Light Microscopy in 4 Dimensions: Modeling of Spatial Information for Image Segmentation

Jussi Tohka
`jussi.tohka@tut.fi, http::/www.cs.tut.fi/˜jupeto`

Institute of Signal Processing
Tampere University of Technology

About the lectures

Goal: Give an overview of basic techniques used in automatic processing and segmentation of 3-D biomedical images

- Only general principles are presented, references are offered for technical details.
- No particular problem domain (e.g. light microscopy) is considered, techniques presented will be quite general.
- Most techniques presented can be applied to process also plane images, but the presentation and symbolics will refer to three dimensions.

Overview

- **Basics**: Some 3D imaging modalities. The definition of image, differences between natural and biological images, differences between plane and 3-D images.
- **Basic image processing**: Image filtering and smoothing in three dimensions. Edge detection in three dimensions.
- **Segmentation**: The definition of the image segmentation. Image models and pattern classification.
- **Typical imaging artifacts**: Intensity non-uniformity. Partial volume effect and blurring.
- **Spatial information for image segmentation**: Markov Random Fields, Deformable models, Gradient flows.

Notation

- Because indexing voxels with three indices is a pain, I will often use only a single index
- The image domain $\mathcal{D} = \{\mathcal{V}_1, \ldots, \mathcal{V}_K\}$. The set of voxel indices $\{1, \ldots, K\}$ is denoted by S. Each voxel $\mathcal{V}_i$ has an intensity value a_i. The image intensities (the data) is $\mathbf{a} = [a_1, \ldots, a_K]$. The random variable (RV) relating to the intensity of the voxel i is A_i.
- When segmenting images, we assign a label for each voxel based on the intensity values. The label says that this voxel is part of some structure or background. The set of labels includes whatever we want to extract from images and background.
- Integers $1, \ldots, L$ denote the labels. The segmented image is then $\mathbf{b} = [b_1, \ldots, b_K]$, where b_i - the label of the voxel $\mathcal{V}_i$ - is some integer from 1 to L.

Limitations of image thresholding

- A major disadvantage of the segmentation techniques based on global image thresholding is that it does not take into account the spatial nature of the data.
- For example, if we know that all the neighbors of a voxel belong to some structure, it is highly probable that also the voxel is part of this same structure; independent on the intensity value at that voxel.
- In other words, the (global) thresholding can be viewed as basic pattern classification, and as such it assumes that the voxel labels are independent from each other. But, this is rarely true!

Modeling spatial image content

- There are many techniques capable of incorporating spatial modeling information into image segmentation.
- Two of them, Markov Random Fields (MRFs) and deformable (surface) models are considered here.
- The essential difference between the two is that MRFs are used more conveniently for the volume based segmentation while deformable surface models are, as the name suggests, a tool for the boundary (edge) based segmentation.

MARKOV RANDOM FIELDS

Markov Random Fields

- MRFs are a statistically based models for spatial interaction. They can be used to describe the probability of a segmentation through its local characteristics.
- That is, they model the probability $P(\mathbf{b}) = P(b_1, \ldots, b_K) \neq \prod_i P(b_i)$.
- There is a big difference between placing a prior probability over the whole segmented image and placing a prior probability over each voxel label individually.
- If we have just two structures of interest and a 64×64 image, there are $2^{64^2} \approx 10^{1000}$ possible segmentations.

MRFs: Neighborhood systems

- In the MRF terminology, we refer to voxels as sites. The sites are related to one another via a neighborhood system.
- Write S for the set of sites or voxels. A neighborhood system on S is then defined as

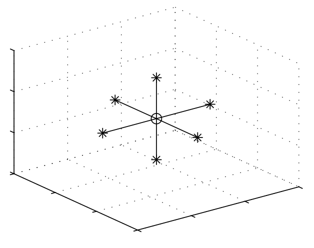
$$N = \{N_i | i \in S\}$$

where N_i is the set of neighbors of the site i. Two conditions are required for N to be a neighborhood system
  - $i \notin N_i$
  - $i \in N_j \Leftrightarrow j \in N_i$

