Chapter 10
Optimization
Variational Approaches and PDEs

Statistical Methods and Learning in Computer Vision
WS 2012

Claudia Nieuwenhuis
Lehrstuhl für Computer Vision and Pattern Recognition
Fakultät für Informatik
Technische Universität München

Overview

1. Optimization

Optimization

The objective of optimization approaches is to find the minimum (or maximum) energy state of a given functional, which describes the 'optimal' solution to the image processing task.

- Discrete Optimization: The problem is modeled as a graph of nodes with specific neighborhoods (Markov Random Field).
- Continuous Optimization: The problem is solved in the continuous function space.

Continuous Optimization - Examples

Let $I : \Omega \to \mathbb{R}^3$ denote the given image. Find the solution u, e.g.

- Denoising: Find the original image $u : \Omega \to \mathbb{R}^3$ without noise given the noisy image I.
- Segmentation: Find an indicator function $u : \Omega \to \{0, 1\}$ which is 1 in the foreground and 0 in the background of the image I.
- Deblurring: Find the original image $u : \Omega \to \mathbb{R}^3$ given the blurred image I.
- 3D-Reconstruction: Find the indicator function $u : \Omega \to \{0, 1\}$ which indicates if the voxel is inside or outside the object given one or several images I.
- Inpainting: Find the repaired image $u : \Omega \to \mathbb{R}^3$ given the image I with holes.
- ...

Denoising

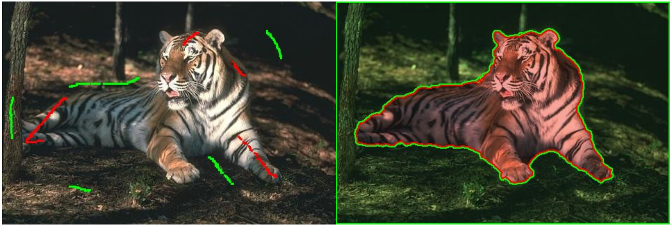
Let $I : \Omega \to \mathbb{R}^3$ denote the given image. Find the original image u without noise given the noisy image I.

noisy I

recovered u

Segmentation

Let $I : \Omega \rightarrow \mathbb{R}^3$ denote the given image. Find an indicator function $u : \Omega \rightarrow \{0, 1\}$ which is 1 in the foreground and 0 in the background of the image I.



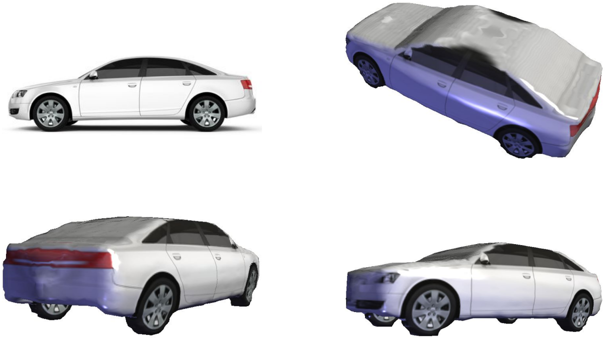
Deblurring

Let $I : \Omega \to \mathbb{R}^3$ denote the given image. Find the original image u given the blurred image I.

Original | I | *recovered u*

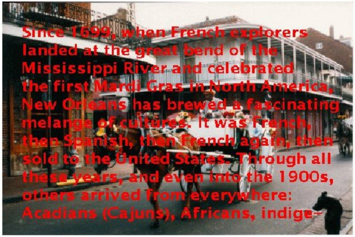
3D-Reconstruction

Find the indicator function $u : \Omega \to \{0, 1\}$ which indicates if the voxel is inside or outside the object given one or several images I.



Inpainting

Let $I : \Omega \to \mathbb{R}^3$ denote the given image. Find the 'repaired' image $u : \Omega \to \mathbb{R}^3$.



