Technical University Berlin

Telecommunication Networks Group

MPEG–4 and H.263 Video Traces for Network Performance Evaluation

Frank H.P. Fitzek and Martin Reisslein

fitzek@ee.tu-berlin.de, reisslein@asu.edu

http://www-tkn.ee.tu-berlin.de/research/trace/trace.html

http://www-tkn.ee.tu-berlin.de/~fitzek/TRACE/trace.html

Berlin, October 2000

TKN Technical Report TKN–00–06

TKN Technical Report Series

Editor: Prof. Dr.–Ing. Adam Wolisz

MPEG–4 and H.263 Video Traces for Network Performance Evaluation

Frank H.P. Fitzek
Technical University Berlin
Dept. of Electrical Eng.
Telecommunication Networks Group
Einsteinufer 25
10585 Berlin, Germany
fitzek@ee.tu-berlin.de
www-tkn.ee.tu-berlin.de/~fitzek

Martin Reisslein *
Arizona State University
Dept. of Electrical Eng.
Telecommunications Research Center
P.O. Box 877206
Tempe, AZ 85287–7206
reisslein@asu.edu
www.public.asu.edu/~mreissl

October 2000

Abstract

MPEG–4 and H.263 encoded video is expected to account for a large portion of the traffic in future wireline and wireless networks. However, due to a lack of sufficiently long frame size traces of MPEG–4 and H.263 encoded videos, most network performance evaluations use currently MPEG–1 encodings. In this technical report we present a publicly available library of frame size traces of long MPEG–4 and H.263 encoded videos, which we have generated at the Technical University Berlin. The frame size traces have been generated from MPEG–4 and H.263 encodings of over 10 video sequences of 60 minutes length each. We also present a thorough statistical analysis of the traces.

Keywords: Frame Size Traces, H.263, Long Range Dependence, MPEG–4, Statistical Analysis, Video Encoding.

1 Introduction

MPEG–4 and H.263 encoded video is expected to account for large portions of the traffic in future wireline and wireless networks. To date the statistical analysis of MPEG–4 and H.263 encoded video has received only little attention in the literature. Similarly, there are only few studies that evaluate networking protocols and resource management schemes with MPEG–4 and H.263 encoded video. This is partly due to a lack of sufficiently long frame size traces

* Most of this work was conducted while Martin Reisslein was with GMD Fokus, Berlin, Germany.

of MPEG–4 and H.263 encoded videos. In fact, most researches currently use the MPEG–1 encodings of Garret [4], Rose [15], or Krunz *et al.* [9].

In the Telecommunication Networks (TKN) Group at the Technical University Berlin we have generated a publicly available library of frame size traces of long MPEG–4 and H.263 encoded videos. The frame size traces have been generated from MPEG–4 and H.263 encodings of over 10 video sequences of 60 minutes length each. We present a thorough statistical analysis of the frame size traces. We study moments and autocorrelations as well as the long range dependence characteristics. We estimate the Hurst parameter of the traces with the R/S statistic.

This technical report is structured as follows. In the following section we give an overview of the encoded video sequences and discuss the grabbing of the uncompressed YUV video information. In Section 2 we discuss the MPEG–4 encodings. We give an overview of MPEG–4 video compression and discuss our encoding procedure in detail. We then conduct a thorough statistical analysis of the MPEG–4 frame size traces. In Section 3 we discuss the H.263 encodings. We briefly discuss the basics of H.263 video compression and describe our encoding procedure in detail. We then conduct a thorough statistical analysis of the H.263 traces. We summarize our contributions in Section 4. In the Appendix we review the statistical methods used in the analysis of the traces.

1.1 Overview of Encoded Video Sequences

We played the videos [1] listed in Table 1 from VHS tapes using a Video Cassette Recorder (VCR). For ease of comparison with the existing MPEG–1 traces we included *Star Wars IV*, which has been MPEG–1 encoded by Garrett [4], and several of the movies that have been MPEG–1 encoded by Rose [15] in our video selection. For each video we grabbed the (uncompressed) YUV information with `bttvgrab` (Version 0.15.10) [17] and stored it on disk.

[1] To avoid any conflict with copyright laws, we emphasize that all image processing, encoding, and analysis was done for scientific purposes. The encoded video sequences have no audio stream and are not publicly available. We make only the frame size traces available to researches.

Movies (rental tapes, German/English movie versions)
Jurassic Park I (G)
Silence of the Lambs (E)
Star Wars IV (E)
Mr. Bean (G)
Star Trek: First Contact (G)
Form Dusk Till Dawn (G)
The Firm (G)
Sports Events (recorded from German cable TV)
Formula 1: Formula 1 car race
Soccer: Soccer game (European championship 1996)
Other TV sequences (recorded from German cable TV)
ARD News: German news (Tagesschau)
ARD Talk: German Sunday morning talk show (Presseclub)
N3 Talk: German late night show (Herman und Tietjen)
Set–top
Office–Cam: Office camera observing person in front of terminal

Table 1: Overview of encoded video sequences.

The YUV information was grabbed at a frame rate of 25 frames/sec in the QCIF format, that is, with a luminance resolution of 176x144 picture elements (pels) and 4:1:1 chrominance subsampling at a color depth of 8 bits. We chose the QCIF format because we are particularly interested in generating traces for the evaluation of wireless networking systems. We expect that hand–held wireless devices of next–generation wireless systems will typically have a screen size that corresponds to the QCIF video format. We note that `bttvgrab` is a high–quality grabber. It is designed to not leave out a single video frame and to overcome temporal delays by buffering several frames. To avoid potential hardware problems due to buffer build–up when grabbing long video sequences, we grabbed the 60 minutes video sequences in segments of 22,501 frames ($\approx$ 15 minutes of video run time). Any 22,501 frame segment of any video gave exactly 855,398,016 Bytes of uncompressed YUV information (= 38,016 Bytes/frame). The stored YUV frame sequences were used as input for both the MPEG–4 encoder and the H.263 encoder. We emphasize that we did not encode in real–time; thus there was no encoder bottleneck.

2 MPEG–4 Video Traces

2.1 Overview of MPEG–4 Video Compression

In this section we provide a brief overview of MPEG–4 video coding; we refer the reader to [3, 8, 16, 13] for details. MPEG–4 provides very efficient video coding covering the range from the very low bit rates of wireless communication to bit rates and quality levels beyond high definition television (HDTV). In contrast to the "frame–based" video coding of MPEG–1 and H.263, MPEG–4 is object based. Each scene is composed of Video Objects (VOs) that are coded individually. (If scene segmentation is not available or not useful, e.g., in very simple wireless video communication, the standard defines the entire scene as one VO.) Each VO may have several scalability layers (i.e., one base layer and one or several enhancement layers) which are referred to as Video Object Layers (VOLs) in MPEG–4 terminology. Each VOL in turn consists of an ordered sequence of snapshots in time, referred to as Video Object Planes (VOPs). For each VOP the encoder processes the shape, motion, and texture characteristics.

The shape information is encoded by bounding the VO with a rectangular box and then dividing the bounding box into Macro Blocks (MBs). Each MB is classified as lying (*i*) inside the object, (*ii*) on the object's border, or (*iii*) outside the object (but inside the bounding box). The "border" MBs are then shape coded. The texture coding is done on a per–block basis similar to the "frame–based" standards, such as MPEG–1 and H.263. In an Intracoded (I) VOP the absolute texture values in each MB are Discrete Cosine Transform (DCT) coded. The DCT coefficients are then quantized and variable–length–coded. In forward Predicted (P) VOPs each MB is predicted from the closest match in the preceding I (or P) VOP using motion vectors. In Bi–directionally predicted (B) VOPs each MB is predicted from the preceding I (or P) VOP and the succeeding P (or I) VOP. The prediction errors are DCT coded, quantized, and variable–length–coded. For the transmission the shape, motion, and texture information is multiplexed at the MB level, i.e., for a given MB the shape information is transmitted first, then the motion information, and then the texture information, then the shape information of the next MB, and so on. To combat the frequent transmission errors typical for wireless communication, MPEG–4 provides a number of error resilience and error concealment features; we refer the reader to [3, 8, 16] for details.

2.2 Our Encoding Approach for MPEG–4

For each video we encoded the YUV information into an MPEG–4 bit stream with the MOMUSYS MPEG–4 video software [5], which has been adopted by MPEG in the MPEG–4 standard, Part 5 — Reference Software. For the complete listing of the parameter setting used in the video encoding we refer the reader to the control files `config_enc.ctl` and the configuration files `config.dat` available on our web site. In summary, we set the number of video objects to one, i.e., the entire scene is one video object. The width of the display is set to 176 pels, the height is set to 144 pels. We used a pel depth of 8 bits per pel. We did not use rate control in the encoding. The single video object was encoded into a single video object layer. We set the video object layer frame rate, i.e., the rate at which video object planes are generated, to 25 frames/sec. The Group of Pictures (GoP) pattern was set to IBBPBBPBBPBB. We encoded each video at three different quality levels: *low*, *medium*, and *high*. For the low quality encoding the quantization parameters were fixed at 10 for I frames (VOPs), 14 for P frames, and 18 for B frames. For the medium quality encoding the quantization parameters for all three frame types were fixed at 10. For the high quality encoding the quantization parameters for all three frame types were fixed at 4.

We note that the MOMUSYS MPEG–4 encoder is limited to encoding segments with a length of at most 1,000 video frames. Therefore, we encoded the YUV frame sequences in segments of 960 frames (= 80 GoPs) each. When encoding a given 80 GoP segment, the last two B frames of the 80th GoP are bi–directionally predicted from the third P frame of the 80th GoP and the I frame of the 81st GoP. Since the 81st GoP is not encoded, the last two B frames of the 80th GoP are not encoded either. As a consequence our trace files were missing two B frames per 960 encoded video frames (= 38.4 seconds of video run time). As a remedy we inserted two B frames at the end of each segment of 958 (actually encoded) frames. We set the size of the inserted B frames to the average size of the B frames in the 958 frame segment. We believe that this error, that is due to the limitations of the MOMUSYS MPEG–4 Reference Software, can be neglected.

For each encoded video and quality level we provide a verbose and terse trace file on our web site. The verbose trace files give Frame Number, Frame Type (I, P, or B), Display Time (in msec), and Frame Size (in bytes) in ASCII format with one frame per line. Note that in the verbose files the frames are ordered in the sequence IPBBPBBPBBIBBP.... This is because for decoding a B frame the decoder needs both the preceding I (or P) frame and the succeeding P (or I) frame. The frames are therefore emitted in the sequence IPBB... by the encoder (and typically transmitted in that sequence in practice). The terse trace files give the Frame Size (in bytes) in the sequence IBBPBBPBBPBBIBBP... following the format of the existing MPEG–1 traces.

