Distributed Sampling and Compression of Scenes with Finite Rate of Innovation in Camera Sensor Networks*

Nicolas Gehrig and Pier Luigi Dragotti
Communications and Signal Processing Group, EEE Department
Imperial College London, Exhibition Road, London SW7 2AZ, U.K.
+44 (0)20 759-46192 | {nicolas.gehrig, p.dragotti}@imperial.ac.uk

Abstract

We study the problem of distributed sampling and compression in sensor networks when the sensors are digital cameras that acquire a 3-D visual scene of interest from different viewing positions. We assume that sensors cannot communicate among themselves, but can process their acquired data and transmit it to a common central receiver. The main task of the receiver is then to reconstruct the best possible estimation of the original scene and the natural issue, in this context, is to understand the interplay in the reconstruction between sampling and distributed compression.

In this paper, we show that if the observed scene belongs to the class of signals that can be represented with a finite number of parameters, we can determine the minimum number of sensors that allows perfect reconstruction of the scene. Then, we present a practical distributed coding approach that leads to a rate-distortion behaviour at the decoder that is independent of the number of sensors, when this number increases beyond the critical sampling. In other words, we show that the distortion at the decoder does not depend on the number of sensors used, but only on the total number of bits that can be transmitted from the sensors to the receiver.

1 Introduction

Sensor networks have been attracting a significant interest in recent years. Advances in wireless communication technologies and hardware have enabled the design of these cheap low-power miniature devices that make up sensor networks. The truly distributed (or decentralized) nature of sensor networks is radically changing the way in which we sense, process and transport signals of interest. In the common "many-to-one" scenario, sensors are spatially deployed to monitor some physical phenomenon of interest, and transmit independently their measurements to a central receiver. The main task of the receiver is then to reproduce the best possible estimation of the observed phenomenon, according to some distortion measure. Sampling in space and distributed coding are clearly two critical issues in sensor networks. For example, deploying to few sensors would lead to a highly aliased reconstruction of the

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

phenomenon, while a too large number of sensors would use all the communication resources by transmitting highly correlated measurements to the receiver. This last issue can actually be addressed by means of distributed source coding techniques [1]. The correlated measurements can then be encoded independently at each sensor with a compression performance similar to what would be achieved by using a joint encoder.

Despite this powerful coding approach, several authors have presented pessimistic results regarding the scalability of sensor networks [2, 3]. Their main argument comes from the fact that, if the total amount of data that can be received by the central decoder is limited, then the throughput at each sensor scales as $\Theta(\frac{1}{N})$ with the number of sensors N. The global performance of the network then goes to zero as $N \to \infty$. More optimistic results were recently proposed in [4, 5], where it was shown that, for a given distortion, an upper-bound (independent of N) on the total information rate can be given. Recent works have also shown that if the observed phenomenon can be represented with a finite number of degrees of freedom, then it is possible to efficiently trade-off the density of sensors with the sensing resolution at each sensor [6, 7].

In this paper, we consider camera sensor networks, where each sensor is equipped with a digital camera and is placed at a certain viewing position around a scene of interest. In this context, the physical phenomenon that has to be transmitted to the receiver is the visual information coming from the scene (or its plenoptic function [8]), and the samples are the different sampled 2-D views acquired by the sensors. The issues of sampling and communication are particularly interesting in this scenario. Several sampling theorems have been proposed to address the question of determining the critical sampling of a visual scene under different model assumptions, such as bandlimited scenes [9] or scenes with finite rate of innovation [10]. Starting from this critical sampling, our main objective is to show how we can arbitrarily increase the number of sensors, while maintaining a constant global rate-distortion behaviour for the reconstruction of the scene at the receiver. This "bit-conservation principle" is achieved by means of our distributed compression scheme for multi-view images proposed in [11, 12].

2 Camera Sensor Network Set-up

We consider the camera sensor network set-up proposed in Figure 1, where N pinhole cameras are placed on a line and observe a common scene in the direction perpendicular to the cameras' line. The distance α between any two consecutive cameras is known, all the cameras have the same focal length f and the distance from the cameras to any object of the scene is at least Z_{min}. The observed scene is made of L Lambertian planar polygons that can be tilted and placed at different depths. Each of these polygons has a certain polynomial intensity (the intensity along the x and y axes varies as a polynomial of maximum degree Q). The perspective projection observed at each camera is therefore given by a 2-D piecewise polynomial function. The difference between the N views is that the pieces are shifted differently according

to their depths (pieces can be linearly contracted or dilated if they correspond to a tilted object, as shown in Figure 2).

