Bottom-Up and Top-Down for Image Segmentation and Object Category Detection

Iasonas Kokkinos, Lotus Hill Institute, 2006
Part I: Global Models for Object Categories

An Expectation Maximization Approach to the Synergy Between Image Segmentation and Object Category Detection
Iasonas Kokkinos and Petros Maragos, ICCV 2005
Synergy Between Image Segmentation and Object Detection

- **Image Segmentation**: Partition of an image into homogeneous regions
- **Homogeneity criteria**: color, texture, motion interpretation based on a common model.

- **Object Detection**: Decision about whether an object is contained in an image region
- **The region is considered as a sliding box**

**Problem Intersection**: Explaining image regions using object models.

**Applications**:
- Top down segmentation
- Validation of bottom up detection
- Robust-to-occlusion model fitting
Graphical models for object detection

- Graphical models:
  - Representation of joint distributions.
  - R.V. dependencies encoded via a graph.
- For object models:
  - Random variables: part locations
    local appearance parameters/features
  - Dependencies: relative locations or appearance constraints
- Object detection: Maximization of

\[
P(L|I) \propto P(I|L)P(L) = \prod_i \Phi_i(I_i|l_i) \prod_{(i,j) \in C} \Psi_{i,j}(l_i, l_j)
\]

\[
L = (l_1, l_2, l_3, l_4, l_5)
\]

- Maximization is tractable for tree-structured graphs
- Belief Propagation:

\[
m_{i,j}(l_j) = \sum_{l_i} \Phi_i(l_i) \Psi_{i,j}(l_i, l_j) \prod_{k \in \mathcal{N}(i) \setminus j} m_{k,i}(l_i)
\]

\[
b_i(l_i) = \sum_{l_i} \Phi_i(l_i) \prod_{j \in \mathcal{N}(i)} m_{j,i}(l_i)
\]
Belief Propagation for Part-Based face detection

\[ m_{ij}(l_i) = \sum_{l_i} \Phi_i(l_i) \Phi_{i,j}(l_i,l_j) \prod_{k \in \{N(i) \cap j\}} m_{k,i}(l_i) \]

\[ b_i(l_i) = \sum_{l_i} \Phi_i(l_i) \prod_{j \in \{N(i)\}} m_{j,i}(l_i) \]
Generative models for object categories

- Object models:
  - PCA for faces (Eigenfaces)
  - Active Appearance Models

Shape: \( S(x; s) = \sum_i s_i S_i(x), \)

Appearance: \( T(x; t) = \sum_i t_i T_i(x) \)

Synthesis: \( I(S(x; s)) \approx T(x; t) \)

- Parameter estimation:
  - Minimization of reconstruction error
    \[ E(s, t) = \sum_x [I(S(x; s)) - T(x; t)]^2 \]
  - Efficient algorithms (CMU group)
Synergy Problem Formulation

- Top-Down approach: object models explain image areas.
- Segmentation can be cast as assigning observations to models.
- ‘Chicken and egg’ problem:
  - Model fitting: uses observations in object support.
  - Region formation: uses fitted model predictions.
- Model similar to mixture modelling:
  \[
  \log P(I|\mathcal{A}) = \sum_{n=1}^{N} \log P(I_n|\mathcal{A}) = \sum_{n=1}^{N} \log \sum_{H_n} P(I_n, H_n | \mathcal{A})
  \]

  \[\mathcal{A}\text{ Parameters} \]
  \[I_n\text{ n-th observation} \]
  \[H_n\text{ hidden assignment variable} \]
  \[K\text{ #components} \]
  \[N\text{ #of observations} \]
  \[H_n = (h_{n,1}, \ldots, h_{n,K}) \]

- Treat segmentation as a field of hidden variables
Expectation Maximization algorithm

- Considering the set of hidden variables known, the likelihood of the full observation set can be written as:

\[
\log P(I, H) = \sum_n \log P(I_n, H_n) = \sum_n \sum_k h_{n,k} \log \pi_k P_k(I_n|\theta_k)
\]

\[
P(I_n, H_n|A) = P(I_n|H_n, A)P(H_n|A) = \prod_{k=1}^K [P_k(I_n|\theta_k) \pi_k]^{h_{n,k}}
\]

- Expectation Maximization algorithm:
  - E Step: Form likelihood expectation using current parameters

\[
\langle \log P(I, H) \rangle = \sum_n \sum_k \langle h_{n,k} \rangle \log \pi_k P_k(I_n|\theta_k) \quad \langle h_{n,k} \rangle = P(h_{n,k} = 1|I_n, A_k^*) = \frac{\pi_k^* P_k(I_n|\theta_k^*)}{\sum_j \pi_j^* P_j(I_n|\theta_j^*)}
\]

  - M Step: Update parameters to maximize the expectation.

