Multiscale Total Variation (MTV) Regularization Models in Image Restoration

Michael Hintermüller

MATHEON – DFG Research Center
&
Department of Mathematics, Humboldt-Universität zu Berlin.
hint@math.hu-berlin.de

Acknowledgment:
FWF under START-grant Y305 "Interfaces and Free Boundaries" and
SFB "Mathematical Optimization and Its Appls. in Biomed. Sciences".

joint work with
Y. Dong (Helmholtz Center, Munich) and M.M. Rincon-Camacho (KFU)

Motivation - Noisy original.

Motivation - Noisy original.

Scalar regularization parameter choice: keep details vs. removing noise in homogeneous features.

Rudin-Osher-Fatemi-model.

Degradation model.

$$K\hat{u} + \eta = z \in L^2(\Omega),$$

$K \in \mathcal{L}(L^2(\Omega))$ and η related to white Gaussian noise with variance σ^2.

Constrained formulation.

$$\begin{aligned} &\text{minimize } J(u) := \textstyle\int_\Omega |Du| \text{ over } u \in BV(\Omega) \\ &\text{subject to } \textstyle\int_\Omega Ku\,dx = \int_\Omega z\,dx, \quad \int_\Omega |Ku - z|^2 dx = \sigma^2 |\Omega|, \end{aligned}$$

with $\int_\Omega |Du|$ the usual BV-seminorm and $|\Omega|$ the volume of Ω.

Unconstrained formulation.

$$\text{minimize } \int_\Omega |Du| + \frac{\lambda}{2} \int_\Omega |Ku - z|^2 dx \text{ over } u \in BV(\Omega)$$

for a suitably chosen $\lambda > 0$.

Localized ROF-model.

We are interested in...

$$\text{minimize} \int_\Omega |Du| + \frac{1}{2}\int_\Omega \lambda(x)|Ku - z|^2(x)dx \text{ over } u \in BV(\Omega), \quad (\mathrm{P}_{\mathrm{uncon}})$$

where $\lambda \in L^\infty(\Omega)$ is determined automatically.

Literature.

"Deterministic" scales.

[Strong, Chan], [Strong, Aujol, Chan].

Discrete local models + statistics.

[Almansa, Ballester, Caselles, Haro], [Facciolo, Almansa, Aujol, Caselles].

Localization.

Normalized filters. $w \in L^\infty(\Omega \times \Omega)$, $w \geq 0$ on $\Omega \times \Omega$ with

$$\int_\Omega \int_\Omega w(x,y)dydx = 1 \quad \text{and} \quad \int_\Omega \int_\Omega w(x,y)\phi^2(y)dy\, dx \geq \epsilon \|\phi\|^2_{L^2(\Omega)}$$

for all $\phi \in L^2(\Omega)$ for some $\epsilon > 0$ (independent of ϕ). May use mean, Gaussian, or Wiener filter.

Localized residual (variance).

$$S(u)(x) := \int_\Omega w(x,y)(Ku - z)^2(y)\, dy.$$

Localized ROF-model – constrained form.

$$\begin{aligned} &\text{minimize } J(u) = \textstyle\int_\Omega |Du| \quad \text{over } u \in BV(\Omega) \\ &\text{subject to } S(u) - \sigma^2 \leq 0 \quad \text{a.e. in } \Omega. \end{aligned} \tag{P}$$

Existence & multipliers.

Assumptions on K.

- K does not annihilate constant functions.
- $K \cdot \mathbf{1} = \mathbf{1}$.

Thm. [coercivity for existence]. $\|u\|_{BV} \to +\infty$ implies $\mathcal{J}(u) \to +\infty$ with

$$\mathcal{J}(u) = J(u) + \int_\Omega \int_\Omega w(x,y)(Ku - z)^2(y)\, dy\, dx.$$

Further properties.

