DIFFERENT - DISTRIBUTED AND FULLY FLEXIBLE IMAGE ENCODERS FOR CAMERA SENSOR NETWORKS

Nicolas Gehrig and Pier Luigi Dragotti

Communications and Signal Processing Group, Electrical and Electronic Engineering Department
Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom
e-mail: {nicolas.gehrig, p.dragotti}@imperial.ac.uk

ABSTRACT

In this paper, we propose a practical coding approach for the problem of distributed compression of multi-view images. Our coding technique is based on a tree structured compression algorithm that guarantees an optimal rate-distortion behaviour for piecewise polynomial signals. We model the different views using a piecewise polynomial function whose singularity positions are shifted from one view to the others according to the constraints imposed by the structure of the plenoptic function [1]. We show that, starting from the optimal tree decompositions of the different views, only partial information from each tree is necessary at the decoder in order to reconstruct all the different approximations. We first present our approach in the more intuitive 1D case and show that it can be used with arbitrary bit-rate allocation. Then, we propose a construction for the case of N different views that satisfies a certain bit conservation principle. Finally, we show how our approach can be extended to the 2D case.

1. INTRODUCTION

Separate lossless encoding of two correlated discrete sources can be as efficient as joint encoding, assuming that the two compressed signals can be jointly decoded. This surprising result is known as the Slepian-Wolf theorem [2] and was proposed more than thirty years ago. However, practical coding approaches based on this (asymptotic and non-constructive) theoretical result have only been proposed recently (see for instance [3, 4, 5, 6]).

We consider the problem of distributed compression in camera sensor networks. Because of the proximity of the cameras, the images obtained from different viewing positions can be highly dependent. According to Slepian-Wolf, it should therefore be possible to develop distributed compression techniques that can exploit this correlation in order to reduce the overall amount of information that has to be transmitted from the sensors to a common receiver, without requiring any collaboration between the different encoders.

Distributed source coding schemes usually rely on the assumption that the correlation of the source is known a-priori. In [7], we showed how the correlation in the visual information, which is related to the structure of the plenoptic function [1], can be estimated using simple geometrical constraints on the scene and on the position of the cameras. We then proposed a coding approach that can exploit this correlation in order to perform distributed compression of the different views. Nevertheless, our approach was based on an a-priory knowledge of the different object boundaries at the encoders and was therefore not directly applicable to encode real multi-view images. In this paper, we show how a particular image coder based on tree structured algorithms [8] can be modified to take advantage of our distributed coding approach. Notice that several other distributed compression approach for multi-view images has been proposed (see [9, 10] for example). The main difference is that our approach tries to estimate the correlation structure in the visual information using some geometrical information. The main properties of our "DIFFERENT" scheme is that a) it allows for an arbitrary partition of the rate between the encoders and b) it satisfies a bit conservation principle. Namely, the reconstruction fidelity at the decoder does not depend on the number of sensors involved, but only on the total bit-budget R.

The paper is organized as follows: The next section introduces the camera sensor network scenario we consider in our work and gives a brief review of some of our previous results. In Section 3, we first introduce the tree structured algorithm we consider in this work and present our coding strategy focusing on the more intuitive 1D case. Then, we show that our approach allows arbitrary bit-rate allocation and we highlight an exact bit conservation principle for the case of more than two cameras. Finally, an extension to the 2D case is highlighted. Section 4 presents the different simulation results and concluding remarks are given in Section 5.

This work is supported in part by DIF-DTC project number 12.6.2. and EOARD 043061.

2. OUR CAMERA SENSOR NETWORK SCENARIO AND REVIEW OF OUR PREVIOUS RESULTS

The camera sensor network scenario we consider is illustrated in Figure 1. We assume that N cameras are placed on a line and that all the objects of the scene have a distance to the cameras that is bounded between a minimum and a maximum values (z_{min}, z_{max}). Assuming that the distance between two cameras is not larger than a certain distance α, this scenario ensures that, regardless of the complexity of the scene, any disparity Δ will be contained in the range: $[\frac{\alpha f}{z_{max}}; \frac{\alpha f}{z_{min}}]$ where f is the (common) focal length of the cameras. Based on this observation, our distributed coding strategy [7] consists in sending only partial information of the positions of the objects from the different encoders, as recalled in Sections 2.1 and 2.2.

