Trace-Based Analysis of MPEG-4 AVC Video Streams

Panagiotis Kizouris[1], Andreas Papadakis[2]

[1]Department of Electrical and Electronics Engineering, University of West Attica, Athens, Greece
[2]Department of Electrical and Electronics Engineering Educators, School of Pedagogical and Technological Education (ASPETE), Athens, Greece
Email: p.kizouris@gmail.com, apapadakis@aspete.gr

Abstract

MPEG-4 AVC encoded video streams have been analyzed using video traces and statistical features have been extracted, in the context of supporting efficient deployment of networked and multimedia services. The statistical features include the number of scenes composing the video and the sizes of different types of frames, within the overall trace and each scene. Statistical processing has been performed upon the traces and subsequent fitting upon statistical distributions (Pareto and lognormal). Through the construction of a synthetic trace, based upon this analysis, our selections of statistical distribution have been verified. In addition, different types of content, in terms of level of activity (quantified as different scene change ratio) have been considered. Through modelling and fitting, the stability of the main statistical parameters has been verified as well as observations on the dependence of these parameters upon the video activity level.

Keywords

Synthetic Traffic, Video Trace, MPEG-4 AVC, I and B Frames

1. Introduction

Multimedia services gradually become more complex and interactive, and they are increasingly based upon the retrieval, transfer and delivery of video. The characteristics of video affect the traffic generated, the network performance and the Quality of Service (QoS) of the services deployed.

Effective provision of multimedia and networked services requires prior knowledge of the content that will be served, the network capabilities upon which it will be delivered, as well as the users' behavior. Video compression me-

chanisms, on the other hand, reduce the bandwidth required for multimedia delivery and the compression ratio can affect the perceived quality by the end user. The bit rate, resulting from the encoding, depends on the selection of encoding parameters and the type of the video itself; so different requirements (in terms of rate and its volatility) are typically expected. Studies have already examined the properties of H.264.SVC single layer video sources, including [1] and [2] in order to provide models and synthetic traffic.

Efficient and homogenized deployment of multimedia services over various networks calls for *complex network dimensioning exercises*, considering dynamically changing properties. A typical example is the IPTV service delivery where multiple users are concurrently involved in the retrieval of video streams. Innovative schemes have been investigated, including caching and reusability of the content [3], making service delivery more efficient. In parallel to that and considering the typically restricted and changing underlying network resources, statistical knowledge of the content properties to be served is necessary.

Simulating the performance of such video provision services can be achieved either with actual traces or with algorithmically generated traces. The first option uses a wide variety of video traces files and there is need of an IO file mechanism which makes the evaluation mechanism slower and less efficient. The second option uses algorithmically generated traces based on the statistical properties of the actual traces. The properties of such traces can support service providers to dimension the bandwidth requirements of their services and facilitate smooth cooperation with the underlying network.

In our current work we investigate the dynamically-changing rate of encoded content focusing on the extraction of statistical properties of the videos through their corresponding video traces. Considering the volume of different types of frames (I and B) and available statistical distributions we endeavor to model the volume of the encoded content and produce synthetic traces.

The field has already met the interest of the scientific community in [4] and [5]. Our work is contributing to 1) the consolidated design of different video trace processing aspects including the statistical parameter extraction, the model creation and the scene separation, and 2) the streamlining and verification of the methodology using tools that extensively deployed in research laboratories (Matlab). Furthermore, 3) we have verified that the selected distributions follow the actual data, while 4) repeating the steps of parameter extraction in a series of video of different activity levels, observations have been produced regarding the dependence of the statistical parameters upon the video activity level.

The structure of the paper is the following: In Section 2 we provide a brief state of the art analysis on video encoding and video trace extraction and processing; the methodological aspects and sources informing our research work are also briefly presented. Section 3 presents the mechanisms and algorithms that have been designed and developed, focusing on scene separation, estimation of the sizes of I and B frames and statistical distribution fitting. The results of

this analysis are presented considering an indicative video trace. Section 4 considers different types of videos, in terms of activity level, and repeats the analysis described in the previous section for these types (which are encoded with same parameters) in order to investigate two key parameters.

2. Brief State of the Art Analysis and Methodology

2.1. Video Encoding Aspects

MPEG4/H264 Advanced Video Coding (H.264/AVC) and Scalable Video Coding (H.264/SVC) provide an efficient way of compressing and encoding videos, with widespread acceptance; AVC (Advanced Video Coding) is used in Blu-ray discs and digital television (DVB-T, DVB-T2). Encoding configuration is related to field coding, the macroblock-adaptive switching between frame and field coding (MBAFF), the weighted prediction, CABAC, and the SI and SP slices. Different approaches involve the computational complexity and the error robustness (from the baseline to the main, the extended and the high profiles).

The videos and traces we consider are based upon H.264/AVC. The main point which differentiates the H.264/AVC from the SVC extension is the ability of the SVC to support different levels of encodings able to determine different video qualities. The H.264/AVC is also referred to as H.264 SVC single layer.

Studies examine the statistical properties of H.264/AVC and SVC streams and models in [6], [7] and [8]. Correlation of quality aspects with encoding parameters and network parameters have also been investigated for various types of videos [9] [10]. In principle, while the encoded streams differ in the resulting volume, common statistical models can be followed for a condensed and flexible form of stream representation.