Assumption. Let $u_1, u_2 \in BV(\Omega)$ denote two solutions of (P) with $u_1 \neq u_2$. If there exist $\delta > 0$ and $\Omega_\delta \subset \Omega$ with $|\Omega_\delta| > 0$ such that

$$\left(\frac{1}{2}K(u_1 + u_2) - z\right)^2 \leq \frac{1}{2}\left((Ku_1 - z)^2 + (Ku_2 - z)^2\right) - \delta \quad \text{a.e. in } \Omega_\delta$$

then there exists $\epsilon_\delta > 0$ such that

$$\int_\Omega w(x,y)\left(\frac{1}{2}K(u_1 + u_2) - z\right)^2(y)dy \leq \sigma^2 - \epsilon_\delta \qquad \text{for almost all } x \in \Omega.$$

Existence & multipliers.

Then...

$$\inf_{c\in\mathbb{R}} \int_\Omega w(x,y)(c-z)^2(y)dy > \sigma^2 \text{ a.e. in } \Omega$$

and for every solution $\tilde{u}$ of (P) $K\tilde{u}$ has the same value.

Multipliers. Define the penalty problem

$$\text{minimize} \quad \mathcal{F}_\gamma(u) := J(u) + \frac{\gamma}{2}\int_\Omega \max(S(u)-\sigma^2,0)^2 dx \quad \text{over} \quad u \in BV(\Omega), \tag{P_{pen}}$$

with $\gamma > 0$ a penalty parameter.

Thm. [consistency]. Problem (P_{pen}) admits a solution $u_\gamma \in BV(\Omega)$ for every $\gamma > 0$, and for $\gamma \to +\infty$ $\{u_\gamma\}$ converges along a subsequence weakly in $L^2(\Omega)$ to a solution of (P).

Further property.

- $\|\max(S(u_\gamma)-\sigma^2,0)\|_{L^2(\Omega)} = \mathcal{O}(1/\sqrt{\gamma})$.

Existence & multipliers.

Define

$$\begin{aligned} \lambda_\gamma^\circ &:= \gamma \max(S(u_\gamma) - \sigma^2, 0), \\ \lambda_\gamma &:= \int_\Omega w(x,y) \lambda_\gamma^\circ(x)\, dx. \end{aligned}$$

Note: λ_γ related to Fréchet-derivative of penalty term in ($\mathrm{P}_{\mathrm{pen}}$).

Thm. [multipliers]. Under appropriate assumptions, $\exists\ \tilde{\lambda} \in L^\infty(\Omega)$, a bounded Borel measure $\tilde{\lambda}^\circ$ and a subsequence $\{\gamma_{n_k}\}$ such that:

(i) $\int_\Omega \lambda_{\gamma_{n_k}} f\, dx \to \int_\Omega \tilde{\lambda} f\, dx$ for all $f \in L^1(\Omega)$ and $\tilde{\lambda} \geq 0$ a.e. in Ω.

(ii) There exists $j(\tilde{u}) \in \partial J(\tilde{u})$ such that

$$\langle j(\tilde{u}), v \rangle_{BV(\Omega)^*, BV(\Omega)} + 2 \int_\Omega (K^* \tilde{\lambda}(K\tilde{u} - z)) v\, dx = 0 \quad \text{for all } v \in BV(\Omega).$$

(iii) $\int_\Omega \varphi \lambda_{\gamma_{n_k}}^\circ\, dx \to \int_\Omega \varphi d\tilde{\lambda}^\circ$ for all $\varphi \in C(\bar{\Omega})$, $\tilde{\lambda}^\circ \geq 0$ and $\int_\Omega \lambda_{\gamma_n}^\circ (S(u_{\gamma_n}) - \sigma^2)\, dx \to 0$.

Existence & multipliers.

Under additional regularity, (iii) equivalent to

$$\tilde{\lambda}^\circ \geq 0 \text{ a.e. in } \Omega, \quad \tilde{\lambda}^\circ = \tilde{\lambda}^\circ + \rho \max(S(\tilde{u}) - \sigma^2, 0)$$

for arbitrary and fixed $\rho > 0$.

Local variance estimator.

Continue in discrete terms...

