Algebric Reconstruction and beyond in Tomography

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

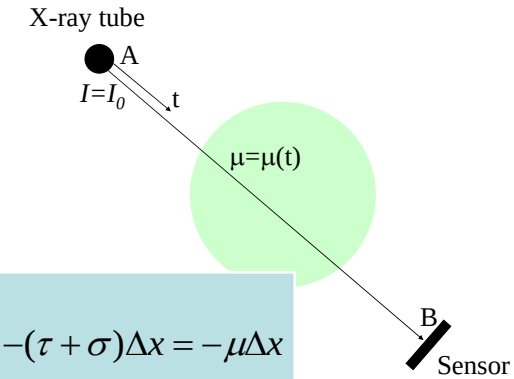
08.June.2011

Overview

- Algebric reconstruction
- Limited angle tomography
- Bayesian Restoration of X-ray images

Description of X-ray interaction

- An X-ray tube emitting monochromatic radiation (from 20 to 150 keV);
- An object absorbing X-rays (absorption + Compton effect);
- A sensor to measure the X-ray intensity.

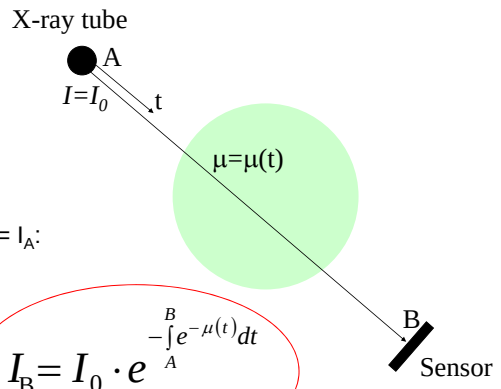


$$\frac{\Delta N}{N} = -(\tau + \sigma)\Delta x = -\mu \Delta x$$

At the limit for Δx -> 0

Tomography as a linear problem

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



With x = A and Δx = B, $I_0 = I_A$:

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

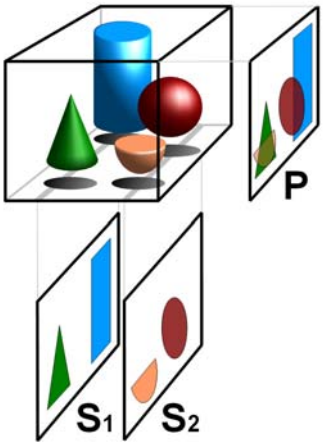
$$\ln\left(\frac{I_B}{I_0}\right) = -\int_A^B \mu dx$$

➔ De Beer's law

μ components

- The X-ray is not monochromatic, but has a (wide) spectrum.
- Absorption is not linear. Lower energy photons are absorbed by soft tissue leaving higher energy photons for bony tissue (Beam hardening).
- Overestimate in few regions (soft tissue) and underestimate in bony regions (underestimate).
- Compton effect does occur.
- Sensors with anti-scatter grid are usually employed.
- **For highly accuracy reconstruction system, in a refinement stage beam hardening could be taken into account.**

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)

CT machines

Skyview - Myray

Somaton Emotion - Siemens

Ricostruzione “continua” del volume

Go to Matlab visualization
SW: Tomografia\3D M\VFW

Slices from 11 images 40°, 60kV,10mA

Approccio analitico

- E' basato sulla trasformata di Radon:

$$P_\theta(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(x\cos\theta + y\sin\theta - t)\, dx\, dy.$$

- E sulla sua antritrasformata filtrata (caso fan-beam).

$$f(x, y) = \int_0^{\pi} Q_\theta(x\cos\theta + y\sin\theta)\, d\theta$$

$$Q_\theta(t) = \int_{-\infty}^{\infty} S_\theta(w)|w| e^{j2\pi wt}\, dw.$$

E viene generalizzato per le diverse geometrie: cone-beam, helicoidal.

Richiede che le immagini siano equi-spaziate. Cambiamento di base. Difficoltà a gestire le peculiarità locali (artefatti da metallo, movimento...).

Tomography – discrete case

- Object discretized into a set of voxels;
- Image discretized into a set of pixels;

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

For each measured X-ray intensity (for each pixel of the sensor and for each position), the group of X-rays that hit the pixel are considered together.

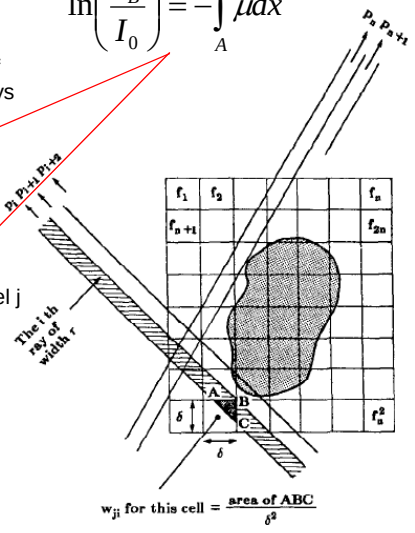
Through logarithms:

$\ln(I_B/I_0) = -\Sigma_j w_{ij} \cdot f_j$ →

- p_i measured value at pixel i.
- w_{ij} length of the path of the X-ray that crosses voxel j and hit pixel i at its center.
- f_j voxel absorption

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

A system of linear equations results.



Strategie di ricostruzione

$\mathbf{w}_i \mathbf{f} = p_i$ è un sistema di dimensioni enormi, molto sparso. La sparsità non è ancora stata ben esplorata e promette accelerazioni interessanti con schede GPU.

CT Cone beam.

512 immagini su 360 gradi, ciascuna di 512 x 512 pixel su 12bpp (8bpp) = 128MB
Volume di: 512 x 512 x 512 voxel = 128 M word (float variables).
N incognite = M misure (equazioni)

Volume di: 256 x 256 x 256 voxel = 16 M word (float variables).
N incognite << M misure (equazioni)

L'inversione diretta del sistema non è pensabile e algoritmi di ottimizzazione iterativa sono stati introdotti.

Algoritmi di ricostruzione algebrici (recentemente diventati popolari)

$p_i = \mathbf{w}_i \cdot \mathbf{f}$

Metodo delle proiezioni

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," Numer. Math., vol. 17, pp. 203-214, 1971

$$\begin{aligned} &w_{11}f_1 + w_{12}f_2 + w_{13}f_3 + \cdots + w_{1N}f_N = p_1 \\ &w_{21}f_1 + w_{22}f_2 + \quad + \cdots + w_{2N}f_N = p_2 \\ &\vdots \\ &w_{M1}f_1 + w_{M2}f_2 + \quad + \cdots + w_{MN}f_N = p_M . \end{aligned}$$

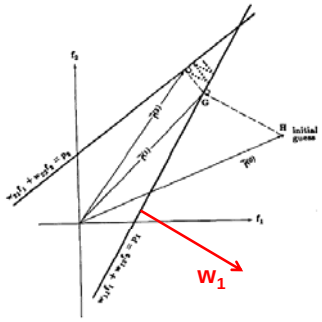
$p_i = \mathbf{w}_i \cdot \mathbf{f}$

Ciascuna misura rappresenta un iper-piano nello spazio N-dimensionale delle N-incognite.

