Markov Random Fields in Image Analysis

Vedrana Dahl, Anders Dahl
DTU Compute

02506 Advanced Image Analysis

March 2020

DTU Compute
Department of Applied Mathematics and Computer Science

02506 Part II: Image analysis (segmentation) with geometric priors

- Week 5 (today): Markov random fields. Practical exercise: MRF modelling (Exercise 5.2) and bone segmentation (Exercise 5.5).
- Week 6: Mumford-Shah functional, Chan-Vese algorithm and snakes. Practical exercise: tracking deformation in a sequence of images.
- Week 7: Layered surface detection. Exercise (tentative): detecting layers in bone data. Combining layered surface detection with deformable models.

Markov Random Fields

What to learn? (Exam question related to this exercise)

- How are MRF used for image segmentation?
- In context of image segmentation using MRF, explain the concepts of likelihood and prior. Which likelihood and prior did you use in the exercise?
- Explain the difference between modelling and optimization when working with MRF. Which optimization did you use in the exercise?
- Explain how graph cuts are used for MRF optimization.

Reading material

- Chapter 5 in course note.
- Additional literature on MRF: S. Z. Li (2009) Markov Random Field Modeling in Image Analysis.
- Lecture and note are supplemented with a few findings from the articles of Boykov-Kolmogorov-Zabith clique – it is not expected that students read these articles.

Relevant problems

Labeling in image analysis (Li 1.1.3)

- Image restauration (denoising): assign a true intensity to each pixel of a noisy image.
- Image segmentation: assign a segment label (e.g. white matter, gray matter, bone, background) to each pixel.
- Stereo: assign a depth (in respect to one camera) to each pixel of a stereo pair.
- Edge detection: assign a label (edge or not edge) to each pixel or to each dual site (a line between the pixels).

Contextual constrains for labeling (Li 1.1.4)

- Without contextual constraints:
  - Labels are independent.
  - Joint probability is the product of the local ones.
  - Optimal global labeling can be computed by considering each label independently.
- With contextual constraints:
  - Labels are mutually dependent.
  - Making global inference using local information is not trivial. Why?
  - Next 4 slides show how Markov Random Fields framework provides a mathematical theory for solving this problem (remember: provides a link between the local appearance and global probability).
  - Context (local appearance) is in MRF defined in terms of neighbors and cliques.

MRF as a solution to labeling problem

Advantages

- A general flexible framework which provides a link between the local appearance and global probability
- Excellent for modeling useful priors: smoothness, piecewise-smoothness. . .
- In certain cases can be efficiently solved using graph cuts

Challenges (also learning challenges)

- In certain cases difficult to solve
- Terminology may be confusing
- Due to extensive use of graph cats for optimizing MRF, many forget that MRF is much more than graph cuts (i.e. distinction between modeling and optimization disappears)

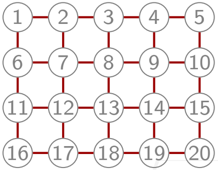
Ultra condensed (and slightly imprecise) MRF theory

Markov Random Fields

- The labeling problem (Li 1.1.1–1.1.2)
  - Sites $\mathcal{S} = \{1, \ldots, m\}$
  - Labeling $f : \mathcal{S} \to \mathcal{L}$
  - Configuration $f = \{f_1, \ldots, f_m\}$
- Neighborhood system (Li 2.1.1)
  - Neighborhood system $\{\mathcal{N}_i \mid \forall i \in \mathcal{S}\}$
  - Property: $i \in \mathcal{N}_j \Rightarrow j \in \mathcal{N}_i$
  - Clique: subsets of sites that are all neighbours to one another
  - Set of all cliques $\mathcal{C}$, one-cliques $\mathcal{C}_1$, two-cliques $\mathcal{C}_2 \ldots$
- Markov random fields (Li 2.1.2)
  - Configuration is a realization of a random process
  - Markovianity: $P(f_i | f_{\mathcal{S} \setminus i}) = P(f_i | f_{\mathcal{N}_i})$

Note that this does not give a probability of a configuration!!!

Example: 4-neighborhood on an image



- Markov-Gibbs equivalence (2.1.4)
  - MRF is GRF with respect to the same neighbourhood
- Gibbs random field (Li 2.1.3)
  - $P(f) \propto e^{-U(f)}$
  - $U(f) = \sum_{c \in \mathcal{C}} U_c(f)$

So: we need to define clique potentials

Example: segmenting a noisy image (Li 3.2.2)

- Data: a noisy image
- Modelling: defining an objective function (i.e. clique potentials) for computing joint probability of any segmentation (image with discrete labels). Therefore we need to define potentials such that the desired segmentation has a low energy.
- Choose a neghborhood system (in exercise: one-cliques and two-cliques) and define clique potentials.
- One-clique potentials, e.g. $V_1(f_i = \text{label 2})$, suggestion: how does the intensity of pixel i fit in the segment 2.
- Two-clique potentials, e.g. $V_2(f_i = \text{label 1}, f_{i'} = \text{label 2})$, suggestion: something that penalizes the changes in labels.
- Total energy is ...

