T1A1 Validation Test Database

Margaret H. Pinson and Arthur Webster

Introduction

In the early 1990s, broadcasters were transitioning from analog to digital systems and ISDN video teleconferencing was an exciting new technology. During 1993-1994, the T1A1 committee conducted an objective video quality metric validation test focused on video teleconferencing applications. T1A1 was a subcommittee of the American National Standards Association (ANSI) accredited Alliance for Telecommunications Industry Solutions (ATIS). T1A1 is now known as PTSC QoSR— Packet Technologies and Systems Committee, Quality of Service and Reliability.

This document summarizes the T1A1 video quality subjective test. We focus on information that a current researcher needs to effectively use this dataset. The T1A1 video sequences and differential mean opinion scores (DMOS) are available on the Consumer Digital Video Library (CDVL, www.cdvl.org, [1]). The test plan and analyses appear in [2]-[4]. A future paper will document this test in more detail.

Scenes and Impairments

The T1A1 validation test analyzed standard definition video as per the NTSC broadcasting standard. The test focused on video teleconferencing applications. The source video sequences (SRC) are all in the public domain. Three are ITU-R Rec. BT.802 standard test sequences. The other 22 SRCs were donated by NTIA/ITS, Delta Information Systems (DIS), PictureTel Corp, and Compression Labs Inc. (CLI).

Twenty-five SRCs were chosen to represent five content categories (see Figure. 1). Most videos were filmed using broadcast quality cameras; however some intentionally included production problems (e.g., light level fluctuation, analog noise, deterioration typical of old film). These 25 SRCs later became the ANSI 801.1 standard test sequences. The digitized videos on CDVL contain occasional analog impairments stemming from the age of the tapes when it became practical to convert the analog Betacam-SP tapes into a digital format (uncompressed AVI, 4:2:2).


A. One person, mainly head and shoulders.


B. One person with graphics and/or more detail..


C. More than one person.


D. Graphics with pointing.


E. High object and/or camera motion (e.g., broadcast TV).

Figure 1. Sample frames of the 25 SRC.

The T1A1 subjective test plan [2] specifies an exact list of the 25 hypothetical reference circuits (HRC) (see Table I). T1A1 used the term HRC to intentionally eliminate vendor information from publications. The HRCs included hardware coder/decoder pairs, a VHS tape dub, and the null impairment (i.e., the original video dubbed from one Betacam-SP recorder to another). All 25 SRC were recorded to a Betacam-SP tape and separated by mid-level grey. The entire SRC tape was played through each HRC, and the output video recorded to another Betacam-SP tape.

Some of the hardware codecs used changed the system delay and frame rate in response to coding difficulty (e.g., longer delay and lower frame rate for difficult-to-code scenes). However, the delay always varied around a single system delay. There were no rebuffering delays or other mean delay changes. The T1A1 video clips include a wider range of dynamic frame rate changes than are found in modern codecs, for example dropping to ≈1 fps during high motion.

The SRC as played into the encoder had three seconds of extra content at the beginning. The extra SRC content ensured that encoding problems and errors would not cause mid-level grey to propagate into the sequence.

Table 1. HRC Descriptions

HRC	*Algorithm (vendor)*	*Resolution*	*Total kbps*	*Audio kbps*	*Video kbps*	*Coding Mode*	*Frame Rate*	*FEC*	*Burst Errors*
1	Null	—	—	—	—	—	30	—	Off
2	VHS	—	—	—	—	—	30	—	Off
3	Proprietary	V. High	45,000	—	—	—	—	—	Off
4	Proprietary	Med.	128	—	—	VQ	—	—	Off
5	Proprietary	High	336	—	—	VQ	—	—	Off
6	Proprietary	Med.	112	—	—	—	—	—	Off
7	Proprietary	Med.	384	—	—	—	—	—	Off
8	Proprietary	Med.	768	—	—	—	—	—	Off
9	Proprietary	High	768	—	—	—	—	—	Off
10	Proprietary	High	1536	—	—	—	—	—	Off
11	H.261 (diff)	QCIF	128	56	70.4	INTER+MC	—	On	Off
12	H.261 (same)	QCIF	128	56	70.4	INTER	10	On	Off
13	H.261 (same)	QCIF	168	48	118.4	INTER+MC	—	On	Off
14	H.261 (diff)	QCIF	384	56	326.4	INTER+MC	—	On	Off
15	H.261 (same)	CIF	112	48	62.4	INTER+MC	—	On	Off
16	H.261 (same)	CIF	128	56	70.4	INTER+MC	—	On	Off
17	H.261 (diff)	CIF	128	48	78.4	INTER+MC	—	On	Off
18	H.261 (same)	CIF	168	48	118.4	INTER+MC	—	On	Off
19	H.261 (same)	CIF	256	56	190.4	INTER+MC	15	On	On
20	H.261 (same)	CIF	384	56	326.4	INTER+MC	—	On	Off
21	H.261 (same)	CIF	384	56	326.4	INTER+MC	—	On	On
22	H.261 (diff)	CIF	768	56	710.4	INTER+MC	—	On	Off
23	H.261 (same)	CIF	768	56	710.4	INTER+MC	—	On	On
24	H.261 (diff)	CIF	1536	56	1478.4	INTER+MC	—	On	Off
25	H.261 (same)	CIF	1536	56	1478.4	INTER+MC	—	On	Off

"Null" is the original SRC recording compared to itself; "VQ" is vector quantization; "FEC" = forward error correction; "INTER" = inter-frame coding; "MC" = "motion compensation"; "Burst Errors" = bursts of bit-errors; "—" = variable not specified; "same" = same coder and decoder manufacturer; and "diff" = different coder and decoder manufacturers.