2.2. Synthetic Video Traces

The idea of using synthetic video traces for the evaluation of service provision has been explored in the past. A typical example is presented at [11], where network performance forecasting for IPTV streaming is investigated using video traces. For the evaluation multiple video traces have been generated corresponding to different usage scenarios.

Similarly in [12] evaluation of access network performance for IPTV streaming has been explored using video traces. Such IPTV characteristics include packet loss, packet delay and buffer size. Based upon such evaluations, the IPTV service provider can create different scenarios with types of video provided to the users at the same time, taking decision related to unicasting and multicasting. Content management becomes more efficient, through simulations of multiple bandwidth demanding programs using synthetic traces.

In [13] a collection of video traces for network performance evaluation is being provided as well as tools that support the generation of such traces. Traces are provided from videos encoded with MPEG-4 Part 2, H.261, H.263, H.264/AVC, and H.264/SVC. This is the source of traces that we consider for

our experimentation. The traces follow a simple and homogenized format, including the following fields: frame number (sequence number), the time of frame appearance (in seconds), and frame type (I, P or B) and frame size (in Bytes). Other aspects may include quality-related parameters of the frames such as the PSNR (Pease Signal to Noise Ratio).

2.3. Methodology

As encoding mechanism, we consider MPEG-4 part 10, H.264/AVC, due to its widespread acceptance (in IPTV and terrestrial satellite broadcasting delivery schemes) and its capability of achieving an effective trade-off between compression rates and quality.

The methodology of our work includes:

1) The separation of the video into scenes based upon important changes in the size of the I frames of the included Group of Pictures (GoP) and the subsequent calculation of the scenes length in terms of frames; in addition the scene separation scheme is fitted into a Pareto distribution as suggested by previous works;

2) The estimation of the size of I and B frames in a) the overall trace, b) a specific scene and c) a Group of Pictures; in addition the frame size scheme are fitted into statistical distribution (lognormal) as suggested from previous best practices.

Having established the main mechanism, we investigate the stability of the main parameters of the statistical distributions, considering different types of content, based on the level of movement as quantified through the scene ratio.

3. Trace Analysis

We leverage a video trace with high definition resolution of 1920 × 1080, automatic QP (Quantization Parameter) cascading G16B15 and Quantization Parameters 35 and 24 frames per second in order to perform our design and implementation. The trace has only I and B frames and the GoP consists of I frame and 15 B frames. The trace has been encoded using H.264/AVC, by the JSVM (9.19.14) encoder. The test video (Harry Potter) is the retrieved from [13].

The total amount of frames in the trace and the overall sequence are depicted in Figure 1, where the horizontal axis indicates the frame number while the vertical axis indicates the size of each frame.

We can observe the variability of the frames included in the sequence as well as the correlation of the size of frames that are temporally near; these frames are considered to belong to the same scene.

Indicative frames populated with the fields (sequence number, time, type and size) characterizing the trace are presented in the following Table 1.

3.1. I and B Frames

The trace consists of a series of Group of Pictures (GoP). The structure of the

GoP is well-defined for each specific encoding scheme, followed upon the overall trace; the pattern is one I followed by 15 B frames. The separation of I and B frames can be verified by the sizes of the frames: the size of the first frame is larger than the size of the following fifteen (as depicted in **Figure 2** for the first fifty frames).

The trace has approximately 5300 I frames with a maximum size of 130 Kbytes as depicted in **Figure 3**. It can be observed that the frame sequence is coarsely composed by intervals in which I frames are relatively close in size. These I frames are thematically connected forming a scene.

Similar parsing of the trace file, in order to identify and extract the volume of the B frames is performed. As depicted in **Figure 4**, approximately 80,000 B frames are included in the video. This number is in accordance with the number of I frames.

Figure 4 also depicts the size of the B frames. Specifically, the maximum size of B frames reaches the 50 Kbytes (with some exception at s almost 70 KBytes) which is approximately one third of that of I frames. The size of the B frames is related with each other within a GoP as well as with the initial I frame (as indicatively depicted in **Figure 2**).

3.2. Scene Separation

Changes in scenes are identified through significant alterations of I frames size. We can define the scene length as the subsequence of I frames (and corresponding GoPs) which have similar sizes, *i.e.* with size differences below a specific threshold. Considering the theoretical analysis in [8], a scene separation mechanism can operate identifying a threshold change in the size of I frames.

Specifically, we can consider that I Frame is the initial frame of a scene if

$$\frac{\left|I(k+1)-I(k)\right|}{\dfrac{\sum_{j=1}^{n} I(j)}{n}} > T_1 \tag{1}$$

where, I is the sequence of I frames, k indicates the number of I frame and n is the summary number of I frames. T_1 is the threshold which represents the permissible frames difference and it is equal 0.2 (20%) respectively. **Figure 5** presents the scenes identified and their length in relation to the number of the contained I frames. The scenes are grouped by the number of I frames, and the sums are divided by the total number of scenes in order to find the probability for each number of I frames.

Table 1. Video trace format.