2.1. Asymmetric Encoding

Assume X and Y are the (discrete) positions of a specific object on the images obtained from two consecutive cameras. We know that their difference (or disparity) is contained in the range

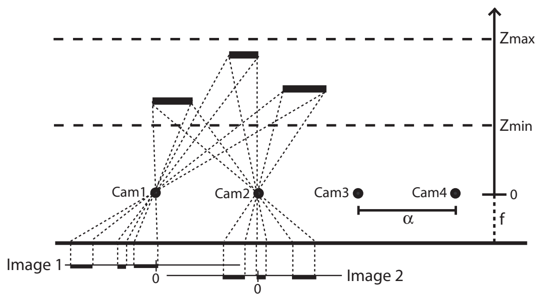

Fig. 1. Our camera sensor network scenario

$[\frac{\alpha f}{z_{max}}; \frac{\alpha f}{z_{min}}]$. Assume that X and Y are uniformly distributed and that each of them can be perfectly encoded independently using R bits. In [7], we showed that sending the full information (R bits) from one of the encoders and only the R_{min} least significant bits from the other one is sufficient to allow for a perfect reconstruction at the decoder ($R_{min} = \lceil \log_2(\delta + 1) \rceil$, where $\delta = \lceil \alpha f(\frac{1}{z_{min}} - \frac{1}{z_{max}}) \rceil$).

2.2. Symmetric Encoding

The coding strategy presented in Section 2.1 can be extended to allow for a more flexible allocation of the bit-rates amongst the encoders. Our coding approach [7] suggests that each encoder should send the last R_{min} bits of the object's positions along with a certain subset of their first $R - R_{min}$ bits. The subsets should be chosen such that they are complementary. This strategy can theoretically achieve the Slepian-Wolf bound and gives us a precise intuition of how distributed compression should be applied to multi-view images. However, it is not directly applicable to encode real multi-view images. In the next section, we show how an existing image coder can be extended to take advantage of our distributed coding approach.

3. DISTRIBUTED COMPRESSION USING TREE STRUCTURED ALGORITHMS

Since the correlation model used by our distributed coding approach is related to the object's positions on the different views, we need to develop coding algorithms that can efficiently represent these positions. Our approach consists in representing the different views using a piecewise polynomial model. The main advantage of such a representation is that it is well adapted to represent real images and that it is able to precisely catch the discontinuities between objects. Two different views can therefore be modeled using a piecewise polynomial signal where each discontinuity is shifted according to the correlation model $\Delta_i \in \{\Delta_{min}, \Delta_{max}\}$. Moreover, if we assume that the scene is composed of lambertian planar surfaces only and that no occlusion occurs in the different views, then we can claim that the polynomial pieces are similar for the different views.[1]

[1] With non-lambertian surfaces, or with the presence of occlusions, the polynomial pieces can differ for the different views. Our simple correlation model should therefore be modified in this case. For the sake of simplicity, we will however only consider this simple model to present our coding approach.

3.1. The prune-join tree decomposition algorithm

In [8], Shukla et al. presented new coding algorithms based on tree structured segmentation that achieve the correct asymptotic rate-distortion (R-D) behaviour for piecewise polynomial signals. Their method is based on a prune and join scheme that can be used for 1D (using binary trees) or for 2D (using quadtrees) in a similar way. We give here a sketch of their compression algorithm for 1D signals (Algorithm 1) and encourage the reader to refer to the original work [8] for more details.

Algorithm 1 Prune-Join binary tree coding algorithm [8]

1: Segmentation of the signal using a binary tree decomposition up to a tree depth J_{max}.
2: Approximation of each node of the tree by a polynomial $p(t)$ of degree $\leq P$.
3: Rate-Distortion curves generation for each node of the tree (scalar quantization of the Legendre polynomial coefficients).
4: Optimal pruning of the tree for the given operating slope $-\lambda$ according to the following Lagrangian cost based criterion: Prune the two children of a node if $(D_{C_1} + D_{C_2}) + \lambda(R_{C_1} + R_{C_2}) \geq (D_p + \lambda R_p)$.
5: Joint coding of similar neighbouring leaves according to the following Lagrangian cost based criterion: Join the two neighbours if $(D_{n_1} + \lambda R_{n_1}) + (D_{n_2} + \lambda R_{n_2}) \geq (D_{n_{Joint}} + \lambda R_{n_{Joint}})$.
6: Search for the desired R-D operating slope (update λ and go back to point 4).

