Correlation model selection for interactive video communication

Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy

CORRELATION MODEL SELECTION FOR INTERACTIVE VIDEO COMMUNICATION

Navid MAHMOUDIAN BIDGOLI[⋆†] *Thomas MAUGEY*[⋆] *Aline ROUMY*[⋆]

[⋆] INRIA Rennes Bretagne-Atlantique
[†] University of Rennes 1

ABSTRACT

Interactive video communication has been recently proposed for multi-view videos. In this scheme, the server has to store the views as compact as possible, while being able to transmit them independently to the users, who are allowed to navigate interactively among the views, hence requesting a subset of them. To achieve this goal, the compression must be done using a model-based coding in which the correlation between the predicted view generated on the user side and the original view has to be modeled by a statistical distribution. In this paper we propose a framework for lossless fixed-length source coding to select a model among a candidate set of models that incurs the lowest extra rate cost to the system. Moreover, in cases where the depth image is available, we provide a method to estimate the correlation model.

***Index Terms*—** Correlation model selection, Depth-Image-Based rendering, Interactive video, Lossless coding, Fixed-Length interactive source coding

1. INTRODUCTION

With the advancement of technology in broadcasting, Interactive Video Communication (IVC) has attracted considerable attention recently. One possible application in this scenario is Free viewpoint television (FTV) in which users are allowed to view a 3D scene in an interactive manner [1]. Beyond being a highly desirable application, FTV poses a new problem that is Massive Random Access (MRA) to subsets of compressed-correlated large database [2].

For example, during broadcasting a soccer game, numerous high resolution cameras are capturing the scene (*large database*) [3] and at the same time a large number of spectators are watching the game on television or on the Internet (*Massive users*), while each one may want to view the scene from different viewpoints (*Random Access*). Since there exist strong redundancies within the captured views, all the data should be compressed jointly to decrease the storage cost [4]. Unfortunately, with most of the standard algorithms, this implies that the whole database must be sent to the user which would be infeasible in an interactive scenario.

Recently [5, 6] proposed an optimistic results for IVC. They derived the information theoretic bounds and showed that interactivity incurs loss in a priori storage only and not in the transmission rate, i.e. we have to store the data corresponding to the worst user navigation but, after its actual request, one sends him only what is needed [7]. Based on the results obtained in [5], interactivity can be seen as a problem of source coding with Side Information (SI). As a result, unlike predictive coding, in which the residuals within frames are compressed, IVC requires a correlation model for the error and uses a Slepian-Wolf (SW) coder to perform compression. The goal of this paper is to determine the proper correlation model for IVC. Note that this question has been tackled in Distributed Video Coding (DVC) [8] but the methodology is not adapted to IVC. Indeed, in DVC, estimation of the model is done at the decoder knowing the SI only [9], instead IVC performs this estimation at the encoder, knowing the source and all the SIs.

Given a set of candidate models, which one would be better for IVC? In this paper, we propose a method to evaluate the efficiency of the chosen model if we use Fixed-Length (FL) coding for Interactive Source Coding (ISC). From an IVC/ISC point of view, we need a methodology to select the model which provides less average description length of the code than others. After recalling the basic ideas of IVC in section 2, we show that source coding can be used optimally in IVC using concatenation of FL codewords. In section 3 we propose a criterion by making the link between Kullback–Leibler Divergence (KLD) and the extra rate cost that penalizes the compression rate if we use a wrong distribution for the correlation model. In section 4 we provide a framework to evaluate different approximations of the correlation models. In section 5 we introduce a correlation model for the case when the input views are multi-view plus Depth images. Finally, we compare the compression cost for several statistical distributions in section 6 and show that Laplace distribution is the best choice when we consider both compression rate cost and computation time of distribution parameters together.

