Course: Model, Learning, and Inference: Lecture 7

Alan Yuille
Department of Statistics, UCLA
Los Angeles, CA 90095
yuille@stat.ucla.edu

Abstract

Snakes and Region Competition and DDMCMC
NOTE: NOT FOR DISTRIBUTION!!

1 Introduction

Snakes and Region Competition and DDMCMC.

2 Snakes

The paper on Snakes (Witkin, Terzouplos, Kass) isone of the most cited papers in vision. It was intended as an interactive tool and had many applications. It also helped motivated important work on object tracking.

2.1 Basic Snakes

A snake is a contour by $\Gamma(s)$ (s is not the arc length). Typically $\Gamma(s)$ is a closed contour. It is defined by an energy function:

$$E[\Gamma] = \int_\Gamma \frac{1}{2}\{\alpha|\vec{\Gamma}_s(s)|^2 + \beta|\vec{\Gamma}_{ss}|^2 - \lambda|\vec{\nabla}I(\vec{x}(s))|^2\}ds. \quad (1)$$

Here $_s$ denotes partial derivative with respect to s. $\Gamma_s = \frac{\partial\vec{\Gamma}}{\partial s}$.

To run a Snake, you initialize it by placing it near, or surrounding, and object (e.g. a face). Then minimize $E[\Gamma]$ by steepest descent. This gives update equations:

$$\frac{d\vec{\Gamma}(s)}{dt} = -\alpha\vec{\Gamma}_{ss} + \beta\vec{\Gamma}_{ssss} + \lambda\vec{\nabla}|\vec{\nabla}I|^2. \quad (2)$$

The first two terms try to make the curve short and have small curvature (NEED AN APPENDIX ON CURVATURE!!). The third term tries to move the snake to regions of large intensity gradient.

NOTE: BETTER ALGORITHMS THAN STEEPEST DESCENT ARE AVAILABLE!!

Snakes can be modified in several different ways. Basic snakes try to minimize their areas (consequence of the prior terms). But we can have *balloons* by putting in prior terms that try to make the area as big as possible. We can also add regional terms (check Ishihara and Geiger!!).

2.2 Bayesian Interpretation of Snakes

Bayesian interpretation. There is a natural Bayesian interpretation of snakes. This shows that the energy imaging term $-|\vec{\nabla} I|^2$ is not very sensible (as was realized in practice fairly soon). Technically this requires taking the continuous formulation and discretizing it (APPENDIX – DISCUSSION!!).

Specify a generative model for the image intensity gradients (see notes for lecture 1):

$$P(\{|\vec{\nabla} I|\}|\Gamma) = \prod_{\vec{x}\in\Gamma} P_{on}(|\vec{\nabla} I(\vec{x})|) \prod_{\vec{x}\in\Omega/\Gamma} P_{off}(|\vec{\nabla} I(\vec{x})|). \quad (3)$$

$$P(\Gamma) = \frac{1}{Z}\exp\{-\int_\Gamma \frac{1}{2}\{\alpha|\vec{\Gamma}_s(s)|^2 + \beta|\vec{\Gamma}_{ss}|^2\}\}. \quad (4)$$

MAP estimation of Γ corresponds to minimizing:

$$-\log P(\{|\vec{\nabla} I|\}|\Gamma) - \log P(\Gamma), \quad (5)$$

where we see that the second term reduces to the spatial terms in the energy function for the snakes.

We now concentrate on the first term which can be re-expressed as:

$$-\sum_{\vec{x}\in\Gamma}\log P_{on}(|\vec{\nabla} I(\vec{x})|) - \sum_{\vec{x}\in\Omega/\Gamma}\log P_{off}(|\vec{\nabla} I(\vec{x})|) = -\sum_{\vec{x}\in\Gamma}\log\frac{P_{on}(|\vec{\nabla} I(\vec{x})|)}{P_{off}(|\vec{\nabla} I(\vec{x})|)} - \sum_{\vec{x}\in\Omega} P_{off}(|\vec{\nabla} I(\vec{x})|). \quad (6)$$

We see that the second term is independent of Γ and hence can be dropped during MAP estimation. We then see that $\log\frac{P_{on}(|\vec{\nabla} I(\vec{x})|)}{P_{off}(|\vec{\nabla} I(\vec{x})|)}$ corresponds to the term $-|\vec{\nabla} I(\vec{x})|^2$. But we know what the typical shape of the log-likelihood function is from lecture 1 – and it certainly is not $-|\vec{\nabla} I(\vec{x})|^2$! Instead it rises slowly from a minimum at $|\vec{\nabla} I| = 0$ and (roughly) asymptotes. This is because if $|\vec{\nabla} I|$ is large then it is almost certainly an edge – but $-|\vec{\nabla} I(\vec{x})|^2$ rewards big edges too much at the expense of small edges.