3.2. Our distributed coding strategy for 1D piecewise polynomial functions

Let $f_1(t)$ be a piecewise polynomial signal defined over $[0; T]$ consisting of $S+1$ polynomial pieces of maximum degree P each, and bounded in amplitude in $[0; A]$. Let $\{t_{1_i}\}_{i=1}^{S}$ represent the S distinct discontinuity locations of $f_1(t)$. We define $f_2(t)$ as another piecewise polynomial function over $[0; T]$ having the same polynomial pieces than $f_1(t)$, but whose set of discontinuity locations $\{t_{2_i}\}_{i=1}^{S}$ is chosen such that: $\Delta_{min} \leq t_{2_i} - t_{1_i} \leq \Delta_{max}, \forall i \in \{1, \ldots, S\}$. The relationship between $f_1(t)$ and $f_2(t)$ is therefore given by the range of possible disparities $[\Delta_{min}; \Delta_{max}]$ which corresponds to the plenoptic constraints we consider in our camera sensor network scenario.

Assume that these two signals are independently encoded using the algorithm presented in the previous subsection for a given distortion target. The total information necessary to describe each of them can be divided in 3 parts: R_{Tree} is the number of bits necessary to code the pruned tree and is equal to the number of nodes in the tree. $R_{LeafJointCoding}$ is the number of bits necessary to code the joining information and is equal to the number of leaves in the tree. Finally, R_{Leaves} is the total number of bits necessary to code the set of polynomial approximations.

Figure 2 presents a prune-join tree decompositions of two piecewise constant signals, having the same set of amplitudes and having their sets of discontinuities satisfying our plenoptic constraints. Because of these constraints, we can observe that the structure of the two pruned binary trees present some similarities. Our distributed compression algorithm uses these similarities in order to transmit only the necessary information to allow for a complete reconstruction at the decoder. It can be described as follows (asymmetric encoding):

- Send the full description of signal 1 from encoder 1. ($R_1 = R_{Tree_1} + R_{LeafJointCoding_1} + R_{Leaves_1}$)
- Send only the subtrees of signal 2 having a root node at level J_Δ along with the joining information from encoder 2, where $J_\Delta = \lceil \log_2(\frac{T}{\Delta_{max}-\Delta_{min}+1}) \rceil$. ($R_2 = (R_{Tree_2} - R_{\Delta_2}) + R_{LeafJointCoding_2}$ where R_Δ corresponds to the number of nodes in the pruned tree with a depth smaller than J_Δ.)

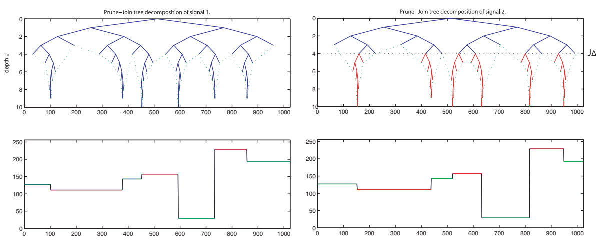

Fig. 2. Prune-Join binary tree decomposition of two piecewise constant signals satisfying our correlation model.

At the decoder, the original position of the subtrees received from encoder 2 can be recovered using the plenoptic constraints (i.e. $\Delta \in [\frac{\alpha f}{z_{max}}; \frac{\alpha f}{z_{min}}]$) and the side information provided by encoder 1. The full tree can then be recovered and the second signal can thus be reconstructed using the set of amplitudes received from encoder 1.

3.3. Arbitrary bit-rate allocation

The construction proposed in the previous subsection is asymmetric since encoder 1 has to transmit its whole information whereas encoder 2 only transmits missing information from its pruned tree structure and its joining information. In order to allow for an arbitrary bit-rate allocation between the two encoders, we propose the following strategy:

- Send all the subtrees having a root node at level J_Δ from both encoders, along with their joining information.
- Send complementary parts of the two upper trees (depth $< J_\Delta$).
- Send complementary subsets of the polynomial approximations.

This coding approach presents the same total rate-distortion behaviour than the asymmetric approach, but allows for an arbitrary bit-rate allocation (see Figure 3 and Table 1 for simulation results).

