Chapter 6
Image Segmentation II: Variational Approaches

Variational Methods for Computer Vision

Winter 2013/14

Prof. Daniel Cremers
Chair for Computer Vision and Pattern Recognition
Department of Computer Science
Technische Universität München

Overview

Variational Image Segmentation

We already studied a number of segmentation algorithms. They were based on two complementary concepts:

- Detecting discontinuities of the brighness function, or
- Grouping pixels of similar brightness (color, texture, etc.)

Most of the approaches discussed so far lack a clear optimization criterion: Edge regions are heuristically fused to connected lines (Perkins, Canny), or pixels are iteratively merged to regions (region merging, region growing).

Toward the end of the 1980s, the first variational formulations for image segmentation emerged, in particular:

- the Snakes (Kass, Witkin, Terzopoulos, *Int. J. of Comp. Vision '88*),
- the Mumford-Shah Functional (Mumford, Shah, *J. Appl. Math. '89*).

Snakes

In 1988, Kass, Witkin and Terzopoulos proposed to minimize the following functional:

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

with an external energy

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

and an internal energy

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

Here, $I : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ denotes the input image, and $C : [0,1] \to \Omega$ denotes a parametric curve. C_s and C_{ss} denote the first and second derivative of the curve C with respect to its parameter s.

Snakes: External Energy

The external energy:

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

measures for a given curve C how well it coincides with the maxima of the brightness gradient $|\nabla I|$.

Thus rather than first searching for these maxima and then grouping them to a curve one defines a cost function which measures the “edge strength” along any conceivable curve.

Subsequently, the optimal curve $\hat{C}$ is determined by minimizing the total energy:

$$\hat{C} = \arg\min_C \; E(C)$$

Gradient descent on this energy induces an evolution of the curve toward locations of large image gradient.

Snakes: Internal Energy

The internal energy is a regularizer which induces some smoothness on the computed curves:

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

It consists of two components, weighted by parameters $\alpha \geq 0$ and $\beta \geq 0$, which penalize the elastic length and the stiffness of the curve.

Minimizing the total energy

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

leads to curves which are short and stiff while passing through locations of large gradient.

Snakes: Gradient Descent

The Snakes energy

$$E(C) = -\int_0^1 |\nabla I(C)|^2 \, ds \; + \; \int_0^1 \left\{ \frac{\alpha}{2} \left| C_s(s) \right|^2 + \frac{\beta}{2} \left| C_{ss}(s) \right|^2 \right\} ds$$

is of the canonical form

$$E(C) = \int \mathcal{L}(C, C_s, C_{ss}) ds$$

The corresponding Euler-Lagrange equation is given by:

$$\frac{dE}{dC} = \frac{\partial \mathcal{L}}{\partial C} - \frac{d}{ds} \frac{\partial \mathcal{L}}{\partial C_s} + \frac{d^2}{ds^2} \frac{\partial \mathcal{L}}{\partial C_{ss}} = -\nabla |\nabla I(C)|^2 - \alpha C_{ss} + \beta C_{ssss} = 0.$$

Consequently, the gradient descent equation reads:

$$\frac{\partial C(s,t)}{dt} = -\frac{dE(C)}{dC} = \nabla |\nabla I(C)|^2 + \alpha C_{ss} - \beta C_{ssss}$$

Some Comments on the Snakes

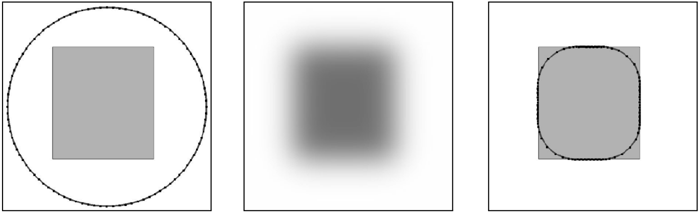
The Snakes are among the most influential publications in image processing. To date (Nov ’13) they have acquired more than 15000 citations. In 2005, the three authors were awarded an Academy Award for realistic simulations of textiles.