Metodo delle proiezioni - 2

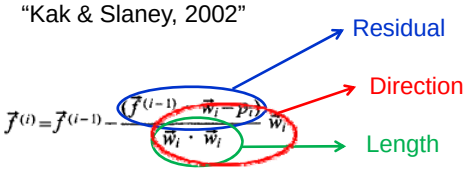
2 incognite -> problema piano -> 2 rette.

$$w_{11}f_1 + w_{12}f_2 = p_1$$
$$w_{21}f_1 + w_{22}f_2 = p_2.$$



First starting with an initial guess, projecting this initial guess on the first line, re-projecting the resulting point on the second line, and then projecting back onto the first line, and so forth. **If a unique solution exists, the iterations will** always converge to that point.

"Kak & Slaney, 2002"



Algebric Reconstruction Technique (ART)

ART iterates thorugh the followins steps:
1) Choose randomly projection image,
2) Choose randomly a pixel in the image,
3) Back project the ray in the volume.
4) Compute the current absorption along the ray.
5) Compute the error on the image (projection error), and smear the error back along the ray.

• It has been applied first in the domain of CAT.

Algebric representation of the update equations:

$$\mathbf{f}^{(i)} = \mathbf{f}^{(i-1)} - \alpha \frac{(\mathbf{f}_\mathbf{i}^{(i-1)} \cdot \mathbf{w}_\mathbf{i} - p_i)}{\| \mathbf{w}_\mathbf{i} \|^2} \mathbf{w}_i$$

It is an iterative optimization algorithm offirst order!!
$0 < \alpha < 1$ is a damping factor.

Problemi con ART

Il calcolo on-the-fly dei coefficienti w diventa proibitivo. Si sostituiscono con valori binari: attraversa non attravesa il voxel i-esimo.

(Damping piccolo -> tempi lunghi) R. Gordon, “A tutorial on ART (algebraic reconstruction techniques),” IEEE Trans. Nucl. Sci., vol. NS-21, pp. 78-93, 1974.

Supersampling (K. T. Smith, D. C. Salmon, and S. L. Wagner, “Practical and mathematical aspects of the problem of reconstructing objects from radiographs,” Bull. Amer. Math. Sot., vol. 83, pp. 1227-1270, 1977).

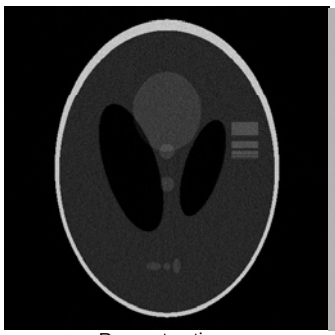
Sostituzione del termine $\mathbf{w} \cdot \mathbf{w}$ con la lunghezza del raggio che attraversa il tessuto (A. Kak and M. Slaney, Principles of Computerized Tomographic Imaging, Classics in Applied Mathematics, SIAM edition, IEEE Press, New York, 2001).

Inoltre....

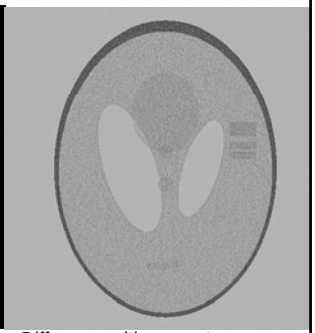
ART does produce a noisy reconstruction

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

Original



Reconstruction
(noise is added to the original)



Difference with respect
to the true image

The solution would be to “regularize” the solution considering more than a single ray at a time.
-SIRT (Simultaneous Iterative Reconstruction Technique) updates the absorption coefficients after considering the rays from **all** the pixels from **all** the projection images.
Time becomes prohobitive.

SIRT algebric interpretations

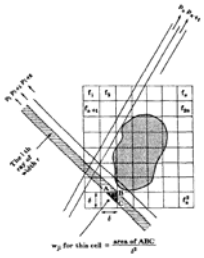
$$\boldsymbol{\mu}^{k+1} = \boldsymbol{\mu}^{k} + \lambda \cdot \mathbf{W}^{T} \cdot \mathbf{D} \cdot \left(\mathbf{p} - \mathbf{W} \cdot \boldsymbol{\mu}^{k}\right) \quad SIRT$$

Gradiente

1) $\mathbf{W} \cdot \boldsymbol{\mu}^{k}$ *projection*

2) $\left(\mathbf{p} - \mathbf{W} \cdot \boldsymbol{\mu}^{k}\right)$ *projection error*

3) $\mathbf{W}^{T} \cdot \mathbf{D} \cdot \left(\mathbf{p} - \mathbf{W} \cdot \boldsymbol{\mu}^{k}\right)$ *retro – projection of the error*



$p_i = \mathbf{d}_{ii}\, \mathbf{w}_i \cdot \mathbf{f} + \nu$

SIRT è **un metodo del gradiente (primo ordine)**.

Si può dimostrare che SIRT minimizza **l'errore quadratico pesato sulle proiezioni**: $||D\,(p - W\mu^k)||^2 = \min \nu^2$ pesato dalla matrice diagonale **D** che contiene gli inversi delle lunghezze al quadrato dei raggi.

Si può anche vedere SIRT come un metodo del gradiente scalato con la matrice **D**.

Video Matlab sul confronto con i 3 metodi.

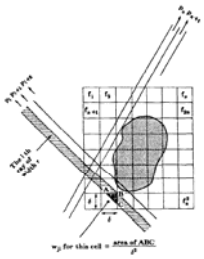
SIRT statistical interpretation

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

1) $\mathbf{W} \cdot \boldsymbol{\mu}^{k}$ *projection*

2) $\left(\mathbf{p} - \mathbf{W} \cdot \boldsymbol{\mu}^{k}\right)$ *projection error*

3) $\mathbf{W}^{T} \cdot \mathbf{D} \cdot \left(\mathbf{p} - \mathbf{W} \cdot \boldsymbol{\mu}^{k}\right)$ *retro – projection of the error*



$p_i = \mathbf{w}_i \cdot \mathbf{f} + \nu$

SART è **un metodo del gradiente (primo ordine)**. E' riconducibile alla **stima a massima verosimiglianza dei parametri**, sotto l'ipotesi di **rumore Gaussiano a media nulla**.

Le osservazioni (misure) vengono pesate con $\mathbf{D}_{ii}$.

Costo di SIRT e Ruttiman per tomografia

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

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

Ruttiman esplicita l'effetto di un piano ricostruito sugli altri piani.

Il costo della singola iterazione dipende dalla dimensione della matrice **W**, pari a
M righe (numero di dati acquisiti) per N colonne (numero di voxel),
con N = NLxNLxNL, $256^3 = 2^{24}$ voxel.
Sia NI il numero di proiezioni acquisite, di dimensione NRxNC,
M=NIxNRxNC (256 x 512 x 512) = 2^{26} pixel.

Costo di SIRT (si suppone campionato il raggio nei voxel con passo NL/2.
Proiezione, 2NL ∀pixel, costo totale: 2xNLxM. ***Errore***. Costo pari ad M.
Retroproiezione, 2xNLxM ***Costo totale***. Mx(4NL+1)=O(MxNL)=2^{36}

Costo di Ruttiman
Calcolo blurring (effetto di ciascun pixel su ciascun piano): N x NL x NI
Calcolo errore, aggiornamento: N ciascuno. ***Costo totale***: Nx(NL x NI + 2)=2^{42}

Si preferisce quindi implementare ed utilizzare SIRT, invece di tomosintesi + Ruttiman

La tomosintesi pura (retroproiezione) può essere utilizzata come **inizializzazione** di SIRT.

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

Three main characteristics:
1) Use of a better approximation of the coefficient matrix.
2) Updating with a "per epoch" style, analogous to that of SIRT: all the ray in one image at a time are considered.
3) Hamming windowing in the longitudinal dimension to emphasize the correction coefficients at the volume center.