3.4. An exact bit conservation principle

Our distributed compression approach can be extended to more than two cameras. We consider the multi-camera scenario presented in Figure 1 with N cameras. If we consider that there is no occlusion in these different views, knowing the discontinuity locations from only two cameras is sufficient to reconstruct any view in between. Assume that we have a total bit budget of R_{tot} bits to allocate between the different encoders and that only the two most distant cameras are transmitting information to the decoder. In this case, we know that each encoder should first compute a representation of its input signal using $R = R_{Tree} + R_{LeafJointCoding} + R_{Leaves}$ bits such that $(R_{Tree} - \frac{1}{2}R_\Delta) + R_{LeafJointCoding} + \frac{1}{2}R_{Leaves} = \frac{1}{2}R_{tot}$ (symmetric encoding). At the decoder, the two views can be reconstructed and any new view in between can be interpolated with a certain fidelity.

Assume now that we want to transmit some information from other cameras as well. We can show that, as long as the subtrees (depth $\geq J_\Delta$) are transmitted from the two extreme cameras, the rest of the information can be obtained totally arbitrarily from any set of cameras. In particular, since the polynomial pieces are similar for each view, they can be transmitted from any camera, all this without impairing reconstruction quality. This result gives us an exact bit conservation principle (see Figure 4 and Table 2 for simulation results).

3.5. Extension to 2D using quadtree decompositions

The prune-join binary tree decomposition used in our approach has an intuitive extension to the 2D case, where the binary tree segmentation is replaced by the quadtree segmentation and the polynomial model is replaced by a 2D geometrical model. Although our approach becomes more involved in the 2D case, the intuitions remain the same. The geometrical model used in 2D corresponds to two 2D polynomials separated by a 1D polynomial boundary. Notice that the quadtree compression algorithm proposed in [8] outperforms Jpeg2000. For this reason, we are confident that its use in the multi-view context will lead to good simulation results. This is, however, part of our ongoing work.

4. NUMERICAL RESULTS

4.1. Piecewise polynomial signals

We have applied our distributed compression approach to different sets of piecewise polynomial signals in order to highlight the arbitrary bit-rate allocation and the bit conservation principle presented in the previous section. In Table 1, we show that an independent encoding of the two signals presented in Figure 3 requires a total bit-rate of .304 bpp to achieve a distortion target (SNR) of 26 dB. Using our distributed compression approach, we can see that a quarter of the total bit-rate can be saved for an identical reconstruction fidelity. Moreover, this result remains constant for any choice of bit-rate allocation (SW asym. or SW sym.). Notice that a similar reduction of bit-rate with an independent encoding strategy would see the total SNR drop of about 10 dB.

Table 1. Arbitrary bit-rate allocation.

Coding Strategy	$\mathbf{R_1}$ (bpp)	$\mathbf{R_2}$ (bpp)	$\mathbf{R_{tot}}$ (bpp)	$\mathbf{D_{tot}}$ (SNR) (dB)
Independent	.307	.301	.304	26.01
SW asym. 1	.307	.148	.228	26.01
SW asym. 2	.154	.301	.228	26.01
SW sym.	.227	.227	.227	26.01

In Table 2, we highlight our bit conservation principle, by applying our compression approach to the three views shown in Figure 4. We know that the two extreme views represent sufficient information to allow the reconstruction of any view in between with a comparable fidelity. Applying our SW approach to these two extreme views, a total of 235 bits is necessary to achieve a distortion

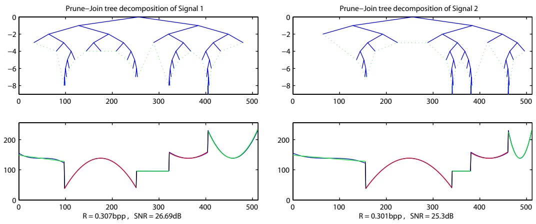

Fig. 3. Join-Prune tree decomposition of two piecewise polynomial signals with shifted discontinuities for a given distortion target (26 dB).

(SNR) of about 26 dB for the three reconstructed views. A similar global rate-distortion behaviour holds when part of the information is transmitted from the central view.[2] This result highlights our bit conservation principle. In other words, if we assume that the sensors transmit their compressed data to the central decoder using a multi-access channel, the fidelity of the reconstructed views only depends on the global capacity of this channel and not on the number of sensors used!

Table 2. An exact bit conservation principle.