2.3 Statistical Analysis of MPEG–4 Traces

In this section we conduct a thorough statistical analysis of the generated MPEG–4 frame size traces. For the analysis we introduce the following notation. Let N denote the number of video frames in a given trace. Let t denote the frame period (display time) of a given frame. Note that for all our MPEG–4 traces approximately $N = 90.000$ and $t = 40$ msec, which corresponds to a video runtime of about 60 minutes. Let $X_n,\ n = 1, \ldots, N$, denote the number of bits in frame n, that is, the frame size of frame n. Let G denote the number of frames per Group of Pictures (GoP). Let $Y_m,\ m = 1, \ldots, N/G$, denote the number of bits in GoP m, that is, the size of GoP m. Clearly, $Y_m = \sum_{n=(m-1)G+1}^{mG} X_n$. Tables 2 and 3 give an overview of the statistical properties of the generated MPEG–4 traces. We refer the reader to the Appendix for the formal definitions of the mean and coefficient of variation. The compression ratio is defined as the ratio of the size of the entire uncompressed YUV video sequence (in bit) to the size of the entire MPEG–4 compressed video sequence (in bit). Comparing encodings at different

Quality	Trace	Compr. ratio YUV:MP4	Frame Size Mean $\bar{X}$ [kbyte]	Frame Size CoV $S_X/\bar{X}$	Frame Size Peak/Mean $X_{\max}/\bar{X}$	Bit Rate Mean $\bar{X}/t$ [Mbps]	Bit Rate Peak $X_{\max}/t$ [Mbps]
High	*Jurassic Park I*	9.92	3.8	0.59	4.37	0.77	3.3
	Silence of the Lambs	13.22	2.9	0.80	7,73	0.58	4.4
	Star Wars IV	27.62	1.4	0.66	6.81	0.28	1.9
	Mr. Bean	13.06	2.9	0.62	5.24	0.58	3.1
	First Contact	23.11	1.6	0.73	7.59	0.33	2.5
	From Dusk Till Dawn	11.16	3.4	0.58	4.62	0.68	3.1
	The Firm	24.53	1.5	0.75	6.69	0.31	2.1
	Formula 1	9.10	4.2	0.42	3.45	0.84	2.9
	Soccer	6.87	5.5	0.41	3.24	1.10	3.6
	ARD News	10.52	3.6	0.70	4.72	0.72	3.4
	ARD Talk	13.95	2.7	0.63	5.72	0.54	3.1
	N3 Talk	13.76	2.8	0.60	6.17	0.55	3.4
	Office–Cam	19.16	2.0	1.09	4.99	0.40	2.0
Medium	*Jurassic Park I*	28.4	1.3	0.84	6.36	0.27	1.7
	Silence of the Lambs	43.43	0.88	1.21	13.6	0.18	2.4
	Star Wars IV	97.83	0.39	1.17	12.1	0.08	0.94
	Mr. Bean	41.34	0.92	0.97	8.23	0.18	1.5
	First Contact	70.56	0.54	1.02	11.03	0.11	1.2
	Formula 1	26.11	1.50	0.64	4.83	0.29	1.4
	Office–Cam	68.13	0.56	2.27	9.29	0.11	1.0
Low	*Jurassic Park I*	49.46	0.77	1.39	10.61	0.15	1.6
	Silence of the Lambs	72.01	0.53	1.66	21.39	0.11	2.3
	Star Wars IV	142.52	0.27	1.68	17.57	0.053	0.94
	Mr. Bean	66.60	0.57	1.55	13.25	0.11	1.5
	First Contact	110.94	0.34	1.50	17.08	0.069	1.2
	Formula 1	43.51	0.87	1.12	8.05	0.17	1.4
	Office–Cam	84.20	0.45	2.87	11.48	0.09	1.0

Table 2: Overview of frame statistics of MPEG–4 traces

quality levels we observe that lower quality encoding achieves higher compression ratios, as is to be expected. Interestingly, these higher compression ratios come at the expense of increased variability of the encoded video streams. We observe that relatively high compression ratios are achieved for the *Star Wars IV* and *First Contact* movies even for high quality encodings. This is probably due to the long scenes with dark backdrops and little contrast in these movies. For the *Formula 1* and *Soccer* videos, on the other hand, only relatively small compression ratios are achieved. These videos feature many small objects that move rapidly. This results in high mean bit rates and relatively small peak–to–mean ratios of the encoded frame sizes. Comparing the frame statistics and the GoP statistics we observe that smoothing the videos over one GoP (= 0.48 sec of video runtime) is quite effective in reducing the variability and

Quality	Trace	GoP Size Mean $\bar{Y}$ [kbyte]	GoP Size CoV $S_Y/\bar{Y}$	GoP Size Peak/Mean $Y_{\max}/\bar{Y}$	Bit Rate Mean $\bar{Y}/(Gt)$ [Mbps]	Bit Rate Peak $Y_{\max}/(Gt)$ [Mbps]
High	*Jurassic Park I*	46	0.47	3.15	0.77	2.4
	Silence of the Lambs	35	0.71	6.22	0.58	3.6
	Star Wars IV	17	0.38	4.29	0.28	1.2
	Mr. Bean	35	0.40	3.73	0.58	2.2
	First Contact	20	0.56	5.02	0.33	1.7
	From Dusk Till Dawn	41	0.43	3.40	0.68	2.3
	The Firm	19	0.52	4.64	0.31	1.4
	Formula 1	50	0.30	2.40	0.84	2.0
	Soccer	66	0.33	2.72	1.10	3.0
	ARD News	43	0.51	3.04	0.72	2.2
	ARD Talk	33	0.31	3.60	0.54	2.0
	N3 Talk	33	0.36	3.18	0.55	1.8
	Office–Cam	24	0.10	1.74	0.40	0.6
Medium	*Jurassic Park I*	16	0.57	3.92	0.27	1.0
	Silence of the Lambs	11.0	0.99	10.07	0.18	1.8
	Star Wars IV	4.7	0.52	6.29	0.08	0.49
	Mr. Bean	11.0	0.50	5.14	0.18	0.34
	First Contact	6.5	0.64	6.71	0.11	0.72
	Formula 1	17.0	0.35	2.86	0.29	0.83
	Office–Cam	6.7	0.14	2.11	0.11	0.24
Low	*Jurassic Park I*	9.20	0.53	4.05	0.15	0.62
	Silence of the Lambs	6.30	0.92	10.48	0.11	1.10
	Star Wars IV	3.20	0.46	5.31	0.053	0.28
	Mr. Bean	6.9	0.46	5.08	0.11	0.58
	First Contact	4.1	0.54	6.03	0.069	0.41
	Formula 1	10.0	0.32	2.77	0.17	0.48
	Office–Cam	5.4	0.09	1.80	0.09	0.16

Table 3: Overview of GoP statistics of MPEG–4 traces

the peak rate.

In the following we provide plots to illustrate the statistical properties of the following three MPEG–4 traces: (a) *Star Wars IV* encoded at high quality, (b) *Jurassic Park I* encoded at medium quality, and (c) *Silence of the Lambs* encoded at low quality. We refer the reader to our web site for the corresponding plots for the other generated traces. Figure 1 gives the frame size traces, i.e., the frame size X_n (in bytes) as a function of the frame number n. We observe from the plots that the *Star Wars IV* encoding at high quality is relatively smooth. The *Silence of the Lambs* encoding at low quality, on the other hand, exhibits extreme changes in the frame sizes. Inspecting this trace closely, we are able to identify periods during which the frame sizes stay roughly at a fixed level; these periods appear to correspond to distinct scenes in the movie.

Figure 2 gives the histograms of the frame size X_n. The histogram plots reflect again the general tendency that lower quality encodings (i.e., a higher compression ratios) result in more variability of the encoded video stream. The difficulty in modeling the frame size distributions is illustrated by the histogram for the *Silence of the Lambs* encoding at low quality; it has a pronounced gap around the frame size of 60 Byte.

Figure 3 gives the autocorrelation coefficient $\rho_X(k)$ (see Appendix for the formal definition) of the frame size sequence $X_n,\ n = 1, \ldots, N$, as a function of the lag k (in frames). The frame size correlations exhibit a periodic spike pattern that is superimposed on a decaying slope. The periodic spike pattern reflects the repetitive GoP pattern. The large positive spikes are due to (the typically large) I frames. An I frame is followed by two (typically small) B frames, which appear as small negative spikes. The subsequent P frame (typically of mid–size) shows up as a small positive spike. The decaying slope is characteristic of the long term correlations in the encoded video. To get a clearer picture of these long term correlations we show in Figure 4 the autocorrelation coefficient $\rho_Y(k)$ of the GoP size sequence $Y_m,\ m = 1, \ldots, N/G$, as a function of the lag k (in GoPs). We observe from the figure that the GoP autocorrelation function of the *Jurassic Park I* encoding at medium quality decays roughly exponentially. This indicates that the GoP size process is memoryless. The other two curves clearly decay slower than an exponential function. This slow decay of the GoP autocorrelation is particularly pronounced for the *Silence of the Lambs* encoding at low quality, which has an autocorrelation coefficient of roughly 0.2 for a lag of 230 GoPs (approximately 110 sec).

For a more accurate characterization of the long range dependences in the encoded MPEG-4 videos we give in Table 4 the Hurst parameters as a function of the aggregation level a. We estimated the Hurst parameters from pox plots of the R/S statistic as outlined in the Appendix. Figure 5 gives some pox plots of R/S for an aggregation level of $a = 1$. Generally speaking, time series without long range dependence have a Hurst parameter of 0.5. Hurst parameters between 0.5 and 1 for all aggregation levels indicate that a given trace has long

		Aggregation level a [frames]										
Quality	Trace	1	12	50	100	200	300	400	500	600	700	800
High	*Jurassic Park I*	0.973	0.830	0.795	0.774	0.737	0.753	0.666	0.653	0.705	0.591	0.622
	Silence of the Lambs	1.007	0.894	0.872	0.868	0.894	0.852	0.819	0.741	0.765	0.728	0.771
	Star Wars IV	0.903	0.838	0.808	0.785	0.776	0.765	0.756	0.758	0.752	0.727	0.722
	Mr. Bean	0.933	0.866	0.866	0.861	0.824	0.740	0.792	0.819	0.765	0.793	0.810
	First Contact	0.931	0.831	0.807	0.791	0.763	0.760	0.726	0.747	0.777	0.776	0.784
	Form Dusk Till Dawn	0.909	0.829	0.806	0.781	0.754	0.726	0.733	0.773	0.728	0.786	0.795
	The Firm	0.969	0.889	0.864	0.859	0.862	0.805	0.785	0.817	0.764	0.760	0.790
	Formula 1	0.867	0.732	0.682	0.600	0.571	0.515	0.601	0.646	0.497	0.445	0.527
	Soccer	0.837	0.701	0.642	0.639	0.652	0.610	0.672	0.632	0.651	0.666	0.694
	ARD News	0.967	0.852	0.709	0.602	0.565	0.567	0.535	0.405	0.409	0.333	0.362
	ARD Talk	0.903	0.863	0.796	0.768	0.772	0.719	0.663	0.684	0.634	0.580	0.532
	N3 Talk	0.882	0.840	0.846	0.868	0.772	0.719	0.663	0.684	0.634	0.580	0.532
	Office–Cam	0.607	0.886	0.850	0.858	0.858	0.884	0.927	0.859	0.940	0.973	0.960
Medium	*Jurassic Park I*	0.948	0.821	0.776	0.756	0.722	0.732	0.630	0.633	0.664	0.549	0.579
	Silence of the Lambs	0.997	0.891	0.867	0.866	0.896	0.864	0.849	0.765	0.781	0.736	0.760
	Star Wars IV	0.847	0.846	0.822	0.798	0.787	0.786	0.770	0.785	0.823	0.815	0.803
	Mr. Bean	0.890	0.868	0.833	0.830	0.815	0.746	0.804	0.833	0.772	0.802	0.801
	First Contact	0.901	0.833	0.803	0.789	0.767	0.740	0.731	0.747	0.743	0.737	0.760
	Formula 1	0.826	0.741	0.695	0.598	0.552	0.521	0.585	0.668	0.510	0.502	0.588
	Office–Cam	0.532	0.882	0.812	0.810	0.810	0.833	0.856	0.809	0.861	0.906	0.883
Low	*Jurassic Park I*	0.881	0.824	0.771	0.752	0.729	0.726	0.628	0.636	0.642	0.538	0.580
	Silence of the Lambs	0.935	0.887	0.863	0.858	0.882	0.867	0.842	0.769	0.777	0.744	0.764
	Star Wars IV	0.770	0.844	0.814	0.795	0.785	0.790	0.769	0.787	0.833	0.817	0.805
	Mr. Bean	0.816	0.871	0.823	0.824	0.813	0.735	0.796	0.809	0.750	0.790	0.791
	First Contact	0.828	0.834	0.802	0.795	0.774	0.741	0.744	0.742	0.740	0.746	0.779
	Formula 1	0.736	0.739	0.685	0.586	0.550	0.512	0.574	0.674	0.542	0.565	0.606
	Office–Cam	0.441	0.919	0.796	0.860	0.872	0.854	0.900	0.822	0.853	0.929	0.887

Table 4: Hurst parameters of MPEG–4 traces estimated from pox diagram of R/S as a function of the aggregation level a.

range dependence properties. Additionally, larger Hurst parameters indicate a higher degree of long range dependence. We observe from the table that the encodings of *Silence of the Lambs*, *Star Wars IV*, *Mr. Bean*, *First Contact*, *From Dusk Till Dawn*, and *The Firm* have Hurst parameters larger than 0.72 for all aggregation levels. This indicates a high degree of long range dependence. The *Formula 1* and *ARD news* encodings have large Hurst parameters for aggregation levels of 50 frames and less; for aggregation levels of 200 frames and larger, however, the Hurst parameters are around 0.5. These results are thus not a strong indication of long range dependence. It is also interesting to note that the Hurst parameters for the *Jurassic Park I* encodings do not give a strong indication of long range dependence properties. This corroborates the observation that the GoP autocorrelation functions decay almost exponentially; thus indicating the memoryless property.