2. MODEL-BASED CODING IN INTERACTIVE COMMUNICATION

The goal of IVC is to provide coding for Multi-view Videos (MV) while having user interactivity. From the user point of view, interactivity means that the user is able to switch within views to see the scene from different viewpoints. From the communication point of view the server should take into account that the user can use the set of frames in its memory, to predict the newly requested view. The server thus only has to send the information needed to correct this prediction. This correction can not be done using classical predictive (P-frame) approach, since user's navigation is random from the server's point of view. However, the proofs in [5, 6] suggest an alternative approach: model-based coding.

The idea of using model-based coding in video compression was addressed in DVC which relies on Slepian and Wolf's theorem [10] and its later extension of Wyner and Ziv [11]. In this paradigm, part of the frames, called Wyner-Ziv (WZ) frame, are encoded independently by applying a systematic channel code and transmit only a set of parity bits. At the decoder side the prediction called Side Information (SI), which can be seen as a noisy version of the actual WZ frame, is first computed. Then the decoder corrects the SI using the corresponding parity bits [12]. The interest of parity information is

This work was partially supported by the Cominlabs excellence laboratory with funding from the French National Research Agency (ANR-10-LABX-07-01) and by the Brittany Region (Grant No. ARED 9582 InterCOR).

that it is able to correct an error wherever it appears as long as it follows a pre-estimated model.

While in DVC the SI is only available at the decoder and not at the encoder, in IVC a set of possible SIs is also available at the encoder. In other words, in IVC, during encoding a given view X drawn from probability distribution $P_X(x)$, a set of potential views $\Psi(X) = \{\psi_1(X), ..., \psi_K(X)\}$ is available. This set represents all possible views that users are authorized to switch from there to X. Using $\Psi(X)$ the system can provide predictions of X and use the predictions as SIs during the compression and storage. The compressed (stored) stream of the source should be general enough such that any user coming from any views in $\Psi(X)$ can decode the prediction they are able to generate given their specific navigation. Assumption of having this possible set of SIs is not far-fetched, because users normally try to navigate through the scene smoothly and continuously and the system can prevent the user to jump from one view to a completely different view.

In IVC schemes, the SI are usually generated using Depth-Image-Based Rendering (DIBR) [13]. In this algorithm a given view is rendered from the texture image and its corresponding depth image of another view. Therefore, in IVC, the DIBR generated views are used as estimates of other views (SI). Let X represents the current original frame to be compressed/transmitted and $\Upsilon = \{Y_1..Y_i..Y_K\}$ represents the set of all possible SIs for this frame generated from $\Psi(X)$ using DIBR algorithm. For a given user's navigation able to generate SI Y_{i^*}, the region of achievable (rate, storage) pair in lossless transmission is [5]:

$$R \geq H(X|Y_{i^*}) \tag{1}$$

$$S \geq \max_{j \in \{1..i..K\}} H(X|Y_j) \tag{2}$$

where $H(X|Y)$ is the conditional entropy of X given Y, R and S denote the transmission rate and storage respectively.

Equations (1) and (2) hold when the IVC encoder is split into two different phases: compression and transmission. It is possible to show that we can encode the source optimally with a concatenation of FL codewords. Figure 1 (a) shows how to encode using FL coding in IVC/ISC. For simplicity, we assume that 2 SIs are available to encode source $X^n = (X_1, X_2, .., X_n)$ of length n with $H(X|Y_1) \leq H(X|Y_2)$. From the definition of weak joint typicality ([14]) we know that $n_1 = (H(X|Y_1) + \epsilon)n$ bits are sufficient to encode X^n given Y_1^n. Similarly we need $n_2 = (H(X|Y_2) + \epsilon)n$ bits to encode X^n given Y_2^n ($n_2 > n_1$). We denote by $A_\epsilon^n(X|Y_1)$ and $A_\epsilon^n(X|Y_2)$ the sets of typical sequences of X^n given Y_1^n and Y_2^n respectively. In the compression phase, the source is stored with n_2 bits using the concatenation of FL codewords as follows:

- Encode $x^n \in A_\epsilon^n(X|Y_1)$ which are shown by blue triangles and red circles in Figure 1 (a) by assigning a codeword of length n_1 to each sequence. This code set is called $\mathcal{C}_1$.
- Upon encoding $x^n \in A_\epsilon^n(X|Y_2)$:
  - if $x^n \in A_\epsilon^n(X|Y_1) \cap A_\epsilon^n(X|Y_1)$ which are shown by red circles in Figure 1 (a): use exactly the same n_1-length codeword of x^n computed in $\mathcal{C}_1$ and pad the rest $n_2 - n_1$ bits with zero.
  - if $x^n \notin A_\epsilon^n(X|Y_1)$ which are shown by green squares in Figure 1 (a): use the remaining n_2-length codewords to distinguish the remaining typical sequences.

Then in the transmission phase, when the previously requested source and consequently Y_{i^*} is known, the server sends exactly the required number of bits from the stored codeword to recover X from Y_{i^*}.

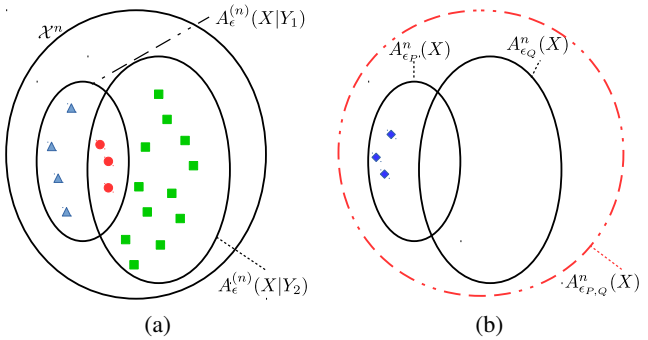

Fig. 1. *a)*: Typical sets of X^n given SIs. $\mathcal{X}^n$ represents the finite alphabet of X^n. The blue triangles show the typical sequences of X^n which are only available in $A_\epsilon^n(X|Y_1)$. The red circles represent the typical sequences of X^n which are available in both $A_\epsilon^n(X|Y_1)$ and $A_\epsilon^n(X|Y_2)$. The green squares show the typical sequences of X^n which are only available in $A_\epsilon^n(X|Y_2)$. *b)*: When we have uncertainty in the true distribution, in order to have vanishing error probability we have to pay extra rate to code also all the typical sequences of $A^n_{\epsilon_P}(X)$ which are not included in $A^n_{\epsilon_Q}(X)$ (shown by blue diamonds) by enlarging the set to $A^n_{\epsilon_{P,Q}}(X)$.

As in DVC, the correlation in IVC between the original source and the SI can be seen as a virtual channel modeled by a statistical distribution. In the following section, we determine the excess rate that occurs when the modeled distribution differs from the true distribution. This will allow us to optimize the modeled distribution such that it minimizes the excess rate.

3. COST OF USING APPROXIMATE DISTRIBUTION

In the IVC model-based encoder, the encoder knows the realizations of X and the set of possible SIs Y_i generated from $\Psi(X)$ for a given X, thus it is able to compute the empirical distributions $P_X(x)$, $P_{Y_i}(y)$ and $P_{X|Y_i}(x|y)$ at the encoder, where x and y are the realizations of X and Y_i. In order to avoid sending all the probabilities to the decoder, we can provide an estimation of true $P_{X|Y_i}(x|y)$ by a statistical distribution $Q_{X|Y_i}(x|y)$ with parameters Θ_i. During the transmission, the server is aware of the user's previous request (i^*), thus it sends parameters Θ_{i^*} and the compressed version of the data to the decoder. The decoder knows the realization of Y_{i^*}, thus it is able to compute the empirical distribution $P_{Y_{i^*}}(y)$. By receiving the parameters Θ_{i^*} it can approximate the joint distribution $P_{X,Y_{i^*}}(x,y)$ by $Q_{X,Y_{i^*}}(x,y) = P_{Y_{i^*}}(y) \; . \; Q_{X|Y_{i^*}}(x|y)$. It performs decoding according to this joint approximate distribution.