Coding Strategy	R_1 (bits)	R_2 (bits)	R_3 (bits)	R_{tot} (bits)	D_{tot} (SNR) (dB)
Independent	157	124	157	438	26.13
SW - 2 views	117	0	118	235	25.68
SW - 3 views	82	71	82	235	26.13

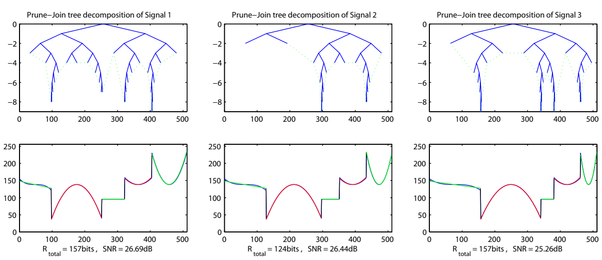

Fig. 4. Join-Prune tree decomposition of three piecewise polynomial signals for a given distortion target (26 dB).

4.2. Results on scan lines of stereo images

In order to justify the correlation model we use in our distributed compression scheme, we tried to apply it to a set of scan lines of real multi-view images. We present a simulation on scan lines of a pair of stereo images (Figure 5) using a piecewise linear model and a symmetric encoding strategy. The reconstructed signals present a good level of accuracy for the discontinuity locations. However, since the assumption of lambertian surfaces does not hold for this scene, the polynomial pieces can sometime be slightly different from one view to the other. A solution to this problem is to include it in our correlation model.

[2]The small variations in the SNR values are only due to different quantization errors.

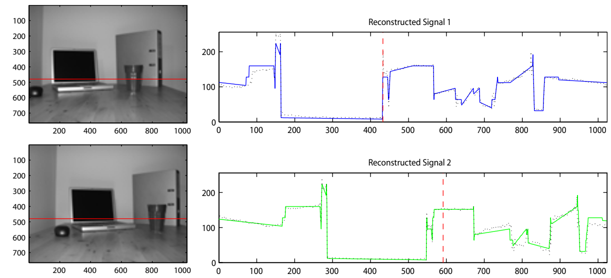

Fig. 5. (left) Stereo images of a real scene where the objects are located between a minimum and a maximum distance from the cameras. (right) Reconstructed scan lines using a piecewise linear model for the binary tree decomposition and a symmetric distributed compression.

5. CONCLUSIONS AND ONGOING RESEARCH

We have shown how our distributed compression approach proposed in [7] can be used with a real image coder using tree structured algorithms. We have shown that our approach allows for an arbitrary bit-rate allocation, and we have highlighted an exact bit conservation principle. Ongoing research focuses on an efficient implementation of our distributed compression approach for the 2D case with more complex correlation models.

6. REFERENCES

[1] E.H. Adelson and J.R. Bergen, “The plenoptic function and the elements of early vision,” in *Computational Models of Visual Processing*, M. Landy and J. Anthony Movshon, Eds. 1991, pp. 3–20, MIT Press, Cambridge, MA.

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

[3] S.S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): Design and construction,” in *Data Compression Conference*, 1999, pp. 158–167.

[4] J. Garcia-Frias, “Compression of correlated binary sources using turbo codes,” *IEEE Communications Letters*, vol. 5, no. 10, pp. 417–419, October 2001.

[5] A. Aaron and B. Girod, “Compression with side information using turbo codes,” in *IEEE International Conference on Data Compression*, April 2002, pp. 252–261.

[6] A.D. Liveris, Z. Xiong, and C.N. Georghiades, “Compression of Binary Sources With Side Information at the Decoder Using LDPC Codes,” *IEEE Communications Letters*, vol. 6, no. 10, pp. 440–442, October 2002.

[7] N. Gehrig and P.L. Dragotti, “Distributed compression of the plenoptic function,” in *IEEE Int. Conf. on Image Processing (ICIP)*, October 2004, pp. 529–532.

[8] R. Shukla, P.L. Dragotti, M.N. Do, and M. Vetterli, “Rate-distortion optimized tree structured compression algorithms for piecewise polynomial images,” *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 343–359, March 2005.

[9] X. Zhu, A. Aaron, and B. Girod, “Distributed compression for large camera arrays,” in *Proceedings of the IEEE Workshop on Statistical Signal Processing, SSP-2003*, September 2003.

[10] A. Jagmohan, A. Sehgal, and N. Ahuja, “Compression of lightfield rendered images using coset codes,” in *Asilomar Conf. on Signals and Systems, Special Session on Distributed Coding*, 2003.