



ELSEVIER

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Symmetric distributed coding of stereo omnidirectional images[☆]

Vijayaraghavan Thirumalai^{*}, Ivana Tomic, Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL), Signal Processing Laboratory (LTS4), Lausanne 1015, Switzerland

ARTICLE INFO

Article history:

Received 29 March 2008

Accepted 4 April 2008

Keywords:

Distributed source coding

Omnidirectional images

LDPC codes

Camera networks

Stereo imaging

ABSTRACT

This paper presents a distributed coding scheme for the representation of 3D scenes captured by a pair of omnidirectional cameras with equivalent computational resources and transmission capabilities. The images are captured at different viewpoints and are encoded independently. A joint decoder exploits the correlation between images for improved decoding quality. The distributed coding is built on the multi-resolution representation of spherical images, whose information is split into two partitions. The encoder then transmits one partition after entropy coding, as well as the syndrome bits resulting from the channel encoding of the other partition. The joint decoder exploits the intra-view correlation by predicting one partition from the other partition. At the same time, it exploits the inter-view correlation using block-based disparity estimation between images from different cameras. Experiments demonstrate that the distributed coding solution performs better than a scheme where images are handled independently. Furthermore, the coding rate stays balanced between the different cameras, which permits to avoid hierarchical relations between vision sensors in camera networks.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Camera networks find widespread usage in several applications that rely on the effective representation of scenes or the analysis of 3D information. These networks normally consist of several cameras distributed in the scene, and pose several problems like the coding of multi-view images, the reconstruction of the 3D structure from multiple views, or the multi-view object recognition, for example. This paper focuses on the compression of multi-view images and particularly stereo omnidirectional images. The images captured from different viewpoints are usually correlated, which permits to reduce the coding rate by exploiting efficiently the redundancy between the different views. Instead of joint encoding that unfortunately requires communication between cameras, we rely

on the Slepian–Wolf theorem [30] and design a distributed coding scheme where images are encoded independently, but decoded jointly in order to exploit the correlation between the images, as illustrated in Fig. 1.

Most of the research carried out on distributed coding for multi-view images or videos propose solutions based on coding with side information. In this case, one of the cameras is chosen as the primary source and its output is encoded independently. The other cameras represent secondary sources whose rate can be drastically reduced if the joint decoder uses the primary source as side information. Such a coding scheme obviously does not balance the transmission rate between the encoders. However, it is often interesting in practice to rather avoid hierarchical relations between sensors and to distribute the coding and transmission cost equally among the sensors. In this paper, we therefore concentrate on symmetric coding scheme, where all cameras are equally important in the representation of the 3D scenes.

We consider a scenario, where two catadioptric cameras are distributed in the 3D scene, as shown in Fig. 1. Each catadioptric camera samples the plenoptic

[☆] This work has been partly supported by the Swiss National Science Foundation, under grant 20001-107970/1.

^{*} Corresponding author.

E-mail addresses: vijayaraghavan.thirumalai@epfl.ch (V. Thirumalai), ivana.tomic@epfl.ch (I. Tomic), pascal.frossard@epfl.ch (P. Frossard).

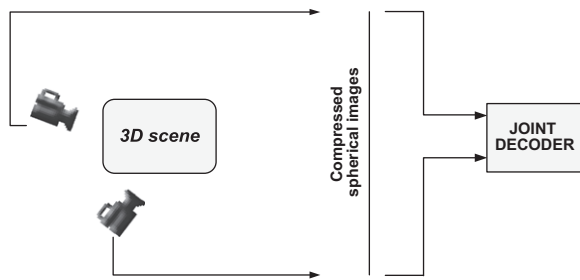


Fig. 1. Distributed coding of the 3D scenes. The correlated images are compressed independently and are decoded jointly.

function that represents the entire visual information seen by the observer [3]. The catadioptric camera is realized by a convex, reflective parabolic mirror placed above and parallel to the camera approximating an orthographically projecting lens. In such a case, the ray of light incident with the focus of the parabola is reflected to a ray of light parallel to the parabola's axis. This construction is equivalent to a purely rotating perspective camera [13]. We propose to work directly in the spherical domain by appropriately mapping the captured omnidirectional images on the sphere through inverse stereographic projection. This mapping permits us to obtain the light field in its natural radial form and to avoid the potential discrepancies that may arise due to Euclidean assumptions in perspective imaging. As the spherical image captures full 360° view, it is particularly suitable for representing the 3D scene. The wide field of view permits to reconstruct the associated 3D structure by processing a significantly smaller number of multi-view images, compared to the number of views in a perspective camera network performing the same task. The reduction of the total number of views is furthermore advantageous for determining the camera arrangement that achieves the complete representation of the 3D scene. Certainly, the quality of the reconstructed 3D structure depends on the resolution of the catadioptric sensor. In order to maintain a good reconstruction quality, the omnidirectional camera requires higher resolution image sensors [5]. Fortunately, in modern camera technologies the resolution of the image sensors has been drastically increased during the last years, offering a variety of high resolution devices at affordable costs.

In this paper, we propose a transform domain symmetric distributed coding scheme for representing a 3D scene captured by the stereo omnidirectional cameras. The correlated omnidirectional images initially undergo a multi-resolution decomposition based on the spherical Laplacian pyramid (SLP), which brings the advantage of shift invariance. The resulting sets of coefficients are quantized and then split into two correlated partitions. The quantized coefficients of the first partition are entropy coded, and sent to the decoder. The second partition is encoded using the nested scalar quantization (NSQ) [41], which is a binning scheme that encompasses a scalar quantizer and a coset encoder. It outputs the coset bin indexes and permits to reduce the coding rate compared to encoding the quantized coefficients directly. The coset

bin indexes are further encoded using a Slepian–Wolf encoder based on multi-level LDPC codes [20,21], in order to achieve further compression. The resulting syndrome bits are finally transmitted to the joint decoder.

The joint decoder estimates the quantized coefficients of the second partition from the quantized coefficients of the first partition, by exploiting intra-view correlation. Furthermore, the joint decoder takes benefit of the correlation between views by performing block-based disparity estimation (DE) on the sphere [35], which matches similar blocks of solid angles from two omnidirectional images, directly in the spherical domain. Therefore, the proposed scheme efficiently combines the intra and inter Wyner–Ziv image coding, which allows for a balanced coding rate between cameras. Such a strategy proves to be beneficial with respect to independent processing of omnidirectional images and shows only a small performance loss compared to joint encoding of the different views. Moreover, we exploit the inter-view correlation by block-based DE, which estimates the displacement between the corresponding objects without using epipolar geometry constraint. Hence, the block-based DE technique used in our scheme does not require any camera parameters which are usually required in the techniques based on epipolar geometry, to perform the correspondence matching (e.g., [4]). This is certainly beneficial in camera networks where the camera parameters are not given or when camera network calibration is not achievable in practice. The proposed scheme therefore provides a low-complexity coding solution for the representation of 3D scenes, which does not require complex setup nor hierarchical organization between vision sensors.

The rest of the paper is organized as follows. Section 2 overviews the related work in distributed coding with a special focus on camera networks. Section 3 presents the distributed coding algorithm adapted to omnidirectional images. Section 4 presents in more details the Wyner–Ziv coding strategy, while Section 5 describes the joint decoding scheme. Section 6 finally presents the experimental results that demonstrate the benefits of the proposed solution. Section 7 concludes this paper.

2. Related work

The first information-theoretical results on distributed source coding (DSC) appeared already in the late seventies. In particular, it has been shown that independent coding of correlated sources can achieve the same rate-distortion bound as joint encoding if a joint decoder can efficiently exploit the correlation between the sources [30]. Rate-distortion bounds have been established later for the particular case of coding with side information [38]. However, most results presented in [30,38] have remained non-constructive for about three decades. Practical DSC schemes have been designed only recently, by establishing a relation between the Slepian–Wolf theorem and channel coding [25]. Subsequently, several practical DSC systems have been presented using different channel codes, e.g., Turbo codes [10,1] or LDPC codes [19].

These results have been used to construct low-complexity video coders by assuming that adjacent frames represent correlated sources. These frames are categorized into key frames and Wyner–Ziv frames, where the key frames are encoded independently and the Wyner–Ziv frames are Slepian–Wolf encoded [2,26,14].

