Distributed Visual Information Processing in Camera Sensor Networks*

Pier Luigi Dragotti

Communications and Signal Processing Group
Imperial College London

June 1, 2006

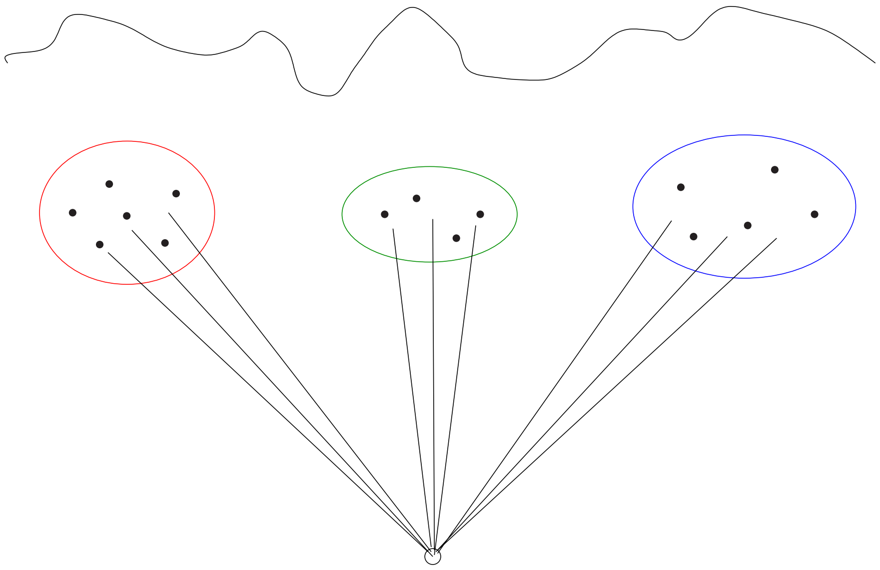
---

*This is a joint work with N. Gehrig (ICL), L. Baboulaz (ICL), J. Berent (ICL), M. Gastpar (UC Berkeley) and M. Vetterli (EPFL & UC Berkeley). Sponsors include DIF-DTC, SEAS-DTC, EOARD.

Outline

- Motivation: Sensor Network
- Structure of the Data and its Sampling: the Plenoptic Function
- Distributed Compression in Camera Sensor Networks
  - Background: Theoretical Foundation and Constructive Codes
  - Lossless Compression: The asymmetric and symmetric cases
  - Lossy case: Quantization followed by Slepian-Wolf coding
- Data Fusion
  - Image Registration and Super-Resolution
  - Scene Interpretation and the Level-set Method
- Fundamental trade-offs in Sensor Networks and joint Source/Channel coding
- Conclusions

Motivation: Sensor Networks



- The source (phenomenon) is distributed in space.
- In our case, sensors are digital cameras and are battery powered.
- The number of sensors can be very large (well beyond stereo imaging or stereophonic sound).
- Communication is critical.
- A central receiver or a leader node fuses the data transmitted by the sensors.

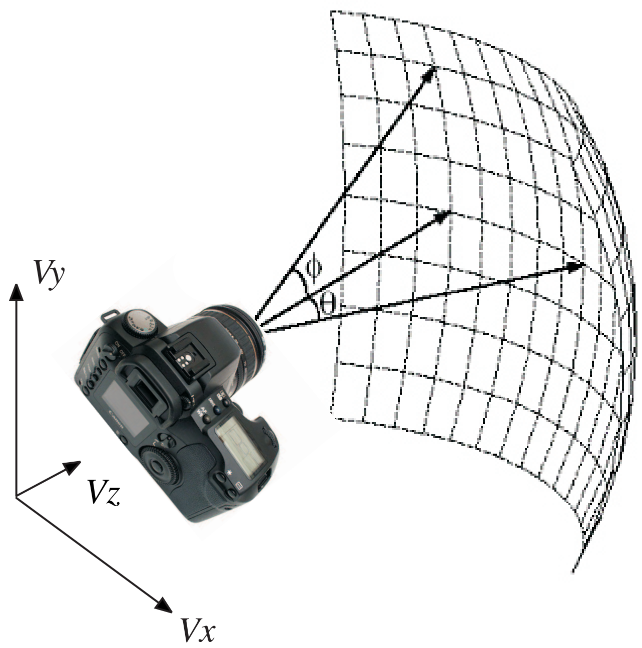
Motivation: Sensor Networks

Open questions:

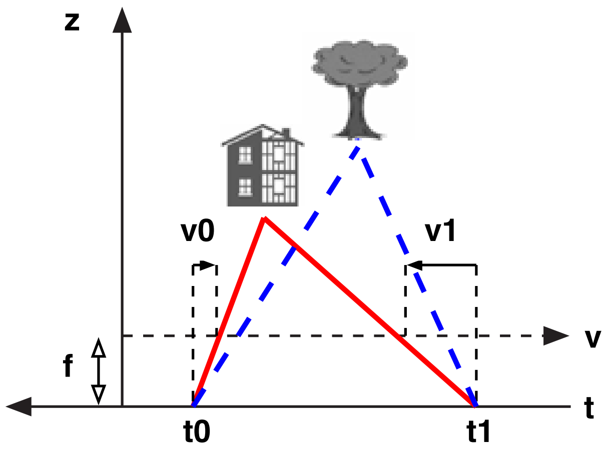
- The observed phenomenon has a particular spatio-temporal structure. Can we understand it? Can we sample it?
- Sensors observe correlated data. We want to perform compression but we want to avoid communication among sensors. How are we going to compress this data?
- No separation principle. Joint-source channel coding?
- Reconstruction:
  - Many sensors, but with very low-resolution. Accurate registration of the data is vital.
  - The receiver has to handle a huge amount of data. Efficient methods for unsupervised data analysis are crucial.
  - New trade-offs between acquisition precision, number and location of sensors, compression rate, delay, complexity.

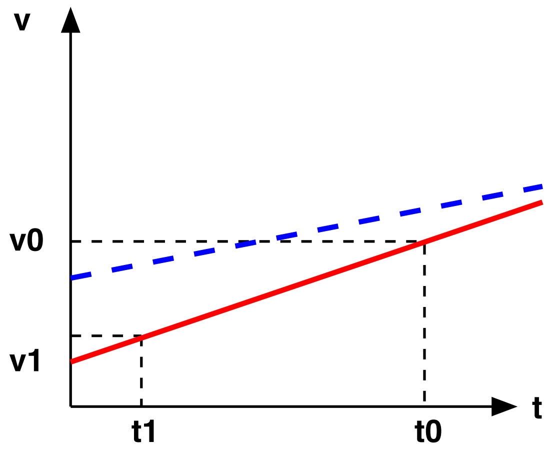
Structure of the data: The Plenoptic Function

The plenoptic function introduced by Adelson and Bergen [AdelsonB:91] describes the intensity of each light ray that reaches any point in space at any time. It is therefore characterized by 7 parameters, namely the viewing position, the viewing directions, time and wavelength. $P(\theta, \phi, V_x, V_y, V_z, \lambda, t)$:



Sampling the Plenoptic Function





If the depth of field is bounded, that is, the distance of the objects to the cameras is bounded between z_{min}, z_{max}, the plenoptic function is approximately bandlimited and can be sampled [Chai-Tong-Chan-Shum:00].

