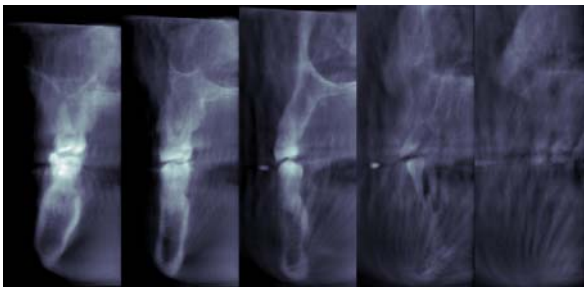
Optimized limited angle tomography Filtering through regularization

Applied Intelligent Systems Laboratory
Department of Computer Science
N. Alberto Borghese



Overview

- Algebric reconstruction and tomosynthesis
- Limited angle tomography

Where does it start from?

For producing a panoramic image, emitter and sensor rotate together.

Why not using multiple rotations for multiple images?

Hyperion
CEFLA (MyRay)

Motivation for limited angle tomography

Dose reduction (ALARA principle)
Costs reduction

MAMMOMAT - Siemens

Hyperion
CEFLA (MyRay)

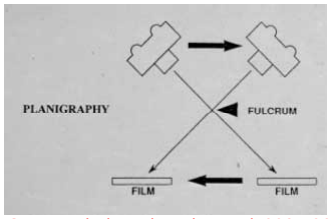
For several applications, few images are sufficient to recover the information that is needed by clinicians.

Where did it come from? (Tiggelen 2002)

Bocage A.E.M.: Procédé et dispositif de radiographie sur plaque en mouvement. Brevet français N° 534 464, 1922.

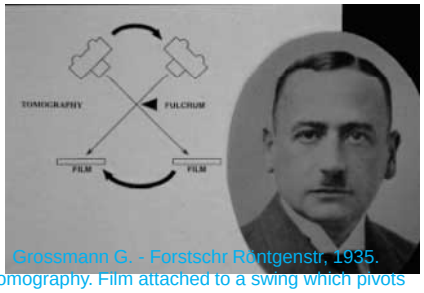
At the same time C. Baese from Florence, developed a tool to identify the depth of bullets inside soldiers.

Vallebona Alessandro Radiol Med, 1930: stratigraphy



George Ziedses des Plantes (1902-1993), Acta Radiol., 1932 replicates slicing in microscopes with slices on film.

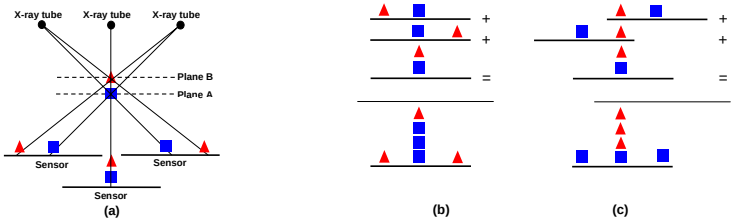
1 Slice only can be imagined for each acquisition.



Grossmann G. - Forstschr Röntgenstr, 1935. Tomography. Film attached to a swing which pivots around a horizontal axis over the patient..

5/79

Tomosynthesis (retroproiezione, shift and add)

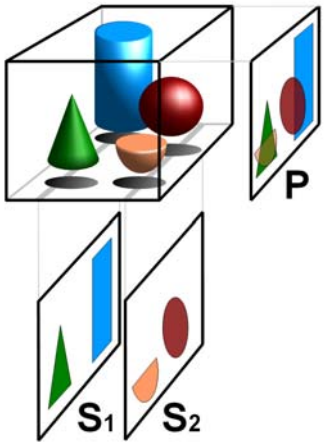


- It allows reconstructing a set of parallel planes.
- Classical tomosynthesis (Shift & add) is equivalent to **backprojection (blurring of out of focus structures)**…
- Characteristics:
  - **limited angle of view** and limited number of images.
  - Images are not necessary equally spaced

http://ais-lab.dsi.unimi.it

6/81

Tomography is a typical inverse problem

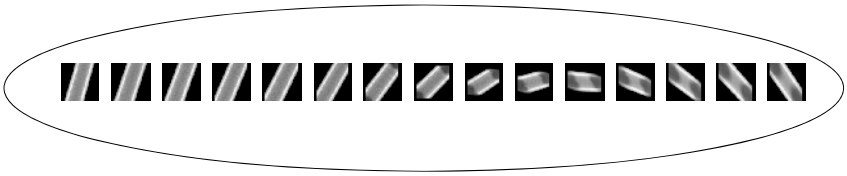


We start from N images of MxP pixels each

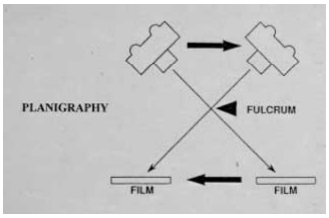
Volume as a set of parallel planes or voxels.

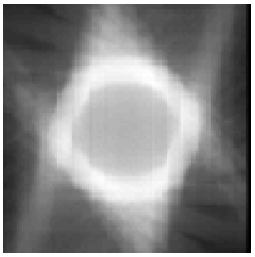
Interpolation between adjacent planes (Marching cubes)

Results with Tomosynthesis



Section of a cylinder acquired in the previous 15 images.





Impulse response of Tomosynthesis

Blurred reconstruction

Tomosinthesis limitations - 1

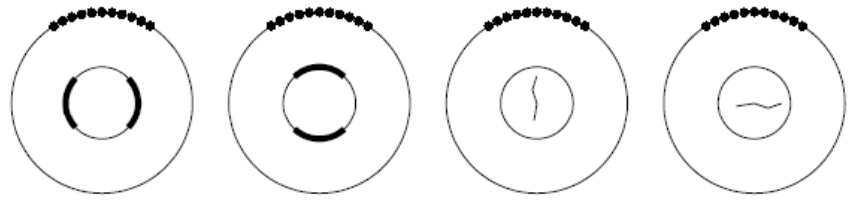
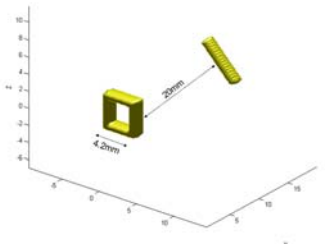


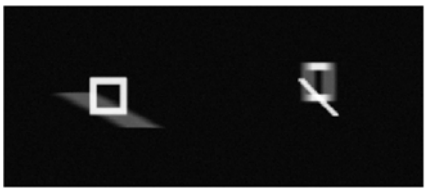
Figure 3. Illustration of discontinuities that can and cannot be recovered based on limited-angle projection data. The black dots denote the locations of the x-ray source in the acquisition of the projection data. For each such location, the detector is thought to be located opposite to the x-ray source as in figure 1.

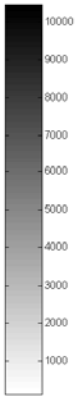
Siltanen et al., Medical Physics, 2003

•Structures orthogonal to the rays cannot be seen
•Good resolution in planes orthogonal to the principal X-ray, poor resolution parallel to the X-ray.

Tomosinthesis limitations - 2







Test objects

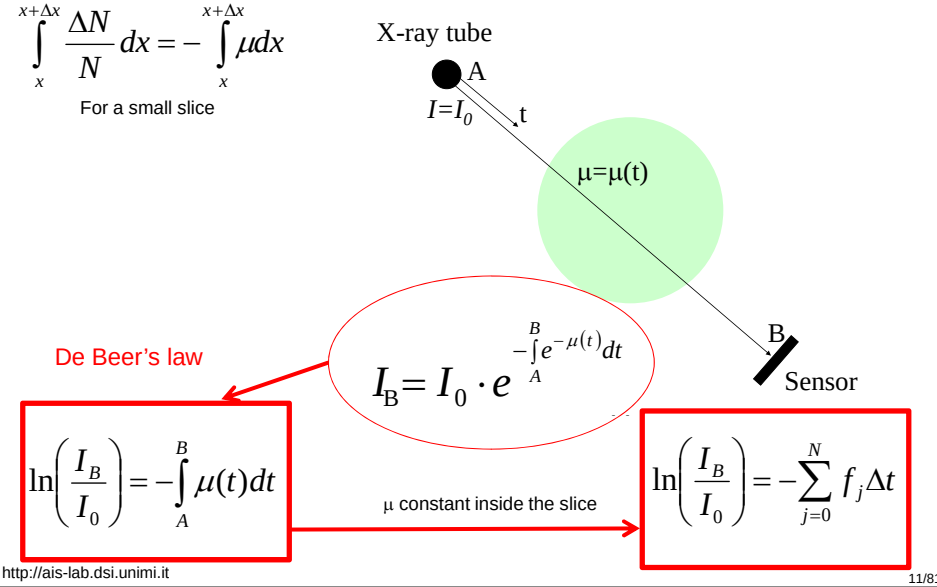
2 sections tomosynthesized:
- The central section of the square
- The central section of the line.

Shadows are evident. How can we remove them? (Filtering)

Tomography as a linear problem

$$\int_{x}^{x+\Delta x} \frac{\Delta N}{N} dx = -\int_{x}^{x+\Delta x} \mu dx$$

For a small slice



De Beer's law

$$I_B = I_0 \cdot e^{-\int_A^B \mu(t)dt}$$

$$\ln\left(\frac{I_B}{I_0}\right) = -\int_A^B \mu(t)dt$$

μ constant inside the slice

$$\ln\left(\frac{I_B}{I_0}\right) = -\sum_{j=0}^{N} f_j \Delta t$$

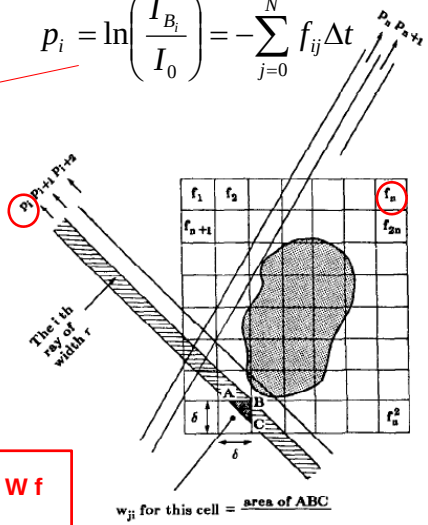
Tomography – discrete case

Voxels and pixels are introduced.

$$p_i = \ln\left(\frac{I_{B_i}}{I_0}\right) = -\sum_{j=0}^{N} f_{ij} \Delta t$$

Integrate over voxels and over pixels.