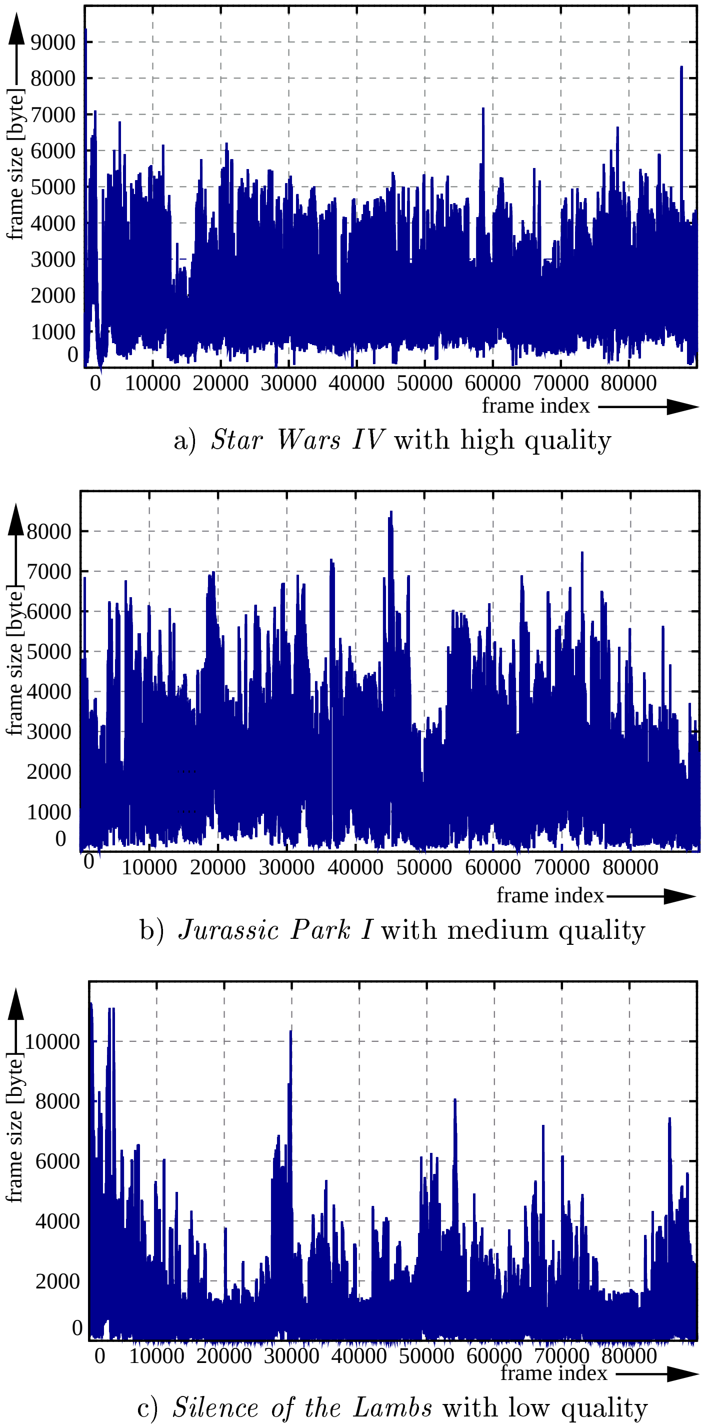
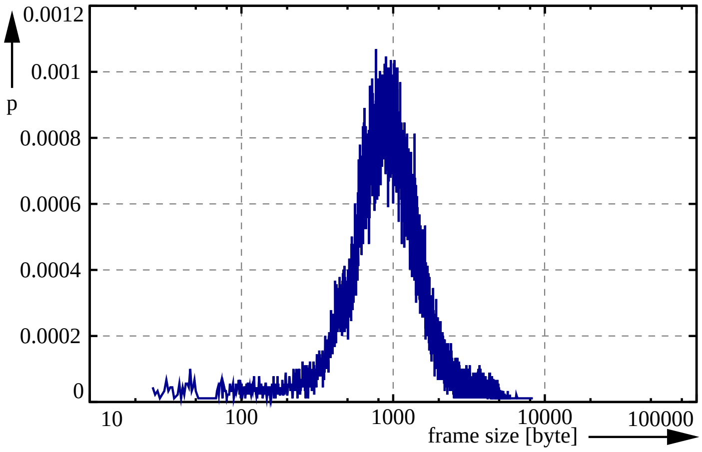
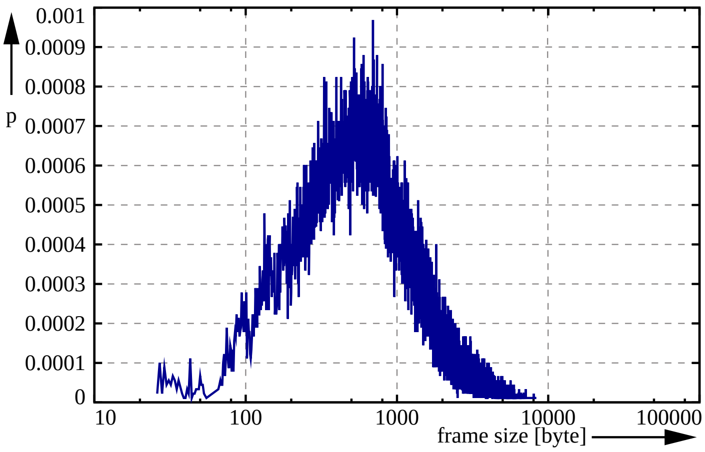


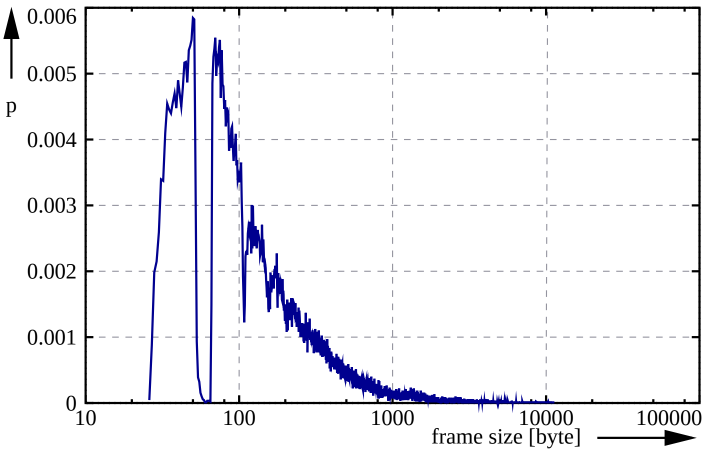
Figure 1: MPEG–4 frame size traces.



a) *Star Wars IV* with high quality

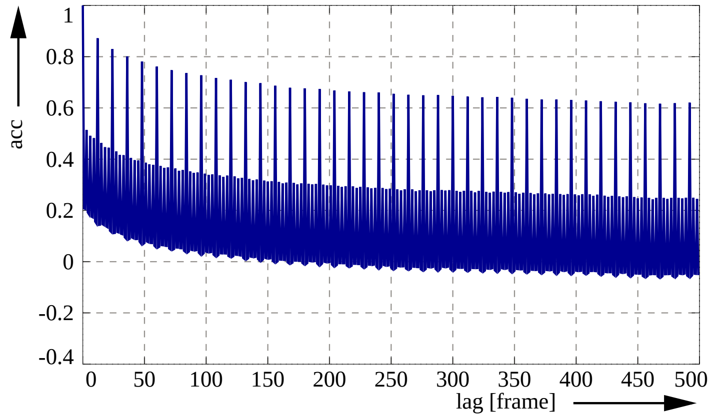


b) *Jurassic Park I* with medium quality

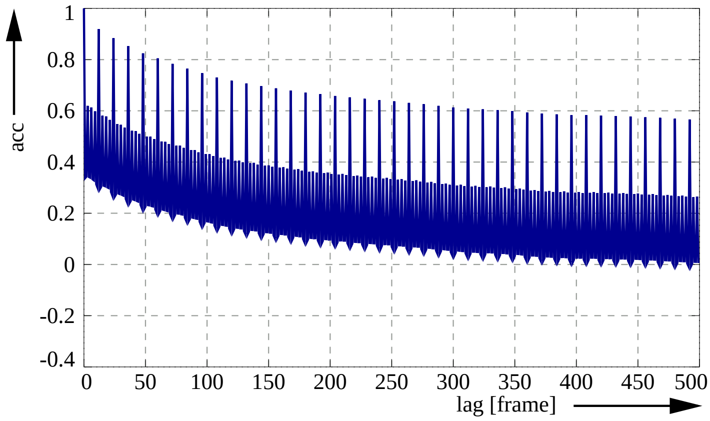


c) *Silence of the Lambs* with low quality

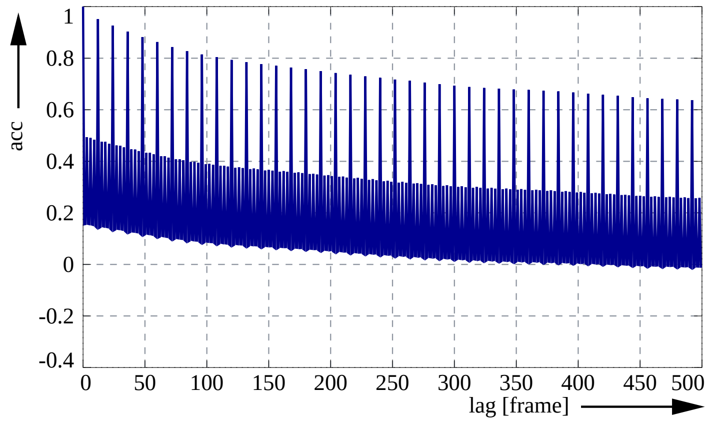
Figure 2: MPEG–4 frame size histograms.



a) *Star Wars IV* with high quality

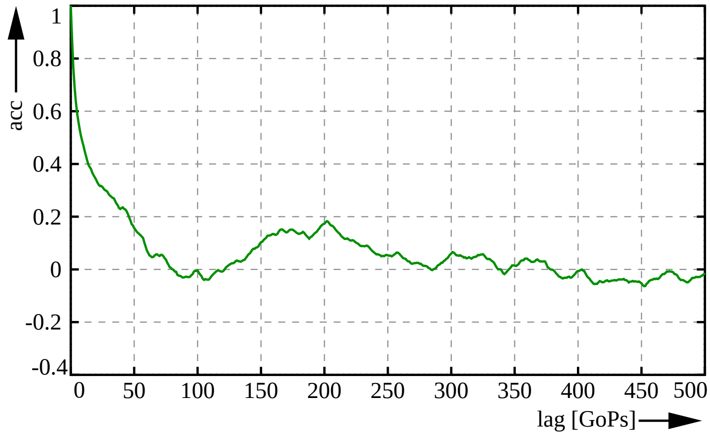


b) *Jurassic Park I* with medium quality

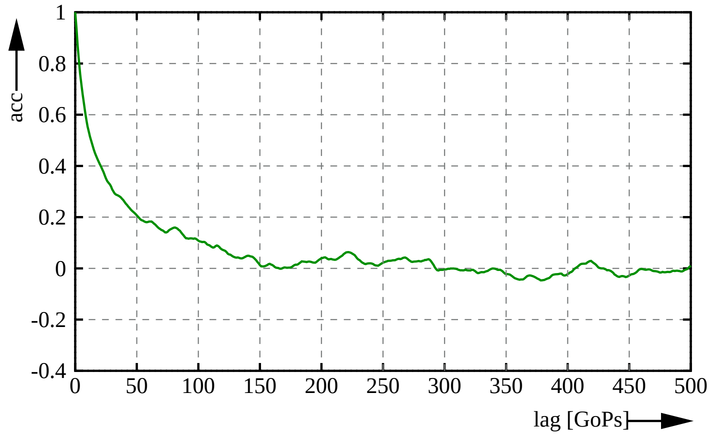


c) *Silence of the Lambs* with low quality

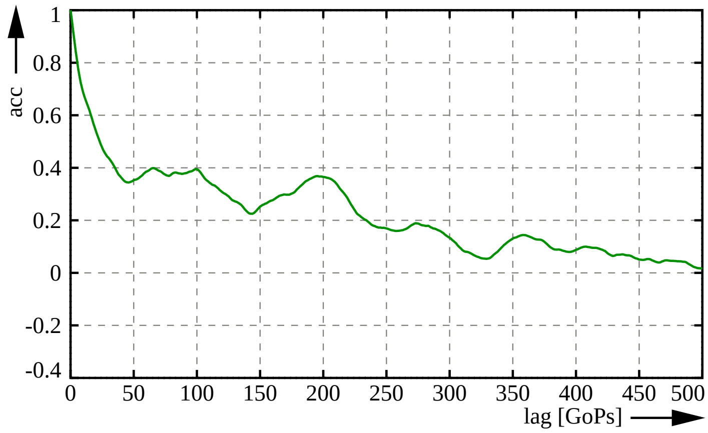
Figure 3: Autocorrelation of MPEG–4 frame size traces.



