SSIP 2013 – Summer School on Image Processing, July 4–13, Veszprém, Hungary

Regularized image denoising

Joakim Lindblad
joakim@cb.uu.se

SSIP 2013, Veszprém

Problem formulation

Consider a scene that is imaged by some imaging device.

Due to imperfections in the imaging process, the true/perfect image of the scene is degraded by noise.

u

f

Task: Recover the original image u from the observed image f.

Model

Assume a greyscale 2D digital image $u = [u_{ij}]$.

$$u : \Omega \subset \mathbb{Z}^2 \to [0, 1]$$

The image u is degraded by uncorrelated(white) additive noise η from a normal distribution with zero mean and variance σ^2.

This gives us the observed image $f = u + \eta$.

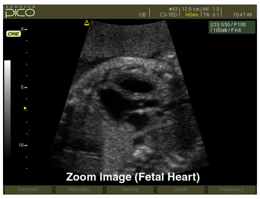
Task: To estimate $\hat{u}$ given the noisy image f.

Ill posed problem.
We need to rely on some a priori information about u and η.

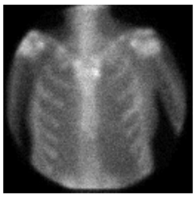
- Behaviour of the noise, size, “colour”, dependence on data. . .
- Spatial (and temporal) relation between image parts.

Noise

Many types and origins of image noise. . .



Ultrasound image of fetal heart



Emission image from nuclear imaging

Very important to have a correct model of the noise.

For simplicity, we start with Additive Gaussian White Noise.

Quantitative performance measure

Since we are doing science, we need some objective (quantitative) measure of performance.

One useful measure for this task is **Signal to Noise Ratio** (SNR)

$$\begin{aligned} SNR_{dB}(u,\hat{u}) &= 20\log_{10}\frac{\text{Amplitude signal}}{\text{Amplitude noise}} \\ &= 10\log_{10}\frac{\text{var}(u)}{\text{MSE}(\hat{u})} = 10\log_{10}\frac{\sum(u_{ij}-\bar{u})^2}{\sum(\hat{u}_{ij}-u_{ij})^2} \end{aligned}$$

Industry standard - ISO 12232 - (defined from ISO film speed equivalent);

- SNR: 32 dB = excellent image quality and
- SNR: 20 dB = acceptable image quality.

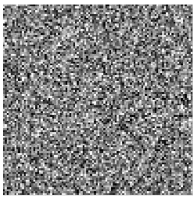
Qualitative performance measures

- **Visible structure in the estimation error**
  The magnitude of the estimation error $\epsilon = \hat{u} - u$ should of course be as small as possible.
  Observing structure in ϵ can help identifying systematic errors of the method.
- **Visible structure in the estimated noise**
  For the uncorrelated additive model, the estimated noise $\hat{\eta} = f - \hat{u}$ should have no structure.
  Note that $\hat{\eta}$ does not rely on knowledge of u, and is thus available for the method.

Cameraman

Additive Gaussian White Noise

u



$\eta = N(0, \sigma)$, $\sigma = 0.1$

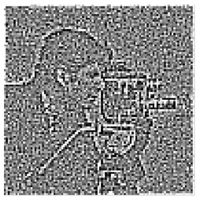
$f = u + \eta$

Initial SNR: 9.75dB

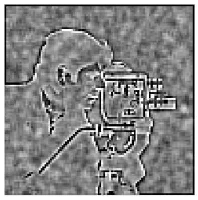
Mean 3×3 filter

u

f



$\hat{\eta} = \hat{u} - f, \hat{\sigma} = 0.122$



$\hat{u} - u$

Mean 3×3 – SNR: 10.89dB, $\hat{\sigma}$: 0.122

Too much blur! $\hat{\sigma} > \sigma$

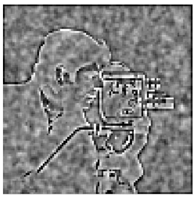
Gaussian filter, $\sigma_g = 0.40$

Tuned to give $\hat{\sigma} = \sigma$

u

$\hat{u}$

$\hat{\eta} = \hat{u} - f, \hat{\sigma} = 0.101$



$\hat{u} - u$

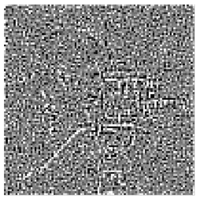
Gaussian – SNR: 12.28dB, $\hat{\sigma}$: 0.101

Good on flat regions in the image. Smooths sharp edges.