Local window.

$$\Omega_{i,j}^{\omega} = \left\{(s+i, t+j) : -\frac{\omega-1}{2} \leq s, t \leq \frac{\omega-1}{2}\right\}.$$

Local residual/variance (mean filter).

$$S_{i,j}^{\omega} := \frac{1}{\omega^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} \left(z_{s,t} - (Ku)_{s,t}\right)^2 = \frac{1}{\omega^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} \left(r_{s,t}\right)^2 \qquad \text{(LVE)}$$

with $r = z - Ku \in \mathbb{R}^{m^2}$.

Suitable upper bound on local variance. Random variable

$$T_{i,j}^{\omega} = \frac{1}{\sigma^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} (\eta_{s,t})^2$$

has χ^2-distribution with ω^2 degrees of freedom, i.e. $T_{i,j}^{\omega} \sim \chi^2_{\omega^2}$.

Local variance estimator.

Brief discussion...

If $u = \hat{u}$ satisfies $\eta = z - K\hat{u}$, then

$$S_{i,j}^{\omega} = \frac{1}{\omega^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} (z_{s,t} - (K\hat{u})_{s,t})^2 = \frac{1}{\omega^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} (\eta_{s,t})^2 = \frac{\sigma^2}{\omega^2} T_{i,j}^{\omega}.$$

On the contrary, if u is an over-smoothed restored image, then the residual $z - Ku$ contains details, and we expect in detail regions

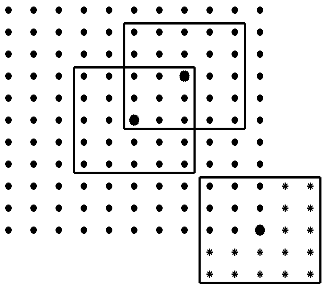
$$S_{i,j}^{\omega} = \frac{1}{\omega^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} (z_{s,t} - (Ku)_{s,t})^2 > \frac{1}{\omega^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} (\eta_{s,t})^2 = \frac{\sigma^2}{\omega^2} T_{i,j}^{\omega}.$$

Thus. Search for bound B such that $S_{i,j}^{\omega} > B$ for some pixel (i,j) implies that details left in residual in neighborhood of this pixel.

For ease of notation: $T_k^{\omega} := T_{i,j}^{\omega}$ with $k = i + (m-1)j$ for $i, j = 1, \dots m$.

Local variance estimator.

Graphical sketch...



$$S_{i,j}^{\omega} := \frac{1}{\omega^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} \left(z_{s,t} - (Ku)_{s,t}\right)^2$$

Distribution of max.

Bound B depends on size of the image ($m \times m$) and on size of window ($\omega \times \omega$) and should be satisfied by "all" local residuals. Thus,

$$B^{\omega,m} := \frac{\sigma^2}{\omega^2}\mathfrak{E}(\max_{k=1,\ldots,m^2} T_k^\omega),$$

with $\mathfrak{E}$ the expected value of a random variable.

Distribution of max. We have T_k^ω, $k = 1, \ldots, N := m^2$. Let $\mathfrak{f}$ be the χ^2-distribution with ω^2 degrees of freedom, and

$$\mathfrak{F}(T) = \int_{-\infty}^{T} \mathfrak{f}(z)dz. \qquad \text{(cumulat. distrib. fctn.)}$$

Maximum value of N observations distributed along $\mathfrak{f}$ denoted by $T_{\max}$.

Goal. Describe distribution $\mathfrak{f}_{\max}$ of maximum value.

Distribution of max.

Distribution [Gumbel; 1935].

$$\mathfrak{f}_{\max}[y(T_{\max})] = N\mathfrak{f}(T_{\text{dom}})e^{-y(T_{\max})-e^{-y(T_{\max})}}, \tag{1}$$

with

- $y(T) = N\mathfrak{f}(T_{\text{dom}})(T - T_{\text{dom}})$ the standardization of T.
- T_{dom} so-called dominant value

$$\mathfrak{F}(T_{\text{dom}}) = 1 - \frac{1}{N}.$$

Cumulative distribution function.

$$\mathfrak{F}_{\max}(T) = P(T_{\max} \leq T) = e^{-e^{-y(T)}}.$$

Expected value and standard deviation of the standardized variable $y(T_{\max})$.

$$\mathfrak{E}(y(T_{\max})) = \kappa \quad \text{and} \quad \mathfrak{d}(y(T_{\max})) = \frac{\pi}{\sqrt{6}},$$

respectively, where $\kappa = 0.577215$ is the Euler-Mascheroni constant

.