a) *Star Wars IV* with high quality

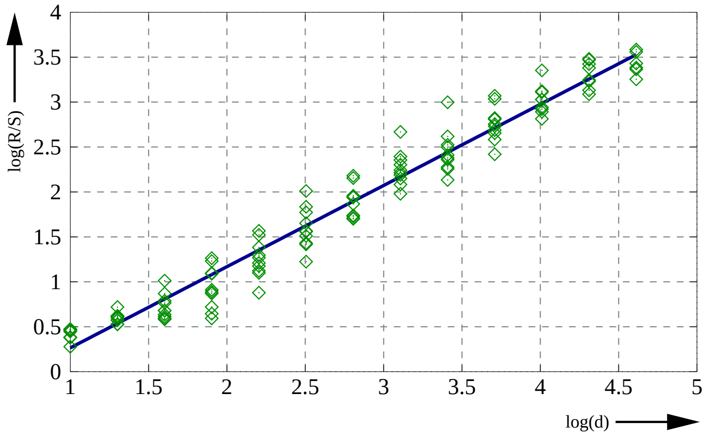


b) *Jurassic Park I* with medium quality

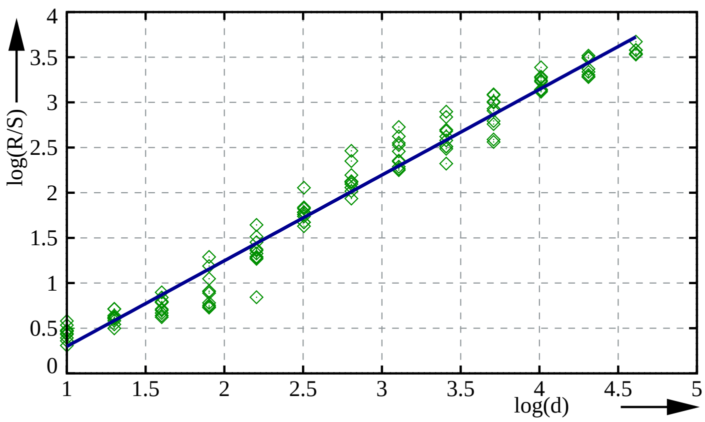


c) *Silence of the Lambs* with low quality

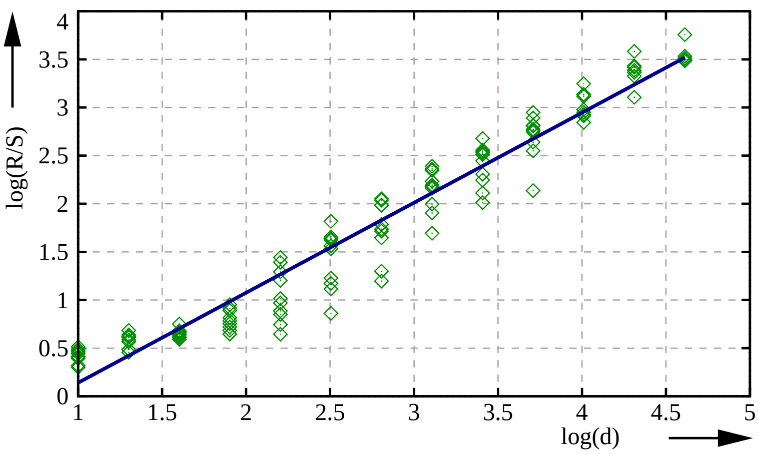
Figure 4: Autocorrelation of MPEG–4 GoP size traces.



a) *Star Wars IV* with high quality



b) *Jurassic Park I* with medium quality



c) *Silence of the Lambs* with low quality

Figure 5: Pox plots of R/S for MPEG–4 traces with aggregation level $a = 1$.

3 H.263 Video Traces

3.1 Overview of H.263 Video Compression

The basic structure of the H.263 video source coding algorithm [7, 14, 13] has been adopted from ITU–T Recommendation H.261 [6]. It uses (1) inter picture prediction to reduce the temporal redundancy and (2) Discrete Cosine Transform (DCT) coding of the residual prediction error to reduce the spatial redundancy. After the DCT coding, the prediction error is quantized and the resulting symbols are variable–length–coded and transmitted. For the interpicture prediction each video frame is divided into macro block and one motion vector is transmitted per macro block. In contrast to H.261, half pixel prediction is used for the motion vectors in H.263. The bit rate of the compressed video stream is controlled by adjusting several encoder parameters, such as quantizer scales and the frame rate. H.263 provides four advanced coding options. Unrestricted motion vectors, advanced prediction, and PB–frames are options that improve the inter–picture prediction. The fourth option is to use the more efficient arithmetic coding instead of variable–length–coding. These four options improve the video quality at the expense of increased video codec complexity. We refer the reader to [7, 14] for details. Roughly speaking, the unrestricted motion vector option allows motion vectors to point outside the video frame. The edge pels are used instead of prediction pels that lie outside the frame. This allows for more efficient compression, especially when there is motion near the frame border and the frame format is small. With advanced prediction the motion predicted blocks overlap and a pel is interpreted as the weighted average of the overlapping blocks. This reduces artifacts in the decoded video frames and increases the perceived video quality. The PB–frames option increases the frame rate without significantly increasing the bit rate. A PB–frame consists of two consecutive frames that are encoded as one entity. Specifically, a PB–frame consists of a P–frame, which is predicted from the preceding P–frame, and a B–frame, which is bi–directionally predicted from the preceding P–frame and the P–frame being part of the PB entity. When the reconstruction of the PB–frame is complete, the B–frame is displayed first and then the P–frame.

3.2 Our Encoding Approach for H.263

We encoded the uncompressed YUV information into an H.263 bit stream with the `tmn` encoder (Version 2.0) [11]. (We did not use the H.263 encoder of `bttvgrab` because it is not fully compliant with the H.263 standard; it inserts additional sequencing and synchronization information into the H.263 bit stream.) We emphasize that we did not encode in real–time; thus there was no encoder bottleneck. The parameters of the `tmn` encoder were set to the values given in Table 5. We did not enable unrestricted motion vectors, syntax–based arith-

Coding format	QCIF (176 x 144 pels)
Integer pel search window	15 pels
Quantization parameter QP	3
QP for first frame	3
Frames to skip between each encoded frame	0
Reference frame rate	25 frames/sec
Use of unrestricted motion vectors	OFF
Use of syntax–based arithmetic coding	OFF
Use of advanced prediction	OFF
Use of PB–frames	ON
BQUANT parameter	2

Table 5: Parameter settings of `tmn` H.263 encoder.

metic coding, and advanced prediction, since we observed that these features bring only little improvement in the video quality while slowing down the encoder dramatically. We did enable PB–frames. We encoded each video at four different target bit rates: (1) 16 kbit/sec, (2) 64 kbit/sec, (3) 256 kbit/sec, and (4) Variable Bit Rate (VBR), i.e., without setting a target bit rate.

3.3 Statistical Analysis of H.263 Traces

In this section we conduct a thorough statistical analysis of the generated H.263 frame size traces. Let N denote the number of frames in a given video trace. Let $X_n,\ n = 1, \ldots, N$, denote the number of bits in frame n, i.e., the frame size of frame n. Let $t_n,\ n = 1, \ldots, N$, denote the frame period (display time) of frame n in msec. Let $T_n,\ n = 1, \ldots, N$, denote the cumulative display time up to (and including) frame n, i.e., $T_n = \sum_{k=1}^{n} t_k$ (define $T_0 = 0$). The H.263 trace files available from our web site give on line $n,\ n = 1, \ldots, N$, the cumulative display time T_{n-1} (up to frame $n-1$), the type (I, P or PB) of frame n, and the frame size X_n in bytes. Table 6 gives the first ten lines of the trace file of the *Silence of the Lambs* video encoding with a target bit rate of 256 kbit/s. As illustrated by the trace file, the T_n's

0	I	12539
360	P	3981
600	PB	6203
760	PB	5884
1000	PB	6749
1160	PB	6425
1400	PB	7849
1640	PB	5983
1800	PB	6183
2040	PB	7052

Table 6: Excerpt of H.263 trace file of *Silence of the Lambs* encoding.

are integer multiples of the basic (reference) frame period $\Delta = 40$ msec of the H.263 encoder. Notice, however, that some frames are skipped by the encoder striving to meet the specified target bit rate [13, p. 67]. This results in variable frame periods. Figure 6 gives the probability mass functions $P(t_n = l \cdot \Delta),\ l = 1, 2, \ldots$, of the frame periods of three generated H.263 traces. (The other probability mass function plots are available from our web site.) We observe from the plots the general tendency that smaller target bit rates result in larger frame periods, i.e., more frames are skipped. On the other hand, for VBR encodings, i.e., without specified target bit rate, the H.263 encoder typically does not skip any frames. Nevertheless, as we observe from Figure 6 a), most encoded frames have a frame period of 2Δ. This is because the encoder produces mostly PB frames, i.e., two consecutive frames are encoded as one entity. If no frame is skipped, the PB frame has a frame period of 2Δ when emitted by the encoder; at the decoder, however, the B frame is displayed first for a period of Δ and then the P frame for a period of Δ.

Table 7 gives an overview of the statistics of the frame sizes X_n. First, we note that the H.263 encoder meets a given target for the average bit rate of the encoded video stream. To see this recall that the uncompressed YUV video stream has a bit rate of 38,016 byte/frame $\cdot$ 25 frames/sec = 950.4 kbit/sec. Also, recall that the compression ratio is defined as the ratio of the sum of the sizes of all unencoded YUV frames of the video to the sum of the sizes of all encoded frames emitted by the encoder. (Keep in mind that the H.263 encoder may (*i*) skip frames and (*ii*) encode two frames into one PB frame; therefore the number of encoded frames may be smaller than the number of unencoded YUV frames.) To achieve a given target for the average bit rate of the encoded video stream the encoder enforces the same compression ratio for all videos. Even though, for a given target rate all encoded videos have the same average bit rate, their average frame sizes are different. For the 16 kbps target rate, for instance, the *Soccer* encoding has an average size of 655 bytes, while the *Silence of the Lambs* encoding has an average frame size of 370 bytes. Nevertheless, the encoder meets the target bit rate by skipping more frames of the *Soccer* video; i.e., the average frame period of the *Soccer* encoding is larger.

Comparing the 256 kbps target rate encodings with the VBR encodings we observe that some VBR encodings have higher compression ratios that the corresponding 256 kbps target rate encodings. The VBR encoding of *Star Wars IV*, for instance, has a compression ratio of 65, while the 256 kbps encoding has a compression ratio of 29.7. The more efficient VBR encoding, however, has a larger variability of the frame sizes. It is important to note that for variable frame period H.263 encoded video, the frame sizes are only one component of the video stream statistics. For the complete picture we need to consider the frame sizes in conjunction with their associated frame periods. Clearly, if the larger frame sizes of the VBR H.263 encodings were associated with larger frame periods, and vice versa, then the larger frame periods could be used to smooth out the larger frames. We shall see shortly that this is to a limited extend possible.

Figure 7 gives the histograms of the frame size X_n for three traces. We observe that the 256 kbps target rate encoding of *Jurassic Park I* has a pronounced bi–modal distribution of the frame sizes. This is because the encoder typically produces (i) P frames with an average size of roughly 3 kbytes, and (ii) PB frames with an average size of 6 kbytes. Similar observations hold for the depicted frame size histogram of the 16 kbps target rate encoding of *Silence of the Lambs*, as well as the other encodings on our web site.

