Simultaneous Tomographic Reconstruction and Segmentation with Class Priors

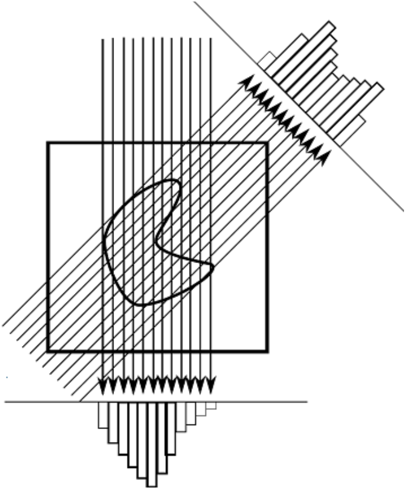
Yiqiu Dong

DTU Compute
Technical University of Denmark
Cooperators: Mikhail Romanov, Anders Bjorholm Dahl, and Per Christian Hansen

Technical University of Denmark

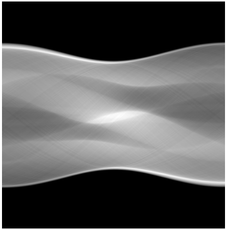
HIGH - DEFINITION TOMOGRAPHY

Transmission Computed Tomography



The principle in parallel-beam tomography: send parallel rays through the object at different angles, measure the damping.

CT Reconstruction and Segmentation



Measurement

$\Longrightarrow$

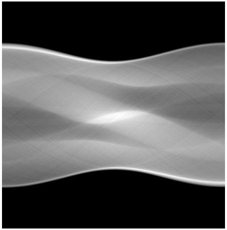
Reconstruction

$\Longrightarrow$

Segmentation

- Reconstruction methods: FBP, ART, variational methods, etc.
- Segmentation methods: thresholding-based methods, level-set methods, graph-cut methods, etc.

Discrete Tomography



Measurement

$\Longrightarrow$

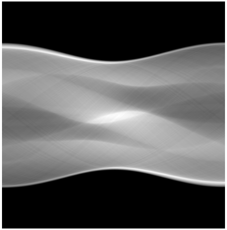
Segmentation

- Idea: allow only a small number of pixel values
- Characteristics: there is no reconstruction
- Methods: DART (Batenburg, et al. 2011), etc.

Simultaneous Reconstruction and Segmentation (SRS)

Reconstruction

$\Longleftarrow$



Measurement

$\Longrightarrow$

Segmentation

- Mumford-Shah level-set methods (Ramlau, et al. 2007)
- Hidden Markov measure field model (HMMFM)-based methods (Van de Sompel, et al. 2008)

Hidden Markov Measure Field Model (HMMFM)

	Class 1	$\cdots$	Class k	$\cdots$	Class K
Pixel 1	δ_{11}	$\cdots$	δ_{1k}	$\cdots$	δ_{1K}
$\vdots$	$\vdots$		$\vdots$		$\vdots$
Pixel j	δ_{j1}	$\cdots$	δ_{jk}	$\cdots$	δ_{jK}
$\vdots$	$\vdots$		$\vdots$		$\vdots$
Pixel N	δ_{N1}	$\cdots$	δ_{Nk}	$\cdots$	δ_{NK}

- δ_{jk} denotes the probability of the pixel j belonging to the class k
- $\sum_{k=1}^{K} \delta_{jk} = 1$ for all $j = 1, \cdots, N$
- The HMMFM, $\delta = \{\delta_{jk}\}$, gives segmentation information

Our Model

$$
\begin{aligned}
\max_{\mathbf{x},\delta} \quad & p(\mathbf{x}, \delta | \mathbf{b}) \\
\text{s.t.} \quad & \sum_{k=1}^{K} \delta_{jk} = 1,\ \delta_{jk} \geq 0,\ j = 1, \cdots, N,\ k = 1, \cdots, K.
\end{aligned}
$$

- $\mathbf{x} \in \mathbb{R}^N$ includes the attenuation coefficients of the object for all pixels
- $\mathbf{b} \in \mathbb{R}^M$ is the measurement
- $\delta \in \mathbb{R}^{N \times K}$ is the set of probabilities in the HMMFM for each class of the object and for each pixel
- $p(\mathbf{x}, \delta | \mathbf{b})$ is the posterior probability density function for the image and the HMMFM with the given data

