Evaluation of side information effectiveness in distributed video coding

Thomas Maugey, *Member, IEEE*, Jérôme Gauthier, Marco Cagnazzo, *Senior Member, IEEE* and Béatrice Pesquet-Popescu, *Fellow, IEEE*

***Abstract*—The rate-distortion performance of a distributed video coding system strongly depends on the characteristics of the side information. One could naively think that the best side information is the one with the largest PSNR with respect to the original corresponding image. However, previous works have shown that this is not always the case and a reduction of the side information MSE does not always translate into better rate-distortion performance for the complete system. The scope of this paper is to explore a set of metrics other than the PSNR and explicitly designed to classify the side information with respect to its impact on the end-to-end compression performance. A first contribution is to define an experimental framework that can be used to meaningfully compare different metrics for side information evaluation. As a second contribution, our analysis allows to understand why in some cases PSNR-based metrics provide a fairly reliable estimation of the side information quality, while in other cases they do not. This analysis also allows us to introduce a set of new metrics that are better adapted for side information effectiveness evaluation, and that are based on a suitable power of the absolute difference between side information and the original image, or on the Hamming distance between the respective transform coefficients. Besides their theoretical interest, these new metrics can also improve the rate-distortion performance of some distributed video coding systems such as the hash-based ones. We observe improvement up to 74 % rate reduction in a simple study case.**

***Index Terms*—Distributed video coding, side information, quality evaluation**

I. Introduction

Based on theoretical results obtained in the 70's, [1], [2], distributed video coding (DVC) [3], [4] schemes allow to avoid the complex motion estimation process at the encoder and to shift most of its computational load to the decoder, without affecting the rate-distortion (RD) performance bound with respect to the case of joint encoding, provided that joint decoding of correlated frames can be performed. The interest in distributed coding has been revitalized when the first practical implementations of DVC appeared in 2002 [5], [6]. They decrease the computational burden of the encoder, and also allow to encode multiple sources without need of communication among them, which may be essential in some applications such as wireless sensor networks [7], multi-camera systems, multiterminal networks [8], [9] and medical imaging [10]. Moreover, DVC techniques have recently found a new exciting application in the context of interactive multi-view video streaming [11], [12] since they allow to encode a new image without knowing which images are available to the decoder (this happens when the user is allowed to interactively change the displayed view).

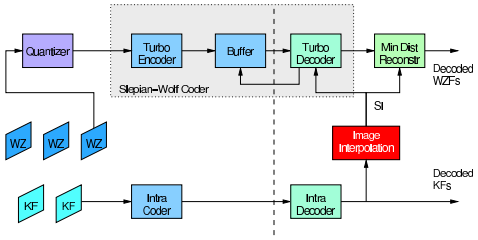


Fig. 1. DISCOVER DVC architecture.

In this work we consider one of the most popular schemes for DVC, firstly proposed in [6]. This scheme has become very popular also thanks to its adoption within the DISCOVER project [13], [14]. Very briefly, the principle of DISCOVER architecture (represented in Fig. 1) relies on the so-called Wyner-Ziv (WZ) coding of a subset of the video frames. Decoding a WZ frame consists in two steps: first an estimation of it, called *side information* (SI) is generated at the receiver; second, the side information is corrected using parity bits from a suitable channel code, such as a turbo code [15]–[17]. One of the major issues of such a scheme is to produce a "good" SI [18]–[20]. Indeed, this has a large impact on the performance, since the more the side information is similar to the original frame (or Wyner-Ziv Frame, WZF), the less correction bits will be sent to decode it. On the other hand, if the SI is a bad estimation of the WZF, many correction bits will be necessary to recover it, inconveniently increasing the coding rate. The effectiveness of the SI is usually measured in terms of Mean Square Error, MSE (or equivalently in terms of PSNR) with respect to the original WZF, and it is observed that an improvement of this metric often corresponds to an improvement in end-to-end RD performances.

Yet, this is not always true: sometimes, a higher SI PSNR does not translate into better end-to-end RD performance. For example, Kubasov [21] presented some cases where one side information has a better PSNR than another, but after decoding, the second one achieves a better reconstruction for a lower bit rate. We present also an artificial example, using the *foreman* sequence (CIF, frame number 10). We generate two images, playing the role of side information: the first one is obtained as a motion-compensated interpolation of the previous and following frames (Fig. 2 (a)); the second image is obtained by adding uniform random noise to the original

one (Fig. 2 (b)). Then each of the two images is used as side information in the DISCOVER decoder and then we compute the number of parity bits that the decoder needs in order to correct the two images. Results are presented in Fig. 2 and show that, despite an almost identical PSNR, the two images used as side information have quite different decoding performance: the first one needs only 137 kbits to achieve a PSNR of 39.29 dB, while for the second, even after sending more than 192 kbits, the PSNR only attains 35.40 dB.

Since the side information quality is a major issue in DVC, we would like to understand whether in practical conditions the PSNR could fail in differentiating between a good and a bad SI, whether in this case more reliable metrics exist, and whether we can foretell which metric is best suited to a given situation. The scope of this paper is to answer to these interesting questions, which at the authors best knowledge have not yet been fully addressed in the scientific literature, the only partial exception being the work by Kubasov [21]. We propose new metrics based on a suitable power of the absolute difference between the original image and the side information, or on the Hamming distance between the corresponding transform coefficients. The conclusions of our work may be useful for future studies aiming at improving the general DVC performance. For example, using the appropriate distortion measure may help to obtain a better description of the error, *e.g.*, a more accurate noise correlation at the Slepian-Wolf decoder. Moreover, the proposed metrics can be used to improve the rate-distortion performance of a DVC system: we show in this work a simple but realistic scheme that is able to exploit the new metrics. The basic idea is that knowing a reliable quality measure between the estimation and the original frame leads to sensible improvements in the hash-based DVC schemes [22] where some information about the original WZ frame is sent to the decoder for enhancing the motion interpolation and the Slepian-Wolf decoding. There are many hash-based systems in the scientific literature [10], [22]–[38] and potentially all of them could take advantage from the new metrics. More generally, the proposed metrics can be used when both the original image and the side information (or an estimation of it) are available. Besides the hash-based systems, also the Witsenhausen-Wyner video coding systems [39] and the DVC with shared encoder/decoder complexity [40]–[42] could benefit from our study.

The remainder of this paper has the following structure. First we introduce a novel framework allowing to compare different SI metrics. Then we propose our own measures and we show through some experiments that they are more reliable than the PSNR. This is followed by a discussion highlighting the conditions which favor one metric or the other. Finally, we depict one potential application of the new metrics and the corresponding rate-distortion improvements, and we conclude the paper by outlining some possible future developments of this work.

II. METHODOLOGY FOR METRIC COMPARISON

In this section, we introduce a new methodology to evaluate the reliability of side information metrics. In opposition to the usual video quality metrics, which have to be compared with human subjective experiments for reliability assessment, side information quality measures should be correlated with the rate-distortion performance of the codec. In particular, a good metric should be able to classify an estimation of a given image with respect to the resulting RD performance obtained when the estimation is used as side information in the DVC scheme.