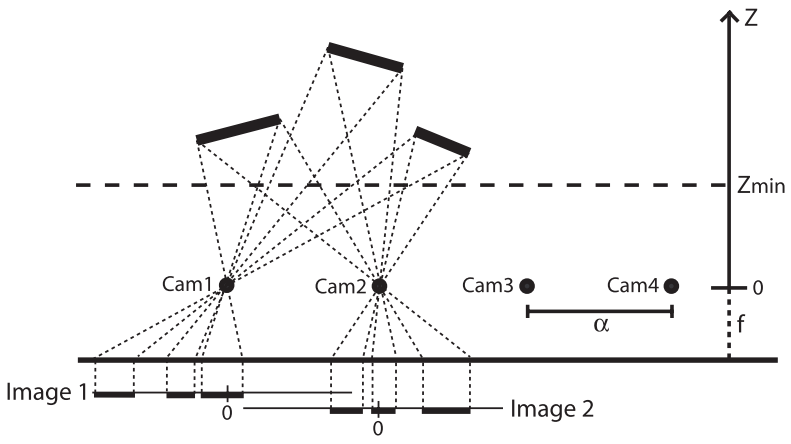

Figure 1. Our camera sensor network configuration.

Since the cameras are placed on an horizontal line, only the horizontal parallax has an effect on the correlation between the N different views. We can therefore reduce the sampling and compression problems to the 1-D case without loss of generality. Throughout the paper, we will thus focus our attention on one particular horizontal scanline for the different views. Figure 2 shows an example of two correlated 1-D views with three pieces of constant intensities.

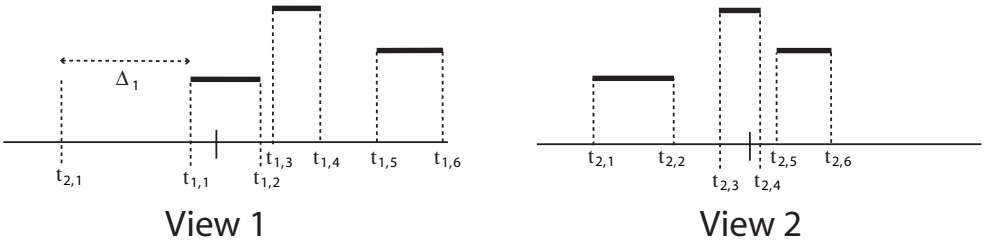

Figure 2. Two correlated views of the same scene observed from two different viewing positions. Each discontinuity is shifted according to the epipolar geometry. The set of disparities is given by $\{\Delta_i\}_{i=1}^{2L}$, where $\Delta_i = t_{1,i} - t_{2,i}$.

The N cameras then communicate to a central station through a multi-access channel with fixed capacity C. The natural questions we want to address are the following: a) Is there a sampling result that guarantees that perfect reconstruction of the visual scene is possible from a finite number of blurred and sampled projections? b) Since the observed projections have to be transmitted through a channel with fixed capacity, is the number of cameras going to influence the reconstruction fidelity at the decoder?

We show in the next sections that an exact sampling theorem for this scenario exists

and, most important, we show that there exists a practical distributed coding strategy that allows for a reconstruction of the scene with a distortion that is independent of the number of sensors and depends only on the total number of bits that can be transmitted through the channel.

The rest of the paper is organised as follows: Section 3 describes our distributed sampling strategies for the scene model with finite rate of innovation that we use. In Section 4, we describe our distributed compression approach and we highlight our "bit-conservation principle" in the context of a rate-distortion analysis. Simulation results are proposed in Section 5 and we conclude in Section 6.

3 Distributed Sampling

The signals observed at the sensors are piecewise polynomial signals and can be classified as signals with *Finite Rate of Innovation (FRI)*. Recently, new sampling methods for these classes of non-bandlimited signals have been proposed [13, 14]. They allow for a perfect reconstruction using only a finite number of samples. The sampling can be done using sinc or Gaussian kernels, or any function that can reproduce polynomials. In other terms, each sensor observes a blurred and sampled version of the original piecewise polynomial projection, and is able to reconstruct exactly the original parameters of the view (exact discontinuity locations and polynomial coefficients). Extension of this sampling approach for 2-D signals with FRI has been proposed recently [15, 16].