More recent results: for certain simple non-bandlimited scenes, the plenoptic function can be sampled exactly [Chebira-Dragotti-Vetterli-Sbaiz:03].

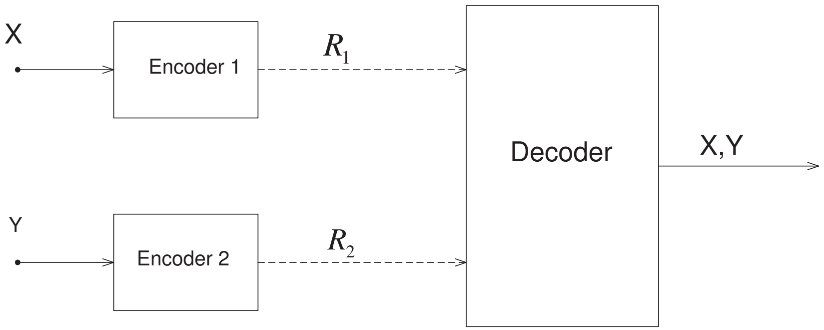
Distributed Source Coding

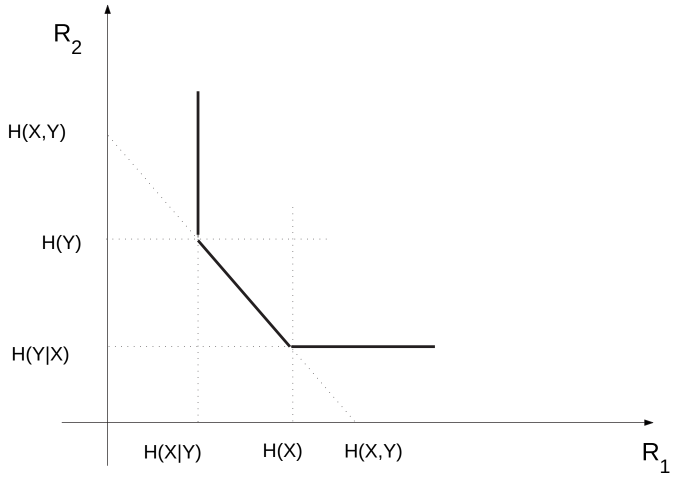
Data acquired by sensors are correlated. Communication among sensors needs to be limited.

- Fundamental theoretical results
  - Slepian-Wolf 1973. Lossless coding.
  - Wyner and Ziv 1976. Lossy coding with side information at the decoder.

- Constructive codes:
  - DISCUS [Pradhan-Ramchandran:99].
  - Extensions using advanced channels coded [Garcia-Frias:01, Aaron-Girod:02, Liveris-Xiong-Georghiades:02, Stankovic et al.:04, Stankovic et al.:06]

Background: Lossless Coding of Correlated Sources





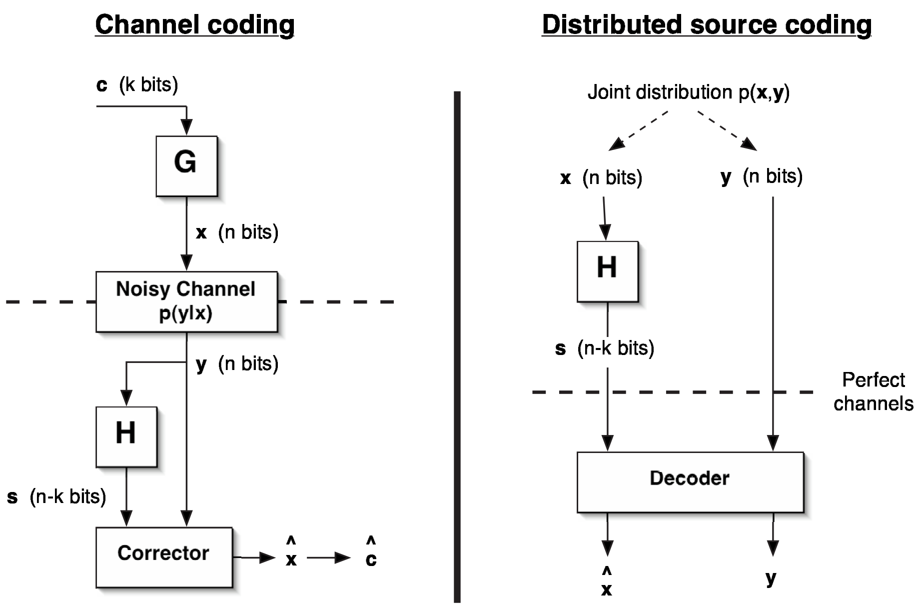
- X and Y are two discrete correlated sources to be encoded at rates R_1 and R_2, respectively.
- Lossless separate coding can be achieved if and only if R_1 and R_2 satisfy

$$\begin{aligned} R_1 &\geq H(X|Y) \\ R_2 &\geq H(Y|X) \\ R_1 + R_2 &\geq H(X,Y) \end{aligned}$$

$$H(X,Y) = H(X|Y) + H(Y|X) + I(X,Y)$$

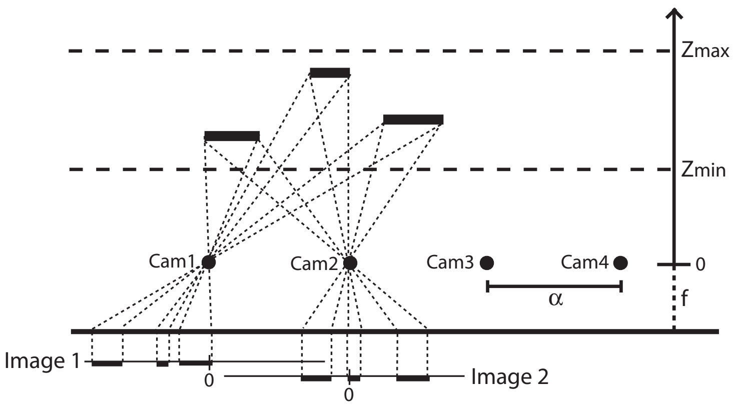
Background: Constructive Codes

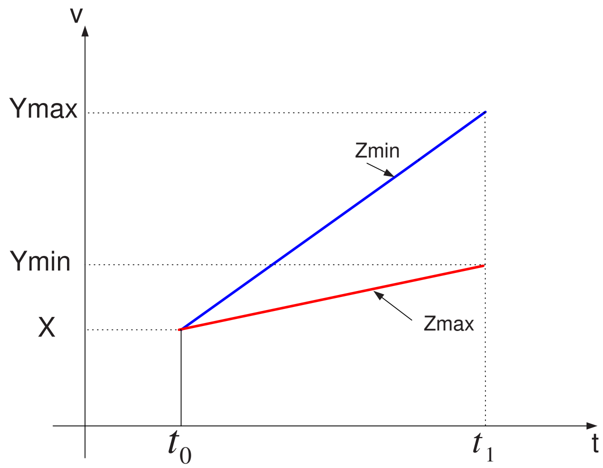
Basic principle: X and Y are the input and output of a noisy channel [Pradhan-Ramchandran:99, Garcia-Frias:01, Aaron-Girod:02, Liveris-Xiong-Georghiades:02].