Distribution of max + choice of bound.

Expected value and standard deviation of $T_{\max}$.

$$\mathfrak{E}(T_{\max}) = T_{\mathrm{dom}} + \frac{\kappa}{\beta_{\max}} \quad \text{and} \quad \mathfrak{d}(T_{\max}) = \frac{\pi}{\beta_{\max}\sqrt{6}} \text{ with } \beta_{\max} = N\mathfrak{f}_{\max}(T_{\mathrm{dom}}).$$

Confidence levels. Probability that the maximum value is below or equal $\mathfrak{E}(T_{\max})$ is

$$P(T_{\max} \leq \mathfrak{E}(T_{\max})) = e^{-e^{-y(\mathfrak{E}(T_{\max}))}} = e^{-e^{-\kappa}} = 0.57037$$

and that the maximum value is not larger than $\mathfrak{E}(T_{\max}) + \mathfrak{d}(T_{\max})$ is

$$P(T_{\max} \leq \mathfrak{E}(T_{\max}) + \mathfrak{d}(T_{\max})) = e^{-e^{-y(\mathfrak{E}(T_{\max})+\mathfrak{d}(T_{\max}))}} = e^{-e^{-\kappa-\frac{\pi}{\sqrt{6}}}} = 0.85580.$$

Choice of bound.

$$B^{\omega,m} := \frac{\sigma^2}{\omega^2}\left(\mathfrak{E}(T_{\max}) + \mathfrak{d}(T_{\max})\right) \qquad (> \sigma^2).$$

Selection of λ.

Decision based on $S^{\omega}_{i,j}$.

Noise only – no details, if

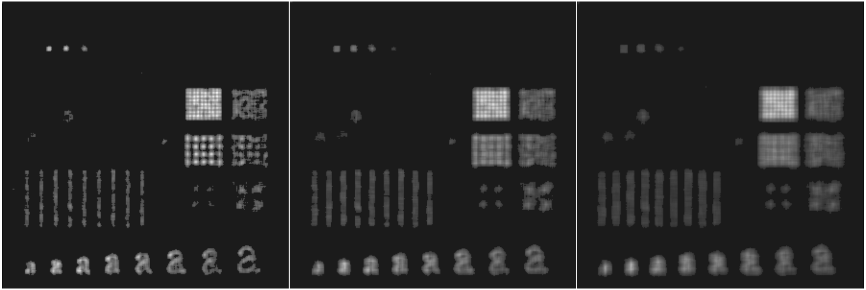
$$S^{\omega}_{i,j} \in [0, B^{\omega,m})\,.$$

Rule: Keep current λ-value at such pixels.

Modified local variance estimator $\tilde{S}^{\omega}$. Defined by

$$\tilde{S}^{\omega}_{i,j} := \begin{cases} S^{\omega}_{i,j} & \text{if } S^{\omega}_{i,j} \geq B^{\omega,m}, \\ \sigma^2 & \text{otherwise}\,. \end{cases}$$

Selection of λ.



$\tilde{S}$ for different window sizes: $\tilde{S}^5$ (left), $\tilde{S}^7$ (middle), $\tilde{S}^9$ (right).

Selection of λ and MTV.

Update rule.

$$\begin{aligned}(\tilde{\lambda}_{k+1})_{i,j} &:= (\tilde{\lambda}_k)_{i,j} + \rho \max((\tilde{S}_k^\omega)_{i,j} - \sigma^2, 0) = (\tilde{\lambda}_k)_{i,j} + \rho((\tilde{S}_k^\omega)_{i,j} - \sigma^2), \\ (\lambda_{k+1})_{i,j} &= \frac{1}{\omega^2} \sum_{(s,t)\in\Omega_{i,j}^\omega} (\tilde{\lambda}_{k+1})_{s,t},\end{aligned}$$

where $\rho > 0$.