Neighborhood systems:An example

- We switch from single index to triple indices for clarity.
- A simple neighborhood system is

$$\begin{aligned} N_{ijk} &= \{(i-1,j,k),(i+1,j,k),(i,j-1,k), \\ &\quad (i,j+1,k),(i,j,k-1),(i,j,k+1)\}. \end{aligned}$$



Definition of MRFs

- Denote by

$$\Omega = \{\mathbf{b} = (b_1, \ldots, b_K) | b_i = 1, \ldots, L\}$$

the set of all possible configurations, i.e. voxel labelings. Each voxel label b_i is then one of the possible voxel labels $1, \ldots, L$.
- Let B_i denote the random variable related to labeling of the voxel i and $B = (B_1, \ldots, B_K)$. We are interested in $P(B = \mathbf{b})$ - the probability that $B = \mathbf{b}$.
- B is then an MRF with respect to N if
  - $P(B = \mathbf{b}) > 0$ for all $\mathbf{b} \in \Omega$
  - $P(B_i = b_i | B_j = b_j, j \neq i) = P(B_i = b_i | B_j = b_j, j \in N_i)$

The Gibbs connection

- A nasty thing in the formulation above is that P must be a probability measure on Ω.
- In other words, if we are given a function depending on a voxel label as well as labels of the neighboring voxels, how do we know that this function is $P(B_i = b_i | B_j = b_j, j \in N_i)$ for some P.
- There is a nice result dubbed as the Hammersley-Clifford theorem which says that Gibbs distributions and MRFs are equivalent.

Gibbs distributions

- The subset C of S is a clique if every distinct pair of sites in C are neighbors.
- The distribution $P(\mathbf{b})$ is Gibbsian with respect to N if it can be written as

$$P(\mathbf{b}) = \frac{1}{Z} \exp(-\beta \sum_{C} V_C(b_i : i \in C)),$$

  where Cs denote cliques, V_C are clique potentials, $\beta > 0$ is a parameter and Z is a normalizing constant.
- Clique potentials may be arbitrary.

Hammrsley-Clifford theorem

- **Hammersley-Clifford theorem**: $P(B = \mathbf{b})$ is Gibbsian with respect to N if and only if B is an MRF with respect to N.
- A simple proof can be found in Besag: Spatial interaction and statistical analysis of lattice systems, Journal of Royal Statistical Society, vol 36 , no 2, 1974.
- The theorem is named after Hammersley and Clifford because they were first to conceive it (around 1970), but they did not publish it. The reason was that they did not like the positivity condition of MRFs. The positivity condition is, as Moussuris later showed, necessary for the Gibbs-MRF connection.

Ising Model

- A simple MRF for ferromagnetism (Ising 1925).
- Suppose that we want to segment a given volumetric image into two distinct regions and some spatial dependence is assumed.
- We label voxels by triple indices and select a simple neighborhood system, where neighbors of the voxel ijk are

$$\begin{aligned} N_{ijk} &= \{(i-1,j,k),(i+1,j,k),(i,j-1,k),\\ &\quad (i,j+1,k),(i,j,k-1),(i,j,k+1)\}. \end{aligned}$$

Ising model

- The cliques are now i) sets of single voxels and ii) sets of two voxels where one voxel index differs by one and the others are the same. There are no other cliques.
- For example, the voxel $(5,5,5)$ belongs to the cliques $\{(5,5,5)\}, \{(5,5,5),(5,5,6)\}, \{(5,5,5),(5,5,4)\},$ $\{(5,5,5),(5,6,5)\}, \{(5,5,5),(5,4,5)\},$ $\{(5,5,5),(6,5,5)\}, \{(5,5,5),(4,5,5)\}.$
- In boundaries of the image domain, voxels do not have all the neighbors. This can be corrected easily assuming that the voxel labels outside the image domain are known.