Notice: Original approaches consider only the asymmetric scenario $(R_1, R_2) = (H(X), H(Y|X))$ or $(R_1, R_2) = (H(X|Y), H(Y))$. More recent approaches allow to cover the entire Slepian-Wolf rate-region [Stankovic et al:04-06], [Gehrig-Dragotti:05].

Plenoptic Constraints

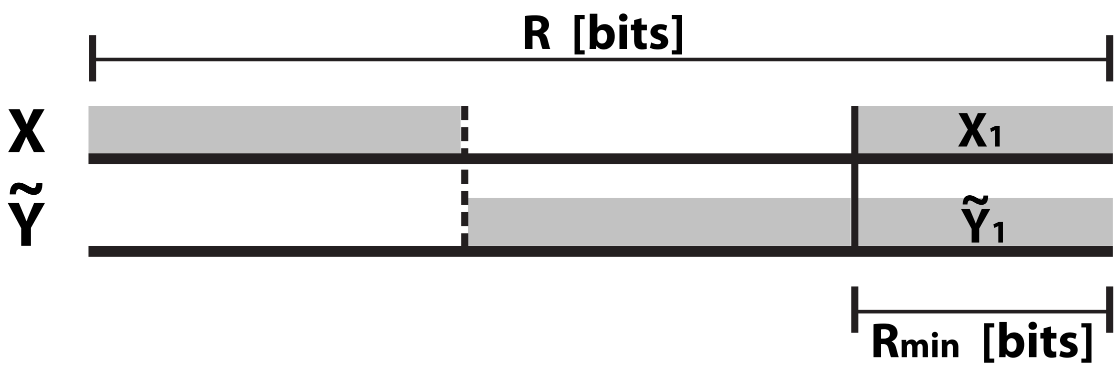




- Assume X and Y are the positions of a point on two different images
- Key insight: Whatever the complexity of the scene, if the depth of field is bounded, the disparity is also bounded.
- Use the information z_{min}, z_{max} and the camera locations, to develop distributed compression algorithms
- Notice that a similar intuition was used by Shum to sample the plenoptic function [Chai-Tong-Chan-Shum:00].

Distributed Encoding of the Disparity

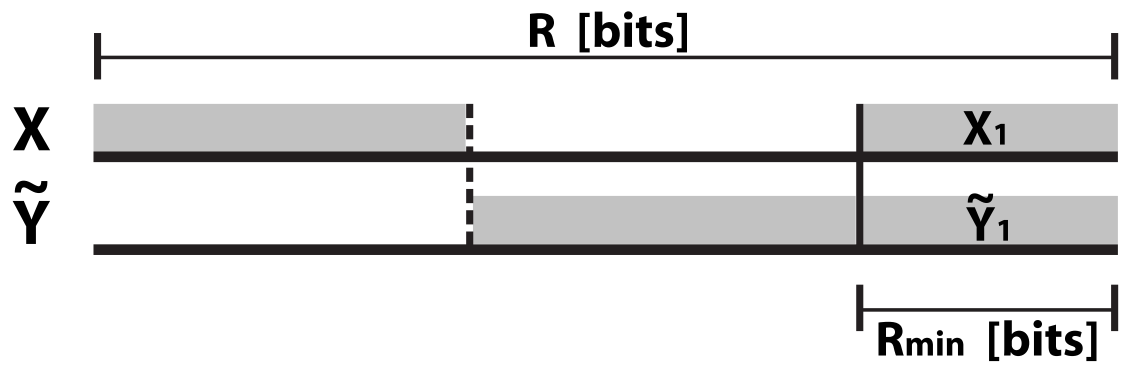
- Given, z_{min}, z_{max} and $\alpha = t_1 - t_0$, it follows that $Y \in [X + \frac{\alpha f}{z_{max}}, X + \frac{\alpha f}{z_{min}}]$
- Asymmetric coding approach (we assume that X and Y are discrete sources)
  - Send X losslessly
  - Modulo encode Y: $Y \mod [\alpha f(\frac{1}{z_{min}} - \frac{1}{z_{max}})]$



$$R_{min} = \left\lceil \log_2 \left(\alpha f (\frac{1}{z_{min}} - \frac{1}{z_{max}}) \right) \right\rceil$$

Distributed Encoding of the Disparity

- The difference between X and Y is bounded, if we look at their representation in bits the differ only in the less significant bits.
- Symmetric coding approach
  - Send the less significant bits of X (i.e., $H(X/Y)$)
  - Send the less significant bits of Y (i.e., $H(Y/X)$)
  - Send complementary subsets of the most significant bits (i.e., share $I(X,Y)$)



Notice that this is true for any memoryless binary source [Gehrig-Dragotti:05].

Intermezzo: Slepian-Wolf Coding of Binary Sources

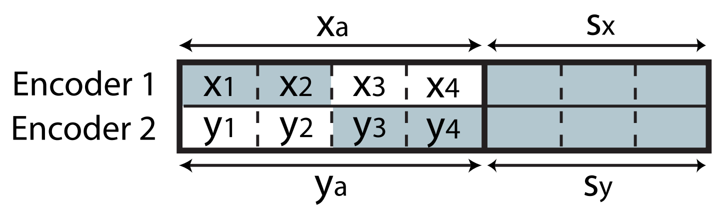
- X is an i.i.d. binary n-sequence with $P(x_i = 1) = 1/2$,
- $Y = X \oplus E$ and $P(e_i = 1) = p$.
- If n is large enough $P(X - Y \leq np) \to 1$. Therefore, the difference is bounded as in the previous case!
- The difficulty is that we do not know the locations of the bits that differ.
- Use syndrome coding to determine these locations

Intermezzo: Slepian-Wolf Coding of Binary Sources

- Design a linear code that can correct up to np errors
- The syndrome s_x of X indicates the distance of X to the all-zeros vector
- The syndrome s_y indicates the distance of Y to the all-zeros vector
- Thus $s_x \oplus s_y$ indicates the difference between X and Y

Coding strategy

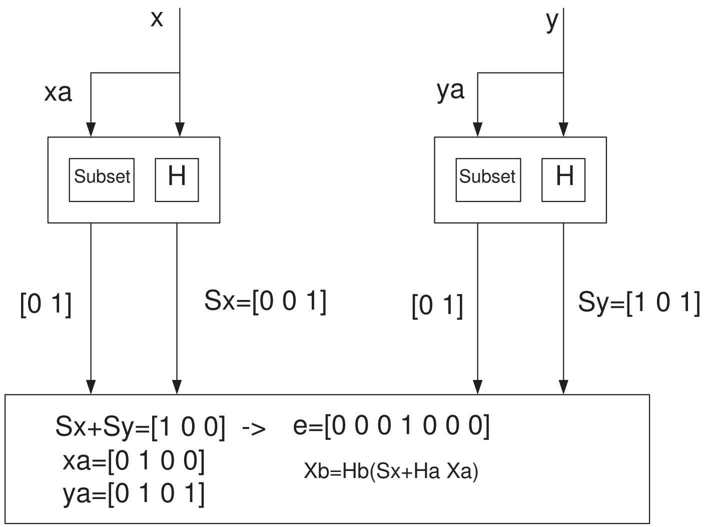
- Send s_x from Encoder 1
- Send s_y from Encoder 2
- Send only complementary sets of the most significant bits of X and Y.