Our Model

$$\max_{\mathbf{x},\delta} \quad \frac{p(\mathbf{b}|\mathbf{x},\delta)p(\mathbf{x}|\delta)p(\delta)}{p(\mathbf{b})}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \delta_{jk} = 1,\ \delta_{jk} \geq 0,\ j = 1,\cdots,N,\ k = 1,\cdots,K.$$

- $p(\mathbf{b}|\mathbf{x},\delta) = p(\mathbf{b}|\mathbf{x})$ is the probability of obtaining the data **b** given the image **x**; the data does not depend on the segmentation of the image
- $p(\mathbf{x}|\delta)$ is the probability of **x** given the probabilities of each pixel in each class
- $p(\delta)$ expresses our belief in the HMMFM
- $p(\mathbf{b})$ is a normalization constant

Our Model

$$
\begin{aligned}
\max_{\mathbf{x},\delta} \quad & \log p(\mathbf{b}|\mathbf{x}) + \log p(\mathbf{x}|\delta) + \log p(\delta) \\
\text{s.t.} \quad & \sum_{k=1}^{K} \delta_{jk} = 1,\ \delta_{jk} \geq 0,\ j = 1, \cdots, N,\ k = 1, \cdots, K.
\end{aligned}
$$

- $p(\mathbf{b}|\mathbf{x}, \delta) = p(\mathbf{b}|\mathbf{x})$ is the probability of obtaining the data **b** given the image **x**; the data does not depend on the segmentation of the image
- $p(\mathbf{x}|\delta)$ is the probability of **x** given the probabilities of each pixel in each class
- $p(\delta)$ expresses our belief in the HMMFM
- $p(\mathbf{b})$ is a normalization constant

Data Fitting Term, $p(\mathbf{b}|\mathbf{x})$

The measured data is usually a sum of several terms:

1. Data received from the X-ray illumination of the object, with Poisson noise
2. Poisson noise of the measuring equipment and from external sources
3. Gaussian noise caused by the electronics and the conversion from an analog signal to digital data

Data Fitting Term, $p(\mathbf{b}|\mathbf{x})$

The measured data is usually a sum of several terms:

1. Data received from the X-ray illumination of the object, with Poisson noise
2. Poisson noise of the measuring equipment and from external sources
3. Gaussian noise caused by the electronics and the conversion from an analog signal to digital data

Assumption: The signal-to-noise level is high enough that we can approximate the total noise by additive white Gaussian noise $\mathcal{N}(0, \sigma^2_{\text{noise}})$

Data Fitting Term, $p(\mathbf{b}|\mathbf{x})$

The measured data is usually a sum of several terms:

1. Data received from the X-ray illumination of the object, with Poisson noise
2. Poisson noise of the measuring equipment and from external sources
3. Gaussian noise caused by the electronics and the conversion from an analog signal to digital data

Assumption: The signal-to-noise level is high enough that we can approximate the total noise by additive white Gaussian noise $\mathcal{N}(0, \sigma^2_{\mathrm{noise}})$

$$\log p(\mathbf{b}|\mathbf{x}) = -\frac{\|A\mathbf{x} - \mathbf{b}\|_2^2}{2\sigma^2_{\mathrm{noise}}} - \frac{1}{2}\log(2^M \pi^M \sigma^{2M}_{\mathrm{noise}})$$

Class Fitting Term, $p(\mathbf{x}|\delta)$

Assumption:

1. The object is composed of a set of K different phases, and each phase has the similar attenuation coefficient everywhere. Here, K is given and $K << N$
2. The distribution of the attenuation coefficients within a class is Gaussian with mean value μ_k (the expected attenuation coefficient) and a small standard deviation σ_k. All μ_k and σ_k for $k = 1, \cdots, K$ are known

Class Fitting Term, $p(\mathbf{x}|\delta)$

Assumption:

1. The object is composed of a set of K different phases, and each phase has the similar attenuation coefficient everywhere. Here, K is given and $K << N$
2. The distribution of the attenuation coefficients within a class is Gaussian with mean value μ_k (the expected attenuation coefficient) and a small standard deviation σ_k. All μ_k and σ_k for $k = 1, \cdots, K$ are known

$$p(\mathbf{x}|\delta) = \prod_{j=1}^{N} p(x_j|\delta_j) = \prod_{j=1}^{N} \sum_{k=1}^{K} p(x_j|\mathrm{class} = k)p(\mathrm{class} = k|\delta_j)$$