Ising model

- Let the two labels be 1 and -1.
- We select $V_C(b_i) = 0$ for singleton cliques - both labels are equally probable.
- For doubleton cliques, we select $V_C(b_i, b_j) = b_i b_j$.
- The Gibbs distribution is

$$\begin{aligned} P(\mathbf{b}) &= \frac{1}{Z} \exp(\beta \sum_i \sum_j \sum_k \\ & b_{i,j,k} b_{i,j,k+1} + b_{i,j,k} b_{i,j+1,k} + b_{i,j,k} b_{i+1,j,k}). \end{aligned}$$

- And $P(b_{i,j,k} | b_l \in N_{i,j,k}) = \frac{\exp(\beta \sum_{l \in N_{i,j,k}} b_l b_{i,j,k})}{1+\exp(\beta \sum_{l \in N_{i,j,k}} b_l)}$

MRFs and images

- The image data must be taken somehow into account when segmenting images.
- This is conceptually straight-forward, but computationally very challenging.
- We need to find the labeling $\mathbf{b}^*$ that maximizes

$$P(\mathbf{b}|\mathbf{a}) \propto P(\mathbf{b}) p(\mathbf{a}|\mathbf{b}) = P(\mathbf{b}) \prod_i p(a_i|b_i),$$

  where $p(a_i|b_i)$ denotes the likelihood of observing intensity a_i if the voxel label is b_i.
- In other words, we want to find *the most probable labeling given the intensity values*.

MRFs and images

- Computational difficulties follow from the prior term $P(\mathbf{b})$.
- We cannot compute values of it for all possible segmentations and
- it does not factor in a straight-forward manner as the likelihood term does.

ICM algorithm

- Julian Besag (J Roy Stat Soc 1986) suggested a straight-forward method for maximizing (locally) the probability of the segmentation.
- Denote the current pixel label by b_i^t. The algorithm is as follows:
  1. Set $b_i^1 = \arg\max_b p(a_i|b)$ for all i set $t = 1$.
  2. Set $t = t + 1$
  3. Set $b_i^t = \arg\max_b P(b|b_j^{t-1}, j \in N_i) p(a_i|b)$
  4. If some $b_i^t \neq b_i^{t-1}$ go step 2, otherwise finish.

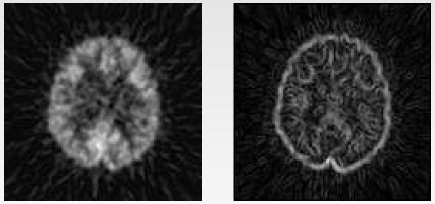
References

- Besag: Spatial interaction and statistical analysis of lattice systems, Journal of Royal Statistical Society, vol 36 , no 2, 1974.
- Geman and Geman: Stochastic relaxation, Gibbs distributions , and Bayesian restoration, IEEE-PAMI, vol 6 no 6, 1984.
- Besag: On statistical analysis of dirty pictures. Journal of Royal Statistical Society, vol 48 no 3, 1986. (approx 750 citations in ISI)
- Li: Markov Random Fields in Image Analysis, Springer-Verlag, 2001.

DEFORMABLE SURFACE MODELS

Boundary based segmentation

- Sometimes it is not possible or it is too complicated to model voxel intensities statistically.
- Then, it can be assumed - with a variable level of success - that the intensity values have jumps at the structure boundaries.
- That is we are detecting edges, but by including knowledge of the spatial configuration.



Deformable models: Idea

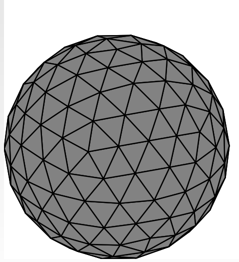
- The idea of deformable surface models is to start from a template surface and deform it according to image data.
- Physics based interpretation: An elastic surface is placed into the potential field and it is then allowed to deform until it stabilizes.
- Probabilistic interpretation: Find the surface that has the maximal probability given the image data.
- Two somewhat related issues need to be considered:
  - What kind is the template should we have?
  - How to implement the deformation?