Only a few studies have been reported about the application of distributed coding principles to camera networks. The works reported in the literature are generally based on coding with side information, where one camera is used as a reference to decode the information from the other cameras. For example, in [42,36,12] the cameras are categorized into reference and Wyner–Ziv cameras and the correlation among views is exploited at the joint encoder using DE based on epipolar geometry, which usually requires camera parameters. When camera parameters are not available and calibration is not possible, the joint decoder can rather use block-based DE to exploit the redundancy between images [34,33]. These schemes, however, introduce a hierarchical relation among the sensors and the coding rate is therefore not balanced.

In practice, however, it is often interesting to avoid hierarchical relations between sensors and to further balance the coding and transmission costs among them. One of the first works that addresses balanced rate allocation in distributed coding is based on time sharing mechanism [37], which is, however, hard to implement due to node synchronization issues. The first practical scheme for symmetric coding based on channel code partitioning has been proposed in [24]. This scheme has been later extended to multiple sources using systematic channel codes by Stanković et al. [31]. It is based on horizontally splitting the generator matrix of the channel code into two sub-generator matrices. Codewords are then generated using the sub-matrices, and are assigned to each encoder. The compression rate of each encoder is determined by the number of rows retained in the corresponding sub-matrix. The advantage of this system is the need for only one channel code. However, this framework is limited to systematic channel codes. The authors in [32] have developed a symmetric DSC using a general linear channel code (which includes both systematic and non-systematic channel codes) and their framework is based on algebraic binning concept of the channel code. Simulation results have shown that almost the entire Slepian–Wolf region can be covered with this coding algorithm.

Symmetric distributed coding can also be achieved by information partitioning. Sartipi et al. [29] have considered the compression of two sources at the symmetric rate by information partitioning, where half of the source bits are transmitted directly while the corresponding syndrome bits are generated on the other half (complementary part) of the source bits. Similar to [32], the authors show that they can approach the entire SW region and thus the decoding error can be made insensitive to arbitrary rate allocation among the encoders. However, both schemes are based on capacity approaching channel codes that usually approach the Slepian–Wolf bound only for long source length (typically 10^4). Grangetto et al. [15]

have proposed a balanced coding scheme for small block length binary sources. The algorithm is based on time sharing version of distributed Arithmetic codes that perform better than the Turbo code-based DSC scheme in the considered framework.

The authors in [40] propose a rate balanced DSC scheme for video sequences. In this scheme each frame is divided into two partitions and one partition is then transmitted directly. In addition, each frame is Wyner–Ziv encoded and the side information is eventually generated using motion estimation. This scheme permits to avoid hierarchical relations between frames. However, it results in high coding rates, since one of the partitions in each frame is encoded using both Wyner–Ziv and independent coding. Finally, a balanced distributed coding scheme for camera networks has been proposed in [11], based on linear channel code construction that can achieve any point in the Slepian–Wolf region. The developed linear codes were not, however, applied to the practical coding of images in camera networks.

In this paper, we present a balanced distributed coding scheme for multi-view image coding. We focus on omnidirectional images and partition the multi-resolution information in different correlated subsets. The coefficients are Slepian–Wolf encoded, and both intra- and inter-view correlation are exploited at the joint decoder. Block-based DE is used to generate the side information, which does not require explicit knowledge of the camera parameters nor precise calibration.

3. Distributed coding scheme

3.1. Overview

This section presents the overview of the symmetric distributed coding scheme, illustrated in Fig. 2. We consider omnidirectional images that can be exactly mapped on the sphere, as those captured by catadioptric mirrors [13]. Since these images have a single center of projection, they permit to capture the plenoptic function [3] in its radial form without discrepancies that usually appear due to Euclidean assumptions for planar images. The stereo omnidirectional images X_1 and X_2 undergo multi-resolution decomposition based on SLP, whose coefficients are then quantized. As we target a balanced rate allocation between cameras, the information of both images are partitioned in a similar way. The quantized coefficients are split into two subsets or partitions: A_1 and B_1 from image X_1 , and partitions A_2 and B_2 from image X_2 . These partitions are generally correlated due to the simple partitioning process, which puts quantized coefficients alternatively in both partitions. The first partitions from both images (e.g., partitions A_1 and A_2) are transmitted directly to the joint decoder after entropy coding. The second partitions B_1 and B_2 are coset encoded and the resulting coset indexes are Slepian–Wolf encoded with a multi-level LDPC code. Hence, each encoder transmits one half of the quantized coefficients, and only the syndrome bits for the quantized coefficients of the second partition.

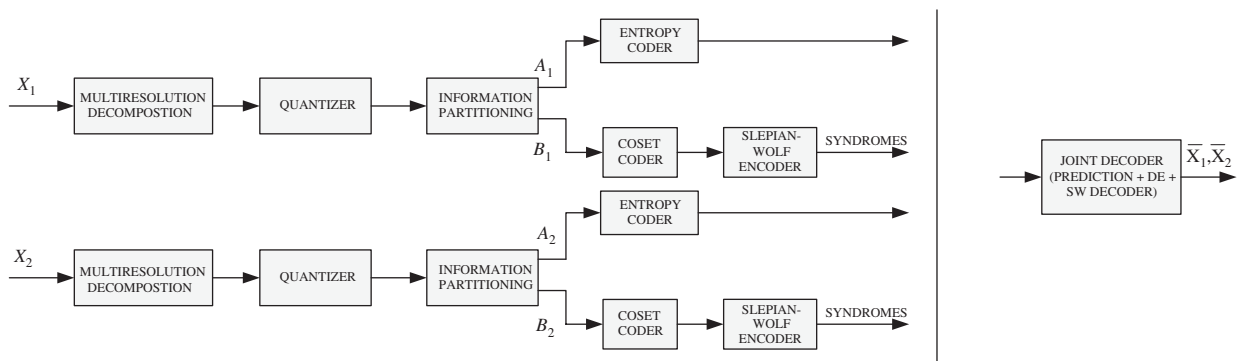


Fig. 2. Overview of the symmetric distributed coding scheme.

The joint decoder tries to exploit intra- and inter-view correlation for improved decoding performance. It estimates the quantized coefficients of the partitions B_1 and B_2 by using the coefficients of the partitions A_1 and A_2 , respectively. Under the assumption that the images X_1 and X_2 are correlated, the predicted result is further refined using block-based DE on the sphere. DE permits to compensate for the displacement of the objects captured from different viewpoints. Prediction and DE together lead to effective side information, which permits to reduce the channel rate of the Slepian–Wolf encoder. The coset indexes corresponding to the partitions B_1 and B_2 are further recovered after correcting the virtual channel noise in the side information using the corresponding syndromes. The SLP coefficients are then estimated from the recovered coset index, and the images are finally reconstructed by inverting the SLP transform.

3.2. Spherical Laplacian pyramid

Multi-resolution analysis is an efficient tool that permits to decompose a signal at progressive resolutions and perform coarse to fine computations on the data. The two most successful embodiments of this paradigm are the various wavelet decompositions [22] and the Laplacian pyramid (LP) [7]. As the shift invariance represents an interesting property for distributed coding in camera networks, we have chosen to use the LP on our scheme. It proves to be beneficial for predictive coding based on DE. Furthermore, since we work with omnidirectional images, we propose to use the LP on the sphere, which is presented below.

Let first $L^2(S^2, d\mu)$ denote the Hilbert space of the square integrable signals on the 2D sphere S^2 , where $d\mu(\theta, \varphi) = \sin \theta d\theta d\varphi$ represents the rotation invariant Lebesgue measure on S^2 . Any spherical signal $F \in L^2(S^2)$ can be expanded using the spherical harmonics $Y_{m,n}$ [8], whose Fourier coefficients are given by

$$\check{F}(m, n) = \int_{S^2} d\mu(\theta, \varphi) Y_{m,n}^*(\theta, \varphi) F(\theta, \varphi), \quad (1)$$

where $Y_{m,n}^*$ is the complex conjugate of the spherical harmonic of order (m, n) . The omnidirectional images are sampled on the nested equi-angular grids on the sphere

described as

$$\mathcal{G}_j = \left\{ (\theta_{jp}, \varphi_{jq}) \in S^2 : \theta_{jp} = \frac{(2p+1)\pi}{4W_j}, \varphi_{jq} = \frac{q\pi}{W_j} \right\}, \quad (2)$$

$p, q \in \mathcal{N}_j \equiv \{n \in \mathbb{N} : n < 2W_j\}$, for a range of bandwidth $W = \{W_j \in 2\mathbb{N}, j \in \mathbb{Z}\}$. These grids permit to perfectly sample any band-limited function $F \in L^2(S^2)$ of bandwidth W_j , i.e., such that $\check{F}(m, n) = 0$ for all $m > W_j$. This class of sampling grids are advantageously associated to a Fast Spherical Fourier Transform [16], which permits rapid transformation.

