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 $u : \Omega \to \mathbb{R}$, e.g.

- Denoising: Find the original image u 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 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.
- ...

Denoising

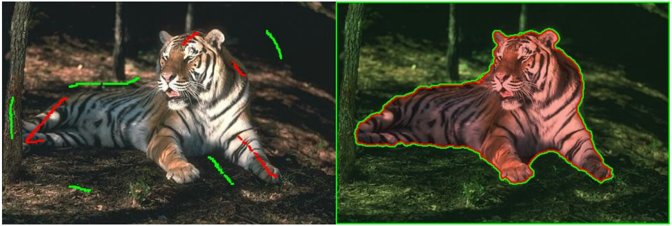
Let $I : \Omega \to \mathbb{R}^3$ denote the given image. Find the original image u without noise given the noisy image I.

noisy I

recovered u

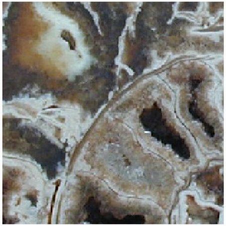
Segmentation

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

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



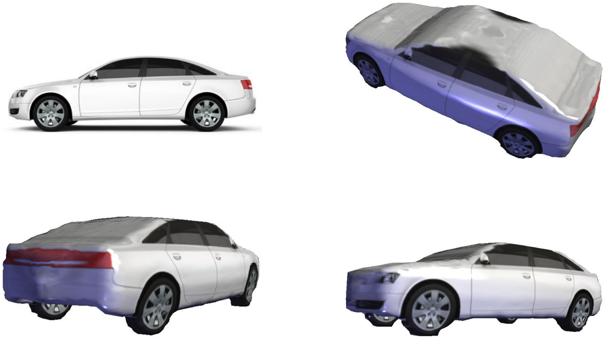
Original

I

recovered u

3D-Reconstruction

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



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 (with scribbles indicating color samples).
Objective:
Find $u : \Omega \to \{0, 1\}$ indicating foreground and background.

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 $u : \Omega \to [0, 255]^3$ 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

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

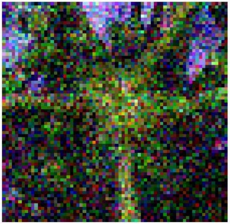
Denoising - Data term and Regularizer

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

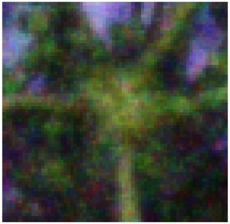
$$\underset{u}{\operatorname{argmax}}\, P(u|I) =$$

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

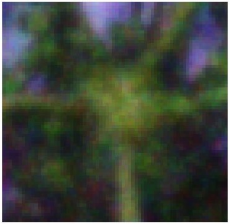
Original



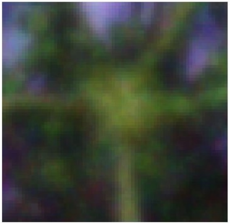
Noisy



$\lambda = 2.5$

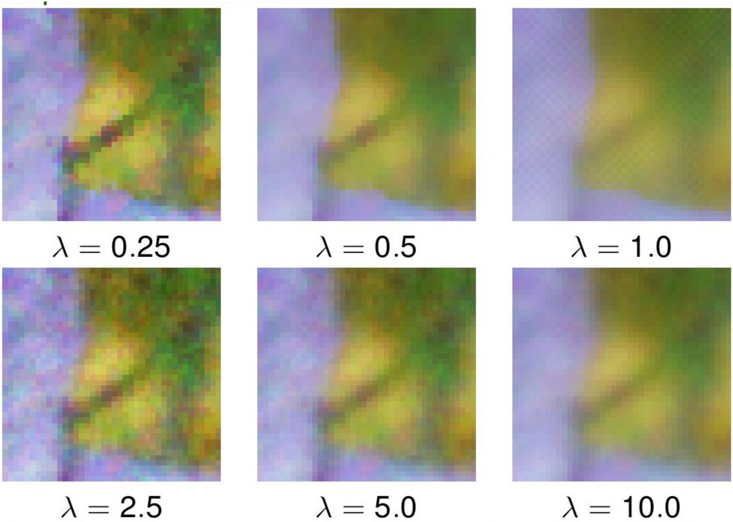


$\lambda = 5.0$



$\lambda = 10.0$

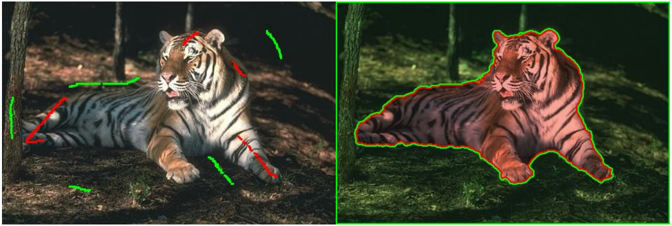
The larger λ is chosen the smoother becomes the denoised image.