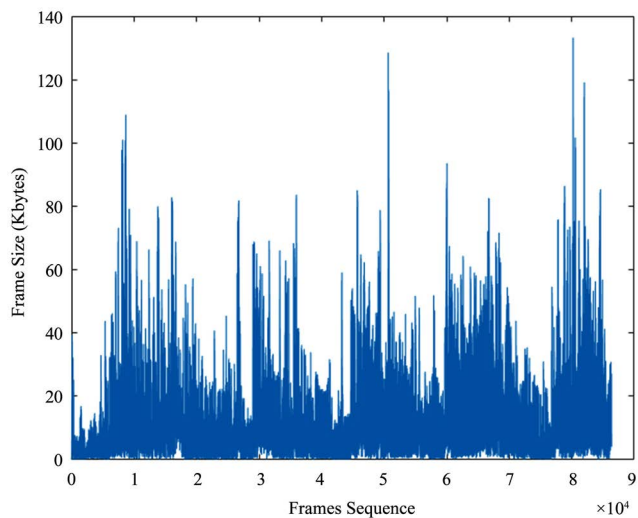| Seq. Num.	Time (Sec)	Type (I, P, B)	Size (Bytes)
1	0	I	443
2	0.0833	B	81.84
3	0.125	B	81.84

Figure 1. Size of frames composing the sequence.
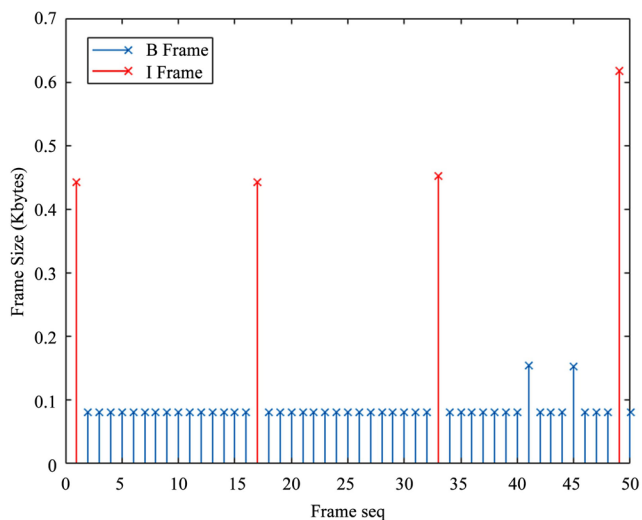

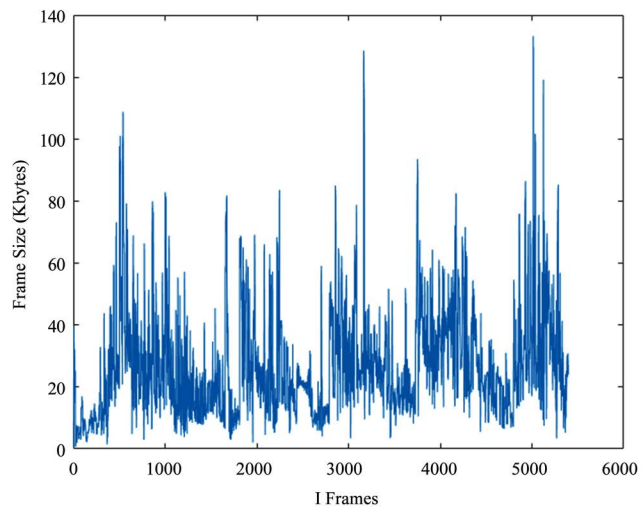**Figure 2.** Frame sequence from 1st to 50th.

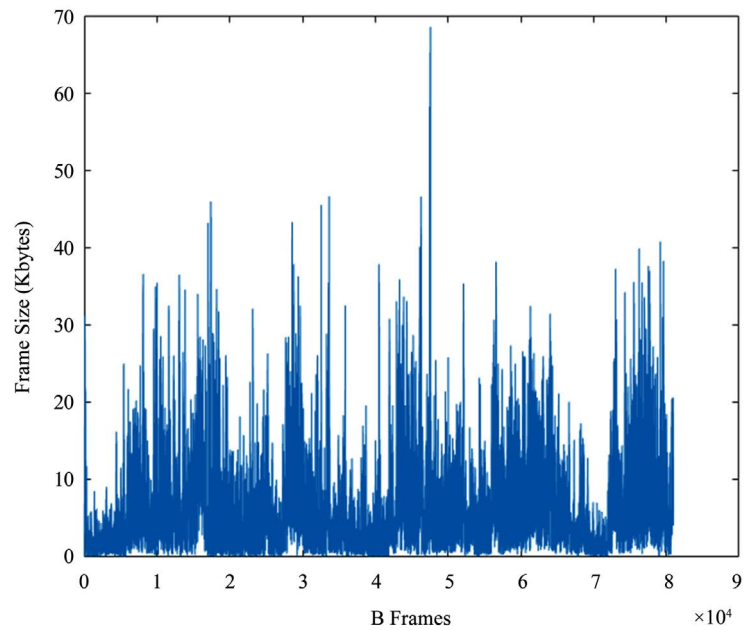**Figure 3.** Sequence of I frames.
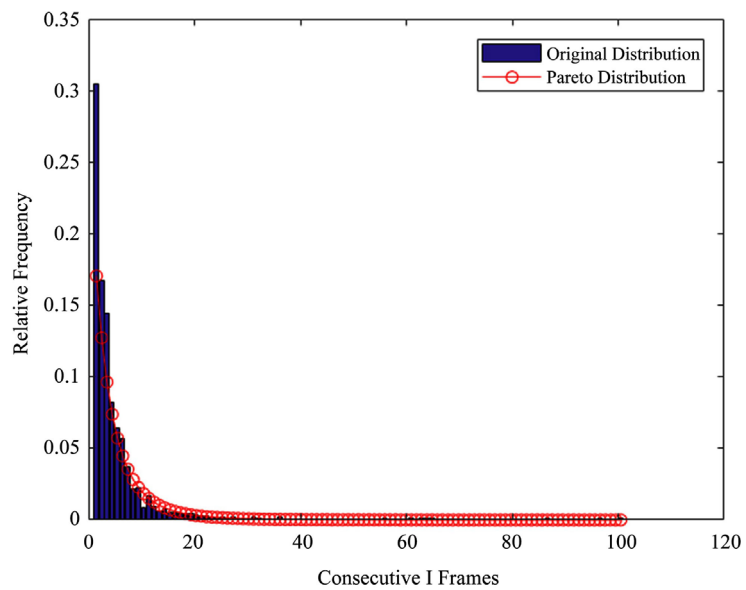
Figure 4. Sequence of B frames.