Our framework is the following. The problem is how to compare two images that are to be used as side information of a third one. The latter is typically a WZ frame in a DVC system, while the former two are typically produced using motion-compensated interpolation or extrapolation. The two SIs are compared by running the DVC system twice, using the first or the second image as side information. The difference in global RD performance allows us to rank the two images according to their effectiveness as side information. This process is detailed in the following.

We compare two estimations of a given WZ image by evaluating the average coding bitrate variation resulting by the use of one or the other estimation. The coding bit-rate is the end-to-end bit-rate, comprehensive of KF and WZ coding rates. The average is computed over four RD points using the well-known Bjontegaard metric [43]. This is arguably the most meaningful comparison between two estimations, and then we use it as a kind of "ground-truth": practical metrics should be as correlated as possible to this one.

The first element in our framework is the association between an image, used as SI for an original frame, and an RD curve. Let us define a decoder function dec which takes two images as input: the first, denoted as I_0, is the original frame, and the second, denoted as I_1, is the side information. The original image is first compressed using 4 given quantization steps, then channel-decoded using I_1 as side information. This results in a set of four rate-distortion points, $[(R_i^1, d_i^1)_{i=1\ldots4}]$, which is the output of the dec function.

The next step consists in associating an average bit-rate variation to a couple of RD point sets. This is easily achieved by using the Bjontegaard metric. The Bjontegaard metric allows to compute the average bit-rate reduction of a set of four RD points with respect to another set of four RD points, provided that the rate intervals are not disjoint. More precisely, if $(R_i^1, d_i^1)_{i=1\ldots4}$ and $(R_i^2, d_i^2)_{i=1\ldots4}$ are two sets of 4 RD points, we can use the Bjontegaard metric to compute:

$$\Delta R = \mathrm{bjm}((R_i^1, d_i^1)_{i=1\ldots4}, (R_i^2, d_i^2)_{i=1\ldots4})$$

If ΔR is positive, the curve sampled at points $(R_i^1, d_i^1)_{i=1\ldots4}$ can be considered as "better" than the second curve, since the latter requires a positive increment in bit-rate in order to achieve the same average quality in the intersection of the two coding rate intervals.

Now we can introduce our reference quality measure, referred to as r. It compares two SI images, I_1 and I_2, *i.e.* two estimations of a given reference image I_0, as follows:

$$r(I_0, I_1, I_2) = \mathrm{bjm}\left[dec(I_0, I_1), dec(I_0, I_2)\right]$$

The reference metric is our "ground-truth" that can be used to evaluate the quality of a generic metric $m(\cdot,\cdot)$, for instance the

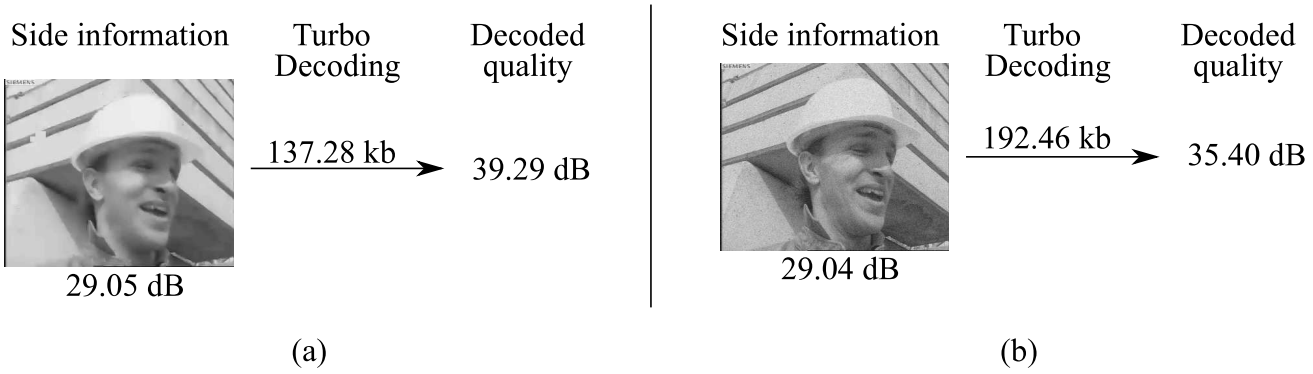


Fig. 2. The two side informations and their decoding results : (a) DISCOVER and (b) uniform random noise. Similar initial PSNR lead to two very different decoded quality.

PSNR, by introducing a confidence criterion. For a given set of images I_1, I_2 and I_0, the metric m respects the confidence criterion if and only if $r(I_0, I_1, I_2)$ and $m(I_0, I_1) - m(I_0, I_2)$ have the same sign. This is easy to interpret. The r function is positive if the side information I_1 leads to better RD performance than the side information I_2. In this case, a metric should assign a better score to I_1 than to I_2, otherwise it is not able to correctly point out the best estimation of the image I_0. Moreover it could be shown that a metric satisfying this "same-sign" condition, also satisfies a stronger condition on the values: in other words, large values for r imply large differences between corresponding metric values. Finally we observe that the confidence criterion can be used only provided that the rate intervals produced by the dec function for the two images to be compared are not disjoint. However this condition is very often satisfied in practice, as we observed in our experiments.

III. Alternative metrics for side information evaluation

A. Family of SIQ$_a$ metrics

An MSE-related metric such as the PSNR depends on the second power of the error. As shown later, this can be ineffective for the quality evaluation of side information, so we introduce a class of metrics where the power of the error is a parameter. We consider the class of metrics SIQ$_a$, with $a \in (0, 1]$, defined as

$$\mathrm{SIQ}_a(I_0, I) = 10 \log_{10} \frac{255^2}{\frac{1}{NM} \sum_{\mathbf{p}} |I_0(\mathbf{p}) - I(\mathbf{p})|^a}$$

where I and I_0 are respectively the side information and the original (reference) images, of size $N \times M$ pixels. These metrics are defined as a strictly decreasing function of the a-norm of the error $e(\mathbf{p}) = |I_0(\mathbf{p}) - I(\mathbf{p})|$. In its turns, the a-norm $||e||_a^a = \sum_{\mathbf{p}} |e(\mathbf{p})|^a$ is a concave function of e in $\mathbb{R}^{N \times M}$ (strictly concave if $a < 1$). As a consequence the proposed metrics are convex (strictly convex if $a < 1$) functions of the error image. This property is interesting since at at the receiver side, the side information is used to produce the log-likelihood ratio (LLR) used by the channel decoder for the bit planes reconstruction. In particular, if X is a decoded bit from a certain bit plane of a transform coefficient, and p is the estimated probability that X is equal to zero, the LLR is defined as $LLR = \log \frac{p}{1-p}$. Now, this function is strictly increasing, and for reasonable values of p (i.e. $p < 0.5$), it is concave: it varies fast for low values of p and slower for high values. This behavior is therefore similar to the a-norms, which for $a \in (0, 1]$ are concave as well; on the contrary the quadratic norm (the MSE) is convex, and then tends to emphasize large errors.