Similar to classical LP decomposition [7], the SLP proceeds first by low-pass filtering the spherical signal in the Fourier domain for speeding up the computations. Suppose the original data F is bandlimited, i.e., $\check{F}(m, n) = 0, \forall m > W_0$, and is sampled on \mathcal{G}_0 . We capture the low frequency information by using an half-band axisymmetric filter \check{H} . A spherical function is said to be axisymmetric if it is invariant to rotation with respect to the principal spherical axis, i.e., such a function is independent of the variable n and is represented by $\check{H}(m)$. The bandwidth of the filter $\check{H}(m)$ is chosen such that it is numerically close to a perfect half-band filter. The signal F is low-pass filtered using $\check{H}(m)$, and the filtered data are then downsampled on the nested sub-grid \mathcal{G}_1 , which gives F_1 the low-pass channel of our SLP. The high-pass channel of the pyramid F_0 is computed as usual, that is by first upsampling F_1 on the finer grid \mathcal{G}_0 , low-pass filtering it with \check{H} and taking the difference with F . Coarser resolutions are computed by iterating this algorithm on the low-pass channel F_1 . By repeating this step k times, the original spherical signal F can be decomposed into the LL subband F_k and detailed subbands $F_{k-1}, F_{k-2}, \dots, F_0$. The parameter k denotes the level of decomposition used to decompose the image F .

The coefficients of the SLP need to be quantized with efficient rate distribution among the subbands. We follow the algorithm proposed for the LP in [28]. The rate allocation can be computed by Lagrange's multipliers method when the quantizers are uniform. Unsurprisingly, the rate in the different subbands is chosen to be proportional to the variance of the coefficients.

4. Slepian–Wolf encoder

4.1. Coefficient partitioning

We describe now in more detail the proposed Wyner–Ziv encoding scheme illustrated in Fig. 3. The omnidirectional images X_1 and X_2 undergo multi-resolution decomposition based on SLP using k decomposition levels. The generated subband coefficients are represented by $\{X_{v,j}\}$ for $v = 1, 2$, and $j = 0, \dots, k$. In the analysis that follows in the rest of this paper, v represents the image (view) index and j represents the corresponding subband index. For example, $X_{v,j}$ represents the set of coefficients in the j th subband of the image X_v . The generated subband coefficients are then quantized uniformly with optimal rate allocation among the subbands, as described above. The quantized coefficients of each image are then distributed alternatively into two correlated partitions that form a kind of a checkerboard pattern. Let (p, q) denote the position indexes of a point $(\theta_{jp}, \varphi_{jq})$ on the equi-angular spherical grid \mathcal{G}_j , as defined in Eq. (2). The quantized coefficient $X_{v,j}$ at a point $(\theta_{jp}, \varphi_{jq})$ on the spherical grid is put in the partition $A_{v,j}$ if $((p \bmod 2) \text{ XOR } (q \bmod 2)) = 0$. Otherwise it is put in the partition $B_{v,j}$. For illustration, Fig. 4 shows the partitioning strategy used in our scheme for the subband $X_{v,j}$ of the bandwidth $W_j = 2$, i.e., of the size 4×4 . It is clear that coefficients at the positions marked in white and black color belong to the partitions $A_{v,j}$ and $B_{v,j}$, respectively. Finally, the quantized coefficients from each subband $j = 0, \dots, k$ of the pyramid are split into two partitions $A_{v,j}$ and $B_{v,j}$, following the same partitioning strategy.

The quantized coefficients in the partitions $\{A_{v,j}\}$ for all $v = 1, 2$ and $j = 0, \dots, k$ are compressed using the Arithmetic encoder and the compressed bits $\{D_{v,j}\}$ are transmitted directly to the joint decoder. The partitions $\{B_{v,j}\}$, however, further undergo coset and Slepian–Wolf coding to save bit rate in the distributed coding scheme. The quantized coefficients in partition $B_{v,j}$ are put in different

cosets. The cosets group coefficients from $B_{v,j}$ that are separated by a distance $d_{v,j}$. Only the coset indexes are eventually encoded, which provides some significant rate savings. The coset distance is estimated as $d_{v,j} = 2^{\lceil \log_2(2E_{v,j}+1) \rceil}$, where $E_{v,j}$ is the maximum error between the original and the side information images in the j th subband. Scalar quantization and coset encoding together behave similarly to NSQ.

4.2. Multi-level LDPC coding

Even after coset encoding, some correlation still exists between the coefficients and the side information available at the decoder. We propose to achieve further compression by encoding the coset indexes with multi-level LDPC codes [39]. In other words, instead of sending the coset indexes to the decoder, the encoder only transmits the syndrome bits $\{S_{v,j}\}$ resulting from the LDPC encoding. We propose to use irregular LDPC codes and we follow the procedure described in [21,23] in order to construct the parity check matrix. We describe now in

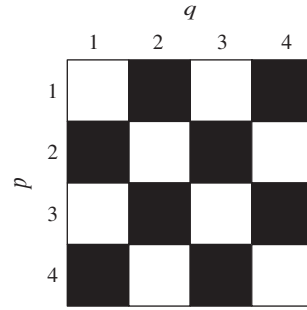


Fig. 4. Checkerboard partition strategy for the subband $X_{v,j}$ of the bandwidth $W_j = 2$. The top most left quantized coefficient is indexed by $p = 1, q = 1$. The partitions $A_{v,j}$ and $B_{v,j}$ are marked by the white and black color, respectively.

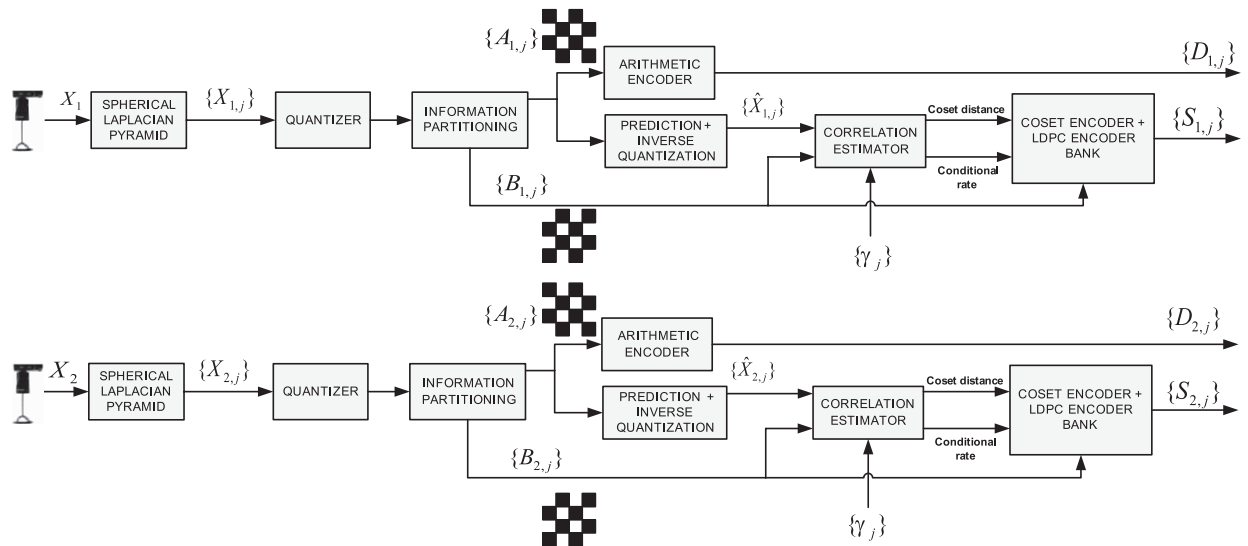


Fig. 3. Detailed block scheme of the Wyner–Ziv encoder with correlation estimation.