Figure 5. Distribution of trace scenes in respect with the number of I frames included.

According to [4], the Pareto distribution can be used to model the scenes length. Feeding the available data of our example into Matlab mathematical package and using the distribution fitting functionality (*fitdist*), a fitted distribution is provided with k = 0.1990, σ = 3.6651, and θ = 0. *K* is the shape parameter, σ is the scale parameter and θ is the threshold parameter.

Figure 5 presents the original data and the Pareto distribution.

3.3. Size of I Frames

Regarding the size of I frames, the values (as provided from the video trace) are

separated into groups and the frequency of appearance is calculated per group. After normalization a probability distribution is produced. In [14] different distributions are compared for modeling I frames sizes; it is concluded that the most suitable choice for MPEG frame modeling is the Lognormal distribution.

For the video of our example, after extracting the distribution parameters through optimal fitting (using Matlab mathematical package), the Lognormal distribution has a mean value equal to 3.0890 and a standard deviation equal to 0.6556. Figure 6 illustrates the actual data and the fitting lognormal distribution; the actual data are well-fitted upon the distribution.

While I frames follow the lognormal distribution across different scenes, I frames within a scene are related in terms of sizes, *i.e.* I frames of a scene follow in size that of the first I frame. The differences of I frames in each scene with respect to the first I frame are fitted to a normal distribution. The parameters for the normal distribution for the video example are the following mean value = −0.2713 and sigma = 4.2472. The actual data and the modelled distribution are presented in Figure 7.

3.4. Size of B Frames

The size of the B frames in the overall trace follow a lognormal distribution (similar to the case of I frames). After fitting the actual data upon the lognormal, the parameters that occur are: mean value = 0.8795 and Sigma = 1.0309. Figure 8 depicts the actual data and the fitted lognormal distribution.

While in the case of I frames we could use the distribution of the overall trace in order to have a value of the first I frame of a scene, B frames in a GoP are directly connected with the corresponding I frame. In fact, B Frames have information only for the changes between the current frame and both the previous and the next frame; so their size should be smaller than the size of the corresponding I frames. Considering this relationship we calculate the differences between the B frames and the I frame in all GoP (performing normalization). Figure 9 depicts the relation between the I frame size of the GoP with respect to the corresponding B frames.

3.5. Overall Statistical Parameters

In Table 2 the statistical parameters for scenes length, I frames sizes and B frames sizes, are presented. The threshold for scene alteration is 20%.

Table 2. Statistical parameter for scenes length, I frames' sizes and B frames' sizes.

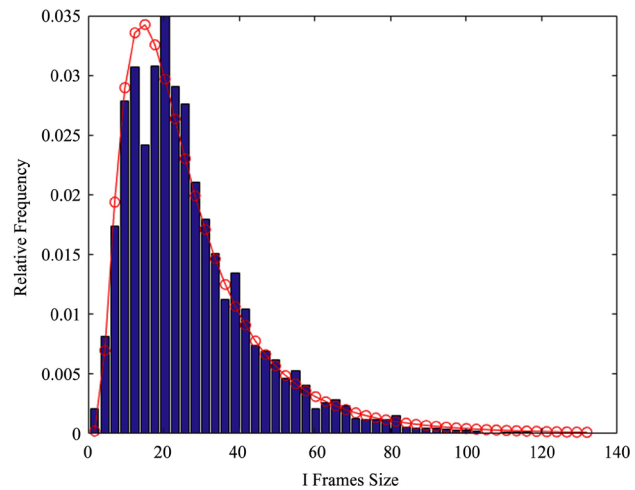|	Video Trace Features				
	Scenes	I Frames (Per Trace)	I Frames (Per Scene)	B Frames (Per Trace)	B Frames (Per GoP)
Distribution	Generalized Pareto	Lognormal	Normal	Lognormal	Lognormal
Parameters	k = 0.1990	mu = 3.0890	mu = -0.2713	mu = 0.8795	mu = 2.0058
	$\sigma = 3.6651, \theta = 0$	sx = 0.6556	sx = 4.2472	sx = 1.0309	sx = 0.6563