To facilitate the analysis of the H.263 video correlations and long range dependence characteristics, we define for a given H.263 frame size trace, two different traces that are derived from the frame sizes X_n and the frame periods t_n. First, we consider a "stuffed" frame size trace $F_m,\ m = 1, \ldots, T_N/\Delta$, obtained by "stuffing" zeros for the skipped frames into the generated frame size trace $X_n,\ n = 1, \ldots, N$. Formally,

$$F_m = \begin{cases} X_n & \text{for } m = \frac{T_n}{\Delta}, \quad n = 1, \ldots, N \\ 0 & \text{for } m \notin \{\frac{T_1}{\Delta}, \ldots, \frac{T_N}{\Delta}\}\ . \end{cases}$$

Secondly, we introduce the rate trace $r(t),\ 0 \leq t \leq T_N$. We convert the discrete frame size trace $X_n,\ n = 1, \ldots, N$, to a fluid flow by transmitting the frame of size X_n at the constant rate X_n/t_n over its frame period, i.e.,

$$r(t) = \frac{X_n}{t_n} \quad \text{for } T_{n-1} < t \leq T_n,\ n = 1, \ldots, N.$$

Note that $r(t)$ changes its value only at integer multiples of the reference frame period Δ. A more convenient representation of $r(t)$ is thus obtained by "sampling" at Δ–spaced intervals. We define the sampled rate trace as

$$R_m = r(m \cdot \Delta),\ m = 1, \ldots, T_N/\Delta.$$

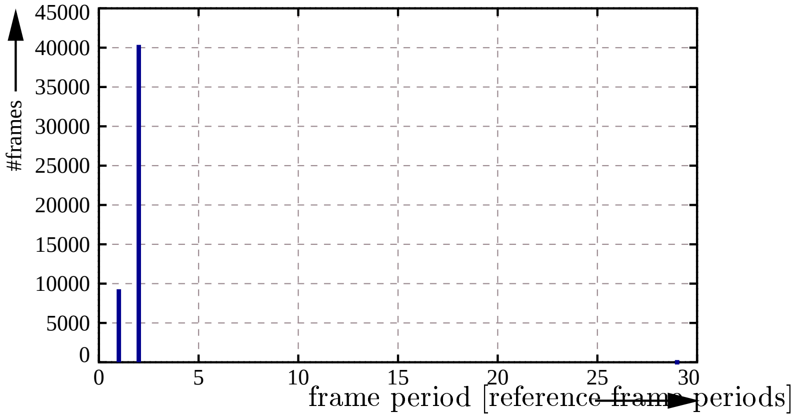
In the following we study the statistical properties of the frame size traces $F_m,\ m = 1, \ldots, T_N/\Delta$, and the rate traces $V_m,\ m = 1, \ldots, T_N/\Delta$, obtained from the generated H.263 frame size traces. Table 8 gives an overview of the statistics of the "stuffed" frame size traces and the sampled rate traces. First, we compare the frame size statistics from Table 7 with the sampled rate trace statistics in Table 8. We observe that for target rate encodings transmitting each encoded frame at a constant rate over its frame period significantly reduces the variability of the encoder output. This is because some extremely large frames are associated with large frame periods. Nevertheless, the peak–to–mean ratios of the rate traces with fixed target rates are typically five and larger. On the other hand, for VBR encodings the rate traces have larger variability than the frame sizes. We also observe from the statistics of the "stuffed" frame size traces that transmitting each frame at a constant rate over one reference period of length Δ gives extremely variable encoder output for small target rates.

As alluded to above, some VBR encodings have significantly smaller average bit rates than the 256 kbps target rate; see, for instance, the *Star Wars IV* encoding. The more efficient VBR encoding, however, entails more variability in the encoded video stream. Loosely speaking, with VBR encoding the encoder produces high output rates when they are needed to encode complex scenes without reducing the video quality. We note, however, that a detailed study of the video stream statistics in conjunction with the perceived video quality is beyond the scope of this technical report.

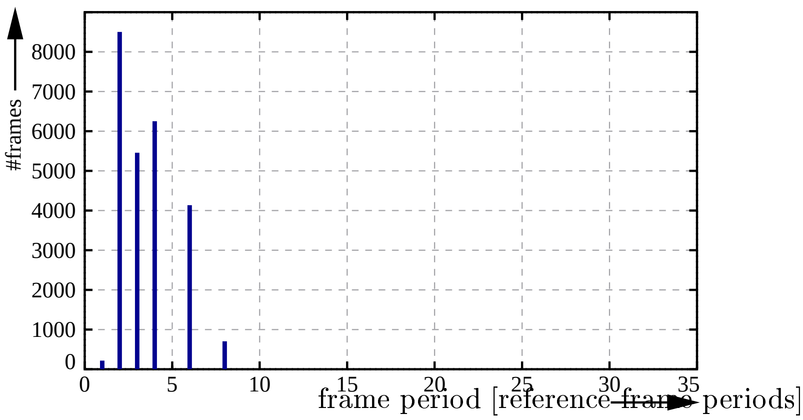
Figure 8 gives the "stuffed" frame size trace F_m as a function of the index m (in reference frame periods of length Δ) for the generated H.263 traces. (The other plots are available on our web site.) Figure 9 gives the autocorrelation coefficient of the "stuffed" frame size trace $\rho_F(k)$ and the autocorrelation coefficient of the sampled rate trace $\rho_R(k)$ as a function of the lag k (in reference frame periods of length Δ) for the generated H.263 traces over 14 reference frame periods of length Δ. We observe that the "stuffed" frame size traces have rather "jerky" autocorrelation functions. This is because the zeros in the "stuffed" traces give negative spikes. The sampled rate traces, on the other hand, have smooth autocorrelation functions. To get a better picture of the long term correlations we give in Figure 10 the autocorrelation functions over 500 reference frame periods. We observe that the autocorrelation function of the VBR encoding (i.e., without specified target rate) of *Star Wars IV* decays very slowly; for a lag of $d = 500\Delta$ the correlation coefficient of the sampled rate trace is roughly 0.2. The autocorrelations of the depicted target rate encodings, on the other hand, decay quickly to zero.

Table 9 gives the Hurst parameters of the sampled rate traces as a function of the aggregation level a. Figure 11 gives pox plots of R/S for an aggregation level of $a = 1$. We notice from the pox plots given here and on our web site that two problems arise when applying the R/S statistic to the H.263 traces. First, some pox plots for the aggregation level $a = 1$ have outliers for small lags d. One strategy could have been to remove those outliers; this would have given larger estimated for the Hurst parameter for the aggregation level $a = 1$. We chose not to do so in order to keep the least–squares fit estimation simple and automated. In interpreting the results in Table 9 we ignore the column $a = 1$ and focus on the larger aggregation levels instead. Secondly, the pox plots for aggregation levels of $a = 200$ and larger for encodings with a specified target bit rate, typically do not settle down around a straight "street". We suspect that this is due to the fact that the H.263 encoder typically skips many frames to meet a specified rate target. As a result these traces might not have a sufficiently large number of values to estimate the Hurst parameter for large aggregation levels.

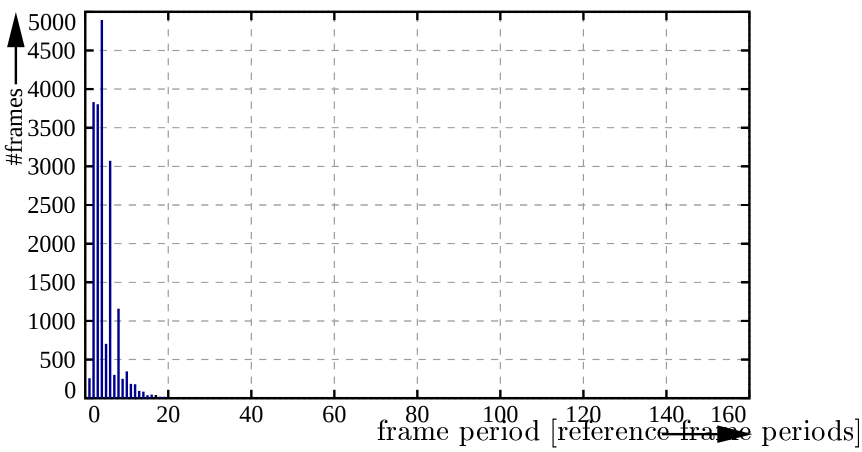
Nevertheless, we observe that all VBR encodings have Hurst parameters above 0.7 for all aggregation levels of $a = 12$ and higher. This indicates a high degree of long range dependence in the VBR traces. We also observe from the table that the encodings with a specified target bit rate have Hurst parameter above 0.7 for the aggregation levels $a = 12$, $a = 50$, and $a = 100$. This gives some indication of long range dependence properties. However, more studies on the long range dependence properties of H.263 traces are needed.



a) *Star Wars IV* without target rate



b) *Jurassic Park I* with target rate 256 kbps

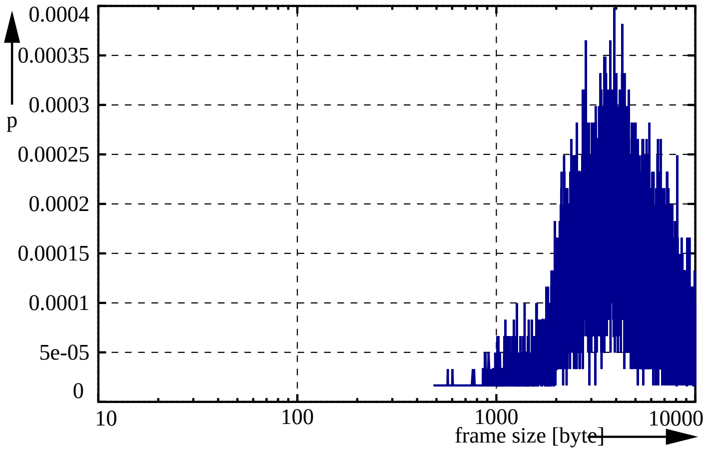


c) *Silence of the Lambs* with target rate 16 kbps

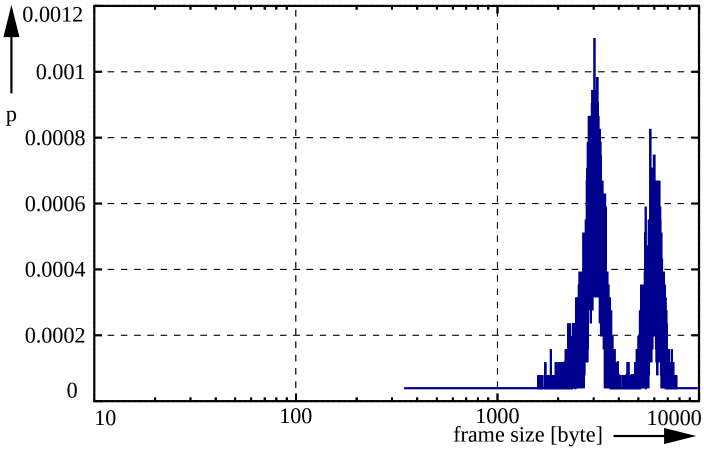
Figure 6: Probability mass functions of frame periods of H.263 traces.