In our implementation, we will consider only 1) and 2), as in the most recent literature.

From ART to: SART

$$\mu^{(t+1)} = \mu_j^{(t)} + \lambda\left(p_i - \mathbf{a_i} \cdot \boldsymbol{\mu}\right) \qquad \mu^{(t+1)} = \mu_j^{(t)} + \lambda \frac{\sum_i a_{ij} \frac{p_i - \mathbf{a_i} \cdot \boldsymbol{\mu}}{}}{\sum_i a^2{}_{ij}}$$

Where a_{ij} is a weight that measures the "closenss" of the voxel center and the ray through the pixel center. In ART $a_i = \{0, 1\}$

I coefficienti di SART

Interpolazione dei valori vicini: interpolazione trilineare nel volume e bilineare sulle immagini.
Dato un punto P=(X,Y,Z), definendo (x_k,y_k,z_k) le coordinate del vettore distanza da P al centro del voxel V_k, normalizzate tra 0 e 1, il coefficiente a_k, relativo a V_K viene calcolato come:

$$a_k = (1-x_k)(1-y_k)(1-z_k)$$

con:

$$\mu^{(t+1)} = \mu_j^{(t)} + \lambda \frac{\sum_i a_{ij} \frac{p_i - \mathbf{a_i} \cdot \boldsymbol{\mu}}{\sum_i a_{ij}}}{\sum_i a_{ij}}$$

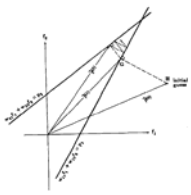
$$\boldsymbol{\mu}^{\mathbf{k+1}} = \boldsymbol{\mu}^{\mathbf{k}} + \lambda \cdot \mathbf{W}^{\mathbf{T}} \cdot \mathbf{D} \cdot (\mathbf{p} - \mathbf{W} \cdot \boldsymbol{\mu}^{\mathbf{k}}) \qquad \mathbf{D}_{\mathbf{ii}} = \frac{1}{\sum_j a^2_{ij}}$$

Sceglieremo di interpolare nel **volume (in CUDA viene realizzata in HW)** e quindi **"pettineremo**" le immagini, tracciando i raggi sui quali calcolare l'assorbimento solamente in corrispondenza dei centri dei pixel.

OS-SART

Invece di aggiornare i coefficienti dopo la singola immagine (SART puro) o dopo tutte le immagini (SIRT), si sceglie un compromesso aggiornando dopo avere esaminato K immagini < M. K rappresenta un sub-set delle immagini.

che può essere scelto in diverso modo (ordered-subset, OS):
- Sequenziale.
- Massima ortogonalità tra le immagini (projection algorithm)
- Random.



Il parametro di damping può rivelarsi critico, da impostare sperimentalmente.

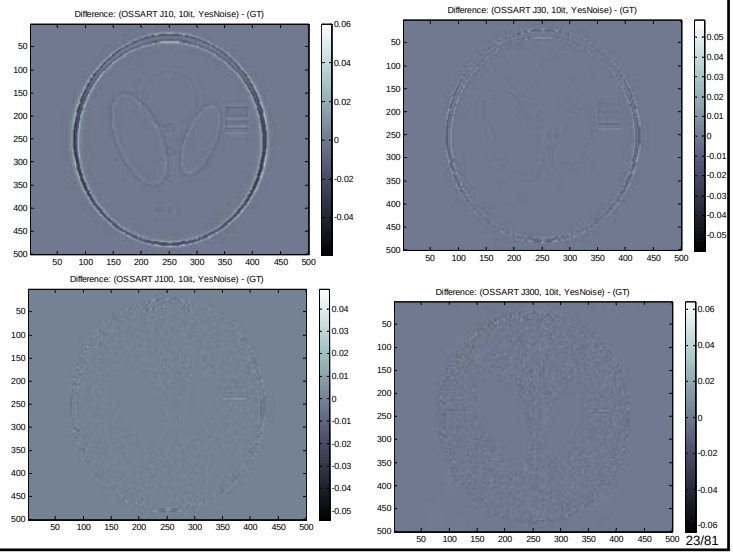
$$\mu^{(t+1)} = \mu_j^{(t)} + \lambda \frac{\sum_i a_{ij} \frac{p_i - \mathbf{a_i} \cdot \boldsymbol{\mu}}{\sum_i a_{ij}}}{\sum_i a_{ij}}$$

Speed-up con OS-SART

Ricostruzione a partire da 900 immagini su 360 gradi.
Sottocampionamento delle immagini (e.g. J10, 1 immagine ogni 10).
10 iterazioni complete sul data set (tempo pari ad una iterazione con tutte le immagini).

OS con poche immagini, risultati simili ad ART

OS con tante immagini, risultati simili a SIRT



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

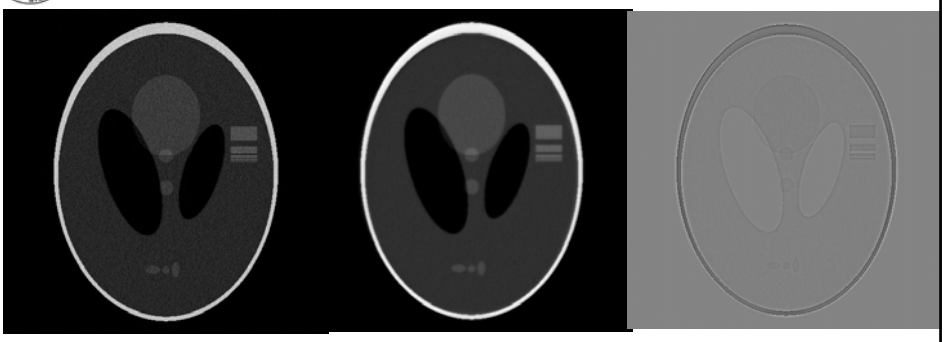
Simluations to assess ART problems

Shepp-Logan phantom (Shepp and Logan, 1974): 2D image, brisk edges, homogenoues regions, 500 x 500 pixels.

From Matlab.

Standard image for benchmarking in tomography and radiographic imaging

SART produces a low-pass filtered image



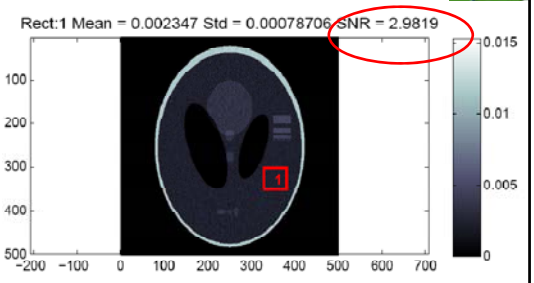
Original noisy image Reconstructed imge Difference

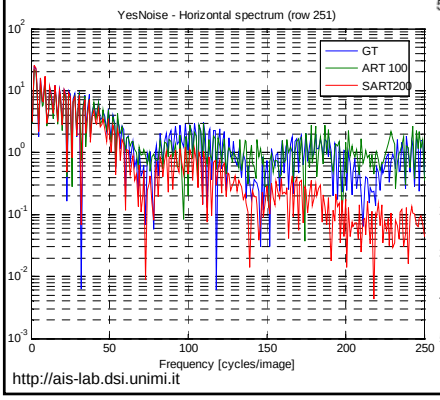
Low-pass filtering is evident at the border of the objects.