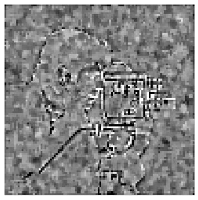
Median 3×3

u

$\hat{u}$



$\hat{\eta} = \hat{u} - f, \hat{\sigma} = 0.115$



$\hat{u} - u$

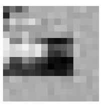
Median 3×3 – SNR: 11.77dB, $\hat{\sigma}$: 0.115

Too smooth! $\hat{\sigma} > \sigma$

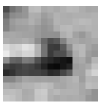
Center Weighted Median

Tuned to give $\hat{\sigma} = \sigma$

u

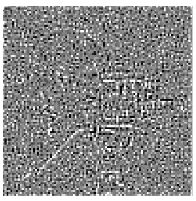


u

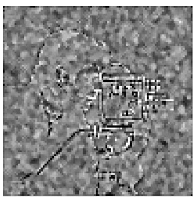


$\hat{u}$

$\hat{u}$



$\hat{\eta} = \hat{u} - f, \hat{\sigma} = 0.097$



$\hat{u} - u$

Weighted Median – SNR: 12.85dB, $\hat{\sigma}$: 0.097

Not too bad. Rounds of corners.

Maximum aposteriori estimator

The Bayesian approach for solving inverse problems

- Model for the imaged data: $f = u + \eta$
- An a priori probability density for "perfect" original signals $P(u)$, i.e., a model for the pure data in itself (not knowing the particular image).

Then the aposteriori probability for u knowing f is computed from Bayes' rule

$$P(u|f) = \frac{P(f|u)}{P(f)} P(u)$$

For additive noise, the conditional density $P(f|u)$ is just the density of the noise

$$P(f|u) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\sum|f_{ij}-u_{ij}|^2}$$

The MAP estimate then becomes

$$\hat{u} = \arg\max_u \; P(u|f) = \arg\max_u \; P(u)P(f|u) = \arg\min_u \; -\log P(u) - \log P(f|u)$$

$$= \arg\min_u \; -\log P(u) + \frac{1}{2\sigma^2}\sum_{ij\in\Omega}|f_{ij} - u_{ij}|^2$$

Maximum aposteriori estimator

Assuming that a priori probability P(u) is Markovian, the discrete MAP estimate can be addressed using theory of Markov Random Fields. . .

We will take another path today.

The variational approach

Forget about the discrete nature of our images and pretend that our images are instead continuous functions on $\Omega \subset \mathbb{R}^2$.

$$f : \Omega \subset \mathbb{R}^2 \to [0, 1]$$

The minimization problem from two slides back can then be written

$$\hat{u} = \underset{u \in L^2(\Omega)}{\arg\min}\ \mu F(u) + \frac{1}{2\sigma^2} \int_\Omega (f(x) - u(x))^2\, dx$$

where F is a functional corresponding to the a priori probability density $P(u)$ and where μ is a weight balancing the two terms.

Which regularization to pick?

$$\hat{u} = \underset{u \in L^2(\Omega)}{\arg\min}\ \mu F(u) + \frac{1}{2\sigma^2} \int_\Omega (f(x) - u(x))^2\, dx$$

F can be seen as a regularization term (i.e. additional information introduced for solving an ill posed problem). The most common regularization, Tikhonov regularization, will usually consider a quadratic F. The advantage of such a choice is that the corresponding problem to solve (Euler-Lagrange eq.) becomes linear.

