Lecture 1: Introductory Lecture

Anand Mishra

Center for Visual Information Technology
IIIT Hyderabad
http://researchweb.iiit.ac.in/∼anand.mishra/

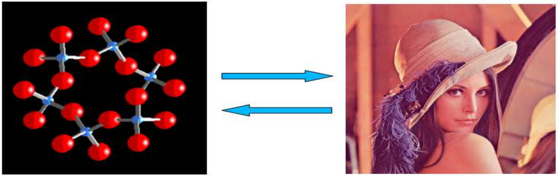
March 2017

Objective of today's Tutorial

At the end of this tutorial I hope you will become ⌣̈

- motivated enough to think in direction of optimization in vision
- Familiar with some basic terminologies going to be frequently used in subsequent tutorials.
- Contents:
  - Motivation
  - The Labelling Problem
  - Optimization in vision
  - MAP formulation

One to One correspondence between physical systems and Images



One to One correspondence between physical systems and Images

Physical System	Images
Molecules, Atoms	Sites (Pixels, edge, image regions)
Higher Temperature	Lower Prior Probability
Potential energy of system (interaction between atoms/molecules)	Clique potential (Interaction between sites)

Energy of Image

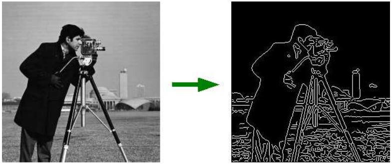
So as least energy physical system is the most stable one, we can define an objective function (due to analogy we refer it as energy function) for given image and problem such that minimum of which is of our interest(or solution of the problem)

From Heuristic image Analysis to Probabilistic image analysis

- How heuristic image analysis works?
- Let us take a popular image analysis example of edge detection.
- How the traditional image analysis address this problem?

From Heuristic image Analysis to Probabilistic image analysis

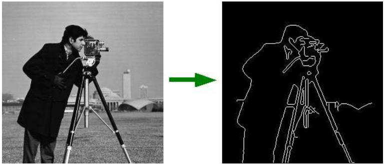
Let us take an example popular cameraman image.We have to find out edges in this image.
Apply canny operator with some threshold:



No these don't looks fine. Sad. Let us change the threshold.

From Heuristic image Analysis to Probabilistic image analysis

Apply canny operator with some other threshold:



Great this looks better. Happy.

From Heuristic image Analysis to Probabilistic image analysis

- This is how we deal with any image analysis problem typically.
- Manual tuning of parameters is the worst part of it.
- No way I can add constraint so that I get only those edges which belong to person
- Question: Can't we think some alternative?

From Heuristic image Analysis to Probabilistic image analysis

- Yes! Probabilistic Image Analysis is one such alternative.
- Probabilistic Image Analysis models probabilistic dependency of neighbouring pixels/patches.
- Thus we can use prior information about the problem.
- Gives flexibility of using training images to automatically learn the parameters.

From Heuristic image Analysis to Probabilistic image analysis

- We should move from heuristic image analysis to probabilistic image analysis.
- How to model contextual constraints in image.

Why MRF?

Markov Random Field is a branch of probability theory provides a foundation for contextual constraints and derivation of the probability distribution of interacting features.

MRF applications

- Low level vision problems
  - edge detection,
  - binarization
  - surface reconstruction
  - image restoration,
  - segmentation etc
- High level vision problems
  - Object matching
  - Object recognition etc

The Labelling Problem

- Many Image analysis and vision problem can be posed as labelling problem
- Example: Segmentation, Edge detection, Stereo, Object detection
- Labelling problem is specified in terms of Sites and Labels.

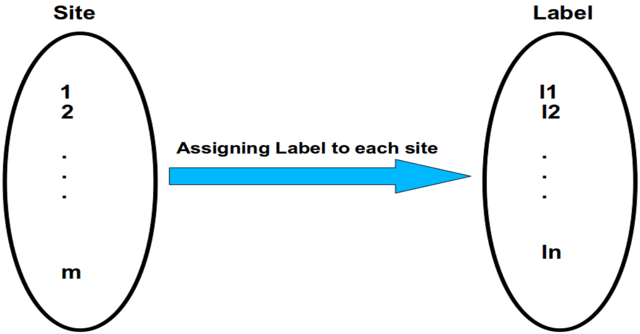
What are sites in Image?

