Astronomical Data Analysis III

28 April - 1 May, 2004 Sant'Agata sui due Golfi, Italy

OBSERVABILITY AND PREDICTIBILITY IN SUPER-RESOLUTION

L.D. Alvarez[+], J. Mateos[+], R. Molina[+], and A. K. Katsaggelos[*]

[+] *Department of Computer Science and Artificial Intelligence, University of Granada, Spain.*

* *Department of Electrical and Computer Engineering, Northwestern University, Evanston, USA.*

Outline

1. Introduction.
2. Acquisition System.
   1. LowResolution ↔ HighResolution.
   2. CompressedLowResolution ↔ HighResolution
3. Regularization.
4. Why are not all the pixels valid to reconstruct the HR image?
   1. Observability problem.
   2. Predictibility problem.
5. Calculating observability maps.
6. Estimating high resolution images.
7. Experimental results.

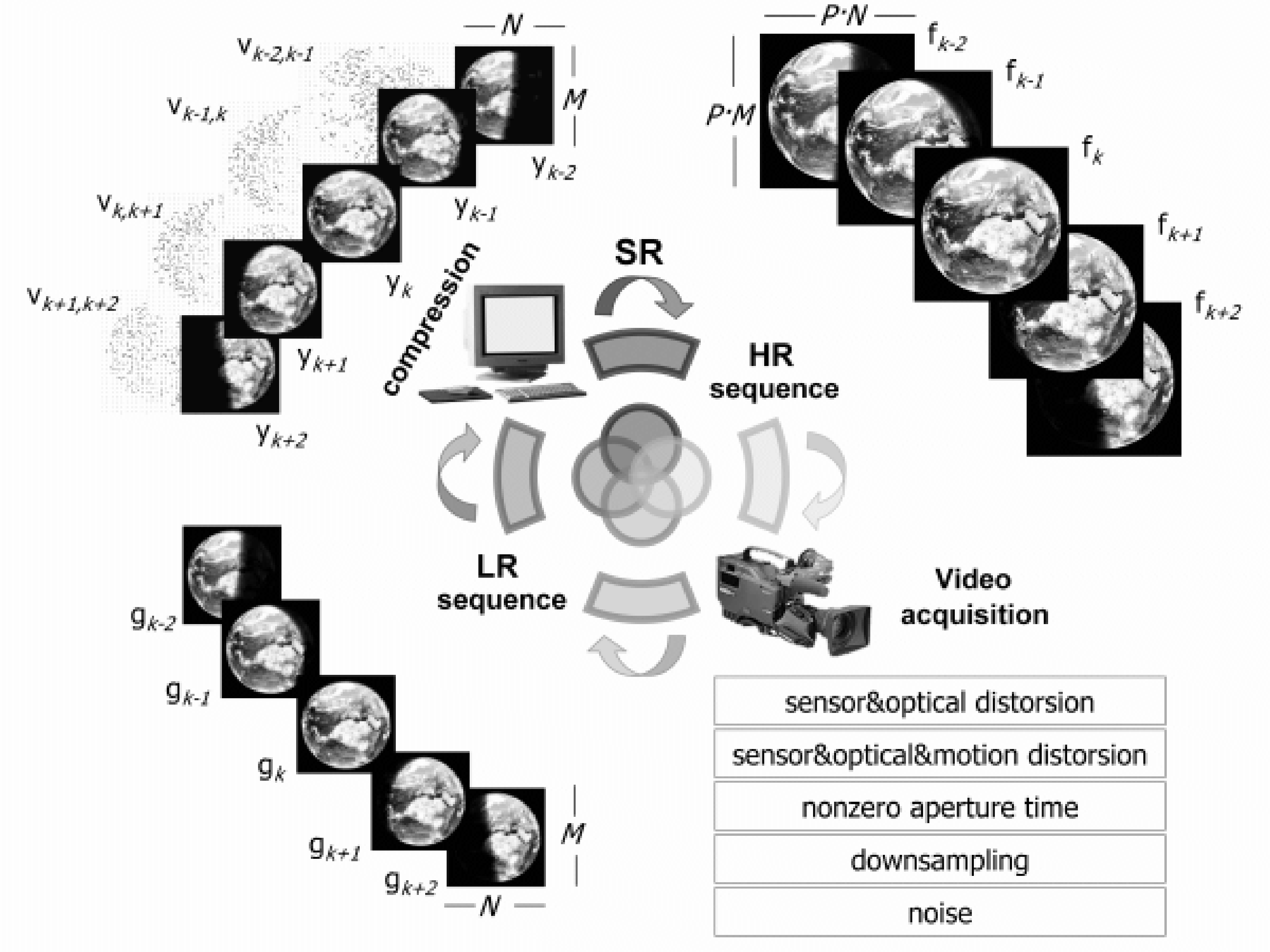
1.- Introduction

- Objective: To estimate a high resolution (HR) video sequence ($\mathbf{f}$) from a compressed low resolution (CLR) video sequence ($\mathbf{y}$ and $\mathbf{v}$) using only observable and predictable pixels ($\mathbf{y}^o$ and $\mathbf{v}^o$) in the reconstruction (the subscript ^{o} means ‘observable or predictable’).

- There will be useful and unuseful (concept to be defined later) pixels in the reconstruction process. How do we identify them?

2.- Acquisition system.

- We have a compressed low resolution video sequence (CLR) {$\mathbf{y}_l$ of size MxN}, that comes from the captured low resolution sequence (LR) {$\mathbf{g}_l$ of size MxN}. The original sequence is a high resolution sequence (HR) {$\mathbf{f}_l$ of size $P\cdot MxP\cdot N$}.
- The sequence is compressed using a hybrid motion compensated video compression method.


N
M
P·N
P·M
$v_{k-2,k-1}$
$v_{k-1,k}$
$v_{k,k+1}$
$v_{k+1,k+2}$
y_{k-2}
y_{k-1}
y_k
y_{k+1}
y_{k+2}
f_{k-2}
f_{k-1}
f_k
f_{k+1}
f_{k+2}
g_{k-2}
g_{k-1}
g_k
g_{k+1}
g_{k+2}
compression
SR
HR sequence
LR sequence
Video acquisition
sensor&optical distorsion
sensor&optical&motion distorsion
nonzero aperture time
downsampling
noise

2.1.- LowResolution ↔ HighResolution

- Objective: to estimate the HR video sequence $\mathbf{f}$, where $\mathbf{f}=\{\mathbf{f}_1, \mathbf{f}_2,\ldots, \mathbf{f}_L\}$ ($\mathbf{f}_l$ es (PMxPN)x1)
- Images are taken in a fast sequence so they are related by:

$$f_l(a,b)=f_k(a+d_{l,k}^x(a,b),\ b+d_{l,k}^y(a,b))+n_{l,k}(a,b) \ \equiv\ \mathbf{f}_l=\mathbf{C}(\mathbf{d}_{l,k})\mathbf{f}_k+\boldsymbol{\rho}_{l,k}$$

