ECSE-626

Statistical Computer Vision

Bayesian Regularization

Introduction

- We have described several ill-posed inverse problems in early vision, and shown how they can be addressed using standard regularization.
- The main idea behind the approach is to restrict the class of admissible solutions by introducing suitable *a priori* knowledge.

Limitations of Standard Regularization Theory

- Standard regularization methods lead to satisfactory solutions to problems in early vision, but have trouble with general problems such as:
  - Discontinuities
  - Fusion of information from multiple modules

J. Marroquin, S. Mitter, and T. Poggio, “Probabilistic Solution of Ill-posed Problems in Computer Vision”, Journal of the American Statistical Association, March 1987, Vol. 82, No. 397, Theory and Methods, pp76-89.

Limitations of Standard Regularization Theory

- Standard regularization with linear models leads to solutions that are too smooth, and incorrect at *discontinuities*.
- These areas are the most critical locations for problems such as optical flow computation and stereo.
- Furthermore, linear Euler-Lagrange equations results in outputs of different modules being combined in a linear way.

Limitations of Standard Regularization Theory

- We would like to have a more general, comprehensive theory capable of dealing with these problems using formal mathematics.
- We would also like to be able to enumerate and represent all sources of uncertainty in the process.

Probabilistic Inference

- For the rest of this course, we will outline probabilistic approaches to addressing the ill-posedness of vision problems based predominantly on Bayes' Theory and Information Theory.
- The main advantages include:

Advantages

1. The presentation of a formal theory for **enumerating all sources of uncertainty** and representing each as a *probability density function*.

   For computer vision, this is important in that data measurements, models and inference are uncertain. We need to represent the uncertainties rather than ignore them.

Advantages

2. All **assumptions are made explicit** so that they can be examined to see if they are valid.

   Assumptions can be changed without altering the framework.

   Nothing is "swept under the rug".

Advantages

3. The presentation of a mathematical framework for **merging all sources of information** and presenting the result in the form of a *probability density function*.

   Rather than obtain a single solution, the confidence in the competing hypotheses are made explicit.

   Higher level processes can then examine the result to see if further processing is required.

Advantages

The sources of information are merged in a fashion that is guaranteed to be **consistent**.

This implies that should the same data and models be presented at a later date, the same posterior distribution will be obtained.

Furthermore, should the random variables be presented in a different order, the same conclusions should be attained.

Bayesian Regularization

- Let us revisit the inverse problem formulation presented by regularization papers and examine the Bayesian solution to the ill-posed inverse problem.
- We will then examine the parallelism to the Tikhonov approach.

Bayesian Regularization

Consider the inverse problem discussed previously:
We wish to determine the solution z to the inverse problem $Az = u$.

Given the uncertainty in the process, and the inherent non-uniqueness of most ill-posed problems, this approach seems unnecessarily restrictive to a Bayesian.

Bayesian Regularization

Rather than obtain a single z solution for inverse problem, the Bayesian approach is to obtain a posterior density function for z given the data u:

Bayesian Regularization

Rather than obtain a single z solution for inverse problem, the Bayesian approach is to obtain a posterior density function for z given the data u:

$$p(z \mid u) = \frac{p(u \mid z)\, p(z)}{p(u)}$$

Likelihood Function

$p(u \mid z)$ is the conditional probability density function for the data u given the solution z.

It is referred to as the *likelihood function*.

This probability represents our knowledge of the image formation process.

It also models the process by which uncertainty is introduced into the data.

Prior Information

$p(u)$ is the (*a priori*) probability density of the data before the measurement is made.

$p(z)$ is the (*a priori*) probability density of the solution before the measurement is made. This probability represents our prior knowledge of the solution. It specifies which solutions are more probable than others.

Posterior Information

$p(z|u)$ is known as the conditional *posterior* density function for z given the data, as it is based on information available *after* a measurement is made.

The solution to the inverse problem that results from finding the z that maximizes $p(z|u)$ is known as the *Maximum A Posteriori (MAP)* solution. *MAP* estimation minimizes the probability of error.

Maximum Likelihood

Maximum likelihood methods estimate the *true* solution as follows:

The solution to the inverse problem that results from finding the z that maximizes $p(u|z)$ is known as the *Maximum Likelihood (MLE)* solution.

Bayesian vs. ML

- Bayesians often argue that you really need to look at the entire likelihood function (and, of course, the prior) to get solution. Otherwise, you can get stuck in a local maximum.
- Bayesians like to look at the entire posterior distribution as output so you can make sense of it, and then decide what to do.

Bayesian Regularization

- In many cases in computer vision, we will see that we cannot feasibly estimate the entire posterior distribution.
- In these cases, we can use optimization techniques to try to locate the MAP solution.

Decision Theory

- Recall that the goal of inference is, given a defined hypothesis space and a particular set of data observations, assign probabilities to the hypotheses.
- Decision theory typically chooses between alternative actions on the basis of these probabilities, typically to minimize the expectation of a *loss function*.

Relationship Between the Tikhonov and Bayesian Approaches to Regularization

The Tikhonov approach to regularization involves minimizing a functional:

$$M^{\alpha}(z,u) = \rho_U(Az,u) + \alpha\Omega[z]$$

whereas the Bayesian approach often involves maximizing a conditional probability:

$$p(z\,|\,u) = \frac{p(u\,|\,z)\,p(z)}{p(u)}$$

Tikhonov and Bayesian Regularization

These two approaches are equivalent if $p(z|u)$ is in the form of a ***Gibbs*** distribution:

$$p(z \mid u) = \frac{\exp(-\beta E(z,u))}{Z}$$

Where Z is a normalization term, often referred to in statistical physics literature as the *partition function*.

Tikhonov and Bayesian Regularization

In a physical interpretation $E(z,u)$ is the free energy of a system corresponding to a configuration z.

$P(z|u)$ is the probability of the system configuration z, given certain measurements u.

The most likely system configuration is the one that minimizes the free energy of the system.

Tikhonov and Bayesian Regularization

We can associate the regularization functional $M^{\alpha}(z,u)$ with the energy functional $E(z,u)$ of the Gibbs distribution.

Thus we can write:

$$p(z\,|\,u) = \frac{\exp(-\beta M^{\alpha}(z,u))}{Z}$$

$$p(z\,|\,u) = \frac{\exp(-\beta \rho_U(Az,u))\exp(-\beta\alpha\Omega[z])}{Z}$$

Tikhonov and Bayesian Regularization

We can therefore make the following associations:

$$P(u \mid z) = \frac{\exp(-\beta\rho_U(Az, u))}{Z_1}$$

$$P(z) = \frac{\exp(-\beta\alpha\Omega[z])}{Z_2}$$

$$P(u) = \frac{Z}{Z_1 Z_2}$$

Tikhonov and Bayesian Regularization

It can been seen that $P(u \mid z)$ involves the forward operator A and the data consistency metric ρ_U. Thus it should be specified in a way which reflects our knowledge of the image formation process.

Likewise, $P(z)$ involves the regularization stabilizing function $\Omega[z]$, and should therefore reflect our *a priori* knowledge of the solution.

Tikhonov and Bayesian Regularization

Note that minimization of the regularization function corresponds to maximizing the Gibbs distribution.

Thus we can solve the regularized version of the ill-posed problem by finding the solution that maximizes the Gibbs distribution.

Tikhonov and Bayesian Regularization

The Bayesian approach involves explicit modeling of the prior information relevant to the particular inference task at hand.

In the Tikhonov regularization, the specification of the stabilizing functional is often done for mathematical convenience, and therefore may not be reflective of the actual experimental conditions.