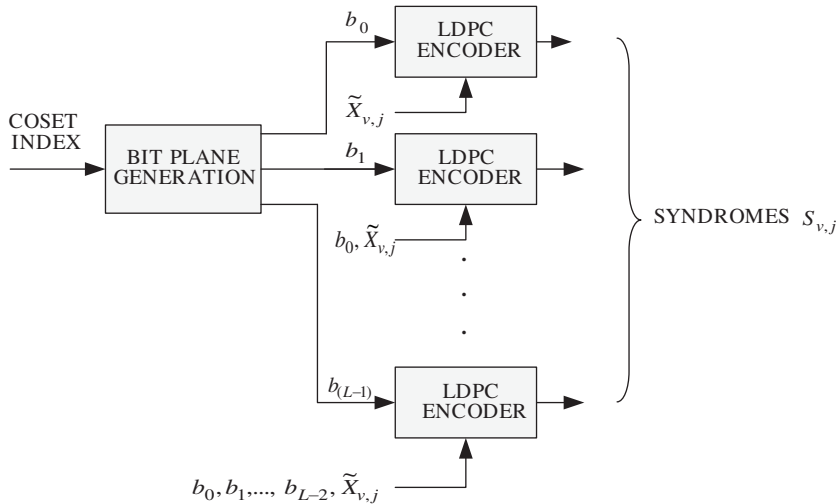


Fig. 5. Block scheme of the LDPC encoder bank, where the bitplanes are encoded starting from b_0 . Each bit plane b_l ($0 \leq l \leq L-1$) is encoded based on the previously encoded $l-1$ bitplanes and the estimated side information $\tilde{X}_{v,j}$.

more details the multi-level LDPC coding, and the channel rate estimation.

For the image X_v , the coset indexes generated from each subband j are first decomposed into L bit planes b_0, b_1, \dots, b_{L-1} , where b_0 represents the most significant bit (MSB) plane and b_{L-1} represents the least significant bit (LSB) plane. Each bitplane b_l ($0 \leq l \leq L-1$) is encoded by the LDPC encoder, starting from the bit plane b_0 , as illustrated in Fig. 5. The LDPC encoding rate is chosen by assuming that the error between the SLP subbands and the corresponding side information follows a Laplacian distribution, as shown later. The subsequent bit planes are also encoded with LDPC codes, where the code rate is, however, adapted based on the previously encoded bitplanes.

In more details, the coding rate for encoding the l th bitplane of the subband $X_{v,j}$ is estimated as follows. First, the conditional probability $P(i) = \Pr(b_l(i) = 1 | \tilde{X}_{v,j}(i), b_0(i), b_1(i), \dots, b_{l-1}(i))$ is calculated for each bit i , where $\tilde{X}_{v,j}$ denotes the side information for decoding the subband $X_{v,j}$. The rate of the LDPC encoder is then chosen to be equal to the following conditional entropy:

$$H(b_l | \tilde{X}_{v,j}, b_0, b_1, \dots, b_{l-1}) \simeq \frac{1}{M} \sum_{i=1}^M H(P(i)), \quad (3)$$

where M denotes the number of bits in the bit plane [9]. However, the side information is unfortunately not available at the encoder, and the conditional probability P has generally to be estimated as described below.

4.3. Noise models

One of the main difficulties in distributed coding is the estimation of the correlation between the sources, or equivalently the construction of noise models at the encoder for the proper design of the Slepian–Wolf encoder. The encoder has to estimate the noise distribution in order to determine the coset distance, and the

LDPC coding rate. Unfortunately, the side information that is used for joint decoding is only present at the decoder, and the encoder can only predict the noise distribution.

We assume that the error $E_{v,j}$ between the SLP subband $X_{v,j}$ and the corresponding side information subband $\tilde{X}_{v,j}$ follows a Laplacian distribution, of the form $f_{E_{v,j}}(e) = (1/2\lambda_{v,j}) \exp(-|e|/\lambda_{v,j})$. The Laplacian distribution is a common assumption in such a scenario, and it provides a good approximation of the actual distribution of the error. In this case, the rate $R_{v,j}$ necessary to code the error $E_{v,j}$ is equivalent to the conditional entropy $H(X_{v,j} | \tilde{X}_{v,j})$. When the quantization is uniform with step size $\delta_{v,j}$ for the subband $X_{v,j}$, the rate depends only on the variance of the Laplacian distribution [17]. It can be written as

$$R_{v,j} = H(X_{v,j} | \tilde{X}_{v,j}) = \alpha_j \log_2 \left(\frac{\lambda_{v,j}}{\delta_{v,j}} \right) + \beta_j, \quad (4)$$

where α_j and β_j are constants that can be estimated offline on test image sets, and therefore are not dependent of the image X_v . The construction of the noise model for a proper choice of the coding parameters therefore consists in estimating the parameters $\{\lambda_{v,j}\}$ of the Laplacian distribution, for all $v = 1, 2$ and $j = 0, \dots, k$.

In our scheme, the side information is actually built on prediction and DE steps. We can thus model separately the effect of spatial prediction of the coefficients, and the benefit of DE. In the first case, the encoder can estimate the rate $R'_{v,j}$ that is necessary to correct the error due to spatial prediction of the coefficients. The encoder can implement the coefficient prediction step, since it does not depend on the information from the other sensors. The residual error between the coefficients in the subband $X_{v,j}$ and the corresponding subband computed by coefficient prediction $\hat{X}_{v,j}$ can then be modeled with a Laplacian distribution. The parameter of the distribution $\lambda'_{v,j}$ is finally estimated from the prediction error. The rate $R'_{v,j}$

needed to code the prediction error can also be computed by the conditional entropy, given as

$$R'_{v,j} = H(X_{v,j}|\hat{X}_{v,j}) = \alpha_j \log_2 \left(\frac{\lambda'_{v,j}}{\delta_{v,j}} \right) + \beta_j. \quad (5)$$

However, the side information is not only built on coefficient prediction, since DE is used at the decoder in order to exploit the correlation between the images from different sensors. We propose to compute a conservative approximation of the gain due to DE, expressed as $\gamma_j = R'_{v,j}/R_{v,j}$. It can be computed by offline encodings of several test images, where the complete side information is made available at the encoder. The offline estimation of γ_j finally permits to estimate at the encoders the parameter of the complete noise model. The parameter $\lambda_{v,j}$ of the Laplacian distribution can be expressed as

$$\lambda_{v,j} = \lambda'_{v,j} 2^{(-R'_{v,j}(1-1/\gamma_j)/\alpha_j)} \quad (6)$$

by combinations of Eqs. (4) and (5). As we have now an approximation of the distribution of the error induced by the side information, we can estimate the side information for subband $\hat{X}_{v,j}$ at the encoder. It permits to estimate the error $E_{v,j}$ and hence the coset distance $d_{v,j}$ for coding the quantized coefficients. Finally, we can estimate the LDPC coding rate by computing the probability $P(i)$, and the conditional entropy given in Eq. (3). The complete encoding scheme is illustrated in Fig. 3.

5. Joint decoding

5.1. Overview

The joint decoder exploits the correlation between the images in order to reconstruct the views of the 3D scenes. The decoding scheme is illustrated in Fig. 6. The quantized coefficients from the partitions $\{A_{v,j}\}$ for $v = 1, 2$, and $j = 0, \dots, k$, are easily recovered by Arithmetic decoding. The quantized coefficients of the partitions $\{B_{v,j}\}$ for $v = 1, 2$, and $j = 0, \dots, k$, however, have to be reconstructed by Slepian–Wolf decoding. These missing quantized coefficients are first predicted in each subband $B_{v,j}$, with the help of the coefficients from the corresponding partition $A_{v,j}$. The unknown coefficients are simply predicted by interpolation from the four nearest neighbors in

the partition $A_{v,j}$. This simple spatial prediction exploits the correlation among successive coefficients. The quantized coefficients further undergo inverse quantization to obtain the j th predicted subband $\hat{X}_{v,j}$. The side information is then built by refining the value of the predicted coefficients by disparity compensation between the approximations of the different images. The decoder implements DE on the sphere to exploit the redundancy between the omnidirectional images from different cameras. Next, the coset indexes that correspond to the coefficients in partitions $\{B_{v,j}\}$ are recovered by using the syndrome bits of the LDPC code, as well as the side information created by prediction and disparity and compensation. Finally, the SLP coefficients are recovered by coset decoding with help of the side information. The main steps of the joint decoder algorithm are detailed in the rest of this section.