- During the acquisition there are two processes over the sequence: downsampling ($\mathbf{A}$) and blurring ($\mathbf{H}$). The relationship between a HR image ($\mathbf{f}_l$) and its corresponding LR image ($\mathbf{g}_l$) is:

$$\mathbf{g}_l=\mathbf{AH}\mathbf{f}_l+\boldsymbol{\eta}_l$$

- Combining the previous equations we obtain the relationship between the l-th LR image ($\mathbf{g}_l$) and the k-th HR image ($\mathbf{f}_k$):

$$\mathbf{g}_l=\mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k+\varsigma_{l,k} \quad \textit{where } \varsigma_{l,k}=\boldsymbol{\rho}_{l,k}+\boldsymbol{\eta}_l \ \textit{ is the acquisition noise}$$

2.2.- CompressedLowResolution ↔ HighResolution

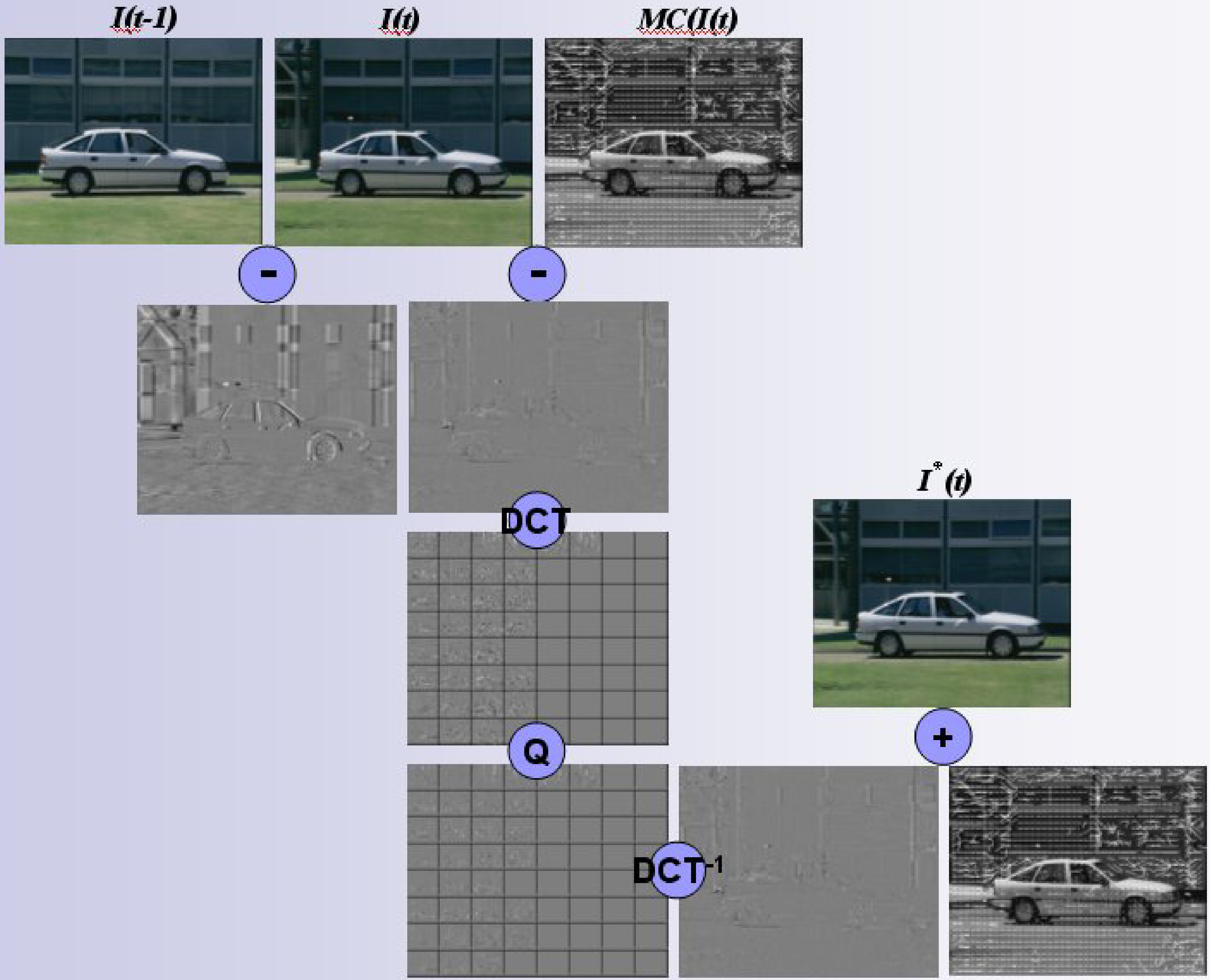
- There are several posibilities to encode a sequence: Intraframe coding; 3-D coding; Interframe coding; **Hybrid coding** (extends the interframe coding adding spatial compression, it is also called *motion compensated video coding*).
- The sequence is divided in groups of frames (GOPs), and each frame is encoded as: I (intra), P (predicted) and B (bidirectionally predicted). We will use I and P frames.
- P frames are predicted using previous frames; this is done by encoding the difference between the frame to be encoded and its prediction. This difference is transformed and quantified and added to the prediction. It is also sent to the decoder one motion estimation for each block.

$$\mathbf{y}_l = T^{-1}\{Q[T(\mathbf{g}_l - \mathbf{m}_l)]\} + \mathbf{m}_l$$

$\mathbf{y}_l$ □ l-th compressed low resolution image (CLR)

$\mathbf{g}_l$ □ l-th low resolution image (LR)

$\mathbf{m}_l$ □ l-th low resolution image obtained via motion compensation


I(t-1)
I(t)
MC(I(t))
-
-
DCT
I*(t)
Q
+
DCT-1

$$\mathbf{y}_l = \mathrm{T}^{-1}\{\, \mathrm{Q}[\, \mathrm{T}(\mathbf{g}_l - \mathbf{m}_l)\,]\,\} + \mathbf{m}_l$$

- If $\mathrm{T}^{-1}\{\mathrm{Q}[\mathrm{T}(\mathbf{g}_l - \mathbf{m}_l)]\}$ is approximated by $\mathbf{g}_l - \mathbf{m}_l + \boldsymbol{\varepsilon}_{Q,l}$ (where $\boldsymbol{\varepsilon}_{Q,l}$ is the prevailing noise, the quantization noise) then:

$$\mathbf{y}_l = \mathbf{g}_l - \mathbf{m}_l + \boldsymbol{\varepsilon}_{Q,l} + \mathbf{m}_l = \mathbf{g}_l + \boldsymbol{\varepsilon}_{Q,l}$$

- The quantization noise is due to Q, which is applied in the transformed domain T, its covariance matrix is $\mathrm{K}_{\mathrm{transformed.domain}}$. As $n_{spatial.domain} = T^{-1}\, n_{transformed.domain}$ then:

$$cov(n_{spatial.domain}) = T^{-1}\, K_{transformed.domain} (T^{-1})^T$$

- The quantization noise $\boldsymbol{\varepsilon}_{Q,l}$ follows a normal distribution with mean 0 and covariance matrix $\mathrm{K}_{Q,l}$:

$$\boldsymbol{\varepsilon}_{Q,l} \sim N(0, \mathrm{K}_{Q,l}) \quad \text{where} \quad \mathrm{K}_{Q,l} = T^{-1}\, \mathrm{K}_{transformed.domain}\, (T^{-1})^T$$

- $\mathrm{K}_{Q,l}$ can be obtained from the compressed bit-stream.