If the encoding and decoding processes are performed using the approximate distribution $Q_{X,Y_{i^*}}$, in order to have vanishing error probability we need to pay extra cost for the compression rate. Since the coding of sequences of X^n given SIs Y_i^n is a particular case of encoding X^n without any SI, for simplicity and without loss of generality, we eliminate the notation of Y_i in the following. We denote the set of typical sequences of the true distribution and the approximate distribution with $A^n_{\epsilon_P}(X)$ and $A^n_{\epsilon_Q}(X)$ respectively. Since some of the true typical sequences of $A^n_{\epsilon_P}(X)$ may not be included in $A^n_{\epsilon_Q}(X)$, the probability of error does not vanish to zero because some of the true typical sequences are not encoded (Figure 1 (b)).

In order to have vanishing probability of error, we need to enlarge the set $A^n_{\epsilon_Q}(X)$ to a larger set $A^n_{\epsilon_{P,Q}}(X)$ that contains all the typical sequences of $A^n_{\epsilon_P}(X)$. Using the Asymptotic Equipartition property (AEP), i.e. $-\frac{1}{n}\log p(X^n) \to H_P(X)$ as $n \to \infty$ ([14]), it is possible to show that:

$$-\frac{1}{n}\log Q(X^n) = -\frac{1}{n}\log P(X^n) + \frac{1}{n}\log\frac{P(X^n)}{Q(X^n)} \leq H_P(X) + D_{KL}(P||Q) + 2\epsilon \quad (3)$$

where $H_P(X)$ is the entropy w.r.t. the true distribution P and D_{KL} is the KLD between the distribution P and approximate distribution Q. If we represent

$$A^n_{\epsilon_{P,Q}}(X) = \{x^n : |-\frac{1}{n}\log Q(X^n) - H_P(X) - D_{KL}(P||Q)| < \epsilon\} \quad (4)$$

Using equations (3) and (4) together with AEP we can derive $A^n_{\epsilon_P}(X) \subset A^n_{\epsilon_{P,Q}}(X)$.

Therefore, if the sequences x^n generated with the distribution P, are encoded with the wrong distribution Q, such that all sequences of typical set $A^n_{\epsilon_{P,Q}}(X)$ have different indexes, then all true typical sequences in $A^n_{\epsilon_P}(X)$ will have different indexes and there will be no error. Using equation (3) the rate required to code all $A^n_{\epsilon_{P,Q}}(X)$ is $H_P(X) + D_{KL}(P||Q)$. In other words, there is an extra rate of $D_{KL}(P||Q)$.

To generalize the extra rate while we have SI, one can use a fixed-length source per realization of the SI sequence, then the extra rate is:

$$D_{KL}(P_{X|Y=y}||Q_{X|Y=y}) = \sum_x P_{X|Y}(x|y)\log\frac{P_{X|Y}(x|y)}{Q_{X|Y}(x|y)}. \quad (5)$$

Computing the rate w.r.t. all realizations y, i.e. $\sum_y P_Y(y)$ we have:

$$D_{KL}(P(X,Y)||Q(X,Y)) = \sum_y P_Y(y)\sum_x P_{X|Y}(x|y)\log\frac{P_{X|Y}(x|y).P_Y(y)}{Q_{X|Y}(x|y).P_Y(y)} \quad (6)$$

Therefore the compression rate cost ΔR_i of using approximate distribution is equal to the KLD from joint approximate distribution to empirical joint distribution using Equation (6).

4. FRAMEWORK FOR CORRELATION MODEL PERFORMANCE EVALUATION IN IVC

The DIBR algorithm produces two different regions for predicted pixels: 1- the dis-occluded region, in which DIBR can not provide any information of the pixel and 2- the predicted pixels. Therefore, the virtual channel which is used for the coder can be modeled as cascade of two channels as shown in Figure 2. More precisely, the correlations of signals X and Y_i is modeled as if X is first passed through a channel with $p(w|x)$ which presents the predicted pixels[1]. Then the signal W is erased (erased symbol corresponds to dis-occluded area). Using the mutual information equation of cascade

[1]The model $p(w|x)$ together with $p(x)$ completely determine the joint distribution $p(x,w)$. However, in SW theorem since $p(w)$ is available at the decoder, the distribution which is used is $p(x|w).p(w)$. Therefore, practically $p(x|w)$ is approximated by $q(x|w)$ and is sent to the decoder.