The Snakes are considered the first variational approach to image segmentation.

In comparison to modern segmentation methods, however, they are only of limited practical use:

- Real images typically have many gradient maxima which induce local minima in the cost function E. In practice, the curve must therefore be initialized sufficiently close to the desired solution. Alternatively one can presmooth the input image (to remove spurious local minima). Yet, the smoothing also removes possibly important edge information.
- The evolution of parametric curves is a numerically challenging problem as one needs to avoid self-intersections and instabilities.

Problem with Initialization



Input (square) and initial curve | smoothed input | final segmentation

Presmoothing makes the edge information “visible” from a larger distance. Yet it removes finer structures like the corners of the square.

(Author: D. Cremers)

Possible solutions

Local minima of the Snakes can be avoided in various ways:

- Ballooning: One extends the cost functional by the balloon energy

$$E_{balloon}(C) = \gamma \int_{\Omega_{int}(C)} d^2x$$

(Cohen & Cohen, Balloons, 1991), which induces the curve to contract (for $\gamma > 0$), or to expand (for $\gamma < 0$), because the balloon energy simply measures the area of the region Ω_{int} inside the curve C.

- Coarse-to-fine optimization: One minimizes the Snakes energy starting with a coarse (smoothed image) and iteratively reducing the smoothness, on each level initializing with the previously estimated curve. (See Blake & Zisserman, Graduated Non-Convexity, 1987).

- Global optimization: One reformulates the optimization problem and computes globally optimal solutions (using graph cut methods or convex relaxation methods).

The Mumford-Shah Approach

In 1989, Mumford and Shah proposed to compute a piecewise smooth approximation u of the input image $I : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ by minimizing the functional:

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

jointly with respect to an approximation $u : \Omega \to \mathbb{R}$ and a one-dimensional discontinuity set $C \subset \Omega$. The three terms have the following meaning:

- The data term assures that u is a faithful approximation of the input I.
- The smoothness term, weighted by $\lambda > 0$, assures that u is smooth everywhere except for the discontinuity set.
- A further regularizer, weighted by $\nu > 0$, assures that this discontinuity set has minimal length $|C|$.

The Piecewise Constant Mumford-Shah

For increasing values of the weight λ, the approximation u is forced to be smoother and smoother outside of C. In the limit $\lambda \to \infty$ we obtain a piecewise constant approximation of the image I:

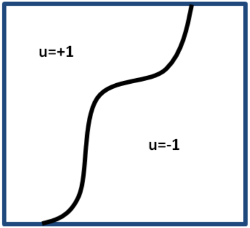
$$E(u, C) = \int_{\Omega} \big(I(x) - u(x)\big)^2 \, dx \; + \; \nu |C|,$$

where $u(x)$ is constant in each of the regions separated by the boundary C. If we denote these regions by $\{\Omega_1, \dots, \Omega_n\}$ and the constants by u_i, this can be rewritten as:

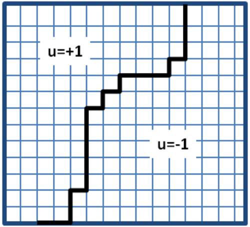
$$E(\{u_1, \dots, u_n\}, C) = \sum_{i=1}^{n} \int_{\Omega_i} \big(I(x) - u_i\big)^2 \, dx \; + \; \nu |C|,$$

For the case of two regions, a spatially discrete formulation of this energy is known as the Ising model (Lenz 1920, Ising 1925, Heisenberg 1928). It models the phenomenon of ferromagnetism and is among the most studied models in statistical physics.

Discrete Approximation



continuous representation



discrete representation

The length of the curve C can be approximated as:

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

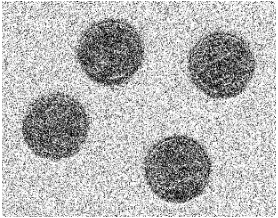
with summation over neighboring pixels i and j. This leads to:

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

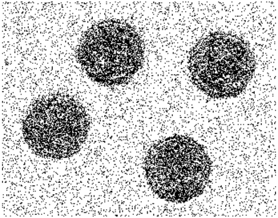
E. Ising, “Beitrag zur Theorie des Ferromagnetismus”, 1925.