Since we wish to reduce noise, it seems natural to impose a smoothing term. That could, e.g., be to minimize the gradient magnitude. For a quadratic F we have

$$F(u) = \int_\Omega |\nabla u|^2\, dx$$

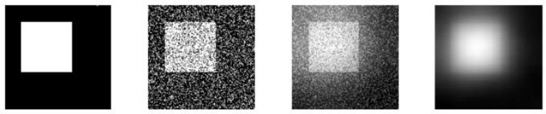
and the Euler-Lagrange equation to solve becomes

$$-\mu \Delta u + u - g = 0\,.$$

Quadratic regularization

$$\hat{u} = \underset{u}{\arg\min} \int |\nabla u|^2 + \frac{\lambda}{2} \int_\Omega (f(x) - u(x))^2\, dx$$

However, it turns that the spatial regularization (smoothing) is too strong with this choice of F, and the resulting image becomes too blurred. In particular, edges are not preserved.



We need something else!

Mumford and Shah model

Suggested by D. Mumford and J. Shah in 1989
to essentially turn off the smoothing whenever we have an edge.

Use some kind of edge-indicator function K
and optimize the following problem

$$\hat{u} = \underset{u}{\arg\min}\ \mu \int_{\Omega \setminus K} |\nabla u|^2\, dx + \nu\, \text{length}(K) + \int_\Omega (f(x) - u(x))^2\, dx$$

However, besides being very difficult to analyze mathematically, this approach is also very complicated numerically and requires solving a difficult non-convex problem with, in general, no good initial guess for K.

The ROF TV-model

Edge preserving smoothing

Replacing the square norm on gradient magnitude $|\nabla u|^2$ with the **Total variation** as a regularizer turns out to give a good balance between the two contradictory aims; smoothing of noise and preservation of edges.

$$F(u) = \int_\Omega |\nabla u|\, dx$$

The Total Variation was introduced for image denoising and reconstruction in the end of the 80's (or 1992) by Rudin, Osher, and Fatemi (ROF).

$$\hat{u} = \underset{u \in L^2(\Omega)}{\arg\min} \int |\nabla u| + \frac{\lambda}{2} \int_\Omega (f(x) - u(x))^2\, dx \tag{1}$$

with $1/\lambda = \mu\sigma^2$.

Optimization

The unconstrained problem (1) is strictly convex, and, hence, admits a unique minimum. This facilitates the creation of fast and robust methods for its solution, and over the years very many such have been proposed.

But first, let's see how it performs...

Cameraman

Results of simple methods

f $\hat{u}$ Gaussian $\hat{u}$ CWMF

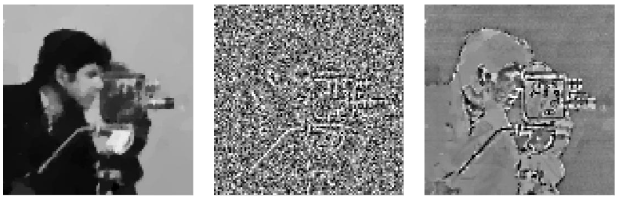
Initial SNR: 9.75dB

Method	SNR[dB]	$\hat{\sigma}$
Gaussian	12.28	0.101
Weighted Median	12.85	0.097

Total Variation Denoising

Tuned to give $\hat{\sigma} = \sigma$

u

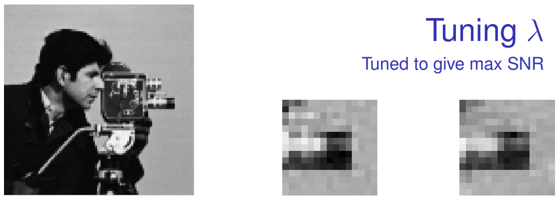


$\hat{u}$ $\hat{\eta} = \hat{u} - f, \hat{\sigma} = 0.100$ $\hat{u} - u$

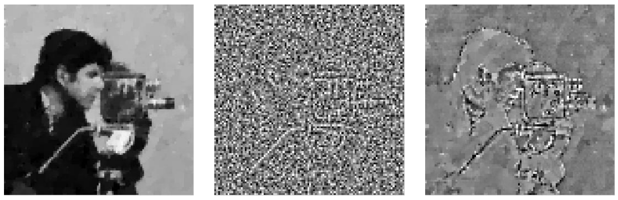
ROF TV denoising, $\lambda = 10$ – SNR: 15.21dB, $\hat{\sigma}$: 0.100