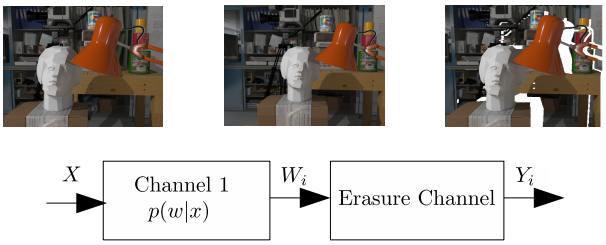

Fig. 2. *top left*: original view X, *top middle*: a reference view available in $\Psi(X)$, *top right*: predicted view. white area corresponds to dis-occluded area. *bottom image*: cascade virtual channel model used in our framework

channels: $I(X;Y) = (1-\alpha)\ I(X;W)$ in which α is the probability of erasure [14], we have the following formula to compute $H(X|Y_i)$:

$$H(X|Y_i) = \frac{N_p}{N_t}.H(X|W_i) + \frac{N_t - N_p}{N_t}.H(X) \quad (7)$$

where N_p and N_t stand for the number of predicted pixels and total number of pixels in the image respectively.

Since using the wrong distribution for correlation model will only affect the predicted pixels in DIBR, the total extra rate cost per pixel for a given view i, when using model $Q(X|W_i)$ instead of the true distribution $P(X|W_i)$, is equal to:

$$\Delta R_i = \frac{N_p}{N_t}\ .\ D_{KL}(P(X,W_i)||Q(X,W_i)) \quad (8)$$

We now express this cost in the framework of IVC. We first assume that the users are navigating among the different views of a static multi-view dataset. To simulate the interactivity we have separated the framework into 2 phases: storage & transmission. For each view X we define a set of authorized views $\Psi(X)$, that users are allowed to switch from them to X. To measure the extra storage cost, ΔS_X, if we use wrong (approximate) distribution for the correlation model, first we try to predict the view from each view available in $\Psi(X)$ to provide possible SIs Υ. Then we sort these SIs w.r.t. their empirical conditional entropies $H(X|Y_i)$ and take the extra rate cost related to the one which has higher conditional entropy than others as storage cost (since we are storing data in the worst case assumption). Therefore we have:

$$\Delta S_X = \max_j [H(X|Y_j) + \Delta R_j] - \max_j H(X|Y_j) \quad (9)$$

For the transmission phase, we generate users' navigation paths by randomly selecting a view among the set of available views (from a uniform distribution) as the first frame which is displayed to the user. Then for the successive frames X_i we select the next view by allowing the user to select a view randomly among predefined $\Psi(X_i)$ for each X_i. Here we allow the user to navigate to the preceding neighboring view or stay at the current view or go to the next view by giving uniform probability to these views and select one of them. We repeat the experiment several times to simulate different user navigation and compute the average extra rate ΔR per pixel as the cost that we penalize if we use approximate distribution. For each of the candidate model, we compute the amount of extra rate and extra storage that we should spend. The model which provides less cost in term of storage and rate among all navigation (in average per pixel) is selected to model the correlation.

5. ADDITIVE MODEL FOR DIBR

The parameters of the distribution $P(x|w)$ need to be sent per realization of w. Since this needs a large number of parameters, we need to look for a relation between original frame X and the DIBR predicted image W to avoid sending to many parameters per realization of w. In this test we have used the Tsukuba dataset [15]. The dataset consists of 1800 synthetic stereo pairs with ground truth depth images in which the stereo camera is navigating through a static scene. We have also tested our experiments on the real dataset Ballet [16]. This dataset includes a sequence of 100 images captured from 8 cameras.

