DISTRIBUTED CODING OF SPHERICAL IMAGES WITH JOINTLY REFINED DECODING

Tammam Tillo*, Barbara Penna*, Pascal Frossard† and Pierre Vandergheynst†

*CERCOM - Center for Multimedia Radio Communications

Dipartimento di Elettronica - Politecnico di Torino - Corso Duca degli Abruzzi 24 - 10129 Torino - Italy

† Ecole Polytechnique Fédérale de Lausanne (EPFL)

Signal Processing Institute, CH-1015 Lausanne, Switzerland

E-mail: {tammam.tillo,barbara.penna}@polito.it, {pascal.frossard,pierre.vandergheynst}@epfl.ch

Abstract— This work addresses the coding of 3-dimensional scenes, as captured by distributed vision sensors with catadioptric cameras. Spherical images allow for avoiding distortion due to the common euclidian assumption in the representation of the plenoptic function. We consider here low complexity encoding of the sensor outputs, in a framework where the cameras could be placed anywhere in the scene, and where the sensors do not communicate to each other. Since multiple spherical images of the same scene most probably provide a redundant representation, we propose to have different compression ratios for different cameras, in order to reduce the overhead of information. The decoder performs a joint decoding of the multiples images, by motion estimation, and joint refinement by consistent inverse quantization. It is finally shown that, even in the absence of any information about the scene or the position of the cameras, the proposed scheme offers improved performance with respect to an independent encoding of the spherical images, especially at low coding rate.

I. INTRODUCTION

EFFICIENT representation and coding of 3-D scenes has recently gained a lot of attention from the research community, fostered by the development of emerging applications in exploration, movie production, virtual reality or even surveillance. While most of the work in this area is focusing on image-based rendering methods, this paper proposes to address the representation of the plenoptic function directly in the spherical domain, under the assumption of perfect vision sensors. This choice presents the main advantage of avoiding the potential discrepancies due to Euclidean approximations in image-based rendering.

In this paper, we consider a framework where catadioptric cameras are distributed arbitrarily in a 3-dimensional scene. The spherical images captured by the multiple cameras are likely correlated, especially the representation of objects that are quite far from the cameras, relatively to the inter-camera distance. In most practical deployment of such systems, sensors often are constrained to implement low complexity tasks,

due to stringent power constraints. In particular, we focus here on a system where cameras do not communicate together, but only with the central decoder. An efficient representation of the complete scene therefore relies on low complexity distributed coding, but joint decoding of the images sent by the multiple cameras: this approach is well known as the distributed source coding (DSC) paradigm [4].

The most closely related work is certainly the DSC coding scheme proposed in [5], which however relies on the a priori knowledge about objects position and their distance to the cameras, in a euclidian framework. The traditional DSC paradigm is generally based on Slepian-Wolf coding, with coset codes for example [1], and allows to achieve good coding performance when the correlation between sources can be accurately modelled. In our framework, we however consider that no a priori information is available about the scene, or the sensor positions. In this context, we have work without source correlation model, and we therefore suggest to use a scheme which is based on successive quantization refinement. Each cameras provides a multi-resolution representation of the scene, possibly with a different fidelity. The multi-resolution representation is implemented with a spherical Laplacian Pyramid (sLP), whose main advantage resides in its isotropic characteristics. A central decoder finally tries to identify the redundant information between multiple images by motion estimation, and reconstruct the scene by joint refinement with consistent inverse quantization.

The rest of this paper is organized as follows. Section II presents the particular framework considered in this paper. Section III presents the distributed encoding strategy, and Section IV describes the joint decoding of spherical images. Simulation results are presented in Section V, and Section VI concludes in proposing directions for improvements of the coding scheme.

