Concepts and Algorithms of Scientific and Visual Computing
–Image Segmentation–

CS448J, Autumn 2015, Stanford University

Dominik L. Michels

Given a continuous function

$$\mathbb{R}^2 \supset \Omega := [0,1]^2 \ni \mathbf{x} \mapsto q(\mathbf{x}) \in [0,1] \subset \mathbb{R},$$

a gray-scale raster graphics image I can be defined as a discretization

$$\{0,\ldots,n-1\} \times \{0,\ldots,m-1\} \ni \mathbf{x} := (x_0, x_1) \mapsto I(\mathbf{x}) := q\left(\frac{x_0}{n-1}, \frac{x_1}{m-1}\right) \in [0,1] \subset \mathbb{R}$$

of q. In this context, n denotes the width and m denotes the height of the image I in terms of the number of pixels. $I(\mathbf{x})$ denotes the gray-scale value intensity of the image I in the pixel at position $\mathbf{x}$ in the range from 0 (black) to 1 (white). Analogously, a colored image is given as a triple $\mathbf{I} := (R, G, B)$, in which each component is defined similar to I. These components correspond to the red, green, and blue channels.[1]

[1] More precisely, each component usually corresponds to the red, green, and blue channels of an image (non-linear sRGB color space) or the luminance and chrominance channels of a video (linear YUV color space). In the case of a non-linear color space, we transform the colors into a linear one, so that we can do calculations on our color values.

Variational-based Image Processing

In the field of image processing, variational methods are common state of the art approaches to solve various tasks. First, we will illustrate this in the context of image denoising. This is usually done by smoothing the image, but if this is carried out without taking care of the details, is it likely, that the interesting features are not preserved.

The general idea is to specify an energy

$$E(q(\mathbf{x}), q'(\mathbf{x})) = \int_\Omega L(q(\mathbf{x}), q'(\mathbf{x})) \, d\mathbf{x},$$

whose argument

$$(q, q')^* = \underset{q \in ([0,1]^2 \to [0,1])}{\arg\min} (E(q, q'))$$

of the minimum is the continuous representation of the smoothed image.

The ensure a high-quality result, the energy E should increase with increasing error of q compared to the inertial representation u as well as if q is too noisy. A common choice is given by

$$L := \frac{1}{2}(q - u)^2 + \frac{\tilde{\lambda}}{2}|q'|^2,$$

in which the first summand penalizes the error and the second one the noise. The parameter λ controls the strength of the smoothness.

Substitution of L into the Euler-Lagrange equation leads to

$$\delta_q L = \partial_q(L) - d_t(\partial_{q'} L) = q - u + \tilde{\lambda} \Delta q.$$

Because the energy is convex, the Euler-Lagrange equation is a sufficient condition for a global minimum.

Hence q^* can easily be numerically computed with a gradient descent procedure

$$\partial_t q = -\delta_q L.$$

This leads to the iteration scheme

$$q(t + \epsilon) \leftarrow q(t) + \epsilon(u - q(t)) + \lambda \Delta q(t)$$

with $\lambda := -\epsilon \tilde{\lambda}$.

The first part $(u - q)$ ensures a minimal shift from the original image during the smoothing process which is controlled by the intensity of the parameter λ.

In the implementation, the Laplace operator can be approximated by

$$\Delta q(x_0, x_1) \approx \frac{1}{\epsilon^2} \left(q(x_0 - \epsilon, x_1) + q(x_0 + \epsilon, x_1) + q(x_0, x_1 - \epsilon) + q(x_0, x_1 + \epsilon) - 4q(x_0, x_1) \right).$$

(Please note, that the variable t describes the numerical time parameter of the gradient descent method and does not model physical time.)

Variational-based Image Processing

Figure : Smoothing of the Lena image*. The first row shows the original image (left) after smoothing for ten times (middle) and 25 times (right) with $(\lambda, \epsilon) = (0.2, 0.05)$. In the second row, $(\lambda, \epsilon) = (0.2, 0)$ are used which corresponds to a pure diffusion process.

*This image of the Swedish playmate Lena Söderberg taken by the American photographer Dwight Hooker appeared in the November 1972 issue of the Playboy magazine. Since then it is widely used as a standard test image in the field of image processing. Since gender equality is nowadays considered as more important compared to the nineteen seventies, the use of it is controversially discussed today.

Variational-based Image Segmentation

Variational-based methods have a strong impact in the context of image segmentation.

In general, the segmentation problem aims for the separation of the area of one object (or several objects) from the background (or from the other objects) of the image. Without loss of generality, we will mainly focus on the binary case, in which fore- and background separation is considered.

It is usually distinguished between edge-based and region-based segmentation approaches. Methods of the first class typically detect discontinuities of the brightness function as [Canny 1986], whereas the latter ones group similar parts together, see [Nock 2013]. We exemplary consider variational-based approaches of each class here, in particular, the edge-based "Snakes" active contour model from [Kass 1988] and region-based algorithms using a minimization of the Mumford-Shah functional, see [Mumford 1989].