- p_i measured value at pixel i: number of photons that hit the i-th pixel.
- f_j voxel absorption: average value of μ inside the voxel j.
- w_{ij} portion of the voxel j-th crossed by the X-rays from the emitter that reach the i-th pixel (impact of the voxel on the measurement of pixel i).



From Kak & Slaney 2002

A system of linear equations results: **p = W f**

Parallel X-rays (fan -> parallel geometry)

Role of w_{ij}

Proiezione

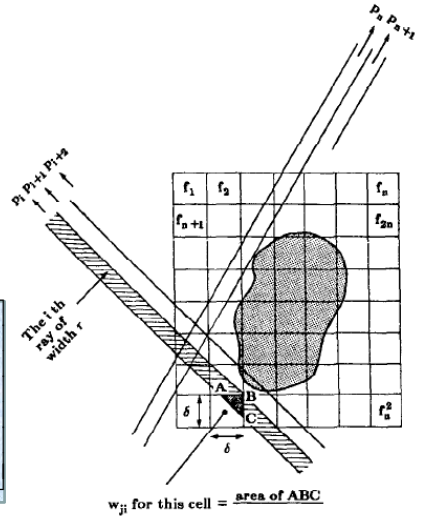
$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \\ p_5 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ & & & \\ w_{51} & w_{52} & w_{53} & w_{54} \end{bmatrix} \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}$$

$W_{*,i}$ è sparsa

Retro-Proiezione

$$\begin{bmatrix} f^*_1 \\ f^*_2 \\ f^*_3 \\ f^*_4 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{21} & w_{31} & w_{41} & w_{51} \\ w_{12} & w_{22} & w_{32} & w_{42} & w_{52} \\ w_{13} & w_{23} & w_{33} & w_{43} & w_{53} \\ w_{14} & w_{24} & w_{34} & w_{44} & w_{45} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix}$$

$W^T_{*,i}$ è quindi anch'essa sparsa



Observations on the Role of w_{ij}

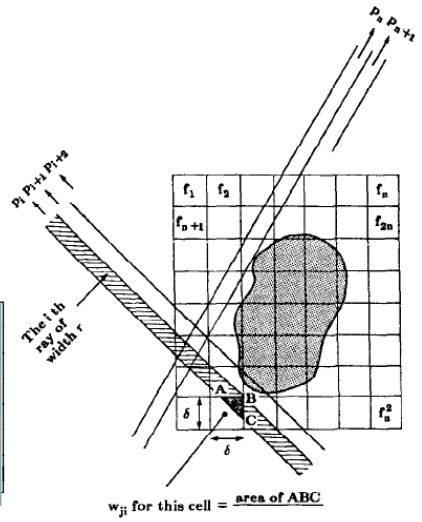
Proiezione

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \\ p_5 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ & & & \\ w_{51} & w_{52} & w_{53} & w_{54} \end{bmatrix} \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}$$

$W_{*,i}$ è sparsa

Retro-Proiezione

$$\begin{bmatrix} f^*_1 \\ \\ f^*_3 \\ f^*_4 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{21} & w_{31} & w_{41} & w_{51} \\ & & & & \\ w_{13} & w_{23} & w_{33} & w_{43} & w_{53} \\ w_{14} & w_{24} & w_{34} & w_{44} & w_{45} \end{bmatrix} \begin{bmatrix} \\ \\ \\ \\ \\ \end{bmatrix}$$



f_{i*} is the tomosynthesis image and it is obtained as the weighted sum of the contribution of the different pixels.

Properties of $p = W^T f$

Proiezione

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \\ p_5 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ & & & \\ w_{51} & w_{52} & w_{53} & w_{54} \end{bmatrix}$$

$W_{*,i}$ è sparsa

Retro-Proiezione

$$\begin{bmatrix} f^*_1 \\ f^*_2 \\ f^*_3 \\ f^*_4 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{21} & w_{31} & w_{41} & w_{51} \\ w_{12} & w_{22} & w_{32} & w_{42} & w_{52} \\ w_{13} & w_{23} & w_{33} & w_{43} & w_{53} \\ w_{14} & w_{24} & w_{34} & w_{44} & w_{45} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix}$$

$W^T_{*,i}$ è quindi anch'essa sparsa

p = **W f** $\quad$ $\mathbf{W^T p} = \mathbf{f^*}$ $\quad$ sostituisco **p**

$\mathbf{f^*} = \mathbf{W^T W f}$ $\quad$ **f*** is not equal to **f** because of the ghost images (blurring)

It could be made equal if no blurring would have occurred.

$f^*_1 = (w_{11}\, w_{11} + w_{21}\, w_{21} + w_{31}\, w_{31} + w_{41}\, w_{41} + w_{51}\, w_{51})\, f_1$

Some numbers on W

Voxels number = 256 x 256 x 256 = 2^{24}

Number of images = 128 = 2^7
Images dimension = 512 x 512 = 2^{18}

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_j \\ p_M \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{1N} \\ w_{21} & w_{22} & w_{23} & w_{2N} \\ w_{31} & w_{32} & w_{33} & w_{3N} \\ w_{j1} & w_{j2} & w_{jk} & w_{jN} \\ w_{M1} & w_{M2} & w_{Mk} & w_{MN} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_k \\ f_N \end{bmatrix}$$

Size of **W** is N x M

In this case:

2^{24} x 2^{25}
≅
16 x10^6 x 32 x 10^6
W is sparse

: **p = W f** $\quad$ Cannot be solved through algebra

How to remove artifacts?

Selective plane removal. Analytical solution for three planes. Gosh-Roy et al, Med Phys. 1985)

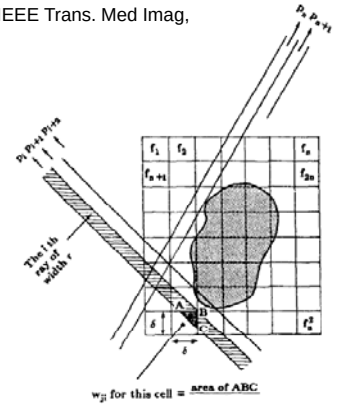
General approach to deblurring in tomosynthesis (Ruttimann et al., IEEE Trans. Med Imag, 1984)

$\mathbf{p} = \mathbf{W}\,\mathbf{f}$ → $\mathbf{f} = \mathbf{W}^T \cdot \mathbf{p}$

Projection

Back-projection (tomosynthesis)

$$\mathbf{f}^{t+1} = \mathbf{f}^t + \lambda \cdot \left(\mathbf{W}^T\mathbf{p} - \mathbf{W}^T\mathbf{W} \cdot \mathbf{f}^t\right) \quad \textit{Ruttiman}$$



Comparison of the backprojected measured images with the backprojected images obtained from the projection of the actual volume. The error is used to update f_{ij}.

A more efficient implementation by matrix rearrangement

Voxel number = 256x256x256 = 2^{24}

Size of **W** is 2^{24} x 2^{25}

Number of images = 128 = 2^7
Images dimension = 512 x 512 = 2^{18}

O(N x M x N) = 2^{24} x 2^{25} x 2^{24} = 2^{73}

$$\mathbf{f}^{t+1} = \mathbf{f}^t + \lambda \cdot \left(\mathbf{W}^T\mathbf{p} - \mathbf{W}^T\mathbf{W} \cdot \mathbf{f}^t\right) = \quad \textit{Ruttiman}$$

$$\mathbf{f}^{t+1} = \mathbf{f}^t + \lambda \cdot \mathbf{W}^T \cdot \left(\mathbf{p} - \mathbf{W} \cdot \mathbf{f}^t\right) = \quad \textit{Algebric method}$$

2 x O(N x M) = 2 x 2^{24} x 2^{25} = 2^{50}

Besides there is a computational and memory problem to store **W**.

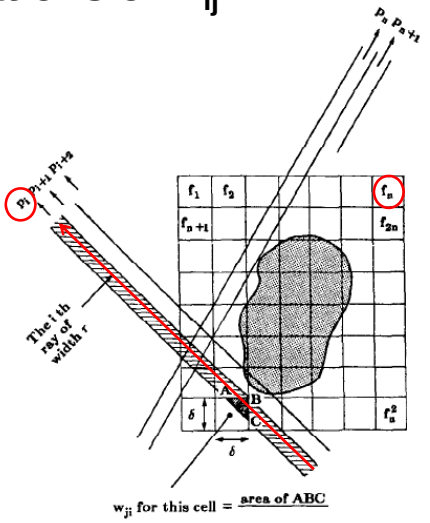
W is computed on-the-fly

w_{ij} are computed on the fly. Approximations of w_{ij}

w_{ij} can be approximate with a binary value (cross / not cross) as in **ART**

w_{ij} can be approximated as the length of the intersection of the central axis.

w_{ij} can be approximated as the inverse of the distance between the ray and the voxel center.



From Kak & Slaney 2002

Parallel X-rays (fan -> parallel geometry)

SART - Simultaneous Algebric Reconstruction Technique (Anderson & Kak, 1984)

It is borne on projections method to solve iteratively linear systems:

S.Kaczmarz, "Angenaherte auflosung von systemen linearer gleichungen," Bull. Acad. Pol. Sci. Lett. A, vol. 6-8A, pp. 355-357, 1937.

K.Tanabe, "Projection method for solving a singular system," Num. Math., vol.17, pp.203-214, 1971

$$p = W f$$

$$p_i = \sum_{j=1}^{N} w_j f_j$$

Modify f_j such that the equations are satisfied.
1 equation at a time is considered.

Back-projection

Error

$$f_j^{(t+1)} = f_j^{(t)} + \frac{w_{ji}\left(p_i - \mathbf{w_i}\cdot\mathbf{f^{(t)}}\right)}{\sum_i w^2_{ij}}$$

SART scales the update for each voxel

$$\mathbf{f}^{t+1} = \mathbf{f}^t + \lambda\cdot\left(\mathbf{W}^{\mathrm{T}}\mathbf{p} - \mathbf{W}^{\mathrm{T}}\mathbf{W}\cdot\mathbf{f}^t\right) = \quad Ruttiman$$

$$\mathbf{f}^{t+1} = \mathbf{f}^t + \lambda\cdot\mathbf{W}^{\mathrm{T}}\cdot\left(\mathbf{p} - \mathbf{W}\cdot\mathbf{f}^t\right) = \quad Algebric\ method$$

SART general formulation

$$f_j^{(t+1)} = f_j^{(t)} + \frac{w_{ji}\left(p_i - \mathbf{w_i}\cdot\mathbf{f}\right)}{\sum_i w^2{}_{ij}}$$

SART iterates through the following steps:

1) Choose randomly a projection image, i, (Alternatively Ordered-subsets can be used). (if we choose a single pixel -> ART, if we choose all the images -> SIRT)

2) Compute the current absorption along each ray of the image ($\mathbf{w_i}\,\mathbf{f}$) – forward projection.

