Noname manuscript No.
(will be inserted by the editor)

A Novel Approach for Characterizing Multimedia 3D Video Streams by means of Quasi-Periodic Processes

Raoul R. Nigmatullin · Domenico Striccoli · Gennaro Boggia · Cristiano Ceglie

Received: date / Accepted: date

Abstract This work proposes a novel methodology to describe the Low-Frequency behaviour of compressed 3D video streams, i.e., their average fluctuations on longer time scales. This study is innovative for two reasons. First, it proves that the Low-Frequency behaviour of the video data belongs to the class of Quasi-Periodic processes. Second, it proposes an innovative approach to describe the long-term behaviour through a set of parameters directly derived from the Quasi-Periodic analysis. Reported results show that the proposed approach is effective in a wide variety of simulation scenarios. Furthermore, it can be easily generalized to other kinds of compressed two-dimensional (2D) streams, whatever the adopted algorithm, the compression degree, video resolution and format. This opens new unexplored possibilities in the field of 3D video characterization, identification and classification.

Keywords 3D video · Quasi-Periodic processes · Stream Characterization · Data Fitting

1 Introduction

Processing the information associated to multimedia three-dimensional (3D) videos is becoming more and more an attractive issue in current scientific literature due to the fast development of last-generation (wireless and wired) networks and technologies [7]. A 3D video includes the representation of the spatial depth information of a

Raoul R. Nigmatullin
Radio-Electronic and Informative Measurements Technics Department, Kazan National Reseach Technical University (KNRTU-KAI), Karl Marx str., 10, 420111, Kazan, Russia
E-mail: renigmat@gmail.com

Domenico Striccoli (corresponding author), Gennaro Boggia and Cristiano Ceglie
Department of Electrical and Information Engineering (DEI), Politecnico di Bari, Via E. Orabona, 4, 70125, Bari, Italy
Tel.: +39-080-5963301
Fax: +39-080-5963410
E-mail: {domenico.striccoli,gennaro.boggia,cristiano.ceglie}@poliba.it

scene, which is totally absent in 2D videos. This added value can be fruitfully adopted in several fields like entertainment, medicine, industry, and so on. The drawback is the increased amount of data needed to describe the extra information of 3D frames. To overcome this problem, several efficient video compression and coding techniques and formats have been implemented to provide 3D video services on transmission links with a relatively limited bandwidth [12], which complicate the design of efficient bandwidth management and estimation schemes, needed to reach the Quality of Service (QoS) and Quality of Experience (QoE) guarantees for end users [12] [8].

The study performed in this work focuses on the Low-Frequency (LF) behaviour of compressed streams because, as well known from literature, 2D and 3D compressed videos are typically Variable Bit Rate (VBR) streams [3] [12], and they are typically characterized by "low-frequency" components for bandwidth allocation and management, and QoS guarantees [6] [17] [9]. For this reason, the high frequency part of the spectrum of VBR streams (i.e., in the time domain, strong data variations over smaller time scales) is usually removed before modeling VBR sources [3].

For all these reasons, a punctual and synthetic description of the LF properties that characterize 3D video sequences is of paramount importance. Unlike the most common approaches aiming to describe statistically the properties of compressed video flows [4] [2], or referring only to specific video formats [14] [16], the method proposed in this study aims to *deterministically* characterize the LF behaviour of 3D video flows, through a small set of parameters, whatever the type, format and compression degree of 3D streams. This finding, in the authors' opinion, can be of great help for network researchers, in their work of research of new methods for video classification, retrieving and mining, and for designers, in their work of traffic engineering, network resources assignment, and QoS management.

To give more strength to this claim, the proposed approach has been also compared with the Hidden Markov Model (HMM) approach. The HMM model is based on a Markov process modeling the size of the frames forming the Group of Pictures (GoP) (the periodic structure of compressed frames characterizing the video flow). This Markov-based model has been chosen for comparison since, as described in detail in [16], it appears to be more accurate than other types of models in describing the frame size distribution and the correlations among video frames.