Class Fitting Term, $p(\mathbf{x}|\delta)$

Assumption:

1. The object is composed of a set of K different phases, and each phase has the similar attenuation coefficient everywhere. Here, K is given and $K << N$
2. The distribution of the attenuation coefficients within a class is Gaussian with mean value μ_k (the expected attenuation coefficient) and a small standard deviation σ_k. All μ_k and σ_k for $k = 1, \cdots, K$ are known

$$
\begin{aligned}
p(\mathbf{x}|\delta) = \prod_{j=1}^{N} p(x_j|\delta_j) &= \prod_{j=1}^{N} \sum_{k=1}^{K} p(x_j|\mathrm{class} = k) p(\mathrm{class} = k|\delta_j) \\
&= \prod_{j=1}^{N} \sum_{k=1}^{K} \delta_{jk} \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x_j-\mu_k)^2}{2\sigma_k^2}}
\end{aligned}
$$

Class Fitting Term, $p(\mathbf{x}|\delta)$

Assumption:

1. The object is composed of a set of K different phases, and each phase has the similar attenuation coefficient everywhere. Here, K is given and $K << N$
2. The distribution of the attenuation coefficients within a class is Gaussian with mean value μ_k (the expected attenuation coefficient) and a small standard deviation σ_k. All μ_k and σ_k for $k = 1, \cdots, K$ are known

$$\log p(\mathbf{x}|\delta) = \sum_{j=1}^{N} \log \left(\sum_{k=1}^{K} \delta_{jk} \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x_j - \mu_k)^2}{2\sigma_k^2}} \right)$$

Regularization Term, $p(\delta)$

In regularization term, we specify the prior knowledge about the behavior of the segmentation.

$$\log p(\delta) = -\sum_{k=1}^{K} R(\delta_k)$$

where $\delta_k = \{\delta_{1k}, \cdots, \delta_{Nk}\}$ is the set of probabilities for class k.

Regularization Term, $p(\delta)$

In regularization term, we specify the prior knowledge about the behavior of the segmentation.

$$\log p(\delta) = -\sum_{k=1}^{K} R(\delta_k)$$

where $\delta_k = \{\delta_{1k}, \cdots, \delta_{Nk}\}$ is the set of probabilities for class k.

- $R_{\mathrm{TV}}(\delta_k) = \sum_{j=1}^{N} \left((\delta_{jk} - \delta_{j'k})^2 + (\delta_{jk} - \delta_{j''k})^2\right)^{\frac{1}{2}}$
  Allow discontinuities in the probabilities for the classes associated with neighboring pixels
- $R_{\mathrm{Tik}}(\delta_k) = \sum_{j=1}^{N} \left((\delta_{jk} - \delta_{j'k})^2 + (\delta_{jk} - \delta_{j''k})^2\right)$
  Enforce the smoothness of the probabilities among the classes associated with neighboring pixels

Full Model

$$\min_{\mathbf{x},\delta} \quad \lambda_{\text{noise}}\|A\mathbf{x}-\mathbf{b}\|_2^2 + \lambda_{\text{class}}\sum_{k=1}^{K} R(\delta_k)$$

$$-\sum_{j=1}^{N}\log\left(\sum_{k=1}^{K}\delta_{jk}\frac{1}{\sqrt{2\pi}\sigma_k}e^{-\frac{(x_j-\mu_k)^2}{2\sigma_k^2}}\right)$$

$$\text{s.t.} \quad \sum_{k=1}^{K}\delta_{jk}=1,\ \delta_{jk}\geq 0,\ j=1,\cdots,N,\ k=1,\cdots,K.$$

- λ_{noise} and λ_{class} are positive regularization parameters
- The model is non-convex
- The objective function is convex w.r.t δ, but non-convex w.r.t $\mathbf{x}$

Simplification

Non-convexity: For each pixel, it is a multi-modal function, which consists of a sum of Gaussian functions.

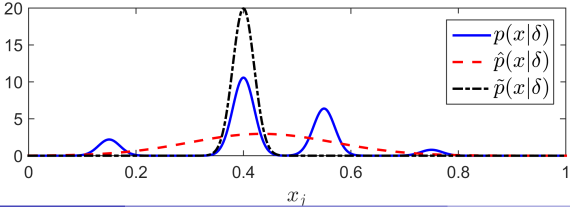
Simplification

Non-convexity: For each pixel, it is a multi-modal function, which consists of a sum of Gaussian functions.

