Medical Image Analysis and Processing

Image Noise Filtering

Total Variation

Emad Fatemizadeh

Distance/online Course: Session 11

Date: 04 April 2021, 15th Farvardin 1400

Contents

Bayesian Interpretation

- Lets consider discrete image model:

$$v(i,j) = u(i,j) + \eta(i,j), \qquad (i,j) \in [1,M] \times [1,N]$$

- $v(i,j)$: noisy observation (data)
- $u(i,j)$: clean image
- $\eta(i,j)$: “zero-mean”, “known variance” $i.i.d$ additive noise

Bayesian Interpretation

› Assume $\eta(i,j)$ is $N(0,\sigma^2)$:

$$p\big(\eta(i,j)\big)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(\eta(i,j)-0)^2}{2\sigma^2}} \Rightarrow p\big(v(i,j)|u(i,j)\big)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{\left(v(i,j)-u(i,j)\right)^2}{2\sigma^2}}$$

› Using $i.i.d$ assumption, image joint distribution will be:

$$p(\boldsymbol{v}|\boldsymbol{u})\propto\prod_{i=1}^{M}\prod_{j=1}^{N}e^{-\frac{\left(v(i,j)-u(i,j)\right)^2}{2\sigma^2}}=e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(v(i,j)-u(i,j)\right)^2}$$

$$p(\boldsymbol{v}|\boldsymbol{u})\propto e^{-\frac{1}{2\sigma^2}\|\boldsymbol{v}-\boldsymbol{u}\|_F^2}$$

F: Frobenius Matrix norm

Bayesian Interpretation

› Using *i.i.d* assumption:

$$p(\boldsymbol{v}|\boldsymbol{u}) \propto e^{-\frac{1}{2\sigma^2}\|\boldsymbol{v}-\boldsymbol{u}\|_F^2}$$

› Now assume:

$$p(\boldsymbol{u}) \propto e^{-f(\boldsymbol{u})}$$

› Bayes rule:

$$p(\boldsymbol{u}|\boldsymbol{v})p(\boldsymbol{v}) = p(\boldsymbol{v}|\boldsymbol{u})p(\boldsymbol{u}) \propto e^{-f(\boldsymbol{u})}e^{-\frac{1}{2\sigma^2}\|\boldsymbol{v}-\boldsymbol{u}\|_F^2}$$

Bayesian Interpretation

› MAP (Maximum a Posteriori) Estimation:

$$\boldsymbol{u}^*_{MAP} = \underset{\boldsymbol{u}}{\operatorname{argmax}}\, p(\boldsymbol{u}|\boldsymbol{v}) = \underset{\boldsymbol{u}}{\operatorname{argmax}} \frac{p(\boldsymbol{v}|\boldsymbol{u})p(\boldsymbol{u})}{p(\boldsymbol{v})} = \underset{\boldsymbol{u}}{\operatorname{argmax}}\, p(\boldsymbol{v}|\boldsymbol{u})p(\boldsymbol{u})$$

› Note: $\boldsymbol{v}$ is observed data

› Thus:

$$\boldsymbol{u}^*_{MAP} = \underset{\boldsymbol{u}}{\operatorname{argmin}} -\{log\big(p(\boldsymbol{u})\big) + log\big(p(\boldsymbol{v}|\boldsymbol{u})\big)\}$$

$$\boldsymbol{u}^*_{MAP} = \underset{\boldsymbol{u}}{\operatorname{argmin}} \left\{f(\boldsymbol{u}) + \frac{1}{2\sigma^2}\|\boldsymbol{v} - \boldsymbol{u}\|_F^2\right\}$$

Bayesian Interpretation

- MAP (Maximum a Posteriori) Estimation:

$$\boldsymbol{u}^*_{MAP} = \underset{\boldsymbol{u}}{\operatorname{argmin}} \left\{ f(\boldsymbol{u}) + \frac{1}{2\sigma^2} \|v - u\|_F^2 \right\}, \quad p(\boldsymbol{u}) \propto e^{-f(\boldsymbol{u})}$$

- ROF formulation in discrete images:

$$\boldsymbol{u}^*_{TV} = \underset{\boldsymbol{u}}{\operatorname{argmin}} \left\{ TV(\boldsymbol{u}) + \frac{\lambda}{2} \|v - u\|_F^2 \right\}$$

- ROF formulation in continuous images :

$$\min_u \int_\Omega \left(|\nabla u| + \frac{\lambda}{2} (u - v)^2 \right)$$

- ROF Solution is MAP estimator for gaussian additive noise (note that we ignore normalization factor in $\propto$