We test the linear dependency between the two variables X and the corresponding DIBR predicted view W (inside the non-occluded area) by using sample Pearson Correlation Coefficient (PCC). The PCC indicates the degree of linear dependence between the variables and has a values between -1 and +1, where ± 1 represents perfect linear relationship [17]. In the test, several views have been randomly selected and synthesized using the neighboring views. Figure 3 shows the mean and 95% confidence interval (CI) on the mean of PCC values as a function of distance between reference view and the requested view. This shows that the more we get closer to the original view, the relation between the SI and original view tends to be more linear. Thus we conclude that the assumption of having an additive model, which is also widely used in classical video coding literature like DVC [9], is reasonable also in DIBR context for IVC (although it is not perfect). The amount of difference w.r.t. 1 can be modeled as measurement error [18]. This error could have different causes: pixel quantization, Lambertian effect, interpolation of pixel intensities, etc. during rendering of synthesized image in DIBR. Finally, the correlation between original frame and DIBR images in the non-occluded area can be described as an additive model:

$$X = W + Z \tag{10}$$

where Z is the error between original source and side information. Now, the goal in correlation model selection is to approximate the distribution of Z, which represents $p(x|w)$ by a statistical distribution $q(x|w)$.

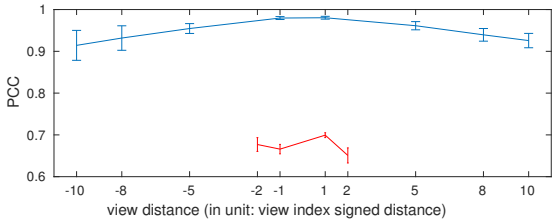

Fig. 3. Mean and its 95% CI of PCC values. The horizontal axis shows how much the reference view is far from the original view (in terms of number of views available in the database) to reconstruct the view. The blue curve is for Tsukuba and the red curve is for Ballet dataset

6. EXPERIMENTAL RESULTS

Using the additive model, we can compute the distribution of error easily by computing the difference of intensity values between the ground truth image and synthesized image inside the non-occluded area and provide a set of possible correlation models for the distribution of error between X and Y to see which one fits better to the error distribution. The set of candidates that we have provided for our experiments are: Laplace distribution, Normal distribution, Gaussian Mixture Models (GMM) and Erasure Channel (EC) Model. The EC model means that the synthesized DIBR images can exactly recover the ground truth image, thus we do not have any error inside the non-occluded area and we only have to recover the dis-occluded pixels [5].

Table 1. Extra rate and extra storage cost in bit per pixel. Inf means infinity

	Tsukuba		**Ballet**	
	ΔS	ΔR	ΔS	ΔR
laplace	0.634	0.610	0.552	0.3
normal	1.43	1.401	0.996	0.555
GMM	0.211	0.2	0.364	0.19
EC	Inf	Inf	Inf	Inf

For the parameter estimation of GMM, Expectation-Maximization (EM) algorithm is used. Since the optimal number of components are unknown, we tuned the number of components by using Akaike's and Baysian Information Criterion (AIC & BIC) [19]. Lower AIC or BIC values indicate better fitting models. Based on our experiments GMM model with 3 components is sufficient to fit to the error distribution. For the Laplace and normal distribution maximum-likelihood estimation is used to estimate the distribution parameters. EC model is actually a deterministic model which has probability equal to one at zero error (and zero probability elsewhere).

We implement the framework presented in section 4 for grayscale color values of Tsukuba and Ballet dataset. For Tsukuba we allowed the user to navigate through neighboring views in left and right images. Therefore, we have 6 possible navigations (either stay at the current view or switch to the other five views). For Ballet, each frame is coded separately and users are allowed to navigate through neighboring view among the 8 views available in each frame. Results are summarized in Table 1. As it can be seen from the results of EC model, it is not possible to perform lossless coding while we are assuming that DIBR predicted pixels can perfectly predict the requested view. Laplace distribution which is also widely used in predictive coding and DVC provides lower cost than normal distribution. GMM is the best, but parameter estimation of GMM takes much more time to compute than Laplace, as GMM requires an iterative EM algorithm, while Laplace parameters is estimated by simple arithmetic operations.