- We had from the acquisition process the equations

$$\mathbf{y}_l = \mathbf{g}_l + \boldsymbol{\varepsilon}_{Q,l} \quad \text{and} \quad \mathbf{g}_l = \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k + \boldsymbol{\varsigma}_{l,k}$$

then:

$$\mathbf{y}_l = \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k + \boldsymbol{\varsigma}_{l,k} + \boldsymbol{\varepsilon}_{Q,l}$$

$\boldsymbol{\varepsilon}_{Q,l} >> \boldsymbol{\varsigma}_{l,k}$, so we can rewrite this equation as:

$$\mathbf{y}_l = \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k + \boldsymbol{\varepsilon}_{Q,l}$$

And this equation relates the k-th HR frame and the l-th LR frame.

- The probability distribution, $P(\mathbf{y}_l|\mathbf{f}_k,\mathbf{d}_{l,k})$, is:

$$P(\mathbf{y}_l|\mathbf{f}_k,\mathbf{d}_{l,k}) = exp[-½(\mathbf{y}_l - \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k)^T \, K_{Q,l}^{-1} \, (\mathbf{y}_l - \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k)]$$

$$\mathbf{y}_l = \mathrm{T}^{-1} \{ \mathrm{Q} [\mathrm{T}(\mathbf{g}_l - \mathbf{m}_l)] \} + \mathbf{m}_l$$

- $\mathbf{m}_l$ is the motion compensated prediction from $\mathbf{g}_l$.
- Given $\mathbf{f}_k$ and $\mathbf{d}_{l,k}$:

$$\mathbf{m}_l = \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k + \boldsymbol{\tau}_{MV,l}$$

where $\boldsymbol{\tau}_{MV,l}$ is the noise due to the prediction; $\mathrm{K}_{MV,l}$ is the covariance matrix that describes the error between the uncompressed LR frame and its estimation via motion compensation.

$\boldsymbol{\tau}_{MV,l} \sim N(0, \mathrm{K}_{MV,l})$ where $\mathrm{K}_{MV,l} = T^{-1}\mathrm{K}_{transformed.domain} (T^{-1})^T$

- An estimation of $\mathrm{K}_{MV,l}$ can be obtained from the compressed bit-stream.
- The probability distribution of $\mathbf{m}_l$, $\mathrm{P}(\mathbf{m}_l|\mathbf{f}_k,\mathbf{d}_{l,k})$, is:

$$\mathrm{P}(\mathbf{m}_l|\mathbf{f}_k,\mathbf{d}_{l,k}) = exp[-\tfrac{1}{2}(\mathbf{m}_l - \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k)^{\mathrm{T}}\mathrm{K}_{MV,l}^{-1}(\mathbf{m}_l - \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k)]$$

3.- Regularization.

- The distribution of $\mathbf{f}_k$ reflects that images are smooth within homogeneous regions and free of blocking artifacts:

$$P(f_k) = \exp\left[-\tfrac{1}{2} \left(\lambda_1 \|\mathbf{Q}_1 \mathbf{f}_k\|^2 + \lambda_2 \|\mathbf{Q}_2 \mathbf{A}\mathbf{H}\mathbf{f}_k\|^2 \right)\right]$$

while $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are high pass filters and λ_1 and λ_2 weights controlling the influence of each term.

- Assuming independent displacements between frames we can write:

$$P(\mathbf{d}) = \Pi_l \, P(\mathbf{d}_{l,k})$$

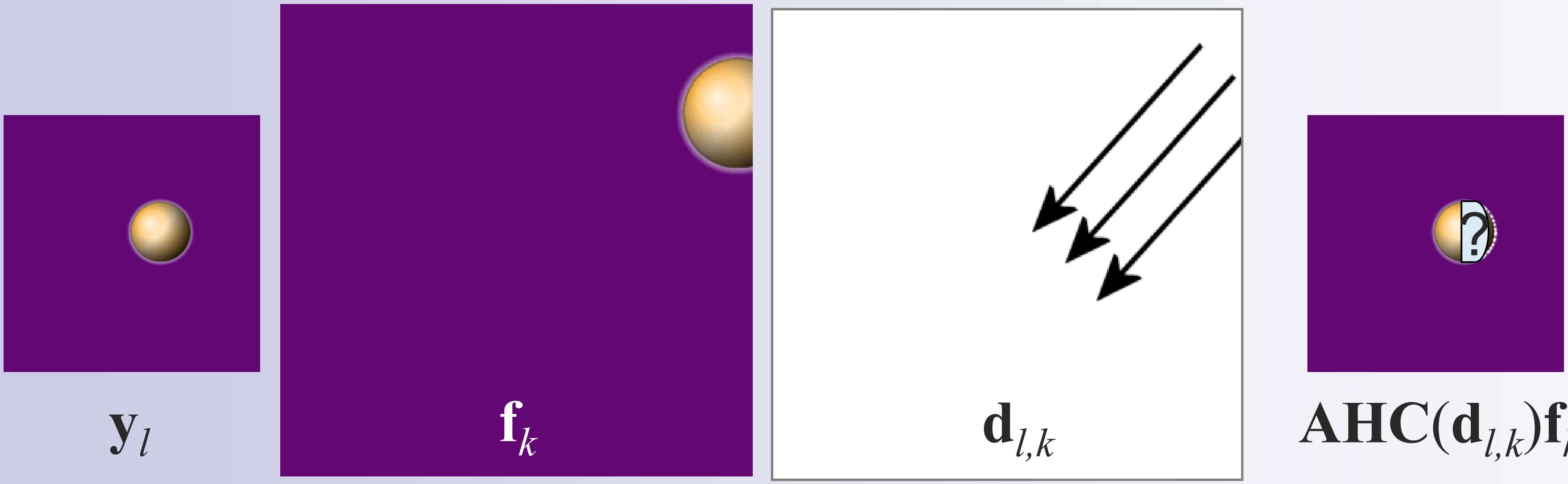
- We enforce $\mathbf{d}_{l,k}$ to be smooth:

$$P(\mathbf{d}_{l,k}) = \exp\left[-\tfrac{1}{2} \left(\lambda_3 \|\mathbf{Q}_3 \mathbf{d}_{l,k}\|^2 \right)\right]$$

4.- Why are not all the pixels valid to reconstruct the HR image?

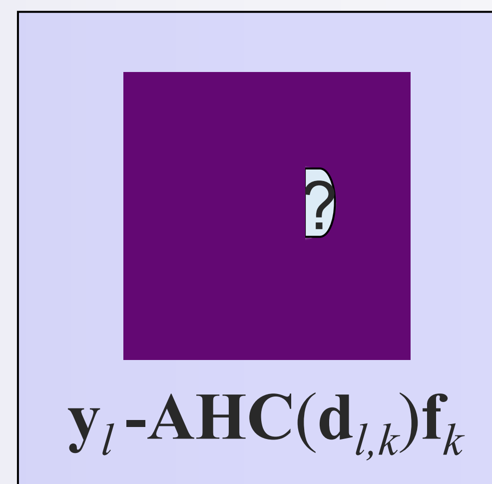
- Not all pixels are valid to reconstruct the HR image. We have two different problems:
  - Observability.
  - Predictibility.