II. DISTRIBUTED CODING OF SPHERICAL IMAGES

We consider a framework where catadioptric cameras are distributed in a 3-dimensional scene, without any restriction about their particular locations. We further assume that the system does not benefit from any a priori information about
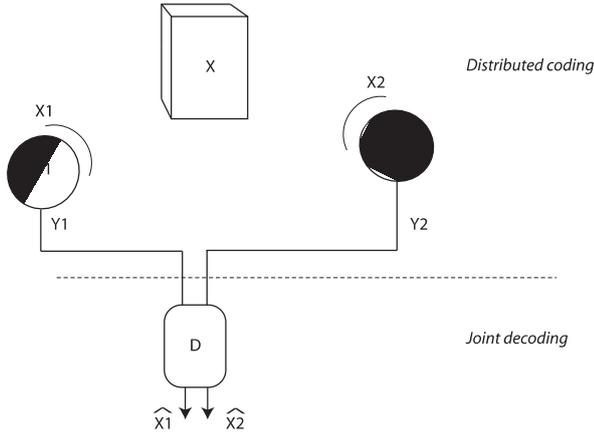
Fig. 1. Distributed coding of spherical images.

the scene itself. The vision sensors provide spherical images of the scene, which are likely to be highly correlated in most practical scenarios. The cameras do not communicate together due to strict constraints in power consumption, but only exchange information with a central decoder. The framework is represented in Figure 1 in the case of two sensors, S_1 and S_2. These sensors capture spherical images X_i of the scene X, whose compressed versions Y_i are sent to the central decoder.

Under the previous assumptions, the problem consists in providing an accurate representation of the scene X, reconstructed from the decoder outputs $\hat{X}_1$ and $\hat{X}_2$. This is a typical distributed coding scenario, that is however further constrained by the absence of information about the scene, and the sensor positions. The sensors independently encode the spherical images, that are jointly decoded by the central receiver, which tries to efficiently identify the correlation between the different images.

We concentrate now on the particular problem where the bandwidths between the sensors and the central decoder are also constrained, with unbalanced rate allocation. In this case, the compression ratio varies between sensors, and the lowest rate images benefit from the information in other images to refine their quality at the central decoder. In other words, one camera, e.g., S_1, serves as the main camera, which sends a fine representation Y_1 of the scene to the decoder; all other cameras ($S_n; n > 1$) send a coarser granularity version of the scene, which will be improved at the decoder by joint decoding with other images, in particular Y_1. Note that the extension of the problem to symmetrical scenarios where the rate constraints are equivalent for each sensor is straightforward. Even if we focus here on the unbalanced rate constraint case, the coding scheme proposed in this paper is flexible enough to be adapted to the symmetrical case, where the problem interestingly becomes quite similar to a balanced multiple description coding paradigm.

III. DISTRIBUTED ENCODERS

The multiple spherical views acquired by recent catadioptric cameras are generally characterized by an enormous amount of information. Among methods adapted to handle this type of data, we can mention the spherical wavelet [2], which is a multi-resolution approach to handle the images directly on the sphere, and the Laplacian Pyramid [3] scheme that can be extended to the spherical case. The spherical wavelet is well suited to analyze the spherical data, but due to the lower redundancy of the spherical Laplacian pyramid, we use the latter for compression.

The spherical Laplacian pyramid (sLP) decomposition performs as follows. A spherical image $\mathcal{S}$ is first convolved with a lowpass filter and downsampled to provide a low resolution representation of the original image, $\mathcal{S}_a^1$. Subtracting the upsampled and interpolated version of $\mathcal{S}_a^1$ from the initial image $\mathcal{S}$ results in the detail sphere $\mathcal{S}_d^1$. This process can be further iterated to produce the sub-spheres $\mathcal{S}_a^2$ and $\mathcal{S}_d^2$ from $\mathcal{S}_a^1$, and eventually leads to a multiresolution representation of the spherical image. The multiresolution representation is particularly interesting in the unbalanced rate constraints scenario, and quite beneficial for a hierarchical motion estimation at the decoder, as explained later. It is worth pointing out that this representation is redundant; in fact, the redundancy can be simply evaluated as $\rho = \frac{1-4^{-L}}{3}N$, where N represents the original number of pixels, and L is the number of levels in the decomposition.