Surface representation

- Consider first the selection of template.
- That is, we need to be able to represent the surfaces in a manner that supports computations.
- In addition, the representation should cover rather many kinds of surfaces: For example, if we consider just spheres, we can extract only spheres. This would be of limited use when segmenting biological images.
- Still many choices exist.

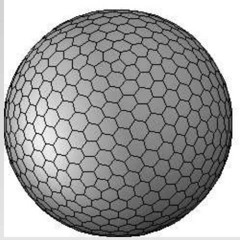
Triangle meshes

- One possibility is to represent surfaces by meshes.
- Triangle meshes consist of triangles glued together along their edges. Triangles are defined by their edges, and edges are defined by their endpoints (vertices).
- Deformation of a triangle mesh is implemented by changing the positions of the vertex coordinates, or mexels as we call them.



Simplex meshes

Simplex meshes are formed by the set of mexels $\mathbf{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_N\}$ and the adjacency relations between mexels. The adjacency relations are modeled by a graph. The nodes of the graph are $1, \ldots, N$ and if the graph has the edge ij then mexels i and j are neighbors. Each mexel has three neighbors in a simplex mesh and the neighbors of the mexel $\mathbf{w}_i$ are denoted by $\mathbf{w}_{i_j}$.



Deformation

- The problem is now following: Given a template surface mesh move its mexels so that the surface mesh corresponds to the image data.
- Two alternatives: Force-based meshes (related to the physics based interpretation) and energy based meshes (related to the probabilistic interpretation).

Find $\mathbf{W}$ such that $F_{ext}(\mathbf{W}) + F_{int}(\mathbf{W}) = 0$	Find $\mathbf{W}$ such that $E(\mathbf{W}) < E(\mathbf{V})$ for all $\mathbf{V}$.

Energy-based approach

- The problem: minimize the energy $E(\mathbf{W})$.

$$E(\mathbf{W}) = \underset{\substack{\uparrow \\ \text{surface shape}}}{E_{internal}(\mathbf{W})} + \underset{\substack{\uparrow \\ \text{image}}}{E_{external}(\mathbf{W})}$$

- The minimization problem is challenging due to many local minima and large number of variables.
- For a simplex mesh $\mathbf{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_N\}$ the energy can be

$$E(\mathbf{W}) = \sum_{i=1}^{N} \left(\frac{||\mathbf{w}_i - \frac{1}{3}\sum_{j=1}^{3} \mathbf{w}_{i_j}||^2}{A(\mathbf{W})} + (1 - ||\nabla I(\mathbf{w}_i)||) \right).$$

Force-based approach

- We are trying find $\mathbf{W}$ that is stable, i.e. $F_{ext}(\mathbf{W}) + F_{int}(\mathbf{W}) = 0$.
- F_{ext} is the force component due to image data and F_{int} is the force component due to surface shape.
- Idea: Simulate the forces starting from an initial mesh $\mathbf{W}^0$, that is iterate

$$\mathbf{W}^{t+1} = \mathbf{W}^t + F_{ext}(\mathbf{W}^t) + F_{int}(\mathbf{W}^t).$$

- Difficulty: How to obtain the external force based on image data?
- $\mathbf{F}_{ext}(\mathbf{w}_i) = \nabla(I * G)(\mathbf{w}_i^t); \mathbf{F}_{int}(\mathbf{w}_i) = \frac{1}{3}\sum_{j=1}^{3} \mathbf{w}_{i_j} - \mathbf{w}_i$ for a simplex mesh.