4.1.- Observability problem.

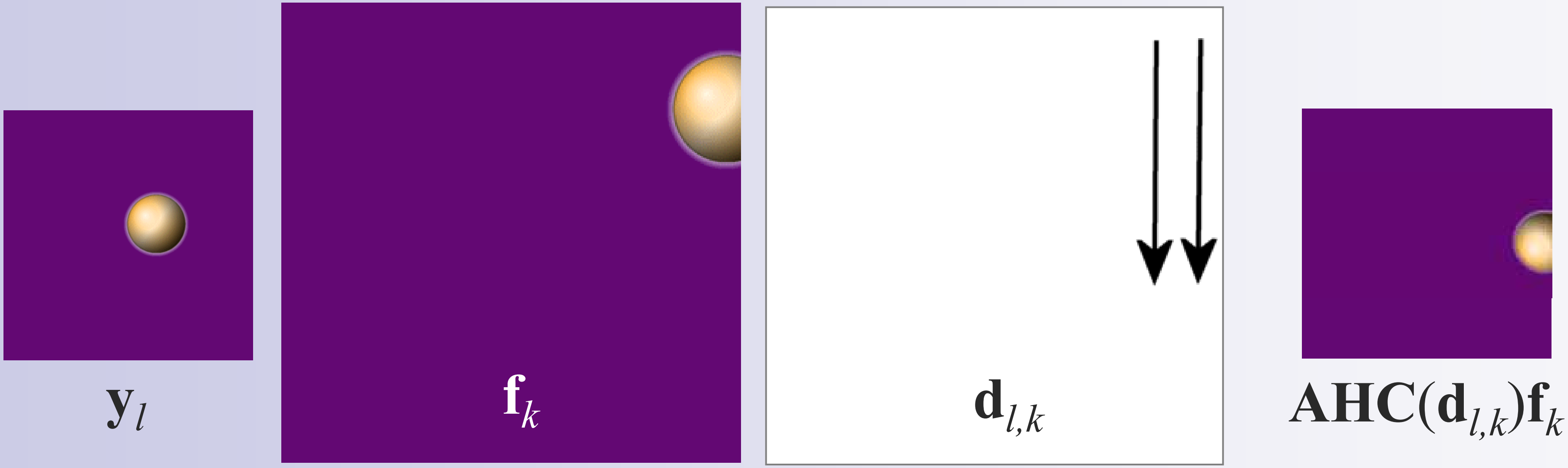


There are certain regions of $\mathbf{y}_l$ (or $\mathbf{m}_l$) that are not observable from $\mathbf{f}_k$ via $\mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k$.

Should we use these zones in our HR reconstruction?

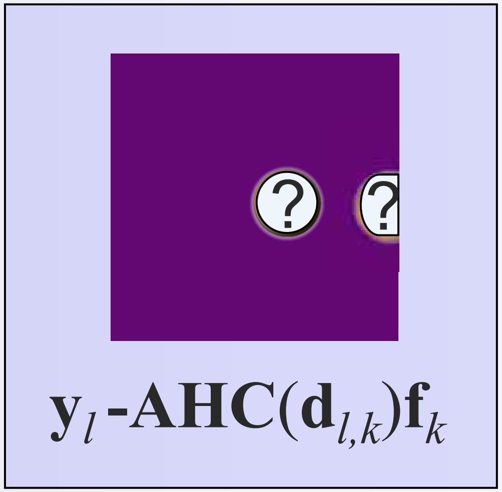


4.2.- Predictibility problem



If our motion estimation ($\mathbf{d}_{l,k}$) is not good enough there will be a problem with the difference between $\mathbf{y}_l$ (or $\mathbf{m}_l$) and its motion compensation prediction $\mathbf{f}_k$ ($\mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k$).

Should we use the corresponding pixels in our HR estimation?



- Our aim is to determine which pixels of the compressed LR sequence should be utilized to estimate the HR images.
- In the simplest scenario, to estimate the HR k-th frame we make use of the previous l-th LR image. But there will be regions in the l-th image that will not be observable or predictable from the k-th frame.
- Not observable or not predictable pixels do not provide useful information to the super-resolution method, so they will be removed.
- From now on, we will denote observability and predictibility as observability.

5.- Calculating observability maps.

Pixel observability:

$$DFD\left(\bar{\mathbf{f}}_{\mathbf{y}_l},\tilde{\mathbf{d}}_{l,k},\tilde{\mathbf{f}}_k\right)(m,n)=\left|\bar{\mathbf{f}}_{\mathbf{y}_l}(m,n)-\left(C\left(\tilde{\mathbf{d}}_{l,k}\right)\tilde{\mathbf{f}}_k\right)(m,n)\right|<T_y \quad \Rightarrow \quad (m,n) \quad observable$$

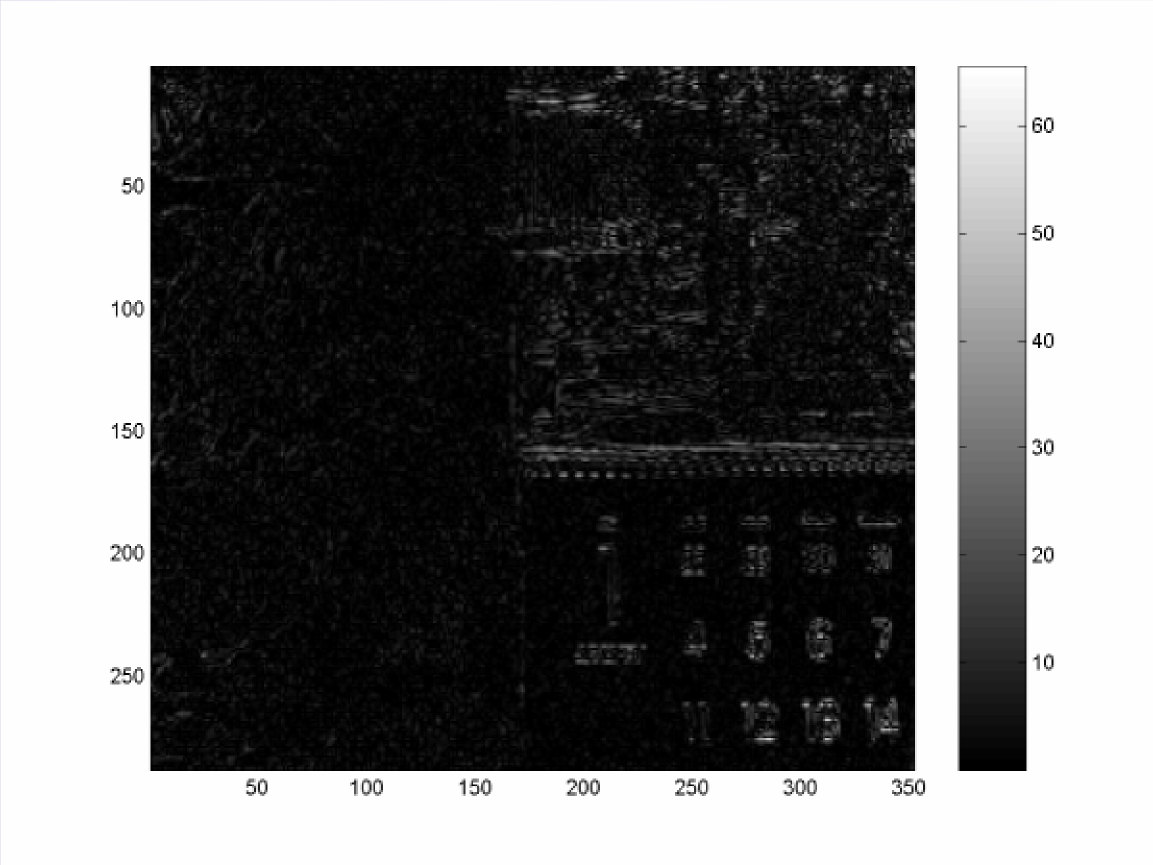
$$DFD\left(\bar{\mathbf{f}}_{\mathbf{m}_l},\tilde{\mathbf{d}}_{l,k},\tilde{\mathbf{f}}_k\right)(m,n)=\left|\bar{\mathbf{f}}_{\mathbf{m}_l}(m,n)-\left(\mathbf{C}\left(\tilde{\mathbf{d}}_{l,k}\right)\tilde{\mathbf{f}}_k\right)(m,n)\right|<T_m \quad \Rightarrow \quad (m,n) \quad observable$$