- A site often represents a point or a region in Euclidean space.
- Could be an image pixel, image patch, a line segment or a corner
- Can be categorized into regular and irregular sites
- We represent site by a finite set $S = \{1, 2, ..., m\}$ where each element represent index of the site.

And what about labels?

- A label is an event that may happen to site.
- Labels depend on the problem. (We will discuss it soon)
- Let us denote Set of labels as $L = \{l_1, l_2, ..., l_n\}$

The Labelling Problem

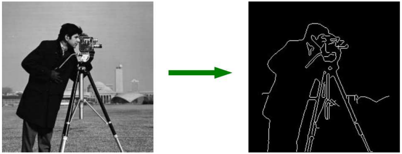


Vision problems as labelling problem

- As we discussed many vision/image analysis problems can be formulated as labelling problem.
- How?
- Let us have a close up look on it.

Edge detection

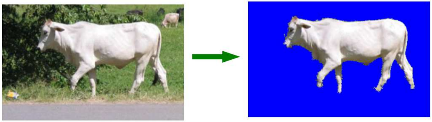
Problem: Given an Image find out edge/non-edge pixels



- Sites: pixels
- Labels: $\{$edge(1), non-edge(0)$\}$

Binary Segmentation

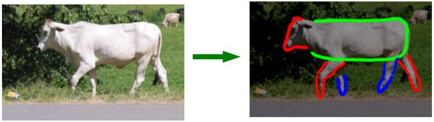
Problem: Given an Image find foreground/background



- Sites: pixels
- Labels: $\{$foreground, background $\}$

Object parts detection

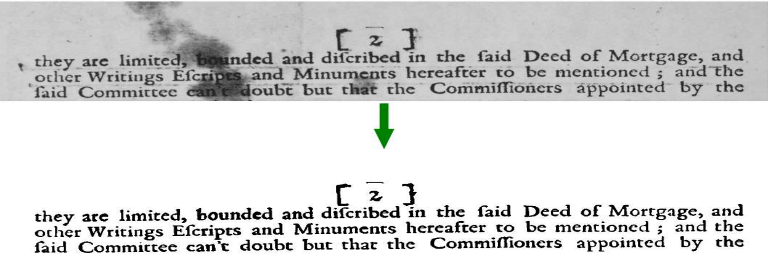
Problem: Given an Image find out the parts of it



- Sites: pixels
- Labels: {'Head', 'Torso', 'Legs'}

Document Image binarization

Problem: Given a Gray document image find out the binary image.


[z]
they are limited, bounded and difcribed in the faid Deed of Mortgage, and other Writings Efcripts and Minuments hereafter to be mentioned ; and the faid Committee can't doubt but that the Commiffioners appointed by the

- Sites: pixels
- Labels: $\{0(\text{text}), 1(\text{non-text})\}$

Word recognition

Problem: Given a word recognize it

- Sites: character windows
- Labels: {English Alphabets}

Combinatorial size of configurations

- Each mapping from Site to Label is known as configuration
- Let F be the set of all possible configurations
- What will be the cardinality of F?

  $$n \times n \times \cdots n \times n\,(m\,times) = n^m$$

  Note that cardinality of site and labels are m and n respectively.
- Take an example of 256×256 image and assume we are dealing foreground - background segmentation problem as labelling problem.
- The cardinality of the configuration is $2^{256 \times 256}$ which is huge.
- So any brute force algorithm to search optimal configuration will take exponential time.

Combinatorial size of configurations

Question: Are we interested in all these n^m configuration?

- the answer is No.
- In fact only a small number of them are good solutions and are of our interest.
- But which ones?
- This leads to questions like which solutions are the optimal ones, what is the optimality criteria?