Basic M(ultiscale)**T**(total)**V**(variation)**-Algorithm**.

1: Initialize $\lambda_0 := \tilde{\lambda}_0 \in \mathbb{R}_+^{m\times m}$ and set $k := 0$.

2: Let u_k denote the solution of the discrete version of the minimization problem ($\mathrm{P}_{\mathrm{uncon}}$) with discrete $\lambda = 2\lambda_k$.

3: Update λ_{k+1} based on u_k and the rule above.

4: Stop, or set $k := k + 1$ and return to step 2.

Hierarchical decomposition.

Generalize work by [Tadmor, Nezzar, Vese] (TNV-method)...

Considering dyadic scales, hierarchical decomposition works as follows:

(i) Choose $\lambda_0 > 0$, $\lambda_0 \in L^\infty(\Omega)$ and compute

$$u_0 := \arg \min_{u \in BV(\Omega)} \int_\Omega |Du| + \frac{1}{2} \int_\Omega \lambda_0 (Ku - z)^2 dx.$$

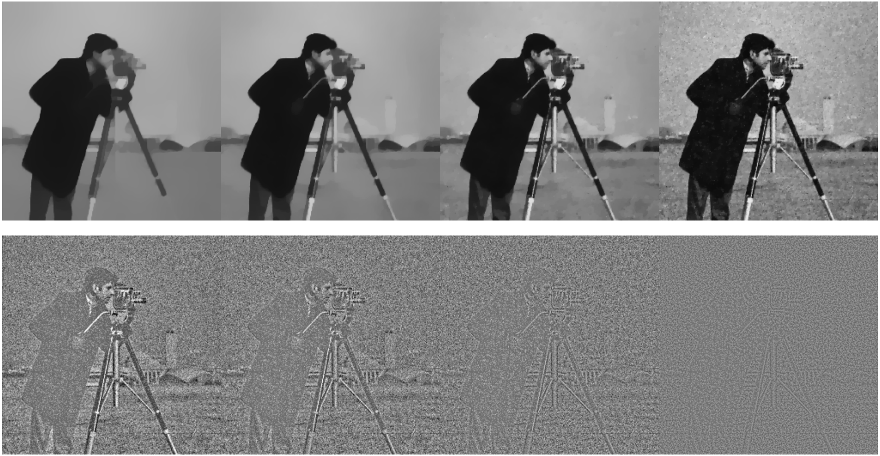
(ii) For $j = 0, 1, 2 \ldots$ set $\lambda_j = 2^j \lambda_0$ and $v_j = z - Ku_j$. Then compute

$$\hat{u}_j := \arg \min_{u \in BV(\Omega)} \int_\Omega |Du| + \frac{1}{2} \int_\Omega \lambda_{j+1} (Ku - v_j)^2 dx, \quad u_{j+1} := u_j + \hat{u}_j.$$

Note: for simplicity u_0 and $\hat{u}_j$, $j = 0, 1, \ldots$, assumed unique.

Hierarchical decomposition & its performance.

Iterations 1,2,3,4... (reconstruction / residual)



Hierarchical decomp. ∞ MTV $\Rightarrow$ Spatially Adapted TV.

SA-TV-Algorithm.

1: Initialize $u_0 = 0 \in \mathbb{R}^{m\times m}$, $\lambda_0 = \tilde{\lambda}_0 \in \mathbb{R}_+^{m\times m}$ (small) and set $k := 0$.

2: If $k = 0$, compute u_0 from discrete version of ROF-problem, else compute $v_k = z - Ku_k$ and solve discrete version of

$$\min_{u\in BV(\Omega)} \int_\Omega |Du| + \frac{1}{2}\int_\Omega \lambda(Ku - v_k)^2 dx$$

with discretization of λ equal to λ_k. Let $\hat{u}_k$ denote corresponding solution.