3) Compute the error on the image (projection error), compute the error.

4) Back-project the error along the ray to the voxels ($\mathbf{W^T}$.), retro-project the error.

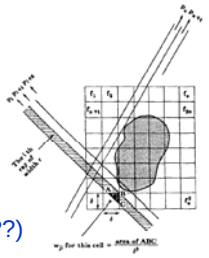
5) Updated voxel j as a fraction of the error captured by the voxel. Update.

SART interpretation

Given: $\mathbf{p} = \mathbf{W}\,\mathbf{f}$

Let us suppose that we minimize: $\left[\mathbf{D}\left(\mathbf{p} - \mathbf{W}\cdot\mathbf{f}\right)\right]^2$

with **D** diagonal



We apply gradient descent optimization and obtain: (Lanweber??)

$$f^{t+1} = f^t + \lambda\cdot\mathbf{W}^{\mathbf{T}}\cdot\mathbf{D}\cdot\left(\mathbf{p} - \mathbf{W}\cdot f^t\right)$$

$$\mathbf{f}^{t+1} = \mathbf{f}^t + \lambda\cdot\mathbf{W}^{\mathbf{T}}\cdot\left(\mathbf{p} - \mathbf{W}\cdot\mathbf{f}^t\right)$$

$$f_j^{(t+1)} = f_j^{(t)} + \frac{w_{ji}\left(p_i - \mathbf{w_i}\cdot\mathbf{f}\right)}{\sum_i w^2{}_{ij}}$$

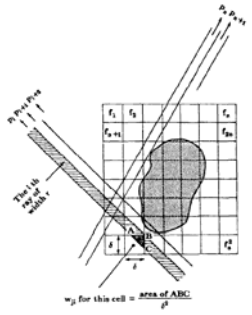
$$\lambda\,\mathbf{D_{jj}} = \frac{1}{\sum_i w^2{}_{ij}}$$

It is a form of scaled gradient.

SART statistical interpretation

$\mathbf{w_i} \cdot \mathbf{f} = p_i + v_i$ with v_i noise on pixel i.

We weight each pixel i-th by $\sum_j w^2_{ij}$



Through interative least squares

$$f^{t+1} = f^t + \lambda \cdot \mathbf{W}^{\mathrm{T}} \cdot \mathbf{D} \cdot \left(\mathbf{p} - \mathbf{W} \cdot f^t\right)$$

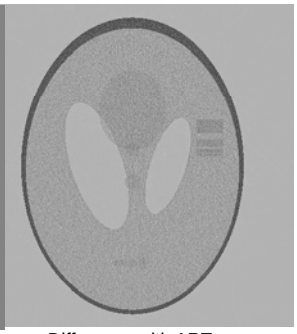
Maximum likelihood estimate of the $\Delta\mathbf{f_k}$, under the hypothesis that noise is Gaussian, zero mean.

SART allows filtering noise

Risults obtained from 300 images equally spaced over 300 degrees. Phantom.

Original
(noise is added to the original)

Difference with SART



Difference with ART

Maximum likelihood estimate -> filtering

Pros and Contra of SART and limited angle projection

Pro

- Easy incorporate information on each projection
- Easy management of tomography from a few projections
- Filtering is incorporated

Contra

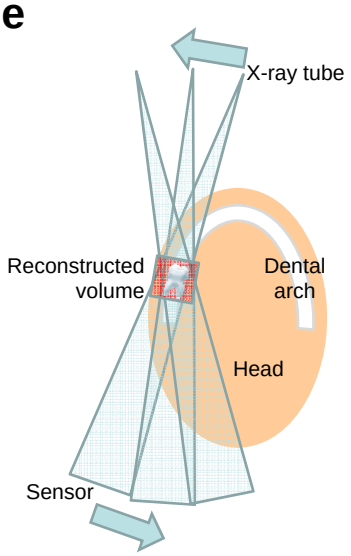
- Limited resolution of structures perpendicular to...
- **Huge memory resources that limited in the past its use.**
- **When limited volumes are reconstructed, truncation artifacts arise (also when using analytical solutions).**

Overview

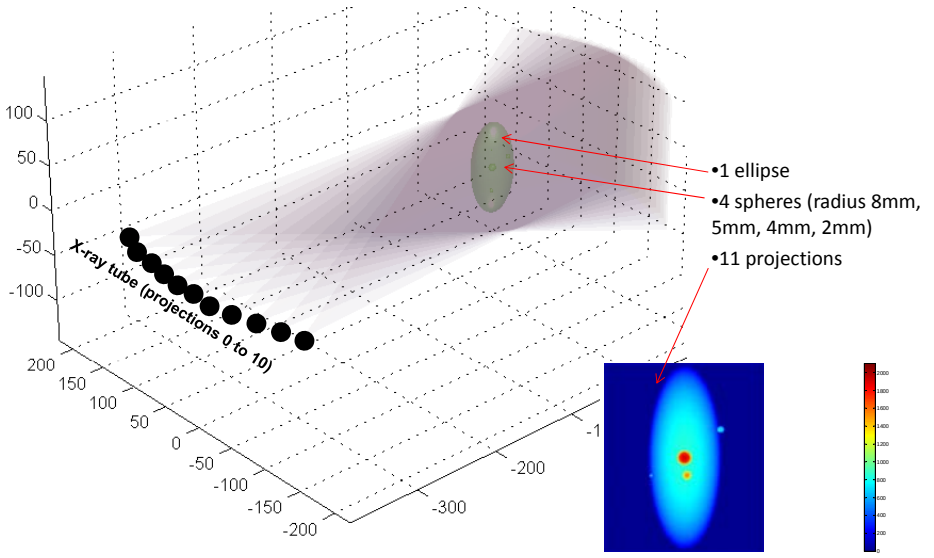
- Algebric reconstruction and tomosynthesis
- Limited angle tomography

Limited angle acquisition of a small volume

- Acquisition angle = -30°… +30°
- Acquisition time = 60s
- Xray duration = 26s (1,5s x 11)
- Reconstruction time =
  - 60s (OS-SART, 5 iterations)
  - 6s in CUDA
- Number of projections = 11
- Images size = 1536x562 pixel
- Pixel size 0,096mm.
- Volume = 5cm x 4cm x 10.5cm
- Volume dimension in voxels = 50 x 267 x 700
- Voxel dimension: 1mm x 0.15mm x 0.15mm



Simulations



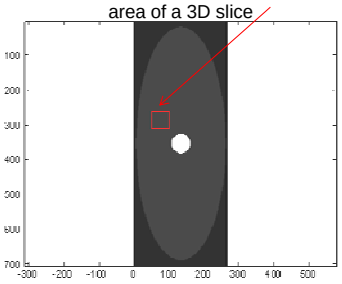
- 1 ellipse
- 4 spheres (radius 8mm, 5mm, 4mm, 2mm)
- 11 projections

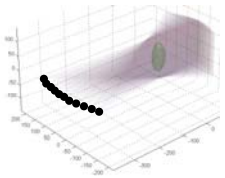
Approximation of W, W^T

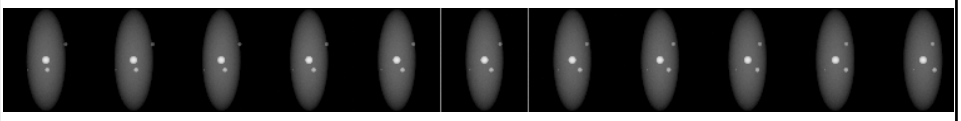
Images with noise (Poisson, max 10000 photons per pixel, added before log is computed)

Use of unmatched projectors (*Zeng, Gullberg, Unmatched Projector/Backprojector Pairs in an Iterative Reconstruction Algorithm, IEEE Trans. Med. Imag., 2000*):
- Voxel based backwards projection.
- Pixel based forward projection.

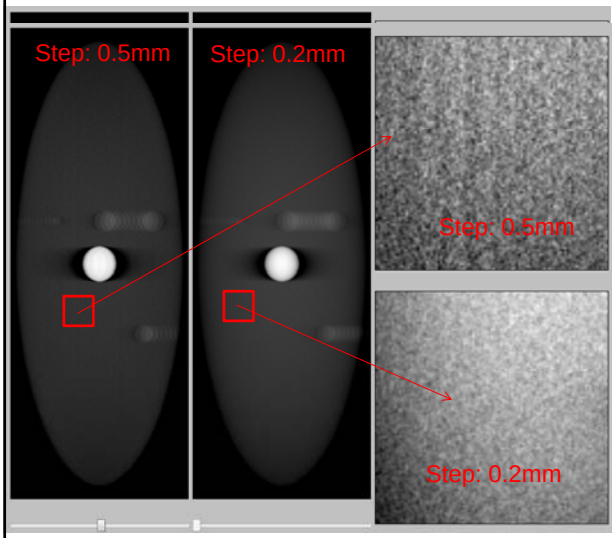
Evaluation of noise inside a uniform area of a 3D slice







Approximation of W and W^T with noise

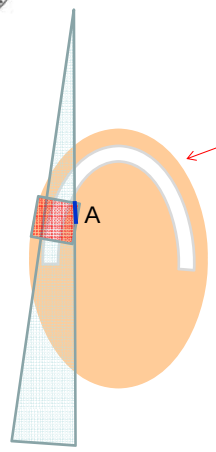


Sampling step forwards
Stampling step bakwards
Bilinear interpolation
Trilinear interpolation

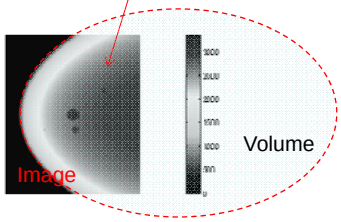
• More uniform appearance for shorter step.

• Trilinear interpolation (forward projection) / blinear interpolation (backwards proection) does not significantly improve the image quality in this case.

Truncation artifacts



- The abosption volume extends beyond the sensor field of view (local tomography - the head is larger than the reconstructed volume)
- White bands are generated in correspondence of short rays (ray A) crossing the volume.