Idea: We approximate this multi-modal function by a single Gaussian function, i.e.,

$$-\log p(\mathbf{x}|\delta) = \sum_{j=1}^{N} \frac{(x_j - \mu_j^{new})^2}{2(\sigma_j^{new})^2}.$$

First, we use a relatively flat Gaussian function to roughly approximate the original model. When we are close to the solution, we expect that most of the pixels clearly belong to one specific class, then we switch to a sharp Gaussian function to approximate the original model.



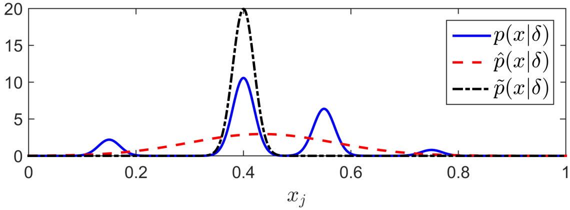
Simplification

- **Stage 1:** We set for $j = 1, \cdots, N$

$$\mu_j^{new} = \sum_{k=1}^{K} \delta_{jk}\mu_k, \quad (\sigma_j^{new})^2 = \sum_{k=1}^{K} \delta_{jk}(\sigma_k^2 + \mu_k^2) - (\mu_j^{new})^2$$

- **Stage 2:** We set for $j = 1, \cdots, N$

$$\delta_{jk} = \begin{cases} 1, & \text{if}\, k = k_j \\ 0, & \text{otherwise} \end{cases}, \quad \mu_j^{new} = \mu_{k_j}, \quad \sigma_j^{new} = \sigma_{k_j}$$



Algorithm

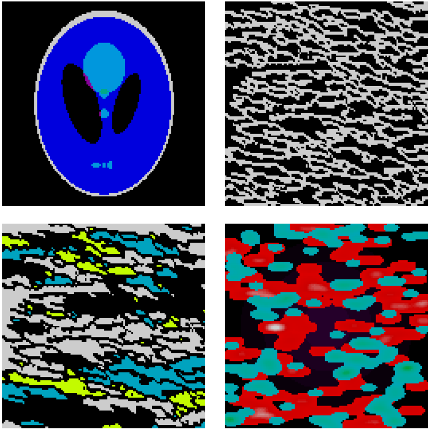
- **Initialization:** $\delta^0_{jk} = \frac{1}{K}$ for all j, k and $\mathbf{x}^0 = \mathbf{0}$
- **Stage 1:**
  - For $n = 1, \cdots, n_1$
    1. Calculate $\mu^n_j = \mu^{new}_j$ and $\sigma^n_j = \sigma^{new}_j$ in order to obtain the "flat" Gaussian simplification
    2. Update $\mathbf{x}^n = \arg\min_{\mathbf{x}} \lambda_{\text{noise}} \|A\mathbf{x} - \mathbf{b}\|^2_2 + \sum_{j=1}^N \frac{(x_j - \mu^n_j)^2}{2(\sigma^n_j)^2}$ by the CGLS algorithm with initial guess $\mathbf{x}^{n-1}$
    3. Update

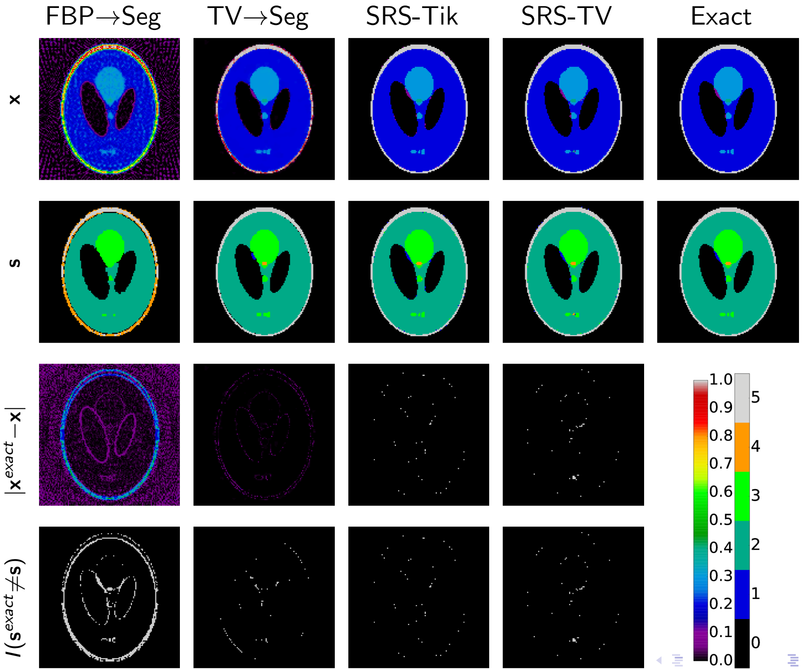
       $$\delta^n = \arg\min_\delta \lambda_{\text{class}} \sum_{k=1}^K R(\delta_k) - \sum_{j=1}^N \log\left(\sum_{k=1}^K \frac{\delta_{jk}}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x^n_j - \mu_k)^2}{2\sigma^2_k}}\right)$$