If a pixel is not observable, it has to be removed from the images $\mathbf{y}_l$ and $\mathbf{m}_l$ (they depend on $\mathbf{f}_l$). How do we remove these pixels in the compressed low resolution images?

Procedure to calculate the observability maps:

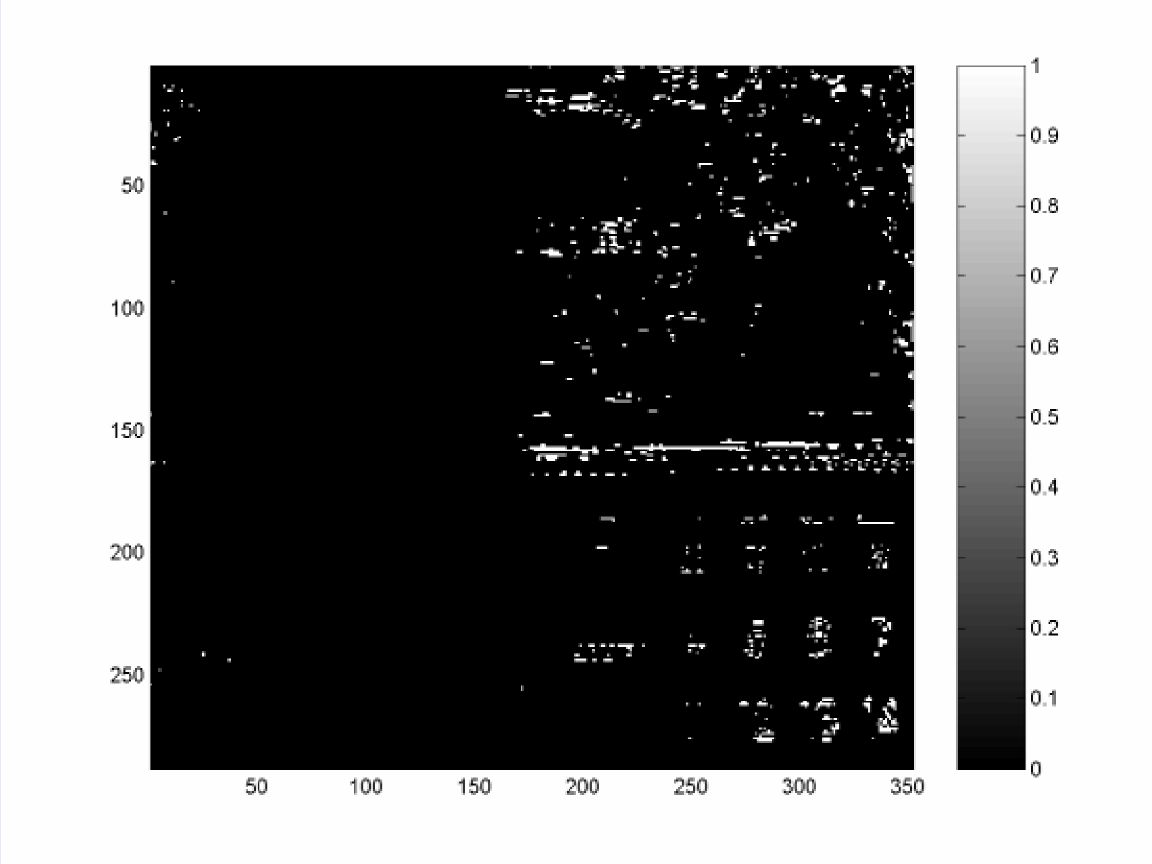
- Once the DFDs have been calculated the observability maps will be obtained from them. The procedure is the following:

  1. Calculate DFD_{H}R (as shown).
  2. Obtain $ErrorsMap_{HR}$ that contains ones if $DFD_{AR} > T$ (and zeros elsewhere).
  3. Downsample $ErrorsMap_{HR}$ using a mean filter to obtain a new errors map but in LR ($ErrorsMap_{LR}$).
  4. If a certain pixel in $ErrorsMap_{BR}$ is lower than 0.25 it will be observable (predictable).