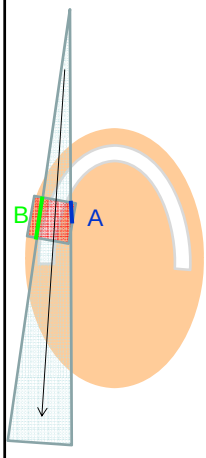


AN van Daatselaar SM Dunn, HJW Spoelder, DM Germans, L Renambot, HE Bal, PF van der Stelt, Feasibility of local CT of dental tissues, Dentomaxillofacial Radiology (2003) 32, 173–180.

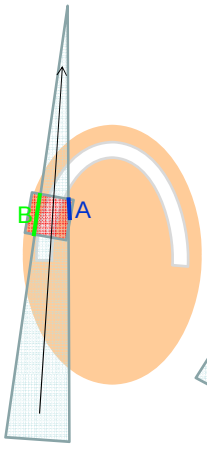
http://ais-lab.dsi.unimi.it

31/81

Truncation artifacts: generation

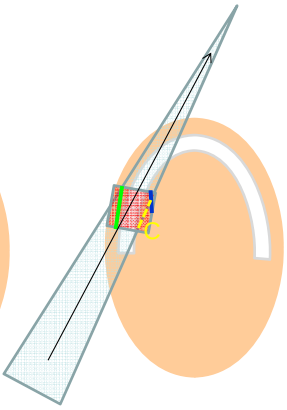


Absorption measured
Equal in A and B
A = B = 100.



Backprojection
Voxels on A: 100 / length(A)
Voxels on B: 20 / length(B)

A sees a larger absorption



Backprojection (other view)
Voxels on C: underestimated.

Black bands On C.
Correct reconstruction on B.

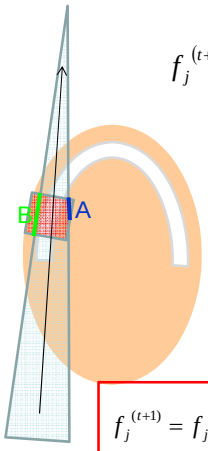
Slice 1 / 50

Slice 25 / 50

http://ais-lab.dsi.unimi.it

32/81

Truncation artifacts: correction



$$f_j^{(t+1)} = f_j^{(t)} + \lambda \frac{w_{ji}\left(\mathbf{w_i} \cdot \mathbf{f^{(t)}} - p_i\right)}{\sum_i w^2_{ij}}$$

$$p_i = \mathbf{w_i} \cdot \mathbf{f}$$

$$f_j^{(t+1)} = f_j^{(t)} + \lambda \left(\sum_j w^2_{ij}\right) w_{ji}(\mathbf{w_i} \cdot \mathbf{f}^{(t)} - p_i)$$

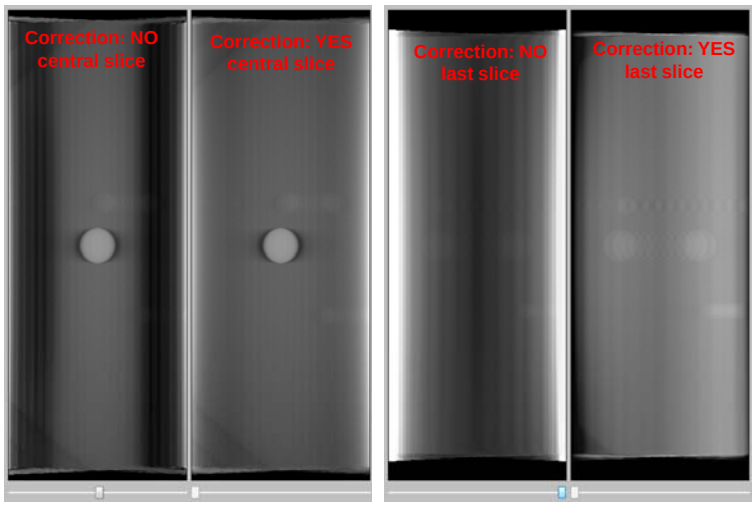
- For a short ray (A), $\sum_i w^2_{ij}$ is small;
- If absorbing structures external to the actual volume are present along the short ray, the absorption coefficients inside the actual volume are overestimated.
- To avoid this bias, short rays should be considered as "less important" in the recontruction process.
- In the least squares estimate this is equivalent to adding a weight proportional to: $\sum_i w^2_{ij}$

This is a very different weighting scheme with respect to original SART formulation:

$$f_j^{(t+1)} = f_j^{(t)} + \lambda \frac{w_{ji}(\mathbf{w_i} \cdot \mathbf{f} - p_i)}{\sum_i w^2_{ij}}$$

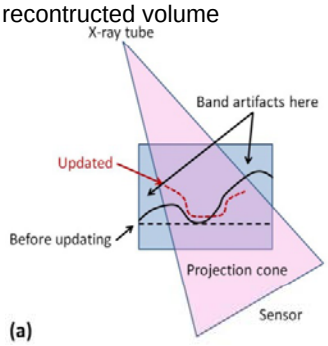
Different scaling matrix.

Truncation artifact: results



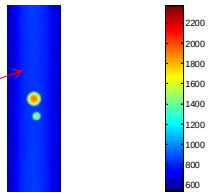
Band artifacts

Band shape artifacts due to the fact that the cone associated to each projection only partially crosses the recontructed volume

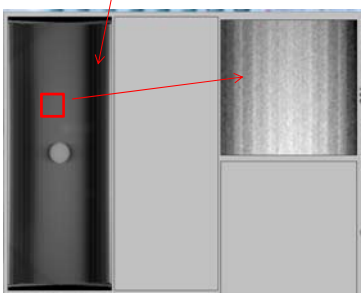


Spurious discontinuities at the projections boundary do arise in the reconstructed volume.

• Projections are taken using a narrow sensor;



•Band artifact are visible in the reconstructed volume.



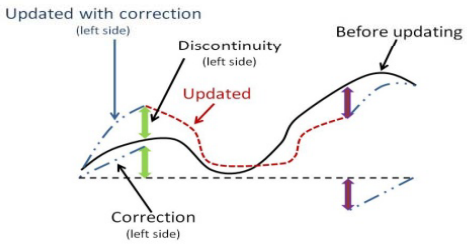
Band artifacts: correction

$$\boldsymbol{\mu}^{k+1} = \boldsymbol{\mu}^{k} + \lambda \cdot \mathbf{W}^{T} \cdot \mathbf{D} \cdot \left(\mathbf{p} - \mathbf{W} \cdot \boldsymbol{\mu}^{k}\right) \quad SART$$

Local equalization (inspired by W. Zbijewski, F. J. Beekman, 2004)

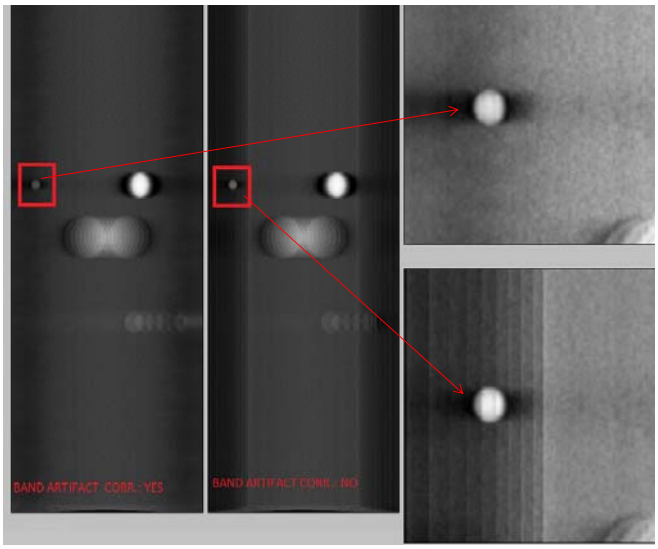
Linear correction of the absorption outside the projecting cone:
$x_{ijk} \leftarrow x_{ijk} + \Delta_k \cdot (1 - \delta_{ijk}/\Delta_k)$

δ_{ijk}/Δ_k ration between 1) the distance of voxel ijk (outside the projecting cone) and the volume border and 2) between the projecting cone and the volume border



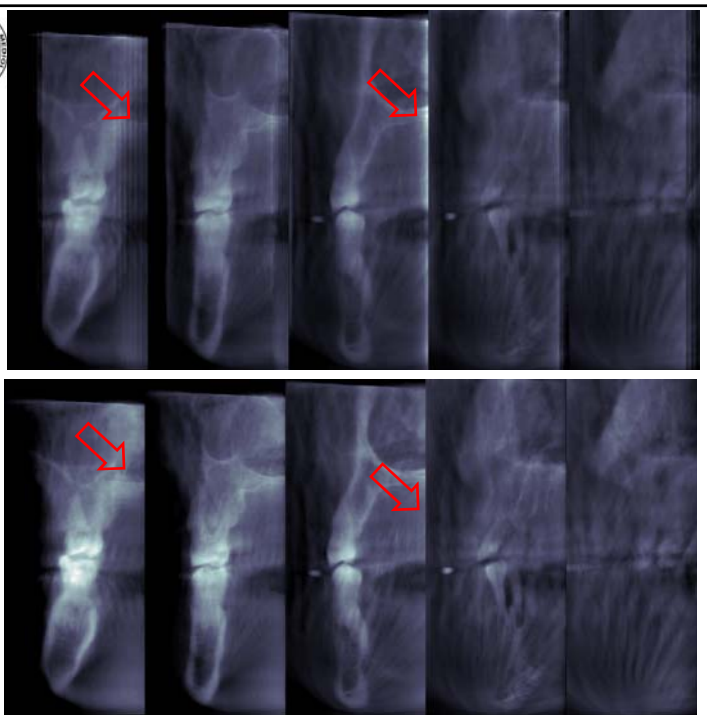
(b)

Band artifacts: results



Some results

Notice the increase in the dynamics



No artifacts correction

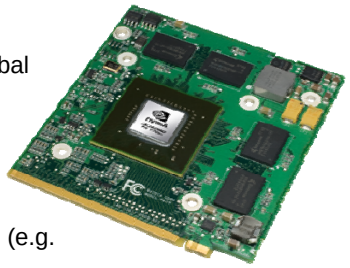
Artifacts corrected

CUDA implementation

CUDA extends the C language and it furnishes a set of functions that permits the usage of the GPU as a parallel multiprocessor.