Bayesian Interpretation

- Hence ROF formulation can be extended to other noise model”
- Laplace noise (ℓ_1-norm instead ℓ_2 norm):

$$\min_u \int_\Omega (|\nabla u| + \lambda |u - v|)$$

- Poisson noise (X-Ray based imaging systems):

$$\min_u \int_\Omega \left(|\nabla u| + \lambda\big(u - v log(u)\big)\right)$$

One Dimensional Discrete TV

- To keep simple, we address 1D signal
- Total variation (TV) of N-points signal:

$$x(n), 1 \leq n \leq N$$

- defined as:

$$TV(\mathbf{x}) = \sum_{i=2}^{N} |x(n) - x(n-1)|$$

- The total variation of ($\mathbf{x}$) can also be written as:

$$TV(\mathbf{x}) = \|D\mathbf{x}\|_1$$

One Dimensional Discrete TV

- The total variation of $(\mathbf{x})$ can also be written as:

$$TV(\mathbf{x}) = \|D\mathbf{x}\|_1$$

- D is a matrix of size $(N-1) \times N$:

$$\mathrm{D} = \begin{bmatrix} -1 & +1 & & \cdots & & \\ & -1 & +1 & & & \\ & & & & & \\ \vdots & & & \ddots & & \vdots \\ & & & \cdots & -1 & +1 \end{bmatrix}$$

One Dimensional Discrete TV

- Noisy image model:

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \qquad \mathbf{x}, \mathbf{y}, \mathbf{n} \in \mathbb{R}^N$$

- TV objective function:

$$J(x) = \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \|D\mathbf{x}\|_1$$

- The optimal value of the objective function is denoted

$$J(x)_* = \min_{\mathbf{x}} \{\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \|D\mathbf{x}\|_1\}$$

- The problem is complicated by the fact that the ℓ_1 norm is not differentiable.

One Dimensional Discrete TV

- Dual form is a possible approach:
- Some facts:

$$|x| = \max_{|z|\leq 1} zx\,, \qquad x \in \mathbb{R}$$

- Non-differentiability of the $|x|$ is transferred to the feasible set.
- Likewise:

$$\|\mathbf{x}\|_1 = \max_{|\mathbf{z}|.\leq 1} z^T x\,, \qquad x \in \mathbb{R}^N$$

- Condition $|\mathbf{z}|. \leq 1$is taken elementwise

One Dimensional Discrete TV

- Therefore, we can write the objective function

$$J(x) = \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \max_{|\mathbf{z}|_\cdot \le 1} \mathbf{z}^T D\mathbf{x}$$

- or:

$$J(x) = \max_{|\mathbf{z}|_\cdot \le 1} \{\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \mathbf{z}^T D\mathbf{x}\}$$

- Optimal solution is:

$$J_*(x) = \min_{\mathbf{x}} \max_{|\mathbf{z}|_\cdot \le 1} \{\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \mathbf{z}^T D\mathbf{x}\}$$

- Convex in **x** and concave in **z**,

One Dimensional Discrete TV

- we can exchange the order of the maximization and minimization (from optimization theory):

$$J_*(x) = \max_{|\mathbf{z}|.\leq 1} \min_{\mathbf{x}} \{\|\mathbf{y}-\mathbf{x}\|_2^2 + \lambda \mathbf{z}^T D\mathbf{x}\}$$

- This is dual formulation of TV denoising
- Inner minimization problem is easy to solve:

$$\frac{\partial(\|\mathbf{y}-\mathbf{x}\|_2^2 + \lambda \mathbf{z}^T D\mathbf{x})}{\partial \mathbf{x}} = -2(\mathbf{y}-\mathbf{x}) + \lambda D^T\mathbf{z} = \mathbf{0}$$

$$\Rightarrow \mathbf{x} = \mathbf{y} - \frac{\lambda}{2} D^T \mathbf{z}$$

One Dimensional Discrete TV

- Outer maximization problem becomes:

$$J_*(x) = \max_{|\mathbf{z}|_\cdot \leq 1} \left\{ \left\| \frac{\lambda}{2} D^T \mathbf{z} \right\|_2^2 + \lambda \mathbf{z}^T D \left(\mathbf{y} - \frac{\lambda}{2} D^T \mathbf{z} \right) \right\}$$

- Using: $|\mathbf{z}|_2^2 = \mathbf{z}^T \mathbf{z}$, we have

$$J_*(x) = \max_{|\mathbf{z}|_\cdot \leq 1} \left\{ -\frac{\lambda^2}{4} \mathbf{z}^T D D^T \mathbf{z} + \lambda \mathbf{z}^T D \mathbf{y} \right\}$$