Notice, a similar approach have been proposed by [StankovicLXG:04] as well.

A Simple Example

X and Y differ of at most one bit $\longrightarrow$ use Hamming $(7,4)$ code

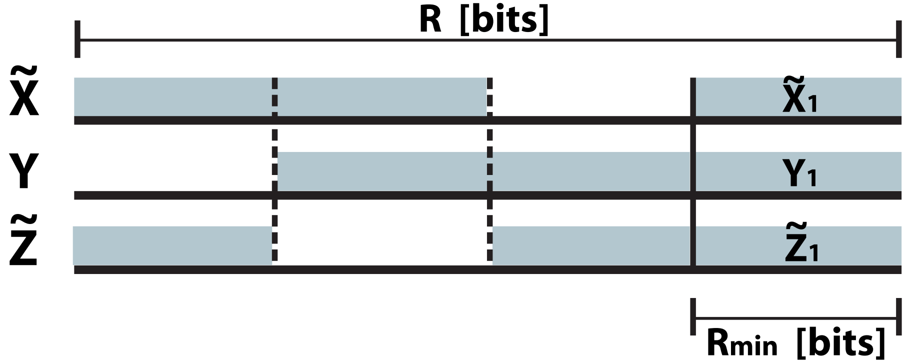


- $x = [\underbrace{0100}_{x_a}\underbrace{110}_{x_b}] \qquad y = [\underbrace{0101}_{y_a}\underbrace{110}_{y_b}]$
- Each encoder sends 5 bits $\longrightarrow$ total number of bits sent 10
- $H(X) = H(Y) =$ 7bits, $H(X/Y) = H(Y/X)$=3bits, $\longrightarrow H(X,Y) = 10 \longrightarrow$ optimal!

The Occlusion Problem

- Three cameras with one possible occlusion.

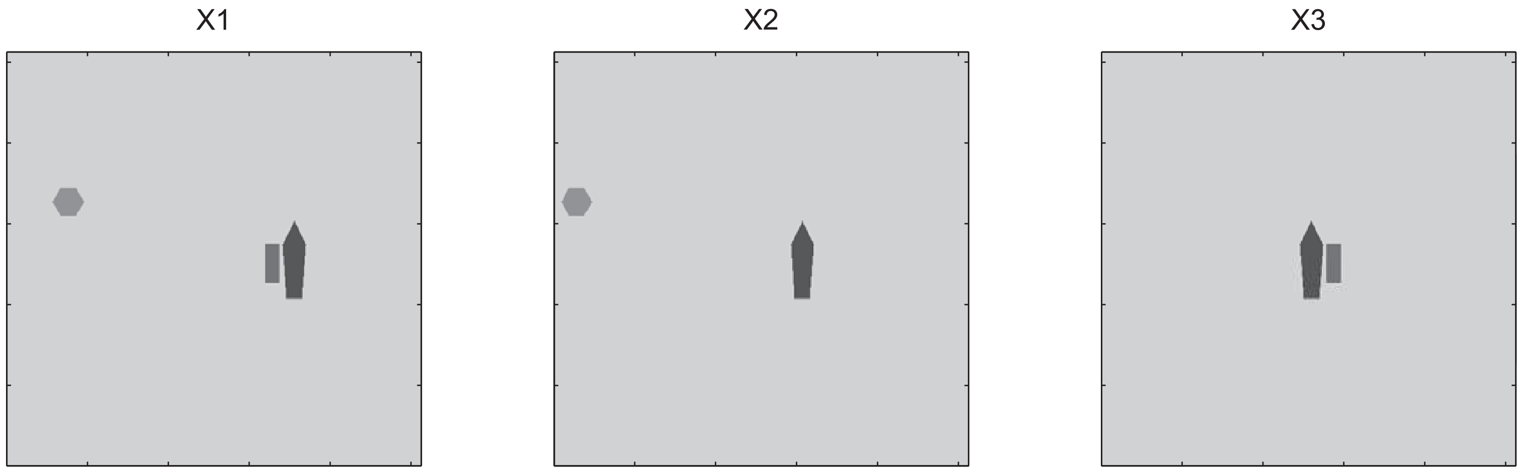
- The information provided by any pair of encoders must be sufficient to allow for a reconstruction at the receiver.



Easy to extend to the case of N cameras and M occlusions ($M \leq N - 2$).

Simulation Results

Three views of a simple synthetic scene



- Independent coding: 18bits per vertex
- Slepian-Wolf with no occlusions: 10bits per vertex
- Slepian-Wolf with occlusions: 14bits per vertex

Lossy Coding in Camera Sensor Networks

Aim:

- Use a compression algorithm that preserves geometry.
- Extend the findings of the lossless case to the lossy case.

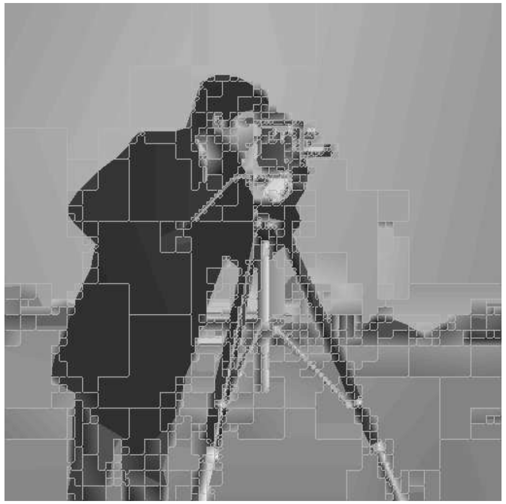
Our Approach:

- Model images as piecewise polynomial signals.
- Use quadtree-based compression algorithms.
- Quantize (compress) first and use distributed lossless compression then.

Natural questions:

- Can we keep simplicity and flexibility of the lossless case?
- Any optimality claims?