Since each sensor is able to retrieve precisely its original perspective projection, we can show that a finite number of sensors is sufficient to reconstruct exactly the original scene using back-projection techniques. The goal of these techniques is to find all the disparity correspondences between the different views. Once this disparity matching problem is solved, the exact depth of any object can be retrieved and the original scene can be reconstructed exactly. We consider here two scene scenarios leading to two different sampling techniques:

A: We first consider the case where the L planar objects are separated and visible from all the N cameras without occlusion, and keep the same "left to right" ordering (the k^{th} piece on the i^{th} view corresponds to the k^{th} piece on the j^{th} view, for any $k \in \{1; L\}$ and $i, j \in \{1; N\}$). With this hypothesis, only two of the N views are necessary in order to reconstruct the original scene (the correspondence problem between the two views is straightforward in this case).

B: We then consider the case where the L objects are separated and visible from all the N cameras without occlusion, but where the "left to right" ordering is not guaranteed anymore. In order to solve the disparity matching problem at the decoder, we need to have at least $L+1$ views of the scene. We can then back-project the $L+1$ left extremities of the pieces and retrieve the L real locations of these extremities [10]. The same procedure is then repeated with the $L+1$ right extremities. Notice that

this approach only relies on the discontinuity locations and not on the intensity of the pieces. This general sampling result is therefore sufficient in any case (even if all the pieces have the same intensity), but is not always necessary (two views are in theory sufficient if all the pieces have different intensities).

For these two scenarios, the minimum number of cameras corresponds to the critical sampling[1]. In the next section, we will show how each sensor can quantize its parameters and use distributed compression to maintain a constant global rate-distortion behaviour at the receiver, independent on the number of sensors involved.

4 Distributed Compression and Bit Conservation

The distributed compression algorithm applied at each sensor can be summarized as follows: First, the original projection is reconstructed from the observed sampled version using an FRI reconstruction method. The original view parameters retrieved are then quantized according to some target distortion value for the reconstructed view. Finally, each quantized parameter is S-W encoded according to the known correlation structure between the different views.

In this section, we first describe the rate-distortion behaviour of our view model when the view parameters are encoded independently. Then, we introduce precisely the correlation between the different views. Finally, we present our distributed compression approach and show that it guarantees a bit-conservation principle.

4.1 R-D for each projection

A 1-D view is modelled by a piecewise polynomial function defined on $[0;T]$ with L independent pieces of maximum degree Q, bounded in amplitude in $[0,A]$, and $2L$ discontinuities. Assume that such a function is quantized using R_t and R_p bits to represent each discontinuity and polynomial piece respectively (the parameters are quantized using a uniform scalar quantizer). It is possible to show that the distortion (MSE) of its reconstruction can be bounded with the following expression [17]:

$$D(R_p,R_t) \leq A^2 LT((Q+1)^2 \ 2^{-\frac{2}{Q+1}R_p} + 2^{-R_t}) \tag{1}$$

For a total number of bits $R = L(2R_t + R_p)$, the optimal bit allocation is given by: $R_p = \frac{Q+1}{Q+5}\frac{R}{L} + G$ and $R_t = \frac{2}{Q+5}\frac{R}{L} - G$, where $G = 2\frac{Q+1}{Q+5}(\log(Q+1)+2)$. This allocation leads to the following optimal rate-distortion behaviour:

$$D(R) \leq \underbrace{A^2 LT((Q+1)^2 \ 2^{-\frac{2}{Q+1}G} + 2^{-G})}_{c_0} 2^{\frac{-2}{L(Q+5)}R} \tag{2}$$

[1]In the presence of occlusions, it is also possible to derive an exact sampling result, assuming that each object of the scene can be occluded in at most a certain given number of views. However, for the sake of simplicity, we do not address the case of occlusions in this paper.

4.2 The correlation model

Assume $f_1(t)$ and $f_2(t)$ are two piecewise polynomial views obtained from two different cameras that are at a distance α apart. The two signals are defined for $t \in [0;T]$ and are bounded in amplitude in $[0;A]$. If there is no occlusion, the two views are exactly represented by L polynomials of maximum degree Q, and $2L$ discontinuities (the signals are equal to zero between the pieces). The shift of a discontinuity from one view to the other (its disparity) is given by the epipolar geometry and can be defined as: $\Delta_i = \frac{\alpha f}{z_i}$, where z_i is the depth of the i^{th} discontinuity (the depth of the object at its extremity). The range of possible disparities for a scene is therefore given by: $\Delta \in [0; \frac{\alpha f}{Z_{min}}]$ (see Figures 1 and 2).