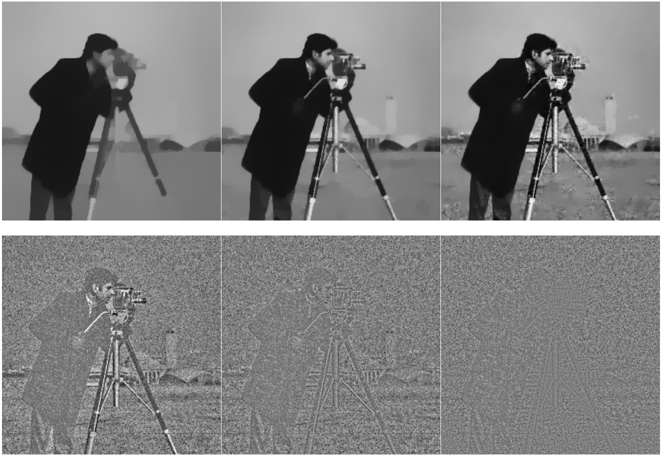
3: Update $u_{k+1} = u_k + \hat{u}_k$.

4: Based on u_{k+1} and λ-update rule, update λ_{k+1}.

5: Stop, or set $k := k + 1$ and return to step 2.

SA-TV vs. TNV.

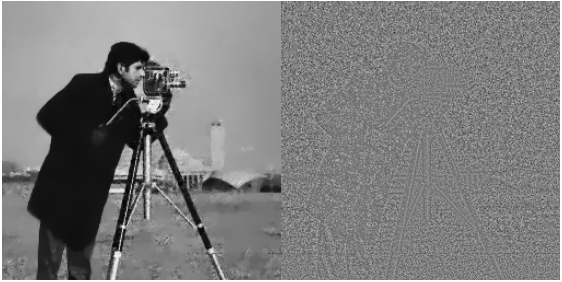
Iterations 1,2,3. (reconstruction / residual)



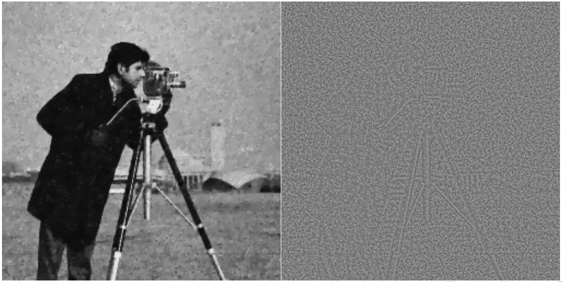
SA-TV vs. TNV.

Comparing the results.

SA-TV



TNV



Semismooth-Newton-TV-solver in a nutshell.

Fenchel-dual of regularized primal problem

$$\max_{\substack{\vec{p} \in \mathbf{L}^2(\Omega) \\ |\vec{p}(x)| \leq 1 \ \text{ a.e. in } \Omega}} \quad -\frac{1}{2}|||K^*\lambda z - \mathrm{div}\vec{p}|||^2_{H^{-1}} + \frac{1}{2}\int_\Omega \lambda z^2 \, dx,$$

with $|||v|||^2_{H^{-1}} = \langle H_{\mu,K} v, v \rangle_{H_0^1, H^{-1}}$ and $H_{\mu,K} = (K^*\lambda K - \mu\triangle)^{-1}$; $\mu > 0$.

For uniqueness of dual variable.

$$\max_{\substack{\vec{p} \in \mathbf{L}^2(\Omega) \\ |\vec{p}(x)| \leq 1 \ \text{ a.e. in } \Omega}} \quad -\frac{1}{2}|||K^*\lambda z - \mathrm{div}\vec{p}|||^2_{H^{-1}} + \frac{1}{2}\int_\Omega \lambda z^2 \, dx - \frac{\beta}{2}\int_\Omega \|\vec{p}\|^2_{\mathbf{L}^2}.$$

where $\beta > 0$.

Semismooth Newton TV-solver in a nutshell.

Nonsmooth optimality system.

$$-\mu\triangle\bar{u} + K^*\lambda K\bar{u} - \operatorname{div}\bar{\vec{p}} = K^*\lambda z \quad \text{in } H^{-1}(\Omega),$$
$$\max(\beta, |\nabla\bar{u}|_2)\bar{\vec{p}} - \nabla\bar{u} = 0 \quad \text{in } \mathbf{L}^2(\Omega).$$

In practice: $\mu = 0$.

