ECSE-626

Statistical Computer Vision

Markov Random Fields Lecture #1

Introduction

- We define an optimization approach to solving difficult vision problems whereby contextual information is used to constrain the possible solutions.
- We will examine Markov Random Field (MRF) theory and show how it provides a convenient and consistent way of modeling context-dependent information such as spatial information.

Introduction

- Theory permits modeling vision problems using a Bayesian framework – mathematically sound, and tractable.
- From computational perspective, algorithms can be implemented locally and in parallel.

Introduction

- MRFs have been used to solve vision problems at all levels, although mostly at low level:
  - Image segmentation
  - Image restoration,
  - Surface reconstruction
  - Edge detection
  - Texture analysis
  - Optical flow
  - Shape-from-X,
  - Data Fusion, etc.

Introduction

- On a higher level:
  - Image registration, alignment,
  - Image matching,
  - Object recognition,
  - Classification, etc.

Introduction

- MRF theory tells us how to model the probability of contextual information (e.g. how features should be arranged).
- MRF theory is often used with statistical decision theory and estimation theory – formulate objective functions using standard optimization principles.

Introduction

- Most popular statistical criteria for optimality in vision literature – MAP probability.
- This gave rise to MAP-MRF modeling:
  - Derive form of posterior distribution,
  - Design optimization algorithm for finding maximum of posterior distribution.
- Theory enables development of algorithms for variety of vision problems using mathematical theory rather than relying on ad-hoc heuristics.

Introduction

- The notes for these lectures are taken mainly from the text:
- **Markov Random Field Modeling in Computer Vision,** Stan Z. Li, Springer-Verlag, Berlin, 1995
- Chapter 1 of this text is available online at:

http://www.nlpr.ia.ac.cn/users/szli/MRF_Book/MRF_Book.html

Motivation

In vision, the quantities solved for often have high dimensionality.

For example, a stereo vision algorithm might provide a dense depth map, defined on every pixel in the image. For a 512x512 image, we would therefore have a state vector which has 1/4 million elements!

Motivation

We would like to be able to represent solutions to vision problems that don't need such high dimensionality.

One approach is to represent the high dimensional state using a ***parametrized generative model***, where the parameter vectors have a much lower dimensionality than the quantities being modeled.

Concept

We can try to factor the *global* statistical model, $p(z_1, z_2, ...z_N)$, where N is the number of state variables into a product of *local* "interaction" factors.

Within the local models, each variable directly depends on only a few other "neighboring" variables. From a global perspective all variables are mutually dependent, but only through successive local interactions.

Concept

The notion of local interactions between variables can be formalized by considering a graph representation of the neighborhood structure:

Sites i and j are considered to be neighbors if they both appear within the same local component of the factorization.

Notation and Definitions

A site often represents the location of a pixel in an image array, but can also be more specific, such as the location of an edge or corner feature.

In MRF theory, sites are usually taken to be unordered (even if in some cases there is a natural ordering).
Inter-relationships between sites are maintained by "neighborhood systems".

Notation and Definitions

We will consider the general vision problem as a "Labelling" problem where we want to assign a label, l, from the label set, L, to each of the sites, $s \in S$.

For example, a label could be a binary edge indicator, a depth value, an optic flow vector, an object class, etc.

Notation and Definitions

In the terminology of random fields,
a labelling is called a *configuration*.

In vision applications, a configuration
refers to things like images, depth maps,
optic flow fields, image segmentations, etc.

Neighbourhood Systems

The sites in the set S are related to one another via a neighbourhood system.

A neighbourhood system for S is defined as:

$$N = \{N_i \mid \forall i \in S\}$$

where N_i is the set of sites neighbouring i.