Tuning λ

Tuned to give max SNR



u u $\hat{u}$



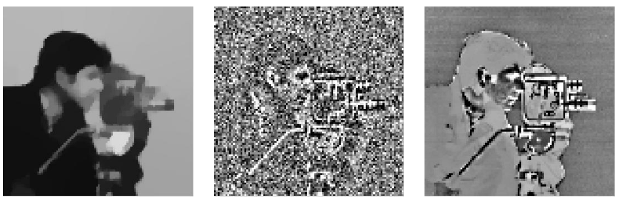
$\hat{u}$ $\hat{\eta} = \hat{u} - f, \hat{\sigma} = 0.008$ $\hat{u} - u$

ROF TV denoising, $\lambda = 14.3$ – SNR: 16.03dB, $\hat{\sigma}$: 0.088

Varying λ

Strong regularization (small λ)

u



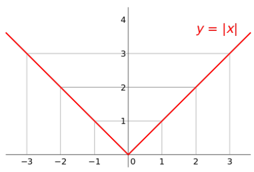
$\hat{u}$ $\hat{\eta} = \hat{u} - f, \hat{\sigma} = 0.100$ $\hat{u} - u$

ROF TV denoising, $\lambda = 5$ – SNR: 12.29dB, $\hat{\sigma}$: 0.118

Optimization methods

Originally ROF suggested to solve (1) by finding the steady state solution of its discrete gradient descent PDE.

However, due to the properties of the problem, including the fact that absolute function, $|\cdot|$, is non-differentiable at 0, that is for smooth regions $\nabla u \approx 0$, the step length has to be very small and the convergence becomes slow.



Optimization

MATLAB vs. the real stuff

- `fminsearch...` – doesn't move at all ☹
- `fminunc` without gradients – very slow, but moves at least.
- `fminunc` with gradients – it starts to look better. . . ☺
  Algorithm: Quasi-Newton line search, 88s (on Intel i7)
- `GPBB-NM` – FAST! ☺
  Algorithm: Non-monotone Barzilai-Borwein Gradient Projection on the **dual problem**, $< 0.1s$!
- And many more. . .

Regularized image denoising
Joakim Lindblad
Preliminaries
Baseline performance – simple methods
The ROF TV-model
Problems and solutions
Variations

Optimization

Problem: Absolute value is non-differentiable at zero.

- Solve the smoothed problem using $|x| \approx \sqrt{\varepsilon + x^2}, \varepsilon > 0$.
- Solve the **dual problem**.

. . . which will not fit in this lecture. . .

- Excellent reference:
  Mingqiang Zhu, Stephen J. Wright, Tony F. Chan. "Duality-based algorithms for total-variation-regularized image restoration", Computational Optimization and Applications, November 2010, Volume 47, Issue 3, pp 377-400
- Good but a bit heavy sometimes:
  "An introduction to Total Variation for Image Analysis" by: Antonin Chambolle, Vicent Caselles, Matteo Novaga, Daniel Cremers, Thomas Pock

Regularized image denoising
Joakim Lindblad
Preliminaries
Baseline performance – simple methods
The ROF TV-model
Problems and solutions
Variations

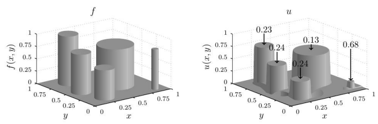
Implementation

- Excellent online material by Mingqiang Zhu, Stephen Wright, and Tony Chan:
  `http://pages.cs.wisc.edu/~swright/TVdenoising/`
- as well as the nice work of Pascal Getreuer:
  `http://www.getreuer.info/home/tvreg`
- GPU implementation ⇒ cirka 100 times faster.

