Chapter 7

Image Segmentation III: Bayesian Inference

Computer Vision I: Variational Methods

Winter 2016/17

Prof. Daniel Cremers
Chair for Computer Vision and Pattern Recognition
Departments of Informatics & Mathematics
Technical University of Munich

Overview

1. **Bayesian Inference**
2. **ML versus MAP**
3. **Mode versus Mean**
4. **Probabilistic Image Segmentation**
5. **Mumford-Shah as MAP Estimation**
6. **Statistical Shape Priors**

Review: The Bayesian Formula

For two random variables A and B we have:

$$\mathcal{P}(A \cap B) = \mathcal{P}(A \mid B)\, \mathcal{P}(B) = \mathcal{P}(B \mid A)\, \mathcal{P}(A).$$

where:

$\mathcal{P}(A \cap B)$ = probability that both events A *and* B occur.

$\mathcal{P}(A \mid B)$ = conditional probability for A given B.

More precisely: $\mathcal{P}(A) \equiv \mathcal{P}(\hat{A} = A)$ denotes the probability that the random variable $\hat{A}$ takes on the value A.

Rewriting the above equation we obtain the Bayesian formula:

$$\boxed{\mathcal{P}(A \mid B) = \frac{\mathcal{P}(B \mid A)\, \mathcal{P}(A)}{\mathcal{P}(B)}}$$

It is the foundation of numerous statistical approaches (statistical inference, Bayesian decision theory).

Thomas Bayes

Reverend Thomas Bayes (1702 – 1761)

- “Essay towards solving a problem in the doctrine of chances”, *Phil. Trans. Roy. Soc. of London* (1764),
- ‘Proposition 9’ is referred to as the Bayesian formula,
- Fellow of the Royal Society (1742).

Statistical Inference

Assume we are interested in the state S of a system and obtain the measurement M containing some information on S. In our case S may be a segmentation (or even semantic decomposition of a scene) and M the image (color values at all pixels). We have:

$$\mathcal{P}(S \mid M) = \frac{\mathcal{P}(M \mid S)\, \mathcal{P}(S)}{\mathcal{P}(M)}$$

where:

$\mathcal{P}(S \mid M)$:	posterior probability,
$\mathcal{P}(S), \mathcal{P}(M)$:	prior probabilities,
$\mathcal{P}(M \mid S)$:	likelihood function.

The likelihood function $\mathcal{P}(M \mid S)$ describes the image formation process (probability of measuring the image M given the system state S). The Bayesian formula provides an inversion of the image formation process in a statistical framework.

Max. Likelihood vs. Max. A Posteriori

Maximum Likelihood (ML) estimation:

Assume we know for any given state S how likely respective measurements M are. The we can assign to a measurement M the state S for which the probabilty of M is largest:

$$S_{ML} = \arg\max_S \mathcal{P}(M \mid S)$$

Maximum A Posteriori (MAP) estimation:

If, in addition, we know the a priori probability of different system states S, we can estimate the state S which is most likely for a given measurement M:

$$S_{MAP} = \arg\max_S \mathcal{P}(S \mid M) = \arg\max_S \frac{\mathcal{P}(M \mid S)\, \mathcal{P}(S)}{\mathcal{P}(M)}$$

For $\mathcal{P}(S) = \text{const.}$ ("uniform prior"), ML and MAP are equivalent.

Mode versus Mean?

While the posterior provides the probability for all conceivable interpretations of the data, in practice a deterministic answer is often desirable. For example, when tracking an object one may want to know its exact location $x(t)$ at time t, rather than some probability distribution $p(x,t)$.

Two popular methods to do this are based on computing the mode S_{mode} of the distribution (i.e. the state of maximum probability) (MAP) or the mean S_{mean} of the distribution:

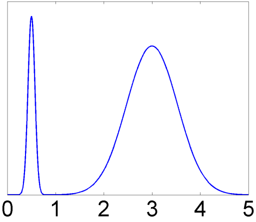
$$S_{mode} = \arg\max_S \mathcal{P}(S \mid M), \qquad S_{mean} = \int S\, \mathcal{P}(S \mid M)\, dS.$$

In general these values are different:

For example:

$S_{mode} = 0.5,$

$S_{mean} = 2.65$



Probabilistic Image Segmentation

We will derive the piecewise constant Mumford-Shah functional in a Bayesian framework (discrete: Besag 1974, continuous: Zhu, Yuille 1996, Brox, Cremers 2009).