The neighbouring relationship has the following properties:

(1) A site i is not a neighbour to itself: $i \notin N_i$

(2) The neighbouring relation is mutual: $i \in N_j \Leftrightarrow j \in N_i$

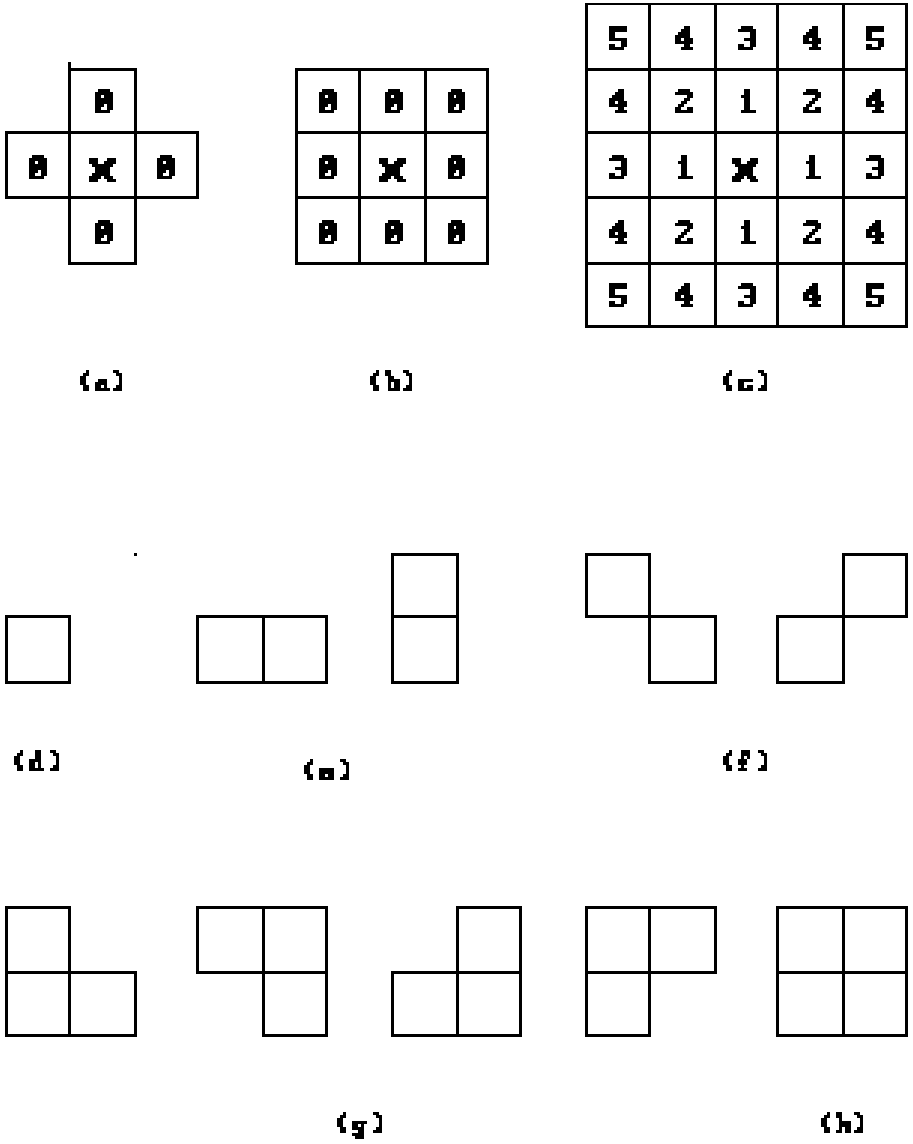
Neighbourhood Systems

For a regular lattice S, the neighbouring set of i is defined as the set of nearby sites within a distance r :

$$N_i = \{ j \in S \,|\, [\text{dist}(\text{pixel}_j, \text{pixel}_i)]^2 \leq r,\ j \neq i \}$$

where dist(A,B) denotes the Euclidean distance between A and B, and r takes on an integer value.

Examples of neighbourhood structures for regular lattices are shown on the next slide.



The numbers in figure (c) indicate the order of a particular neighborhood system.
For example, the first order neighborhood system is shown in figure (a). This is often called the 4-neighborhood.
The second order (or 8-) neighborhood system is shown in figure (b).