- Energy is the sum of clique potentials, for one and two clique

$$E(f) = \sum_{i \in \mathcal{C}_1} V_1(f_i) + \sum_{\{i,i'\} \in \mathcal{C}_2} V_2(f_i, f_{i'})$$

- Segmenting a noisy image may be defined as

$$E(f) = \alpha \sum_{i \in \mathcal{C}_1} (\mu(f_i) - d_i)^2 + \beta \sum_{\{i,i'\} \in \mathcal{C}_2} (1 - \delta(f_i - f_i'))$$

where α and β give a trade-off between the data and the context-free information and contextual information.

Modelling choices

Prior

- Our preferable prior: smoothness prior for discrete labels
  - Definition

$$V_2(f_i, f_{i'}) = (1 - \delta(f_i - f_{i'}))$$

$$= \begin{cases} 0 & \text{if } f_i = f_{i'} \\ 1 & \text{otherwise} \end{cases}$$

  - Penalizing neighbouring labels being different

$$\sum_{\{i,i'\} \in \mathcal{C}_2} (1 - \delta(f_i - f_{i'}))$$

$$= \#\{f_i \neq f_{i'} | \{i, i'\} \in \mathcal{C}_2\}$$

- Other prior potentials (also for continuous labels): quadratic, truncated quadratic (modeling discontuities), see Li 3.2.1

Likelihood

- Our preferable likelihood: squared distance likelihood
  - Definition

$$V_1(f_i) = \alpha(d_i - \mu(f_i))^2$$

- Other likelihood potentials: allow for classes with different standard deviation, proportions..., truncated versions.

$$V_1(f_i) = \frac{1}{\sigma(f_i)\sqrt{2\pi}} + \frac{1}{2}\left(\frac{d_i - \mu(f_i)}{\sigma(f_i)}\right)^2$$

MAP-MRF framework (Li 1.3)

- MRF are related to Bayesian framework, so it is usefull to get acquainted with Bayesian terminology.
- Finding most probable MRF labeling corresponds to finding maximal a posterior (MAP) solution.

Bayes rule

$$\underset{\text{posterior}}{P(f|d)} = \frac{\overset{\text{likelihood}}{P(d|f)} \cdot \overset{\text{prior}}{P(f)}}{\underset{\text{evidence}}{\cancel{P(d)}}}$$

- f is estimate, d is observation
- $P(d|f)$ – likelihood probability
- $P(f)$ – prior
- $P(f|d)$ – posterior probability
- $P(d)$ – evidence (constant and not used)

- Bayes: $P(f|d) \propto P(d|f)P(f)$
- Gibbs: $P(f) \propto e^{-U(f)}$
- This leads to

$$U(f|d) = U(d|f) + U(f)$$

  - $U(d|f)$ – likelihood energy, often in terms of one-clique potentials
  - $U(f)$ – prior energy, often in terms of higher-order clique potentials
  - $U(f|d)$ – posterior energy, sum of all clique potentials

Exercise notation as in Li 3.2.2

$$U(f|d) = U(d|f) + U(f)$$

$$E(f) = \sum V_1 + \sum V_2$$

Modelling summary

Summary of the MAP-MRF modelling approach (Li 1.3.4)

- Pose a labeling problem (with regular or irregular sites, discrete or continuous labels).
- Define a neighborhood structure and a set of cliques.
- Define prior clique potentials (2-cliques).
- Define likelihood energy (1-cliques).
- Add prior and likelihood to yield posterior.
- Find configuration with highest posterior.

How to come up with the model?

- Define local potentials which penalize improbable configurations:
  - Prior (has to do with how the desired solution should look like)
  - Likelihood (has to do with how the observations are made - data creation model)

Validation of modeling (Li 1.4). Is model formulation good?

- Does the energy minimum of objective function correspond to the correct solution?
- If you (somehow) got hold of the desired configuration, would its probability be higher than the probability of the undesired configuration?