Deformable meshes: Pros and cons

- Pros:
  - Surface topology can be controlled
  - Shape can be smoothed effectively: Robustness to noise
  - Physics based meshes are simple and effective to implement
- Cons:
  - Initialization problem
  - Parameter sensitivity and over-regularization
  - Topology adaptation is (somewhat) complicated

References

- Kass, Witkin, Terzopoulos: Snakes: Active Contour Models. Int J. Computer Vision 1:321-331 1988.
- Delingette: General Object Reconstruction Based on Simplex meshes. Int J. Computer Vision, 32:111-142, 1999.
- Xu and Prince: Generalized gradient Vector Flows. Signal Processing, 71: 131 - 139, 1998
- Tohka, Mykkänen: Deformable Mesh for Automated Surface extraction from Noisy Images, Int J. of Image and Graphics, 4:405 - 432, 2004.

Level sets and gradient flows

- Idea: If the surfaces are represented by a level set of a higher dimensional function, the changes in surface topology and connectivity are automatically handled.
- The surface is $\mathcal{B} = \{(x, y, z) | \Psi(x, y, z) = 0\}$.
- We can also represent the surface parametrically $\mathcal{B} = \{b(r, s) : r \in [0, 1], s \in [0, 1]\}$
- Our purpose is the minimize the weighted area of the surface:

$$\int_{\mathcal{B}} \phi(I(b(r, s)))$$

Gradient flow

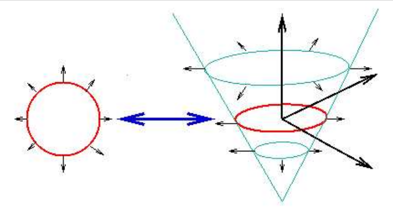
- Assume first that we are trying to minimize simply the area of the surface and wish the evolve the surface in the direction where the surface area decreases most rapidly.
- Such an evolution equation can be computed

$$\mathcal{B}_t = H\mathcal{N},$$

where H is the mean curvature and $\mathcal{N}$ is the unit normal.
- If we minimize the image weighted surface area, we get

$$\mathcal{B}_t = (\phi H - \nabla\phi \cdot \mathcal{N})\mathcal{N}.$$

Level sets



The figure is from the homepage of J. A Sethian

http://math.berkeley.edu/˜sethian/

Level sets

- Evolve Ψ instead of $\mathcal{B}$.
- $\mathcal{B} = \{(x, y, z) | \Psi(x, y, z) = 0\}$.
- The flow becomes

$$\frac{d\Psi}{dt} = \phi(I)|\nabla\Psi|\mathrm{div}(\frac{\nabla\Psi}{||\Psi||}) + \nabla\phi(I) \cdot \nabla\Psi.$$

- $\phi(I) = \frac{1}{1+|\nabla I|}$
- Note that the level set flow is defined on the zero-level set only that causes some numerical challenges. Particularly, an initial Ψ defined in $\mathbb{R}^3$ has to be constructed.

Pros and cons

- Pros
  - Topology changes are handled implicitly. Surfaces can break, merge , develop handles etc. Therefore, level sets are extremely powerful technique for extracting structures with complex topology.
  - Related advantage is that discretization is in the domain similar to the image domain and not in the surface domain, i.e. parameterization problems are not as big as with deformable meshes.
- Cons:
  - The surface topology cannot be controlled.
  - Related numerical schemes are quite involved, both computationally and conceptually.
  - Parameter sensitivity.

References 4

- Malladi, Sethian,Vemuri: Shape Modeling with Front Propagation, IEEE-PAMI, vol 17, no 3, 1995.
- Caselles et al. Minimal Surfaces: A geometric three dimensional segmentation approach, Numer math, 77:423 - 451, 1997.
- Siddiqi et al. Area and Length Minimizing Flows for Shape Segmentation, IEEE-TIP, vol 7, no 3, 1998.
- Sethian: Tracking Interfaces with Level Sets, American Scientist, May-June, 1997 (Available from http://math.berkeley.edu/˜sethian)
- Sarti A., Ortiz de Solorzano C., Lockett S.J, Malladi R: Computer-Aided Cytology: A Geometric Model for 3D Confocal Image Analysis. IEEE T Biomedical Engineering 47(12): 1600-1609, 2000.