Figure 6. Distribution of I frames sizes in different scenes (lognormal distribution).
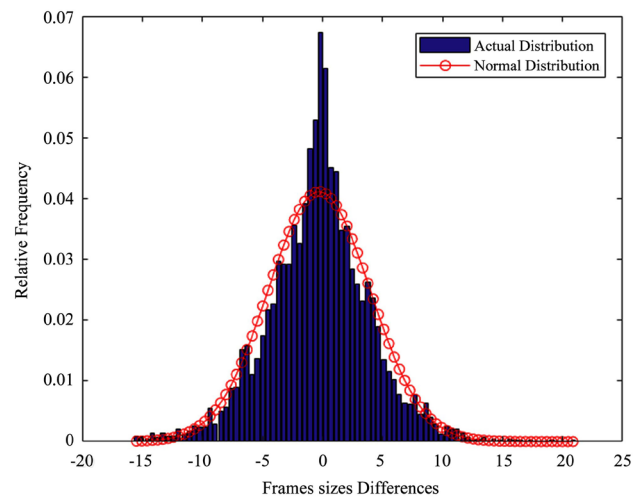

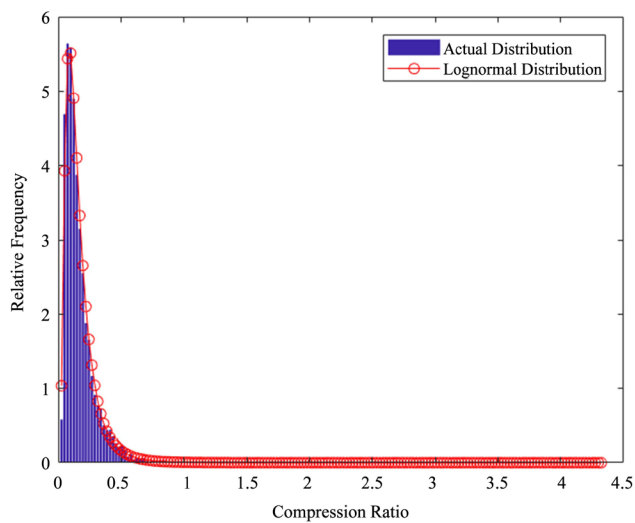**Figure 7.** Distribution of size differences among I frames in the same scene (normal distribution).

Figure 8. Distribution of B Frames sizes in the overall trace (lognormal distribution).
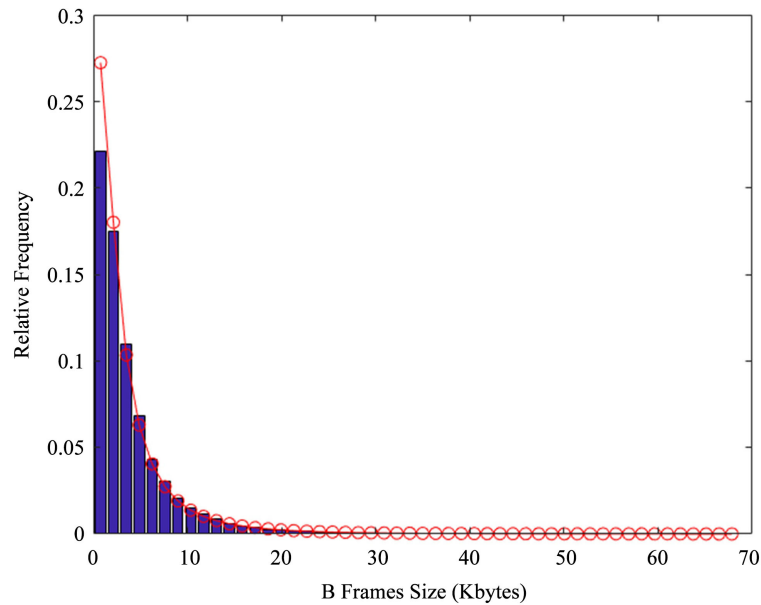
Figure 9. Distribution of size differences among B frames in the same GoP (normal distribution).

Following the distributions and parameters identified and calculated, we have created a synthetic trace and contrasted it with the actual trace of our example; as graphically depicted, in **Figure 10**, the two traces, the actual and synthetic traces have similar frame sizes.

Figure 11 and **Figure 12** depict the statistical properties of the synthetic trace as compared with those of actual trace. Specifically, **Figure 11** presents the distribution of the sizes of I frames of synthetic trace in comparison with the actual distribution and **Figure 12** depicts the B frame sizes of the synthetic trace. Both I and B frames sizes follow, closely enough, the original distribution of actual trace. Specifically, for the synthetic trace the I frame distribution is lognormal with mu equal to 3.0156 and sigma equal to 0.8724, while the B Frames lognormal Distribution has mu equal to 1.0044 and sigma equal to 1.0946.

4. Parameters of Traces of Different Content Types

Measurements on Different Types

In this Section we follow the methodology described for different types of videos. We have considered, as the differentiating parameter, the level of action in the video as expressed through the ratio of scene changing.