Why optimization in vision

- Existence of uncertainties in every vision process
- What are the source of uncertainties?
  1. Noise
  2. Degradation
  3. Appearance
  4. Pose
  5. Visual interpretation etc

Three steps in optimization based vision

- Problem Representation
- Formulating Objective function
- Applying optimization algorithm

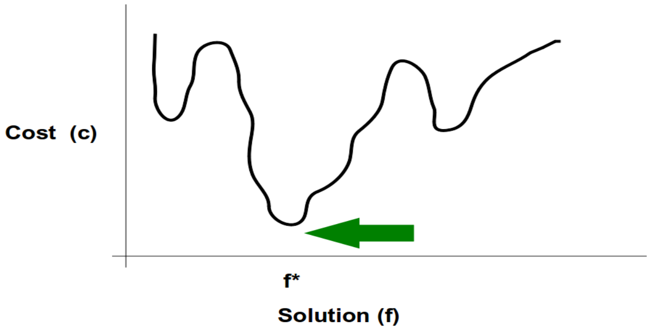
Problem representation

- Selecting sites and labels for the problem
- Consider problem of binary segmentation: what will be our choice for sites? And What about labels?

Formulating objective function

- is a function between solution space and the goodness of solution
- Mostly non-convex nature of objective function
- How to encode image features, pixel intensities in the function?

Typical nature of objective function



Applying optimization algorithm

- Many optimization algorithms exist to solve the problem
- Like: Graph cut, Belief propagation, ICM, Simulated Annealing, Highest confidence first, Convex relaxation etc
- We will cover some of them in detail in later lectures.

Back to Objective/energy function again

- A configurations f can be thought of as random vector taking values from the set of labels.
- Given observation d we have to infer random variable f.
- Clearly In order to find best solution we have to maximize probability term $P(f|d)$

$$P(f|d) = \frac{P(f)p(d|f)}{p(d)}$$

Back to Objective/energy function again

- As we discussed in the start of the lecture higher probable configurations are of our interest.
- we can always define an energy functions which encodes the posterior probability in such a way that higher posterior probability corresponds to lesser energy.

Optimality criteria

- **Maximum Likelihood:** Likelihood is available but prior is not.

$$f^* = \arg\max_f p(d|f)$$

- **Maximum Entropy:** Only prior is available.

$$f^* = \arg\max_f \{-\sum_{i=1}^{m} P(f_i) \log P(f_i)\}$$

- **Maximum A Posteriori:** When both prior and likelihood are known: Bayesian estimation

MAP Estimation

- Given an observation d, posterior probability of configuration f is:

$$P(f|d) = \frac{P(f)p(d|f)}{p(d)} \tag{1}$$

- Risk of estimating solution f^*:

$$\int_{f \in F} C(f, f^*)P(f|d)df \tag{2}$$

MAP Estimation(Contd.)

Cost functions

$$C(f, f^*) = ||f^* - f||^2 \tag{3}$$

$$C(f, f^*) = \begin{cases} 0 & |f^* - f| \leq \delta \\ 1 & \text{otherwise} \end{cases} \tag{4}$$

MAP Estimation(Contd.)

We can write equation (2) as:

$$R(f^*) = \int\limits_{|f^* - f| \leq \delta} C(f, f^*) P(f|d) df + \int\limits_{|f^* - f| > \delta} C(f, f^*) P(f|d) df \quad (5)$$

For simplicity let us assume we use cost function (4) then

$$R(f^*) = 1 - \int\limits_{|f^* - f| < \delta} P(f|d) df \quad (6)$$

MAP Estimation(Contd.)

If we take $\delta \to 0$ then risk can be rewritten as:

$$R(f^*) = 1 - kP(f|d) \tag{7}$$

Where k is volume of space containing all points for which $|f^* - f| < \delta$ We have to minimize this risk i.e. we have to maximize $P(f|d)$

MAP Estimation(Contd.)

Which leads to conclusion that the optimum solution f^* is one which maximizes posterior $P(f|d)$ or from equation 1 we can write optimal configuration is:

$$f^* = \arg\max_f p(d|f)P(f)$$

- Since images are not complete random. Thus pixels are mutually dependent.
- **Question:** How to model such dependency?
- **Answer:** We can treat Image as MRF and model such contextual constraints.

References

- Stan Z. Li, Markov Random Field Modeling in Image Analysis , Spriger 3rd Ed. , chapter -1, 2009
- S. Geman and D. Geman, ”Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” IEEE Trans. Pattern Anal. Mach. Intell, 6, 721-741, 1984 (only for motivation part)