Solution via Graph Cuts

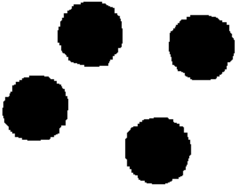
Original f



noisy: $I = f + \eta$



thresholding: $I > \theta$



$\arg\min E(u)$

Minimization of the discrete two-region model using graph cuts

Ernst Ising

1925

1995

Ernst Ising * 1900 in Cologne, † 1998 in Peoria, Illinois)

Doctoral thesis with Wilhelm Lenz in Hamburg, emigration to the US in 1947.

The Piecewise Constant Mumford-Shah

For $\lambda \to \infty$ one obtains a special case of the Mumford-Shah functional known as the piecewise constant approximation:

$$E(\{u_1, \dots, u_n\}, C) = \sum_{i=1}^{n} \int_{\Omega_i} \left(I(x) - u_i\right)^2 dx \; + \; \nu |C|,$$

This functional is of interest for several reasons:

- It is the spatially continuous version of the discrete spin models (Lenz 1920, Ising 1925, Potts 1956, ...).
- It is quite powerful, yet mathematically well understood.

In the following, we will therefore discuss several aspects of this functional in more detail:

- Some important mathematical results.
- Euler-Lagrange equations and possible implementations.
- A statistical interpretation.

Mathematical Insights

For a given boundary C, the minimizing constants u_i are uniquely determined. The are given by the average brightness in each region:

$$\frac{\partial E}{\partial u_i} = 2 \int_{\Omega_i} \big(I(x) - u_i\big) dx = 0 \;\Rightarrow\; u_i = \frac{\int_{\Omega_i} I(x) dx}{\int_{\Omega_i} dx}$$

As a result, the cost function is merely a function of the boundary C: $E(C) \equiv \min_u E(C, u)$. In particular, the segmentation has the same mean intensity as the input image.

Existence of minima: There exist minima of the functional $E(C)$. The minimizing boundaries C are closed and differentiable up to discontinuities of the following type:

- Three boundary segments meet at equal angles (120°).
- The boundary meets the domain boundary at a 90° angle.

Minimia of $E(C)$ are generally not unique.

Euler-Lagrange Equations

Unfortunately, the Mumford-Shah functional in its original formulation is not in a canonical form, since the variable of interest (the boundary C) appears in the integrand.

There exists an entire research community dedicated to such optimization problems known as shape optimization or shape sensitivity analysis.

In the following we will derive the Euler-Lagrange equation using Green's theorem (following S.C. Zhu '95).

Assume we are given an energy of the form

$$E(C) = \int\limits_{int(C)} f(x, y)\,dxdy,$$

where $int(C)$ denotes the region inside a curve C. Let $C : [0, 1] \to \mathbb{R}^2$ be a parametric closed curve, with $C(s) = \big(x(s),\, y(s)\big)$.

Euler-Lagrange Equations

Green's Theorem: For a vector field of the form $\vec{v} = (a(x, y),\ b(x, y)) \in \mathbb{R}^2$ and a closed boundary $C \subset \Omega$ we have:

$$\int\limits_{int(C)} (\nabla \times \vec{v}) d^2x = \int\limits_{C} \vec{v} ds,$$

where the rotation of v is defined as $\nabla \times \vec{v} \equiv \partial_x b - \partial_y a$. Thus:

$$\int\limits_{int(C)} (b_x - a_y) dxdy = \int\limits_{C} adx + bdy$$

Chosing a vector field $\vec{v}$ such that $f = (b_x - a_y)$, we can rewrite the energy in the canonical form:

$$E(C) = \int\limits_{int(C)} f\, dxdy = \int\limits_{C} adx + bdy = \int\limits_{0}^{1} (a\dot{x} + b\dot{y}) ds \equiv \int\limits_{0}^{1} \mathcal{L}(x, \dot{x}, y, \dot{y}) ds.$$

where $\dot{x} \equiv \frac{dx(s)}{ds}$ and $\dot{y} \equiv \frac{dy(s)}{ds}$.

