An Information theoretical problem in interactive Multi-View Video services

Aline Roumy
Inria, Rennes, France
aline.roumy@inria.fr

1. Introduction

Free-viewpoint television (FTV) [1] is a novel system for watching videos which allows interaction between the server and the user. More precisely, different views of the same 3D scene are proposed and the user can choose its viewpoint freely. Moreover, the user can change its viewpoint at any time, leading to a free navigation within the 3D scene. (Figure 1 shows one navigation path within the views). The goal of FTV is thus to propose an immersive sensation, although the visualization remains 2D.

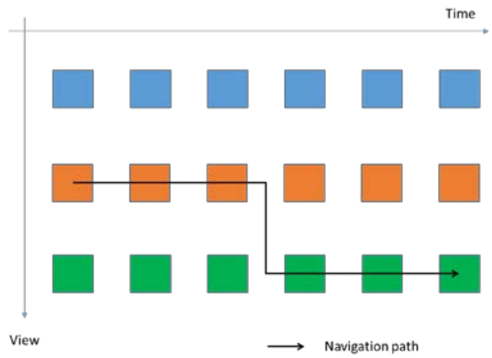

Fig. 1. FTV: the user can navigate from one view to another view during playback.

FTV presents nice features that are interesting in sport event like soccer. Indeed, in its application for the 2022 FIFA World Cup[1], Japan announced that the stadium would be equipped with FTV technology.

This soccer example immediately suggests the technical constraints on FTV. First, FTV implies the use of a *large database*. Indeed, the MPEG working group in charge of FTV [2] considers 100 views in order to allow smooth and nice navigation within the scene, which leads to video needing about 120 Gbps (uncompressed HD videos with 50 frames per second and 100 views). In addition, *the number of users is potentially huge* as seen in the soccer example. Finally, in such a scenario, only one view is needed at a time, and the choice of this view depends on the user. Therefore, each user *requests a subset of the data*, and this request can be seen as *random* from the sender perspective, as it only depends on the user choice.

2. A novel problem: massive random access to subsets of compressed correlated data

In fact, FTV raises a novel question that has not been yet studied in depth in information theory. This problem can be called massive random access to subsets of compressed correlated data and is illustrated in Figure 2.

This problem can be defined as follows. Consider a database that is so large that, to be stored on a single server, the data have to be compressed efficiently, meaning that the redundancy/correlation inside the data have to be exploited. The compressed dataset is then stored on a server and made available to users. We consider a scenario in which users want to access only a subset of the data. This is typically the case in FTV. Since the choice of the subset (i.e. the view in FTV) is user-dependent, the request (of a data-subset) can be modeled as a random access. Finally, massive requests are made, meaning that a lot of users may want to access some data at the same time. Consequently, upon request, the server can only perform low complexity operations (such as bit extraction but no decompression -compression).

The novelty of this problem lies in the study of data compression, while adding a new constraint: namely massive and random access. Indeed, classical unconstrained compression does not allow to access part of the compressed data. This is a consequence of the way the encoding map is constructed but also of the optimality that occurs only in the asymptotic regime. More precisely, the source coding map consists in associating to each input sequence an index that is by definition non-separable. Second, optimality occurs when infinite length sequences are processed. Therefore, accessing part of the data in the classical compression framework, can only be performed in the signal domain but not in the compressed domain.

3. Solutions to the massive random access problem based on classical compression algorithms

Even if incompatible, existing compression algorithms can be adapted to meet the random access constraint. This can be done in one of these three ways:

1. *Send the whole database* as shown in Figure 3. This is the most efficient solution from the compression perspective, but the least efficient from the communication one. Moreover, it might also be infeasible as the required transmission datarate might be larger than the capacity link in many scenario. As an example, sending 80 views compressed with the best known compression algorithm for Multiview images (3D-HEVC)