Regularized image denoising
Joakim Lindblad
Preliminaries
Baseline performance – simple methods
The ROF TV-model
Problems and solutions
Variations

Problems

- Loss of contrast
  - For a white disk of radius r, the contrast is reduced by $\frac{1}{\lambda r}$.



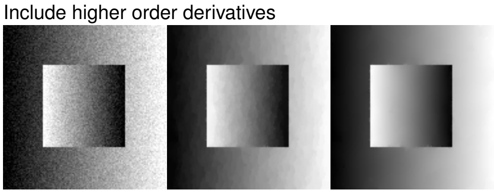
- Loss of geometry
  - Cutting off a corner may reduce the overall energy.
- Staircasing
  - A tendency to produce flat regions with artificial edges.
- Loss of texture
  - Failure to differentiate fine texture details from noise.

See also: "Recent developments in total variation image restoration" by T. Chan , S. Esedoglu , F. Park , A. Yip In Mathematical Models of Computer Vision (2005)

Regularized image denoising
Joakim Lindblad
Preliminaries
Baseline performance – simple methods
The ROF TV-model
Problems and solutions
Variations

Solutions

- Iterated refinement
  - Compensate for loss of signal by adding back removed signal in next iteration of ROF.
- Total Generalized Variations
  - Include higher order derivatives



- L^1 norm for data term (and other variants)
- Potential function on the regularization term

Regularized image denoising
Joakim Lindblad
Preliminaries
Baseline performance – simple methods
The ROF TV-model
Problems and solutions
Variations

L^1 norm

Modifying the data term

Loss of contrast as well as loss of shape is partly due to different power of the two terms, one quadratic and one linear, where for small values, the linear part dominates.

One solution is to change to L^1 norm in the data fidelity term.

$$\hat{u} = \underset{u \in L^2(\Omega)}{\arg\min} \int |\nabla u| + \frac{\lambda}{2} \int_\Omega |f(x) - u(x)| \, dx$$

This approach has been successfully used for other types of noise, e.g., Poisson noise or salt and pepper noise. Further this norm turns out to be useful for denoising of *shapes*, i.e., when the u is a characteristic function of a shape.

Potentials

Modifying the regularization term

Another solution is to change the other side, that is the regularization term. This can be done by wrapping the term by a s.c. **potential function** φ,

$$\hat{u} = \underset{u \in L^2(\Omega)}{\arg\min} \int \varphi(|\nabla u|) + \frac{\lambda}{2} \int_{\Omega} (f(x) - u(x))^2 \, dx \qquad (2)$$

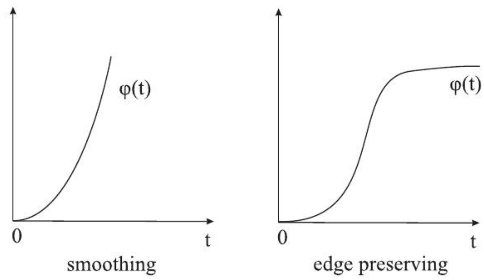
Clearly, choosing $\varphi(x) = x$ gives us back Eq. (1), and using $\varphi(x) = x^2$ gives us the quadratic term that does not give edge preservation. Are there other options which are better?

The problem (2) is convex if φ is convex (sum of convex functions).

To reduce the problem of loosing contrast and shape, we want φ to be close to quadratic for small values (small gradients). That also reduces the problem of non-differentiability of $|\cdot|$ at zero.

To preserve edges, the function should not grow too fast away from zero.

Potentials



Example shapes of smoothing and edge-preserving potentials.

Potentials

We tested a number of potential functions, some convex, some non-convex, for different images and noise conditions.

Empirical results fit theoretical expectations.