The rest of the paper is organized as it follows. Sec. 2 describes the proposed methodology. Sec. 3 illustrates the effectiveness of the proposed approach . Finally, in Sec. 4 some conclusions are drawn.

2 The proposed methodology

The main scope of this section is to describe the mathematical background which the technique for capturing the LF properties of a video data set is based on. In this section, the key-aspect of the proposed method is shown, that is, how the LF behaviour of compressed streams can be effectively described through a Quasi-Periodic (QPr) process.

Ti this aim, the data classification procedure explained in [10] is followed, where a QPr process with a memory of only one period (T) follows the functional equation:

$$F(t+T) = a \cdot F(t) + b \tag{1}$$

In this case, the solution of (1) can be expressed in a closed form:

$$F(t) = B + E_0 e^{\lambda t/T} + \sum_{i=1}^{K} \left[A_{c_i} \phi_i(t/T) + A_{s_i} \psi_i(t/T) \right] \ , \tag{2}$$

where $\lambda = \ln(a)$, $B = \frac{b}{(1-a)}$, and $\phi_i(x) = e^{\lambda x} \cos(2\pi i x)$ and $\psi_i(x) = e^{\lambda x} \sin(2\pi i x)$ are the so called "fitting modes" (see [10] for further details).

In [10] it is shown that the so-called criterion of the reduction to three incident points is a sufficient condition to detect the presence of the Quasi-Periodic property of a data set. So, the application of this criterion to the LF trend of the video data will guarantee its description through a QPr process. This general (even though approximate) criterion can be resumed in what follows. The data set is subdivided into disjoint subsets of the same size, say of m samples each. In each subset, the maximum, mean, and minimum values of the m values are derived, forming three sequences in a "compressed" domain (with a compression factor of m). The criterion says that if the three so obtained sequences keep numerically close each other, this is an indication of the invariance of the data set under analysis, with respect to self-similar transformations [11]. Consequently, these three sequences can be considered as three strongly-correlated realizations of a QPr process with the functional equation (1), which admits the solution (2), and whose parameters a and b can be easily derived by averaging the relative parameters of the three realizations [10]. So, eq. (2) can be chosen as the fitting function that quantitatively describes the unknown curve in terms of $2K + 3$ parameters: B, E_0, T, and the decomposition coefficients A_{c_i}, A_{s_i} $(i = 1, 2, ..., K)$. All these parameters can be found by the Linear Least Square Method (LLSM) [5] [13]. The proposed LF analysis is now explained in detail. The goal is to accurately fit the LF part of the video data under test with the analytical function (2), finding the optimal values of the 2K+3 parameters mentioned above.

To this aim, the first step is a low-pass filtering performed on the video data, to extract the only LF behaviour of the video sequence. This is done through an integration with the trapezoidal rule (the most simple type of low-pass filtering, widely utilized in the numerical analysis [15]):

$$J_{y_i} = J_{y_{i-1}} + \frac{1}{2} \left(t_i - t_{i-1} \right) \cdot \left(D_{y_i} + D_{y_{i-1}} \right), \ 1 \leq i \leq N. \tag{3}$$

where $\overline{y}$ is the sample mean of the original data set, $D_{y_i} = y_i - \overline{y}$, and J_{y_i} is the resulting i-th filtered sample. In this manner, this very simple, even if approximated, method cuts the high frequency components.

The second step aims to find the parameters of Eq. (2) starting from the values of T and K (the guidelines for their ranges of variation can be found in [10]).