Rate	Trace	Comp. ratio YUV:H.263	Mean $\bar{X}$ [byte]	CoV $S_X/\bar{X}$	Peak/Mean $X_{\max}/\bar{X}$
16 kbps	*Jurassic Park I*	476.36	476.36	0.67	20.83
	Silence of the Lambs	476.43	369.85	0.67	33.90
	Star Wars IV	476.43	326.02	0.61	11.32
	Mr. Bean	476.38	402.41	0.63	8.11
	First Contact	476.37	318.36	0.64	11.44
	From Dusk Till Dawn	476.30	448.91	0.68	13.24
	The Firm	476.40	336.47	0.52	11.89
	Formula 1	476.07	518.23	0.59	4.90
	Soccer	476.30	654.56	0.60	7.10
	ARD News	475.89	443.47	0.63	4.90
	ARD Talk	476.20	340.57	0.51	11.26
	N3 Talk	476.45	336.35	0.55	12.08
	Office–Cam	476.30	452.12	0.81	11.48
64 kbps	*Jurassic Park I*	118.96	1132.02	0.36	7.89
	Silence of the Lambs	118.95	1129.94	0.41	11.10
	Star Wars IV	118.95	1153.32	0.43	7.11
	Mr. Bean	118.96	1115.56	0.40	3.90
	First Contact	118.95	1099.94	0.46	5.33
	From Dusk Till Dawn	118.95	1056.67	0.42	5.62
	The Firm	118.96	1186.73	0.35	3.53
	Formula 1	118.92	862.65	0.36	5.60
	Soccer	118.94	905.35	0.37	5.13
	ARD News	118.83	1236.72	0.38	5.74
	ARD Talk	118.94	1268.00	0.35	3.38
	N3 Talk	118.96	1145.44	0.39	4.24
	Office–Cam	118.94	1564.69	0.12	3.32
256 kbps	*Jurassic Park I*	29.73	4533.67	0.35	2.61
	Silence of the Lambs	29.73	4453.81	0.39	5.00
	Star Wars IV	29.73	4563.53	0.33	3.58
	Mr. Bean	29.73	4424.96	0.37	2.25
	First Contact	29.73	4420.26	0.36	2.47
	From Dusk Till Dawn	29.73	4183.71	0.39	3.00
	The Firm	29.73	4690.26	0.33	1.94
	Formula 1	29.72	3431.64	0.34	3.25
	Soccer	29.73	3397.87	0.37	3.29
	ARD News	29.70	4878.85	0.36	2.13
	ARD Talk	29.73	5062.28	0.34	2.04
	N3 Talk	29.73	4546.89	0.38	3.87
	Office–Cam	29.73	6199.49	0.11	1.37
VBR	*Jurassic Park I*	17.08	3993.31	0.64	4.55
	Silence of the Lambs	25.19	2703.46	0.99	10.27
	Star Wars IV	65.79	1048.21	0.66	8.58
	Mr. Bean	25.00	2662.61	0.60	6.09
	First Contact	44.64	1512.82	0.78	7.68
	From Dusk Till Dawn	19.30	3378.48	0.58	4.81
	The Firm	57.17	1241.80	0.80	7.39
	Formula 1	14.25	3826.30	0.47	3.69
	Soccer	10.18	5583.62	0.50	4.05
	ARD News	20.17	3442.46	0.77	4.45
	ARD Talk	30.70	2374.17	0.55	5.59
	N3 Talk	27.96	2545.62	0.57	5.48
	Office–Cam	84.01	903.78	0.36	5.74

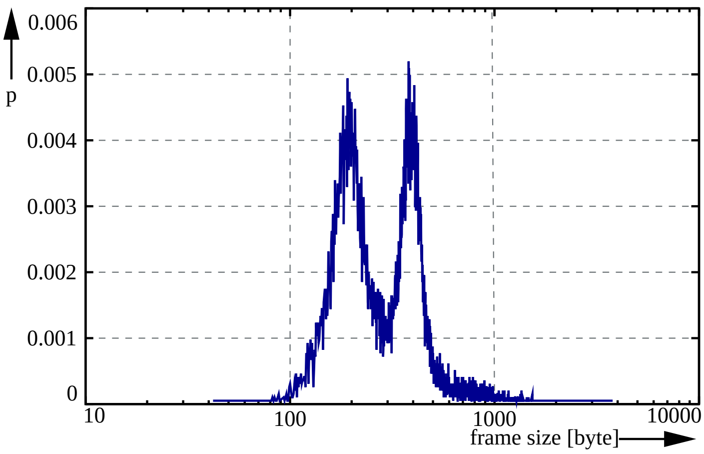
Table 7: Overview of frame size statistics of H.263 traces.



a) *Star Wars IV* without target rate



b) *Jurassic Park I* with target rate 256 kbps

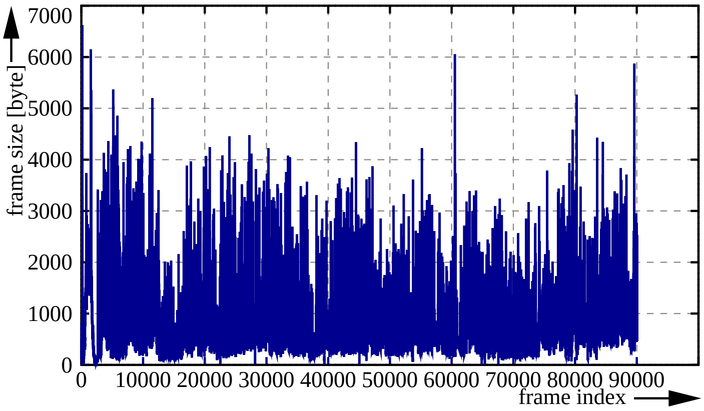


c) *Silence of the Lambs* with target rate 16 kbps

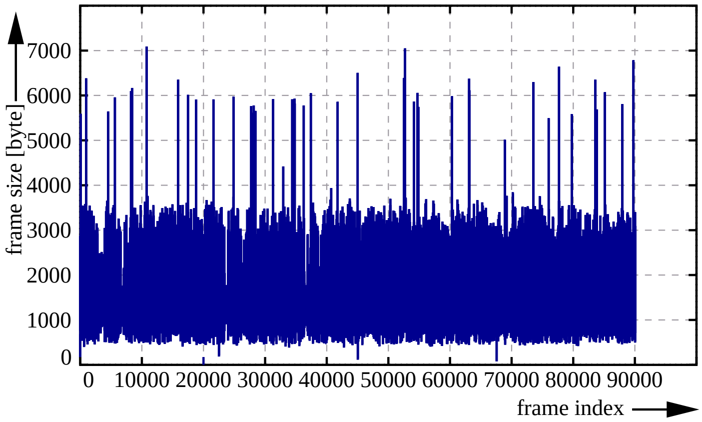
Figure 7: H.263 frame size histograms.

		"Stuffed" Frame Size Trace			Sampled Rate Trace		
		Mean	CoV	Peak/Mean	Mean	CoV	Peak/Mean
Rate	Trace	$\bar{F}$ [byte]	$S_F/\bar{F}$	$F_{\max}/\bar{F}$	$\bar{R}$ [kbit/s]	$S_R/\bar{R}$	$R_{\max}/\bar{R}$
16 kbps	*Jurassic Park I*	79.81	2.48	111.96	16	0.35	5.8
	Silence of the Lambs	79.80	2.38	157.14	16	0.37	6.3
	Star Wars IV	79.81	2.15	46.24	16	0.42	6.1
	Mr. Bean	79.80	2.46	40.91	16	0.34	6.1
	First Contact	79.81	2.15	45.62	16	0.4	5.8
	From Dusk Till Dawn	79.82	2.69	74.46	16	0.3	6.1
	The Firm	79.82	2.69	74.46	16	0.3	6.1
	Formula 1	79,86	2.79	60.50	16	0.21	4.9
	Soccer	79.82	3.18	58.22	16	0.17	4.6
	ARD News	79.88	2.60	58.02	16	0.37	4.9
	ARD Talk	79.83	2.09	48.02	16	0.41	5.7
	N3 Talk	79.79	2.12	50.92	16	0.38	5.8
	Office–Cam	79.82	2.90	65.04	16	0.44	5.3
64 kbps	*Jurassic Park I*	319.59	1.73	27.96	64	0.42	5.7
	Silence of the Lambs	319.60	1.77	39.23	64	0.42	5.7
	Star Wars IV	319.62	1.81	25.66	64	0.45	5.2
	Mr. Bean	319.58	1.74	13.61	64	0.4	8.1
	First Contact	319.62	1.78	18.35	64	0.43	5.8
	From Dusk Till Dawn	319.62	1.70	18.59	64	0.39	6.3
	The Firm	319.59	1.78	13.10	64	0.45	5.4
	Formula 1	319.69	1.43	15.11	64	0.32	5.7
	Soccer	319.63	1.49	14.54	64	0.33	6.0
	ARD News	319.91	1.85	22.21	64	0.42	5.3
	ARD Talk	319.65	1.86	13.41	64	0.43	5.7
	N3 Talk	319.58	1.77	15.20	64	0.41	5.4
	Office–Cam	319.63	1.99	16.24	64	0.44	4.9
256 kbps	*Jurassic Park I*	1278.72	1.72	9.24	256	0.42	5.5
	Silence of the Lambs	1278.61	1.73	17.4	256	0.42	5.9
	Star Wars IV	1278.80	1.72	12.76	256	0.47	5.3
	Mr. Bean	1278.78	1.71	7.79	256	0.40	5.7
	First Contact	1278.80	1.71	8.54	256	0.44	5.8
	From Dusk Till Dawn	1278.79	1.66	9.83	256	0.39	5.9
	The Firm	1278.73	1.75	7.13	256	0.46	5.2
	Formula 1	1279.13	1.41	8.71	256	0.32	5.5
	Soccer	1278.76	1.42	8.73	256	0.35	5.7
	ARD News	1279.99	1.82	8.12	256	0.43	5.6
	ARD Talk	1278.89	1.85	8.08	256	0.43	5.8
	N3 Talk	1278.70	1.75	13.74	256	0.41	5.5
	Office–Cam	1278.88	1.98	6.66	256	0.41	5.5
VBR	*Jurassic Park I*	2225.26	1.23	8.16	450	0.69	7.7
	Silence of the Lambs	1509.14	1.60	18.41	300	1.10	17.0
	Star Wars IV	589.63	1.24	15.25	120	0.76	11.0
	Mr. Bean	1520.68	1.17	10.67	300	0.70	11.0
	First Contact	851.63	1.36	13.65	170	0.84	13.0
	From Dusk Till Dawn	1969.69	1.13	8.25	390	0.68	8.2
	The Firm	665.00	1.44	13.79	130	0.89	13.0
	Formula 1	2667.84	0.86	5.29	530	0.47	5.3
	Soccer	3733.26	0.93	6.06	750	0.53	6.1
	ARD News	1884.48	1.38	8.12	380	0.88	7.8
	ARD Talk	1238.43	1.22	10.72	250	0.61	9.4
	N3 Talk	1359.79	1.22	10.26	270	0.62	9.0
	Office–Cam	452.52	1.12	11.47	91	0.38	11.0

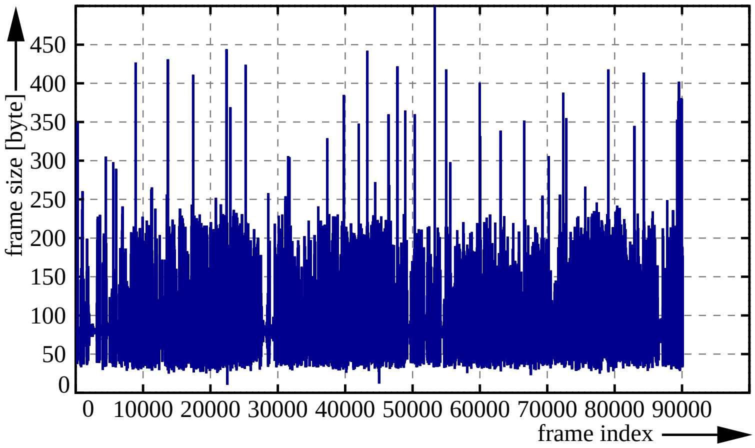
Table 8: Overview of statistics of H.263 "stuffed" frame size traces and sampled rate traces.