Compression of the spherical images is then achieved by scalar quantization of the coefficients in the spherical Laplacian pyramid, and finally entropy coding. Note that, even if the sLP is a redundant transform, simple scalar quantization provides sufficient performance in our framework that mostly targets low bit rate coding of images. The bit budget attributed to a sensor is allocated to the different sub-spheres of the sLP, proportionally to the variance of the coefficients. The bit budget allocated to each sensor is given by the rate constraints of the system, or imposed by the central decoder that tries to efficiently reduce the overall coding rate of the multi-camera system. In the two camera setup considered here, the compressed image Y_1 and Y_2 are respectively allocated a bit budget of R_1 and R_2 bits, with $R_2 < R_1$. The coefficients of each sub-sphere of both spherical Laplacian pyramids are quantized by an entropy constrained quantizer, driven by the coefficient energy distribution.

IV. JOINT DECODING WITH REFINED INVERSE QUANTIZATION

When multiple compressed views are present at the decoder, their correlation can be exploited to enhance the quality of the representation of the 3-dimensional scene. In particular, the coarsely quantized image Y_2 can be refined by using information present in the image with higher coding rate, Y_1. In order to identify the correlation between spherical images,

we propose to use a motion estimation procedure at decoder [6]. The motion estimation tries to identify similar objects present in different images, and allows to create a motion compensated image $\tilde{Y}_2$ from the high resolution image Y_1, which provides an approximation of the low resolution image Y_2.

In our scheme, motion vector estimation is performed on approximation sub-spheres, in order to refine the detail sub-spheres in the sLP. It is achieved by initially dividing the approximation sub-sphere of the second view $\mathcal{Y}_{2a}^{L}$ into macro-blocks. Macro-blocks on the second view are then matched with the best approximating macro-blocks in the first view, in a mean square error sense. The displacement between corresponding macro-blocks is considered as a motion vector. In order to get a finer representation of the detailed sub-sphere $\mathcal{Y}_{2d}^{L}$, we use the previously obtained motion vector to build a motion compensated sub-sphere using the first view, as $\tilde{\mathcal{Y}}_{2d}^{L} = MC(\mathcal{Y}_{1d}^{L})$. The compressed image sub-sphere $\mathcal{Y}_{2d}^{L}$, and its motion compensated approximation $\tilde{\mathcal{Y}}_{2d}^{L}$ are merged together to refine the decoding of the low coding rate sub-sphere, according to the merging procedure explained here-below. The refined sub-sphere is used with the approximation sub-sphere $\mathcal{Y}_{2a}^{L}$ to reconstruct the higher level approximation sub-sphere $\mathcal{Y}_{2a}^{(L-1)}$. In parallel, the corresponding sub-sphere in the first view is also generated as $\mathcal{Y}_{1a}^{(L-1)}$. This procedure is further iterated on the successive levels of the sLP, until the second view $\hat{X}_2$ is fully reconstructed. The first view $\hat{X}_1$ is decoded by the simple inversion of the Laplacian pyramid decomposition.

The refinement stage that merges information from motion-compensated and low coding rate sub-spheres finally performs as follows. Assume that x_1 and x_2 are quantized coefficients of similar views of the same object, according to the motion estimation step. They however belong to two different spherical images, and their values after quantization are thus likely to be different. If they however represent the same coefficient x, we can consider in a first approximation that $x_1 \approx x_2 = x$. Since this coefficient x has been quantized with two different quantizers having different step sizes Δ_1 and Δ_2, the dequantized values at the decoder are:

$$y_1 = round(\frac{x - \mu_1}{\Delta_1})\Delta_1 + \mu_1$$
$$y_2 = round(\frac{x - \mu_2}{\Delta_2})\Delta_2 + \mu_2,$$

where μ_1 and μ_2 represent the mean values of the two sub-spheres. Since $x - 0.5 < round(x) \leq x + 0.5$, we have:

$$x - 0.5\Delta_1 < y_1 \leq x + 0.5\Delta_1$$
$$x - 0.5\Delta_2 < y_2 \leq x + 0.5\Delta_2,$$

that combine in:

$$-0.5(\Delta_1 + \Delta_2) < y_1 - y_2 < 0.5(\Delta_1 + \Delta_2). \qquad (1)$$
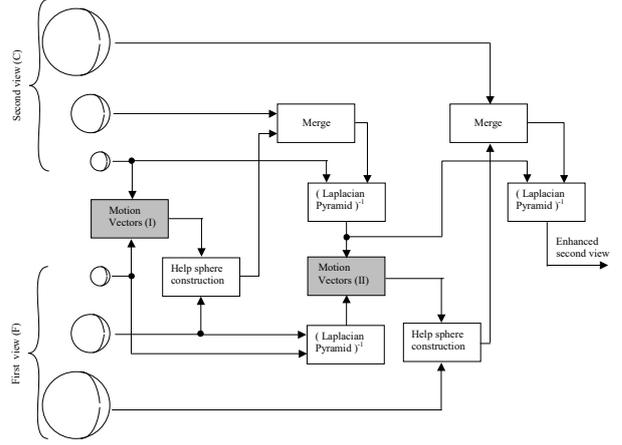

Fig. 2. The full block diagram of the joint decoder; $(\mathsf{Laplacian\ Pyramid})^{-1}$ represent one level Laplacian Pyramid composition.
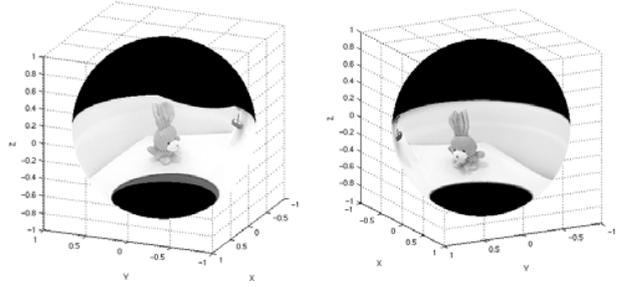

Fig. 3. The two real spherical images acquired in the laboratory.

It means that dequantized coefficients y_1 and y_2, taken from first and second view respectively, which represent the same coefficient x, should satisfy the inequality given in Eq. (1). Therefore the refinement stage simply evaluates the difference between dequantized versions of corresponding coefficients, and refines the value of coarsely quantized coefficient if the inequality given in Eq. (1) is satisfied. In this case, both y_1 and y_2 are indeed considered to represent the same coefficient. When Eq. (1) is not verified, the value of the quantized coefficients is simply left unchanged. Typically, the inequality Eq. (1) does not hold for coefficients that represent different parts of the same object, or when the motion vector estimation fail to identify object translation between two views. The joint decoding procedure with refinement is illustrated in Figure 2.

V. SIMULATIONS AND EXPERIMENTAL RESULTS

The performance of the distributed coding algorithm with joint refinement at decoder is now evaluated in the symmetrical two-sensor scenario. The described distributed source coding scheme has been applied to two spherical images (Fig. 3)
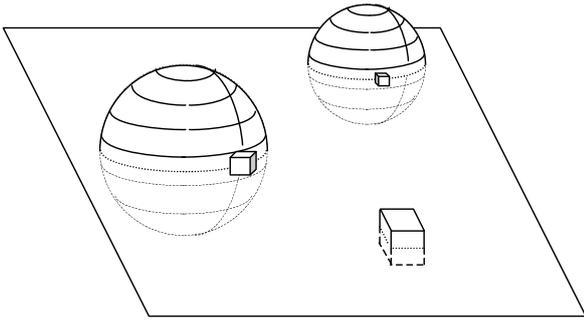
Fig. 4.   Stripe-based motion estimation on spherical images.

Fig. 5.   Rate distortion curve of: (a) the first view; (b) the second view with the proposed DSC scheme and the no joint decoding approach.

acquired in the laboratory, which represent an object on the table. The spherical Laplacian Pyramid has been used to decorrelate the views with three level of decomposition, using Haar filters adapted to the spherical framework. Furthermore, a stripe based motion estimation strategy is adopted in order to reduce the complexity of motion estimation stage: macro-blocks are searched in a portion of a stripe rather than with a full search approach, as illustrated in Figure 4. Also, the quantization of the approximation sub-spheres in both images is identical, in order to ensure that the motion estimation at the first approximation level is accurate. We use adaptive arithmetic coder as the entropy coding stage.