Quad-Tree Algorithm with Prune and Join Steps [ShuklaDDV:05]



Prune-Join Quad-Tree Decomposition

P-J Tree PSNR 28.9dB, 0.11bpp

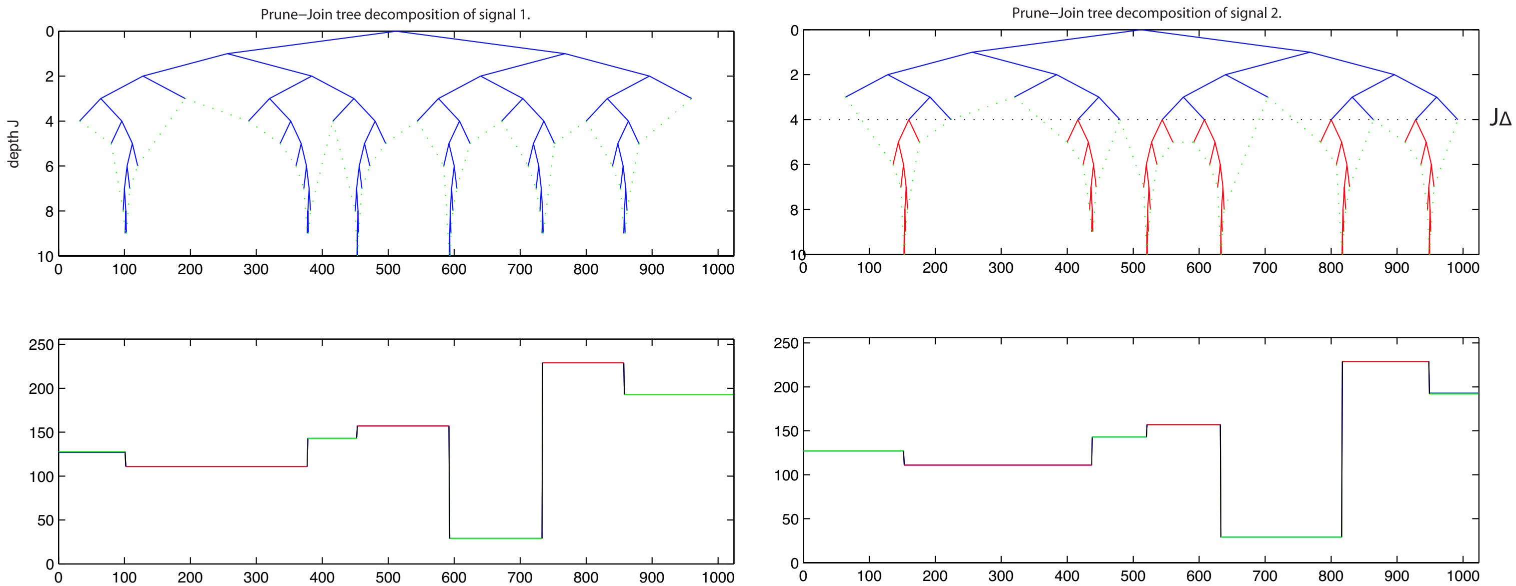
JPEG2000 PSNR 27.8dB 0.11bpp

- Tiles can be pruned or merged.
- Each tile is made of two polynomials divided by a straight-line.