[1] See the advertisement video at: https://www.youtube.com/watch?v=KrmbMHJQ_u4

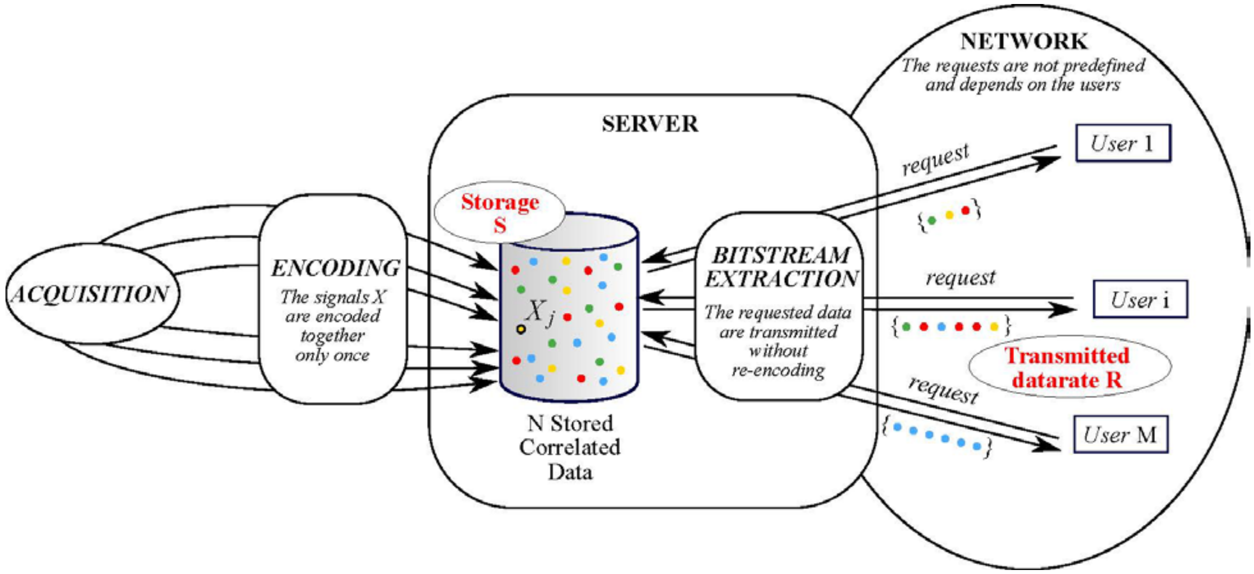

Fig. 2. Random access to a database: the user can choose any subset of the compressed correlated data

requires about 100 Mbit/s [3].

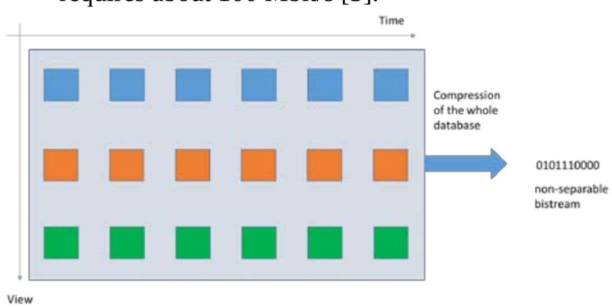

Fig. 3. Compression of the whole database.

2. *Split the database into chunks and compress each chunk separately* as shown in Figure 4. To allow a smooth navigation, the chunk may only contain a single frame.

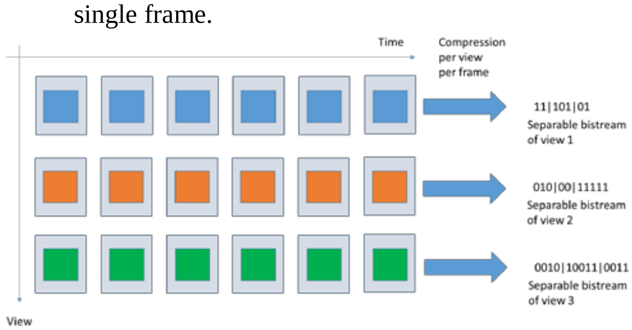

Fig. 4. Compression of separate chunks (a chunk is a frame in FTV).

This scheme is inefficient not only from the storage but also communication perspective. As for the communication inefficiency, this is because the redundancy between the successive frames is not exploited. This requires about *80% more transmission rate*. This loss corresponds to the discrepancy between the inter and intra framerates averaged over the Class A and B sequences of the HEVC corpus. Inter coding was performed with GOPsize 8 and IntraPeriod 64.
As for the storage, neither the inter-frame correlation nor the inter-view correlation is exploited. The former incurs a loss of 80%, while the latter brings an additional cost of about 40%. Indeed, the latter loss corresponds to the discrepancy observed between simulcast (separate encoding of each view) and multiview encoding [3]. Therefore, about *152% more storage* is required.

3. *Decode the whole database and re-encode the request only*. This is optimal from both compression and communication perspectives. But, in case of massive access, this approach is too complex for the server.

Therefore, classical (unconstrained) solutions are either not feasible (case 1 and 3) or suboptimal (case 1 and 2) in a context of massive random access. Therefore, there is a need to design new compression algorithms that take into account the massive random access constraint.