Noise in ART and SART

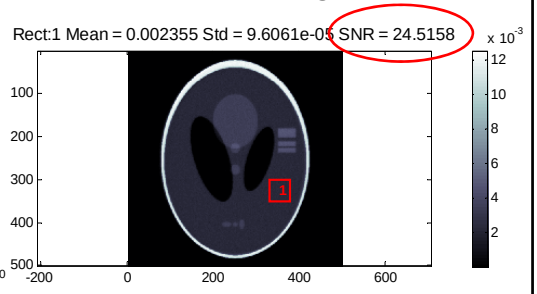
The noise on reconstructed images through ART is evident. Noise is Poisson.

ART

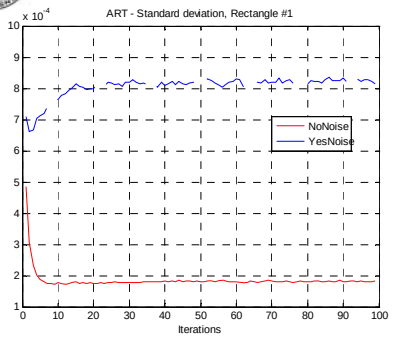


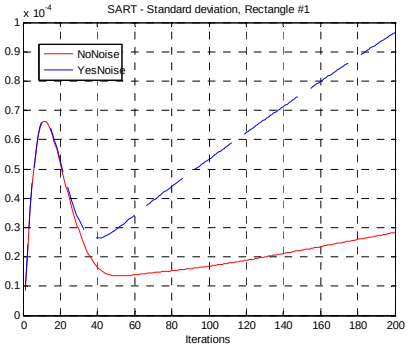


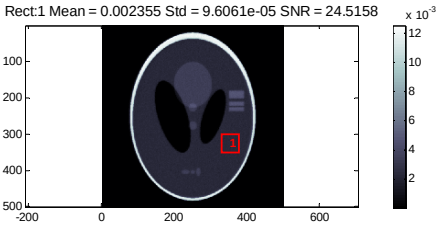
SART



SART and semiconvergence



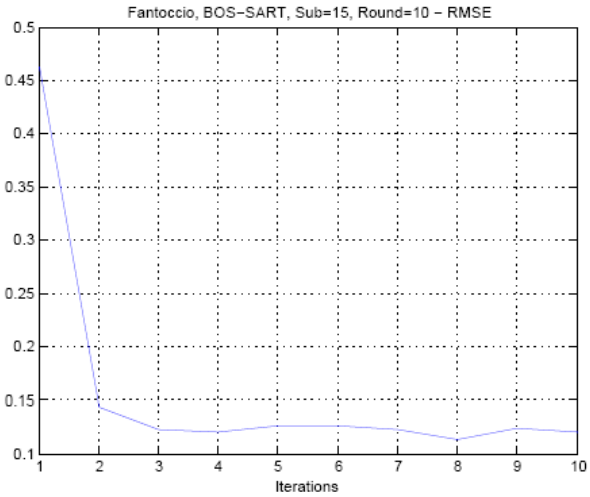




Number of iterations is critical in SART: standard deviation in region 1 decreases up to the 40° iteration and start increasing again: 40 is a good number of iterations.

Increasing the number of iterations SART –> ART.

Effetto sulla convergenza di OS



Con un OS di 15, su 900 immagini, un $\lambda = 0.45$, si ottiene un buon risultato già dopo la prima iterazione, come riportato anche nella letteratura.

Summary on SART

- SART produces a more homogenous image, with noise filtering.
- Its weak points are the convergence time (number of iterations) and the low-pass filtering effect.
- Acceleration can be achieved through OS techniques.
- GPU processing is mandatory to obtain results in a reasonable amount of time.
- Low-pass filtering can be limited by limiting the number of iterations or adopting a regularization approach (that can be computationally heavy).

- Parameters to be optimized empirically:
  - Number of images in the OS
  - Value of damping factor, λ
  - Voxel dimension.

Overview

- Algebric reconstruction
- Limited angle tomography
- Bayesian Restoration of X-ray images

Motivation for limited angle tomography

Dose reduction
Costs reduction

MAMMOMAT - Siemens

Hyperion
CEFLA (MyRay)

Multiple images over 180 degrees -> 3D volumetric images
Multiple images over a smaller angle -> 2.5D volumetric images

Which are the limitations?

- Per effetto dell'allargamento dell'angolo di acquisizione è possibile allargare l'intervallo delle frequenze osservabili (diversi orientamenti) nel volume ricostruito.
- Per contro, un allargamento dell'angolo di acquisizione porta ad una diminuzione della risoluzione nei piani ortogonali rispetto alla direzione di acquisizione principale.
- E' necessario trovare il giusto compromesso…

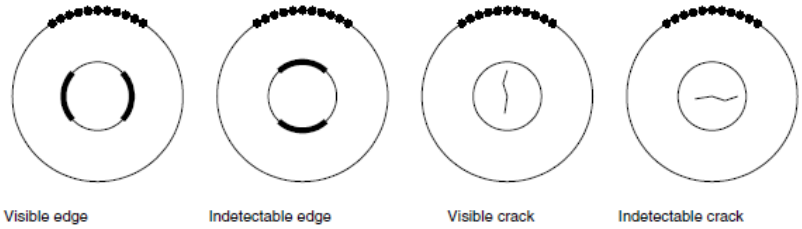


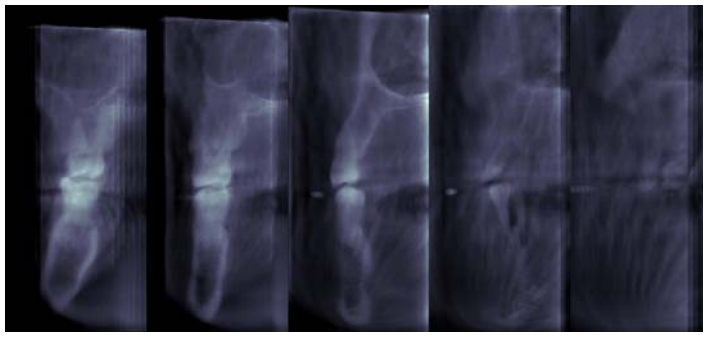
Figure 3. Illustration of discontinuities that can and cannot be recovered based on limited-angle projection data. The black dots denote the locations of the x-ray source in the acquisition of the projection data. For each such location, the detector is thought to be located opposite to the x-ray source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

Somersalo et al., 2003

Which are the results of using limited angle?