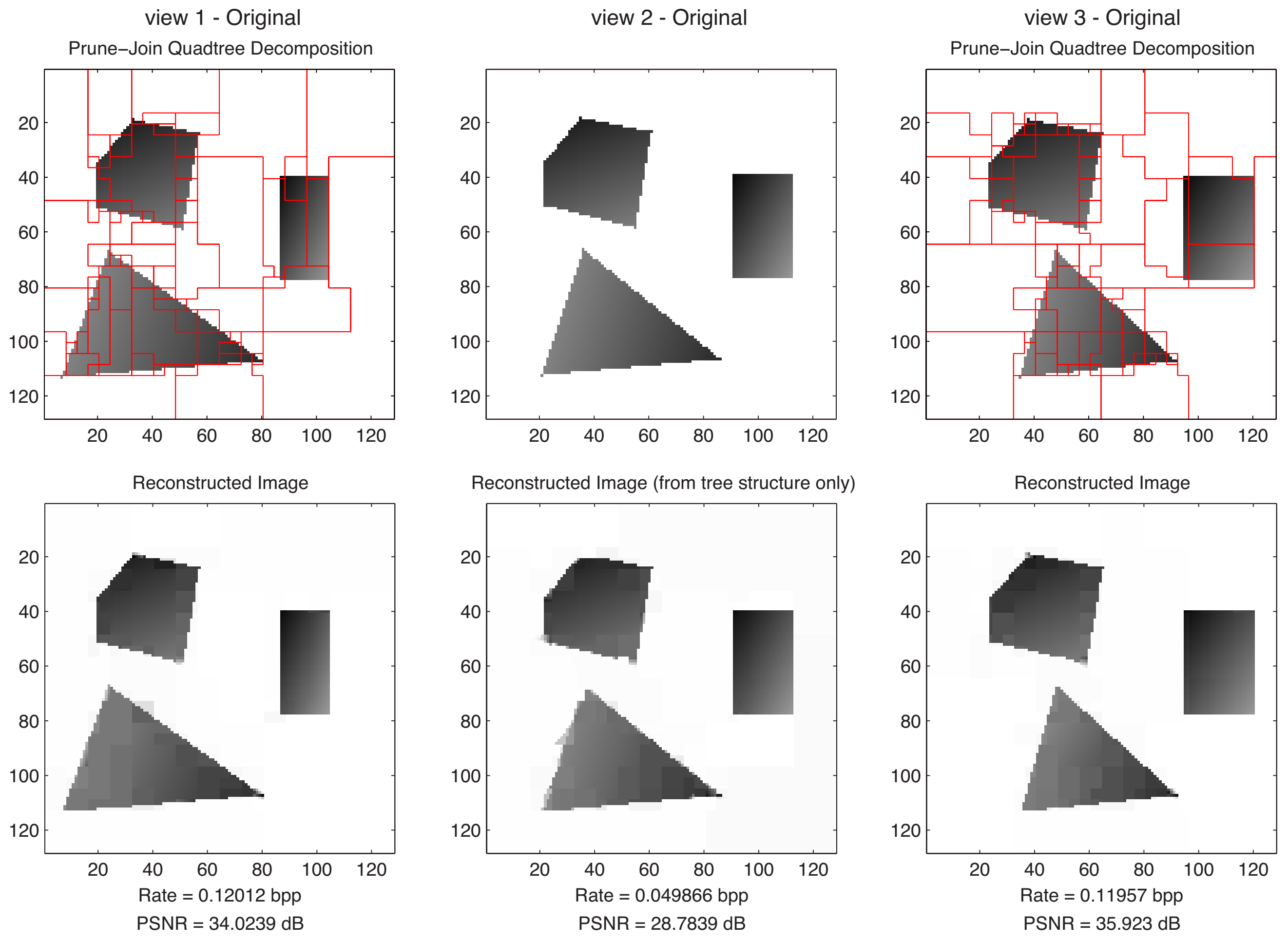
Distributed Lossy Compression

Intuition: In the case of multi-view images

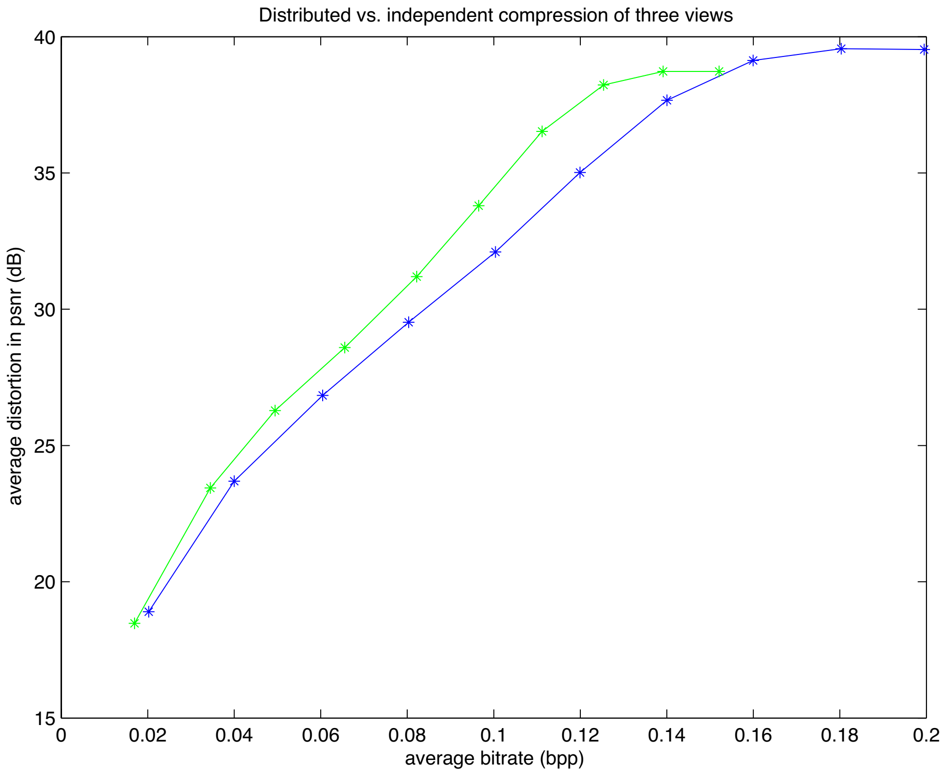
- The structure of the quad-tree follows the disparity constraints.
- If the scene is Lambertian, tiles at different sensors are going to be equal.



Distributed Lossy Compression



Distributed Lossy Compression



Notice: no channel codes used, only epipolar constraints. See also: [Tuncel et al:06].

Distributed Lossy Compression

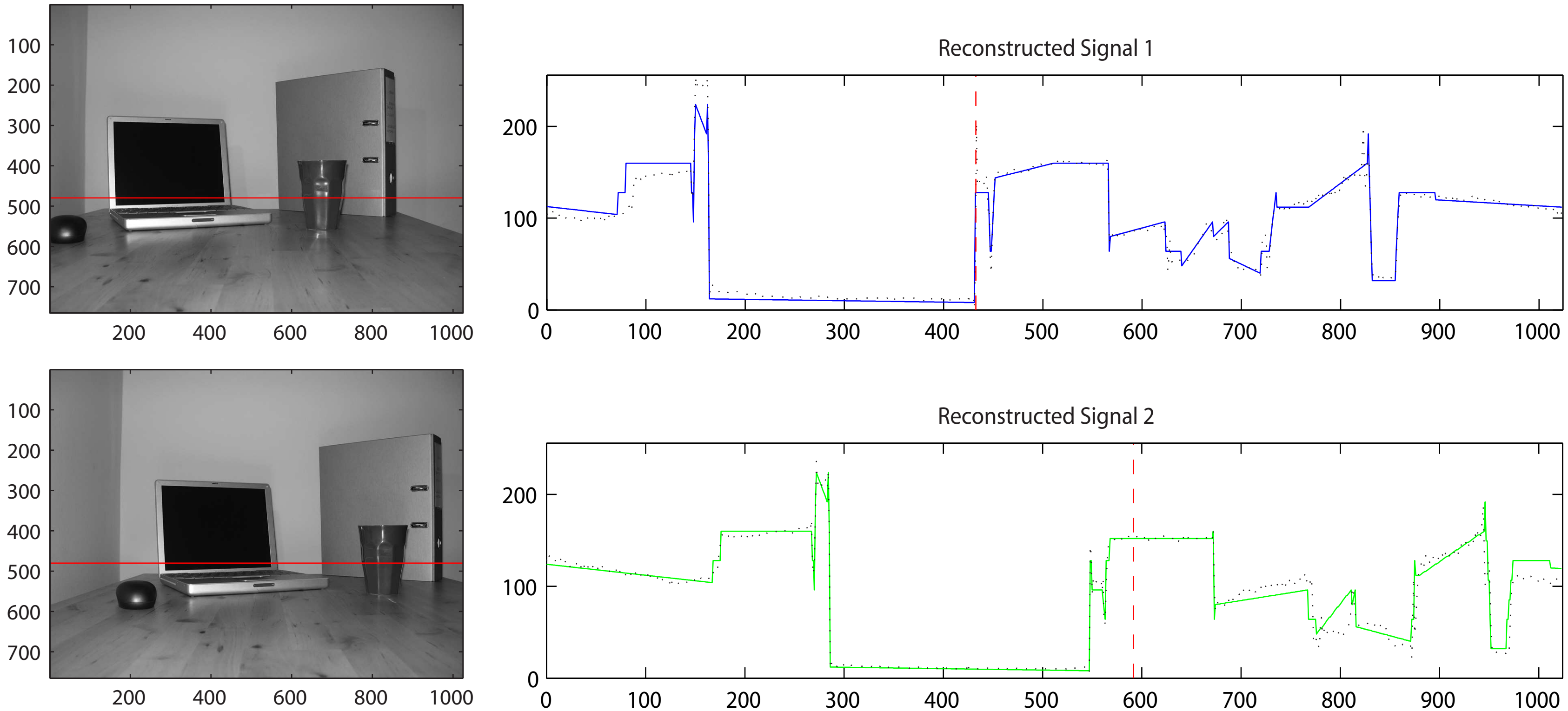
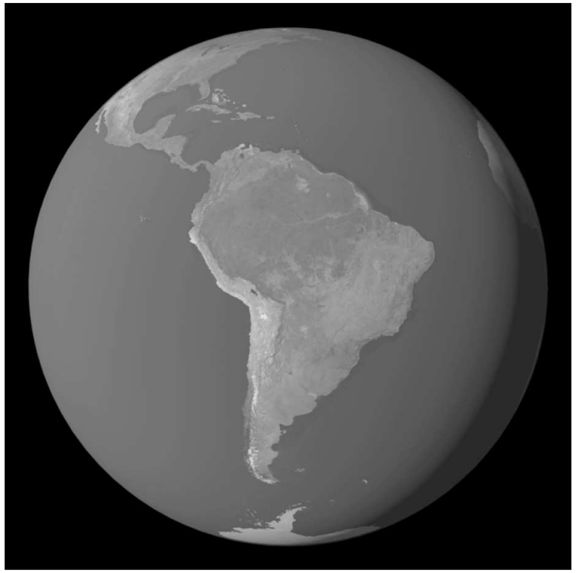
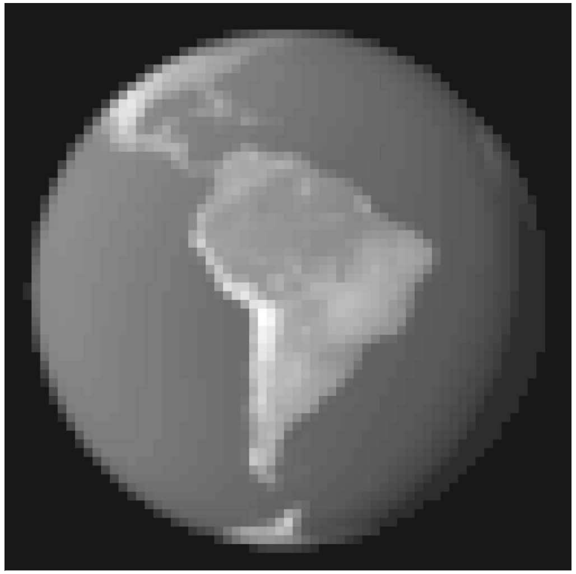


