Variational Image Restoration

Image Restoration: Denoising

Image restoration is a classical inverse problem: Given an observed image $f : \Omega \to \mathbb{R}$ and a (typically stochastic) model of the image formation process, we want to restore the original image $u : \Omega \to \mathbb{R}$.
A prototypical noise model is given by:

$$f = u + \eta, \qquad \eta \sim \mathcal{N}(0, \sigma),$$

which means that the observed image f is equal to the original u plus additive zero-mean Gaussian noise. Given some prior that the true image u is spatially smooth, one can estimate the true image by minimizing the ROF model (Rudin, Osher, Fatemi '92):

$$\min_u \frac{1}{2} \int |u - f|^2 dx + \int |\nabla u| dx.$$

It gives rise to the Euler-Lagrange equation

$$u - f - \text{div}\left(\frac{\nabla u}{|\nabla u|}\right) = 0.$$

Of course one can consider other noise models and other regularizers.

Image Restoration: Denoising

Original

noisy

denoised

(Source: Goldluecke, Cremers, CVPR 2010)

Image Restoration: Deblurring

A prototypical blur model is given by

$$f = A * u + \eta \qquad \eta \sim \mathcal{N}(0, \sigma),$$

where the observed image f arises by convolving the original u with a blur kernel A and adding Gaussian noise. This process can be inverted in a variational setting by minimizing the TV deblurring functional:

$$\min_u \frac{1}{2} \int |A * u - f|^2 dx + \int |\nabla u| dx.$$

For symmetric kernels A the corresponding Euler-Lagrange equation is given by:

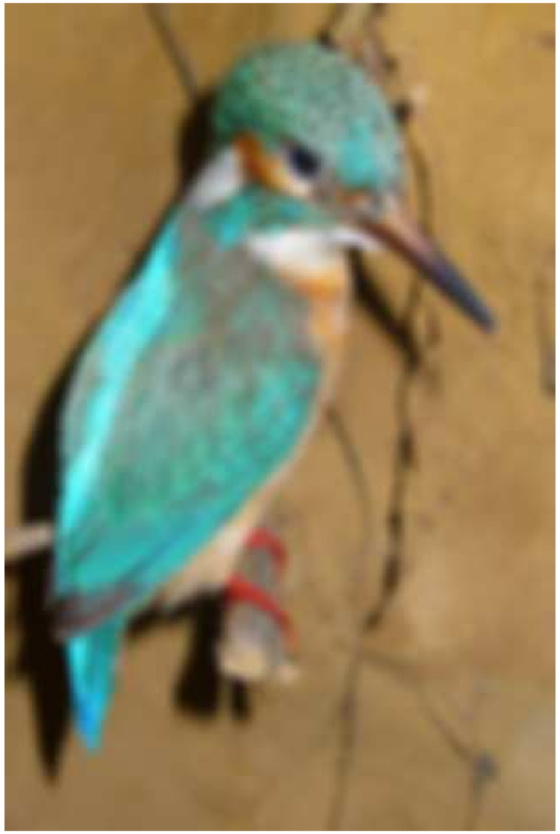
$$A * (A * u - f) - \text{div}\left(\frac{\nabla u}{|\nabla u|}\right) = 0,$$

and the gradient descent equation

$$\frac{\partial u}{\partial t} = -A * (A * u - f) + \text{div}\left(\frac{\nabla u}{|\nabla u|}\right).$$

Image Restoration: Deblurring

Original



blurred and noisy

deblurred

(Source: Goldluecke, Cremers, ICCV 2011)

Inverse Problems and Bayesian Inference

How can one systematically derive functionals associated with different image formation models?
A systematic approach to this question is given by the framework of Bayesian inference. Let u be the unknown true image and f the observed one, then we can write the joint probability for u and f as:

$$\mathcal{P}(u, f) = \mathcal{P}(u|f)\,\mathcal{P}(f) = \mathcal{P}(f|u)\mathcal{P}(u).$$

Rewriting this expression we obtain the Bayesian formula (Thomas Bayes 1887):

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

Using this formula, we can now aim at computing the most likely solution $\hat{u}$ given f by maximizing the posterior probability $\mathcal{P}(u|f)$

$$\hat{u} = \arg\max_u \mathcal{P}(u|f) = \arg\max_u \mathcal{P}(f|u)\,\mathcal{P}(u).$$

In this setting $\mathcal{P}(f|u)$ is called the likelihood and $\mathcal{P}(u)$ the prior.
This method is referred to as Maximum Aposteriori (MAP) estimation.