4. Information theoretical bounds for random massive access: the lossless i.i.d. case.

The observation concerning the incompatibility between classical compression algorithms and massive random access (see Section 3) raises the interesting question of the existence of a theoretical tradeoff between massive random access and compression such that a compression algorithm allowing flexibility in the access to the data will always suffer some sub-optimality in terms of compression.

A partial answer to this question is given in our recent work [4], where we showed that there is theoretically no coding performance drop to expect with massive random access, from the communication perspective. The setup considered in [4] is the following. Let $\{X_i(t)\}_{i,t}$ represents the set of frames to be compressed, where $i \in [1,N]$ and $t \in [1,T]$ are the view index and time index, respectively.

For the sake of analysis, a frame (for fixed i and t) is modeled as a random process $X_i(t)$ of infinite length. This infinite length model is indeed a requirement in order to derive information theoretical bounds as compression rates can only be achieved in the asymptotic regime. Note however that the novel video formats (UHD, HDR) tend to produce frames containing a large number of symbols, such that the infinite length assumption is not restrictive.

Now let us assume that at time instants $t-1$ and t, requests are views j and i, respectively. Let S_i and $R_{i,j}$ be the storage and transmission rate (in bits per source symbol) for view i, when the previous request is j. Let us further assume that, given i and t, $X_i(t)$ is an independent and identically distributed (i.i.d.) random process and that lossless compression is performed. We now compare three different schemes.

1. Encoding with **perfect** knowledge of the previous request, as shown in Figure 5. This case corresponds to either a non-interactive scheme, where the user can not choose the view, or to a non-massive random access, where the data are re-encoded at the server upon request. The latter scenario is Scheme 3 described in Section 3. In this case, the necessary and sufficient storage and data rate for lossless compression are:

$$S_i = R_{i,j} = H(X_i(t) \mid X_j(t-1)), \qquad (1)$$

where $H(.|.)$ stands for the conditional entropy. This case is presented for comparison purpose, as the encoding scheme does not satisfy the random massive access constraints.

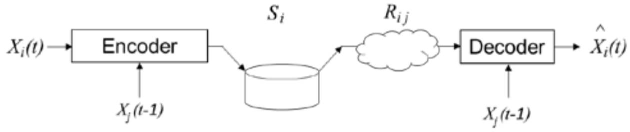

Fig. 5. Encoding with **perfect** knowledge of the previous request at the encoder.

2. Encoding with **partial** knowledge of the previous request, as shown in Figure 6. This case corresponds to an interactive scenario. Indeed, at time instant t, view i is compressed without any explicit reference to a previous request. A trivial upper bound for compression rate is then the unconditional entropy $H(X_i(t))$, since the previous request is not known upon encoding. However, compression can be performed under the assumption that one view among the set $\{X_k(t-1)\}_k$ will be available at the decoder. This allows to reduce both storage and transmission rate from the unconditional entropy $H(X_i(t))$ to the conditional entropy in equation (2):

$$S_i = R_{i,j} = \max_k H(X_i(t) \mid X_k(t-1)), \qquad (2)$$

These rates are necessary and sufficient for lossless source encoding for the scheme depicted in Figure 6, and have been derived in [5]. A practical scheme for video coding has been constructed in [6], based on the insights provided by the achievability part of the theorem in [5].

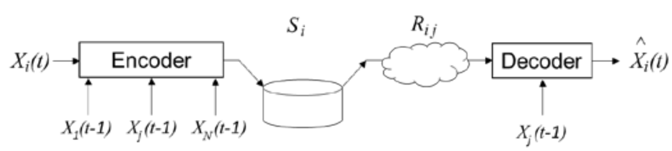

Fig. 6. Encoding with **partial** knowledge of the previous request at the encoder.

3. Encoding with **partial** knowledge of the previous request but sending with **perfect** knowledge, as shown in Figure 7. As in Figure 6, compression occurs without any explicit reference to the previous request. This leads to an interactive scheme and requires the same storage as in (2). However, upon request of a particular view at time instant t, the server knows the previous request of the same user. [4] shows that this information may be used to lower the transmitted datarate from the worst case conditional entropy to the true conditional entropy, see (3), in the case of lossless compression.

$$S_i = \max_k H(X_i(t) \mid X_k(t-1)), \qquad (3)$$

$$R_{i,j} = H(X_i(t) \mid X_j(t-1)). \qquad (4)$$

This results shows that one can efficiently compress the data, while allowing random access to the data. More precisely, it is possible to compress the data in a flexible way such that the server extracts the requested data subset from the compressed bitstream, without a need to decode the whole database and encode the requested subset of data. More surprisingly, the transmission datarate does not suffer any increase even if a flexibility constraint has been added. The constraint incurs an additional cost in the storage only.