We assume Lambertian surfaces for all the planar objects that make up the scene (i.e., the intensity of any point on the surface remains the same when observed from different viewing positions). A polynomial piece corresponding to a tilted planar object can be linearly contracted or dilated in the different views. However, its representation using Legendre polynomials is the same for any view, because of the support normalization on $[-1;1]$ that we use with this basis.

The correlation between the two views is therefore such that, knowing all the parameters of the first view, the only missing information needed to reconstruct perfectly the parameters of the second view is given by the set of disparities $\{\Delta_i\}_{i=1}^{2L}$. In the next section, we will show how this correlation structure can be used to perform distributed compression of the different views.

4.3 Distributed compression

Each sensor has to quantize and transmit its parameters to the central receiver. In the case where there is no occlusion, this information corresponds exactly to $2L$ discontinuity locations and L polynomials. Let R_{t_i} and R_{p_i} be the number of bits used to quantize respectively each discontinuity and each polynomial of the i^{th} view. The total number of bits used to encode this view is therefore given by $R_i = L(2R_{t_i} + R_{p_i})$ and is associated to a distortion D_i.

Assume just for one instant that the N views that have to be transmitted to the receiver are independent. In this case, for an average distortion D_i over all the reconstructed views at the receiver, the total amount of data to be transmitted would be of the order of $NR_i = NL(2R_{t_i} + R_{p_i})$.

We can now show how the correlation model can be exploited to perform distributed compression for the two scene scenarios (A and B) introduced in Section 3:

A: In this scenario, we know that only two views are necessary to retrieve the original scene parameters. Our correlation model is such that each discontinuity is shifted from one view to the other by a certain value $\Delta_i \in [0; \frac{\alpha f}{Z_{min}}]$. Assume that a total of R_1 bits is used to encode the first signal such that each polynomial piece and each discontinuity is using R_{p_1} and R_{t_1} bits respectively ($R_1 = L(R_{p_1} + 2R_{t_1})$). Now, in order to reconstruct the second view at the decoder with a similar distortion,

only $R_{t_{SW}} = R_{t_1} - R_s = \lceil \log_2(\frac{\alpha f}{Z_{min}}) \rceil$ bits are necessary to encode each discontinuity of the second view [11] (R_s corresponds thus to the number of most significant bits of each discontinuity location that does not need to be transmitted from the second view). Since the polynomials are similar for the two views, they do not need to be re-transmitted from the second sensor. The total bit-rate necessary to transmit the two views is therefore given by $R_{tot} = L(R_{p_1} + 2R_{t_1} + 2R_{t_{SW}})$. Now, if we want to transmit some information from more than two encoders, we know that, as long as the two extreme sensors transmit at least $R_{t_{SW}}$ bits from their discontinuities information, the rest of the $R_{tot} - 2R_{t_{SW}}$ bits can be obtained from any subset of the N encoders [12]. In other words, the information available at each sensor can be divided in two subsets. The last $R_{t_{SW}}$ bits of each discontinuity location ($2LR_{t_{SW}}$ bits in total) corresponds to the local innovation of the view according to the correlation structure between the two most distant sensors. This information has to be transmitted from the two most distant sensors. The remaining $L(R_{p_i} + 2R_s)$ bits is the information that is common to all sensors and can be obtained from any subset of the N cameras. The number of sensors used for the transmission has therefore no influence on the reconstruction fidelity at the decoder (the reconstruction fidelity is defined as the mean squared error over all the N reconstructed views). Using an optimal bit allocation and transmitting the information using N sensors, the distortion of any reconstructed view given the total bit budget R_{tot} can be shown to behave as:

$$D(R_{tot}) \leq \underbrace{c_0 2^{\frac{-2(2R_s+3G)}{Q+9}}}_{c_1} 2^{\frac{-2}{L(Q+9)}R_{tot}} \tag{3}$$

Notice that an independent encoding of the N views would lead to the following behaviour:

$$D(R_{tot}) \leq c_0 2^{\frac{-2}{NL(Q+5)}R_{tot}} \tag{4}$$

We can observe that the difference in the decaying factor between the distributed and the independent coding approaches becomes larger as N, L, or Q increases. If there is only two views to transmit and $Q = 0$, the decaying factors compare as $\frac{-2}{9L}$ vs. $\frac{-2}{10L}$ for the distributed and independent approaches respectively.