MAP Estimation in the Discrete Setting

Let us assume n independent pixels. For each the measured intensity f_i is given by the true intensity u_i plus additive Gaussian noise. This corresponds to the likelihood

$$\mathcal{P}(f_i|u_i) \propto \exp\left(-\frac{(u_i - f_i)^2}{2\sigma^2}\right).$$

Since all measurements are mutually independent, we obtain for the entire vector $f = (f_1, \dots, f_n)$ of pixel intensities:

$$\mathcal{P}(f|u) = \prod_{i=1}^{n} \mathcal{P}(f_i|u) = \prod_{i=1}^{n} \mathcal{P}(f_i|u_i) \propto \prod_{i=1}^{n} \exp\left(-\frac{(u_i - f_i)^2}{2\sigma^2}\right).$$

Let us now assume that the apriori probability for each u_i only depends on the neighbor intensities (Markov property):

$$\mathcal{P}(u) = \mathcal{P}(u_1, \dots, u_n) = \mathcal{P}(u_1|u_2 \dots, u_n)\mathcal{P}(u_2, \dots, u_n) \propto \prod_{i=1}^{n-1} \mathcal{P}(u_i|u_{i+1}).$$

MAP Estimation in the Discrete Setting

Assuming a simple smoothness prior, we have:

$$\mathcal{P}(u) = \prod_{i=1}^{n-1} \mathcal{P}(u_i | u_{i+1}) \propto \prod_{i=1}^{n-1} \exp\left(-\lambda |u_i - u_{i+1}|\right).$$

With these assumptions, the maximum aposteriori (MAP) probability is given by:

$$\mathcal{P}(u|f) \propto \prod_{i=1}^{n} \exp\left(-\frac{|f_i - u_i|^2}{2\sigma^2}\right) \prod_{i=1}^{n-1} \exp\left(-\lambda |u_i - u_{i+1}|\right)$$

Rather than maximizing this probability, one can equivalently minimize its negative logarithm (because the logarithm is strictly monotonous). It is given by the energy

$$E(u) = -\log \mathcal{P}(u|f) = \sum_{i=1}^{n} \frac{|f_i - u_i|^2}{2\sigma^2} + \lambda \sum_{i=1}^{n-1} |u_i - u_{i+1}| + \text{const.}$$

MAP Estimation in the Continuous Setting

Similarly one can define Bayesian MAP optimization in the continuous setting, where the likelihood is given by:

$$\mathcal{P}(f|u) \propto \exp\left(-\int \frac{|f(x)-u(x)|^2}{2\sigma^2}dx\right),$$

and the prior is given by

$$\mathcal{P}(u) \propto \exp\left(-\lambda\int |\nabla u(x)|dx\right).$$

Thus the data term in variational methods corresponds to the likelihood, whereas the regularizers corresponds to the prior:

$$E(u) = -\log \mathcal{P}(u|f) = \int \frac{|f(x)-u(x)|^2}{2\sigma^2}dx + \lambda \int |\nabla u(x)|dx + \text{const.}$$

A systematic derivation of probability distributions on infinite-dimensional spaces requires a more formal derivation (introduction of measures etc). This is beyond the scope of this lecture.