Damaged image I

Recovered image u

Bayes Formalism

Bayes Formalism

These problems can be formalized in the Bayes Formalism. We are looking for the most probable function u given the image I.

$$P(u|I) = \frac{P(I|u)P(u)}{P(I)} \tag{1}$$

$P(I|u)$ is the probability of seeing I given the solution u.
$P(u)$ is a prior probability describing the class of solutions (i.e. what typical images or indicator functions look like).
$P(I)$ is constant and can be neglected since it is has no influence on the solution u.

Bayes Formalism

Given:
Original image $I : \Omega \to \mathbb{R}^3$.
Objective:
Find $u : \Omega \to \mathbb{R}^3$ or $u : \Omega \to \{0, 1\}$ respectively.

We have to define

- data term $P(I|u)$: relating the image (data) to the segmentation result u
- regularizer $P(u)$: describing typical indicator functions, i.e. regularizing the solution

Denoising

noisy I

recovered u

Find the original image $u : \Omega \to \mathbb{R}^3$ without noise given the noisy image I.

Denoising - Data term

The data term relates the image to the desired solution. For deblurring the recovered image should be similar to the observed image I.

$$P(I|u) = \frac{1}{C} \exp^{-\|I-u\|^2}$$

We assume that the color values at each pixel are independent of each other.

$$P(I|u) \approx \prod_{x \in \Omega} \frac{1}{C(x)} \exp^{-\|I(x)-u(x)\|^2}$$

Denoising - Regularizer

The regularizer makes assumptions on the denoised image, e.g. a certain smoothness. These assumptions are called priors. What are suitable priors for a denoised image?

1) Squared Gradient

$$P(u) = \frac{1}{C} \exp^{-\lambda \|\nabla u\|^2} \approx \prod_{x \in \Omega} \frac{1}{C(x)} \exp^{-\lambda \|\nabla u(x)\|^2}$$

2) Total Variation

$$P(u) = \frac{1}{C} \exp^{-\lambda \|\nabla u\|} \approx \prod_{x \in \Omega} \frac{1}{C(x)} \exp^{-\lambda \|\nabla u(x)\|}$$

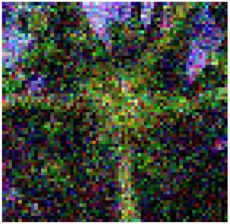
Denoising - Data term and Regularizer

Instead of maximizing $P(u|I)$ we minimize its negative logarithm

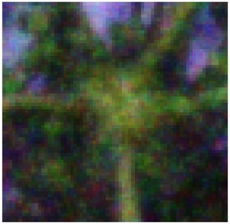
$$\underset{u}{\operatorname{argmax}}\, P(u|I) =$$

$$\underset{u}{\operatorname{argmin}} \int_{\Omega} \|I(x) - u(x)\|^2 \, dx + \lambda \int_{\Omega} \|\nabla u(x)\| \, dx$$

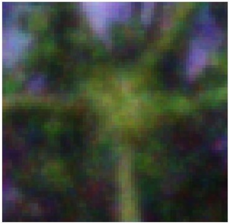
Original



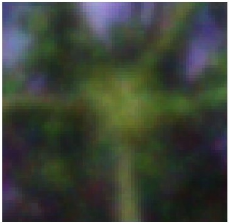
Noisy



$\lambda = 2.5$



$\lambda = 5.0$

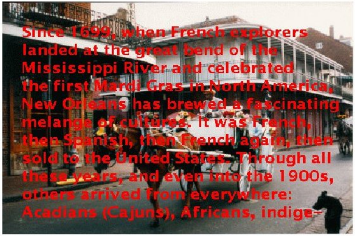


$\lambda = 10.0$

The larger λ is chosen the smoother becomes the denoised image.

Inpainting

Let $I : \Omega \to \mathbb{R}^3$ denote the given image, and let $M \subset \Omega$ denote the set of pixels to be inpainted (i.e. the red text here). Find the 'repaired' image $u : \Omega \to \mathbb{R}^3$.