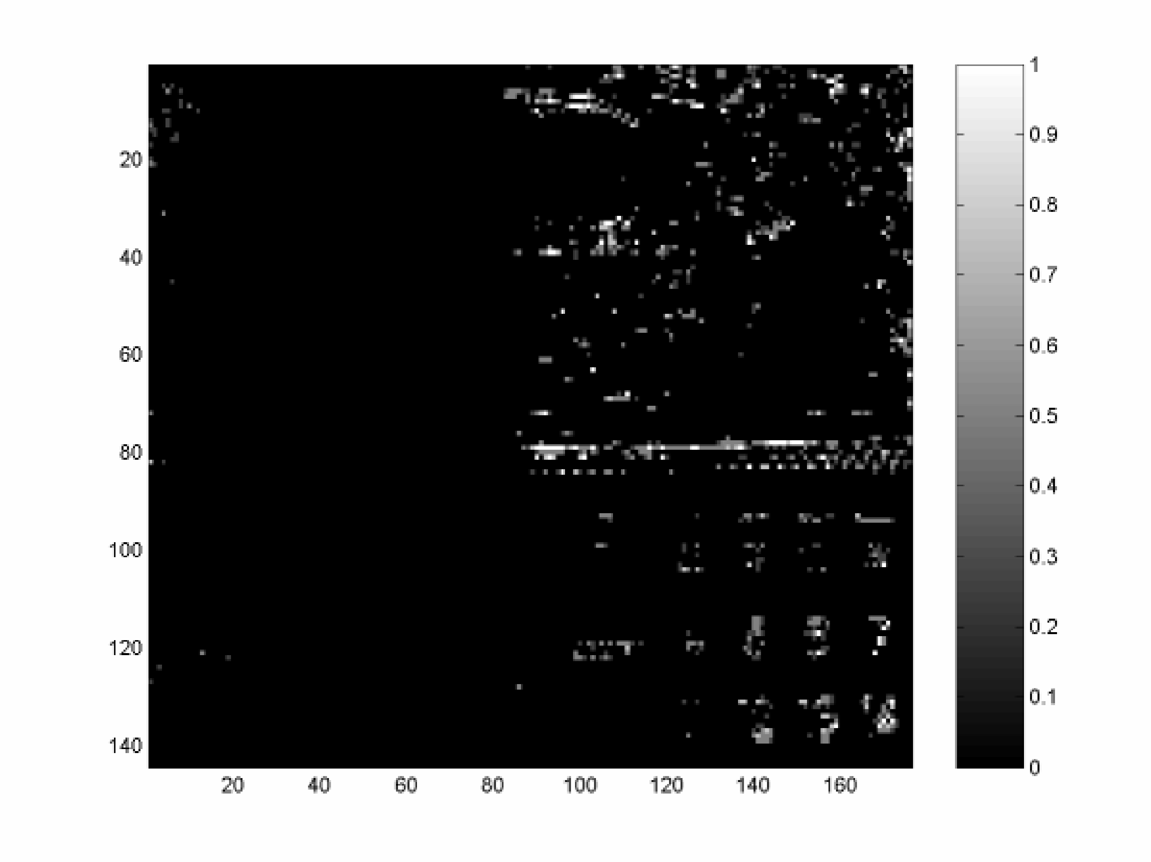
1.- $\mathbf{DFD}_{\mathbf{HR}}$

Difference between the fixed image (k) and the l-th estimated image.



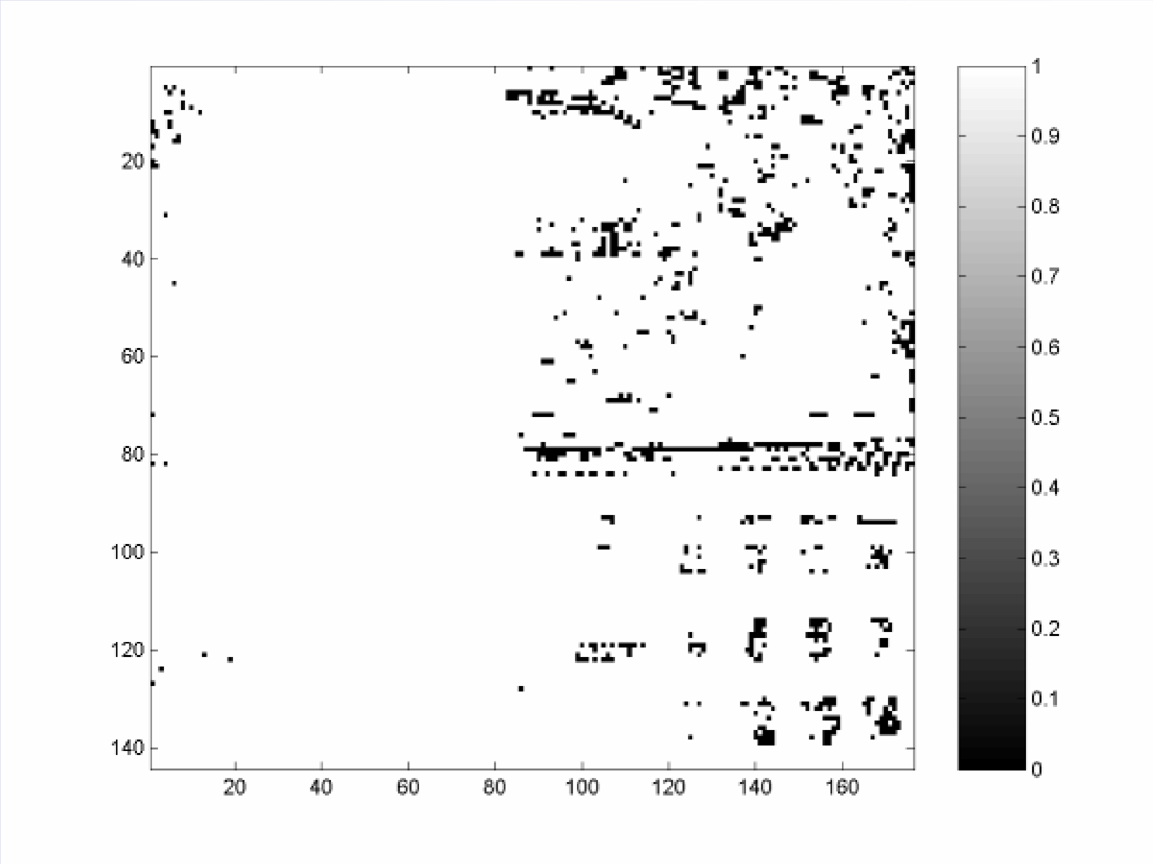
2.- $\mathbf{ErrorMap_{HR}=[DFD_{HR}>T_y\,]}$ $\mathbf{(ErrorMap_{HR}=[DFD_{HR}>T_m\,])}$

A high resolution pixel with a value greater than a given threshold (T_y or T_m) will not be observable.



3.- $\mathbf{ErrorMap_{LR}} = \downarrow[\mathbf{mean}(\mathbf{ErrorMap_{HR}})]$

Use a mean filter on the binary error map and downsample it.



4.- $\text{ObservabilityMap} = \text{ErrorMap}_{LR} < 0.25$

- Once we have calculated the observability maps the images $\mathbf{y}_l^o$ and $\mathbf{m}_l^o$ are available. These images come from $\mathbf{y}_l$ and $\mathbf{m}_l$, selecting only the pixels that are useful for the reconstruction (observable and predictable pixels).
- We had:

$$P(\mathbf{y}_l|\mathbf{f}_k,\mathbf{d}_{l,k})=exp[-½(\mathbf{y}_l–\mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k)^T K_{Q,l}^{-1} (\mathbf{y}_l–\mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k)]$$

$$P(\mathbf{m}_l|\mathbf{f}_k,\mathbf{d}_{l,k})=exp[-½(\mathbf{m}_l–\mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k)^T K_{MV,l}^{-1}(\mathbf{m}_l–\mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k)]$$