Cliques

The pair $(S, N) = G$ constitutes a graph, where S contains the nodes and N determines the links between the nodes.

A *clique*, c, for G is defined as a subset of sites in S. The clique can consist of either a single site $c = \{i\}$, or of a pair of neighboring sites $c = \{i, j\}$, or of a triple of neighboring sites $c = \{i, j, k\}$ and so on.

Cliques

The sites in a clique are ordered, so $\{i, j\}$ is not the same clique as $\{j, i\}$.

The type of a clique for a regular lattice is determined by its size, shape, and orientation. The clique types for the first and second order neighborhood systems of a regular lattice are shown in figures d-h on the previous slide.

Markov Random Fields (MRFs)

In one-dimension problems, such as time-series analysis or audio signal processing problems are often simplified by assuming a Markov model for the joint statistics of the signal values.

The *nth*-order Markov model holds that the conditional probability of a value at time k depends only on the values at a finite set of previous times $\{k\text{-}1, k\text{-}2, \ldots, k\text{-}n\}$.

The global dependence can derived by expanding out the local dependencies.

Markov Random Fields (MRFs)

The Markov random *field* is an extension of the 1-D Markov model to higher dimensions (of the site space).

The analogy to the 1-D local dependencies is that, in the Markov random field model, the conditional probability of a label at a given site depends on the labels at sites in the neighborhood structure.

Markov Random Fields (MRFs)

Let $F = \{F_1, ..., F_m\}$ be a family of random variables defined on the set S, in which each random variable F_i takes on a value $f_i \in L$. The family F is called a *random field*.

We use the notation $F_i = f_i$ to denote:

the event that F_i takes on the value f_i,

and the notation $\{F_1 = f_1, F_2 = f_2, ..., F_m = f_m\}$ to denote the joint event.

Markov Random Fields (MRFs)

For simplicity the joint event is often written $\{F = f\}$ where $f = \{f_1, f_2, ..., f_m\}$ is a *configuration* of F corresponding to a *realization* of the field (or a *sample* of the field).

For a discrete label set, the probability that a random variable F_i takes on a value f_i is denoted $P(F_i = f_i)$, often abbreviated to $P(f_i)$.

For continuous label sets the probability is given by a probability density function, $p(F_i = f_i)$, or just $p(f_i)$.

Markov Random Fields (MRFs)

F is said to be a Markov Random Field on S with respect to a neighborhood system N iff the following two conditions hold:

$$P(f) > 0, \forall f \in \mathcal{F} \qquad \text{(Positivity)}$$

$$P(f_i \mid f_{S-\{i\}}) = ? \qquad \text{(Markovianity)}$$

where $S-\{i\}$ is the set difference, $f_{S-\{i\}}$ is the set of labels at the sites in $S-\{i\}$, $\mathcal{F}$ is the set of possible labels, and

$$f_{N_i} = \{ f_j \mid j \in N_i \}$$

stands for the set of labels at the sites neighboring i.

Markov Random Fields (MRFs)

F is said to be a Markov Random Field on S with respect to a neighborhood system N iff the following two conditions hold:

$$P(f) > 0, \forall f \in \mathcal{F} \qquad \text{(Positivity)}$$

$$P(f_i \mid f_{S-\{i\}}) = P(f_i \mid f_{N_i}) \qquad \text{(Markovianity)}$$

where $S-\{i\}$ is the set difference, $f_{S-\{i\}}$ is the set of labels at the sites in $S-\{i\}$, $\mathcal{F}$ is the set of possible labels, and

$$f_{N_i} = \{ f_j \mid j \in N_i \}$$

stands for the set of labels at the sites neighboring i.

Markov Random Fields (MRFs)

There are two ways of specifying a MRF:

- in terms of the conditional probabilities, $P(f_i \mid f_{N_i})$.
- in terms of the joint probability $P(f)$.