Damaged image I

Recovered image u

Inpainting

$$\underset{u}{\operatorname{argmax}}\, P(u|I) =$$

$$\underset{u}{\operatorname{argmin}} \int_\Omega (1 - 1_M(x))\|u(x) - I(x)\|^2 + \lambda\|\nabla u(x)\|\, dx$$

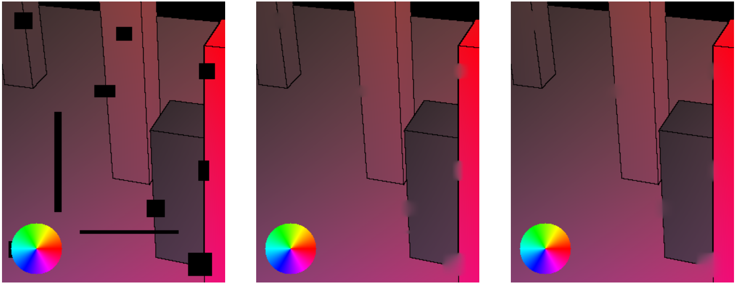
where 1_M is the characteristic function of the set of pixels to be inpainted

$$1_M(x) = \begin{cases} 1, & x \in M \\ 0, & \text{otherwise} \end{cases}$$

The constant $\lambda > 0$ is chosen very small so that smoothing of the image outside the inpainting region M is minimal.

Inpainting

Inpainting result for the image in a) with holes in black.

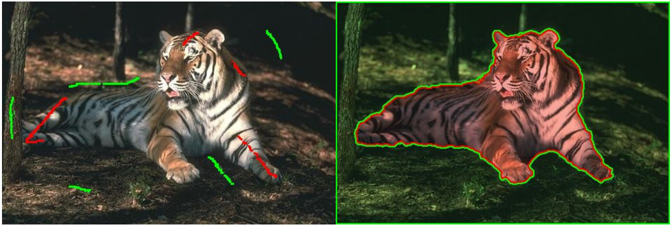


a) mask b) $\int_\Omega \|\nabla u(x)\|^2 \, dx$ c) $\int_\Omega \|\nabla u(x)\| \, dx$

b) Result for inpainting based on the squared gradient regularizer.
c) Result for inpainting based on the total variation regularizer.

We see that TV produces more pronounced edges, whereas the squared gradient oversmoothes the boundaries.

Segmentation



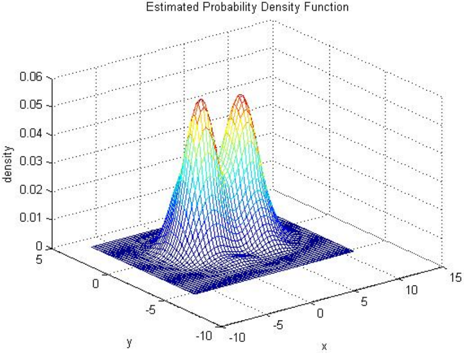
Find an indicator function $u : \Omega \rightarrow \{0, 1\}$ which is 1 in the foreground and 0 in the background of the image I.

Segmentation - Data term

$P(I|u)$ relates the image data to the segmentation result. Under independence assumptions we obtain

$$P(I|u) \approx \prod_{i=0}^{1} \prod_{x \in \Omega_i} P(I|u(x) = i) \tag{2}$$

where $\Omega_i = \{x \in \Omega | u(x) = i\}$. $P(I|u(x) = 1)$ and $P(I|u(x) = 0)$ can be estimated by means of a Parzen density estimator.



Segmentation - Regularizer

$P(u)$ indicates the likelihood for each possible segmentation u.
What does a 'correct' segmentation look like?
The segments should be smooth, but edges should be preserved.

Total Variation

$$P(u) = \frac{1}{C}\exp^{-\lambda\|\nabla u\|} \approx \prod_{x\in\Omega} \frac{1}{C(x)}\exp^{-\lambda\|\nabla u(x)\|}$$