The optimal values of T and K in Eq. (2) can be found minimizing the Relative Absolute Error (RAE) $\rho(T, K)$ between the original data samples J_{y_i} and the corresponding fitting values $F(t_i)$ [1]. It is defined as:

$$\rho(T,K) = \left(\sum_{i=1}^{N} \left| F(t_i) - J_{y_i} \right| \right) / \left(\sum_{i=1}^{N} \left| J_{y_i} - \overline{J} \right| \right) \cdot 100\% \tag{4}$$

where $\overline{J} = \sum_{i=1}^{N} J_{y_i}/N$ is the sample mean of J_{y_i}. In Eq. (4), the relative error depends on T and K due to Eq. (2). The optimal values of T and K should minimize the relative error with respect to these parameters, i.e., $\rho_{opt} = \min_{T,K}\{\rho(T,K)\}$.

It is worth noting that, for the scope of the procedure, it is not necessary to find the absolute minimum relative error, but a set of parameters ensuring that the relative error falls in an acceptable range, i.e., $1\% < \rho_{opt} < 10\%$.

Finally, the last remaining $2K+1$ unknown parameters of Eq. (2), i.e., E_0, A_{c_i} and A_{s_i}, can be easily found applying the LLSM technique.

The total set of these parameters *fully* quantifies the trend of the data set under analysis and can be fruitfully adopted for the accurate fitting of multimedia 3D data sequences.

3 Numerical Results

The goal of this section is twofold. First, the methodology developed in Sec. 2 is applied to 3D streams of different types, to study the main trends of the compressed 3D videos in order to characterize them with this novel approach. In this context, to further highlight the robustness of this methodology, it has been compared with the HMM approach as described in [14] and [16]. Second, to show that this approach can be successfully applied to other types of video data, it has been applied also to several other 2D streams, with different coding algorithms, compression degrees, and GoP patterns.

Regarding the 3D streams, the data sets listed in Table 1 have been chosen, with different formats, encoding schemes, and Quantization Parameters (QPs) (see [12] for further details on the chosen parameters). Each video lasts about 35 minutes.

Table 1 Encoding parameters for different 3D video formats.

<table>
<tr><th>3D Video Format</th><th>Encoding Algorithm</th><th>GoP Structure</th><th>Quantization Parameter</th></tr>
<tr><td>FS</td><td rowspan="2">H264/AVC</td><td>G32B1, G32B7</td><td rowspan="3">24,28,34</td></tr>
<tr><td>SBS</td><td rowspan="2">G16B1, G16B7</td></tr>
<tr><td>MV</td><td>H264/MVC</td></tr>
</table>

To explain step by step the results of the proposed approach, the film “Alice in Wonderland” in Side-By-Side (SBS) video format, with a QP $= 34$ and GoP structure $G16B1$ is considered in what follows. This considered stream contains $N = 51184$ 3D frames, or data sample points.

The first step is the data conditioning as described by Eq. (3). In this case we obtain a more smoothed curve, with slower fluctuations.

The second step is the application of the procedure of the three incident points to the filtered data set. According to the suggestions in [10], to assure both a sufficient number M of intervals, and at the same time a relatively high number m of samples per each segment, the value $m = 250$ is chosen. Thus, the number of segments is $M = \lfloor N/m \rfloor = 204$. After the three curves F^{up}, F^{mn}, and F^{dn} are built, it can be noted that they are very close to each other (this result is not shown for brevity). As described in Sec. 2, this condition, which has been verified for all the chosen 3D streams, guarantees the classification of their LF behaviour as a QPr process.

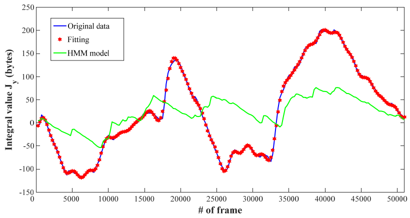
The parameters a and b in Eq.(1) are then computed along with all the parameters of $F(t)$ in Eq. (2). Similar results hold also for the other chosen streams.

For the same trace, and for sake of comparison, the HMM model has also been implemented. As described in [14], it consists of a non-stationary hidden Markov model to characterize the frame size main statistics for 3D streams. Furthermore, to increase the accuracy of the model, the distributions of the I, P, and B frames in each state (which models the level of the activity in the scene), and the time of permanence in the state, have all been modeled by histograms directly derived from the real video trace.