Optimization options

Optimization – standard metods (Li 3.2.2, 9.3.1, 7.1.6)

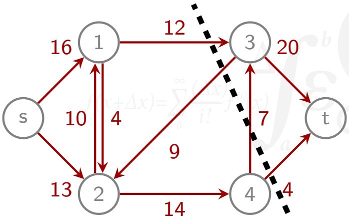
- Iterated conditional modes
  - Initialize (how: randomly or maximum likelihood?)
  - Consider one site at a time and assign an optimal label (Serial or parallel? In which order?)
- Gibbs sampler
  - Initialize using maximum likelihood.
  - For each site compute local probability of each label (requires normalizing probabilities to sum to 1).
  - Replace label with given probability.

MRF and graph cuts (Li 10.4)

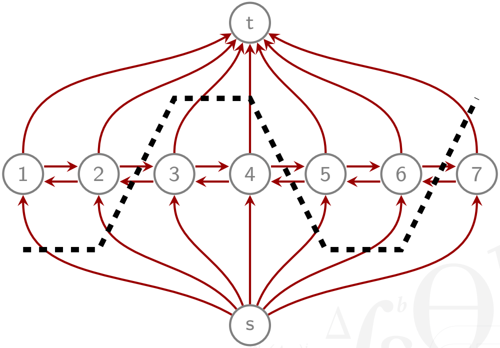
- Binary labels
  - If MRF energy is submodular (I will not formally define this now, roughly equivalent to 'encourage smooth solution') then it can be exactly solved using graph cuts.
- Multiple discrete labels
  - Approximation of the global solution can be found efficiently using graph cuts – iteratively solving multiple binary graph cuts.

Graph cuts (Li 10.4.1)

- s-t graph
  - Directed weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
  - Vertex set with elements i, edge set with elements i, j, positive weights w_{ij}
  - Source s and sink t
- Max flow – min cut
  - What it the maximal flow from s to t? Max-flow.
  - What is the minimal cut dividing s and t Min-cut or s-t cut.
  - Ford-Fulkerson theorem: those problems are equivalent.



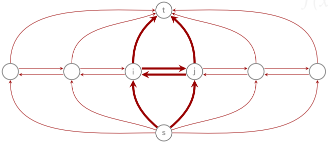
Graph cuts and MRF for image segmentation (Li 10.4.2)



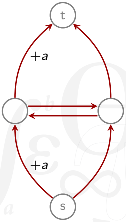
- A s-t graph constructed from to the MRF likelihood function and prior
  - Place one-clique potentials (data term per pixel) for the two labels $U(f_i = 1)$ and $U(f_i = 2)$ on the terminal edges.
  - Place two-clique potentials β on the internal edges in both directions.
- A s-t cut corresponds to MRF energy of the configuration: min-cut corresponds to the optimal configuration.

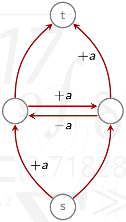
Equivalences (Li 10.4.2)

- Constructing graph from the MRF energy is not unique
- Questions (answerd by Boykov and Zabith in 2001, Kolmogorov and Zabith in 2004, and Boykov and Kolmogorov in 2004):
  - Can we simplify the graph representation (many 0-weight edges)? Equivalences.
  - Which (more general) energies can be represented? Submodularity.
  - How to expand to multiple (discrete) labels? α expansion.
- Consider two sites...



- Simple graph construction
  - Place one-clique potentials on the terminal edges.
  - Place two-clique potentials on the internal edges.
- Equivalences may be used to make this graph more sparse





- Transformation will not change the minimal cut (it will change its value for a).

Multiple labels: α expansion (Li 10.4.3)

Idea

- Chose a label, call it α
- Segment into α and $\bar{\alpha}$
- Chose another label
- Repeat until convergence

Issue: smoothing penalty needs to be adjusted for configuration $(\bar{\alpha}, \bar{\alpha})$

- Auxiliary node (Boykov, Veksler and Zabih 2001)
- Without auxiliary node, just adjusting weights (Kolmogorov and Zabih 2004)

Overview over exercises in MRF from course note. With priorities!

5.1 Example: Gender determination (easy 1D example, no actual work)

5.2 Exercise: MRF Modelling (validating the model, but not finding the result – do this such that you can visualize your results and reflect on the quality of the results)

5.3 Exercise: Iterative optimization for MRF (finding the result using standard iterative methods – optional advanced exercise for gaining understanging on how iterative optimization works)

5.4 Example: Graph cuts for MRF (example, trying out provided code)

5.5 Exercise: Binary segmentation using MRF and graph cuts (binary segmentation of bone image – actual use of MRF for bianry image segmentation)

5.6 Exercise: Multilabel segmentation (try α expansion on a synthetic and bone image – optional advanced exercise for gaining understanding on how α expansion works)