B: The distributed compression strategy in this case consists in sending the discontinuities from $L + 1$ views and each polynomial piece from only one encoder. The total bit-rate necessary is therefore given by: $R_{tot} = L((L + 1)2R_t + R_p)$. If we now want to transmit information from more than this minimum number of sensors $N_{min} = L+1$, we can do it in a very flexible manner: For each new sensor introduced in the system, we can ask it to take the responsibility of transmitting some partial information about the polynomial pieces (therefore reducing the communication task of some other sensors) or to replace one of its two neighbours to transmit some subset of the most significant bits of its discontinuity locations. The distortion of any reconstructed view given the total bit budget R_{tot} can be shown to behave as:

$$D(R_{tot}) \leq \underbrace{c_0 2^{\frac{-2(2L+1)G}{Q+9}}}_{c_2} 2^{\frac{-2(Q+5)}{L(Q+9)(4L+Q+5)}R_{tot}} \tag{5}$$

Again, the distortion at the receiver only depends on the total number of bits transmitted R_{tot} and not on the number of sensors used.

5 Simulation Results

We propose a simple simulation where five cameras are observing a synthetic scene made up of three tilted polygons with linear intensities. The cameras are placed on a horizontal line and observe blurred and undersampled views (32×32 pixels) of the original scene, as illustrated in Figure 3. For each view, knowing that the original scene belongs to a certain class of signals with finite rate of innovation, the sampling results presented in Section 3 can be used to retrieve the 33 original parameters of the view (twelve vertices having two parameters each, and three 2-D linear functions having three parameters each). Using these retrieved parameters, high resolution version of the different views can be reconstructed at each encoder (see Figure 4).

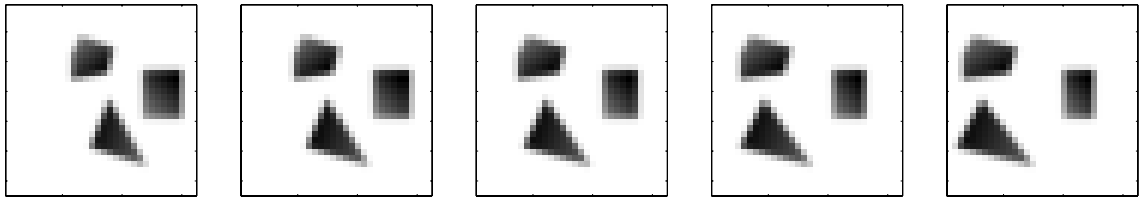

Figure 3. Observation at the sensors. Each sensor observes a blurred and undersampled version of the perspective projection of the scene.

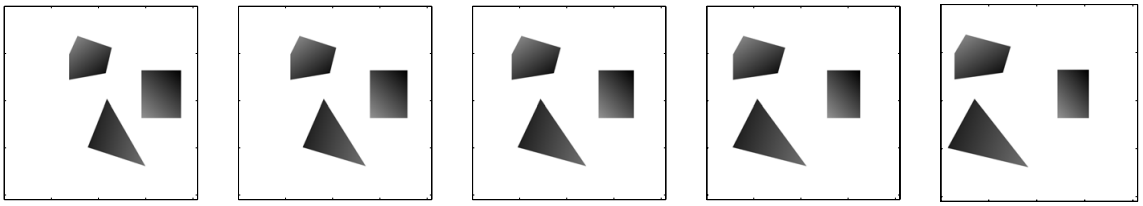

Figure 4. High resolution version of the 5 views reconstructed at the encoders from the samples shown in Figure 3 using an FRI method.

The original views are represented with the following precision: 22 bits for each vertex (view of 2048×2048 pixels) and 9 bits for each polynomial coefficient. One view is therefore exactly represented using $12 \times 22 + 9 \times 9 = 345$ bits. The parameters α, f and Z_{min} are such that $R_{t_{SW}} = 10$ bits (the disparities between the first and the fifth views can be larger than a quarter of the image width). As we have shown in Section 4.3, a total rate of $R_{tot} = 345 + 12 \times 10 = 465$ bits is therefore sufficient to reconstruct all the high resolution views at the receiver.

If the bit budget R_{tot} is smaller than 465, coarser quantization of the parameters has to be done prior to applying distributed compression. Table 1 highlights our

bit-conservation principle for specific bit budgets R_{tot}. In particular, it shows that our approach suffers no rate loss when we increase the number of sensors used for transmission.

Table 1. An exact bit-conservation principle.