- or equivalently,

$$J_*(x) = \min_{|\mathbf{z}|_\cdot \leq 1} \left\{ \mathbf{z}^T D D^T \mathbf{z} - \frac{4}{\lambda} \mathbf{z}^T D \mathbf{y} \right\}$$

One Dimensional Discrete TV

› DD^T is a matrix of size $(N-1) \times (N-1)$:

$$DD^T = \begin{bmatrix} 2 & -1 & & \cdots & \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ \vdots & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}$$

One Dimensional Discrete TV

› Let forget the condition:

$$J_*(x) = \min_{|\mathbf{z}|_\infty \le 1} \left\{ \mathbf{z}^T D D^T \mathbf{z} - \frac{4}{\lambda} \mathbf{z}^T D \mathbf{y} \right\}$$

› Setting the derivative with respect to **z** to zero gives the equation:

$$D D^T \mathbf{z} = \frac{2}{\lambda} D \mathbf{y}$$

› A large system of linear equations that does not yield a solution **z** satisfying the constraint $|\mathbf{z}|_\infty \le 1$

One Dimensional Discrete TV

- Using majorization-minimization (MM) in optimization,

$$\boldsymbol{z}^{(i+1)} = clip\left(\boldsymbol{z}^{(i)} + \frac{1}{\alpha} D\left(\frac{2}{\lambda}\boldsymbol{y} - D^T \boldsymbol{z}^{(i)}\right), 1\right), \quad \boldsymbol{z}^{(0)} = \boldsymbol{0}$$

- where α is greater than or equal to the maximum eigenvalue of DD^T.

- and clip function defined as:

$$clip(b, T) = \begin{cases} b & |b| \leq T \\ T sign(b) & |b| \geq T \end{cases}$$

One Dimensional Discrete TV

› Final Algorithm

Repeat

$$\boldsymbol{z}^{(0)} = \mathbf{0}, \boldsymbol{i} \leftarrow \mathbf{0}$$

$$\mathbf{x}^{(\boldsymbol{i}+\mathbf{1})} = \mathbf{y} - \frac{\lambda}{2} D^T \boldsymbol{z}^{(i)}$$

$$\boldsymbol{z}^{(i+1)} = clip\left(\boldsymbol{z}^{(i)} + \frac{1}{\alpha} D\left(\frac{2}{\lambda}\boldsymbol{y} - D^T \boldsymbol{z}^{(i)}\right), 1\right)$$

$$\boldsymbol{i} \leftarrow \boldsymbol{i} + \mathbf{1}$$

Until convergence

Two Dimensional Discrete TV

- Discrete Definition of TV
- Anisotropic definition:

$$TV_{aniso}(u) = \sum_{i}\sum_{k}\left|u_{i+1,j} - u_{i,j}\right| + \left|u_{i,j+1} - u_{i,j}\right|$$

- Isotropic definition:

$$TV_{iso}(u) = \sum_{i}\sum_{j}\sqrt{\left(u_{i+1,j} - u_{i,j}\right)^2 + \left(u_{i,j+1} - u_{i,j}\right)^2}$$

Two Dimensional Discrete TV

- Discrete Definition of TV
- Upwind TV:

$$TV_u(u) = \sum_i \sum_j \sqrt{\begin{array}{c}\left(u_{i,j} - u_{i+1,j}\right)_+^2 + \left(u_{i,j} - u_{i-1,j}\right)_+^2 + \\ \left(u_{i,j} - u_{i,j+1}\right)_+^2 + \left(u_{i,j} - u_{i,j-1}\right)_+^2\end{array}}$$

- Where $(a)_+ = max(a, 0)$

Two Dimensional Discrete TV

- TV Denoising Problems:
- Denoising:

$$J(x)_* = \min_{\mathbf{x}\in\mathbb{R}^{\boldsymbol{M}\times\boldsymbol{N}}} \{\|\mathbf{y}-\mathbf{x}\|_2^2 + \lambda TV(\mathbf{x})\}$$

- *Denoising+Enhancement* (Known Degradation operator **A**):

$$J(x)_* = \min_{\mathbf{x}\in\mathbb{R}^{\boldsymbol{M}\times\boldsymbol{N}}} \{\|\mathbf{A}\mathbf{y}-\mathbf{x}\|_2^2 + \lambda TV(\mathbf{x})\}$$

- It is hard to solve for $TV_i(u)$ and $TV_u(u)$

The End

› AnY QuEsTiOn?