Limited resolution on depth -> parallel sections

- Mammografia (sezioni parallele al torace)
- Dentale (sezioni parallele al canale mandibolare)

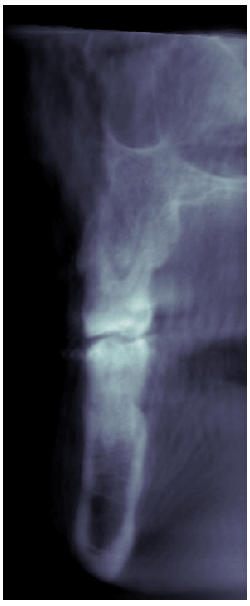


Standard configuration

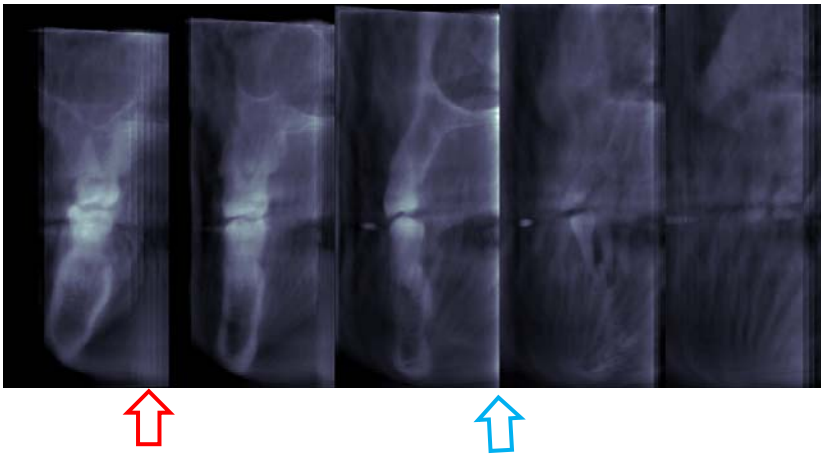
- Acquisition angle = -30°… +30°
- Acquisition time = 60s
- Xray duration = 26s (1,5s x 11)
- Reconstruction time =
  - 118s (OS-SART, 5 iterations)
  - Real time (a few s, tomosynthesis)
- Number of projections = 11
- KV 60-84, mA 1-10, set from panoramic image
- Enlarging factor = 1.43
- Images size = 1536 x 662 pixel

  Useful images size = 1536x**562** 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
- Automatic identification of the volume in the panoramic image.

  Volume is centered slightly inside for frontal teeth.



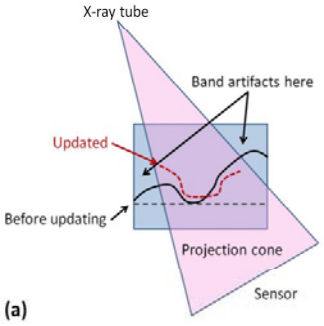
Artefacts with band shape and truncation



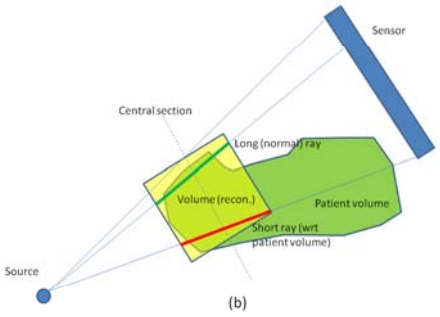
Five sections from SART. Notice vertical lines on the right of the image and the whitening of the image.

Problems analysis

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



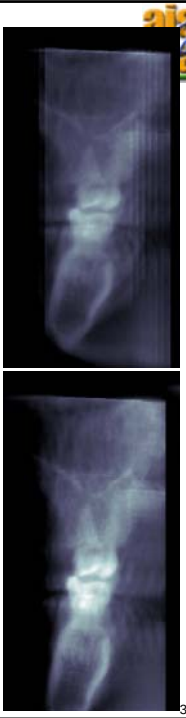
X-ray absorption outside the volume



We can solve these problems modifying the optimization engine used in SART

Solution for band artifacts

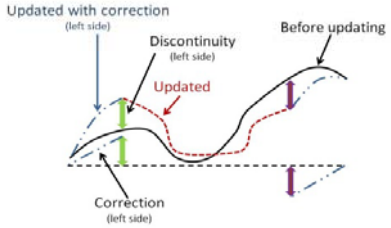
$$\boldsymbol{\mu}^{k+1} = \boldsymbol{\mu}^{k} + \lambda \cdot \mathbf{W}^{T} \cdot \mathbf{D} \cdot \left(\mathbf{p} - \mathbf{W} \cdot \boldsymbol{\mu}^{k}\right) \quad SART$$



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)

Solution for band artifacts

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

$$\mathbf{D}_{ii} = \frac{1}{\sum_{j} a^{2}_{ij}}$$

Estimate of absorption outside volume in voxels with abnormal response. (Zhang et al., J. Comput. Ass. Tomography 2009)

Modify the weights:
The longer the ray, the more it receives error, the more reliable is the update.
→In least square sense, the pixel associated to that ray is more reliable → higher weight.

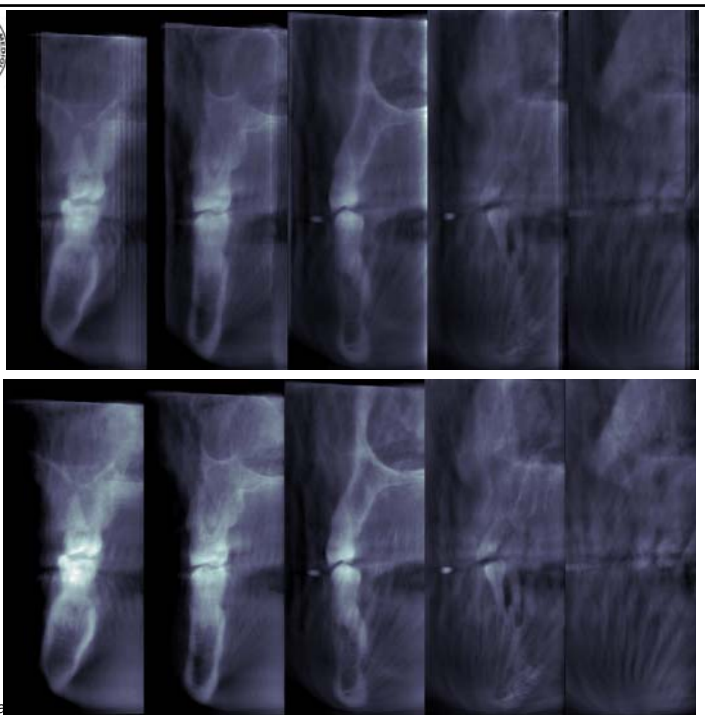
$\boldsymbol{\mu}^{k+1} = \boldsymbol{\mu}^{k} + \lambda\, \mathbf{W}^{T}\, \mathbf{D}^{-1}\, (\mathbf{p} - \mathbf{W}\, \boldsymbol{\mu}^{k})$

A new weighting scheme for limited angle arises
(Frosio and Borghese, in preparation)

$p_i = \mathbf{w}_i \cdot \mathbf{f}$

In this situation the voxels are updated more strongly for a ray that crosses a thicker volume region because it is more realiable that the measured absorption is due to the voxel crossed and not to outside ones (it is a different phylosophy).

Some results

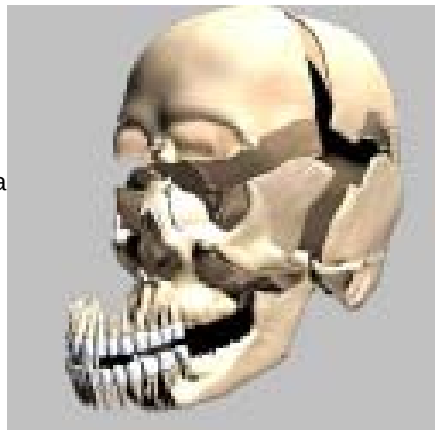


Notice the increase in the dynamics

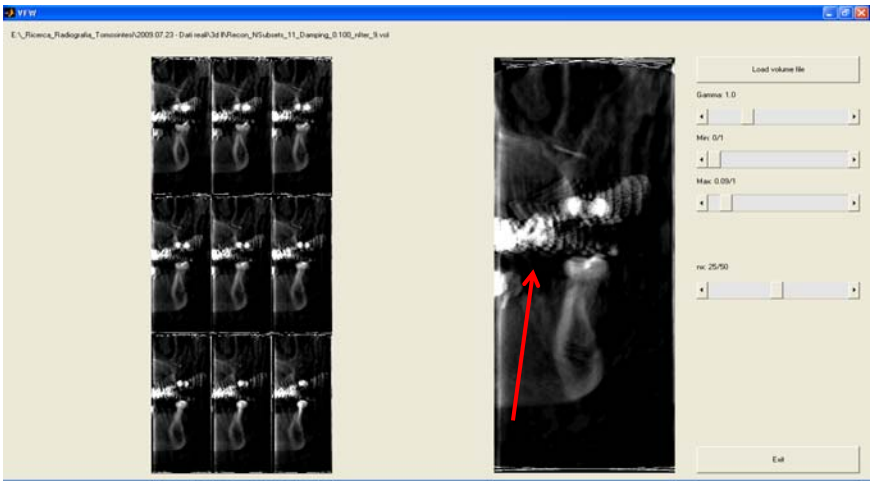
Ottimizzazione del sistema patient-based