Subjective Testing

The T1A1 subjective test was conducted according to ITU-R Rec. BT.500-5 using the double stimulus impairment scale (DSIS). Although the currently in-force BT.500-13 excludes DSIS, this method appears in ITU-T Rec. P.910 under the name degradation category rating (DCR). The test was conducted using Betacam-SP tapes, written scoring sheets, and a broadcast quality CRT monitor.

The entire test includes 625 processed video sequences (PVS), which was too much for any single subject to comfortably rate. Instead, the PVSs were divided into three pools of 10 HRCs

each. Overlapping HRCs promoted consistent scoring between subject pools, but those extra scores were discarded. Three subjective labs (NTIA/ITS, GTE, and DIS) each gathered one-third of the data for each pool. The T1A1 subjective data includes ratings from 30 subjects for each PVS (i.e., ten from each lab). An analysis by Cermak and Fay [3] found that the data from these three labs were not statistically different.

The DMOS scores for HRCs 2 through 25 are available on CDVL with the video sequences. The DMOS scores for HRC 1 (null, now labeled "original") were misplaced. At this time, the raw subjective data is only available from one of the three labs (in Contribution T1A1.5/94-143 [5]).

The videos on CDVL contain, for each PVS, all available content on the HRC tape: mid-level grey frames, 3 sec pre roll, 9 sec sequence, 1 sec post-roll, and mid-level grey frames. A spreadsheet (redistributed with the video sequences) lists the following information for the 600 PVSs:

- DMOS
- Standard deviation of differential opinion scores
- Spatial shift in frame lines vertically and pixels horizontally
- Luma gain & level offset values
- Time aligned segment (start frame & stop frame)

These calibration values were calculated using the NTIA/ITS full-reference temporal registration algorithms [6], followed by a manual inspection. This algorithm finds a typical delay for the entire sequence. The time alignments used for the viewing tape edits were chosen by eye and so differ slightly from the spreadsheet values.

Conclusion

Is the T1A1 dataset valuable today, since it examines 20 year old technology? The NTIA/ITS philosophy is to encourage the development of technology independent metrics. If an

objective model is rooted in the image receptors of the eye, and the visual cortex and image processing centers of the brain, then it should be accurate for the T1A1 dataset. Such flexibility indicates resilience: an objective model whose performance will degrade gracefully as coding technology continues to change.

References

[1] M. Pinson, "The Consumer Digital Video Library [Best of the Web]," IEEE Signal Processing Magazine, vol. 30, no. 4, pp. 172-174, Jul. 2013. doi: 10.1109/MSP.2013.2258265

[2] T1A1.5/94-118 R1 "Subjective test plan (tenth and final draft)," Detailed Test Plan Ad Hoc Group and Data Analysis Ad Hoc Group, (A.C. Morton, Editor) 3 Oct. 1993. See ftp://vqeg.its.bldrdoc.gov/Documents/OLD_T1A1/

[3] T1A1.5/94-148 "T1A1.5 video quality project: GTE Labs Analysis," GTE Laboratories, Inc. (Gregory W. Cermak and David A. Fay), 20 Sep. 1994. See ftp://vqeg.its.bldrdoc.gov/Documents/OLD_T1A1/

[4] T1A1.5/94-152 "Analysis of T1A1.5 subjective and objective test data," NTIA/ITS (Coleen Jones, Ned Crow, Stephen Wolf, Arthur Webster), 3 Oct. 1994. See ftp://vqeg.its.bldrdoc.gov/Documents/OLD_T1A1/

[5] T1A1.5/94-143 "DIS/NCS subjective test data," Delta Information Systems / National Communications Systems (DIS/NCS), 14 Jul. 1994. See ftp://vqeg.its.bldrdoc.gov/Documents/OLD_T1A1/

[6] M. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," IEEE Transactions on Broadcasting, vol.50, no.3, pp. 312-322, Sept. 2004. See http://www.its.bldrdoc.gov/publications/2576.aspx.

Margaret Pinson (top) and Arthur Webster (bottom) participated in the T1A1 validation test as proponents. The NTIA/ITS General Model for video quality (VQM) was trained on this dataset. Both are with the National Telecommunications and Information Administration, Institute for Telecommunication Sciences in Boulder, Colorado.