In conclusion, we consider the idea that the SIQ$_a$ family could better reproduce the characteristics of the decoder. This idea was only explored for the case $a = 1/2$ by Kubasov [21]. One of the contribution of the present work is to extend the analysis to other possible values of a. While it is obviously impossible to test all the values for a, we consider the following cases:

- $a = 1$, which corresponds to the mean-absolute-error (MAE or ℓ_1-norm), commonly used in signal processing. Here, we give up strict concavity, it thus constitutes an interesting limit case to study.
- $a = \frac{1}{2}$, for which Kubasov has given some initial results [21].
- $a = \frac{1}{3}$, in this case we try to further enhance the difference between small error values.

B. A metric based on the Hamming distance

As explained before, the SIQ$_a$ metrics are mostly inspired from the LLR calculated at the WZ decoder. Since the best value for the a parameter could mismatch the correlation noise model implemented in the WZ decoder, we introduce a new metric, which intends to fit the decoder structure. In a typical DVC system, the WZ frame quantized transform coefficients are represented by bit planes. We adopt the same quantization matrix as in DISCOVER [13], [14], which states a number of quantization bins for every band, for 8 levels of quality. All the bit planes are then encoded successively (the most significant first). At the receiver side, the correlation noise (*i.e.*, the difference between SI and original WZF) is modeled as a zero-mean Laplace random process, and the standard deviation is estimated band-per-band, following an approach very popular in literature [44].

For each bit plane, the decoder receives a first set of parity bits and starts the channel decoding algorithm of the corresponding bit plane of the side information (*i.e.* the decoder uses the received parity bits to correct the side information bit planes). The channel decoder uses the so-called extrinsic information (*i.e.*, the probability that a bit is 0 or 1 given the side information) to decode the first bit plane. Then, the residual error probability is estimated. If it is larger than a given threshold (a common value is 10^{-3}), the decoder requests one more set of parity bits (referred to as a "chunk" [13]), and restarts the decoding until the estimated error probability drops under the threshold. When this happens, the current bitplane is considered to have been correctly decoded. Using relatively large chunks allows to reduce the number of parity bits requests and of channel decoder iterations, thus improving the complexity and the rate requirements of the system [45], [46].

Each following bitplane is decoded taking into account the previously decoded ones: the extrinsic information is now the probability that a bit is 0 or 1 given the SI *and* the previous bit planes. These probabilities are easily computed given the distribution of the correlation noise, as described for example in [4] or in [47]. The propagation of extrinsic information through bit planes makes it easier to correct individual bit errors if they correspond to a small difference among the original and the SI coefficient. However each wrong bit may trigger a new chunk of parity bits to be sent, with potentially a large increase of the number of transmitted bits, even if the conditional probability would indicate a low parity rate.

We observe that the usual metrics such as the PSNR do not take into account this decoding process; they compute a sum of differences in the *spatial* domain, emphasizing large error contributions because of the second power in the MSE. Moreover, a pixel-wise difference in the spatial domain is not exactly what the channel decoder will try to correct: the error correction is performed in the transform domain and on quantized coefficients. Therefore we introduce a metric that simply counts the number of wrong bits in the side information, *i.e.* we use a Hamming distance. In particular, we consider the case of a decoder working bit-plane by bit-plane in the transform domain, *i.e.*, exactly on the same data used to compute the parity bits. In this way we also account for the transmission of parity bits in chunk, since each wrong bit in a quantized transform coefficient may trigger the sending of a new chunk.

In order to establish the notation, let $\overline{I}$ and $\overline{I}_0$ be respectively the transformed and quantized versions (with a given quantization step QI) of the SI and of the reference image, N_{bits} the number of bits used to represent $\overline{I}$ and $\overline{I}_0$, and $d_H(\overline{I}, \overline{I}_0)$ the Hamming distance between these two representations. We define the *Hamming Side Information Quality (HSIQ)* metric as:

$$\text{HSIQ}(I_0, I) = 10 \log_{10} \left(\frac{N_{\text{bits}}}{d_H(\overline{I}, \overline{I}_0)} \right), \qquad (1)$$

which is the relative number of wrong bits expressed in a logarithmic scale. The advantage of this metric is that it is more related to the channel decoding process, in the sense that it takes into account the actual data to be corrected (the transformed and quantized coefficients) and, at least in part, the fact that parity bits are transmitted in chunks. Therefore, in contrast to the PSNR and SIQ_a, the HSIQ provides something closer to a rate measure than to a distortion one, and also closer to what the turbo decoder does when establishing an error probability threshold. Nevertheless, it is still a quality metric in the sense that it associates a real, non-negative number to a couple of images. Moreover, another advantage of the HSIQ metric is that it explicitly takes into account the quantization and the transform, while the SIQ only compares the original WZF and the SI. We observe here that pixel-domain DVC systems that directly perform the channel encoding of WZF pixels have been proposed in literature, but they have worse performance than transform-domain systems [44], [48]. These advantages are somewhat counterbalanced by the fact that the Hamming distance is not the best suited for the Min-Distance reconstruction stage. Moreover, in some cases, very close values of transform coefficients may have a relatively large Hamming distance. In this case, estimating the parity bit-rate is rather difficult: the bit plane propagation of extrinsic information suggests that few bits are needed; but since the Hamming distance is large, actually the probability of triggering a new chunk is high. In summary, HSIQ may over-estimate the parity bit rate, but considering only the coefficient difference does not takes into account the chunk effect. However, experimental results clearly show that the benefits of the HSIQ metric usually surpass the drawbacks, therefore this metric has the potential for improving real system performances, as shown in Section VI.

IV. Metric validation

A. Experimental setup

In this section we compare the metrics for SI quality in the proposed framework. In order to perform these tests, we firstly generate some sets of images that will be used as side information. For this experiment, we generate a more realistic data set than the one used in Fig 2. To this end, we observe that the most popular methods for SI generation are based on motion estimation and compensation. More precisely, the side information is virtually always generated by motion compensation of one or two reference images, which in turn are compressed versions (via transformation and quantization) of the original ones.

In conclusion, a realistic SI is potentially affected by two kinds of degradations: errors from quantization (*i.e.* compression) of reference frames, and errors from incorrectly estimated trajectories (*i.e.* motion vectors) [49]. Therefore, in the following, we consider two different sets of images to be used as SI. Both of them are constructed by altering a motion-compensated image generated by a motion estimation algorithm which disposes of the original Wyner-Ziv frame. More precisely, we randomly modify some of the estimated vectors, this emulates the unavoidable errors occurring in the motion estimation process. These errors mainly result from the fact that in SI generation the actual WZF is not available. The only difference between the two sets is that in this first one all the reference images are quantized with the same quantization

step, while in the second these steps are different. Results are given respectively in Sec. IV-B and IV-C.

This method for image generation aims at simulating actual estimations and moreover allows regulating the balance between errors from quantization and errors from bad vectors (blocking errors). It could be argued that randomly placed blocking artifacts are not realistic since motion errors are correlated with the characteristics of object textures, movements, borders, *etc*. However, this observation holds mainly for motion-compensated prediction; in the case of motion-compensated temporal interpolation, the induced errors are much less regular, since we do not have the original WZF, the estimation is based on trajectory interpolation and errors related to occlusions cannot be easily avoided. Therefore trajectory errors and occlusions sum up with motion estimation inaccuracies, giving place to unpredictable errors. Another relevant case of irregular errors in SI is when multi-view DVC is considered [50], [51]. In this case the SI is produced from merging two independent images, resulting in a non-stationary distribution of errors. This results in more frequent and less correlated block errors, as very often observed in practice.