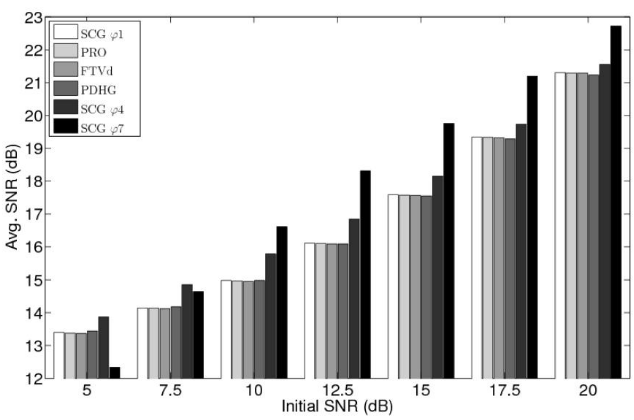
$\varphi(t)$
total variation pot. fun.
$\varphi1(t) = t,$
smoothing pot. fun.
$\varphi2(t) = t^{\alpha},\ 1 < \alpha < 2$
$\varphi3(t) = t^2,$
edge preserving pot. fun.
$\varphi4(t) = \begin{cases} t^2, & t \le \alpha \\ 2\alpha t - \alpha^2, & t > \alpha \end{cases} \quad \alpha > 0$
$\varphi5(t) = \sqrt{\alpha + t^2},\ \alpha > 0$
$\varphi6(t) = \ln\cosh(\alpha t),\ \alpha > 0$
$\varphi7(t) = \dfrac{\alpha t^2}{1 + \alpha t^2},\ \alpha > 0$
$\varphi8(t) = \ln(1 + \alpha t^2),\ \alpha > 0$
$\varphi9(t) = 1 - e^{-\alpha t^2},\ \alpha > 0$
$\varphi10(t) = \begin{cases} \sin(\alpha t^2), & 0 \le t \le \sqrt{\frac{\pi}{2\alpha}} \\ 1, & t > \sqrt{\frac{\pi}{2\alpha}} \end{cases} \quad \alpha > 0$

Potentials

Two potentials stand out as best: The convex *Huber potential*, φ_4 and The non-convex *Geman and McClure*, φ_7.

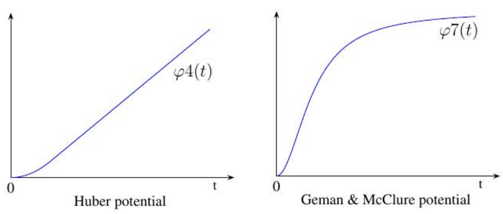
Results

Performance comparison with state of the art methods



T. Lukić, J. Lindblad, and N. Sladoje. "Regularized image denoising based on spectral gradient optimization." Inverse Problems. Vol 27, No. 8, 2011.

Potentials



The Huber potential has the best overall performance, and provide robust and fast (faster than TV) denoising.

The non-convex Geman and McClure provides by far the most accurate reconstruction for low noise levels, however at the cost of slower convergence and possibly non-unique solution (we initialize with the observed image f).

- Both are differentiable at zero.

TV deconvolution

Replace u in (1) with the convolution $k * u$

$$\hat{u} = \underset{u \in L^2(\Omega)}{\arg\min} \int |\nabla u| + \frac{\lambda}{2} \int_{\Omega} (f - k * u)^2 \, dx$$

original image   observed image   proposed method

Picture from: "`deconvtv` - an algorithm for total variation deconvolution", Stanley H. Chan, Philip E. Gill and Truong Q. Nguyen

Texture preserving TV

Adaptive regularization

- Reduces selectively the total-variation of the image. Denoising is strong in smooth regions and weaker in textured regions.
- Preserves better texture and fine-scale details.
- A two-phase process where the noise and textures are first isolated by scalar TV. The method then imposes local power constraints based on local variance measures of the first phase.

Noisy input   Scalar TV   Adaptive TV

Guy Gilboa, Nir Sochen, Yehoshua Y. Zeevi, "Texture Preserving Variational Denoising Using an Adaptive Fidelity Term", in Proc. VLSM 2003, Nice, France, Oct. 2003.