Assume that the observed scene is made up of n regions $R_1, \dots, R_n$ in which the brightness values are independent samples drawn from a Gaussian distribution

$$\forall x \in R_i : \quad \mathcal{P}_{\mu_i,\sigma_i}(I(x)) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(I(x)-\mu_i)^2}{2\sigma_i^2}\right),$$

where μ_i and σ_i denote mean and standard deviation of this distribution. This is a generative model, meaning that for a given set of regions R_i and model parameters μ_i, σ_i, one can synthesize images by sampling from the above distributions.

In order to segment an image $I(x)$ one computes $S \equiv \{R_1, \dots, R_n, \mu_1, \dots, \mu_n, \sigma_1, \dots, \sigma_n\}$ using the Bayes approach:

$$S_{MAP} = \arg\max_S \mathcal{P}(S \mid I) = \arg\max_S \mathcal{P}(I \mid S)\, \mathcal{P}(S)$$

Probabilistic Image Segmentation

Let us first assume that the image plane Ω is a discrete set of pixels. Then the data likelihood can be written as

$$\mathcal{P}(I \,|\, S) = \prod_{x \in \Omega} \mathcal{P}(I(x) \,|S) = \prod_{i=1}^{n} \prod_{x \in R_i} \mathcal{P}_{\mu_i, \sigma_i}(I(x))$$

The prior $\mathcal{P}(S)$ states how likely *a priori* are different decompositions in terms of regions R_i and brightness parameters μ_i and σ_i. We now assume that all brightness values have equal probability and that a decomposition into regions $\{R_1, \ldots, R_n\}$ is more likely if the separating boundary C has a shorter length $|C|$:

$$\mathcal{P}(S) \propto \exp\left(-\nu |C|\right)$$

Up to a constant, we therefore have:

$$E = -\log \mathcal{P}(S \,|\, I) = -\log \left(\prod_{i=1}^{n} \prod_{x \in R_i} \mathcal{P}_{\mu_i, \sigma_i}(I(x)) \right) - \log \mathcal{P}(S)$$

Probabilistic Image Segmentation

$$E(S) = -\log \mathcal{P}(S \mid I) = -\log \left(\prod_{i=1}^{n} \prod_{x \in R_i} \mathcal{P}_{\mu_i,\sigma_i}(I(x)) \right) - \log \mathcal{P}(S)$$

$$= \sum_{i=1}^{n} \sum_{x \in R_i} \left(-\log \mathcal{P}_{\mu_i,\sigma_i}(I(x)) \right) + \nu |C|$$

$$= \sum_{i=1}^{n} \sum_{x \in R_i} \left[\frac{(I(x)-\mu_i)^2}{2\sigma_i^2} + \log \left(\sqrt{2\pi}\sigma_i \right) \right] + \nu |C|$$

In a spatially continuous setting we have:

$$E(S) = -\log \mathcal{P}(S \mid I) = -\log \left(\prod_{i=1}^{n} \prod_{x \in R_i} \left[\mathcal{P}_{\mu_i,\sigma_i}(I(x)) \right]^{dx} \right) - \log \mathcal{P}(S)$$

$$= \sum_{i=1}^{n} \left(\frac{1}{2\sigma_i^2} \int_{R_i} \left(I(x) - \mu_i \right)^2 dx + |R_i| \log \left(\sqrt{2\pi}\sigma_i \right) \right) + \nu |C|$$

The exponent dx is introduced to assure the correct continuuum limit (Cremers et al. IJCV '07).