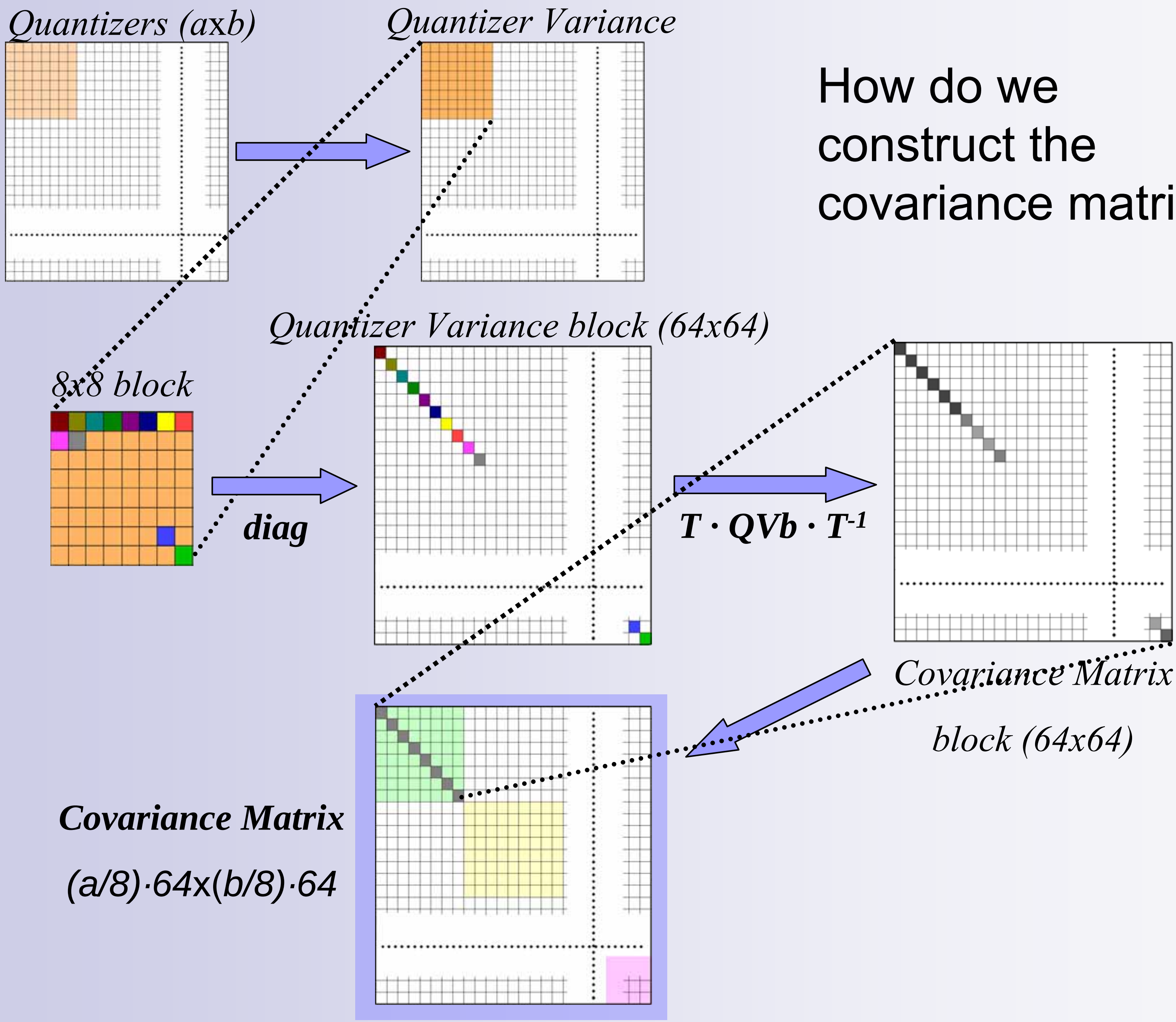
And finally:

$$P(\mathbf{y},\mathbf{m}|\mathbf{f}_k,\mathbf{d}) = \Pi^l [P(\mathbf{y}_1|\mathbf{f}_k,\mathbf{d}_{l,k})\ P(\mathbf{m}_l|\mathbf{f}_k,\mathbf{d}_{l,k})] \quad \text{para } l=1,2,\ldots,L$$

- But we have to introduce here the observability notion.

- Now, we have
  - $\mathbf{y}_l^o$ → l-th CLR frame using only observable pixels.
  - $\mathbf{m}_l^o$ → observable pixels in the prediction, obtained via motion compensation from $\mathbf{y}_l^o$.
- If $\mathbf{y}_l$ is now $\mathbf{y}_l^o$ the difference $\mathbf{y}_l - \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k$ will be $\mathbf{y}_l^o - \mathbf{A}_l^o\mathbf{HC}(\mathbf{d}_{l,k})\mathbf{f}_k$.
- Analogously, from $\mathbf{m}_l - \mathbf{AHC}(\mathbf{d}_{l,k})\mathbf{f}_k$ we have $\mathbf{m}_l^o - \mathbf{A}_l^o\mathbf{HC}(\mathbf{d}_{l,k})\mathbf{f}_k$.
- $\mathbf{A}_l^o$ matrices are different due to the observability maps.
- The covariance matrices are also differents:
  - $K_{Q,l}^{-1}$ → $(K_{Q,l}^{o})^{-1}$.
  - $K_{MV,l}^{-1}$ → $(K_{MV,l}^{o})^{-1}$.

How do we construct the covariance matrix.



- Including the observability maps, the probability distributions become:

$$P(\mathbf{y}_l^o | \mathbf{f}_k, \mathbf{d}_{l,k}) =$$

$$exp[-½(\mathbf{y}_l^o - \mathbf{A}_l^o\ \mathbf{HC}(\mathbf{d}_{l,k})\mathbf{f}_k)^T\ (K_{Q,l}^o)^{-1}\ (\mathbf{y}_l^o - \mathbf{A}_l^o\ \mathbf{HC}(\mathbf{d}_{l,k})\mathbf{f}_k)]$$

$$P(\mathbf{m}_l^o | \mathbf{f}_k, \mathbf{d}_{l,k}) =$$

$$exp[-½(\mathbf{m}_l^o - \mathbf{A}_l^o\ \mathbf{HC}(\mathbf{d}_{l,k})\mathbf{f}_k)^T\ (K_{MV,l}^o)^{-1}\ (\mathbf{m}_l^o - \mathbf{A}_l^o\ \mathbf{HC}(\mathbf{d}_{l,k})\mathbf{f}_k)]$$

6.- Estimating high resolution images.

Our aim is to estimate the high resolution image $\mathbf{f}_k$ and high resolution motion vectors, $\mathbf{d}$, satisfying:

$$\hat{\mathbf{f}}_k, \hat{\mathbf{d}} = \arg\max_{\mathbf{f}_k, \mathbf{d}} \left[P(\mathbf{f}_k, \mathbf{d}) P\left(\mathbf{y}^o, \mathbf{m}^o \mid \mathbf{f}_k, \mathbf{d}\right) \right]$$

<u>Method:</u>

Starting with an initial estimation $\hat{\mathbf{f}}_k^{q=0}$ and $\hat{\mathbf{d}}^{q=0}$

Iterate until convergence:

$\mathbf{y}^o$ $$DFD\left(\bar{\mathbf{f}}_{\mathbf{y}_l}, \hat{\mathbf{d}}_{l,k}^q, \hat{\mathbf{f}}_k^q\right)(m,n) = \left| \bar{\mathbf{f}}_{\mathbf{y}_l}(m,n) - \left(C\left(\hat{\mathbf{d}}_{l,k}\right)\hat{\mathbf{f}}_k\right)(m,n) \right|$$

$\mathbf{m}^o$ $$DFD\left(\bar{\mathbf{f}}_{\mathbf{m}_l}, \hat{\mathbf{d}}_{l,k}^q, \hat{\mathbf{f}}_k^q\right)(m,n) = \left| \bar{\mathbf{f}}_{\mathbf{m}_l}(m,n) - \left(\mathbf{C}\left(\hat{\mathbf{d}}_{l,k}\right)\hat{\mathbf{f}}_k\right)(m,n) \right|$$

$$\hat{\mathbf{d}}^{q+1} = \arg\max_{\mathbf{d}} \left[P(\mathbf{d}) P\left(\mathbf{y}^o, \mathbf{m}^o \mid \hat{\mathbf{f}}_k^q, \mathbf{d}\right) \right]$$

$$\hat{\mathbf{f}}_k^{q+1} = \arg\max_{\mathbf{f}_k} \left[P(\mathbf{f}_k) P\left(\mathbf{y}^o, \mathbf{m}^o \mid \mathbf{f}_k, \hat{\mathbf{d}}^{q+1}\right) \right]$$