B. SI database with reference frames at the same QP

In this first experiment, all reference frames used to build the side information database were compressed with the same quantization step. We generate an image (*i.e.* an estimation) for each of the first $N = 100$ frames of *breakdancer*, *book arrival*, and *outdoor* sequences, and for each quantization step (reference frames are quantized with H.264 Intra at QP 31, 37 and 40 [52]). We have chosen to work with these three sequences as they present very different characteristics (motion activity, camera positions and alignment, indoor/outdoor). Once the entire database is created, for each frame of each sequence at each quantization level, all the SIs $\widehat{I_k^{si}}$ are decoded at 4 different quantization points that correspond to 4 quantization indexes (QI) in the DISCOVER quantization matrix. This matrix gives the number of bits allocated to each band for 8 QI, where 1 is low bit rate and 8 is high resolution. In other words, we compute $\forall k \in \{1, \ldots, N\}$,

$$(R_i^k, d_i^k)_{i=1\ldots4} = dec(I_0, \widehat{I_k^{si}}),$$

where I_0 is the original image we are considering. Note that, in order to compare the RD curves using the method defined in Section II, we need that the intersection of the curve supports is not empty. Therefore the QI and QP parameters were set such that to generate $(R_i^k; d_i^k)_{i=1\ldots4}$ with large support intersections for all $k \in \{1...N\}$. Finally, for each metric m to be tested, and for each couple of image indexes $k, \ell \in \{1, \ldots, N\}$ we compute the percentage of cases in which the following equivalence is satisfied:

$$m(\widehat{I_k^{si}}) \leq m(\widehat{I_\ell^{si}}) \Leftrightarrow r(I_0, \widehat{I_k^{si}}, \widehat{I_\ell^{si}}) \leq 0.$$

The corresponding results are presented in Table I. The percentages correspond to the success rate of the metric: for example, if one estimation has a higher PSNR than a second one, in more than 90% of cases it will also have better rate-distortion performance. We remark that the PSNR, the SIQ$_a$ and the HSIQ obtain similar results, *i.e.* the five metrics are all equally reliable for this type of database. Therefore, when the reference frames have all the same quantization step and the estimation is obtained by motion estimation/compensation, the PSNR metric is a reasonable choice for assessing side information effectiveness. We also observe that the SIQ$_a$ metrics have no clear ranking among them: any of them outperforms the other two in some cases. As we observed before, we interpret this as the fact that the best value for a is related to the correlation noise model.

TABLE I
PERCENTAGE OF VALIDATION OF THE CONFIDENCE CRITERION FOR SEVERAL SEQUENCES AND SEVERAL QUANTIZATION STEPS FOR THE REFERENCE FRAMES USED TO GENERATE THE SI DATABASES.

	breakdancer			*outdoor*			*book arrival*			
QP	31	37	40	31	37	40	31	37	40	Avg
PSNR	**90.1**	**90.4**	**87.3**	**89.9**	92.0	**90.5**	90.9	92.0	90.0	90.5
SIQ$_1$	89.9	89.3	87.0	89.1	93.1	89.0	**92.2**	92.0	**91**.1	90.5
SIQ$_{\frac{1}{2}}$	89.7	89.4	86.0	89.0	93.0	88.9	91.6	92.2	90.6	90.3
SIQ$_{\frac{1}{3}}$	89.0	87.5	87.0	88.3	92.7	88.7	90.0	92.5	89.8	89.6
HSIQ	89.1	87.7	86.5	88.9	**93.3**	89.0	90.4	**92.9**	90.1	90.0

TABLE II
PERCENTAGE OF VALIDATION OF THE CONFIDENCE CRITERION FOR SEVERAL SEQUENCES AND FOR THE SECOND DATABASE.

Sequence	*breakdancer*	*outdoor*	*book arrival*	Average
PSNR	66.09	61.65	74.99	67.57
SIQ$_1$	**92.27**	91.90	**95.66**	**93.27**
SIQ$_{\frac{1}{2}}$	90.83	91.88	95.27	92.66
SIQ$_{\frac{1}{3}}$	90.53	91.81	94.79	92.37
HSIQ	90.84	**93.82**	93.68	92.78

C. SI database with reference frames at variable QP

In these experiments, the reference frames are quantized with different quantization steps (QP = 31, 34, 37, 40). However, the number of blocks affected by vector errors is reduced when the reference frames are roughly quantized, and incremented when they are better represented. The rationale behind this constraint is to avoid freely varying errors, which could result in pathological configurations. However we observe explicitly that we obtained very similar results even when removing this constraint on the introduced error.

Once the database is created, we perform the same kind of experiments as those presented in the previous Section, obtaining the statistics presented in Table II. In this case we notice a remarkable difference among metrics: while SIQ$_a$ and HSIQ are still reliable, this is no longer the case for the PSNR, which estimates the right quality order between two side informations in only two cases over three on the average. On the other hand, HSIQ and SIQ are correct in more than 90% of the cases for the same images. We also remark that SIQ$_1$ performs better than other SIQ$_a$ metrics. This could be ascribed to the fact that using the absolute value in (III-A) does not amplify small errors (as it would be for $a < 1$) or large errors (as for $a > 1$); or to the fact that it better matches the Laplacian law for the correlation error. These results highlight that even in realistic cases, the PSNR could fail in determining the best side information, and therefore the introduction of new metrics could be advantageous. We observe explicitly that Wyner and Ziv proved that the reconstruction quality

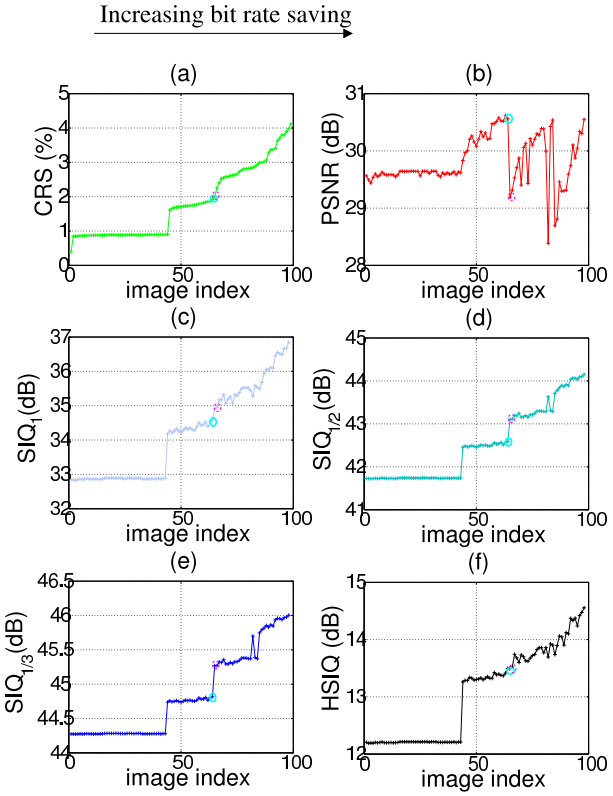


Fig. 3. Metric values as a function of the number of estimations for frame 3 of *outdoor* sequence (512×384). The estimations are sorted in the decoded performance growing order (real quality). Solid cyan and dotted purple circles indicate the examples illustrated in Fig. 4 and 5.

is a decreasing function of the side information MSE [2]. However, we observe that the Wyner-Ziv theorem only applies to the case of memoryless Gaussian sources. Even though the theorem has been extended to other statistical laws [53], [54], the hypothesis of stationary signal must hold in order for these extensions to apply. However, since a generic video signal has statistics that do not perfectly match with theorem hypotheses, and moreover it is not stationary, we conclude that the MSE may be not the best SI quality estimator, and this would not contradict the well-established Wyner-Ziv theory.