a) *Star Wars IV* without target rate

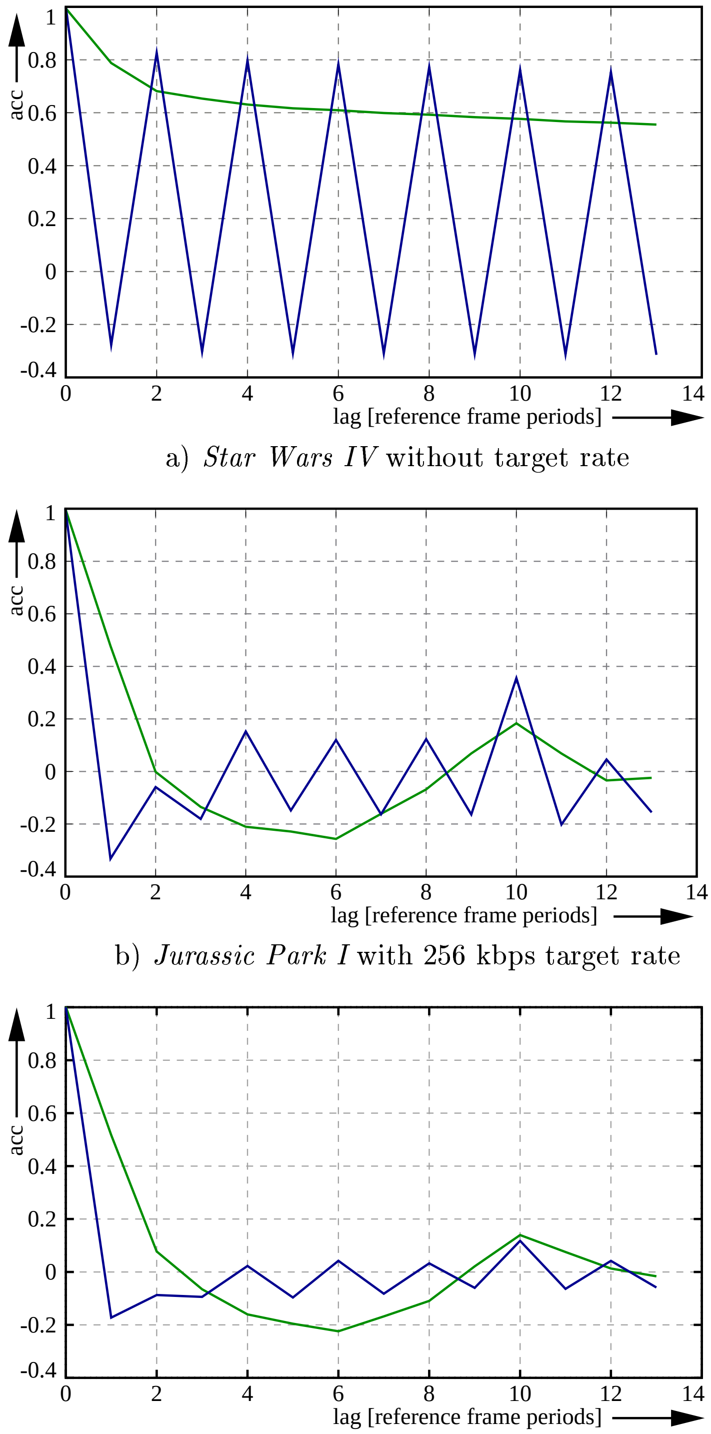


b) *Jurassic Park I* with 256 kbps target rate



c) *Silence of the Lambs* with 16 kbps target rate

Figure 8: H.263 "stuffed" frame size traces.

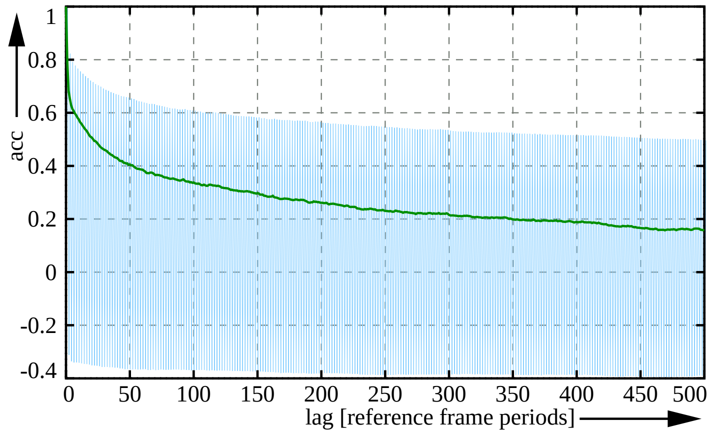


a) *Star Wars IV* without target rate

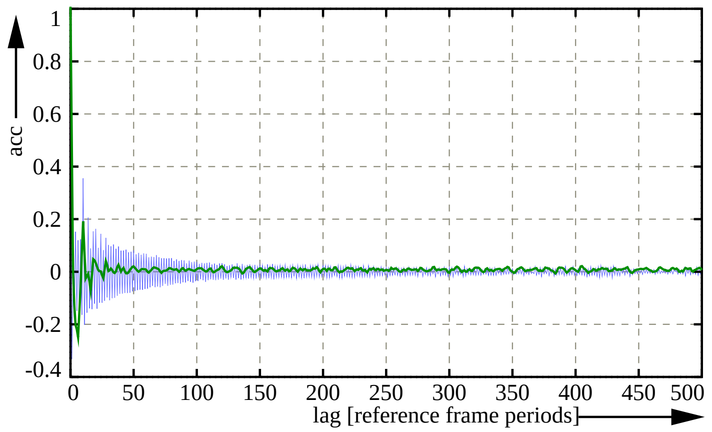
b) *Jurassic Park I* with 256 kbps target rate

c) *Silence of the Lambs* with 16 kbps target rate

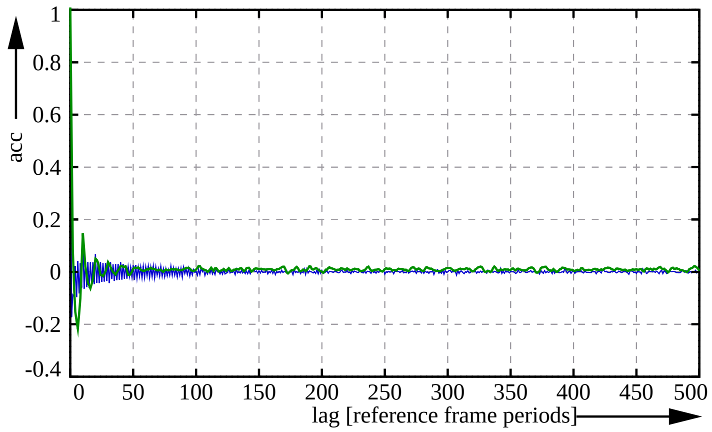
Figure 9: Autocorrelation of H.263 "stuffed" frame size traces and sampled rate traces over 14 reference frame periods.



a) *Star Wars IV* without target rate



b) *Jurassic Park I* with 256 kbps target rate

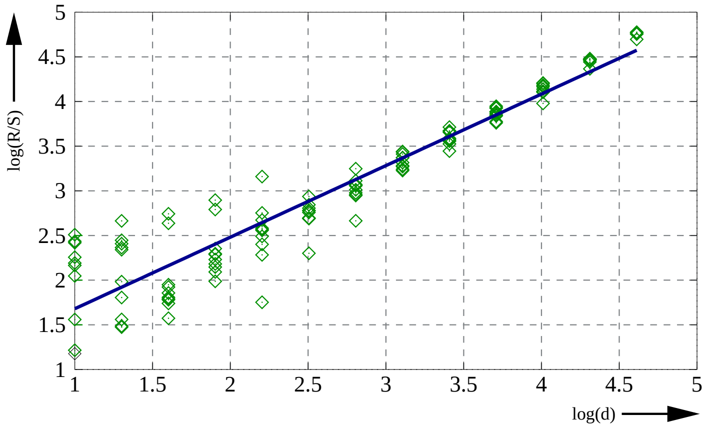


c) *Silence of the Lambs* with 16 kbps target rate

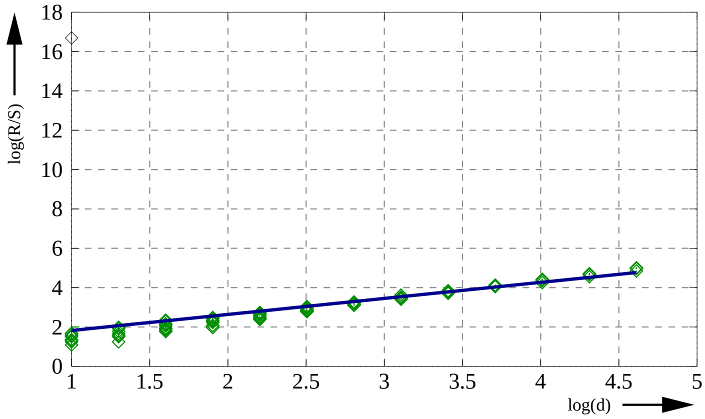
Figure 10: Autocorrelation of H.263 "stuffed" frame size traces and sampled rate traces over 500 reference frame periods.

Rate	Trace	Aggregation level a [reference frame periods Δ] 1	12	50	100	200	300	400	500	600	700	800
16 kbps	*Jurassic Park I*	0.945	0.930	0.843	0.770	0.630	0.409	0.581	0.413	0.083	0.352	0.503
	Silence o/t Lambs	0.262	0.657	0.835	0.721	0.594	0.405	0.488	0.420	0.091	0.215	0.361
	Star Wars IV	0.947	0.944	0.872	0.762	0.621	0.380	0.534	0.391	0.157	0.384	0.511
	Mr. Bean	0.859	0.897	0.783	0.722	0.571	0.439	0.446	0.410	0.201	0.296	0.520
	First Contact	0.313	0.952	0.864	0.771	0.644	0.449	0.484	0.438	0.158	0.216	0.463
	Dusk Till Dawn	0.718	0.861	0.791	0.610	0.497	0.277	0.460	0.387	0.097	0.190	0.480
	The Firm	0.465	0.920	0.851	0.771	0.615	0.376	0.466	0.423	0.119	0.307	0.470
	Formula 1	0.507	0.837	0.699	0.595	0.368	0.078	0.437	-0.073	-0.381	-0.121	0.492
	Soccer	-0.114	0.742	0.676	0.550	0.350	0.187	0.272	0.105	-0.121	-0.047	0.149
	ARD News	0.496	0.870	0.834	0.748	0.768	0.331	1.065	-0.297	-0.126	0.545	-0.086
	ARD Talk	0.493	0.923	0.853	0.758	0.616	0.532	0.411	0.212	0.487	0.456	0.339
	N3 Talk	0.612	0.929	0.850	0.768	0.585	0.342	0.457	0.383	0.161	0.274	0.454
	Office–Cam	0.474	0.968	0.843	0.724	0.611	0.470	0.413	0.149	0.492	0.378	0.220
64 kbps	*Jurassic Park I*	1.005	0.963	0.832	0.748	0.608	0.365	0.503	0.445	0.209	0.323	0.527
	Silence o/t Lambs	0.459	0.985	0.886	0.790	0.608	0.400	0.465	0.383	0.101	0.272	0.467
	Star Wars IV	0.963	0.952	0.903	0.802	0.620	0.348	0.509	0.400	0.150	0.388	0.465
	Mr. Bean	0.960	0.926	0.812	0.728	0.589	0.385	0.404	0.420	0.111	0.275	0.436
	First Contact	0.646	0.979	0.872	0.730	0.603	0.397	0.511	0.418	0.180	0.295	0.513
	Dusk Till Dawn	0.805	0.904	0.831	0.682	0.540	0.364	0.501	0.409	0.101	0.261	0.572
	The Firm	0.933	0.925	0.848	0.758	0.642	0.386	0.485	0.424	0.157	0.322	0.463
	Formula 1	0.772	0.910	0.772	0.616	0.403	0.150	0.534	0.056	-0.107	0.054	0.571
	Soccer	0.954	0.889	0.863	0.752	0.579	0.387	0.415	0.347	0.070	0.200	0.433
	ARD News	0.713	0.866	0.867	0.752	0.800	0.361	1.066	-0.318	-0.157	0.544	-0.047
	ARD Talk	0.955	0.921	0.819	0.691	0.593	0.475	0.449	0.170	0.435	0.428	0.333
	N3 Talk	0.761	0.940	0.864	0.765	0.630	0.368	0.499	0.471	0.116	0.306	0.478
	Office–Cam	0.786	0.985	0.743	0.563	0.396	0.262	0.224	-0.184	0.186	-0.002	0.143
256 kbps	*Jurassic Park I*	0.815	0.960	0.883	0.770	0.575	0.353	0.449	0.360	0.129	0.318	0.497
	Silence o/t Lambs	0.961	0.950	0.867	0.757	0.598	0.407	0.509	0.422	0.112	0.226	0.435
	Star Wars IV	0.777	0.960	0.872	0.743	0.585	0.326	0.495	0.418	0.111	0.328	0.497
	Mr. Bean	0.994	0.883	0.853	0.733	0.611	0.380	0.465	0.417	0.169	0.312	0.469
	First Contact	0.798	0.970	0.871	0.723	0.556	0.361	0.495	0.385	0.122	0.205	0.537
	Dusk Till Dawn	0.797	0.936	0.830	0.703	0.561	0.350	0.420	0.405	0.102	0.243	0.498
	The Firm	0.781	0.926	0.855	0.740	0.605	0.352	0.451	0.382	0.111	0.226	0.453
	Formula 1	0.767	0.907	0.791	0.643	0.414	0.179	0.426	0.003	-0.182	0.110	0.626
	Soccer	0.801	0.919	0.832	0.708	0.540	0.332	0.449	0.370	0.075	0.180	0.445
	ARD News	0.964	0.932	0.880	0.739	0.787	0.392	1.037	-0.356	-0.173	0.485	-0.129
	ARD Talk	0.913	0.973	0.806	0.670	0.595	0.523	0.390	0.154	0.462	0.340	0.354
	N3 Talk	0.970	0.953	0.863	0.755	0.589	0.372	0.468	0.415	0.132	0.272	0.439
	Office–Cam	0.926	0.895	0.716	0.579	0.400	0.274	0.194	-0.138	0.205	0.153	0.154
VBR	*Jurassic Park I*	0.863	0.862	0.902	0.900	0.867	0.816	0.770	0.785	0.805	0.817	0.867
	Silence o/t Lambs	0.575	0.758	0.778	0.815	0.793	0.730	0.753	0.803	0.727	0.738	0.755
	Star Wars IV	0.633	0.893	0.882	0.882	0.882	0.845	0.859	0.864	0.856	0.876	0.865
	Mr. Bean	0.860	0.811	0.803	0.816	0.836	0.842	0.816	0.802	0.753	0.765	0.748
	First Contact	0.650	0.855	0.874	0.870	0.871	0.877	0.888	0.909	0.896	0.944	0.875
	Dusk Till Dawn	0.671	0.852	0.841	0.867	0.915	0.899	0.915	0.877	0.875	0.893	0.891
	The Firm	0.803	0.812	0.854	0.847	0.878	0.831	0.849	0.848	0.844	0.879	0.820
	Formula 1	0.663	0.877	0.840	0.876	0.897	0.834	0.899	0.882	0.856	0.980	1.036
	Soccer	0.726	0.853	0.935	0.925	0.910	0.911	0.934	0.912	0.882	0.897	0.925
	ARD News	0.675	0.702	0.805	0.846	0.914	0.887	1.010	0.876	0.942	0.931	0.931
	ARD Talk	0.791	0.823	0.877	0.908	0.861	0.801	0.813	0.870	0.859	0.974	0.923
	N3 Talk	0.642	0.894	0.821	0.776	0.714	0.705	0.710	0.692	0.619	0.674	0.687
	Office–Cam	0.642	0.894	0.821	0.776	0.714	0.705	0.710	0.692	0.619	0.674	0.687