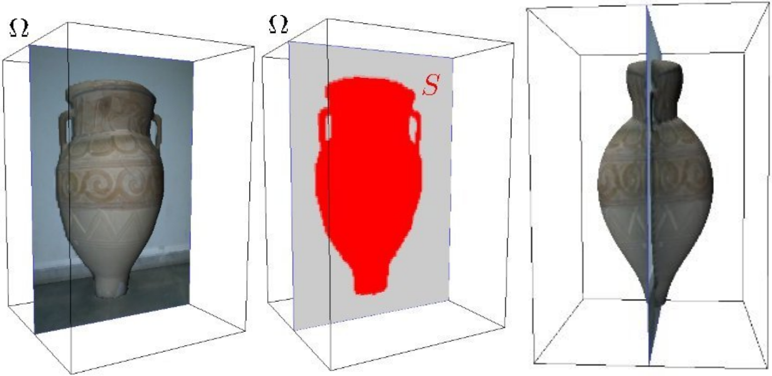
Segmentation - Data term and Regularizer

Instead of maximizing $P(u|I)$ we minimize its negative logarithm

$$
\begin{aligned}
&\operatorname*{argmax}_u P(u|I) = \\
&\operatorname*{argmin}_u \int_\Omega -\log P(I(x)|u(x)=1)u(x) \\
&\qquad\qquad -\log P(I(x)|u(x)=0)(1-u(x))\, dx \\
&\qquad\qquad +\lambda \int_\Omega \|\nabla u(x)\|\, dx
\end{aligned}
$$

3D-Reconstruction

Find the indicator function $u : \Omega \to \{0, 1\}$ which indicates if the voxel is inside or outside the object given one or several images I. Voxels are 'three-dimensional pixels'.


Ω
Ω
S

3D Reconstruction - Data term and Regularizer

Here, we do single view reconstruction, i.e. we only have one image.
1) We assume that the projection of the object onto the 2D image plane equals the segmented object in the image. u_S contains all indicator functions which have this property.
2) The surface of the object ($\int_\Omega \|\nabla u(x)\| \, d^3x$) should be minimal.
3) To avoid the $u = 0$ solution we define the volume V of the object: $\int_\Omega u(x)\, dx = V$

$$\operatorname*{argmin}_u \int \|\nabla u(x)\| d^3x, \quad s.t. u \in U_S, \int_\Omega u(x) dx = V$$

There are several ways to optimize such functionals:

- Gradient Descent
- For convex functionals there are fast primal-dual algorithms based on the dual form of the optimization problem
- Augmented Lagrangian methods, which solve constrained minimization problems by solving the unconstrained problem with additional terms in the energy. The idea is based on Lagrange Multipliers.

Gradient Descent

- Start with any initial guess
- Compute the gradient direction at this point to find out in which direction the functional decreases most quickly
- Move into this direction and update the guess

In general, we can only find local minima with this algorithm. How can we find optimal points of functionals?

To understand the derivation of the Euler-Lagrange Equation we need two things: the divergence theorem of Gauss and the DuBois-Reymond-Lemma

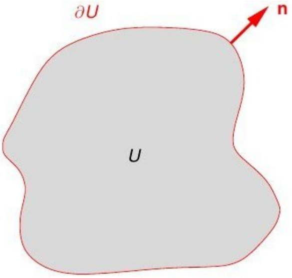
Divergence theorem of Gauss

Let $U \subset \mathbb{R}^n$ be compact with piecewise smooth boundary, $n : \partial U \to \mathbb{R}^n$ the outer normal of U and $\xi : \mathbb{R}^n \to \mathbb{R}^n$ a continuously differentiable vector field. Then

$$\int_U div\ \xi\ dx = \int_{\partial U} \xi n ds.$$

For integration by parts in higher dimensional spaces it follows

$$\int_\Omega \nabla u\ \xi\ dx = -\int_\Omega u\ div\ \xi\ dx + \int_{\partial\Omega} u\ \xi\ n\ ds.$$



DuBois-Reymond-Lemma

Let $u \in \mathcal{L}^1$. If

$$\int_\Omega u(x)h(x)\,dx = 0$$

for all test functions $h \in C_c^1$ then $u = 0$ almost everywhere.

The Euler-Lagrange Equation is a PDE which has to be satisfied by an extremal point u^* of the functional.