In this view, we have included the following types of videos:

- Action movie (high level of action);
- Documentary (medium level of action);
- News broadcasting (low level of action).

For the experimentation we have retrieved video traces with the same encoding parameters. The results are presented in **Figure 13** (where the mean values are included) and **Figure 14** (where the variations are included).
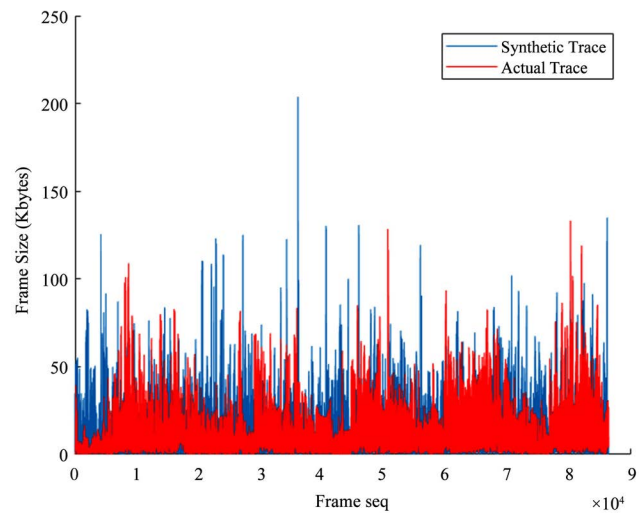
Figure 10. Comparison of trace generated with modelled parameters and authentic trace.
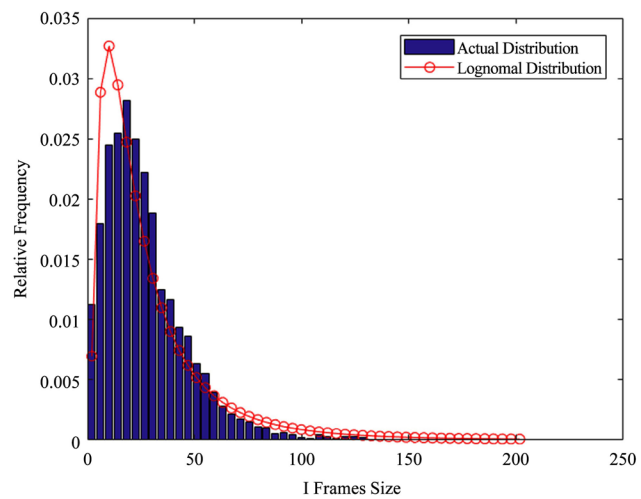

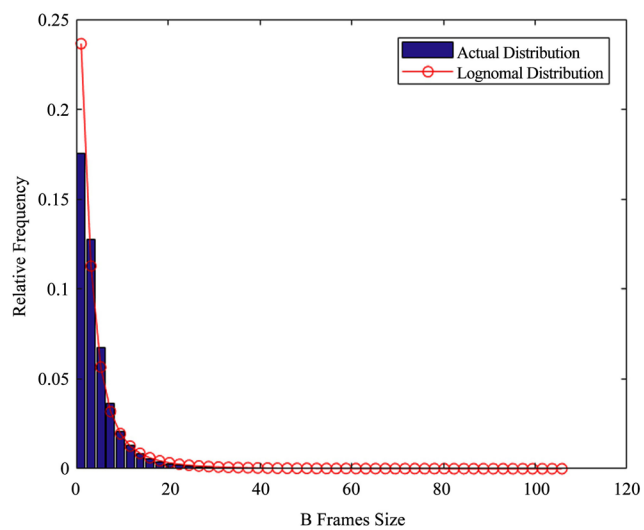**Figure 11.** Distribution of I frames sizes in the synthetic trace.

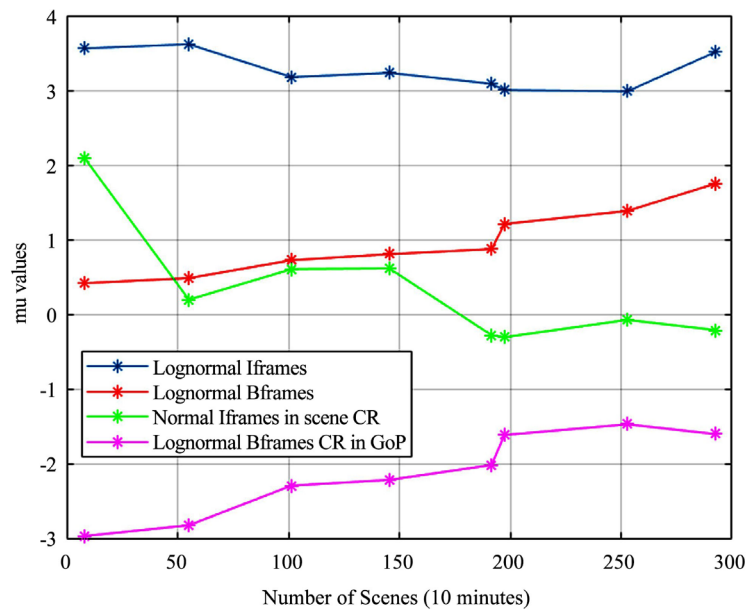**Figure 12.** Distribution of B frames sizes in the synthetic trace.