3 Probabilistic Models of Images

The weak membrane models are models of natural images. We now discuss a richer class of models that were developed by Zhu, Lee, Yuille (1996) and Tu and Zhu (2002). These models lead to challenging inference and learning problems. The region competition model (Zhu et al 1996) is named after the inference algorithm. Tu and Zhu's later model (2002) required data driven Markov Chain Monte Carlo (DDMCMC) in order to perform inference. Both models were later extended to image parsing (Tu, Chen, Yuille, Zhu 2003, 2005) and later work by Zhu's group.

The goal is to decompose the image domain Ω into M non-overlapping sub-domains $\{\Omega_i : i = 1, ..., M\}$ – such that $\bigcup_{i=1}^M \Omega_i = \Omega$ and $\Omega_i \bigcap \Omega_j = 0,\ i \neq j$. The boundaries of sub-region i is specified by $\partial\Omega_i$. We defined Γ to be all the boundaries. The number of regions M is a random variables – we do not know how many regions there will be in any image.

The region competition model (Zhu et al 1996) assumes that the intensity, or intensity features, in each subregion Ω_i are generated by a distribution parameterized by α_i. Formally we write $P(\{I(x,y) : (x,y) \in \Omega_i\}|\alpha_i)P(\alpha_i)$. A simple model requires that the image intensities are drawn independently at each position (x,y) from a Gaussian distribution – i.e. $\alpha_i = (\mu_i, \sigma_i^2)$, $P(\{I(x,y) : (x,y) \in \Omega_i\}|\mu_i, \sigma_i) = \prod_{(x,y)\in\Omega_i} N(I(x,y)|\mu_i, \sigma_i^2)$, where $N(I(x,y)|\mu_i, \sigma_i^2)$ is a Gaussian with mean μ_i and variance σ_i^2. By extending this formulation, we can also include the imaging term for the Mumford-Shah model – α_i corresponds to $\{J(x,y) : (x,y) \in \Omega_i\}$ and there is a Gaussian prior defined on J specified by $(1/Z)\exp\{-\int\int_{\Omega_i} dxdy \vec{\nabla} J(x,y)\cdot\vec{\nabla} J(x,y)\}$. NOTE: Zhu's work on minimax entropy learning was driven by the need to discover other, more realistic, models of images that could be used.

Putting everything together gives a model:

$$P(\{\Omega_i\}, M, \{\alpha_i\}) = \frac{1}{Z} \exp\{-E[\{\Omega_i\}, \{\alpha_i\}, M]\}, \tag{7}$$

where

$$E[\{\Omega_i\}, \{\alpha_i\}, M] = \sum_{i=1}^{M} \frac{\mu}{2} \int ds_{\partial\Omega_i} + \lambda M - \sum_{i=1}^{M} \log P(\{I(x,y) : (x,y) \in \Omega_i\}|\alpha_i) - \sum_{i=1}^{M} \log P(\alpha_i). \tag{8}$$

We can express the model differently in a more insightful way:

$$P(I|\Omega, \alpha) P(\alpha) P(\Omega|M) P(M), \tag{9}$$

where

$$P(M) = \frac{1}{Z_1} \exp\{-\lambda M\},$$
$$P(\Omega|M) = \frac{1}{Z_2} \exp\{-(\mu/2) \sum_{i=1}^{M} \int_{\partial\Omega_i} ds\},$$
$$P(I|\Omega, \alpha) = \prod_{i=1}^{N} P(\{I(x,y) : (x,y) \in \Omega_i\}|\alpha_i). \tag{10}$$

In other words, to generate an image you proceed in four steps: (I) sample M from $P(M)$ – to generate the number of regions in the image, (II) sample the shape of the regions $\{\Omega_i\}$ from $P(\Omega|M)$, (III) sample the parameters α_i for each region from $P(\alpha)$, and (IV) sample the images in each region from $P(\{I(x,y) : (x,y) \in \Omega_i\}|\alpha_i)$.

In practice, it is very hard to do this sampling. The most difficult step is to sample from $P(\Omega|M)$ (which has never been done as far as I know). But Zhu and his collaborators have sampled images from related models assuming that the boundaries are known, or estimated (see Tu and Zhu 2002).