We compare the performance of joint decoding with refinement, with independent decoding of both spherical images. Figure 5 represents the evolution of the quality (PSNR) as a function of the coding rate, for both images. Figure 5-(a) reports the performance of the encoding of the first view, which is the high coding rate image in our scenario. Figure 5-(b) compares the performance of the proposed DSC scheme, and the independent decoding approach for the second image. We can notice that a gain of nearly 4 dB is obtained at low rate by the proposed DSC approach over the independent decoding scheme. The results show that the gain is substantial for low to medium bit rate, whereas at high rate, the gap decreases; this can be explained by the fact that at high rate, any gain due to quantization refinement can be compromised by an unprecise motion estimation, and by the small differences between the same region viewed from different points of view in case of non uniform light conditions. Further improvement could be obtained if the encoded part at low bit rate is exploited by the joint decoding. This can be obtained by using two different quantizers (mid-rise and mid-tread), in order to combine the two different quantizers results into better representation of some values. Improved performance can also be obtained by incorporating the quantizer in the Laplacian pyramid structure, in order to limit the propagation of quantization error throughout the various levels of the pyramid, that becomes non-negligible when coding rate increases.
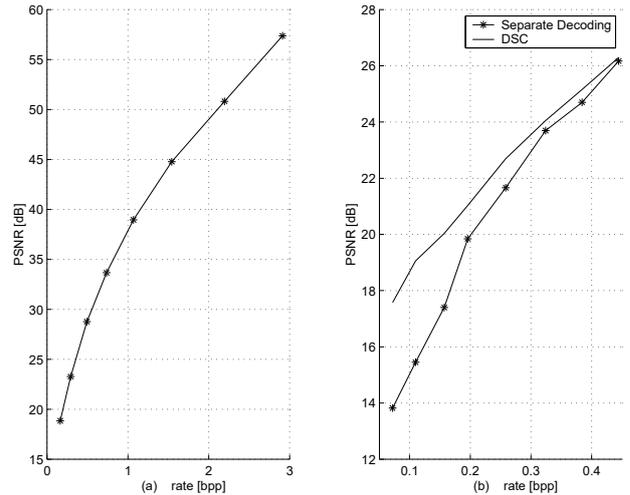
VI. CONCLUSIONS AND FUTURE WORKS

We proposed a scheme for distributed coding of spherical images, that relies on motion compensation for joint refinement at decoder. The rate distortion performance in a framework where encoders with a priori unknown position do not communicate together, exhibits an improvement over independent decoding of the spherical images, especially at low rate. We plan to improve the consistent reconstruction step by generalization to more efficient filters. Finally, the addition of a Slepian-Wolf coding stage adapted to the particular framework considered in that paper, is expected to yet improve the performance towards higher coding rate.

VII. ACKNOWLEDGMENTS

REFERENCES

[1] S.Sandeep Pradhan and Kannan Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction", *Proc. IEEE Data Compression Conference (DCC)*, 1999.
[2] J.P. Antoine, L. Demanet, L. Jacques, P. Vandergheynst, "Wavelets on the sphere :Implementation and Approximation", *Appl. Comput. Harmon. Anal.*, vol. 13, 2002, pp. 177-200.
[3] P.J. Burt et A.E. Adelson, "The Laplacian pyramid as a compactimage code", *IEEE Trans. on Communications*, vol. 31, pp. 532-540, 1983.
[4] D. Slepian and J.K. Wolf, Noiseless coding of correlated information sources, *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471-480, July 1973.
[5] N. Gehrig and P.L. Dragotti, "Distributed compression of the plenoptic function", *Proceedings of the International Conference on Image Processing*, vol. 1, pp. 529 - 532, Oct. 2004.
[6] I. Tosic, I. Bogdanova, P. Frossard and P. Vandergheynst, "Multiresolution Motion Estimation for Omnidirectional Images", *Proceedings of EUSIPCO 2005, accepted for publication.*, Sep. 2005.