- Utilizzo di un modello anatomico per il calcolo dell'esposizione ottimale immagine per immagine.
- Modello parametrizzato automaticamente a partire dalla panoramica del paziente.
- Considerazione dell'effetto del metallo (amalgama, protesi) sia nell'esposizione che nella formulazione della traiettoria.

Risultato: Optimized images (higher contrast and higher dynamics).



Riduzione degli artefatti da metallo



E' un problema che si sposa bene con le tecniche di ricostruzione algebriche

Compensazione del movimento (eliminazione blurring da movimento)

Identificazione del movimento mediante identificazione di feature sull'immagine radiogafica.

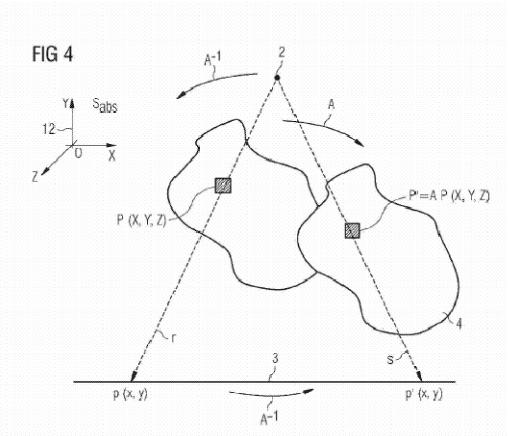
Ricerca di feature adatte per i vari distretti mandibolari.

Determinazione del movimento. Affidabilità della stima.

Studio di fattibilità (video? X-ray? Modifiche alla macchina?)

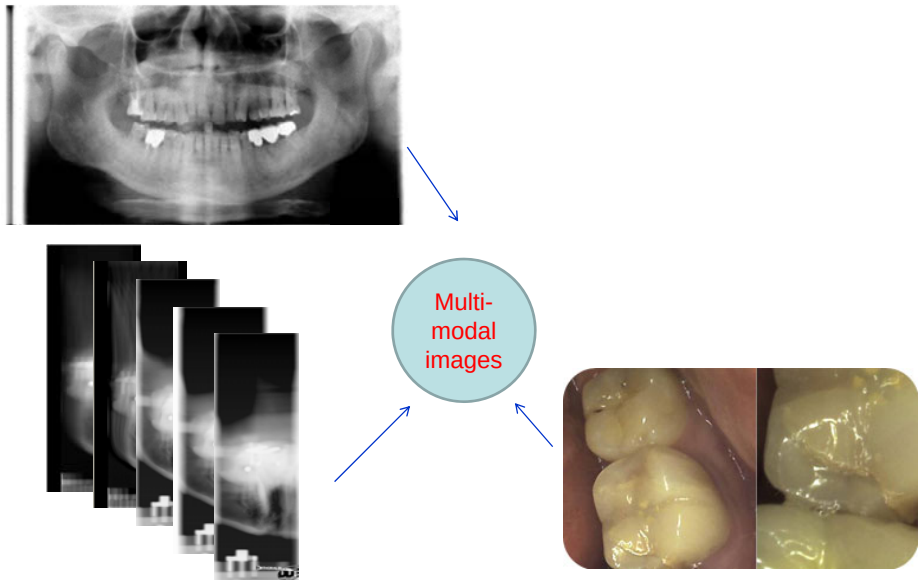
Obbiettivo: Correzione del "blurring da movimento"

Possibilità di correggere anche errori di posizionamento macchina / taglio immagini.



Borghese, Nanni, Rotondo
US Patent 0008559, Jan 2010

Integration of video, 3D images and panoramic imaging (imaging multi-modale)

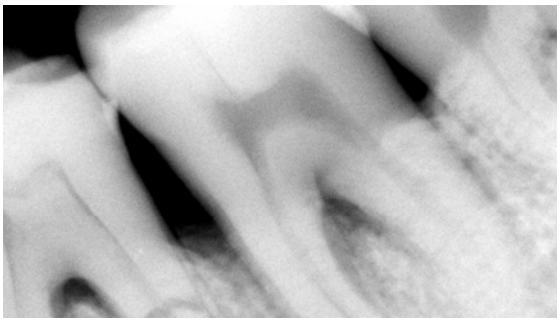


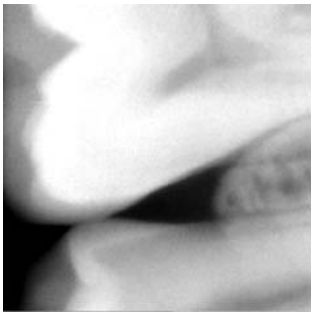
Overview

- Algebric reconstruction
- Limited angle tomography
- Bayesian Restoration of X-ray images

Denoising of a radiographic image

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





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:

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

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:

$A^T Af = A^T g_n$ **That is the least squares solution**.

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

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

Can we solve any problem?

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

Going in the frequency domain: $G(w) = G_o(w) + W(w) = K(w)F_0(w) + W(w)$

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

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

$G(w) - W(w) = K(w)F_0(w)$ ➔

No solution exists as noise W(w) cannot be reproduced by $F_0(w)$.

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

As K(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.

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 observing f and g_n (joint probability) 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)} =

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

Likelihood = adherence to the data

A-priori

Tikhonov regularization

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

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

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

$\| \lambda Pf \|^2$

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

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

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

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

$$\| \lambda Pf \|^2$$

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

Regularization

No analytical solution -> non-linear optimization

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

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

Total variation

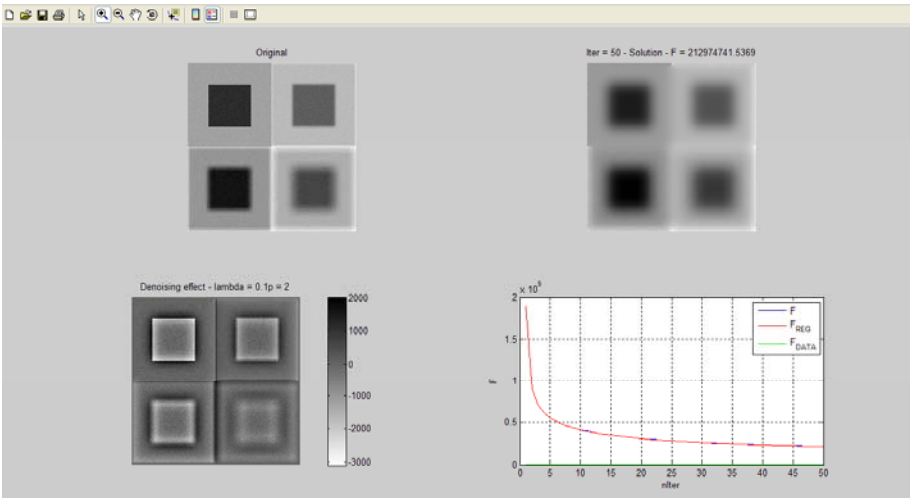
$$\sum_i \sqrt{\sum_{w=1}^{W} p_{f_w}{}^2}$$