Many computational cores (on the GPU) execute the same code (CUDA kernels) on different data (SIMD = Single Instruction Multiple Data architecture).

Each thread (and the corresponding core) has its own (limited) register and memory. Shared memory and global memory are also present (higher latency).



To maximize the efficiency:

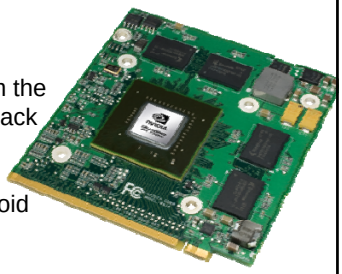
- Avoid CPU / GPU data transfer;
- Data alignment for optimal access;
- Optimize the resource usage for each kernel (e.g. number of registers).

CUDA implementation

Quadro FX 770M – On board memory: 512MByte (2008).
Low cost GPU (less than 60$) for mobile workstations.

Volume data (40MByte = 700 x 267 x 50 float variable) can be entirely contained in the global board memory. 3D texture is used to sample data from the volume (HW trilinear interpolation) during forward projection.

All images (4MByte per projection = 1538 x 592 float variables) can be entirely contained in the on the global board memory. 2D textures are used to sample data from the error image (**p** – **W f**) (HW bilinear interpolation) during back projection.



All the computation is carried out done on the GPU to avoid GPU/CPU data exchange.

CUDA Kernel 1: forward projection and error computing – 1 Thread per ray tracing (per pixel).
CUDA Kernel 2: back projection – 1 Thread per voxel.
CUDA Kernel 3: band artifact correction – 1 Thread per band.

CUDA implementation

CPU (Intel Centrino Duo @ 3.06GHz, RAM 4M)

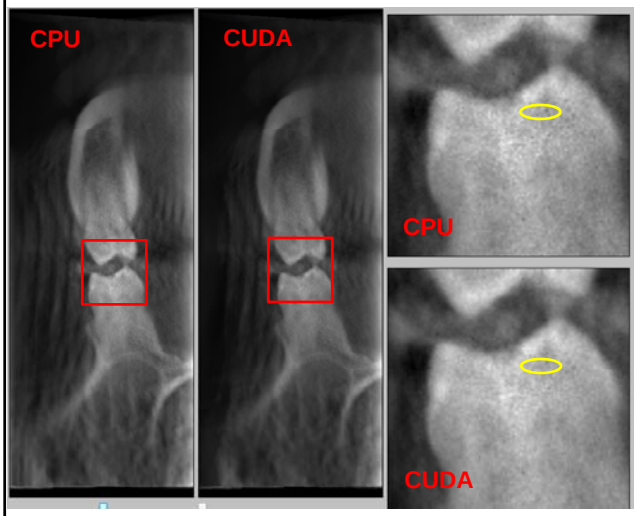
- 11 Projections 1538x562;
- Forward projection: NN interpolation;
- Backprojection: NN interpolation;
- Trunctation artifact correction: YES
- Band artifact correction: YES
- Reconstruction time: 60s.

GPU (Quadro FX 770M, RAM 512M - Low cost for mobile)

- 11 Projections 1538x562;
- Forward projection: TI interpolation;
- Backprojection: BL interpolation;
- Trunctation artifact correction: YES
- Band artifact correction: YES (simpler version – Work In Progress)
- Reconstruction time: 6s.

41/81

CUDA vs. CPU



Reconstructuion time:
60s CPU vs. 6s CUDA.
Sampling step 0.5

Quality:
Sligthly less noise with CUDA (trilinear interpolation and bilinear interpolation are free)

http://ais-lab.dsi.unimi.it

42/81

How to further improve?

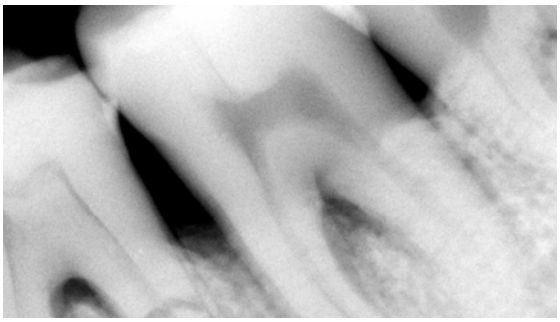
- Metal (streak) artifacts
- Correction for patient movement
- Improving CUDA efficiency on new boards.
- Increase resolution

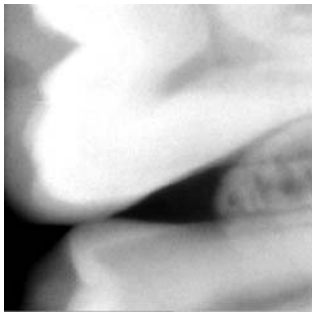
Overview

- Algebric reconstruction and tomosynthesis
- Limited angle tomography

Denoising of a radiographic image

Typical images (digital radiographies @12bpp, 1.5Mpixels – negative images), Poisson noise.





45/79

Can we solve any problem?

$g = Af + noise \Rightarrow g = g_o + noise$

Going in the frequency domain: $G(w) = G_o(w) + N(w) = A(w)F(w) + N(w)$

If we suppose that $A(w) \neq 0$ outside the support of $N(w)$, the problem is ill-posed.

And if we solve the linear system above, in Fourier space, the following nasty things happen:

$G(w) - N(w) = A(w)F(w)$ ➔

No solution exists as noise N(w) cannot be reproduced by F(w).

Moreover the solution is not unique as any function defined outside the support of F(w) can be added.

As N(w) decreases to zero with a transition bandwidth of finite width, noise components tend to be largely amplified, that produce spurious oscillations.

In case of filtering A = I, the restaured image would be equal to the noisy one.

46/79

Statistical formulation of image restauration

Measuring an image g_{noisy} taken from an object, f, (e.g. perspective image, transmission image): g = Af + ***noise*** → f?

(A may or may not contain the blurring of the measuring tool, point spread function)

Each pixel is considered an independent process (white noise). For each pixel therefore, we want to maximize:

$p(g_{noisy}; f)$

Being the pixels independent, the total probability can be written in terms of product of independent probabilities:

$$L(g_{noisy,i}; f_i) = \prod_{i=1}^{N} p(g_{noisy,i}; f_i)$$

L(.) is called likelihood function.

Determine $\{f_i\}$ such that L is maximized. Negative log-likelihood is usually considered to deal with sums:

$$J_0(g_{n,i}; f_i) = -\sum_{i=1}^{N} \ln(p(g_{n,i}; f_i))$$

47/79

Gaussian & Poisson cases

$Noise_i = \|Af_i - g_{ni}\|$

We know the statistical distribution of the noise -> we now the statistical distribution of the second term.

In case of (normalized) Gaussian distribution therefore:

$$J_0 = -\ln(p(g_n, f)) = -\ln\left\{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^m e^{-\frac{1}{2\sigma^2}\|g_n - Af\|^2}\right\} = \text{cost} + K \| g_n - Af\|^2$$

The minimization of this function leads to:

$A^TAf = A^Tg_n$ **That is the least squares solution**.

The same line of reasoning for Poisson noise leads to the KL divergence:

$$J_0 = -\ln(p(g_n, f)) = D_{KL}(g_n, Af) = \sum_i g_{n,i} \ln\left(\frac{g_{n,i}}{Af} + Af_i - g_{n,i}\right)$$

48/79

Observations

Solution has a single global minimum in both discrete and continuous cases (Shepp and Vardi, 1982).

Solution has many local minima in both Poisson and Gaussian cases, in the continuous case (cf. Malthei, 1993 for Poisson case).

Is the solution interesting? No, as it also incorporates noise. Therefore, semi-convergence has been proposed (Bertero and Boccacci, 1998).

The Bayesian framework

We assume that the object f is a realization of the "abstract" object F that can be characterized statistically.

The probability $p(g_n| f)$ becomes a conditional probability:

$J_0 = p(g_n| f = f^*)$

Under this condition, the probability of measuring g_n can be written as the product of the probabilities:

$p(g_n, f) = p(g_n \mid f)\, p_f$

As we are interested in determining f, we have to write the conditional probability of f given g_n through the Bayes theorem:

$$p(f \mid g_n) = \frac{p(g_n \mid f)\, p_f}{p_{g_n}} = L(g_n; f)\frac{p_f}{p_{g_n}}$$

MAP estimate

$p(f \mid g_n) = \dfrac{p(g_n \mid f)\,p_f}{p_{g_n}} = L(g_n; f)\dfrac{p_f}{p_{g_n}}$

We can find f here by maximizing p(f | g), with respect to f. **MAP estimate.** We explicitly observe that the marginal distribution of p_{gn} is not dependent on f.

Again, it is more efficient to go to the logarithms:

f = argmax{p(f | g_n)} =

$$\arg\min_f \left\{-\ln\left(p(g_n \mid f)\,p_f\right)\right\} = \arg\min_f \left\{-\ln\left(p(g_n \mid f)\right) + \ln\left(p_f\right)\right\}$$

Likelihood = adherence to the data

A-priori

51/79

Tikhonov regularization

$$f = \arg\min_f \left\{-\ln\left(p(g_n \mid f)\,p_f\right)\right\} = \arg\min_f \left\{-\ln\left(p(g_n \mid f)\right) + \ln\left(p_f\right)\right\}$$

Gaussian noise model
Squared shape for the log a-priori term

$$K\sum_i \left\|g_{n,i} - Af_i\right\|^2$$

Squared error

$\| \lambda Pf \|^2$

$$f = \arg\min_f \sum_i \|g_n - Af\|^2 + \lambda\|Pf\|^2$$

It is a quadratic cost function

This leads to the classical Tikhonov formulation

$$f = A^T A g_n - A^T A f + \lambda P^T P f$$

Poggio and Girosi, 1990

52/79

What happens when noise is Poisson?

$$f = \arg\min_f -\{\ln(p(g_n \mid f)p_f)\} = \arg\min_f -\{\ln(p(g_n \mid f)) + \ln(p_f)\}$$

Poisson noise model
Squared shape for the a-priori term

$$\sum_i g_{n,i} \ln\left(\frac{g_{n,i}}{Af} + Af_i - g_{n,i}\right)$$

KL discrepancy

$\| \lambda Pf \|^2$

$$f = \arg\min_f \sum_i g_{n,i} \ln\left(\frac{g_{n,i}}{Af} + Af_i - g_{n,i}\right) + \lambda \|Pf\|^2$$