Euler-Lagrange Equations

The functional

$$E(C) = \int_{int(C)} f(x,y)\,dxdy = \int_0^1 (a\dot{x} + b\dot{y})ds \equiv \int_0^1 \mathcal{L}(x,\dot{x},y,\dot{y})ds.$$

is equal to an integral along the curve C and we can compute the functional derivative with respect to $C(s) = (x(s),\, y(s))$:

$$\frac{\partial \mathcal{L}}{\partial x} - \frac{d}{ds}\frac{\partial \mathcal{L}}{\partial \dot{x}} = \frac{\partial a}{\partial x}\dot{x} + \frac{\partial b}{\partial x}\dot{y} - \frac{d}{ds}a = \left(\frac{\partial b}{\partial x} - \frac{\partial a}{\partial y}\right)\dot{y} = f\dot{y}$$

$$\frac{\partial \mathcal{L}}{\partial y} - \frac{d}{ds}\frac{\partial \mathcal{L}}{\partial \dot{y}} = \frac{\partial a}{\partial y}\dot{x} + \frac{\partial b}{\partial y}\dot{y} - \frac{d}{ds}b = \left(-\frac{\partial b}{\partial x} + \frac{\partial a}{\partial y}\right)\dot{x} = -f\dot{x}$$

In summary we obtain the simple functional gradient:

$$\frac{dE}{dC} = f(x,y)\begin{pmatrix}\dot{y}\\ -\dot{x}\end{pmatrix} = f(x,y)\,\vec{n}_C, \quad \text{where } \vec{n}_C = \text{outer normal}$$

Minimizing the Mumford-Shah Functional

The above calculation shows that functionals of the form $E(C) = \int_{int(C)} f(x,y)dxdy$ have the functional derivative:

$$\frac{dE}{dC} = f(x,y)\vec{n}_C.$$

For the piecewise constant Mumford-Shah functional (without boundary length term) and only two regions separated by a curve C we have:

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

so the functional derivative is given by:

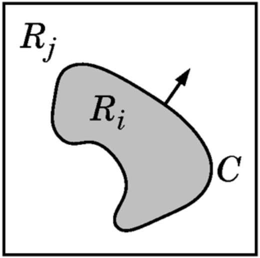
$$\frac{dE}{dC} = \Big((I(x) - u_{int})^2 - (I(x) - u_{ext})^2\Big)\vec{n}_C,$$

because both regions contribute to the gradient and the outer normal for the outside region is simply $-\vec{n}_C$.

Gradient Descent

The gradient descent equation is therefore:

$$\frac{\partial C(s,t)}{\partial t} = -\frac{dE(C)}{dC} = \left((I - u_{ext})^2 - (I - u_{int})^2 \right) \vec{n}_C.$$



At each boundary point:

Displace the curve

- outwards, if $|I - u_{int}| < |I - u_{ext}|$
- inwards, if $|I - u_{int}| > |I - u_{ext}|$

Intuition (Zhu & Yuille, Region Competition, PAMI '96):

If the local brightness $I(x)$ at point x is more similar to the average brightness of the interior then x is assigned to the interior (the curve moves outward) and vice versa.

Gradient Descent with Length Regularity

For the two-region piecewise constant Mumford-Shah with length regularity we get:

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

and the gradient descent reads:

$$\frac{\partial C(s,t)}{\partial t} = -\frac{dE(C)}{dC} = \Big((I - u_{ext})^2 - (I - u_{int})^2 - \nu \kappa_C \Big) \vec{n}_C,$$

where κ_C denotes the local curvature of the curve C.

This means that in addition to separating bright and dark areas, the evolution aims at suppressing large curvature of the curve. This is what leads to a local minimization of the boundary length $|C|$.