NOTE: Tu and Zhu extend this model by having families of models – $P(I \in \Omega_i|\tau, \alpha)P(\alpha|\tau)P(\tau)$, where τ is the type of the model and α are its parameters. This allows us to use models for texture, for smoothly varying intensity, for junk, for smooth gradients. So it represents images much better. But it makes inference much harder and will require the Data Driven Markov Chain Monte Carlo (DDMCMC) algorithm (described later).

We can also interpret this model in terms of encoding the image. Shannon's information theory proposes that if data is generated by a distribution $P(x)$ then an example x should be encoded by $-\log P(x)$ bits. If the distribution has a parameter α, then it is best encoded by finding $\alpha^* = \arg\min_\alpha\{-\log P(x|\alpha)\}$, which is simply the maximum likelihood estimate. (MORE BACKGROUND ON SHANNON's THEORY!!).

3.1 Inference: Region Competition

Suppose the number M of regions is known. Then we can fix M and minimize $E(\Omega, \alpha, M)$ with respect to Ω and α alternatively.

This gives two (alternating) steps:

$$\text{Solve } \alpha_i^* = \arg\max\{\int\int_{\Omega_i} \log P(\alpha_i|I(x,y))dxdy\}, \ \ \forall i = 1, ..., M. \tag{11}$$

$$\frac{d\Gamma(s)}{dt} = \frac{-\mu}{2} \kappa_\nu \vec{n}_\nu + \log \frac{P(I(\nu)|\alpha_k)}{P(I(\nu)|\alpha_{k+1})} \vec{n}_\nu. \tag{12}$$

Here ν denotes a point of the boundary Γ between regions k and $k+1$, $\vec{n}_\nu$ is the normal to the boundary curve. This equation is derived as gradient descent of the functional (NEED some calculus of variations here!!).

The intuitions for these equations are clear. The regions k and $k+1$ compete for "ownership" of the pixels on the boundary between the regions (by the log-likelihood term) and the curvature terms tries to make the boundary as short (straight) as possible.

Example – Gaussian distributions!!

$$\frac{d\Gamma(s)}{dt} = \frac{-\mu}{2}\kappa_\nu \vec{n}_\nu - \frac{1}{2}\{\log\frac{\sigma_i^2}{\sigma_j^2} + \frac{(\overline{I}-\mu_i)^2}{\sigma_i^2} - \frac{(\overline{I}-\mu_j)^2}{\sigma_j^2} + \frac{s^2}{\sigma_i^2} - \frac{s^2}{\sigma_j^2}\}\vec{n}_\nu, \tag{13}$$

where $\overline{I}$ is the intensity mean in the window W and s^2 is the intensity variance.

NOTE: In practice, a window is used to get better estimate of the statistics (Zhu et al) – so we replace the log-likelihood term on the pixel by the log-likelihood over the window. OTHER EXAMPLE: Texture statistics I_x, I_y.

But this algorithm requires dealing with unknown M and requires an initialization.

The strategy (Zhu et al 1996): (i) initialize with many (small) regions, (ii) run the 2-stage iterative algorithm – this will make some regions disappear (get squashed) while others grow large, (iii) remove the squashed regions, (iv) merge bigger regions if their α's are similar and if this decreases the energy (tradeoffs – need to estimate a single α for the combined region, which increases the energy, but this may be compensated for by eliminating the boundary between the regions and the reduced penalty for number of regions).

FIGURE TO ILLUSTRATE THIS!!

3.2 Learning the Models

The work on region competition motivated Zhu to develop a theory for learning models of images and boundaries. He observed that the energy terms in the region competition energy function could be obtained by the maximum entropy principle. . He also observed that the empirical statistics for operators like I_x, I_y (partial derivatives) were non-Gaussian (even though Gaussian models were used in region competition). Empirically, the statistics of these derivatives had longer tails than Gaussians (REFER BACK TO THE STATISTICS AND MAX-ENTROPY IN EARLIER NOTES!!).

This is the standard learning without hidden variables. We define the distributions to be of exponential form and use standard methods from previous lectures. EXAMPLES from Texture, Shading, Junk. Learning the statistics of boundaries is also straightforward (Zhu 1999) but requires stochastic sampling to ensure learning.

Learning these models is straightforward if the data is segmented – i.e. you have examples that all come form the same region. If not, you can in theory apply EM to learn the models from training examples without segmentation. But nobody has down this.

Distributions on image or on image features? Easier to put distribution on features but this is sub-optimal in terms of performance but may be computationally far simpler!!