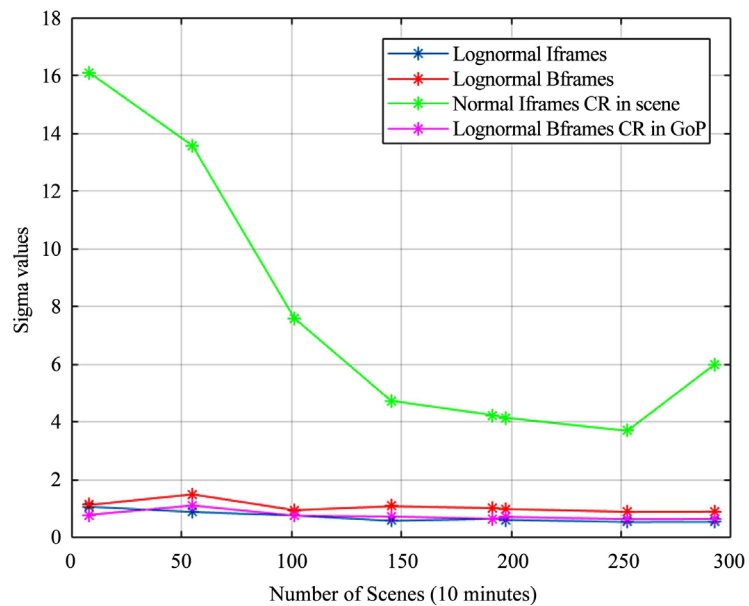**Figure 13.** Comparison of mean values for different activity levels.

Figure 14. Comparison of variation for traces of different activity levels.

In **Figure 13**, regarding the mean values of I frames in the overall trace we can identify some differentiations per activity level but the values of the lower and higher activity are almost identical. For the B frames in the overall trace the mean value is following a slightly increasing pattern which can be justified due to the increasing activity level. The mean value of the differences among I frames in a scene are decreased when the activity level is increased which is justified as the scene change in a more rapid fashion and the number of I frames included is decreased. For the B frames within a GoP the mean values is increased with the activity level.

In **Figure 14**, the variation for I and B frames in the overall trace as well as the variation of B frames in a GoP show a stable behavior. The variation of I frames within a scene is decreased with the enhancement of the activity level, which can be justified considering that the overall number of I frames in a scene I decreased.

Table 3 presents the statistical properties of video traces which are depicted in **Figure 13** and **Figure 14**. The videos are sorted based upon the ratio of scenes per 10 minutes (increasing rate). We consider a proportional characterization of low, medium and high action corresponding to the intervals [0, 100), [100, 200) and [200, 300) changes of scene per 10 minutes.

5. Conclusions

In this work, we explored and implemented the mechanisms to retrieve statistical parameters from MPEG4 video traces, mainly related to the size of the included frames. Specifically, based upon the GoP structure and the types of frames, we have identified significant changes in the size of I frames indicating the boundaries of scenes. A number of scenes included in the overall traces, resulting in a scene ratio, as well as a number of included GoP in a scene, have

Table 3. Statistical parametersof video traces.

Trace	Scenes		I Frames (Per Trace)	I Frames (Per Scene)	B Frames (Per Trace)	B Frames (Per GoP)
		Video Trace Statistical Properties				
	Scenes/10 min	Generalized Pareto	Lognormal	Normal	Lognormal	Lognormal
BluePlanet	8.2	$k = 0.2173$	mu = 3.5616	mu = 2.0894	mu = 0.4240	mu = −2.9551
		$\sigma = 18.5984, \theta = 0$	sx = 1.0776	sx = 16.0949	sx = 1.1388	sx = 0.7976
SonyDemo	55.10	$k = 0.2816$	mu = 3.6174	mu = 0.2038	mu = 0.4915	mu = −2.8116
		$\sigma = 14.4284, \theta = 0$	sx = 0.8964	sx = 13.5679	sx = 1.5063	sx = 1.1205
LakeHouse	101.38	$k = 0.4977$	mu = 3.1779	mu = 0.6114	mu = 0.7307	mu = −2.2812
		$\sigma = 4.5036, \theta = 0$	sx = 0.7721	sx = 7.6021	sx = 0.9587	sx = 0.7661
FindingNeverland	145.37	$k = 0.2364$	mu = 3.2344	mu = 0.6230	mu = 0.8112	mu = 2.2056
		$\sigma = 4.6904, \theta = 0$	sx = 0.5904	sx = 4.7422	sx = 1.0935	sx = 0.7363
HarryPotter	191.59	$k = 0.1990$	mu = 3.0890	mu = −0.2713	mu = 0.8795	mu = 2.0058
		$\sigma = 3.6651, \theta = 0$	sx = 0.6556	sx = 4.2472	sx = 1.0309	sx = 0.6563
Fugitive	197.43	$k = 0.2322$	mu = 3.0052	mu = −0.2978	mu = 1.2149	mu = −1.6048
		$\sigma = 3.4391, \theta = 0$	sx = 0.6234	sx = 4.1566	sx = 0.9988	sx = 0.7232
Speed	252.79	$k = 0.1357$	mu = 2.9875	mu = −0.0668	mu = 1.3897	mu = −1.4622
		$\sigma = 3.0589, \theta = 0$	sx = 0.5478	sx = 3.7017	sx = 0.9007	sx = 0.6460
Transporter II	292.73	$k = 0.0990$	mu = 3.5153	mu = −0.2035	mu = 1.7549	mu = −1.5934
		$\sigma = 2.7515, \theta = 0$	sx = 0.5534	sx = 6.0134	sx = 0.9078	sx = 0.6511

been statistically processed and Pareto-based distribution has been formulated. The size of I and P frames have also been statistically processed in the overall trace (using lognormal distributions) as well as in a single scene and GoP respectively (using the normal and lognormal distributions). Through our work, we have verified that the selected distributions follow the actual data.