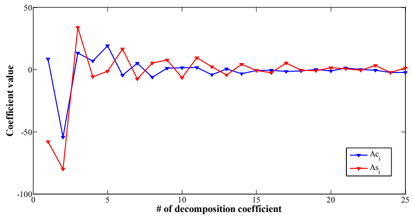
It is known by [16] that the only way to use the Markov-based models for the video trace representation is to generate synthetic traces and subsequently use them in simulation studies. This procedure has been adopted also in this case, by extracting randomly the I, P and B frames of each GoP with exactly the same frame sequence of the real video GoP pattern. Also the time of permanence in each state has been extracted in the same way. Then, 100 synthetic traces have been generated, each time calculating the relative error.

Fig. 1(a) demonstrates the result of the comparison between the original trace and the traces obtained by applying the proposed method and the HMM model. The HMM trace has been chosen as the one with the minimum relative error among the 100 obtained results. The comparison shows the very good fit of the compressed curve obtained through eq. (2), and that the proposed method significantly outperforms the HMM model. This last result holds also for all the other 3D chosen streams (whatever the value of the compression parameters). The relative fitting error is equal to 3.39% and a value of $K = 25$ is enough for providing a good fit. The corresponding error for HMM is instead of 84.3%. It is worthwhile to note that the fitting of the compressed data is *exactly* the same fitting of the totality of uncompressed integral data, due to the scaling properties of $F(t)$ highlighted at the end of Sec. 2. Fig. 1(b) shows the behavior of the decomposition coefficients A_{c_i} and A_{s_i} relative to the considered data example.

In the same manner all the other considered multimedia 3D flows have been fitted. The fitting procedure is applied to the filtered data relative to the other two 3D movies (a piece of the “IMAX Space Station” film, and a piece of “Monsters vs Aliens” cartoon). Figures 2(a) and 2(b) show that the proposed method is able to ensure a very good fitting, keeping the same settings of the “Alice in Wonderland” stream (i.e., QP $= 34$, GoP = $G16B1$ and $K = 25$).

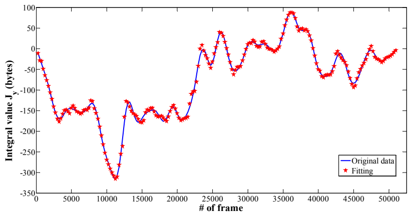


(a) Final fitting of the integrated curve of the video through the proposed method, and comparison with the HMM model.

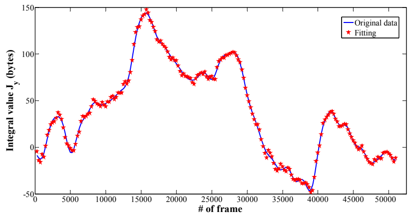


(b) The behavior of decomposition coefficients (A_{c_i} and A_{s_i}) figuring in the fitting function (2) for the example data film.

Fig. 1 Result of the fitting of the "Alice in Wonderland"film.



(a) The fitting of the "IMAX Space Station" movie (SBS, $QP = 34$, GoP = $G16B1$). Relative error equal to 5.69%.



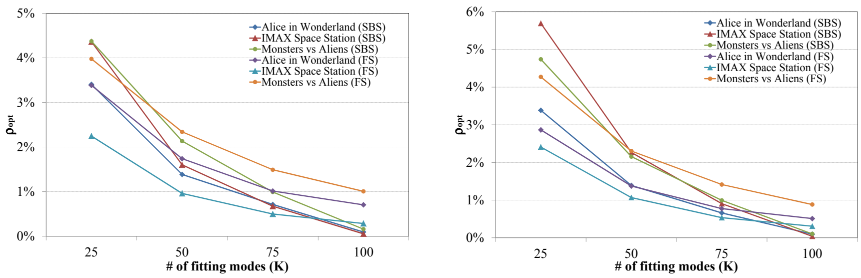
(b) The fitting of the "Monsters vs Aliens" cartoon (SBS, $QP = 34$, GoP = $G16B1$). Relative error equal to 4.74%.