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

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

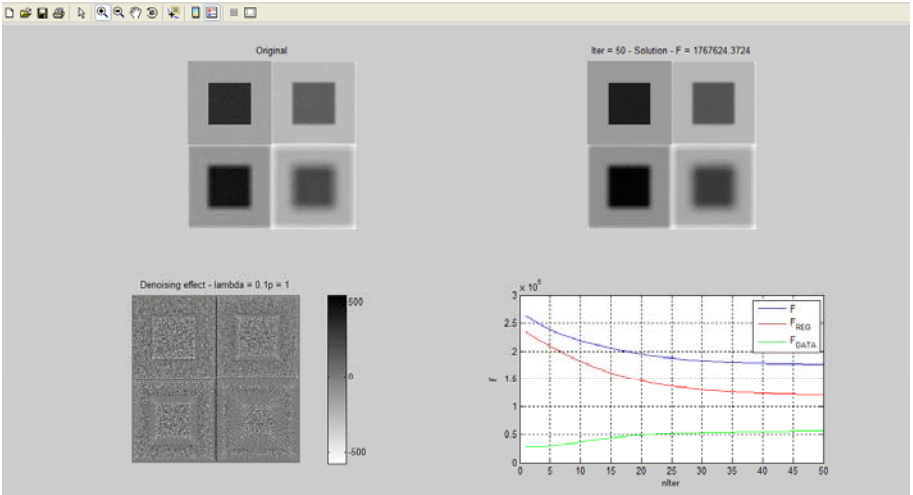
$$f = \arg\min_f \sum_i \left(\|g_n - Af\|^2 + \lambda \sqrt{\sum_1^W p_{f_w}{}^2} \right)$$

Why total variation - simulations



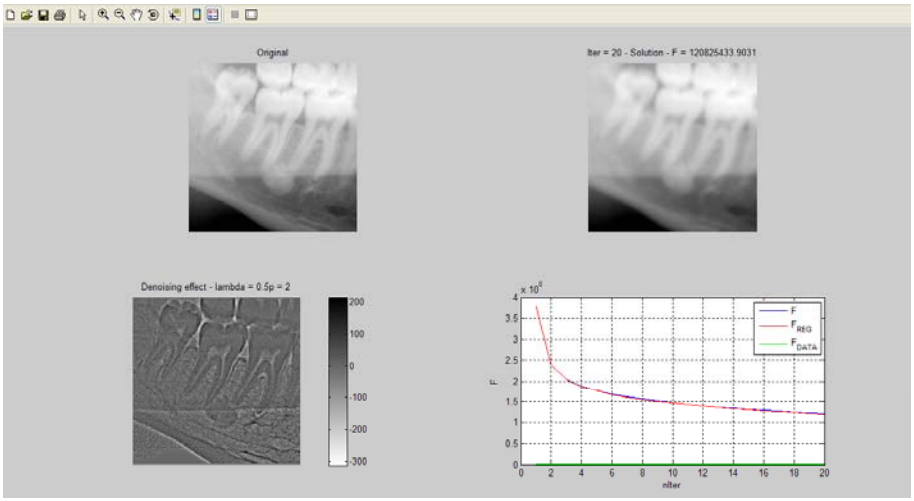
Edge smoothing effect with Tikhonov-like regularization
Poisson noise model – $\lambda = 0.5$
P is the gradient operator

Why total variation - simulations



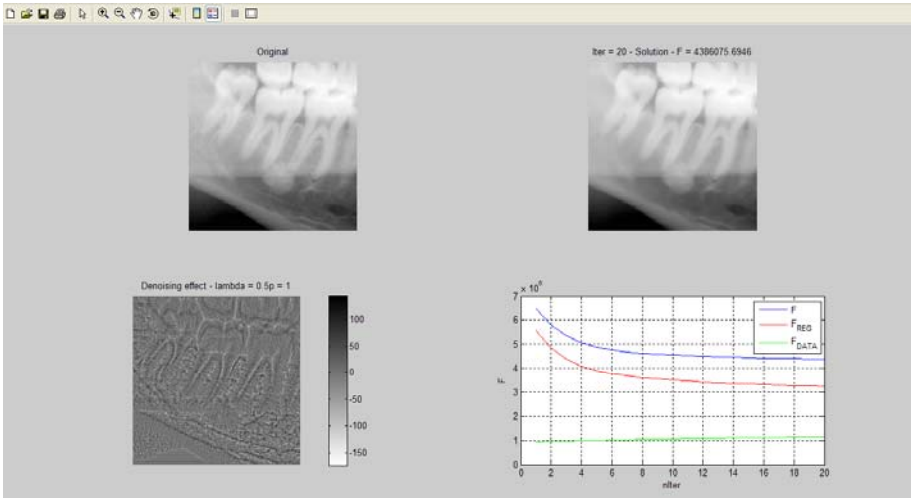
No appreciable edge smoothing with total variation
Poisson noise model - $\lambda = 0.5$
P is the gradient operator

Why total variation – panoramic images



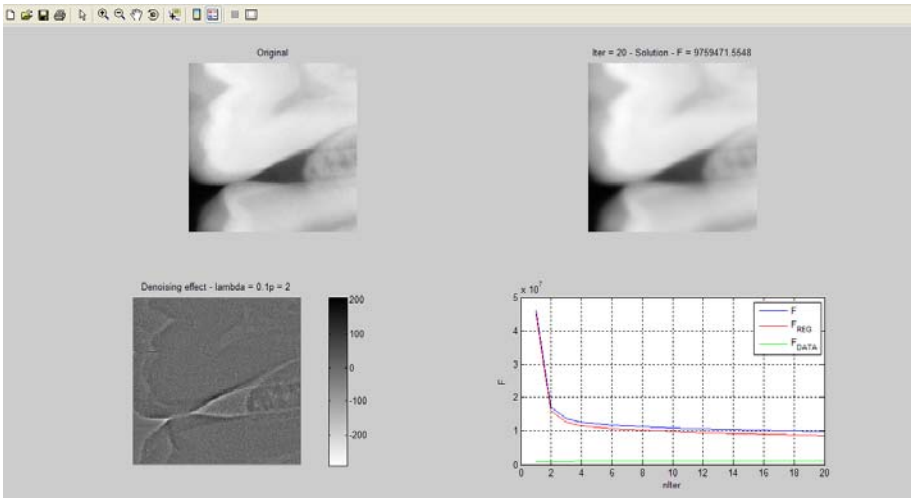
Edge smoothing effect with Tikhonov-like regularization
Poisson noise model - $\lambda = 0.5$
P is the gradient operator

