 4.68 - 0.54 * p_1 - 0.46 * p_2$$

where

$$p_1 = \log_{10}\{\text{mean}_{time}[SI(t_n) * TI(t_n)]\}$$

and

$$p_2 = \log_{10}\{\text{max}_{time}[abs(SI(t_n) - SI(t_{n-1}))]\}.$$

For these equations, the parameters were computed using a time window that was the same as the subjective data (9 seconds). Parameter p1 is a measure of the average value (over time) of the instantaneous product of SI and TI, while parameter p2 is a measure of the maximum instantaneous change in SI from frame to frame. When lots of spatial-temporal gradient energy is present, the scene is difficult to code. Likewise, when the scene has rapidly changing values of SI (perhaps due to scene changes or complex motion), the scene is difficult to code. The performance of the fitted model is given in Table 2 while a plot of the performance is given in Figure 1. The coefficient of correlation between the objective and subjective measures is 0.91.

Table 2 Comparison of subjective and objective measures of scene criticality

Scene	s (subjective measure)	o (objective measure)	Error = o - s
ftball	2.05	2.34	0.29
cirkit	2.16	2.28	0.12
2wbord	2.33	2.56	0.23
rodmap	2.56	2.37	-0.19
smity2	2.56	2.37	-0.19
smity1	2.58	2.41	-0.17
flogar	2.62	2.36	-0.26
washdc	2.63	2.89	0.26
ysmite	2.73	3.01	0.28
fredas	2.73	2.82	0.09
split6	2.77	3.07	0.3
intros	2.8	2.71	-0.09
boblec	2.86	3.09	0.23
3inrow	3.02	2.96	-0.06
vowels	3.1	3.05	-0.05
vtc2zm	3.14	3.04	-0.1
inspec	3.14	3.35	0.21
3twos	3.17	3	-0.17
susie	3.28	3.05	-0.23
5row1	3.37	3.47	0.1
filter	3.51	3.17	-0.34
disgal	3.65	3.49	-0.16
vtc1nw	3.66	3.7	0.04
vtc2mp	3.67	3.58	-0.09
disguy	3.68	3.57	-0.11

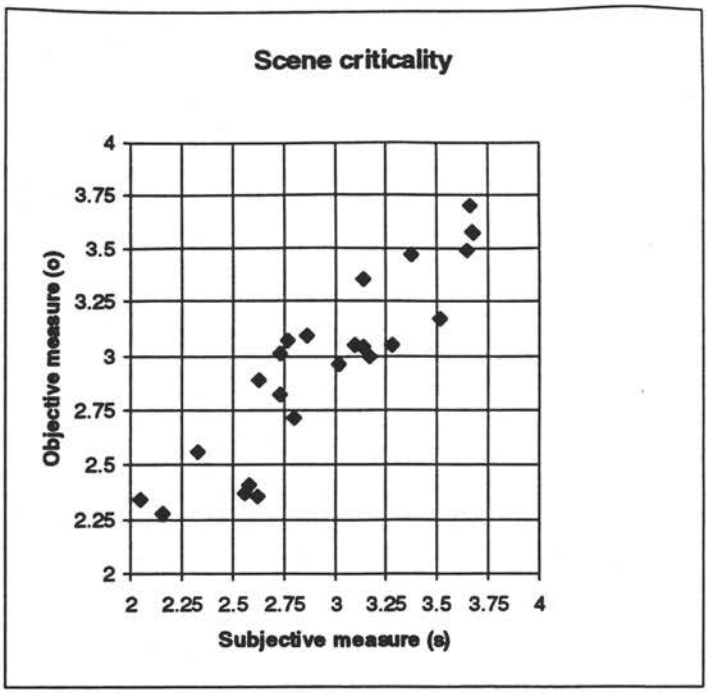

Figure 1 Plot of subjective and objective measures of scene criticality

V. Summary

This paper has presented an objective model of test scene criticality for a set of test scenes that were used to perform subjective video teleconferencing experiments. The performance of the objective model was compared with the "scene main effect", which was obtained from the subjective data by averaging over all viewers and digital video systems that were used in the test. The investigations that lead to the development of this objective model were limited to the best 2 parameter model that could be derived from the SI and TI scalar feature values. These features provide estimates of the spatial detail and temporal motion that are present in the video scene. Although these preliminary results look promising, further testing using independent subjective data from other experiments needs to be examined.

VI. References

[1] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," Recommendations of the ITU (Telecommunication Standardization Sector).

[2] ITU-T Contribution to Question 22/12, COM 12-66-E "Selections from the Draft American National Standard: Digital Transport of One-Way Video Signals - Parameters for Objective Performance Assessment," (USA), January 1996.

[3] ANSI T1.801.03-1996, "American National Standard for Telecommunications - Digital Transport of One-Way Video Telephony Signals - Parameters for Objective Performance Assessment," Alliance for Telecommunications Industry Solutions, 1200 G Street, N. W., Suite 500, Washington DC 20005.

[4] ANSI T1.801.01-1995, "American National Standard for Telecommunications - Digital Transport of Video Teleconferencing/Video Telephony Signals - Video Test Scenes for Subjective and Objective Performance Assessment," Alliance for Telecommunications Industry Solutions, 1200 G Street, N. W., Suite 500, Washington DC 20005.