5.2. Side information generation

The subbands built on the spatial prediction of the missing coefficients are refined by DE on the sphere. The inverse SLP transform is first applied on the predicted subbands $\{\hat{X}_{1,j}\}$ and $\{\hat{X}_{2,j}\}$ for $j = 0, \dots, k$ to generate the predicted images \hat{X}_1 and \hat{X}_2 , respectively. The correlation between these images \hat{X}_1 and \hat{X}_2 is captured by implementing multi-resolution block-based DE on the sphere. We summarize here the key ideas of the DE on the sphere, while more details are available in [35]. Initially SLP decomposition is carried out on the images \hat{X}_1 and \hat{X}_2 using k decomposition levels. The low resolution subband $\hat{X}_{1,k}$ is divided into blocks of uniform solid angles. For each block in $\hat{X}_{1,k}$, a best matching block in the mean square sense is selected in $\hat{X}_{2,k}$. The displacement between the corresponding blocks represents the disparity vector. The generated disparity vectors from the lower resolution are upsampled and are used as the initial estimate for the higher resolution. The initial estimate is further refined using the subband coefficients at the higher resolution. This process is iterated up to the finest resolution and eventually outputs the disparity vectors. The resulting disparity vectors are then used for constructing an estimate \tilde{X}_1 of the image X_1 that serves as side information for decoding the image X_1 . In particular, the side

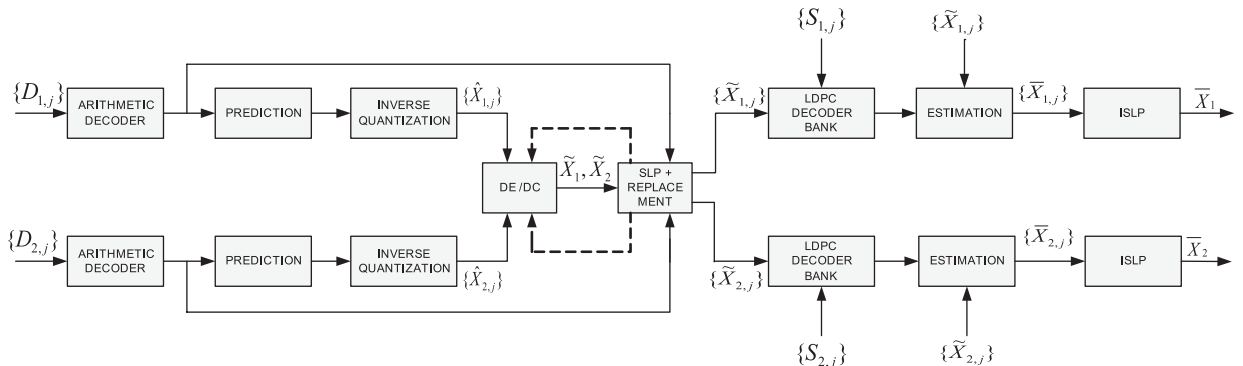


Fig. 6. Detailed block scheme of the Wyner–Ziv decoder.

information image is first constructed by applying disparity compensation from the image \tilde{X}_2 . It then undergoes a SLP decomposition, similar to the transform implemented at the encoder. The coefficients corresponding to the partition $\{A_{1,j}\}$ for $j = 0, \dots, k$ are then substituted by the coefficients that have been correctly received from the encoder, in order to reduce the estimation error. The same process is implemented to generate the side information image \tilde{X}_2 with disparity compensation based on the predicted image \tilde{X}_1 .

The exploitation of the correlation between images by DE permits to refine the values of the predicted result from the partition $B_{v,j}$. The DE process can be repeated on the images \tilde{X}_1 and \tilde{X}_2 in order to further improve the image approximation. We have observed empirically that it is advantageous to repeat the disparity compensation a second time. This step is represented with a dashed line on the block scheme in Fig. 6. Further iterations, however, do not improve significantly the side information. The resulting side information images are transformed with SLP to build the side information subbands that are used for decoding the coset indexes.

5.3. Coefficient decoding

The coefficients from partition $\{B_{v,j}\}$ for $v = 1, 2$ and $j = 0, \dots, k$, are finally recovered by Slepian–Wolf decoding. The generated side information $\tilde{X}_{v,j}$ is used by the LDPC decoder together with the syndromes bits $S_{v,j}$ to decode the coset indexes in each subband $X_{v,j}$. An LDPC decoder bank uses L LDPC decoders to decode each bit plane successively, starting from MSB to LSB bit planes. While decoding the bit plane b_l ($0 \leq l \leq L-1$), the previously decoded $l-1$ bit planes b_0, b_1, \dots, b_{l-1} are used as the additional side information by the LDPC decoder. LDPC decoding is implemented with a Belief propagation algorithm, where the confidence level is initialized at the variable node using the following log likelihood ratio (LLR):

$$\text{LLR} = \log \left(\frac{P(b_l = 0 | \tilde{X}_{v,j}, b_0, b_1, \dots, b_{l-1})}{P(b_l = 1 | \tilde{X}_{v,j}, b_0, b_1, \dots, b_{l-1})} \right). \quad (7)$$

The coset indexes of each j th subband $X_{v,j}$ are reconstructed when all the bit planes are decoded. The coefficients in each subband are finally computed by decoding the coset indexes. The decoded coefficient corresponds to the coefficient in the coset that is the closest to the side information $\tilde{X}_{v,j}$. Once all the subband coefficients are decoded, the image is reconstructed by inverting the SLP transform.

6. Experimental results

6.1. Setup

We evaluate the performance of our system on both synthetic and natural spherical images. Synthetic spherical image set Room is shown in Fig. 7 and the natural spherical image set Lab is shown in Fig. 8.

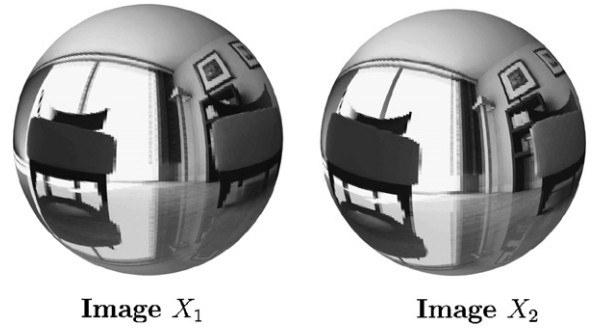


Fig. 7. Original Room images.

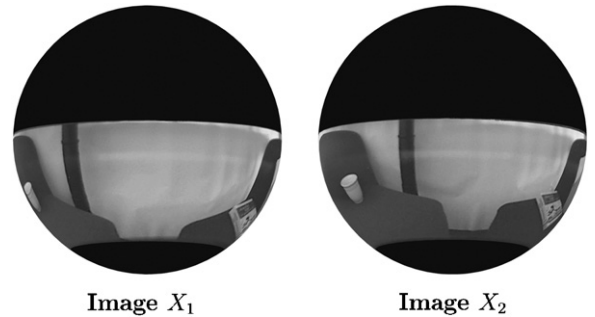


Fig. 8. Original Lab images.

The SLP is implemented using an axisymmetric filter $\check{H}(m)$ constructed from a seven-tap digital filter $h(s) = \{-0.0625 \ 0 \ 0.5625 \ 1 \ 0.5625 \ 0 \ -0.0625\}$. The filter $\check{H}(m)$ is constructed by computing the Fourier transform of $h(s)$ and replicating it for each n such that $\check{H}(m, 1) = \check{H}(m, n)$ for all n . Obviously such a construction results in an axisymmetric filter that is independent of the variable n , and is completely determined by $\check{H}(m)$. The SLP decomposition is further carried out in the Fourier domain in order to speed up the computations. The SLP is implemented with four levels of decomposition ($k = 4$) in the results presented below. The multi-resolution block-based DE at the decoder is carried out on blocks of size 4×4 . Finally, the performance is measured in terms of PSNR, where the mean square error is evaluated using the inner product on the sphere.