In summary, these two experiments allows us to conclude that the HSIQ metric performs very well in all situations (but it is more complex to compute). The SIQ_a metrics have somewhat similar performance, with SIQ_1 being slightly better for the second database. We observe that SIQ_1 has also the best results in the average, but for some cases it is outperformed by $\text{SIQ}_{1/2}$ or $\text{SIQ}_{1/3}$. We conclude that there is not an absolute best value of a, since it depends on the correlation noise model, but the choice $a = 1$ is usually quite robust. Finally the PSNR is reliable only when the errors are homogeneous.

V. Discussion

In the following, we try to understand why the PSNR is sometimes wrong in SI quality estimation by investigating some representative examples from the second set of images. For the sake of conciseness, in the following we only show the results for a specific image; however we observed the same behavior for all the images of all the video sequences in the database.

Let us consider a specific original frame, namely the third one of the *outdoor* sequence, and refer to it as I_0. We consider the $N = 100$ estimations for this image from the second database, and we sort them in the decoding performances growing order; we also number them accordingly: if $i \leq j$ are two natural numbers between 1 and 100, we have $\text{bjm}(dec(I_0, \widehat{I_i^{si}}), dec(I_0, \widehat{I_j^{si}})) \leq 0$, and we refer to the sort index j as *image index*. For the $j-th$ side information, $\widehat{I_j^{si}}$, we calculate its *cumulative rate saving (CRS)* with $\text{CRS}(\widehat{I_1^{si}}) = 0$ and:

$$\forall j > 1, \qquad \text{CRS}(\widehat{I_j^{si}}) = bjm\left((R_i^j, d_i^j)_{i=1\ldots4}, (R_i^{j-1}, d_i^{j-1})_{i=1\ldots4}\right) + \text{CRS}(\widehat{I_{j-1}^{si}}). \tag{2}$$

This is equivalent to say that the CRS value for the j^{th} side information corresponds to the rate saving with respect to the side information of the lowest quality ($\widehat{I_1^{si}}$). In Fig. 3 (a), we plot the CRS values. Between the worst and the best side information, we observe a CRS difference of approximately 4%, which allows to say that the database is significantly wide from the point of view of decoding quality. Fig. 3 (b)-(f) present the plots of respectively the PSNR, SIQ_1, $\text{SIQ}_{\frac{1}{2}}$, $\text{SIQ}_{\frac{1}{3}}$ and HSIQ against the image index. A good metric m should preserve the increasing order, *i.e.*, if $i < j$ then it should be $m(\widehat{I_i^{si}}) \leq m(\widehat{I_j^{si}})$, and thus we should observe increasing curves in Fig. 3 (b)-(f). When this inequality is not verified, the metric m is not able to correctly discriminate between two estimations and the amplitude of the resulting negative gap quantifies this malfunction. We can conclude that HSIQ and the SIQ_a metrics have a satisfying behavior, while the PSNR plot shows many consecutive estimations with a negative variation (sometimes of more than 1 dB) instead of an improvement. Moreover, one can notice that SIQ_1 (Fig. 3 (c)) and HSIQ (Fig. 3 (f)) are the two metrics which evolve the most similarly to the CRS behavior (Fig. 3 (a)), especially in the second part of the plot. One of the possible reason for the slightly-non-monotonic behavior of the HSIQ is the mismatch with the Min-Distance reconstruction stage. However the results shown here and in the following section let us conclude that the introduction HSIQ brings more gains than losses.

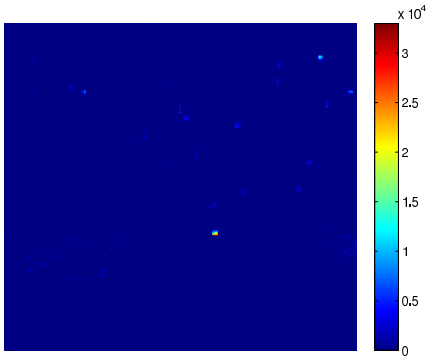
Let us focus now on a specific couple of estimation images, namely $\widehat{I_{64}^{si}}$ and $\widehat{I_{65}^{si}}$. The associated values of the metrics are highlighted respectively by solid cyan and dotted purple circles in Fig. 3. Even though $\text{CRS}(\widehat{I_{64}^{si}}) < \text{CRS}(\widehat{I_{65}^{si}})$, the PSNR gives the opposite order, *i.e.* $\text{PSNR}(\widehat{I_{64}^{si}}) > \text{PSNR}(\widehat{I_{65}^{si}})$, and with a very wide gap of more than 1.4 dB. SIQ_a and HSIQ predict the right order for these estimations.

The two estimations both present distortions coming from key frame quantization and from the motion errors, but not in the same proportion. The blocking artifact errors are more relevant in $\widehat{I_{65}^{si}}$ while $\widehat{I_{64}^{si}}$ has a higher quantization noise (more precisely, the reference frames were quantized with QP = 37

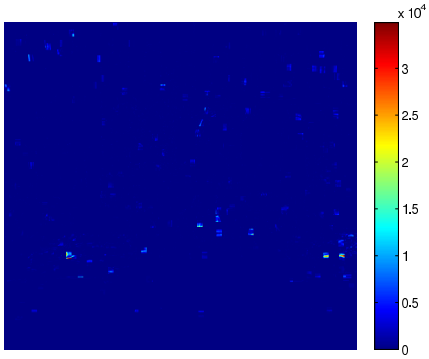
for the latter image and 34 for the former image). Since the PSNR is based on the MSE, we show in Fig. 4 (a) and (b) the square error image. One can see that the square error is heavily affected by high errors associated to blocking artifacts, while the smaller but more widespread quantization error has a smaller impact. This explains why the PSNR of $\widehat{I_{65}^{si}}$ is much smaller than the PSNR of $\widehat{I_{64}^{si}}$.

On the contrary, if we compute the pixel-wise $\mathrm{SIQ}_{\frac{1}{2}}$, for example, Fig. 5 (a) and (b), we can perceive that the quantization error is taken into account almost as much as the block errors, which explains why $\mathrm{SIQ}_{\frac{1}{2}}(\widehat{I_{64}^{si}}) < \mathrm{SIQ}_{\frac{1}{2}}(\widehat{I_{65}^{si}})$. Indeed, the PSNR metric mainly accounts for high amplitude errors present in the image, which in turns mainly come from blocking artifacts. On the other hand, the different SIQ_a metrics are based on a lower error power a, so that they can better take into account smaller but more frequent errors. On its side, the HSIQ, which also fits well the CRS, offers a different compromise, because it considers the magnitude error information only when an error in a bit plane propagates to the next bit plane.