Solved efficiently by **semismooth Newton method**:

- inexact Newton-steps (typically 10-15 iterations);
- only matvecs (3-5 CG steps / Newton iteration);
- may be globalized by line search (not necessary in practice);
- local superlinear convergence;
- [H., Stadler; SISC '06]; [Dong, H., Neri; SIIMS – SIAM Imag. Sci. '09]; [H., Rincon-Camacho; Inv. Prob. '10].

Comparison of methods.

(TV - optim. scalar λ;
PSNR=27.04)

(Bregman it.); PSNR = 27.34

(TNV; PSNR = 26.96)

(SA-TV; PSNR = 27.90)

Comparison of methods.

CPU-ratios and the number of iterations for different methods.

	"Cameraman"		"Barbara"	
	CPU-ratio	k	CPU-ratio	k
Bregman iteration	1.56	4	1.51	4
TNV-method	1.3	4	1.16	4
Basic MTV-method	3.53	22	4.45	37
SA-TV-method	1	3	1	3

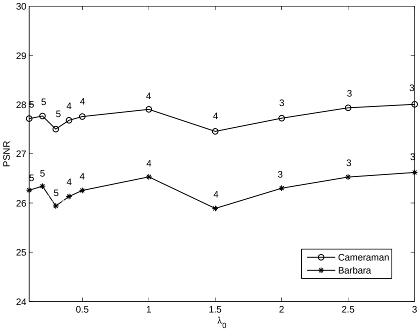
Data

SA-TV

Dependence on initialization of λ.

Further quantitative assessments...

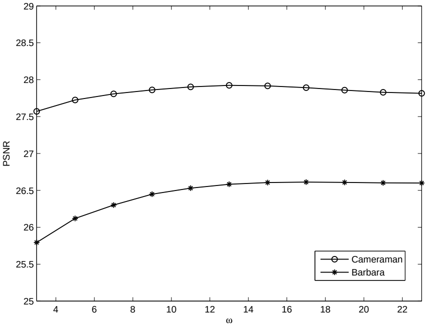
Stability w.r.t. initialization.



PSNR for images restored by our method for different initial $\bar{\lambda}^0$.

Dependence on window size.

Stability w.r.t. window size for localization.



PSNR for results obtained by our method with different ω.

Dependence on window size.

Reconstructions for different window sizes (a) $\omega = 3$, *(b)* $\omega = 7$, *(c)* $\omega = 13$, *(d)* $\omega = 17$.

(a)

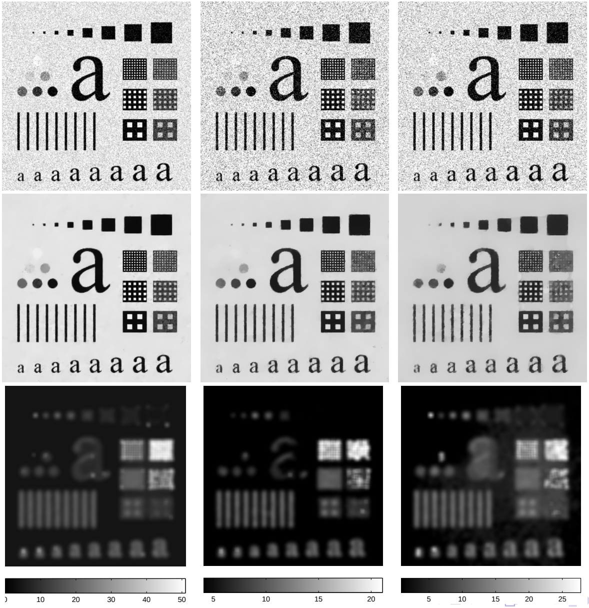
(b)

(c)

(d)

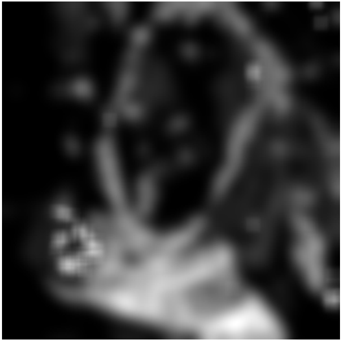
Dependence on noise level.