Implementations

The paper of Mumford and Shah is focused on aspects of existence and uniqueness of solutions and the study of properties of solutions. For example, it is shown that triple junctions can only exist in the minimizer if the contours meet at equal (120°) angles.

The paper of Mumford and Shah does not propose a numerical implementation for finding minimizers.

There exist a number of alternative methods, for example:

- *Koepfler et al. '95, Multiscale Algorithm:* an implementation of the piecewise constant model in the spirit of region merging based on the notion of 2-normal segmentations.

- *Cremers et al. '02, Diffusion Snakes:* Implementation of the piecewise smooth and piecewise constant models using closed parametric spline curves (hybrid of the Mumford-Shah and the Snakes).

2-normal Segmentations

Definition (Koepfler et al.'95):

A 2-normal segmentation is a partitioning of the image plane Ω into pairwise disjoint regions $\Omega_1, \dots, \Omega_n$, such that each segmentation obtained by merging two neighboring regions has a larger or equal energy (in the sense of the piecewise constant Mumford-Shah energy).

The algorithm of Koepfler et al. allows to compute 2-normal segmentations. To this end it iteratively merges neighboring regions until convergence.

Minima of the piecewise constant Mumford-Shah are always 2-normal segmentations.

However: Not all 2-normal segmentations are minimizers of the Mumford-Shah functional.

Two questions arise:

- In which order should one merge neighboring regions?
- How should one select the parameter ν?

Properties of 2-normal Segmentations

Koepfler et al. show a number of properties of 2-normal segmentations which provide an intuitive understanding of the optimization problem.

Proposition: The number n of regions of a 2-normal segmentation is bounded by the following function of the scale parameter ν:

$$n \leq \frac{|\Omega| \text{osc}(I)^4}{c_0 \, \nu^2}.$$

$|\Omega| \equiv$ image size, and $\text{osc}(I) \equiv \sup(I) - \inf(I)$ is called the oscillation of the brightness function (difference between largest and smallest brightness).

In particular, this implies: The smaller Ω and the larger ν, the stronger the constraint on the number of regions.

Thus, the scale parameter ν defines the spatial scale on which segmentation is performed – on a coarse scale for large ν and on a finer scale for smaller values of ν.

Properties of 2-normal Segmentations

Moreover, one can compute bounds on the size and shape of segments.

Proposition: The individual segments of a 2-normal segmentation have a a positive minimal size:

$$|\Omega_i| \geq c_1(I, \nu, \Omega) \quad \forall i.$$

In particular, this implies that the Mumford-Shah segmentation process (for $\nu > 0$) leads to an elimination of small regions. In historical approaches this was introduced through a heuristic post-processing step.

Proposition: For every individual segment Ω_i the length of its boundary $\partial\Omega_i$ is bounded by a mutiple of its area $|\Omega_i|$:

$$|\partial\Omega_i| \leq c_2(I, \Omega)\, |\Omega_i| \quad \forall i.$$

This implies that minimization of the Mumford-Shah functional (for $\nu > 0$) also leads to an elimination of elongated regions.

Multiscale Implementation

The implementation of Koepfler et al. allows to compute a 2-normal segmentation by the following region merging process:

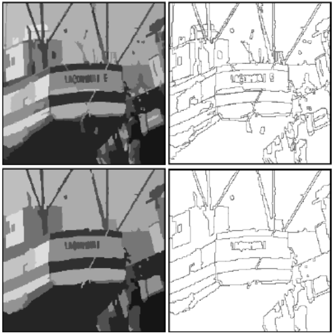
1. Initialize: every pixel is its own region.
2. For all neighboring pairs of regions, compute the change ΔE in energy obtained by merging the two regions. Obviously it is of the form:

$$\Delta E = \Delta E_{region} + \nu \Delta E_{length}$$

3. For all pairs of adjacent regions determine the value $\hat{\nu}$, for which there is an energy decrease ($\Delta E < 0$). This value always exists because $\Delta E_{length} < 0$ and $\Delta E_{region} \geq 0$.
4. In each step, merge the region pair with the smallest value of $\hat{\nu}$.
5. Repeat steps (2.) - (4.) until the desired number of regions or a sufficiently large value of ν is reached.