In conclusion, when the number of block errors are of the same order of magnitude for two estimations, the PSNR gives a reliable quality value, but if the error types are different, *e.g.* a locally highly concentrated error (like the blocking artifacts due to the motion compensation) or a diluted error (like the quantization noise), the PSNR would penalize the highly concentrated errors whereas they would be more easily corrected by the turbo decoder.



(a) $(\widehat{I_{64}^{si}} - I)^2$, 30.63 dB (PSNR)



(b) $(\widehat{I_{65}^{si}} - I)^2$, 29.20 dB (PSNR)

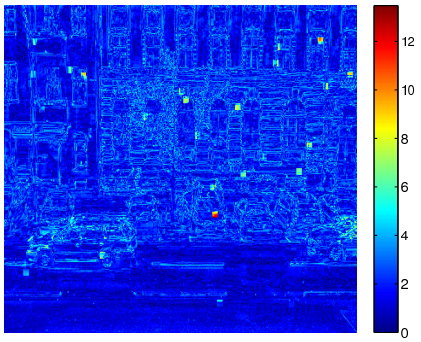
Fig. 4. Pixel domain error image associated to the PSNR measure, for the estimations 64 and 65 (respectively solid cyan and dotted purple circles in Fig. 3).

VI. Hash-based scheme improvement

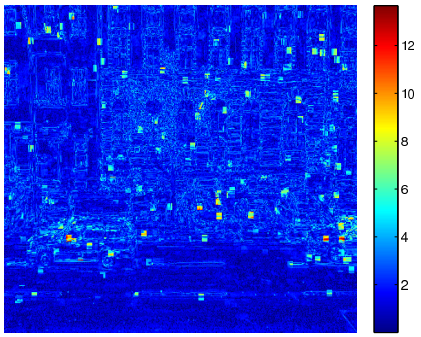
Based on the conclusions of Sec. IV and V, we provide here some experiments which prove that, in addition to being more reliable, the proposed metrics have some interesting practical applications. More precisely, we use the novel metrics, and especially the HSIQ, to modify and improve a hash-based DVC scheme. This kind of schemes [23], [26], [32] is based on the idea that the WZ frame estimation error obtained at the decoder is not stationary and some regions present a very important distortion because of high motion, occlusions, etc. For these regions, hash-based schemes send some additional information (*e.g.* a quantized version of the corresponding macroblock transform coefficients), called *hash*, in order to improve the WZ frame estimation at the receiver. Therefore, a hash-based scheme has to solve two main problems: a) for which regions hash information has to be sent and b) how to use efficiently this information at the receiver. The first problem cannot be solved in the classical DVC framework, where the encoder is completely unaware of the estimation available to the decoder. Therefore the hash based schemes slightly relax the DVC paradigm: the WZF encoder has access to key frames and can produce a very rough estimation of the side information available to the decoder; however it cannot produce the same SI as the decoder. In summary, strictly speaking a hash-based coding scheme should be classified as a low-complexity encoder rather than a DVC system, but it shares with the latter much of its structure.

Coming back to our system, state-of-the-art hash-based schemes produce a very simple estimation of the side information at the encoder, for example by averaging the reference KFs. Then, the MSE (or equivalently the PSNR) between the original frame and this rough estimation is computed, and the hash is sent for the blocks with low PSNR. At this point it is easy to conceive an application of our metrics, using them instead of the PSNR for selecting the blocks to be coded with the hash.

The proposed metrics can also be used to improve the solution to problem b), *i.e.* how to efficiently use the hash at the decoder side. In the state-of-the-art scheme, in the block matching algorithm performed at the decoder in order to find out the object trajectories, instead of calculating the error between the compensated blocks, the transmitted hash is compared with the average of the two compensated blocks. This comparison is usually done with MSE. We can then replace the MSE with the HSIQ, which is much better suited to this end, since the hash are transmitted as quantized DCT coefficients. We explicitly observe that all the candidate blocks are tested with the new metric, i.e. we do not use HSIQ to perform an early stop of the matching algorithm. Our work in the DCT domain constitutes a good complement to those who have already tested the Hamming distance in a wavelet-based hash DVC scheme [55].

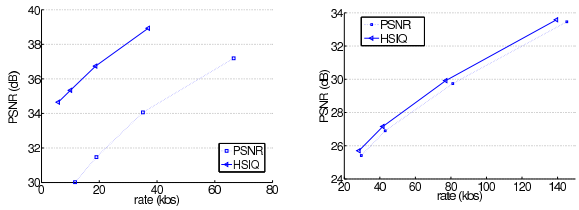


(a) $|\widehat{I^{si}_{64}} - I|^{\frac{1}{2}}$, 42.61 dB ($SIQ_{\frac{1}{2}}$)



(b) $|\widehat{I^{si}_{65}} - I|^{\frac{1}{2}}$, 43.12 dB ($SIQ_{\frac{1}{2}}$)

Fig. 5. Pixel domain error image associated to the $SIQ_{\frac{1}{2}}$ measure, for the estimations 64 and 65 (respectively solid cyan and dotted purple circles in Fig. 3).



(a) *container*, CIF (b) *football*, CIF

Fig. 6. Influence of metrics in the design of hash-based scheme (hash selection + side information generation) in case of 5% of hash transmitted.

In order to validate this approach we have implemented a simple hash-based system, which uses MSE at the encoder and the decoder. This scheme, mainly inspired from [32], is an extension of the well-known reference DVC scheme DISCOVER, and constitutes the reference scheme that we improve by changing the metrics. The hash blocks are quantized DCT coefficients and are selected and used as described before. For the sake of simplicity, we do not test a hash-based scheme with all the proposed metrics. We have experimentally selected the ones fitting best to the problem. In an hash-based system, the encoder has typically a very small available computational power. For this reason, since HSIQ and SIQ_1 obtain similar performance, at the encoder we use SIQ_1 which is much less complex. On the contrary, at the decoder there is typically much more available computational power, so we have chosen the HSIQ, since it performs a better block matching for finding out the object trajectories. In order to perform a more complete analysis, we have also considered the case when the decoder uses SIQ_1. Rate-Distortion curves are shown in Fig. 6 for two CIF sequences. In Tab. III we present the Bjontegaard gains (in rate and in distortion) when using the new metrics with respect to the reference PSNR metric. Both proposed metrics achieves very significant improvements: the bitrate reduction ranges from 4.3 % to 30.5 % for SIQ_1, with an average rate reduction equal to 11.9 %. HSIQ is even better: for any single test sequence it achieves larger rate reductions than SIQ_1, with an average of 26 %. PSNR improvents are in line with rate reductions, with an average gain of 0.5 dB for SIQ_1 and 1.5 dB for HSIQ. These experimental results confirm that the HSIQ is a good metric of SI quality, thanks to the fact that it takes into account the quantized transform coefficients, while SIQ_1 does not.

However, both metrics have large gains with respect to PSNR. This is explained by the fact that the frame estimations that are used as side information have similar characteristics with those of the second database in Sec. IV-C, *i.e.* they contain different types of errors (quantization, motion errors) that the PSNR does not take into account with appropriate weight. It should also be noted that in some cases (*e.g. container*), the gain can be huge because of the very specific motion activity (fixed background and rapid motion in the foreground).

