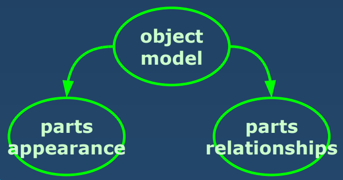
Conditional Random Fields for High-Level Part Correlation Analysis in Images

Giuseppe Passino, Ebroul Izquierdo
{giuseppe.passino,ebroul.izquierdo}@elec.qmul.ac.uk
Multimedia and Vision Lab, Queen Mary, University of London

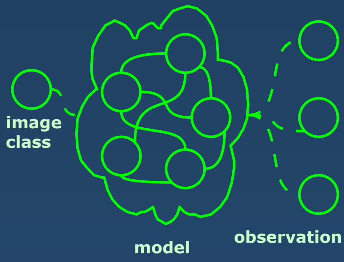
Introduction

The problem is to recognise objects in images considering them as the sum of their parts.



Graphical Modelling

The model is represented as a graph: the relationship between the parts are the link between the corresponding nodes.



Which Probability?

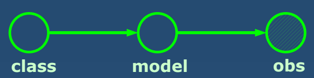
How can a probabilistic function describe such a system?

$P(\text{obs, conf, class}|\theta)$

- obs: extracted observation (given)
- conf: model configuration (hidden)
- class: image class (goal)
- θ: model parameters

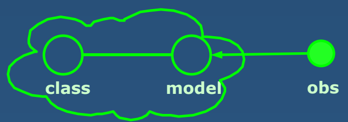
Generative vs Conditional Models

Generative models explain the observation with the configuration of the model.



$$P(\text{class/obs}) \propto \sum_{\text{conf}} P(\text{obs/conf})\, P(\text{conf/class}) \cdot P(\text{class})$$

In the conditional models the observation is embedded in the modelled probability. No cause-effect relationship between object class and model configuration.



$$P(\text{class/obs}) \propto \sum_{\text{conf}} P(\text{class,conf})\, P(\text{conf/obs})$$

What is Modelled?

Generative models: P(obs/conf)

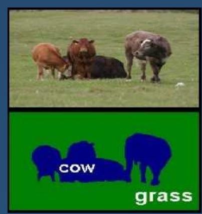
- Model conform to the reality
- If the observation is known, each possible configuration has to be considered

Conditional models: P(conf/obs)

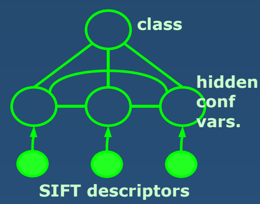
- Probability of a configuration given the observation – efficiency
- Difficult to directly model the dependence of the configuration from the parameter

Conditional Random Fields (CRF)

@Cambridge: pixel-based segmentation



@MIT: hidden parts CRF+SIFT



Where is the Semantics?

The fine-grained elements have very reduced semantic content.

The aim of the research is to exploit the CRF to model the semantic relationships!

Take into account higher parts: more semantics associated

System Structure

Block cascade system:

Feature Extraction → Graph Generation → MAP Class Inference

- The feature extraction is colour-based and is aimed to a semantic image segmentation
- The graph structure is a tree
  - inference via belief propagation
  - only local dependences retained

Anisotropic Diffusion

Segmentation process: non-uniform smoothing (uniform colour regions are isolated).

$$\partial/\partial t\, I(\underline{x},t) = \nabla(c\, \nabla I(\underline{x},t))$$

A non-constant diffusion coefficient makes the process equivalent to a Gaussian blurring with variable variance

First Results

Model	rel. time	accuracy
our	0.13	77%
ref. (MIT)	1.0	90%

Training & Inference

The training process is a standard Maximum Log-likelihood estimation:

$$\theta = \arg\max \{ \sum_{\text{train}} P(\text{class/obs}) \}$$

The parts are not labelled – (hidden model)

Belief propagation + Quasi Newton
(inference) (optimization)

The classification is MAP (straightforward):

$$c = \arg\max_{\text{class}} \{ P(\text{class/obs}) \}$$

Conclusions

- ✔ The system works!
- ✔ Reducing the parts number the inference is easier.
- ✔ It's straightforward to add new features.
- ✗ The probability model choice is not easy.
- ✗ How to segment?

Next: integrate better object models in the image graph.