7. CONCLUSION

In this paper we showed that fixed-length coding suits for interactive source coding and that it is possible to encode optimally the sources in the sense that achievable (rate, storage) region proposed in [5, 6] can be obtained. Then we questioned the choice of the model in IVC and proposed a new criterion to select the model by computing the extra rate cost in the case of using fixed-length source coding when using an approximation for the correlation model. For the case that depth images are available along with the texture images, we proposed a model and parameter estimation framework for DIBR predicted views. We evaluated several statistical distributions and concluded that it is not possible to code a view losslessly by completely relying on DIBR predicted pixels and only using erasure channel to recover dis-occluded area.

8. REFERENCES

[1] M. Tanimoto, "Overview of free viewpoint television," *Signal Processing: Image Communication*, vol. 21, no. 6, pp. 454 – 461, 2006, Special issue on multi-view image processing and its application in image-based rendering.

[2] A. Roumy, "An information theoretical problem in interactive multi-view video services," *IEEE Communications Society Multimedia Communications Technical Committee (ComSoc MMTC) E-Letter*, vol. 11, no. 2, pp. 11–16, March 2016.

[3] "Call for evidence on free-viewpoint television: Super-multiview and free navigation," Tech. Rep., Poland, Warsaw, June 2015, http://mpeg.chiariglione.org/standards/exploration/free-viewpoint-television-ftv/call-evidence-free-viewpoint-television-super.

[4] G. J. Sullivan, J. M. Boyce, Y. Chen, 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, Dec 2013.

[5] A. Roumy and T. Maugey, "Universal lossless coding with random user access: The cost of interactivity," in *2015 IEEE International Conference on Image Processing (ICIP)*, Qubec, Canada, Sept 2015, pp. 1870–1874.

[6] E. Dupraz, T. Maugey, A. Roumy, and M. Kieffer, "Rate-storage regions for Massive Random Access," *arXiv preprint arXiv:1612.07163v1 [cs.IT]*, Dec. 2016.

[7] G. Cheung, A. Ortega, and N. M. Cheung, "Interactive streaming of stored multiview video using redundant frame structures," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 744–761, March 2011.

[8] T. Maugey, J. Gauthier, B. Pesquet-Popescu, and C. Guillemot, "Using an exponential power model for Wyner Ziv video coding," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, Texas, U.S.A, March 2010, pp. 2338–2341.

[9] C. Brites and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain wyner-ziv video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 9, pp. 1177–1190, Sept 2008.

[10] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, Jul 1973.

[11] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, Jan 1976.

[12] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71–83, Jan 2005.

[13] G. Petrazzuoli, M. Cagnazzo, F. Dufaux, and B. Pesquet-Popescu, "Using distributed source coding and depth image based rendering to improve interactive multiview video access," in *2011 18th IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium, Sept 2011, pp. 597–600.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*, A Wiley-Interscience publication. Wiley, 2006.

[15] M. Peris, S. Martull, A. Maki, Y. Ohkawa, and K. Fukui, "Towards a simulation driven stereo vision system," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Tsukuba Science City, JAPAN, Nov 2012, pp. 1038–1042.

[16] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," in *ACM SIGGRAPH*. August 2004, vol. 23, p. 600608, Association for Computing Machinery, Inc.

[17] M. M Mukaka, "A guide to appropriate use of correlation coefficient in medical research," *Malawi Medical Journal*, vol. 24, no. 3, pp. 69–71, 2012.

[18] D. P. Francis, A. J. Coats, and D. G. Gibson, "How high can a correlation coefficient be? effects of limited reproducibility of common cardiological measures," *International journal of cardiology*, vol. 69, no. 2, pp. 185–189, 1999.

[19] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.