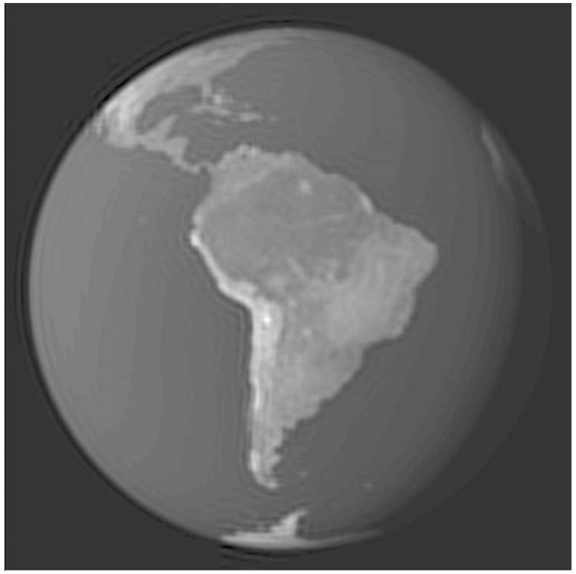
Image Super-Resolution [BaboulazD:06]



(a)Original (2000×2000)



(b) Low-res. (65×65)



(c) Super-res (2000×2000)

- One hundred low-resolution and shifted versions of the original.
- Given the number of cameras and the low resolution, registration is critical.
- Accurate registration is achieved by modelling precisely the acquisition and compression process.
- The registered images are interpolated to achieve super-resolution.

Image Super-Resolution

- A pixel $P_{m,n}$ in the compressed image is given by

$$P_{m,n} = \langle I(x,y), \varphi(x/2^J - n, y/2^J - m) \rangle.$$

- The scaling function $\varphi(x,y)$ can reproduce polynomials:

$$\sum_n \sum_m c_{m,n}^{(l,j)} \varphi(x-n, y-m) = x^l y^j \quad l = 0,1,...,N; j = 0,1,...,N.$$

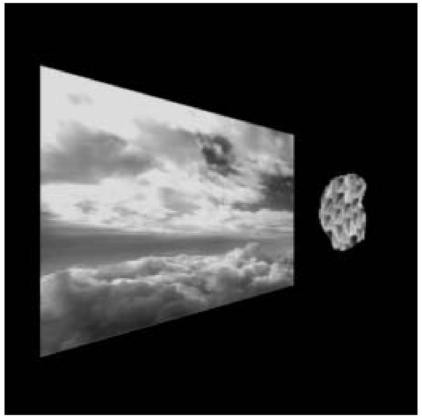
- Retrieve the exact moments of $I(x,y)$ from $P_{m,n}$:

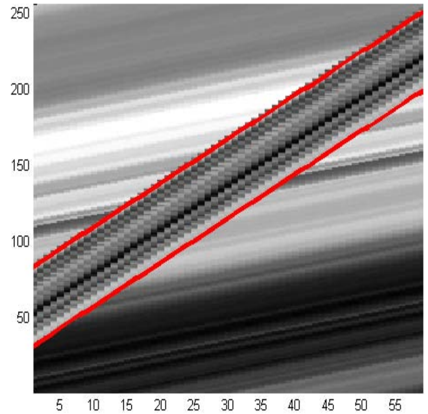
$$\tau_{l,j} = \sum_n \sum_m c_{m,n}^{(l,j)} P_{m,n} = \langle I(x,y), \sum_n \sum_m c_{m,n}^{(l,j)} \varphi(x/2^J - n, y/2^J - m) \rangle = \int \int I(x,y) x^l y^j dx dy.$$

- Register the compressed images using the retrieved moments.

Scene Interpretation

Huge amount of data difficult to analyze. Use level-set method to extract layers and model occlusions explicitly.

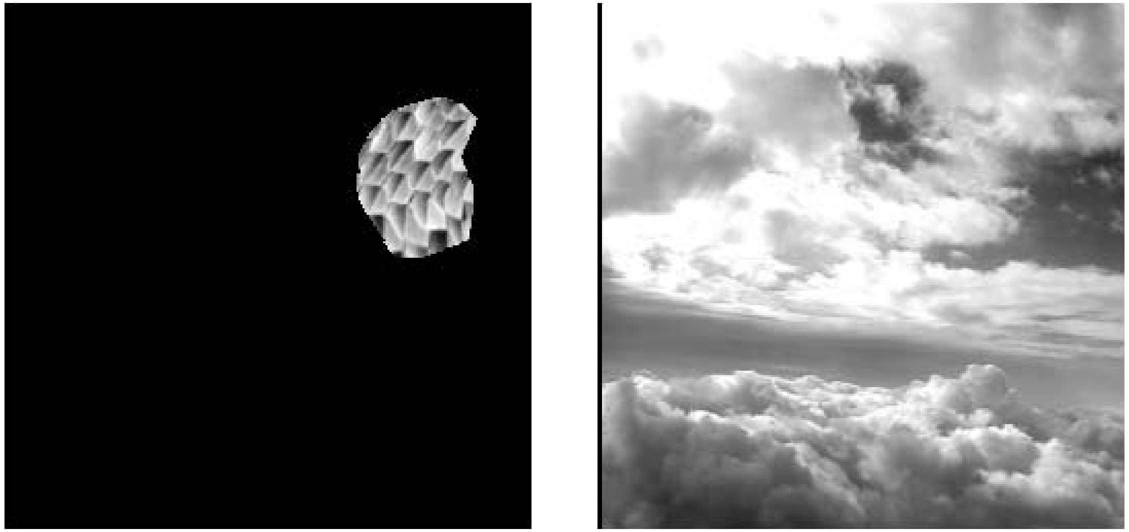




Why level-set method?