Finally, we report here a comparison with some reference schemes, giving the average Bjontegaard rate variation on the six sequences. First, we compared the HSIQ hash-based system with DISCOVER. We expect to achieve some gain with respect to DISCOVER, since the latter does perform a strictly distributed encoding, while hash-based systems have access to both key frames and Wyner-Ziv frames. The experiments confirmed this intuition: the hash-based systems provide an average rate reduction (in the sense of Bjontegaard metrics) of 20.4 %. As far as the complexity is concerned, it is not increased very much in the proposed system with respect to DISCOVER. The complexity increase of the encoder is comparable to the one of any hash-based system; the decoder is more complex than the one of DISCOVER, but with the same order of magnitude. Finally we observe that the decoder complexity is seldom an issue in DVC or in hash-based systems.

We have also compared the hash-based schemes with H.264/AVC in different configuration: INTRA-only, Zero-motion (IPIP), Inter (IPIP). In general, DISCOVER has better performance than H.264/AVC Intra, it is comparable to H.264/AVC Zero-motion and worse the H.264/AVC Inter. We have found that the proposed system using HSIQ has an average rate reduction of 33 % and 4.5 % respectively with respect to H.264/AVC Intra and H.264/AVC Zero-motion, while it increases the bit-rate of only 8.4 % with respect to H.264/AVC Inter.

TABLE III
BJONTEGAARD GAINS OF PROPOSED SCHEMES W.R.T. THE REFERENCE PSNR-BASED ONE

Sequence	rate saving (%)		PSNR improvement (dB)	
	SIQ_1	HSIQ	SIQ_1	HSIQ
city	5.2	12.4	0.24	0.62
coastguard	10.4	23.9	0.43	1.05
container	30.5	74.7	1.41	5.19
eric	13.3	20.8	0.62	1.13
football	4.3	8.3	0.22	0.42
foreman	7.5	16.6	0.33	0.81

VII. CONCLUSION

In this paper we have studied the problem of measuring the quality of the side information in DVC. In the literature, this evaluation is generally performed with the PSNR. Our work firstly demonstrated that the PSNR metric, despite its acceptable behavior in some situations, presents important drawbacks for comparing the qualities of two estimations when different types of errors (high-value spatially concentrated errors or low-value widespread errors) occur. We also propose some alternative metrics: the family of SIQ_a metrics and the Hamming-based HSIQ measure, which provide satisfying results in the different tested settings, since the new metrics better correspond to the turbo decoding behavior (the LLR for the SIQ_a, and the transform plus quantization structure for the HSIQ). From the results shown in this paper, we observe that the SIQ_a and HSIQ have similar performances. Both of these metrics seem to be well suited for measuring the side information quality in a DVC context and are more versatile than the PSNR. We also show a possible practical application of the proposed metrics: they can be used to improve the rate-distortion performance of low-complexity hash-based schemes, with a bit-rate reduction up to 74 % with respect to the case where the PSNR is used to assess the SI quality.

REFERENCES

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Inform. Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[2] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the receiver," *IEEE Trans. on Inform. Theory*, vol. 22, pp. 1–11, Jan. 1976.

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

[4] C. Guillemot, F. Pereira, L. Torres, T. Ebrahimi, R. Leonardi, and J. Ostermann, "Distributed monoview and multiview video coding: Basics, problems and recent advances," *IEEE Signal Processing Magazine*, pp. 67–76, Sep. 2007, spec. Iss. on Sig. Process. for Multiterminal Commun. Syst.

[5] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," in *Proc. of the 40th Allerton Conference on Communication, Control and Computing*, Allerton, IL, USA, Oct. 2002.

[6] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. Data Compression Conference*, Snowbird, UT, USA, Apr. 2002, pp. 252–261.

[7] R. Puri, A. Majumdar, P. Ishwar, and K. Ramchandran, "Distributed video coding in wireless sensor networks," *Proc. IEEE*, pp. 94–106, 2006.

[8] V. Stankovic, A. Liveris, Z. Xiong, and C. Georghiades, "On code design for the slepian-wolf problem and lossless multiterminal networks," *IEEE Trans. on Inform. Theory*, vol. 52, pp. 1495 – 1507, 2006.

[9] Y. Yang, V. Stankovic, Z. Xiong, and W. Zhao, "On multiterminal source code design," *IEEE Trans. on Inform. Theory*, vol. 54, pp. 2278 – 2302, 2008.

[10] N. Deligiannis, F. Verbist, J. Barbarien, J. Slowack, R. Van De Walle, P. Schelkens, and A. Munteanu, "Distributed coding of endoscopic video," in *Proc. Int. Conf. on Image Processing*, Brussels, Belgium, Sep. 2011.

[11] G. Cheung, A. Ortega, and N. Cheung, "Interactive streaming of stored multiview video using redundant frame structures," *IEEE Trans. on Image Proc.*, vol. 3, pp. 744–761, 2011.

[12] 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 *Proc. Int. Conf. on Image Processing*, vol. 1, Bruxelles, Belgium, Sep. 2011, pp. 605–608.

[13] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: Architecture, techniques and evaluation," in *Picture Coding Symposium (PCS)*, Lisbon, Portugal, Nov. 2007.

[14] DISCOVER-website, "www.discoverdvc.org," 2005.

[15] W. E. Ryan, "A turbo code tutorial," New Mexico state University, Tech. Rep., 1997.

[16] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Trans. on Commun.*, vol. 44, pp. 1261–1271, 1996.

[17] C. Brites and F. Pereira, "Encoder rate control for transform domain Wyner-Ziv video coding," in *Proc. Int. Conf. on Image Processing*, San Antonio, TX, USA, Oct. 2007.

[18] J. Ascenso, C. Brites, and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," in *EURASIP Conf. on Speech and Image Process., Multimedia Commun. and Serv.*, Smolenice, Slovak Republic, Jun. 2005.

[19] S. Ye, M. Ouaret, F. Dufaux, and T. Ebrahimi, "Improved side information generation for distributed video coding by exploiting spatial and temporal correlations," *EURASIP J. on Image and Video Proc.*, vol. 2009, pp. 223–228, 2009.

[20] M. Ouaret, F. Dufaux, and T. Ebrahimi, "Iterative multiview side information for enhanced reconstruction in distributed video coding," *EURASIP J. on Image and Video Proc.*, 2009, special issue on Distributed Video Coding.

[21] D. Kubasov, "Codage de sources distribuées : nouveaux outils et application à la compression vidéo," Ph.D. dissertation, Université de Rennes 1, IRISA, Rennes, France, Dec. 2008.

[22] C. Yaacoub, J. Farah, and B. Pesquet-Popescu, "Improving hash-based Wyner-Ziv video coding using genetic algorithms," in *Int. Mobile Multimedia Commun. Conf. (MOBIMEDIA)*, London, UK, Sep. 2009.

[23] A. Aaron, S. Rane, and B. Girod, "Wyner-Ziv video coding with hash-based motion compensation at the receiver," in *Proc. Int. Conf. on Image Processing*, vol. 5, Singapore, Oct. 2004, pp. 3097–3100.