Example: Motion Blur

Assume the camera lens opens instantly and remains open during the time interval $[0, T]$ in which the camera moves with constant velocity V in x-direction. Then the observed brightness is given by

$$g(x,y) = \int_0^T f(x - Vt, y)dt.$$

Inserting $x' \equiv Vt$ this expression can be written as a convolution with a kernel $h(x,y)$:

$$g(x,y) = \int\limits_0^{VT} f(x - x', y)\frac{1}{V}dx' = \int\limits_{-\infty}^{\infty} f(x - x', y - y')h(x', y')dx'dy',$$

where:

$$h(x', y') = \frac{1}{V} \cdot \delta(y') \cdot \chi_{[0,VT]}(x'), \qquad \text{and}$$

$$\chi_{[a,b]}(x') = \begin{cases} 1, & \text{if } x' \in [a,b] \\ 0, & \text{else} \end{cases} \qquad \text{(box filter)}$$

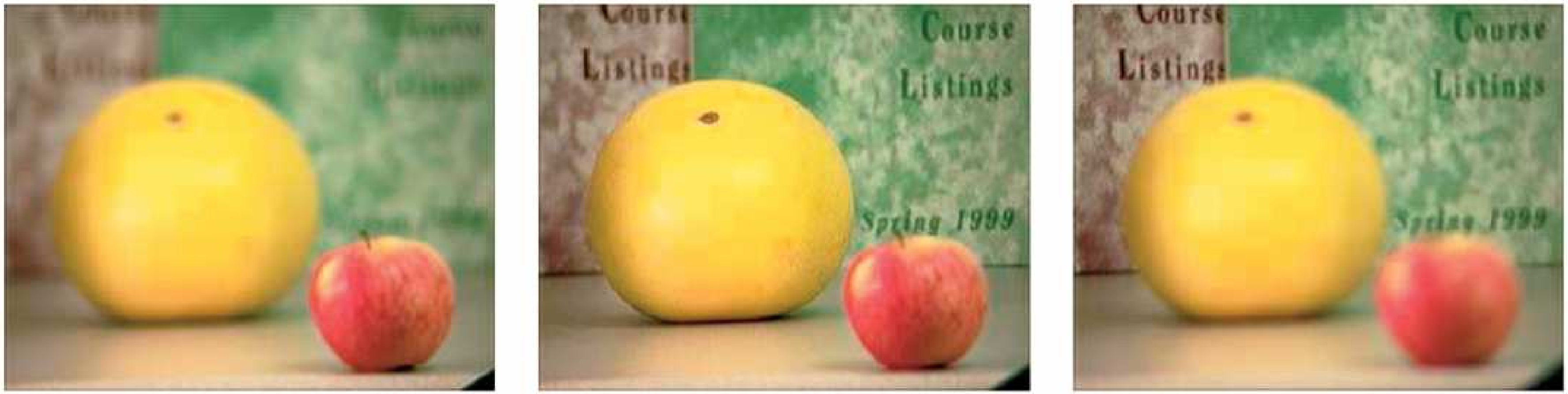
Example: Motion Blur

Original

Motion-blurred

(Author: D. Cremers)

Example: Defocus Blur



Scene captured with three different focal settings.

Space-varying blur depends on the distance from the focal plane.

(Source: Favaro, Soatto, PAMI 2005)

Image Restoration: Super Resolution

The key idea of super resolution from video is to exploit the redundancy available in multiple images of a video. The assumption is that each input image f_i is a blurred and downsampled version of the original high-resolution scene. We can try to recover a high-resolution image u with a variational approach of the form:

$$\min_u \sum_{i=1}^{n} \int |Au(x + w_i(x)) - f_i(x)| \, dx \; + \; \lambda \int |\nabla u| \, dx.$$

Here $w_i : \Omega \to \mathbb{R}^2$ are the motion fields which the original scene undergoes, and A is a linear operator modeling the blurring and downsampling. Again, the variational approach aims at inverting an image formation process of the form:

$$f_i(x) = Au\big(x + w_i(x)\big) + \eta,$$

which states that the observed image is obtained from the “true” image by displacement, blurring and downsampling plus additive Poisson-distributed noise η.

Image Restoration: Super Resolution

One of several input images

Superresolution estimate

(Source: Schoenemann, Cremers, IEEE T. on Image Processing 2012)