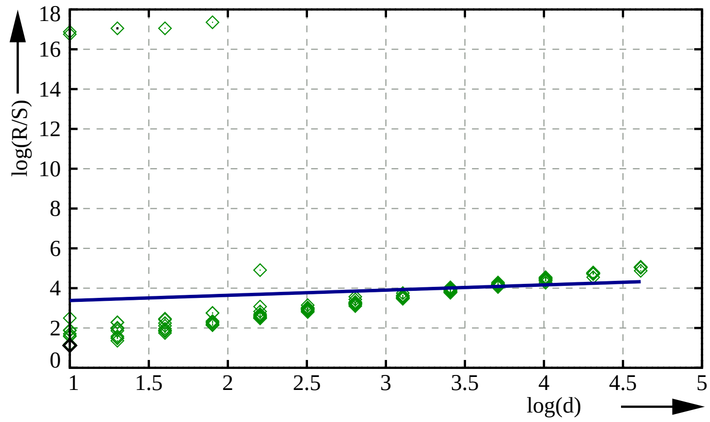
Table 9: Hurst parameters of H.263 sampled rate traces estimated from pox diagram of R/S as a function of the aggregation level a.



a) *Star Wars IV* without target rate



b) *Jurassic Park I* with 256 kbps target rate



c) *Silence of the Lambs* with 16 kbps target rate

Figure 11: Pox plots of R/S for H.263 "stuffed" frame size traces and sampled rate traces with aggregation level $a = 1$.

4 Conclusion

In this technical report we have presented a publicly available library of frame size traces of MPEG–4 and H.263 encoded videos. We have encoded over ten videos of 60 minutes length each. For each video we have generated MPEG–4 and H.263 encodings at several different quality levels. All in all, there are over 70 hours worth of video traces available on our web site: `http://www-tkn.ee.tu-berlin.de/~fitzek/TRACE/trace.html`.

We have conducted a detailed statistical analysis of the generated traces. For the analysis of the H.263 encodings, which have variable frame periods, we have introduced the notion of a *rate trace*. The rate trace facilitates the analysis of the H.263 frame sizes in conjunction with their associated frame periods. We have found that the traces are typically highly variable in their frame sizes and bit rates. Also, many of the traces show clear indications of long range dependence properties.

In our ongoing work we are expanding the available trace library by producing MPEG–4 and H.263 encodings of more videos. We are also generating MPEG–4 encodings with multiple Video Objects and multiple Video Object Layers. Moreover, we are encoding videos using the H.263+ encoder, which incorporates advanced motion prediction and enhanced PB frames.

Acknowledgment: We are grateful to Prof. Adam Wolisz for providing the environment that allowed us to pursue the work presented in this technical report. We are grateful to Moncef Abda Ben, Thomas Kroener, Mohammad Kandil, and Ahmed Salih for assisting in the recording and encoding of the videos during the student project at TKN in summer 2000. We gratefully acknowledge insightful discussions, borrowed video cassettes, and encouragement from Jean–Pierre Ebert, Enno Ewers, Andreas Festag, Andreas Koepsel, Rolf Morich, and Stephan Rein. We are grateful to Guido Heising for explaining the intricacies of the MOMUSYS MPEG–4 software.

Appendix

In this appendix we review the statistical definitions and methods used in the analysis of the generated frame size traces. Recall that N denotes the number of frames in a given trace. Also recall that $X_n,\ n = 1, \ldots, N$, denotes the size of frame n in bit.

Mean, Coefficient of Variation, and Autocorrelation

The (arithmetic) sample mean $\bar{X}$ of a frame size trace is estimated as

$$\bar{X} = \frac{1}{N} \sum_{n=1}^{N} X_n.$$

The sample variance S_X^2 of a frame size trace is estimated as

$$S_X^2 = \frac{1}{N-1} \sum_{n=1}^{N} (X_n - \bar{X})^2.$$

A computationally more convenient expression for S_X^2 is

$$S_X^2 = \frac{1}{N-1} \left[\sum_{n=1}^{N} X_n^2 - \frac{1}{N} \left(\sum_{n=1}^{N} X_n \right)^2 \right].$$

The coefficient of variation CoV is defined as

$$CoV = \frac{S_X}{\bar{X}}.$$

The maximum frame size $X_{\max}$ is defined as

$$X_{\max} = \max_{1 \leq n \leq N} X_n.$$

The autocorrelation coefficient $\rho_X(k)$ for lag $k,\ k = 0, 1, \ldots, N$, is estimated as

$$\rho_X(k) = \frac{1}{N-k} \sum_{n=1}^{N-k} \frac{(X_n - \bar{X})(X_{n+k} - \bar{X})}{S_X^2}.$$

For a given GoP trace $Y_m,\ m = 1, \ldots, N/G$, the sample variance S_Y^2, the coefficient of variation, and the autocorrelation coefficients $\rho_Y(k)$ are estimated in analogous fashion. We refer the reader to Law and Kelton [10] for more details on these definitions.

1.	**For** $d = 10,\ 20,\ 40,\ 80, \ldots$ **do**
2.	$I = K + 1 - \lceil \frac{dK}{N} \rceil$
3.	**For** $i = 1, \ldots, I$ **do**
4.	$t_i = (i-1)\frac{N}{K} + 1$
5.	$\bar{X}(t_i, d) = \frac{1}{d}\sum_{j=1}^{d} X_{t_i+j}$
6.	$S^2(t_i, d) = \frac{1}{d-1}\sum_{j=1}^{d}[X_{t_i+j} - \bar{X}(t_i,d)]^2$
7.	$R(t_i, d) = \max\{0,\ \max_{1 \le k \le d} W(t_i,k)\} - \min\{0,\ \min_{1 \le k \le d} W(t_i,k)\}$
8.	$W(t_i,k) = \left(\sum_{j=1}^{k} X_{t_i+j}\right) - k\bar{X}(t_i,d)$
9.	plot point $\left(\log d,\ \log \frac{R(t_i,d)}{S(t_i,d)}\right)$

Table 10: Algorithm for pox diagram of R/S.

R/S Statistic

We use the R/S statistic [12, 1, 2] to investigate the long range dependence characteristics of the generated frame size traces. The R/S statistic provides an heuristic graphical approach for estimating the Hurst parameter H. Roughly speaking, for long range dependent stochastic processes the R/S statistic is characterized by $E[R(n)/S(n)] \sim cn^H$ as $n \to \infty$ (where c is some positive finite constant). The Hurst parameter H is estimated as the slope of a log–log plot of the R/S statistic.

More formally, the *rescaled adjusted range statistic* (for short *R/S statistic*) is plotted according to the algorithm given in Table 10. The R/S statistic $R(t_i,d)/S(t_i,d)$ is computed for logarithmically spaced values of the lag d, starting with $d = 10$. For each lag value d as many as K samples of R/S are computed by considering different starting points t_i; we set $K = 10$ in our analysis. The starting points must satisfy $(t_i - 1) + d \le N$, hence the actual number of samples I is less than d for large lags d. Plotting $\log[R(t_i,d)/S(t_i,d)]$ as a function of $\log d$

gives the *rescaled adjusted range plot* (also referred to as *pox diagram of R/S*). A typical pox diagram starts with a transient zone representing the short range dependence characteristics of the trace. The plot then settles down and fluctuates around a straight "street" of slope H. If the plot exhibits this asymptotic behavior, the *asymptotic Hurst exponent* H is estimated from the street's slope using a least squares fit.

To verify the robustness of the estimate we repeat this procedure for each trace for different aggregation levels $a \geq 1$. The aggregated trace $X_n^{(a)},\ n = 1, \ldots, N/a$, is obtained from the original trace $X_n,\ n = 1, \ldots, N$, by averaging over non–overlapping blocks of length a, i.e.,

$$X_n^{(a)} = \frac{1}{a} \sum_{j=(n-1)a+1}^{na} X_j.$$

References

[1] J. Beran. *Statistics for long–memory processes.* Chapman and Hall, 1994.

[2] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger. Long–range dependence in variable–bit–rate video traffic. *IEEE Transactions on Communications*, 43(2/3/4):1566–1579, February/March/April 1995.

[3] R. Koenen (Editor). Overview of the MPEG–4 standard, ISO/IEC 14496, May/June 2000.

[4] M. W. Garret. *Contributions toward Real-Time Services on Packet Networks.* PhD thesis, Columbia University, May 1993. ftp address and directory of the used video trace: bellcore.com /pub/vbr.video.trace/.

[5] G. Heising and M. Wollborn. MPEG–4 version 2 video reference software package, ACTS AC098 mobile multimedia systems (MOMUSYS), December 1999.

[6] ITU-T. Recommendation H.261, video codec for audio–visual services at 64 – 1920 kbit/s, 1993.

[7] ITU-T/SG15. Recommendation H.263, video coding for low bitrate communication, 1996.

[8] R. Koenen. MPEG–4 multimedia for our time. *IEEE Spectrum*, 36(2):26–33, February 1999.

[9] M. Krunz, R. Sass, and H. Hughes. Statistical characteristics and multiplexing of MPEG streams. In *Proceedings of IEEE Infocom '95*, pages 455–462, April 1995.

[10] A. M. Law and W. D. Kelton. *Simulation, Modeling, and Analysis*. McGraw–Hill, second edition, 1991.

[11] K. O. Lillevold. tmn, version 2.0, June 1996.

[12] B. B. Mandelbrot and M. S. Taqqu. Robust R/S analysis of long–run serial correlations. In *Proceedings of 42nd Session ISI, Vol. XLVIII, Book 2*, pages 69–99, 1979.

[13] A. Puri and T. Chen. *Multimedia Systems, Standards, and Networks*. Marcel Dekker, New York, 2000.

[14] K. Rijkse. H.263: Video coding for low–bit–rate communication. *IEEE Communications Magazine*, 34(12):42–45, December 1996.

[15] O. Rose. Statistical properties of MPEG video traffic and their impact on traffic modelling in ATM systems. Technical Report 101, University of Wuerzburg, Institute of Computer Science, Am Hubland, 97074 Wuerzburg, Germany, February 1995.

[16] L. D. Soares and F. Pereira. MPEG–4: A flexible coding standard for the emerging mobile multimedia applications. Technical report, MOMUSYS Project, 1999.

[17] J. Walter. bttvgrab. `http://www.garni.ch/bttvgrab/`.