Multiscale Implementation

Input



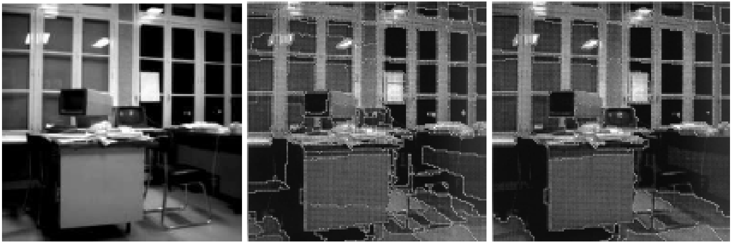
Segmentations

Segmentations for $\nu = 2022$ (above) and $\nu = 6173$ (below)

Koepfler et al., SIAM J. of Numer. Analysis '95

http://www.math-info.univ-paris5.fr/~gk/papers/heidelberg95.pdf

Multiscale Implementation



Original and segmentation with $\nu = 1024$ and $\nu = 4096$

Koepfler, Morel, Solimini '95

Diffusion Snakes

The Diffusion Snakes minimize the functional

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

by alternating two gradient descent evolutions:

$$\frac{\partial C(s,t)}{\partial t} = -\frac{\partial E}{\partial C} = \Big((I-u)^2 + \lambda |\nabla u|^2\Big)\vec{n} + 2\nu C'', \quad \vec{n} = \text{normal}$$

$$\frac{\partial u(x,t)}{\partial t} = -\frac{\partial E}{\partial u} = \lambda \nabla(w_c \nabla u) + (I-u), \quad w_c(x) = \begin{cases} 0, & x \in C \\ 1, & \text{else} \end{cases}$$

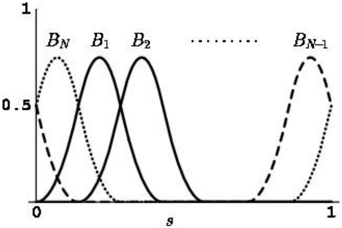
We therefore have a curve evolution in alternation with an inhomogeneous diffiusion process (constant diffusion inside regions, no diffusion acrosss boundary). Thus the Diffusion Snakes simultaneously perform denoising (in each of the separated regions) and boundary estimation.

Diffusion Snakes

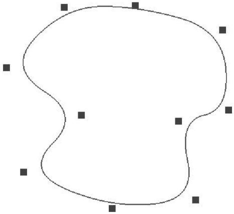
The evolution of the curve C is implemented by evolving a finite number of control points $p_1, \ldots, p_n \in \mathbb{R}^2$:

$$C(s,t) = \sum_{i=1}^{n} p_i(t)\, B_i(s).$$

Here $B_i(s)$ are spline basis functions:

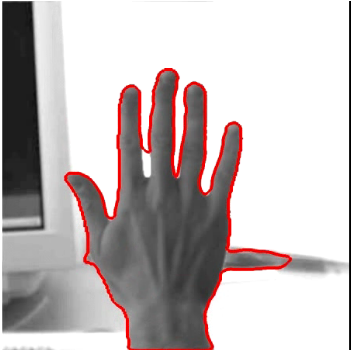


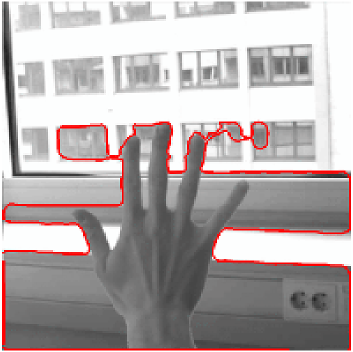
spline basis functions



Spline & control points

Diffusion Snakes





ν large $\qquad$ ν small

D. Cremers et al., Int. J. of Computer Vision, 2002