Edge-based "Snakes" Active Contour Model

In the edge-based "Snakes" active contour model originally introduced in [Kass 1988], the total energy is composed of an external and an internal energy:

$$E(C) = E_{ext}(C) + E_{int}(C),$$

in which $C : [0,1] \to \Omega$ describes the curve which separates the external and the internal part.

Since we are searching for a large gradient at the boundary, we set up a negative external energy

$$E_{ext}(C) = -\int_0^1 \|\nabla I(C(s))\|^2 \, ds$$

and an internal energy

$$E_{int}(C) = \int_0^1 \left(\frac{\alpha}{2} \|\partial_s C(s)\|^2 + \frac{\beta}{2} \|\partial_s^2 C(s)\|^2 \right) ds.$$

The parameters α and β penalize the length of the boundary by penalizing the first and the second derivatives: $\partial_s C$ describing the so-called elastic length, whichs grows smaller for shorter curves and the stiffness $\partial_s^2 C$ of the curve penalizing a winding curve. Substitution of the integrand of E into the more general Euler-Lagrange equation for derivatives of higher order (Exercise 2.4 for $i = 2$) given by

$$\frac{\partial L}{\partial q} + \sum_{i=1}^{n} (-1)^i \frac{\mathrm{d}^i}{\mathrm{d}t^i} \frac{\partial L}{\partial q^{(i)}} = 0$$

leads to the gradient descent step

$$\frac{\partial C}{\partial t} = \nabla \|\nabla I(C)\|^2 + \alpha \partial_s^2 C - \beta \partial_s^4 C.$$

Since E is not convex, the computation of the global minimum $C^* = \arg\min_C(E(C))$ is hard to realize using a gradient descent procedure. Consider a white image with a black square in the middle: if the initial curve is set up outside the black square, it remains almost in its initial state. Otherwise, if it is initially set up inside the black square, the curve will shrink too much

In order to prevent from the first shortcoming, the image can be presmoothed to create a gradient unequal to zero. If the smoothing is realized adaptively in a way that it is reduced over time, this strategy can be interpreted as a graduated non convexity approach, see [Blake 1987]. To prevent from the strong shrinking effect, an additional negative ballooning energy term can be added which scales linearly with the area of the inner part, see [Cohen 1991].

Mumford-Shah Functional

The Mumford-Shah functional is given by

$$E(u, C) = \int_{\Omega} (I(\mathbf{x}) - u(\mathbf{x}))^2 \, d\mathbf{x} + \lambda \int_{\Omega \setminus C} \|\nabla u(\mathbf{x})\|^2 \, d\mathbf{x} + \nu \, \|C\|,$$

in which $u : \Omega \to \mathbb{R}$ describes an approximation of the image I and $C \subset \Omega$ a discontinuity set.

The first term measures the similarity to the original image, the second one penalizes explicitly at the non-boundary regions, and the last one penalizes the length of C.

For increasing λ, the approximation of u is forced to be smoother outside of C. In the limit case, we obtain a piecewise constant approximation of the image, such that u is constant in the separated regions. For constant $u_1, \ldots, u_n$, we get

$$E(\{u_1, \ldots, u_n\}, C) = \sum_{i=1}^{n} \int_{\Omega_i} (I(\mathbf{x}) - u_i)^2 \, d\mathbf{x} + \nu \, \|C\|.$$

Ising Model

For $n = 2$, it can be seen as an analog to the Ising model describing ferromagnetisms in solids, see [Lenz 1920, Ising 1925, Heisenberg 1928]. The two regions correspond to the two spins ± 1 with arbitrary spatial directions.

The length of C can be approximated by summing over all neighbor nodes:

$$\|C\| \approx \frac{1}{2} \sum_{i \sim j} \left(\frac{u_i - u_j}{2} \right)^2 = \frac{1}{8} \sum_{i \sim j} u_i^2 + u_j^2 - 2 u_i u_j = \text{const.} - \frac{1}{4} \sum_{i \sim j} u_i u_j,$$

so that the energy is described by

$$E(u) = \sum_i (I_i - u_i)^2 - \frac{\nu}{4} \sum_{i \sim j} u_i u_j,$$

in which the last term denotes the so-called Ising energy. The spins u should be aligned with an external field described by I. For $n = 2$, the global optimum can efficiently be computed in polynomial time. For $n > 3$, the optimization problem is NP-complete.

Region-based Image Segmentation

According to [Zhu 1995] we consider

$$E(C) = \int_{int(C)} f(x,y)\,\mathrm{d}x\mathrm{d}y$$

for $C : [0,1] \rightarrow \mathbb{R}^2$, $C(s) = (x(s), y(s))$. Green's theorem states that

$$\int_{int(C)} (\nabla \times \mathbf{v})\,\mathrm{d}^2 x = \int_C \mathbf{v}\,\mathrm{d}s$$

holds for a vector field $\mathbf{v} := (a(x,y), b(x,y)) \in \mathbb{R}^2$ and a closed boundary C. The rotation of v is given by $\nabla \times \mathbf{v} = \partial_x b - \partial_y a$ so that we get

$$\int_{int(C)} (b_x - a_y)\,\mathrm{d}x\mathrm{d}y = \int_C a\,\mathrm{d}x + b\,\mathrm{d}y.$$

Changing the vector field **v** such that $f = b_x - a_y$ leads to

$$E(C) = \int_{int(C)} f \, dx \, dy = \int_C a \, dx + b \, dy = \int_0^1 (ax' + by') \, ds =: \int_0^1 L(x, x', y, y') \, ds$$

with $x' := d_s x$ and $y' := d_s y$.