Euler-Lagrange Equations

Let u^* be an extremum of the function $E : C^1 \to \mathbb{R}$ with

$$E(u) = \int_\Omega L(u(x), \nabla u(x), x)\; dx,$$

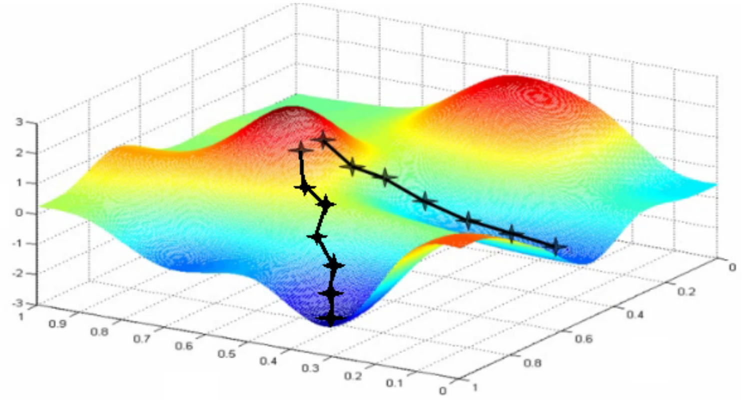
where $L : \mathbb{R} \times \mathbb{R}^n \times \Omega \to \mathbb{R}$, $(a, b, x) \to L(a, b, x)$. Then u^* satisfies the Euler-Lagrange Equation

$$\frac{\partial L(u^*, \nabla u^*, x)}{\partial a} - div_x \left[\frac{\partial L(u^*, \nabla u^*, x)}{\partial b} \right] = 0$$

Important: The left-hand side of the Euler-Lagrange equation can be understood as the gradient of the functional E with respect to the function u, a gradient in function space.

Gradient Descent via Time Stepping

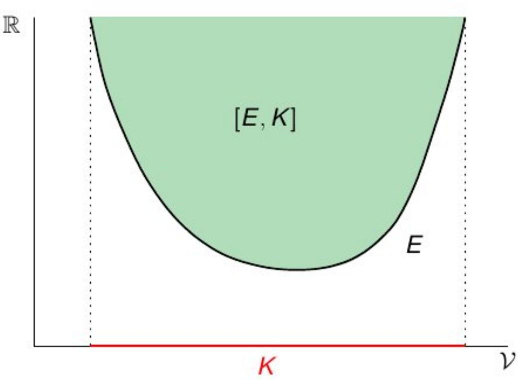
1. choose an initial solution u, i.e. $u = 0$.
2. update your solution $u^{t+1} = u^t - dt\frac{\partial E}{\partial u}$
3. if solution has converged (e.g. $|u^{t+1} - u^t| < \epsilon$) stop, otherwise goto 2



Convex Functionals

A functional $E : K \to \mathbb{R}$ is convex if its epigraph $[E, K]$ is a convex set, i.e. if

$$E(\alpha u + (1-\alpha)v) \leq \alpha E(u) + (1-\alpha)E(v).$$



Convex functionals have nice optimization properties, since they only have one global minimum which can be computed!

Discretization

Discretization of the Gradient

The gradient is computed on the primal variables u and is usually discretized with forward differences:

$$(\nabla u)_{ij} = \begin{pmatrix} u_{i+1,j} - u_{i,j} \\ u_{i,j+1} - u_{i,j} \end{pmatrix}$$

Von Neumann boundary conditions are used, i.e. for grid size $M \times N$

$$(\nabla u)^1_{M,j} = 0 \quad \text{and} \quad (\nabla u)^2_{i,N} = 0$$

Discretization of the Laplace operator

The Laplace operator $\Delta u = \frac{\partial^2 u}{\partial^2 x_1} + \frac{\partial^2 u}{\partial^2 x_2}$ is discretized as follows:

$$(\Delta u)_{ij} = -4u_{i,j} + u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}$$

Von Neumann boundary conditions are used, i.e. for grid size $M \times N$

$$(\Delta u)_{1,j} = 0 \quad (\Delta u)_{M,j} = 0 \quad (\Delta u)_{i,1} = 0 \quad (\Delta u)_{i,N} = 0$$