Furthermore, we have repeated the methodology considering traces of different activity levels. The results have verified the stability of the models. In addition, the results have allowed for observations, regarding the dependence of the statistical parameters upon the video activity level (measured through the scene change ratio).

As for future work, we consider further elaboration of the statistical parameters on specific types of videos (in terms of activities) and the design of mechanisms supporting the identification of events of interest. Such identification can be context-based, especially in the cases of monitoring or surveillance where a typically stable rate of frame size is interrupted, due to activity, with frames of different sizes.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Van der Auwera, G., David, P. and Reisslein, M. (2008) Traffic and Quality Characterization of Single-Layer Video Streams Encoded with the H.264/MPEG-4 Advanced Video Coding Standard and Scalable Video Coding Extension. *IEEE Transactions on Broadcasting*, **54**, 698-718.

[2] Koumaras, H., Skianis, C. and Kourtis, A. (2009) Analysis and Modeling of H.264 Unconstrained VBR Video Traffic. *International Journal of Mobile Computing and Multimedia Communications*, **1**, 14-31. https://doi.org/10.4018/jmcmc.2009072802

[3] Zahariadis, T., Papadakis, A. and Nikolaou, N. (2012) Simulation Framework for Adaptive Overlay Content Caching. *IEEE* 35*th International Conference on Telecommunications and Signal Processing* (*TSP*), Prague, 3-4 July 2012, 62-66.

[4] Reaz, A., Murayama, D., *et al.* (2011) Synthetic Traffic Generation for Streaming Video to Model IPTV. 5*th IEEE International Conference on Advanced Telecommunication Systems and Networks* (*ANTS*), Bangalore, 18-21 December 2011, 1-6. https://doi.org/10.1109/ANTS.2011.6163638

[5] Seeling, P. and Reisslein, M. (2012) Video Transport Evaluation with H.264 Video Traces. *IEEE Communications Surveys and Tutorials*, **14**, 1142-1165. https://doi.org/10.1109/SURV.2011.082911.00067

[6] Wan, F., Cai, L. and Gulliver, T. (2008) A Simple, Two-Level Markovian Traffic Model for IPTV Video Sources. 2008 *IEEE Global Telecommunications Conference*, New Orleans, 30 November-4 December 2008, 1-5. https://doi.org/10.1109/GLOCOM.2008.ECP.288

[7] Liu, R., Rao, D., Huang, Z. and Yang, D. (2012) An Emperical Traffic Model of M2M Mobile Streaming Services. 2012 4*th International Conference on Multimedia Information Networking and Security*, Nanjing, 2-4 November 2012, 400-404.

https://doi.org/10.1109/MINES.2012.60

[8] Krunz, M. and Hughes, H. (1995) A Traffic for MPEG-Coded VBR Streams. ACM SIGMETRICS, Ottawa. https://doi.org/10.1145/223587.223592

[9] Van der Auwera, G., David, P.T. and Reisslein, M. (2008) Traffic and Quality Characterization of Single-Layer Video Streams Encoded with H.264/MPEG-4 Advanced Video Coding Standard and Scalable Video Coding Extension. *IEEE Transactions on Broadcasting*, **54**, 698-718. https://doi.org/10.1109/TBC.2008.2000422

[10] Politis, I., Dounis, L., Tselios, C., Kordelas, A., Dagiuklas, T. and Papadakis, A. (2012) A Model of Network Related QoE for 3D Video. 2012 *IEEE Globecom Workshops*, Anaheim, 3-7 December 2012, 1335-1340.

[11] Bhaumik, P., SayeemReaz, A. and Mukherjee, B. (2012) Performance of 10G-EPON in Streaming IPTV. 2012 *IEEE International Conference on Advanced Networks and Telecommunciations Systems* (*ANTS*), Bangalore, 16-19 December 2012, 30-32.

[12] Bhaumik, P., SayeemReaz, A., Murayama, D., Suzuki, K.-I., Yoshimoto, N., Kramer, G. and Mukherjee, B. (2015) IPTV over EPON: Synthetic Traffic Generation and Performance Evaluation. *Optical Switching and Networking*, **18**, 180-190.

[13] Video Traces for Network Performance Evaluation. Arizona State University, Tempe. http://trace.eas.asu.edu/tracemain.html

[14] Kyungwoo, K. and Haniph, A.L. (2009) Statistical Traffic Modeling of MPEG Frame Size: Experiments and Analysis. *Systemics, Cybernetics, and Informatics*, **7**, 54-59.