TV inpainting

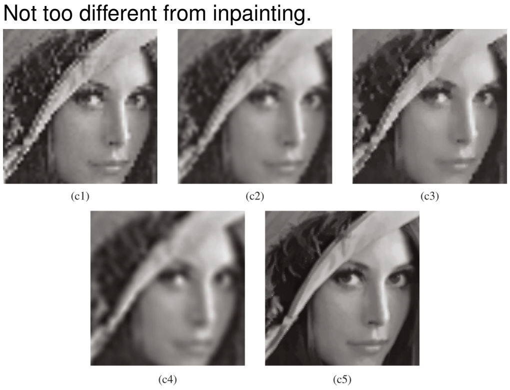
Very (visually) impressive results by selectively setting $\lambda(x) = 0$

Input   Output   Original

P. Getreuer, "Total Variation Inpainting using Split Bregman," Image Processing On Line, 2012.

Zooming / super resolution

Not too different from inpainting.



Yu-Fei Yang ; Ting-Ting Wu and Zhi-Feng Pang "Image-zooming technique based on Bregmanized nonlocal total variation regularization", Opt. Eng. 50(9), September 26, 2011.

Segmentation

A similar approach can be used for segmentation.

As an output we want plateaus of distinct object labels.

- Image intensities can be modelled as a liner mixture (convex combination) of pure class representatives (a.k.a. end members).
  - This as achieved by multiplying with a matrix in the data term.
- We want strong smoothing (to get plateaus) and strong edges, but no slow changes.
  - More similar to the Mumford and Shah model.
  - L^0 norm, **Sparse representation**.
  - Leads to a non-convex problem. Addressed by starting from a convex problem and slowly increasing non-convex parts.
- Gradient magnitude is not well defined for non-smooth image functions.
  - Total variation is defined for non-smooth functions; definition is equivalent to definition of **perimeter of a fuzzy set**.

Coverage segmentation by energy minimization

- To impose sparsity of the solution (only few edge pixels) we slowly increase the cost of fuzzy pixels in the image.
- We aim for a coverage segmentation (no reason to throw away edge information).
  - Allow fuzzy, i.e., partly covered pixels surrounding the different image components.

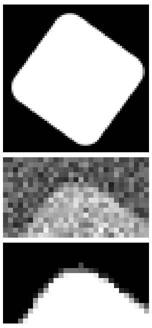
$$\hat{u} = \underset{u \in \mathcal{A}(\Omega)}{\arg\min} \|f - uc\|_2^2 + \mu P(u) + \nu F^*(u)$$

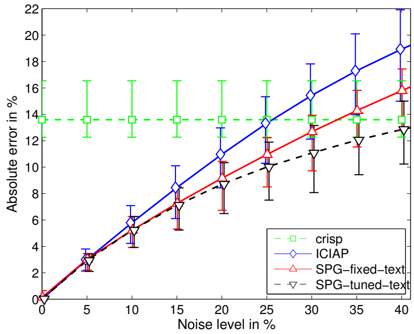
J. Lindblad and N. Sladoje. Coverage Segmentation Based on Linear Unmixing and Minimization of Perimeter and Boundary Thickness. Pattern Recognition Letters. Vol 33, No. 6, pp. 728-738, 2012.

Coverage segmentation by energy minimization

Quantitative evaluation

- noise sensitivity

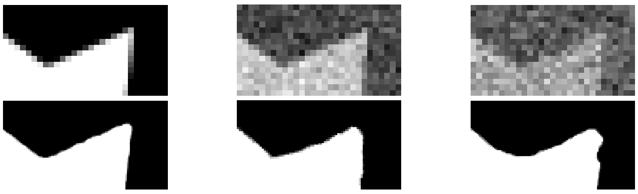




The use of spatial information (perimeter term) leads to further improved segmentation performance.

A flexible framework

- Combine with deconvolution, for segmentation of blurry images.
- Combine with inpainting, for segmentation of partly occluded objects.
- Combine with zooming, for super resolution segmentation.



Segmentation at $4\times$ higher resolution for 0%, 15%, 30% noise.