(where $\bar{\mathbf{f}}_{y_l}$ and $\bar{\mathbf{f}}_{m_l}$ are initial estimates of the corresponding high resolution images)

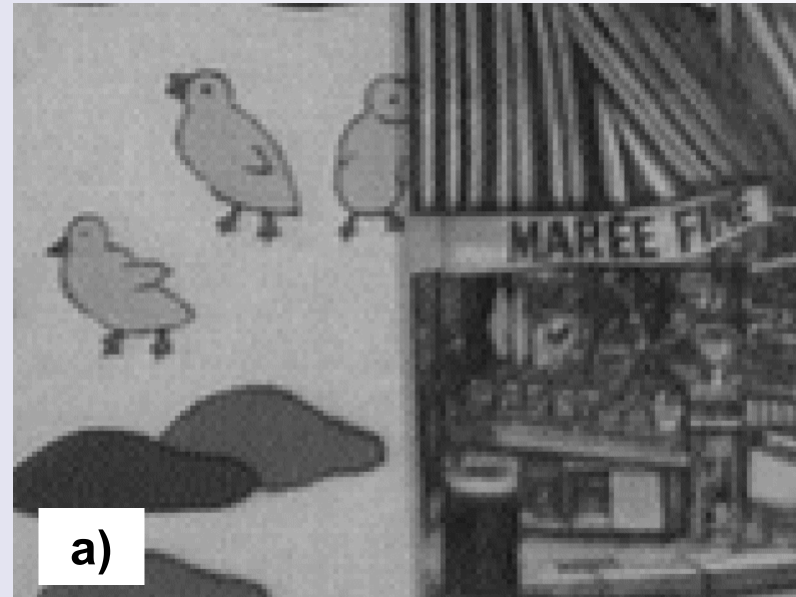
7.- Experimental results.

Reconstruction example

a) HR original image (cropped region).

b) CLR obtained by bilinear interpolation.

c) Reconstruction obtained using the proposed method.



Consecutive CLR frames (numbers: 8, 9, 10) of the sequence.

HR frame number 9.

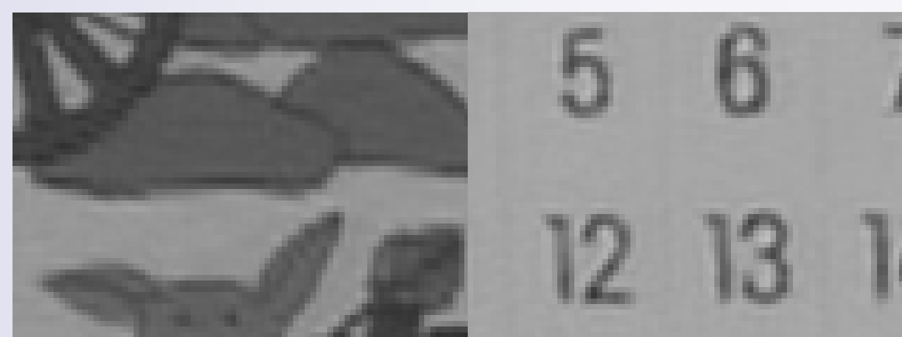
Observable pixel maps for the CLR frames and LR motion compensated numbers 7, 8, 10 and 11 using thresholds $T_y=16$ and $T_m=7$.

Bilinear interpolation of frame 9.

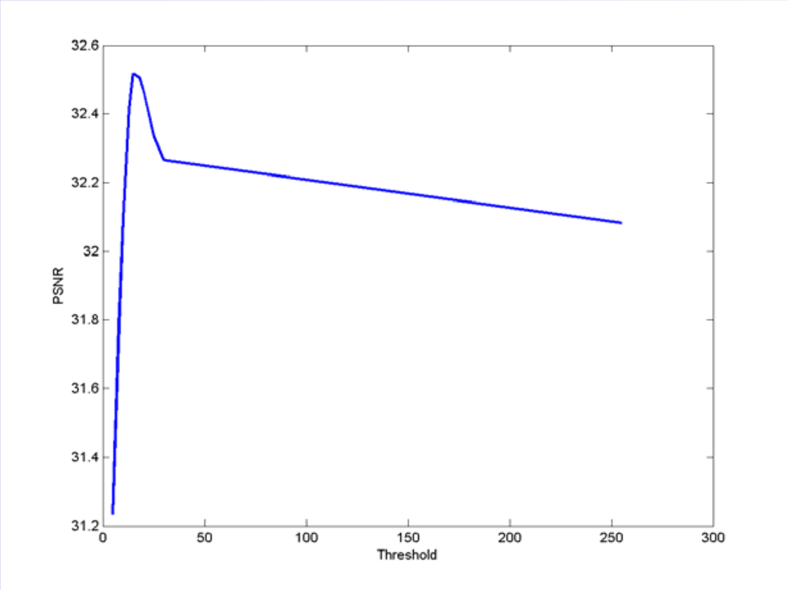
HR estimation without observability maps (fig. A).

HR estimation using observability maps (fig. B).

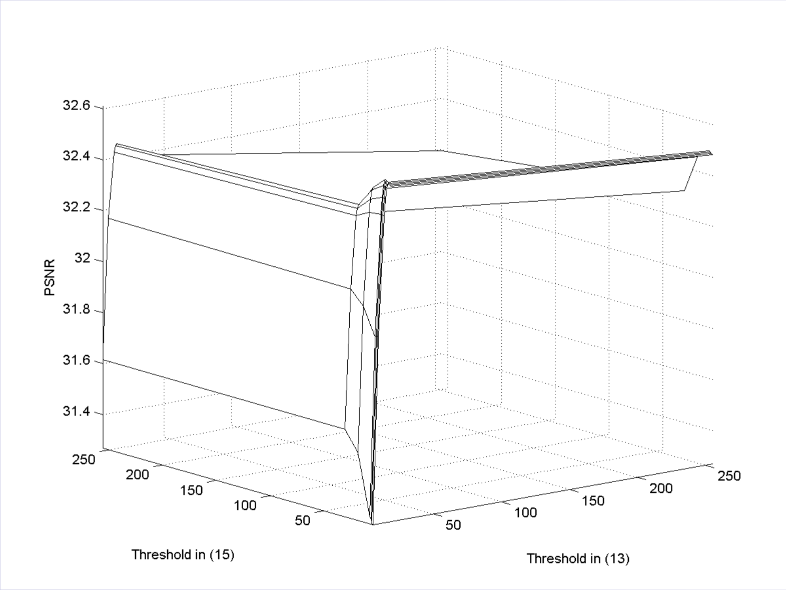
Two details of image in fig. A.



Two details of image in fig. B



PSNR evolution using only observability map for $\mathbf{y}_l$ with different threshold values.



PSNR evolution using observability maps for $\mathbf{y}_l$ and $\mathbf{m}_l$ with different threshold values.