6.2. Channel model evaluation

Before analyzing the performance of the distributed coding scheme, we propose to evaluate the channel model that is used for designing the Slepian–Wolf encoder. We first show in Fig. 9 the distribution of the error between the subband $X_{2,3}$ and the corresponding side information subband $\tilde{X}_{2,3}$ in the Room image data set. The error is computed only on the coefficients of the partition $B_{2,3}$. We can see that the error follows zero-mean Laplacian distribution with $\lambda_{2,3} = 0.0178$, as expected.

Then, we estimate the constants α_j and β_j that are used to compute the conditional entropy in Eq. (4) for both Room and Lab image sets. These constants are the same

for all image sets and differ only in the respective subbands. We have obtained $\alpha_4 = 1.04$ and $\beta_4 = 2.44$ for the LL subband and $\alpha_j = 0.54$ and $\beta_j = 1.92$ ($0 \leq j \leq 3$) for the detail subband.

Finally, we evaluate the constant γ_j for all the subbands ($0 \leq j \leq 4$) that captures the benefit of the DE. We have obtained a value of $\gamma_4 = 1.25$ for the LL subband in both image sets. For the detail subbands the parameter γ_j is found to be 1.6, 1.4, 1.2 and 1.1 starting from the lowest to the highest resolution subbands, respectively. The value of γ_j is decreasing when the resolution increases, since the DE is mostly efficient in capturing the correlation in the low frequency subbands.

6.3. Coding performance

We first compare the performance of the proposed DSC solution (using estimated correlation model) with an independent coding scheme, and a joint encoding scheme. In the independent coding scheme, the images X_1 and X_2 are encoded independently using SLP-based strategy. The images X_1 and X_2 are transformed using four SLP decomposition levels. Compression is achieved by first quantizing the coefficients [28] and further the quantized coefficients are entropy coded (e.g., Arithmetic coding), similarly as the coefficients of the partition $A_{v,j}$ in the distributed coding scheme. The joint encoding scheme is based on disparity compensated predictive coding. One

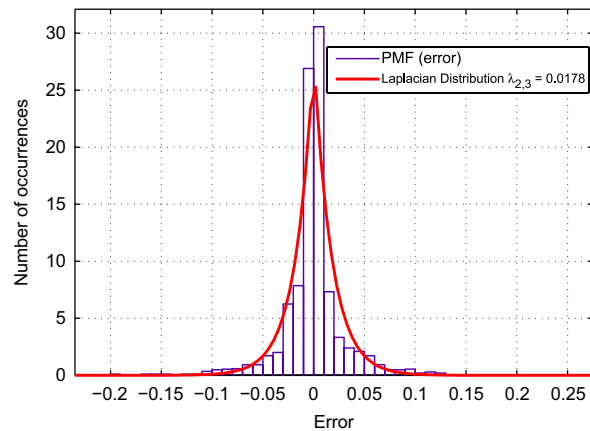


Fig. 9. Distribution of the error $E_{2,3}$, computed on the partition $B_{2,3}$ between the detail subband $X_{2,3}$ and $\hat{X}_{2,3}$ of the Room image set. The fitting curve shows a zero-mean Laplacian distribution with $\lambda_{2,3} = 0.0178$.

image is selected as the reference and it is encoded independently, whereas the other image is predicted from the reference image. In our scheme, the image X_2 is selected as the reference image and X_1 is predicted from X_2 . The reference image is encoded using four SLP decomposition levels. Multi-resolution DE with a three-level SLP decomposition and blocks of size 8×8 is used to predict the image X_1 . The residual error after DE is also encoded using SLP-based strategy. The disparity vectors of the successive resolution levels are differentially encoded. Finally, the rate allocation between the reference and the predicted images is chosen such that the rate-distortion performance is maximized. The corresponding rate distributions are given in Tables 1 and 2, where the bits used for the disparity vectors are included in the budget of the predicted frames.

The comparison between the distributed, independent and joint coding scheme is given in Fig. 10 in terms of rate-distortion performance. We observe that the distributed coding scheme performs close to joint encoding algorithm that is based on the same representation and coding strategy. We see also that our proposed DSC scheme clearly outperforms independent coding scheme. In particular, the gain reaches 1.5 dB for the Room image set and 1.3 dB for the Lab image set. We further compute the rate savings between DSC and independent coding schemes for the same reconstruction quality. Tables 3 and 4 tabulate the percentage of rate saving at different reconstructed qualities for the Room and Lab images, respectively. We could see that bit saving is approximately 25%, for both image sets. The reconstructed Room image \hat{X}_1 is finally shown in Fig. 11 for two sample bit rates. The reconstructed images are shown as planar images in the (θ, φ) coordinates to show all parts of the spherical images.

Table 2

Lab image set: distribution of bits between reference image and the predicted image in joint encoding

Reference image (X_2)		Predicted image (X_1)		Total rate (bpp)	Mean PSNR (dB)
Bits	PSNR (dB)	Bits	PSNR (dB)		
3742	25.26	219	24.11	0.0604	24.64
7759	28.06	707	26.37	0.1292	27.13
11469	29.73	1528	28.02	0.1983	28.79
11469	29.73	3274	29.1	0.2250	29.40

Table 1

Room image set: distribution of bits between reference image and the predicted image in joint encoding

Reference image (X_2)		Predicted image (X_1)		Total rate (bpp)	Mean PSNR (dB)
Bits	PSNR (dB)	Bits	PSNR (dB)		
5293	23.8	297	23.19	0.0853	23.47
13161	26.11	864	25.93	0.214	26
37694	30.5	1887	28.12	0.604	29.15
37694	30.5	5197	28.8	0.6545	29.57

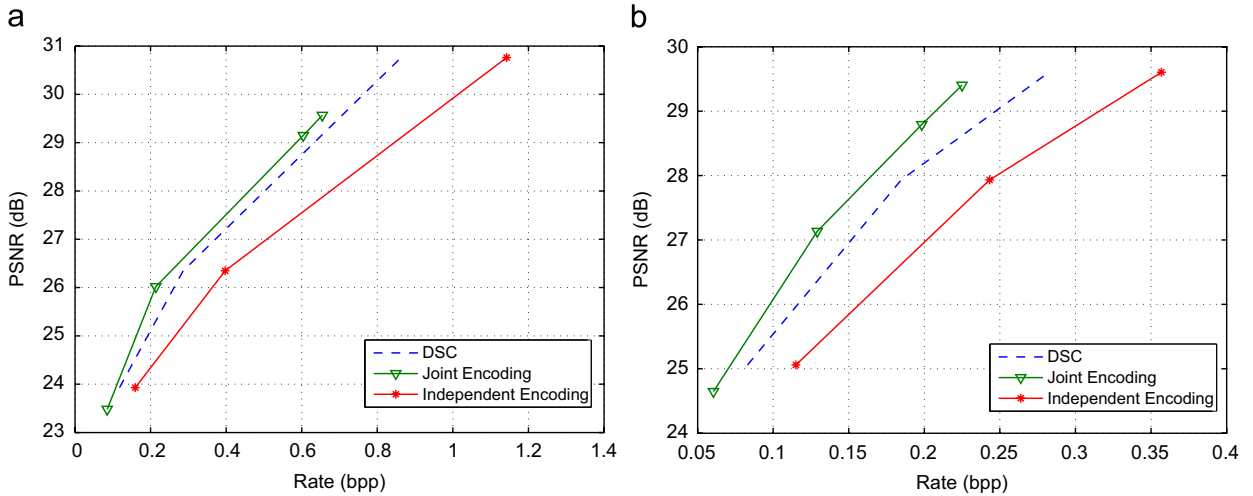


Fig. 10. Rate-distortion performance comparison between the proposed DSC scheme, and joint and independent coding strategies for (a) Room and (b) Lab image sets.