Different noise levels (a) $\sigma = 0.1$, *(b)* $\sigma = 0.2$, *(c)* $\sigma = 0.3$.



Final λ.

(Cameraman)



(Barbara)

Extensions.

- L1-TV.

  M.H., M.M. Rincon-Camacho: *Expected absolute value estimators for a spatially adapted regularization parameter choice rule in L^1-TV-based image restoration*, Inverse Problems, 2010.

$$\text{minimize} \int_\Omega \lambda(x)|Ku - z|(x)dx + \int_\Omega |Du| \quad \text{over } u \in BV(\Omega).$$

  - Random-valued impulse noise or salt-and-pepper noise.
  - Localization: Use expected absolute value

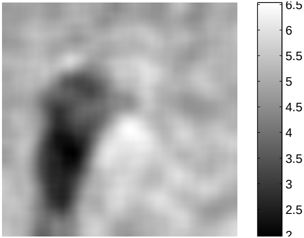
$$E(|\eta|) = \int_{-\infty}^{\infty} |x_\eta| f(x_\eta) dx_\eta,$$

    with f the associated probability density function.

Extensions.

Salt-and-pepper noise (noise level 0.5) with Gaussian blur (9×9 pixel window).

(a)



(b)

(c)

Extensions.

- Color images.

  Y. Dong, M.H., M.M. Rincon-Camacho: *A multi-scale vectorial L^τ-TV framework for color image restoration*, International Journal of Computer Vision, 2010.

$$\text{minimize } \frac{1}{\tau}\int_\Omega \lambda(x)|K\mathbf{u}-\mathbf{z}|_\tau^\tau(x)dx + \int_\Omega |D\mathbf{u}| \quad \text{over } \mathbf{u}\in\mathbf{BV}(\Omega)$$

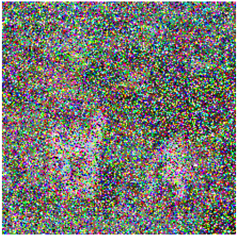
with

$$\int_\Omega |D\mathbf{u}| := \sup\left\{\mathbf{u}\cdot\operatorname{div}\underline{\mathbf{v}} : \underline{\mathbf{v}}\in C_c^1(\Omega;\mathbb{R}^{M\times 2}); |\underline{\mathbf{v}}|_F \le 1 \text{ in } \Omega\right\}.$$

$M\ldots$ number of color channels; $\tau\in[1,2]$.

Extensions.

Salt-and-pepper noise (noise level 0.6) with cross-channel average, Gaussian, and motion blur.



(a)

(b)

(c)

Extensions.

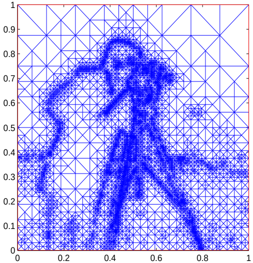
Framework suitable for more general inverse problems:

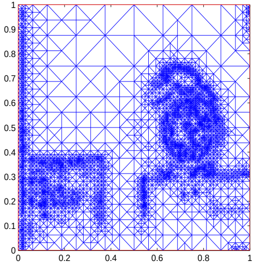
$$\begin{array}{ll} \text{minimize} & J(u) \\ \text{s.t.} & S^\omega(R(u)) \leq m(\eta) \text{ a.e.} \end{array}$$

where J encodes regularization, S^ω localization, R a residual, and m a statistical measure compatible with S^ω.

Extensions.

- A posteriori error estimates for adaptive finite element discretization (AFEM) with M.M. Rincon-Camacho.





Summary.

- Localized regularization in L^2-TV models (localizing residuals through filters).
- Relation: Regularization parameter $\leftrightarrow$ Lagrange multiplier.
- Local variance estimator $+$ statistics (distribution of maximum of M random variables).
- Hierarchical decomposition.
- Semi-smooth Newton solver for L^2-TV (and also L^1-TV).
- Extensions
  - L^1-TV.
  - Color-TV.
  - Generalization potential.
  - AFEM.