top: total variation, bottom: squared gradient

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

Segmentation



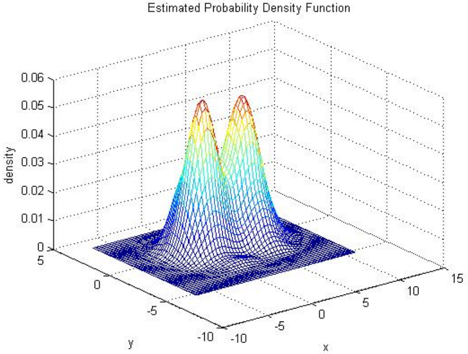
Find an indicator function $u : \Omega \to \{0, 1\}$ which is 1 in the foreground and 0 in the background of the image I.

Segmentation - Data term

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

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

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

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

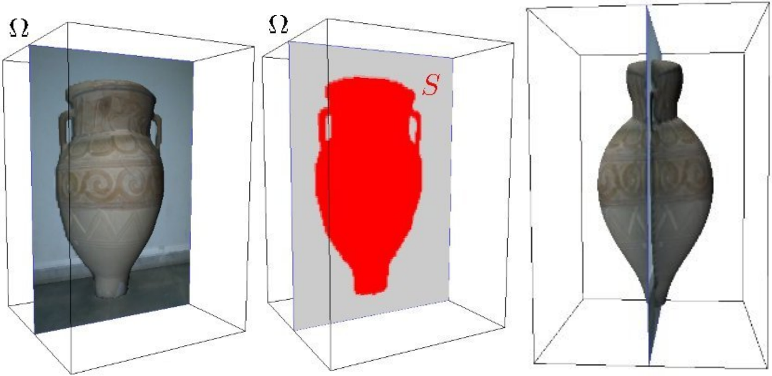
Segmentation - Data term and Regularizer

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

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

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.

$$\operatorname*{argmin}_u \int \|\nabla u(x)\| d^3x, \; s.t. u \in U_S$$

There are several ways to optimize such functionals:

- Gradient Descent
- For convex functionals there are fast algorithms based on the dual form
- Lagrange Multipliers for constrained problems ('augemented Lagrangian')

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?

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

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

To understand the derivation of the Euler-Lagrange Equation we need two things.

Divergence theorem of Gauss

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

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

For partial integration 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.

Total Variation

The Euler-Lagrange Equation of the total variation comes with problems as the denominator can become zero or very small leading to undefined or very large derivatives which cause numerical instabilities. Therefore, the total variation was introduced.

Total Variation

$$TV(u) = \int_\Omega \|\nabla u(x)\|\, dx = \sup_\xi \left\{ \int_\Omega u(x)\, div\, \xi(x) \mid \xi \in C_c^1(\Omega, \mathbb{R}^2), \|\xi\|_{L^\infty} \leq 1 \right\}$$

$\xi : \Omega \to \mathbb{R}^2$ is a vector field in C_c^1, that means its first derivative is continuous and it is defined over a compact set (i.e. outside the set we have $\xi(x) = 0$).

Important: The total variation regularizer is convex!

Segmentation - Data term and Regularizer

$$\begin{aligned}
\operatorname*{argmin}_u \int_\Omega & -\log P(I(x)|u(x)=1)u(x) \\
& -\log P(I(x)|u(x)=0)(1-u(x))\,dx \\
& +\lambda \int_\Omega \|\nabla u(x)\|\,dx = \\
\operatorname*{argmin}_u \sup_{\|\xi\|_{L^\infty}\leq 1} \int_\Omega & -\log P(I(x)|u(x)=1)u(x) \\
& -\log P(I(x)|u(x)=0)(1-u(x))\,dx \\
& +\lambda \int_\Omega u(x)\, div\, \xi(x)\,dx
\end{aligned}$$