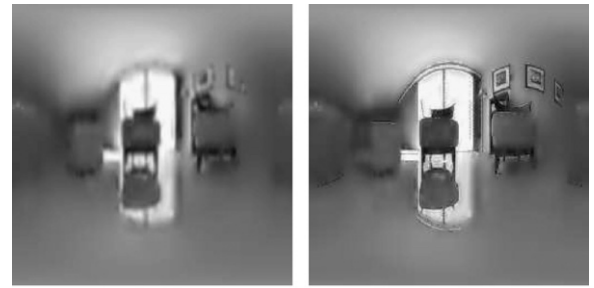
Table 3
Rate savings for Room image X_2

PSNR (dB)	Bits		Bits saved	% Bit saving
	DSC	Independent		
23.8	3928	5293	1365	25.8
26.1	9526	13 161	3635	27.6
30.5	28 166	37 694	9528	25.3

Table 4
Rate savings for Lab image X_1

PSNR (dB)	Bits		Bits saved	% Bit saving
	DSC	Independent		
25.26	2852	3797	945	24.9
28.06	6214	8179	1965	24
29.73	9542	11 920	2378	20

Finally, we compare in Fig. 12 the average performance of the distributed coding scheme with independent coding implemented by the JPEG compression standard, for the Room image set. The equi-angular grid of the spherical image is represented as a 2D planar image. A baseline JPEG scheme is used to encode both unwrapped X_1 and X_2 independently. We can see that both independent and the distributed coding schemes based on the SLP decomposition outperform JPEG at low coding rates, thanks to efficient data processing on the sphere. At higher rates, the mode of representation of the information becomes less important, and JPEG provides improved performance. The degradation of the RD performance of our scheme with respect to JPEG at higher rates could be explained by the use of a simple encoding scheme of the overcomplete LP, which is based on adaptive



Rate: 0.14 bpp, PSNR: 26.1 dB Rate: 0.43 bpp, PSNR: 30.5 dB

Fig. 11. Reconstructed image \bar{X}_1 in the Room scene. The images are represented as planar images in (θ, φ) coordinates.

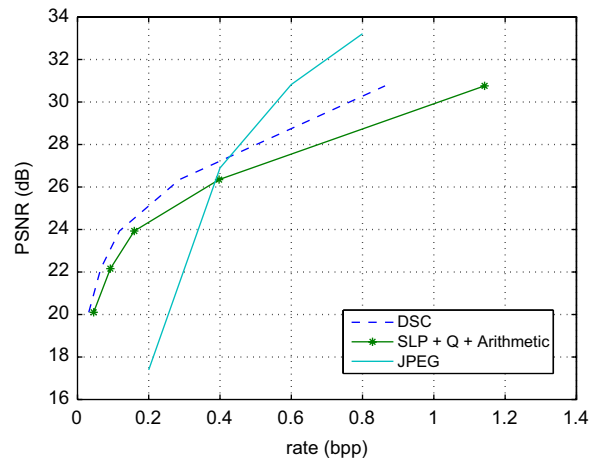


Fig. 12. Average rate-distortion performance for encoding the Room image set using the distributed coding scheme, and independent coding with JPEG.

quantization. Employing more efficient coding methods for LP, like the one proposed in [27], could result in improved RD performance of the proposed scheme, also at higher rates.

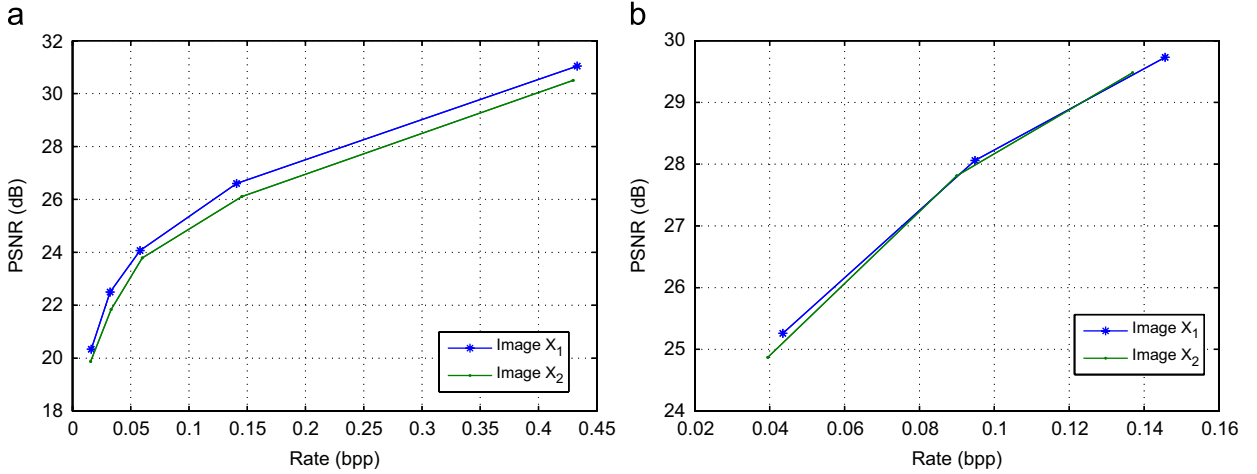


Fig. 13. Rate-distortion comparison between the images X_1 and X_2 , to examine the rate balance among the encoders. (a) Room and (b) Lab.

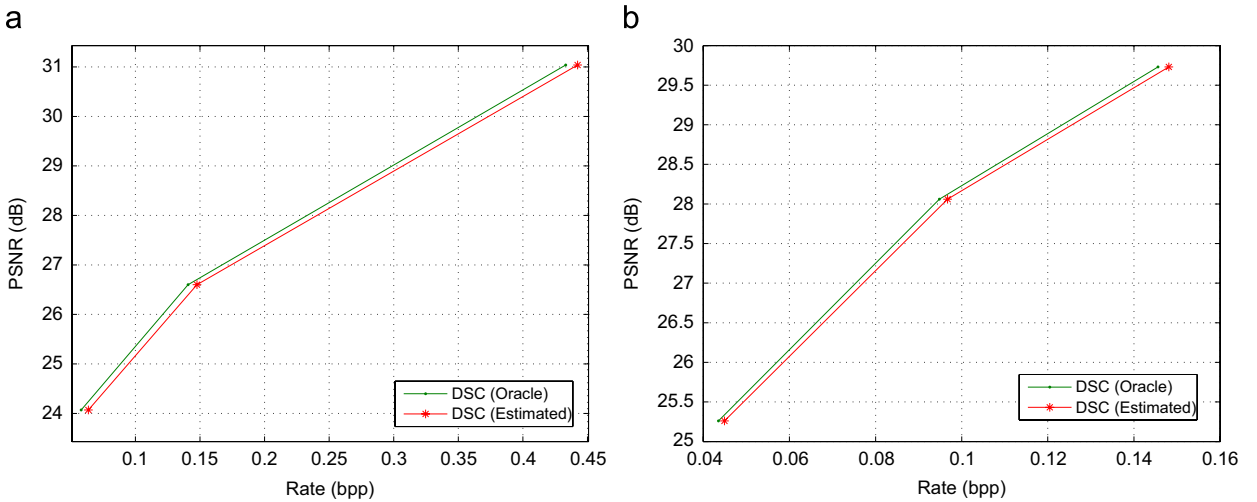


Fig. 14. RD comparison between estimated correlation model vs exact correlation model. (a) Room image X_1 and (b) Lab image X_1 .

6.4. DSC scheme analysis

We analyze in more details the behavior of the distributed coding scheme. In particular, we examine the rate balance between the two encoders, by comparing the RD performance of the images X_1 and X_2 . Fig. 13 shows the RD curves for the test images. As expected, the DSC scheme balances the encoding rates, since the encoding rates between the images X_1 and X_2 are quite similar at a given reconstruction quality.

Lastly, we study the effect of imprecise estimation of the coding rate in the Slepian–Wolf encoder. For both image sets, we compare the rate-distortion characteristics between the estimated correlation model described in this paper, and an exact oracle model where the Laplacian distribution parameters $\{\lambda_{v,j}\}$ are known a priori at the encoder for all v, j . The comparison is presented in Fig. 14 for the image X_1 of the Room and Lab sets. We can see that

the methodology proposed in this paper for estimating the channel rate performs very similar to the exact model. The performance degradation due to inexact rate estimation stays smaller than 0.2 dB.