       subject to $\sum_{k=1}^K \delta_{jk} = 1$ and $\delta_{jk} \geq 0$ for all j, k
       by the iterative Frank-Wolfe algorithm with initial guess δ^{n-1}
- **Stage 2:**
  - For $n = n_1 + 1, \cdots, n_1 + n_2$
    Similar as Stage 1, but with $\mu^n_j = \mu_{k_j}$ and $\sigma^n_j = \sigma_{k_j}$ in order to obtain the "sharp" Gaussian simplification for image update step

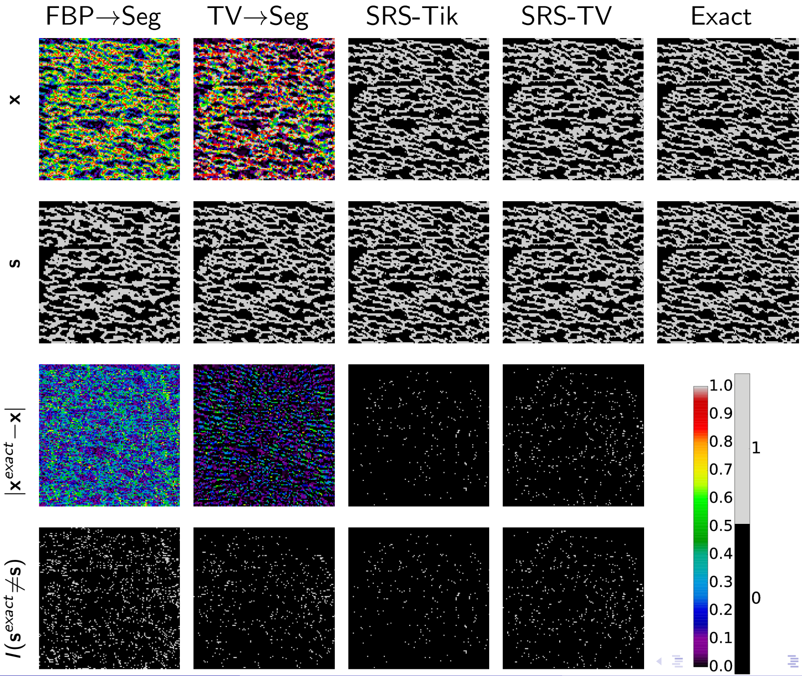
The Test Problems

- Phantoms: Shepp-Logan, binary and porous phantoms with128-by-128 pixels
- Parallel beam tomography
- Angles: 58
- Parallel rays: 181
- Underdetermined rate: $\frac{M}{N} = 0.6$
- Noise Level: 1%