$\mathbf{R_1}$ (bits)	$\mathbf{R_2}$ (bits)	$\mathbf{R_3}$ (bits)	$\mathbf{R_4}$ (bits)	$\mathbf{R_5}$ (bits)	$\mathbf{R_{tot}}$ (bits)	**PSNR** (dB)
345	-	-	-	120	465	∞
276	-	-	-	108	384	38.8
128	-	128	-	128	384	38.8
108	84	-	84	108	384	38.8
84	72	72	72	84	384	38.8
171	-	-	-	60	231	22.8
48	48	39	48	48	231	22.8

6 Conclusion

We have addressed the problem of distributed sampling and compression in camera sensor networks for scenes with finite rate of innovation. In particular, we have shown that using a piecewise polynomial model for the scene, the critical sampling (i.e., the minimum number of sensors needed) can be precisely derived. Then, we have also shown that when using more sensors, our distributed compression approach can be used to maintain a constant global rate usage (independent on the number of sensors N), for any given reconstruction distortion. This result provides us with an exact bit-conservation principle when the number of sensors increases. Our ongoing research focuses on the extension of those results to more complex scene and correlation models, and on its application to real images obtained from dense multi-camera arrays.

References

[1] D. Slepian and J.K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471–480, Jul 1973.

[2] D. Marco, E. Duarte-Melo, M. Liu, and D. Neuhoff. On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data. In *Information Processing in Sensor Networks (IPSN '03)*, April 2003.

[3] D. Marco and D. Neuhoff. Reliability vs. efficiency in distributed source coding for field-gathering sensor networks. In *Information Processing in Sensor Networks (IPSN '04)*, pages 161–168, April 2004.

[4] A. Kashyap, L.A. Lastras-Montano, C. Xia, and L. Zhen. Distributed source coding in dense sensor networks. In *Data Compression Conference (DCC '05)*, pages 13–22, March 2005.

[5] D.L. Neuhoff and S.S. Pradhan. An upper bound to the rate of ideal distributed lossy source coding of densely sampled data. Submitted to *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*.

[6] P. Ishwar, A. Kumar, and K. Ramchandran. Distributed sampling for dense sensor networks: a "bit-conservation principle". In *Information Processing in Sensor Networks (IPSN '03)*, April 2003.

[7] Y.W. Hong, A. Scaglione, R. Manohar, and B. Sirkeci Mergen. Dense sensor networks are also energy efficient: When 'more' is 'less'. In *Milcom 2005*, October 2005.

[8] E.H. Adelson and J.R. Bergen. The plenoptic function and the elements of early vision. In M. Landy and J. Anthony Movshon, editors, *Computational Models of Visual Processing*, pages 3–20. MIT Press, Cambridge, MA, 1991.

[9] J.-X. Chai, S.-C. Chan, H.-Y. Shum, and X. Tong. Plenoptic sampling. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 307–318. ACM Press/Addison-Wesley Publishing Co., 2000.

[10] A. Chebira, P. L. Dragotti, L. Sbaiz, and M. Vetterli. Sampling and Interpolation of the Plenoptic Function. In *IEEE International Conference on Image Processing (ICIP '03)*, September 2003.

[11] N. Gehrig and P.L. Dragotti. Distributed compression of the plenoptic function. In *IEEE Int. Conf. on Image Processing (ICIP '04)*, October 2004.

[12] N. Gehrig and P.L. Dragotti. DIFFERENT - DIstributed and Fully Flexible image EncodeRs for camEra sensor NeTworks. In *IEEE Int. Conf. on Image Processing (ICIP '05)*, September 2005.

[13] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Transactions on Signal Processing*, 50(6):1417–1428, June 2002.

[14] P.L. Dragotti, M. Vetterli, and T. Blu. Exact sampling results for signals with finite rate of innovation using strang-fix conditions and local kernels. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05)*, March 2005.

[15] I. Maravic and M. Vetterli. Exact sampling results for some classes of parametric non-bandlimited 2-D signals. *IEEE Transactions on Signal Processing*, 52(1):175–189, January 2004.

[16] P. Shukla and P.L. Dragotti. Sampling schemes for 2-D signals with finite rate of innovation using kernels that reproduce polynomials. In *IEEE Int. Conf. on Image Processing (ICIP '05)*, September 2005.

[17] P. Prandoni and M. Vetterli. Approximation and compression of piecewise smooth functions. *Phil. Trans. R.Soc.Lond. 1999.*, 357(1760):2573–2591, September 1999.