7. Conclusions

In this paper we have presented a transform-based distributed source coding scheme for the balanced representation of 3D scenes with stereo omnidirectional cameras. The images are decomposed by spherical Laplacian pyramid, and coefficients are partitioned and Slepian–Wolf encoded at independent encoders. The joint decoder efficiently exploits both the intra- and inter-view correlation by coefficient prediction and disparity estimation. Our scheme outperforms independent coding and performs close to a joint encoding solution based on

similar coding principles. The proposed scheme is shown to be quite robust to inexact estimation of the correlation and Slepian–Wolf coding rate. It provides an interesting and low-complexity solution for simple networks of omnidirectional cameras, since it does not require any calibration nor hierarchy between sensors.

Acknowledgments

The authors would like to thank the anonymous reviewers and the guest editor for their careful reviews and valuable suggestions that help to improve this paper.

References

- [1] A. Aaron, B. Girod, Compression with side information using turbo codes, in: Proceedings of the IEEE DCC, April 2002, pp. 252–261.
- [2] A. Aaron, S. Rane, E. Setton, B. Girod, Transform-domain Wyner–Ziv codec for video, in: Proceedings of the SPIE Visual Communications and Image Processing, January 2004, pp. 520–528.
- [3] E.H. Adelson, J.R. Bergen, in: M. Landy, J.A. Movshon (Eds.), Computational Models of Visual Processing, MIT Press, Cambridge, MA, 2001, pp. 3–20.
- [4] Z. Arican, P. Frossard, Dense disparity estimation from omnidirectional images, in: Proceedings of the IEEE AVSS, September 2007, pp. 399–404.
- [5] S. Baker, S.K. Nayar, A theory of catadioptric image formation, in: Proceedings of the IEEE ICCV, January 1998, pp. 35–42.
- [7] P.J. Burt, E.H. Adelson, The Laplacian pyramid as a compact image code, IEEE Trans. Comm. 31 (4) (April 1983) 532–540.
- [8] W.E. Byerly, An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics, with Applications to Problems in Mathematical Physics, Dover, New York, 1959.
- [9] S. Cheng, Z. Xiong, Successive refinement for the Wyner–Ziv problem and layered code design, IEEE Trans. Signal Process. 53 (8) (August 2005) 3269–3281.
- [10] J. Garcia-Frias, Y. Zhao, Compression of correlated binary sources using turbo codes, IEEE Comm. Lett. 5 (10) (October 2001) 417–419.
- [11] N. Gehrig, P.L. Dragotti, Distributed compression in camera sensor networks, in: Proceedings of the IEEE MMSP, September 2004, pp. 311–314.
- [12] N. Gehrig, P.L. Dragotti, Distributed compression of multi-view images using a geometric approach, in: Proceedings of the IEEE ICIP, September 2007, pp. 421–424.
- [13] C. Geyer, K. Daniilidis, Catadioptric projective geometry, Internat. J. Comput. Vision 45 (3) (November 2001) 223–243.
- [14] B. Girod, A. Aaron, S. Rane, D. Rebollo-Monedero, Distributed video coding, Proc. IEEE 93 (January 2005) 71–83.
- [15] M. Grangetto, E. Magli, G. Olmo, Symmetric distributed arithmetic coding of correlated sources, in: Proceedings of the IEEE MMSP, October 2007, pp. 111–115.
- [16] D.M. Healy, D. Rockmore, P. Kostelec, S. Moore, Ffts for the 2-sphere—improvements and variations, J. Fourier Anal. Appl. 9 (3) (July 2003) 341–385.
- [17] A. Jagmohan, A. Sehgal, N. Ahuja, Two-channel predictive multiple description coding, in: Proceedings of the IEEE ICIP, September 2005, pp. 670–673.
- [19] A.D. Liveris, Z. Xiong, C.N. Georghiades, Compression of binary sources with side information at the decoder using LDPC codes, IEEE Comm. Lett. 6 (10) (October 2002) 440–442.
- [20] D. MacKay, Good error-correcting codes based on very sparse matrices, IEEE Trans. Inform. Theory 45 (3) (March 1999) 399–431.
- [21] D. Mackay, R.M. Neal, Near Shannon limit performance of low density parity check codes, Electron. Lett. 33 (6) (March 1997) 457–458.
- [22] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, New York, 1998.
- [23] Methods for constructing LDPC codes. Available in URL: (<http://www.cs.utoronto.ca/pub/radford/LDPC-2001-05-04/pchk.html>).
- [24] S.S. Pradhan, K. Ramchandran, Distributed source coding: symmetric rates and applications to sensor networks, in: Proceedings of the IEEE DCC, March 2000, pp. 363–372.
- [25] S.S. Pradhan, K. Ramchandran, Distributed source coding using syndromes (DISCUS), IEEE Trans. Inform. Theory 49 (3) (March 2003) 626–643.
- [26] R. Puri, A. Majumdar, K. Ramchandran, PRISM: a video coding paradigm with motion estimation at the decoder, IEEE Trans. Image Process. 16 (10) (October 2007) 2436–2448.
- [27] G. Rath, C. Guillemot, Compressing the Laplacian pyramid, in: Proceedings of the IEEE MMSP, October 2006, pp. 75–79.
- [28] J.L. Salinas, R.L. Baker, Laplacian pyramid encoding: optimum rate and distortion allocations, in: Proceedings of the IEEE ICASSP, May 1989, pp. 1957–1960.
- [29] M. Sartipi, F. Fekri, Distributed source coding in the wireless sensor networks using LDPC codes: the entire Slepian–Wolf rate region, in: Proceedings of the IEEE WCNC, March 2005, pp. 1939–1944.
- [30] D. Slepian, J.K. Wolf, Noiseless coding of correlated information sources, IEEE Trans. Inform. Theory 19 (4) (July 1973) 471–480.
- [31] V. Stanković, A.D. Liveris, Z. Xiong, C.N. Georghiades, Design of Slepian–Wolf codes by channel code partitioning, in: Proceedings of the IEEE DCC, March 2004, pp. 302–311.
- [32] P. Tan, J. Li, A practical and optimal symmetric Slepian–Wolf compression strategy using syndrome formers and inverse syndrome formers, in: Proceedings of the Allerton Conference on Communications, Control and Computing, September 2005.
- [33] V. Thirumalai, I. Tomic, P. Frossard, Distributed coding of multi-resolution omnidirectional images, in: Proceedings of the IEEE ICIP, September 2007, pp. 345–348.
- [34] T. Tillo, B. Penna, P. Frossard, P. Vanderghenst, Distributed coding of spherical images with jointly refined decoding, in: Proceedings of the IEEE MMSP, November 2005, pp. 1–4.
- [35] I. Tomic, I. Bogdanova, P. Frossard, P. Vanderghenst, Multiresolution motion estimation for omnidirectional images, in: Proceedings of the EUSIPCO, September 2005.
- [36] D. Varodayan, Y.C. Lin, A. Mavlankar, M. Flierl, B. Girod, Wyner–Ziv coding of stereo images with unsupervised learning of disparity, in: Proceedings of the PCS, November 2007.
- [37] F.M.J. Willems, Totally asynchronous Slepian–Wolf data compression, IEEE Trans. Inform. Theory 34 (1) (January 1988) 35–44.
- [38] A.D. Wyner, J. Ziv, The rate-distortion function for source coding with side-information at the decoder, IEEE Trans. Inform. Theory 22 (1) (January 1976) 1–10.
- [39] Q. Xu, Z. Xiong, Layered Wyner–Ziv video coding, IEEE Trans. Image Process. 15 (12) (December 2006) 3791–3802.
- [40] F. Yang, Q. Dai, G. Ding, Multi-view images coding based on multiterminal source coding, in: Proceedings of the IEEE ICASSP, April 2007, pp. 1037–1040.
- [41] R. Zamir, S. Shamai, U. Erez, Nested linear/lattice codes for structured multiterminal binning, IEEE Trans. Inform. Theory 48 (6) (June 2002) 1250–1276.
- [42] X. Zhu, A. Aaron, B. Girod, Distributed compression for large camera arrays, in: Proceedings of the IEEE SSP, September 2003, pp. 30–33.