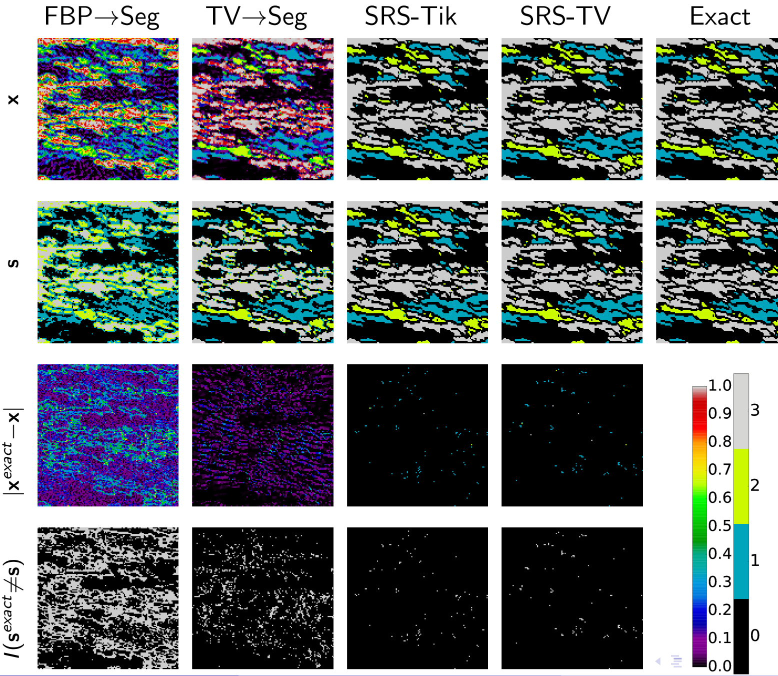


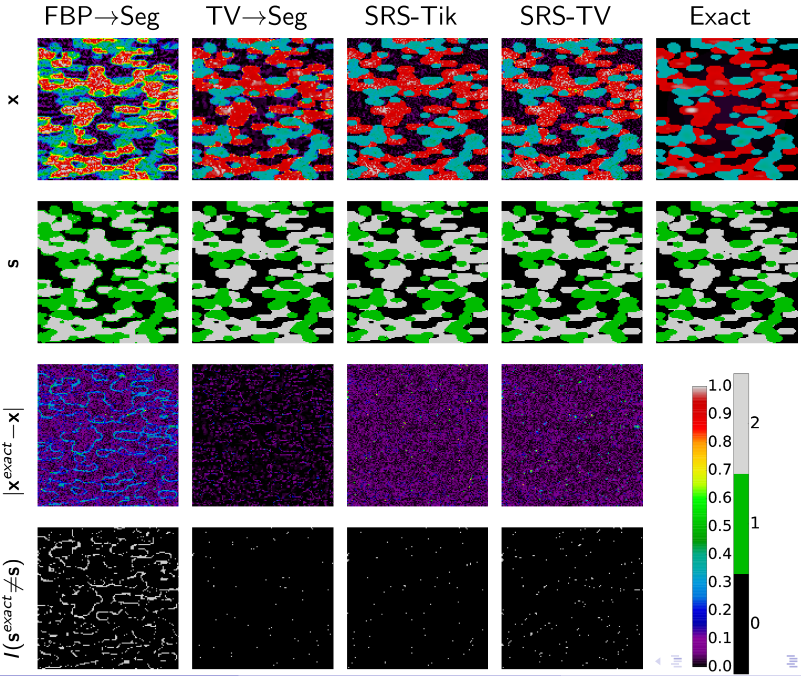


FBP→Seg
TV→Seg
SRS-Tik
SRS-TV
Exact
$\mathbf{x}$
$\mathbf{s}$
$|\mathbf{x}^{exact} - \mathbf{x}|$
$I(\mathbf{s}^{exact} \neq \mathbf{s})$
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
5
4
3
2
1
0



FBP→Seg
TV→Seg
SRS-Tik
SRS-TV
Exact
$\mathbf{x}$
$\mathbf{s}$
$|\mathbf{x}^{exact} - \mathbf{x}|$
$I(\mathbf{s}^{exact} \neq \mathbf{s})$
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1
0


FBP→Seg
TV→Seg
SRS-Tik
SRS-TV
Exact
$\mathbf{x}$
$\mathbf{s}$
$|\mathbf{x}^{exact} - \mathbf{x}|$
$I(\mathbf{s}^{exact} \neq \mathbf{s})$
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
3
2
1
0

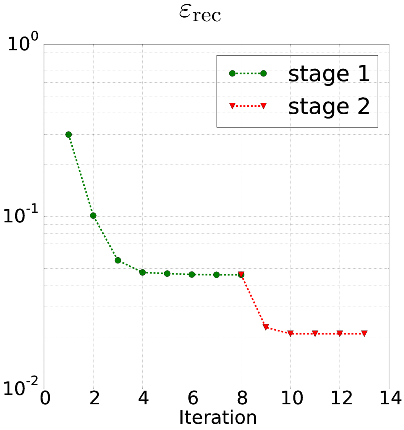

FBP→Seg
TV→Seg
SRS-Tik
SRS-TV
Exact
x
s
$|x^{exact} - x|$
$I(s^{exact} \neq s)$
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
2
1
0