Fig. 2 Result of the fitting of the other films.

Now, the fitting procedure has been tested in a more general scenario. To this aim, the three chosen streams have been considered with two different QPs. Figures 3(a) and 3(b) show the behavior of the relative error for different values of the total number K of linear coefficients. As expected, the relative error decreases as K increases: a higher number of coefficients allows a better fitting of the fluctuations in the video data. Conversely, the complexity of the fitting function increases, together with the computational effort to find the optimal values of all the parameters. Nevertheless, also the lower value of K (i.e., $K = 25$) is sufficient to guarantee a very good fitting, with a relative error always smaller than 6% in all the considered cases. This important result confirms the strength of the proposed approach, in a wide variety of application scenarios.

Table 2 Chosen two-dimensional streams with their encoding parameters

Video Name	Video Format	Encoding Algorithm	GoP Structure	Quantization Parameter
Speed	HD	H.265/HEVC	G24B7	10, 40
Tears of Steel	4k	H.264/H.265	G1B0, G24B0	20, 35
Terminator 1	SVC	H264	G16B1, G16B7	10, 50
Starwars	SVC/CGS		G16B7	48



(a) Relative fitting error vs number of fitting modes, for different movies and formats ($QP = 24$).

(b) Relative fitting error vs number of fitting modes, for different movies and formats ($QP = 34$).

Fig. 3 Relative fitting error for all the films.

Table 3 Relative error (in percentage) for a wide variety of 3D streams

	3D Video Format	QP = 24		QP = 28		QP = 34	
		B1	*B7*	*B1*	*B7*	*B1*	*B7*
Alice in Wonderland	SbS	3.4052	3.4772	3.5172	3.5392	3.3853	3.3981
	FS	3.3838	3.2557	3.1121	3.0368	2.8657	2.8447
	MV	3.4868	3.5018	3.5892	3.5217	3.4331	3.3873
IMAX Space Station	SbS	4.3610	4.6009	4.7518	5.0494	5.6939	5.9619
	FS	2.2462	2.2340	2.2490	2.2355	2.4098	2.3980
	MV	3.2718	3.3557	3.5423	3.6368	4.2204	4.2414
Monsters vs Alien	SbS	4.3725	4.4964	4.4823	4.6085	4.7396	4.8840
	FS	3.9768	3.9986	4.1405	4.1141	4.2717	4.2249
	MV	4.4237	4.5154	4.7117	4.8558	5.2487	5.3470

The procedure has been applied to a wide variety of video streams. Together with the abovementioned 3D streams, several other 2D streams have been added. They are listed in Table 2. Furthermore, to better show the robustness of the fitting analysis, the relative error has been computed for both 54 3D streams, the characteristics of which are shown in Table 3, and for all the 2D streams whose characteristics are listed in Table 2, and whose results are reported in Table 4. Results in Table 4 testify that this method can be extended to 2D compressed streams, whatever the type of video, the QP, the compression algorithm, and the GoP pattern. Finally, the comparison of the proposed approach with the HMM model (not shown here for lack of space) testifies again the much better performance of the former with respect to the latter, like the case of the 3D streams.

Table 4 Relative error (in percentage) for the chosen 2D streams

Name	**Speed**		**Tears of Steel**			**Terminator 1**			**Star Wars IV**		
GoP		**G24B7**		**G1B0**	**G24B0**		**G16B1**	**G16B7**		**G16B3**	**G16B15**
QP	**10**	3.27%	**20**	2.09%	2.24%	**10**	3.96%	4.42%	**48**	0.27%	0.28%
	40	1.84%	**35**	1.03%	1%	**50**	4.42%	4.31%			

4 Conclusions

This work presents a new approach in multimedia video streams characterization by describing their low-frequency behaviour through a Quasi-Periodic process. This is an interesting novelty in the field of compressed streams, and allows to characterize their slowly-varying trends through a small set of parameters. This important result found in this contribution has been widely tested not only for 3D streams, but also for other types of compressed 2D streams, for several video quality parameters. It shows that compressed streams, whatever the type, format, and resolution adopted, can be characterized by the specific class of Quasi-Periodic processes with a high degree of accuracy. This opens new unexplored possibilities, both for 3D video characterization itself, and for several applications ranging from bandwidth management to 3D stream identification and mining.