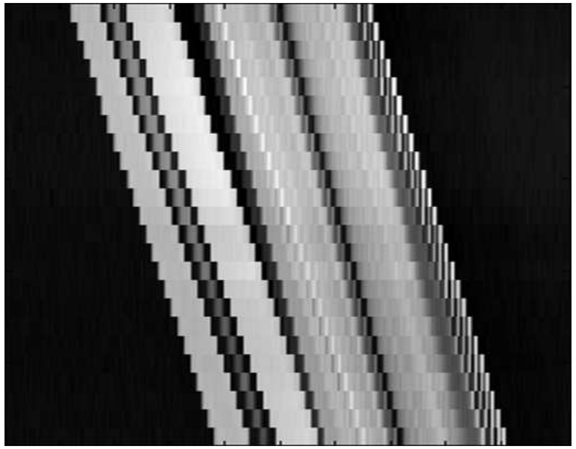
- Can be used for multi-dimensional signals.
- More flexible than active contours.
- Fast algorithms exist.

Scene Interpretation

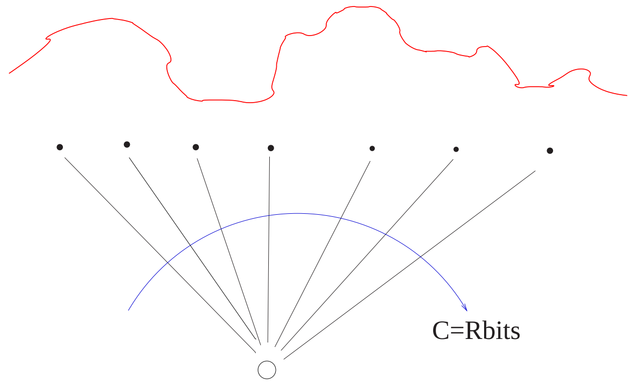


Layers extraction and interpolation.

Scene Interpretation [BerentD:06]



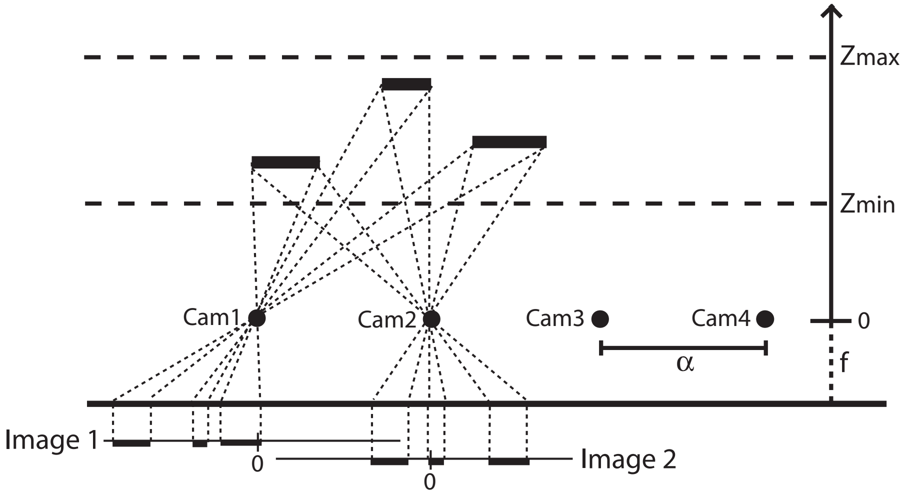
A Glimpse at Fundamental Trade-offs in Sensor Networks



- The phenomenon can be sampled (i.e., N sensors are sufficient to represent it).
- Sensors communicate through a (multi-access) channel with capacity $C = R$ bits
- Fundamental question: what happens when we have $M > N$ sensors but same channel capacity?

Fundamental Performance Bounds in Camera Sensor Networks

For idealized scenarios:



We can develop new sampling schemes that allow for an exact reconstruction of the phenomenon.

We have a distributed encoder that is optimal independently of the number of sensor used (exact 'bit conservation principle') [Gehrig-Dragotti-06].

Moreover [Gastpar-Vetterli-Dragotti-06], if measurements are noiseless

- with separation we achieve $MSE \sim 1/M$,
- an ideal joint source-channel encoder can achieve $MSE \sim 1/M^2$.

If measurements are noisy

- the gap between a separate and a joint encoder is exponential.

Conclusions

- Sensor networks bring new degrees of freedom: large number of acquisition devices, flexibility. But also new constraints: communication is critical, data acquired is difficult to handle.
- In sensor networks it is of fundamental importance to understand the physics of the phenomenon, and the interaction between sampling, compression and transmission.
- In camera sensor networks
  - Use geometric constraints to develop distributed coding algorithms.
  - Use advanced mathematical methods to perform sampling, registration and scene interpretation.

Publications

http://www.commsp.ee.ic.ac.uk/∼pld/publications/

- On Distributed Compression in Camera Sensor Networks
  - N. Gehrig and P.L. Dragotti, DIFFERENT - **DI**stributed and **F**ully **F**lexible image **E**ncode**R**s for cam**E**ra sensor **N**e**T**works, Proc. of the IEEE International Conference on Image Processing (ICIP), Genova, Italy, September 2005.
  - N. Gehrig and P.L. Dragotti, Distributed Compression of the Plenoptic Function, in Proc. of the IEEE International Conference on Image Processing (ICIP), Singapore, October 2004.
  - N. Gehrig and P.L. Dragotti, Symmetric and Asymmetric Slepian-Wolf codes with systematic and non-systematic linear codes, IEEE Communications Letters, vol. 9(1), pp.61-63, January 2005.
- On Fundamental trade-offs in Sensor Networks
  - N.Gehrig and P.L. Dragotti, Distributed Sampling and Compression of Scenes with Finite Rate of Innovation in Camera Sensor Networks, Proc. of IEEE Data Compression Conference, Snowbird, Utah, March 2006.
  - M. Gastpar, M. Vetterli and P.L. Dragotti, Sensing Reality and Communicating bits: a Dangerous Liaison, IEEE Signal Processing Magazine, July 2006.
- On Image Super-Resolution and Scene Interpretation
  - L. Baboulaz and P.L. Dragotti, Distributed Acquisition and Image Super-Resolution based on Continuous Moments from Samples, IEEE International Conference on Image Processing (ICIP), Atlanta, October 2006.
  - J. Berent and P.L. Dragotti, Segmentation of Epipolar-Plane Image Volumes with Occlusion and Disocclusion Competition, IEEE International Workshop on Multimedia Signal Processing (MMSP), Victoria (CA), October 2006.