Mumford-Shah as MAP Estimation

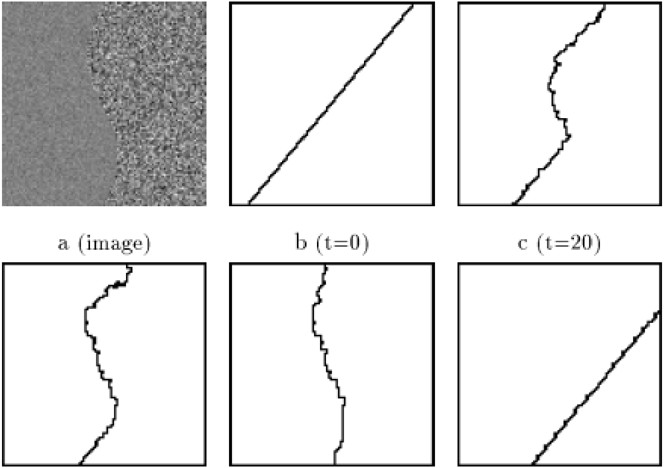
The above derivation shows that under the assumption of Gaussian distributed intensity in each region and a prior favoring small boundary length the maximum aposteriori estimate of the segmentation amounts to the minimizer of the Mumford-Shah-like functional:

$$E(C, \{\mu_i, \sigma_i\}) = \sum_{i=1}^{n} \left(\int_{R_i} \frac{\left(I(x) - \mu_i\right)^2}{2\sigma_i^2} + \log\left(\sigma_i\right) dx \right) + \nu |C|.$$

In this sense, Mumford-Shah segmentation has a clear statistical interpretation. Moreover, we observe that:

- In contrast to the piecewise constant Mumford-Shah functional, in addition to the mean intensities μ_i we also have standard deviations σ_i.
- For different probability models representing the colors in each region, the above MAP estimation gives rise to a variety of different cost functionals.

Segmentation with Identical Means



a (image) b (t=0) c (t=20)

d (t=30) e (t=50) f (without F-test)

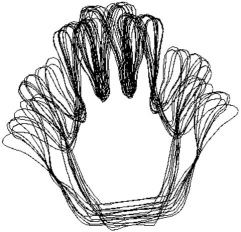
Segmentation of an image with same means μ_i,
but different standard deviations σ_i (Author: S.-C. Zhu, UCLA)

Statistical Shape Priors...

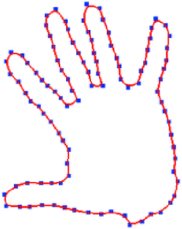
The commonly used boundary length $|C|$ corresponds to a prior favoring shorter boundaries.

When segmenting known objects, it may be better to impose object-specific shape priors. Such priors can be statistically learned from training shapes, mimicking the human capacity to learn from examples.

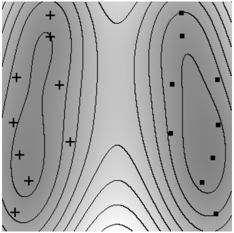
We can represent a set of training shapes by spline curves and estimate a probability density on the space of curves:



training silhouettes



spline encoding



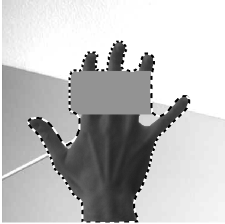
density in 2D projection

Cremers et al., IJCV 2002

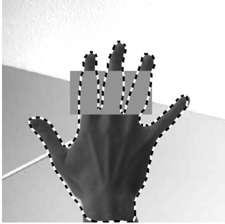
...for Image Segmentation

Imposing this prior in a variational segmentation method leads to segmentations which are a priori similar to the observed training shapes. The statistical shape prior assures that misleading background clutter is ignored and that missing curve parts are "filled in".

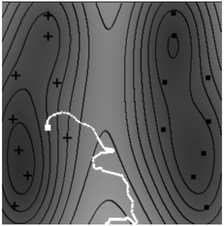
The image on the right shows the gradient descent curve evolution (white trajectory) in shape space.



with length prior



with statistical prior



curve evolution in shape space

Cremers et al., IJCV 2002

Segmentation with Statistical Shape Priors

$$E(C) = E_{diffsnake}(C) + E_{prior}(C) \;\rightarrow\; \min$$

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

Tracking of a 3D Object using only 2D Silhouettes

Cremers et al., ECCV 2002