Regularization

No analytical solution -> non-linear optimization

53/79

Non-quadratic a-priori

$$f = \arg\min_f -\{\ln(p(g_n \mid f)p_f)\} = \arg\min_f -\{\ln(p(g_n \mid f)) + \ln(p_f)\}$$

Gaussian noise model

$$\sum_i \|g_{n,i} - Af_i\|^2$$

Total variation

$$\sum_i \sqrt{\sum_{w=1}^{W} p_{f_w}^2}$$

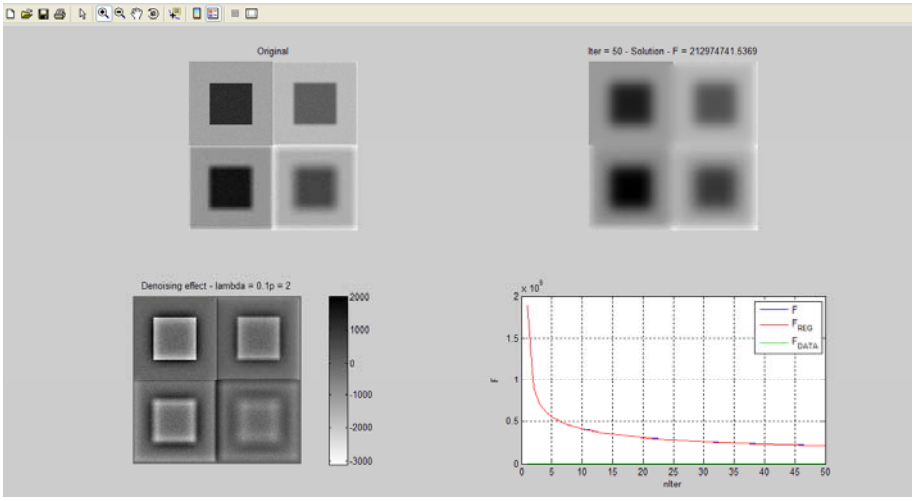
A-priori is a gradient and it is expressed in l_2 norm

$$p_f = \sum_i \sqrt{\left(f_{x,i}^2 + f_{y,i}^2 + f_{z,i}^2\right)}$$

$$f = \arg\min_f \sum_i \left(\|g_n - Af\|^2 + \lambda \sqrt{\sum_1^W p_{f_w}^2} \right)$$

54/79

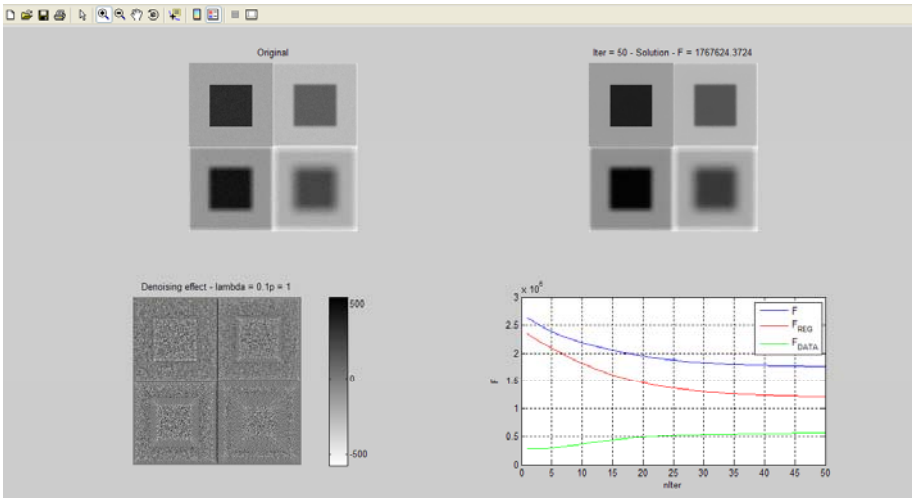
Why total variation - simulations



Edge smoothing effect with Tikhonov-like regularization
Poisson noise model – $\lambda = 0.5$
P is the gradient operator

55/79

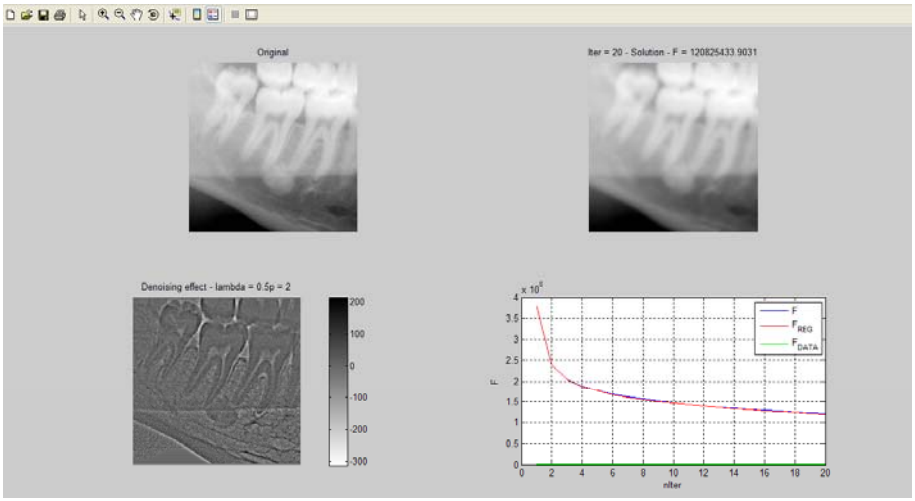
Why total variation - simulations



No appreciable edge smoothing with total variation
Poisson noise model - $\lambda = 0.5$
P is the gradient operator

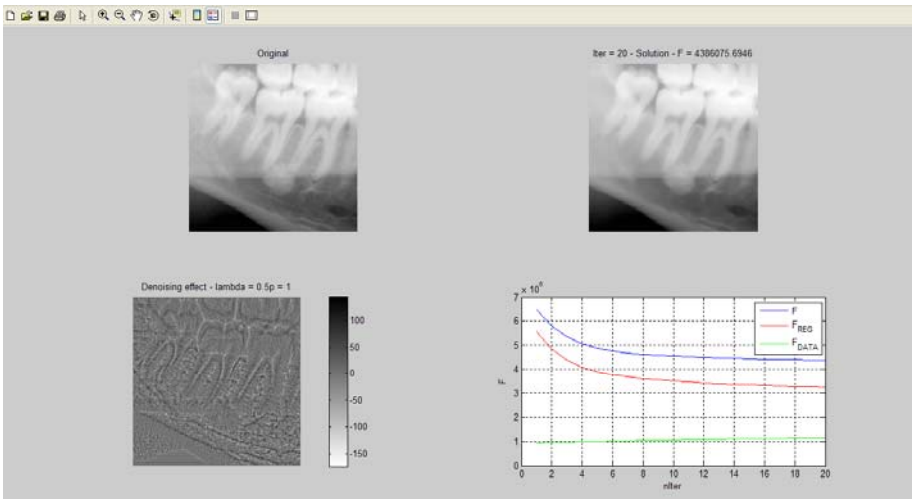
56/79

Why total variation – panoramic images



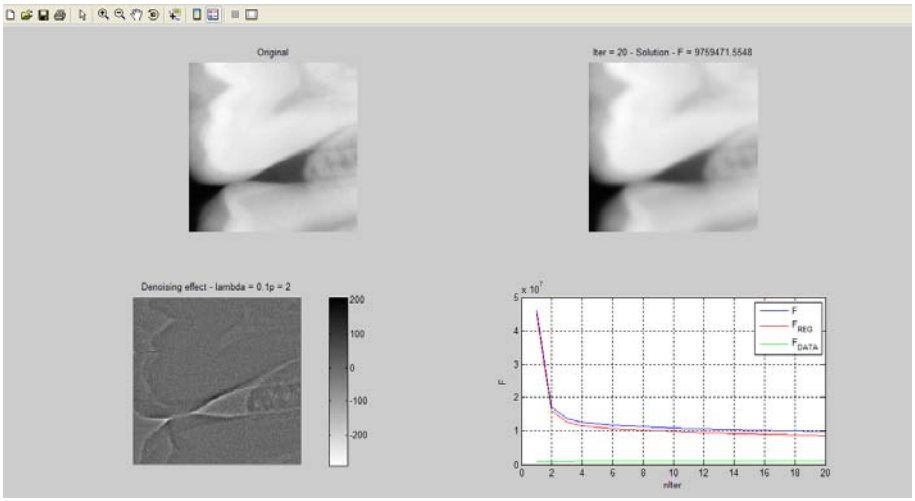
Edge smoothing effect with Tikhonov-like regularization
Poisson noise model - $\lambda = 0.5$
P is the gradient operator

Why total variation – panoramic images



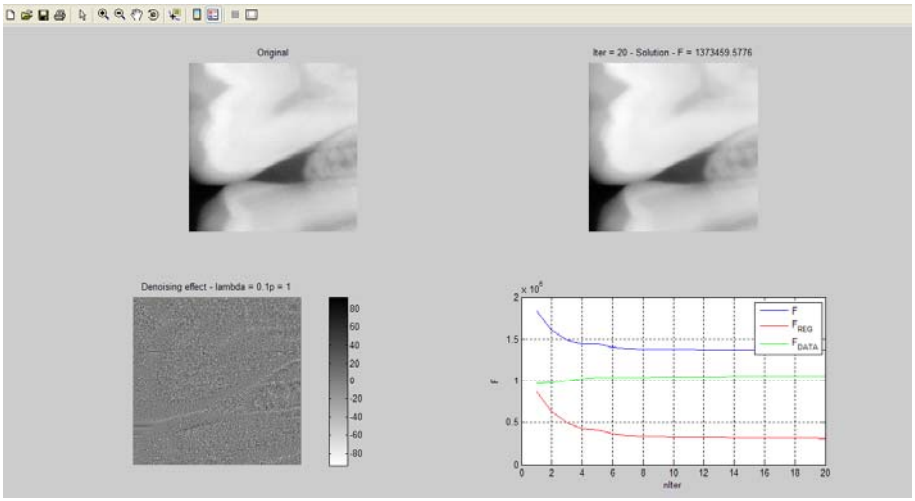
No appreciable edge smoothing with total variation
Poisson noise model - $\lambda = 0.5$
P is the gradient operator