The applications of the Euler-Lagrange approach leads to

$$\delta_x L = fy', \quad \delta_y L = -fx'$$

corresponding to the gradient descent procedure

$$\partial_C E = f(x, y) \cdot \begin{pmatrix} y' \\ -x' \end{pmatrix}.$$

We consider two regions and define the energy by

$$E(u, C) = \int_{int(C)} (I(\mathbf{x}) - u_{int})^2 \, d\mathbf{x} + \int_{ext(C)} (I(\mathbf{x}) - u_{ext})^2 \, d\mathbf{x}$$

with model intensities u_{int} and u_{ext} and simply obtain the gradient descent

$$\partial_t C(s, t) = -\partial_C E = -\left((I(\mathbf{x}) - u_{int})^2 - (I(\mathbf{x}) - u_{ext})^2 \right) \cdot \mathbf{n}_C$$

with region normal $\mathbf{n}_C$.

This describes a region competition in which an outwards motion is performed if $|I - u_{int}| < |I - u_{ext}|$ and an inwards motion if $|I - u_{int}| > |I - u_{ext}|$.

Region-based Image Segmentation

The energy defined in [Zhu 1996],

$$E(C) = \int_{int(C)} (I(\mathbf{x}) - u_{int})^2 \, d\mathbf{x} + \int_{ext(C)} (I(\mathbf{x}) - u_{ext})^2 \, d\mathbf{x} + \nu \, \|C\|,$$

leads to the gradient descent

$$\partial_t C(s, t) = \left((I - u_{ext})^2 - (I - u_{int})^2 - \nu \kappa_C \right) \cdot \mathbf{n}_C.$$

in which $\kappa_C := \|d_s \mathbf{T}\|$ describes the local curvature defined as the change of the unit tangent $\mathbf{T}$.

The so-called diffusion snakes of [Cremers 2002] minimize the energy,

$$E(u, C) = \int_{\Omega} (I(\mathbf{x}) - u(\mathbf{x}))^2 \, d\mathbf{x} + \lambda \int_{\Omega \setminus C} \|\nabla u(\mathbf{x})\|^2 \, d\mathbf{x} + \nu \int_0^1 (C'(s))^2 \, ds,$$

using the gradient descent

$$\begin{aligned}
\partial_t C(s, t) &= -\partial_C E = \left((I - u)^2 + \lambda \|\nabla u\|^2\right) \cdot \mathbf{n}_C + 2\nu \|C''\|, \\
\partial_t C(s, t) &= -\partial_u E = \lambda \nabla(w_c \nabla u) + (I - u)
\end{aligned}$$

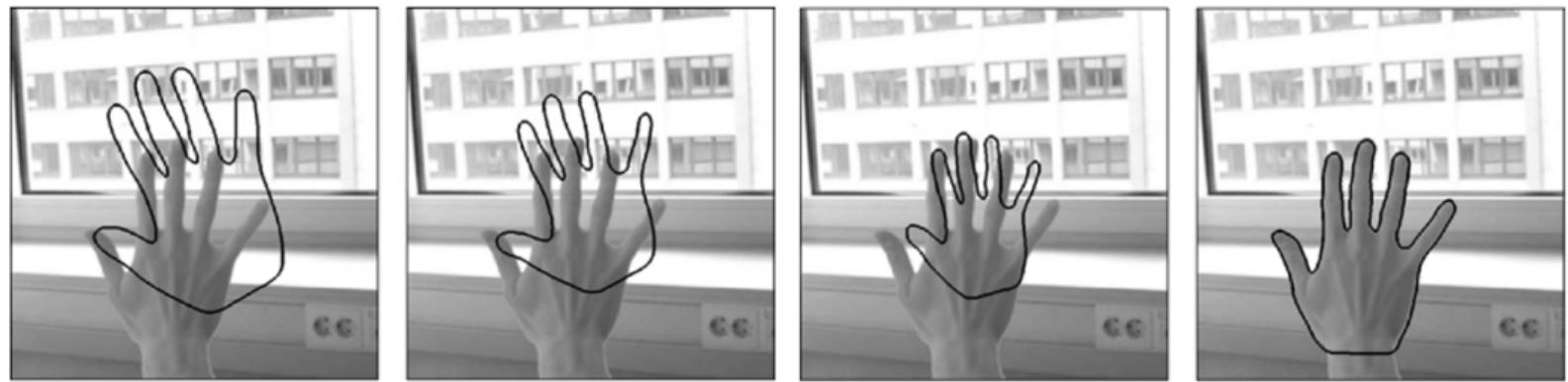with $w_c(\mathbf{x}) = 0$ if $\mathbf{x} \in C$ and $w_c(\mathbf{x}) = 0$ else.

Figure : Illustration of a diffusion snakes-based segmentation process from [Cremers 2002].