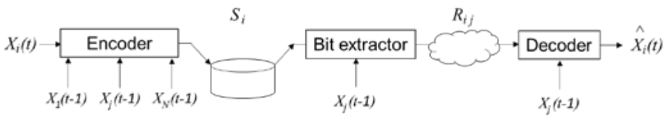

Fig. 7. Mixed: **partial** (**perfect**) knowledge of the previous request at the encoder (sender, respectively).

The necessary and sufficient datarates for the three schemes described in this section are shown in Figure 7.

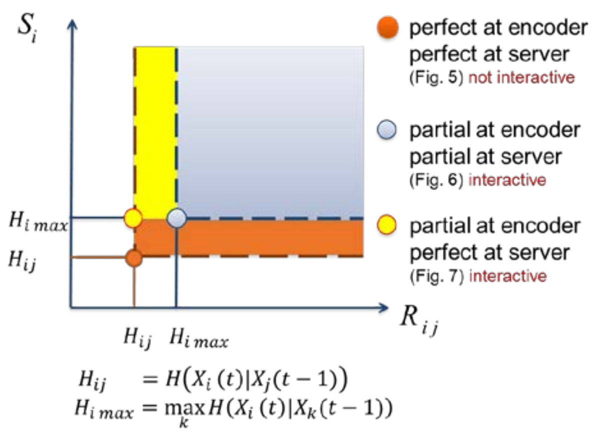

Fig. 8. Comparison of storage and transmission datarates for the schemes with either perfect (Fig. 5), partial (Fig. 6), or mixed (partial at encoder and perfect at transmitter, Fig. 7) knowledge of the previous user request.

5. Conclusion

In this paper, we reviewed some results on data compression for interactive communication well adapted to video services. A novel problem has been defined: massive and random access to subsets of compressed correlated data, and it was shown that FTV can be seen as an instance of this problem. A very surprising result was stated: from the communication prospective, flexible compression with random access to subset of the data achieves the same performance as the very complex scheme that performs whole database decompression and data subset compression upon request. This optimality result is, however, only partial, as it concerns lossless (and not lossy) compression for a simplified model with N correlated sources, where each source is i.i.d. There is now a need to extend this preliminary but very promising result.

Acknowledgement

This work has been partially supported by a French government support granted to the CominLabs excellence laboratory (Project InterCom: interactive communication) and managed by the French National Agency for Research (ANR) in the "Investing for the Future" program under reference Nb. ANR-10-LABX-07-01.

The author would like to thank Elsa Dupraz, Michel Kieffer, and Thomas Maugey for stimulating discussions that helped in recognizing that the FTV scenario in [4] is an instance of the more general problem of massive random access discussed in the present paper.

References

[1] M. Tanimoto, M.P. Tehrani, T. Fujii, T. Yendo, "FTV: Free-viewpoint television," *IEEE Signal Processing Magazine*, vol. 27, no. 6, pp. 555–570, Jul. 2012.

[2] Call for Evidence on Free-Viewpoint Television: Super-Multiview and Free Navigation, ISO/IEC JTC1/SC29/WG11 MPEG2015/N15348, June 2015. http://mpeg.chiariglione.org/sites/default/files/files/standards/parts/docs/w15348_rev.docx

[3] G.J. Sullivan, J.M. Boyce, and Y. Chen and J.-R. Ohm, C.A. Segall, and A. Vetro, "Standardized Extensions of High Efficiency Video Coding (HEVC)," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 1001–1016, 2013.

[4] A. Roumy and T. Maugey, "Universal lossless coding with random user access: the cost of interactivity," in *IEEE International Conference on Image Processing* (ICIP), 2015.

[5] S. C. Draper and E. Martinian, "Compound conditional source coding, Slepian-Wolf list decoding, and applications to media coding," in *IEEE Int. Symp. Inform. Theory*, 2007.

[6] G. Cheung, A. Ortega, and NM. Cheung, "Interactive streaming of stored multiview video using redundant frame structures," *IEEE Trans. on Image Proc*., vol. 3, no. 3, pp. 744–761, Mar. 2011.

Aline Roumy received the Engineering degree from Ecole Nationale Supérieure de l'Electronique et de ses Applications (ENSEA), Cergy, France in 1996, the Master degree in 1997 and the Ph.D. degree in 2000 from the University of Cergy-Pontoise, France. During 2000-2001, she was a research associate at Princeton University, Princeton, NJ. In 2001, she joined Inria, Rennes, France as a research scientist. She has held visiting positions at Eurecom and Berkeley University. Her current research interests include the area of signal and image processing, coding theory and information theory.