Why total variation- endo-oral images



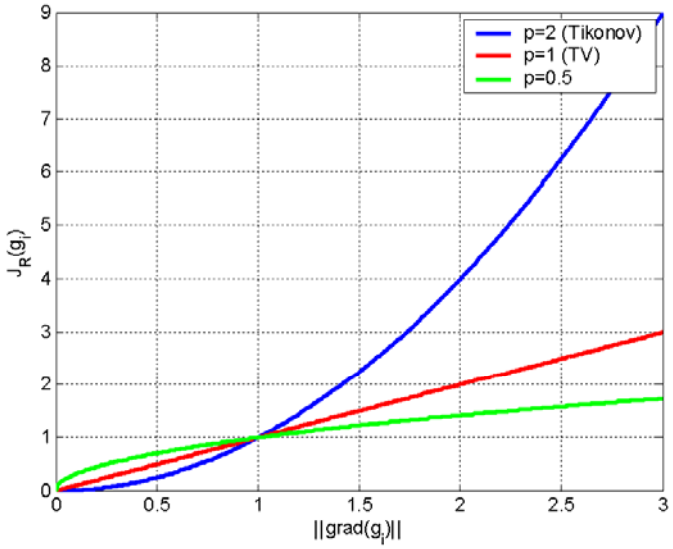
Edge smoothing effect with Tikhonov-like regularization
Poisson noise model - $\lambda = 0.1$
P is the gradient operator

Why total variation – endo-oral images



No appreciable edge smoothing with total variation
Poisson noise model - $\lambda = 0.1$
P is the gradient operator

Cost introduced by the regularzation term



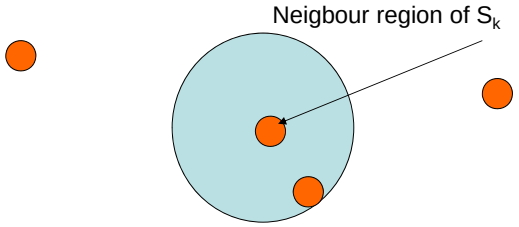
Cost increases quadratically with the local gradient in Tikhonov

61/79

A-priori

We can insert in the a-priori term all the desirable characteristic of the image: local smoothness, edges, piece-wise constancy,….

The idea of defining a neighboring system is a natural one:



Images have a natural neighbouring system: the pixels structure. We want to consider the local properties of the image considering neighboring pixels (in particular differential properties - our vision system is particularly tuning to gradients both spatial and temporal). Ideas have been borrowed from physics.

62/79

Neighboring System

Let P be the set of pixels of the image: $P = \{p_1, p_2, \ldots p_P\}$

The neighboring system defined over P, S, is defined as $H = \{\mathcal{N}_p \mid p, \forall p \in P\}$, that has the following properties:

An element is not a neighbour of itself: $p_k \notin \mathcal{N}_{pk}$

Mutuality of the neighboring relationship: $p_k \in \mathcal{N}_{pj} \leftrightarrow p_j \in \mathcal{N}_{pk}$

(S, P) constitute a graph where P contains the nodes of the graph and S the the links.

Depending on the distance from p, different neighboring systems can be defined:

	o	
o	x	o
	o	

First order neighboring System
4-neighboring System

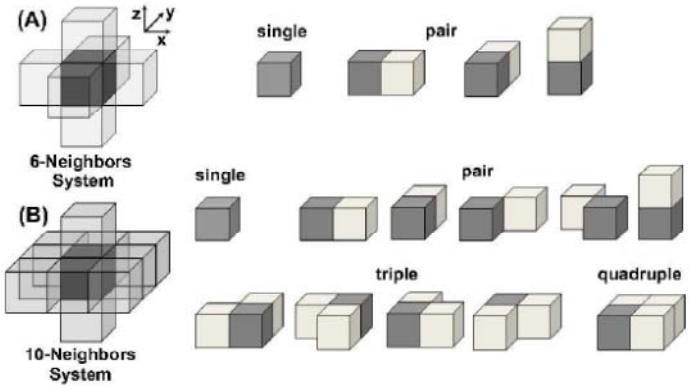
o	o	o
o	x	o
o	o	o

Second order neighboring System
8-neighboring System

63/79

Clique

Borrowed from physics.



A clique C, for (S, P), is defined as a subset of sites in S.

W can consider ordered sets of voxels, that are connected to p through S.

Types of cliques: single-site, pairs of neighboring sites, triples of neighboring sites,… up to the cardinality of $\mathcal{N}_p$

64/79

Markov Random Field

Given a neighbouring system, S, and a set of pixels, P, we can define a set of random values, $\{f_k(p)\}$ for each element defined by S, that is in $\mathcal{N}_p$. Therefore we define a **random field** , $\mathcal{F}$, over S:

$\mathcal{F}(\mathcal{N}_p) = \{f_k(m) \mid m \in \mathcal{N}_p\} \; \forall p$

Under the Markovian hypotheses:

1) $P(f(p)) \geq 0 \; \forall p$ Positivity

2) $P(f(p) \mid g(P-\{p\}) = P(f(p) \mid g(\mathcal{N}_p)\}$ Markovianity

2 expresses the fact that the probability of p assuming a certain value f (e.g. a certain gradient), is the same considering all the pixel of P but p, or only the neighbor pixels, that is the value of f depends only on the gray value of the pixels in $\mathcal{N}_p$.

The random field $\mathcal{F}$ is named **Markov Random Field**.

65/79

Energy in a Markov Random Field

A "potential" function, $\phi(f)$, can be defined for a MRF. This is a scalar value that is a function of the random value associated to the pixels for all the possible elements of a clique:

$$\phi_c(f) = \sum_{j \in c} f(p_j)$$

If we consider all the possible cliques defined for each element p, we can define a potential energy function associated to the MRF:

$$U(f) = \sum_{c \in C} \phi_c(f)$$

The higher is the potential energy, the lower is the probability that the set of random values of the elements of the cliques is realized, that is the higher is the penalization for the associated configuration.

66/79

Gibbs prior

If we consider all the possible cliques defined for each element p, we can define a potential energy function associated to the MRF:

$$U(\mathbf{f}) = \sum_{c \in C} \phi_c(\mathbf{f})$$

The higher is the potential energy, the lower is the probability that the set of random values of the elements of the cliques is realized, that is the higher is the penalization for the associated configuration.

This is well captured by the Gibbs distribution, that describes the probability of a certain configuration to occur. It is a function exponentially decreasing of U:

$$P(\mathbf{f}) = \frac{1}{Z} e^{\left\{-\frac{1}{\beta} U(\mathbf{f})\right\}}$$

P(f) is a Gibbs random field, Hammersley-Clifford theorem (1971). β regulates the decrease in probability and it is associated with temperature in physics. Z is a normalization constant. NB to define Gibbs random fields, P(f) > 0, P(f) → 0 U(f) → ∞: there are not configurations with 0 probability.

67/79

Gibbs priors and Regularization

$$\arg\min_f -\{\ln(p(g_n \mid f)p_f)\} = \arg\min_f -\{\ln(p(g_n \mid f)) + \ln(p_f)\}$$

Likelihood = adherence to the data

A-priori

Gaussian $K(\sigma)\sum_i \left\| g_{n,i} - Af_i \right\|^2$

Poisson $\sum_i g_{n,i} \ln\left(\frac{g_{n,i}}{Af} + Af_i - g_{n,i}\right)$

$$-\ln\left\{\frac{1}{Z} e^{\left\{-\frac{1}{\beta} U(\mathbf{f})\right\}}\right\}$$

$$J(f) = J_o(f) + \lambda J_R(f)$$

$J_R(f) = U(f)$

68/79

Role of λ

$$K(\sigma)\sum_i \left\| g_{n,i} - Af_i \right\|^2 \qquad -\ln\left\{ \frac{1}{Z} e^{\left\{ -\frac{1}{\beta} U(\mathbf{f}) \right\}} \right\}$$

$$\boxed{J(f) = J_o(f) + \lambda J_R(f)}$$

λ incorporates different elements here:
- the standard deviation of the noise in the likelihood
- the "temperature", that is the decrease in the energy of the configurations with their cost
- the normalized constant Z.

λ has been investigated in the classical regularization theory (Engl et al., 1996), but not as deep in the Bayesian framework ➔ λ is set experimentally through cross-validation.

69/79

Choice of the Gibbs priors

We choosed $\| \lambda Pf \|^2$ as a quadratic functional, but not specified P.

P is ofted chosen as a smoothing operator. The rationale is that the noise added to the image is often white (both Gaussian and Poisson) over the image as there is no correlation between adjacent pixels. Therefore its spatial content is unform and with a larger bandwidth that the signal.

As a smoothing operator P is often a differential operator, which penalizes edges.

$$J_R(\mathbf{f}) = \sum_{c \in C} \phi_c(\mathrm{d}^k{}_c \mathbf{f})$$

k is the order of the derivative
ϕ_c can be l_2 norm (total variation), squared (Tikhonov)

k = 2 difference of gradients → piecewise linear areas.
k = 3 difference of Hessian → piecewise squared.
Neighbor of order higher than 2.

70/79

Quadratic Priors with k = 0

k = 0 – No derivative, the same gray level – single site cliques.

$$J_R(\mathbf{f}) = \sum_{c \in C} \phi(\mathrm{d}^{\mathrm{k}}{}_{\mathrm{c}}\mathbf{f}) = \sum_{c \in C} (\mathrm{d}^{0}{}_{\mathrm{c}}\mathbf{f})^2 = \sum_{p \in P} \mathbf{f}(\mathrm{p})^2$$

It has been applied to both Poisson and Gaussian noise models

Reduces bright spots and biases the solution to low intensity values.

71/79

Quadratic Priors with k = 1

k = 1 – First order derivatives – pair-sites cliques.

$$J_R(\mathbf{f}) = \sum_{c \in C} \phi(\mathrm{d}^{1}{}_{\mathrm{c}}\mathbf{f}) = \sum_{p \in P} \sum_{\mathrm{m} \in \mathcal{N}_p} \phi(\mathrm{d}^{1}{}_{\mathrm{c}}\mathbf{f})^2 = \sum_{p \in P} \sum_{\mathrm{m} \in \mathcal{N}_p} \phi\left(\frac{\mathrm{f(p) - f(m)}}{\mathrm{d(p,m)}} \right)$$

d(p,m) takes into account anisotropies in computing the distance.