Why total variation – panoramic images



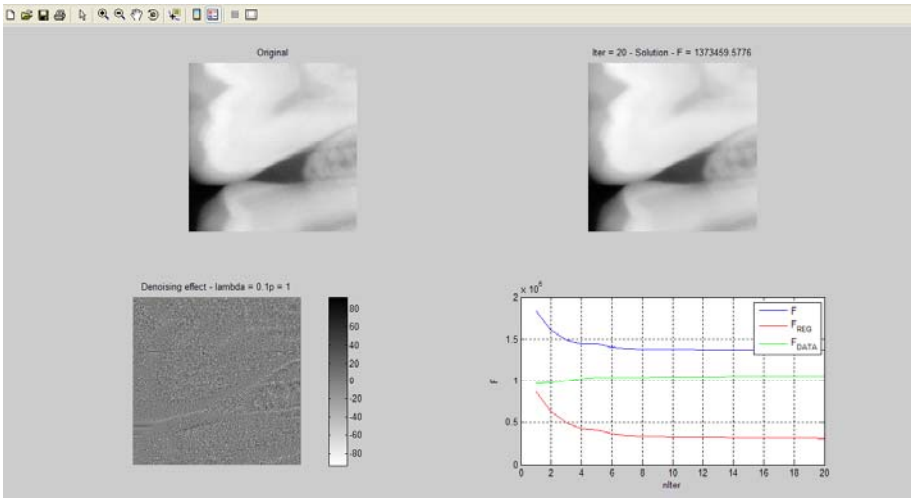
No appreciable edge smoothing with total variation
Poisson noise model - $\lambda = 0.5$
P is the gradient operator

Why total variation- endo-oral images



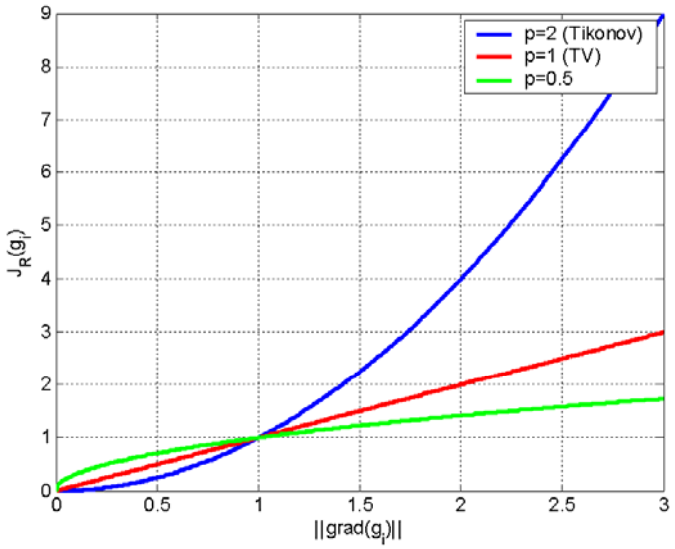
Edge smoothing effect with Tikhonov-like regularization
Poisson noise model - $\lambda = 0.1$
P is the gradient operator

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

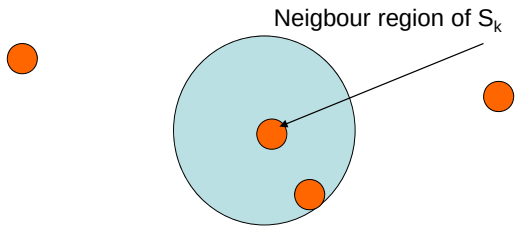


Cost increases quadratically with the local gradient in Tikhonov

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.

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	

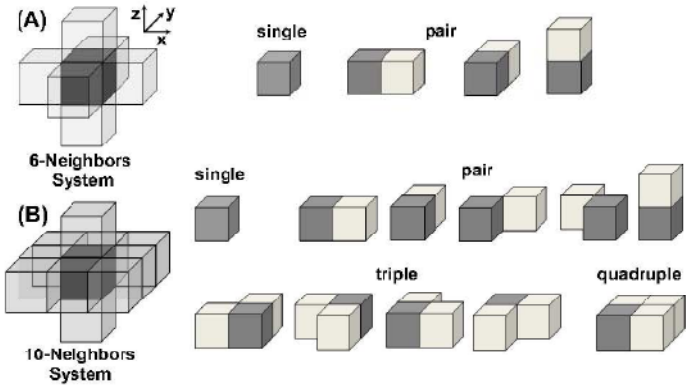
First order neighboring System
4-neighboring System

o	o	o
o	x	o
o	o	o

Second order neighboring System
8-neighboring System

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 $\mathcal{N}_p$

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**.

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.

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.

Gibbs priors and Regularization

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

Likelihood = adherence to the data

A-priori

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

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

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

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

$J_R(f) = U(f)$

Role of λ

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

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

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.

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.

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.

Quadratic Priors with k = 1

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

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

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:

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

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

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).

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.

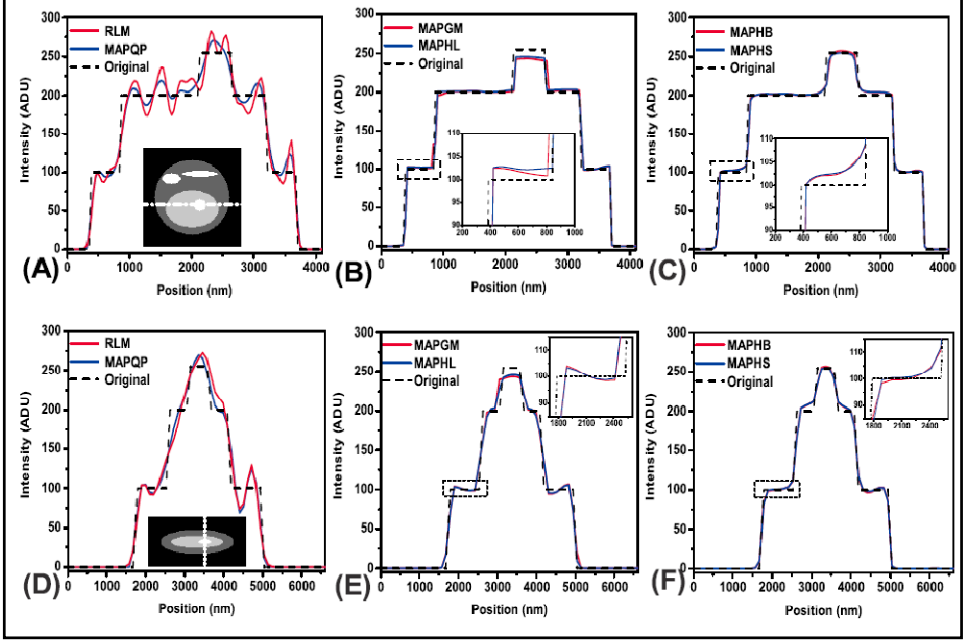
Few non-quadratic potentials (Vicedomini 2008)

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

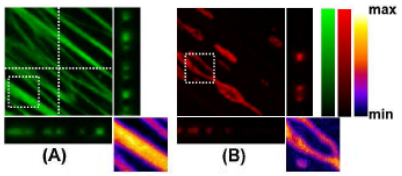
Asymptotic linear behavior

Asymptotic log-like behavior

Results

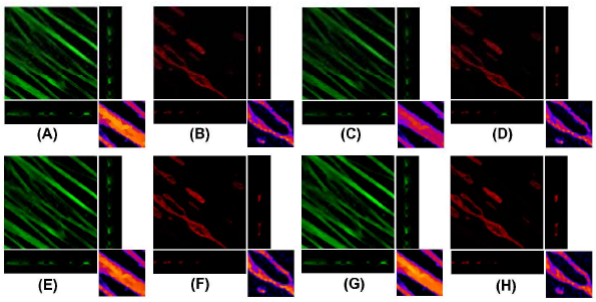


Application to optical microscopy



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)



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.

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.

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.