Further research work will aim at studying also the high-frequency (short term) fluctuations of multimedia streams, to give a more exhaustive video "fingerprint" and more accurately characterize compressed videos.

References

1. Armstrong, J., Collopy, F.: Error measures for generalizing about forecasting methods: Empirical comparisons. International journal of forecasting **8**(1), 69–80 (1992)
2. Beran, J.: Statistics for Long-Memory Processes, vol. 61. Chapman & Hall (1994)
3. Beran, J., Sherman, R., Taqqu, M., Willinger, W.: Long-range dependence in variable-bit-rate video traffic. IEEE Transactions on Communications **43**(234), 1566–1579 (1995)
4. Ceglie, C., Maione, G., Striccoli, D.: Statistical Analysis of Long-Range Dependence in Three-Dimensional Video Traffic. In: Proc. of Int. Conf. on Mathematical Methods In Engineering (MME 2013) (2013)
5. Chatterjee, S., Hadi, A.: Regression Analysis by Example. John Wiley & Sons, Inc. (2012)
6. Chong, S., Li, S.Q., Ghosh, J.: Predictive dynamic bandwidth allocation for efficient transport of real-time VBR video over ATM. IEEE Journal on Selected Areas in Communications **13**(1), 12–23 (1995)
7. Dahlman, E., Parkvall, S., Skold, J.: 4G LTE/LTE-Advanced for Mobile Broadband. Academic Press (2011)
8. Gurler, C.G., Gorkemli, B., Saygili, G., Tekalp, A.M.: Flexible transport of 3-D video over networks. In: Proc. IEEE, pp. 694–707 (2011)
9. Li, S.Q., Chong, S., Hwang, C.L.: Link capacity allocation and network control by filtered input rate in high-speed networks. IEEE/ACM Transactions on Networking **3**(1), 10–25 (1995)
10. Nigmatullin, R., Khamzin, A., Machado, J.: Detection of quasi-periodic processes in complex systems: how quantitatively to describe their properties? Physica Scripta **89**(1), 1–11 (2014)
11. Nigmatullin, R., Machado, J., Menezes, R.: Self-similarity principle: the reduced description of randomness. Central European Journal of Physics **11**(6), 724–739 (2013)
12. Pulipaka, A., Seeling, P., Reisslein, M., Karam, L.: Traffic and statistical multiplexing characterization of 3d video representation formats. IEEE Transactions on Broadcasting **59**(2), 382–389 (2013)
13. Rao, C.R., Toutenburg, H., Shalabh, Heumann, C.: Linear Models and Generalizations: Least Squares and Alternatives. Springer (2007)
14. Rossi, L., Chakareski, J., Frossard, P., Colonnese, S.: A Non-Stationary Hidden Markov Model of Multivew Video Traffic. In: Proc. of 2010 IEEE 17th International Conference on Image Processing (ICIP 2010), pp. 2921–2924 (2010)
15. Scott, L.: Numerical Analysis. Princeton University Press (2011)
16. Tanwir, S., Perros, H.: A survey of VBR video traffic models. IEEE Communications Surveys & Tutorials **15**(4), 1778–1802 (2013)
17. Zhang, Z.L., Kurose, J., Salehi, J., Towsley, D.: Smoothing, statistical multiplexing, and call admission control for stored video. IEEE Journal on Selected Areas in Communications **15**(6), 1148–1166 (1997)