If we consider $\phi(.)$ a squared function, we have another form of Tikhonov regularization:

$$J_R(\mathbf{f}) = \sum_{p \in P} \sum_{\mathrm{m} \in \mathcal{N}_p} \left(\frac{\mathrm{f(p) - f(m)}}{\mathrm{d(p,m)}} \right)^2 \qquad \| \mathrm{Pf} \|^2$$

72/79

Quadratic Priors with k = 1

k = 1 – First order derivatives – pair-sites cliques.

$$J_R(\mathbf{f}) = \sum_{p \in P} \sum_{m \in \mathcal{N}_p} \left(\frac{f(p) - f(m)}{d(p,m)} \right)^2$$

If we consider $\phi(.)$ a squared function, we have another form of Tikhonov regularization:

$|| Pf||^2$

P is the convolution with the Laplacian operator:

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

First order neighboring System
4-neighboring System

$$\begin{bmatrix} -\frac{\sqrt{2}}{2} & -1 & -\frac{\sqrt{2}}{2} \\ -1 & 4+2\sqrt{2} & -1 \\ -\frac{\sqrt{2}}{2} & -1 & -\frac{\sqrt{2}}{2} \end{bmatrix}$$

Second order neighboring System
8-neighboring System

73/79

Non-quadratic potential functions, k = 1

Quadratic functions priors imposes smoothness everywhere. Large true gradients of the solution are therefore penalized → smoothing sharp edges.

In imaging objects tend to be piecewise smooth, but different pieces of objects are separated by more or less sharp edges. We want to smooth inside the object but not the edge. A parallel worthwhile to be investigated is with anisotropic diffusion (Koenderink, 1987; Perona&Malik, 1990).

We search different potential functions (Geman&McClure, 85; Charbonnier et al., 1994, 1997; Hebert&Lehay, 1989).

74/79

Non-quadratic potentials (Charbonier et al., 1997)

1. $\phi(t) \geq = 0 \quad \forall t \qquad \phi(0) = 0$ — Derives from the definition of potential
2. $\Phi(t) \geq = 0 \quad \forall t$
3. $\phi(t) = \phi(-t)$ — Positive and negative gradients are equally considered
4. $\phi(t) \in C^1$ — This is to avoid instability.

Up to now quadratic potentials are OK

5. $\frac{\varphi'(t)}{2t}$ — The potential increase rate should decrease with t.
6. $\lim_{t\to\infty} \frac{\varphi'(t)}{2t} = 0$ — The potential increase rate should decrease for all t (at least for large values of t)
7. $\lim_{t\to 0} \frac{\varphi'(t)}{2t} = \cos t > 0$ — The potential increases at least linearly for t = 0.

75/79

Few non-quadratic potentials (Vicedomini 2008)

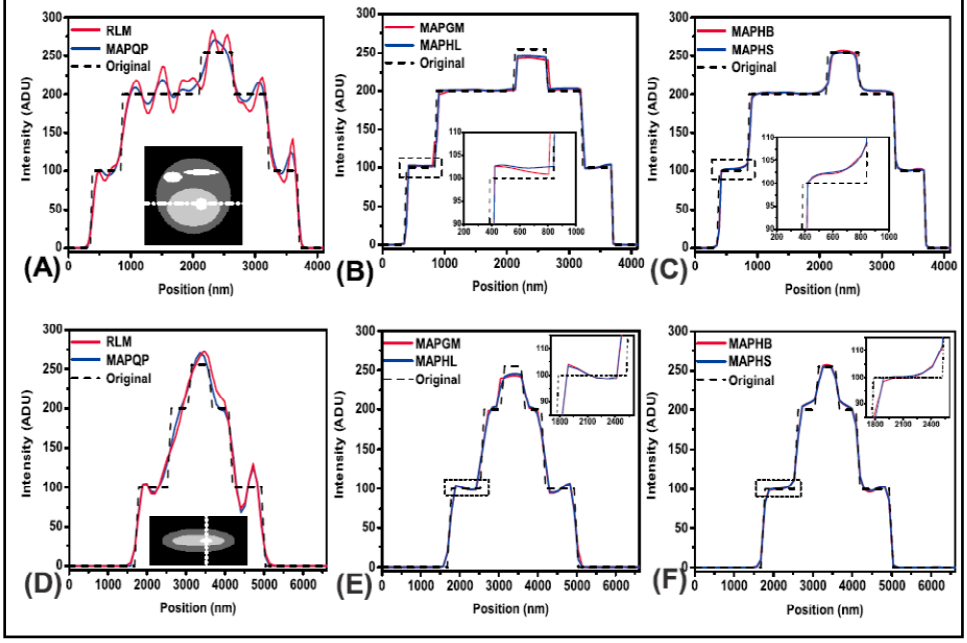
Regularization name	Potential function	Expression of $\varphi(t)$	Expression of $\psi(t) = \varphi'(t)/2t$	Convex
Quadratic-Potential	φ_{QP}	t^2	1	yes
Geman-McClure	φ_{GM}	$\frac{t^2}{1+t^2}$	$\frac{1}{(1+t^2)^2}$	no
Hebert-Leahy	φ_{HL}	$\log(1+t^2)$	$\frac{1}{1+t^2}$	no
Huber	φ_{HB}	$\begin{cases} t^2, & \lvert t\rvert \leq 1 \\ 2\lvert t\rvert - 1, & \lvert t\rvert > 1 \end{cases}$	$\begin{cases} 1, & \lvert t\rvert \leq 1 \\ 1/\lvert t\rvert, & \lvert t\rvert > 1 \end{cases}$	yes
Hyper-Surface	φ_{HS}	$2\sqrt{1+t^2} - 2$	$\frac{1}{\sqrt{1+t^2}}$	yes

Asymptotic linear behavior

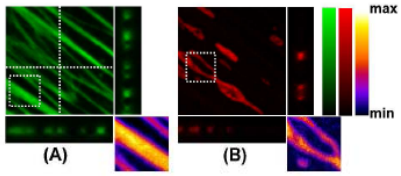
Asymptotic log-like behavior

76/79

Results

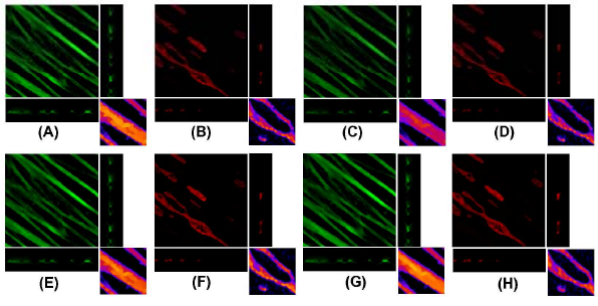


Application to optical microscopy



A) Raw image of actin structures in a cell.
B) Raw image of mito-chondria structures.

Geman & McClure regularization (A,B)
Hebert & Leahy regularization (C,D),
Huber regularization (E,F)
Hyper-surface regularization (G,H).
(From Vicedomini, 2008)



3/79

Summary

MAP estimate can be seen as a statistical version of regularization.

The regularization term can be derived from the potential energy associated to an adequate neighbor system defined over the object (e.g. over the image).

Under this hypothesis the value assumed by the elements of the object to be reconstructed (e.g. restored or filtered image) represent a MRF.

Different neighbor systems and different potential functions allow defining different properties of the object.

For quadratic potential functions, Tikhonov regularizer are derived.

The discrepancy term for the data represents the likelihood and can accommodate different statistical models: Poison, Gaussian or even mixture models.

79/79

Issues to be investigated

- Is the a-priori term really suitable? There is a difference between the metric used in evaluating the distance between the image and the projection and the gradient. Which is the best metric?
- How to set the regularization parameter? We have introduced a generalized discrepancy principle, but can we do something better?
- Optimization and parallelization of the code using CUDA. Real-time reconstruction and filtering.

80/79

References

I. Frosio, N.A. Borghese, (2011) Optimized limited angle tomography, in preparation for *IEEE Trans. on Medical Imaging*.

I. Frosio, N.A. Borghese (2011) Semi-automatic geometric calibration of an ortopantomograph, Proc. CARS 2011

M. Lucchese, I. Frosio, NA Borghese (2010) Optimized Algebric Local Tomography, Proc. ICIAP 2010.

I. Frosio, N. A. Borghese, (2009) Statistical Based Impulsive Noise Removal in Digital Radiography," *IEEE Transactions on Medical Imaging,* Vol.28, No.1, Jan. 2009, pp.3-16.

M. Lucchese and N.A. Borghese, (2009) Denoising of Digital Radiographic Images with Automatic Regularization Based on Total Variation, Image Analysis and Processing - ICIAP 2009, P.Foggia, C.Sansone, M.Vento (eds.), pp. 711-720, Lecture Notes on Computer Science. Elsevier, 2009.

I. Frosio, F. Pedersini, A. Pasini, D. Bianconi, N. A. Borghese, (2009), Algebraic reconstruction methods for GPU cone beam tomography, in Proc. CARS 2009, Berlin (Germany), June 2009.

Frosio I., Abati S., Borghese N.A. (2008) An Expectation Maximization Approach to impulsive Noise Removal in digital Radiography. *Int. J. Computer Assisted Radiology and Surgery.* Vol. 3, No. 1-2, June 2008, pp. 91-96.

Frosio I., Borghese N.A. (2009) Compression and smart coding of offset and gain maps for intraoral digital x-ray sensors, *Medical Physics*, Vol. 36, No. 2, Feb. 2009, pp. 464-79.

Frosio, N. A. Borghese, (2006) A New Real Time Filter for Local Exposure Correction in Panoramic Radiography, *Medical Physics*, Vol. 33, No. 9, Sep. 2006, pp. 3478-88 .

Frosio I., Ferrigno G., Borghese N.A. (2006). Enhancing Digital Cephalic Radiography with Mixture Models and Local Gamma Correction. *IEEE Trans. Medical Imaging.* Volume 25, Issue 1, Jan. 2006 Page(s):113 - 121.

Cerveri P., Forlani C., Borghese N.A., Ferrigno G. (2002), Distortion correction for X-ray image intensifiers: a comparison between local un-warping polynomials and adaptive neural networks. *Medical Physics,* August 2002, Vol. 29, pp. 1759-1771.

N.A. Borghese, E. Nanni, G. Rotondo. Dynamic Error Correction in Tomography. (2009). US Patent, Application number: 12/502,704, 14th July, 2009.

N.A. Borghese, I. Frosio, E. Nanni, G. Rinaldi, G. Rotondo (2009). Method and Apparatus for Radiographic Imaging. US Patent, Application number: 12/479,254, 5th June 2009.