Error Comparison

$$\varepsilon_{\mathrm{rec}} = \frac{\|\mathbf{x}^{exact} - \mathbf{x}\|_2}{\|\mathbf{x}^{exact}\|_2} \qquad \varepsilon_{\mathrm{seg}} = \frac{1}{N}\sum_{j=1}^{N} I(s_j^{exact} \neq s_j)$$

Test problem		FBP→Seg	TV→Seg	SRS-Tik	SRS-TV
1: Shepp-Logan	$\varepsilon_{\mathrm{rec}}$	0.34	0.038	**0.021**	0.023
	$\varepsilon_{\mathrm{seg}}$	0.056	0.0038	**0.0026**	0.0031
2: Binary	$\varepsilon_{\mathrm{rec}}$	0.46	0.33	**0.18**	0.26
	$\varepsilon_{\mathrm{seg}}$	0.096	0.035	**0.015**	0.029
3: 4-class	$\varepsilon_{\mathrm{rec}}$	0.39	0.16	**0.047**	0.055
	$\varepsilon_{\mathrm{seg}}$	0.38	0.077	**0.0057**	0.0064
4: Gray-scale	$\varepsilon_{\mathrm{rec}}$	0.24	0.082	**0.060**	0.087
	$\varepsilon_{\mathrm{seg}}$	0.095	**0.0040**	0.0047	0.0041

Reconstruction and Segmentation Error Historises



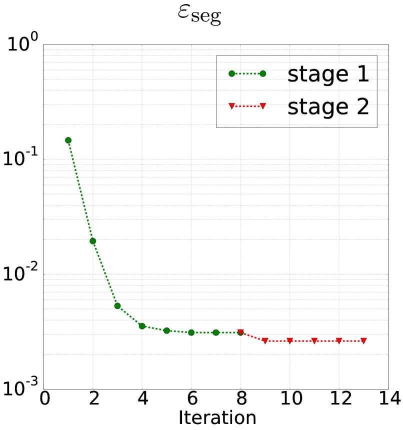
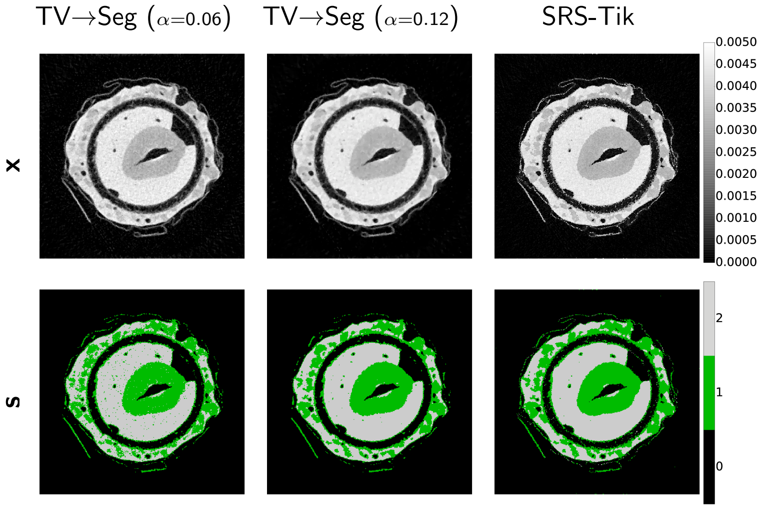


Table : The evolution of ε_{rec} and ε_{seg} during the iterations of the SRS-Tik algorithm on the Shepp-Logan phantom

Real Data


TV→Seg (α=0.06)
TV→Seg (α=0.12)
SRS-Tik
x
s
0.0050
0.0045
0.0040
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
2
1
0

Conclusion

- The classical approach with two separate steps only uses the prior knowledges on classes in the segmentation step. But simultaneous image reconstruction and segmentation method also incorporates such prior in order to increase the quality of the reconstruction
- Due to the communication of reconstruction and segmentation, we are able to produce sharp edges in the reconstruction and obtain high accuracy on the segmentation
- Future work
  - More efficient algorithms in order to deal with large-scale tomographic problems
  - Image regularization term in order to further improve the reconstruction
  - Parameter selection methods

Thank you!