The conditional probability approach has some drawbacks...

Markov Random Fields (MRFs)

The conditional probability approach has some drawbacks:

- No obvious way to compute joint probability from the conditional.
- Conditional probabilities are subject to non-obvious and restrictive consistency conditions.
- The specification of an equilibrium of a statistical process is in terms of the joint probability.

Markov Random Fields (MRFs)

Fortunately, there is an equivalence between so-called Gibbs distributions and MRFs which provides a tractable means of specifying the joint probability of an MRF.

Gibbs Random Fields

A set of random variables F is said to be a Gibbs Random Field (GRF) on S with respect to N iff its configurations obey a *Gibbs* distribution.

Gibbs Random Fields

As we have seen earlier in the course, a Gibbs distribution has the following form:

$$P(f) = \frac{1}{Z}\exp(-\frac{1}{T}U(f))$$

where

$$Z = \sum_{f \in \mathcal{F}} \exp(-\frac{1}{T}U(f))$$

is a normalizing constant called the Partition Function.

Gibbs Random Fields

The parameter T has the interpretation of temperature in a physical system.

The temperature T controls the sharpness of the distribution.

When the temperature is high, all configurations tend to be equally distributed.

Near the zero temperature, the distribution concentrates around the global energy minima.

Gibbs Random Fields

The function $U(f)$ is often referred to as the *energy function* in analogy to the free energy of a physical system. It can be written as a sum of *clique potentials* $V_c(f)$ over all possible cliques, C:

$$U(f) = \sum_{c \in C} V_c(f)$$

The value of $V_c(f)$ depends on the local configuration on the clique c.

Gibbs Random Fields

A GRF is said to be homogenous if $V_c(f)$ is independent of the relative position of the clique c in S.

A GRF is said to be isotropic if $V_c(f)$ is independent of the orientation of the clique.

GRFs are much simpler to specify if they are homogenous and isotropic.

Gibbs Random Fields

A special case of the Gibbs distribution is obtained when there are only cliques of up to size 2.

In this case the energy can be written as:

$$U(f) = \sum_{i \in S} V_1(f_i) + \sum_{i \in S} \sum_{i' \in N_i} V_2(f_i, f_{i'})$$

Gibbs Random Fields

$$U(f) = \sum_{i \in S} V_1(f_i) + \sum_{i \in S} \sum_{i' \in N_i} V_2(f_i, f_{i'})$$

For this case the conditional probability is:

$$P(f_i \mid f_{N_i}) = \frac{\exp(-[V_1(f_i) + \sum_{i' \in N_i} V_2(f_i, f_{i'})])}{\sum_{f_i \in L} \exp(-[V_1(f_i) + \sum_{i' \in N_i} V_2(f_i, f_{i'})])}$$

Equivalence of Markov and Gibbs Random Fields

An MRF is characterized by its local properties (the Markov assumption) while
a GRF is characterized by its global properties (the Gibbs distribution).

Nonetheless, the two are equivalent. This equivalence is established by the *Hammersley-Clifford theorem*…

Hammersley-Clifford Theorem

F is a Markov Random Field on a set of sites S
with respect to a neighborhood structure N
if and only if
F is a Gibbs Random Field on a S with respect to N.

The converse to this theorem is also true:
-i.e. that F is a Gibbs random field
if and only if
F is a Markov random field.

Equivalence of Markov and Gibbs Random Fields

The practical value of the Hammersley-Clifford theorem is that it provides a simple way of specifying the joint probability for a Markov Random Field.

One can specify the joint probability by specifying the clique potential functions of the Gibbs Random Field.

The clique potential functions encode the knowledge or preference about interactions between labels.

Equivalence of Markov and Gibbs Random Fields

To actually *compute* the joint probability of the MRF it is necessary to compute the partition function.

This is usually intractable, since the partition function is a sum over the (combinatorial) number of different configurations of the field.

The explicit computation of the partition function can be avoided if all the parameters defining $U(f)$ are known.