[24] L. Kang and C. Lu, "Low-complexity Wyner-Ziv video coding based on robust media hashing," in *Int. Workshop on Multimedia Sig. Proc.*, Victoria, Bc, Canada, Oct. 2006.

[25] A. Aaron and B. Girod, "Wyner-Ziv video coding with low encoder complexity," in *Picture Coding Symposium (PCS)*, Lisbon, Portugal, Nov. 2007.

[26] J. Ascenso and F. Pereira, "Adaptive hash-based side information exploitation for efficient Wyner-Ziv video coding," in *Proc. Int. Conf. on Image Processing*, vol. 3, San Antonio, TX, Oct. 2007, pp. 29–32.

[27] M. Tagliasacchi and S. Tubaro, "Hash-based motion modeling in Wyner-Ziv video coding," in *IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, Honolulu, HawaI, 2007.

[28] Y. Zhang and C. Zhu, "Full search of side-information in distributed video coding," in *Fourth International Conference on Image and Graphics (ICIG)*, Chengdu, Sichuan, China, Aug. 2007.

[29] L. Kang and C. Lu, "Multi-view distributed video coding with low-complexity inter-sensor communication over wireless video sensor networks," in *Proc. Int. Conf. on Image Processing*, San Antonio, Texas, USA, Sep. 2007.

[30] T. N. Dinh, G. Lee, J.-Y. Chang, and H.-J. Cho, "Side information generation using extra information in distributed video coding," in *IEEE International Symposium on Signal Processing and Information Technology*, Cairo, Egypt, Dec. 2007.

[31] Z. Li, A. Wang, and H. Wang, "Distributed video coding based on conditional entropy hash," in *International Conference on Computational Aspects of Social Networks (CASoN)*, Taiyuan, China, Dec. 2010.

[32] T. Maugey, C. Yaacoub, J. Farah, M. Cagnazzo, and B. Pesquet-Popescu, "Side information enhancement using an adaptive hash-based genetic algorithm in a Wyner-Ziv context," in *Int. Workshop on Multimedia Sig. Proc.*, Saint-Malo, France, Oct. 2010.

[33] N. Deligiannis, M. Jacobs, F. Verbist, J. Slowack, J. Barbarien, R. Van De Walle, P. Schelkens, and A. Munteanu, "Efficient hash-driven Wyner-Ziv video coding for visual sensors," in *Fifth ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, Ghent, Belgium, Aug. 2011.

[34] F. Verbist, N. Deligiannis, M. Jacobs, J. Barbarien, A. Munteanu, P. Schelkens, and J. Cornelis, "Maximum likelihood estimation for the generation of side information in distributed video coding," in *18th International Conference on Systems, Signals and Image Processing (IWSSIP)*, Sarajevo, Bosnia and Herzegovina, Jun. 2011.

[35] F. Verbist, N. Deligiannis, M. Jacobs, J. Barbarien, P. Schelkens, J. Cornelis, and A. Munteanu, "A probabilistic predictor for side information generation in distributed video coding," in *17th International Conference on Digital Signal Processing (DSP)*, Corfu, Greece, Jul. 2011.

[36] N. Deligiannis, J. Barbarien, M. Jacobs, A. Munteanu, A. Skodras, and P. Schelkens, "Side-information-dependent correlation channel estimation in hash-based distributed video coding," *IEEE Trans. on Image Proc.*, vol. 2012, pp. 1934–1949, 2012.

[37] F. Verbist, N. Deligiannis, S. M. Satti, A. Munteanu, and P. Schelkens, "Iterative wyner-ziv decoding and successive side-information refinement in feedback channel-free hash-based distributed video coding," in *Proceedings SPIE 8499 Applications of Digital Image Processing XXXV*, San Diego, CA, Aug. 2012.

[38] N. Deligiannis, F. Verbist, J. Slowack, R. Van De Walle, P. Schelkens, and A. Munteanu, "Joint successive correlation estimation and side information refinement in distributed video coding," in *Proc. Eur. Sig. and Image Proc. Conference*, Bucarest, Romania, Aug. 2012.

[39] M. Guo, Z. Xiong, F. Wu, D. Zhao, X. Ji, and W. Gao, "Witsenhausen-Wyner video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, pp. 1049ï¿ $\frac{1}{2}$–1060, 2011.

[40] T. Clerckx, A. Munteanu, J. Cornelis, and P. Schelkens, "Distributed video coding with shared encoder/decoder complexity," in *Proc. Int. Conf. on Image Processing*, San Antonio, Texas, USA, Sep. 2007.

[41] D. Chul Keun Kim; Min Geon Kim; Suh, "Distributed video coding encoder/decoder complexity sharing method by phase motion estimation algorithm," in *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PacRim)*, Victoria, B.C., Canada, Aug. 2009.

[42] X. H. B. Jeon, "Flexible complexity control solution for transform domain Wyner-Ziv video coding," *IEEE Transactions on Broadcasting*, vol. 58, no. 2, pp. 209–220, Jun. 2012.

[43] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," 13th VCEG-M33 Meeting, Austin, TX, USA, Tech. Rep., Apr. 2001.

[44] C. Brites and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain Wyner–Ziv video coding," *IEEE Trans. on Circ. and Syst. for Video Technology*, vol. 18, no. 9, pp. 1177–1190, Sep. 2008.

[45] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "A distributed source coding technique for correlated images using turbo-codes," *IEEE Communications Letters*, vol. 6, no. 9, pp. 379–381, 2002.

[46] J. Li, Z. Tu, and R. S. Blum, "Slepian-Wolf coding for nonuniform sources using turbo codes," in *Proc. of the Data Compression Conference*. IEEE, 2004, pp. 312–321.

[47] Y. Vatis, S. Klomp, and J. Ostermann, "Inverse bit plane decoding order for turbo code based distributed video coding," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 2. IEEE, 2007, pp. II–1.

[48] C. Brites, J. Ascenso, and F. Pereira, "Improving transform domain Wyner-Ziv video coding performance," in *Proc. Int. Conf. on Acoust., Speech and Sig. Proc.*, vol. 2, Toulouse, France, May 2006, pp. 525–528.

[49] F. Dufaux and J. Konrad, "Efficient, robust, and fast global motion estimation for video coding," *IEEE Trans. on Image Proc.*, vol. 9, pp. 497–501, 2000.

[50] T. Maugey, W. Miled, M. Cagnazzo, and B. Pesquet-Popescu, "Fusion schemes for multiview distributed video coding," in *Proc. Eur. Sig. and Image Proc. Conference*, Glasgow, Scotland, Aug. 2009.

[51] M. Ouaret, F. Dufaux, and T. Ebrahimi, "Fusion-based multiview distributed video coding," in *Proc. ACM Int. Workshop on Video Surveillance and Sensor Networks*, Santa Barbara, California, USA, Oct. 2006.

[52] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Circ. and Syst. for Video Technology*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[53] S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 1181–1203, 2003.

[54] R. Zamir, "The rate loss in the Wyner Ziv problem," *IEEE Trans. on Inform. Theory*, vol. 42, pp. 2073–2084, 1996.

[55] A. Wang, Y. Zhao, Z. Zhu, and H. Wang, "Scalable distributed video coding based on block SW-SPIHT," *Chinese Optics Letters*, vol. 5, pp. 336–339, 2007.