\[
E_{n,k} = \langle h_{n,k} \rangle \quad \pi_k = \frac{\sum_n E_{n,k}}{N}, \quad \theta_k = \text{argmax} \sum_n E_{n,k} \log P_k(I_n|\theta_k)
\]
EM for Bottom-Up/Top-Down

- Object detection to initialize support and model parameters.
- E: Object support estimate using observation likelihood.
- M: Estimation of model parameters using object support
- Variational approach to BU/TD

Introduce AAMs:

Iterations
Implementation of the E step at the image segment level

- Alternative background hypothesis:
  - Constant model in the interior of images.
- Hidden variables assign image segments to the objects.
- E step:

\[
E_{S,\mathcal{O}} = \frac{P(I_S|\mathcal{O})}{P(I_S|\mathcal{O}) + P(I_S|\mathcal{B})} = \frac{1}{1 + \exp \left( \log \frac{P(I_S|\mathcal{B})}{P(I_S|\mathcal{O})} \right)}
\]

- Practical problems:
  - Background model is more complicated
    - Introduction of a penalty term.
  - Correlated reconstruction Error (overcounting of evidence)
    - Division by a normalizing term.
Fragment-Based Top-down Segmentation

- Problems:
  - Inaccurate bottom-up segmentations are inherited to top-down.
  - Jagged boundaries
  - Problematic combination of bottom-up and top-down information.
Top Down Segmentation Using Curve Evolution

Fragment-Based E Step

Curve Evolution

Fragment-Based E Step

Curve Evolution
**M step: modified AAM fitting**

- **Typical AAM fitting criterion:**
  \[
  \sum_x H(x) \left[ I(S(x; s)) - T(x; t) \right]^2
  \]

- **EM approach criterion:**
  \[
  \sum_x E(S(x; s)) D(x; s) H(x) \left[ (I(S(x; s)) - T(x; t))^2 - (I(S(x; s)) - B(S(x; s)))^2 \right]
  \]

- **Automated introduction of segmentation information, robustness to occlusions.**

Typical AAM fitting scenario 

Occlusion 

Without EM

With EM 

Segmentation
Use of Synergy for Object Detection

Two complementary cues of information from top-down:

- Image area assigned to object region:
  - cue for fidelity of observations to model hypothesis predictions.
- Values of AAM model parameters (witnessed by typical/non-typical synthesis results for True/False hypotheses)
  - Bottom-up, top-down helps provide robust parameter estimates.
Synergetic Object Detection

- Use 3 classifiers: bottom-up information, AAM parameters, segmentation support.
- Combination of classifier results $C = \{C_1, C_2, C_3\}$ based on supra-Bayesian approach:
  - Treat classifier outputs as random variables
  - Consider their joint distribution is Gaussian
    \[
    \frac{P(O|C)}{P(B|C)} = c \frac{P(C|O)}{P(C|B)} = c \prod_{k=1}^{K} \frac{P(C_k|O)}{P(C_k|B)}
    \]
- Detection Results:

<table>
<thead>
<tr>
<th>Precision</th>
<th>#True Positives</th>
<th>#Detections</th>
<th>#True Positives</th>
<th>#Objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Repeat Experiment introducing occlusion in the test set:
Part II: Part-Based models for Object Categories

Bottom-Up and Top-Down Object Detection Using Primal Sketch Features and Graphical Models, Iasonas Kokkinos, Petros Maragos and Alan Yuille, CVPR 06
Use of Interest Points for Object Detection

- Major object detection problem: huge number of object location hypotheses:
  (e.g. $10^4$-$10^5$ pixels, 5 scales, 10 orientations)
- Approach: Interest points to extract important image areas

  + Sparse set of points (typically $\sim 10^3$).
  + Facilitate introduction of scale and orientation invariance.
  + Use of machine learning techniques for learning models.
    - Use of thresholds for operator.
    - Blob, corner and junction features do not capture all possible image structures.
Scale-Invariant Edge and Ridge Features

- Scale selection problem for edge detection:
  - Large Scales: robustness to noise.
  - Small Scales: accurate edge localization.
- Scale selection for ridge detection: width of strongest ridge structure.
- Introduction of normalized differential operator to determine scale (Lindeberg).

\[ L_t = I * G_t \]
\[ G_\gamma = t^\gamma \left( \left( \frac{\partial L_t}{\partial x} \right)^2 + \left( \frac{\partial L_t}{\partial y} \right)^2 \right), \quad \gamma = \frac{1}{2} \]
\[ A_\gamma = t^{2\gamma} \left( \left( \frac{\partial^2 L_t}{\partial x^2} \right)^2 + \left( \frac{\partial^2 L_t}{\partial x \partial y} \right)^2 + \frac{1}{4} \left( \frac{\partial^2 L_t}{\partial x \partial y} \right)^2 \right), \quad \gamma = \frac{3}{4} \]
Scale Space Primal Sketch

Scale Space: \( L_t = I \ast G_t, \quad G_t : \text{Gaussian kernel, } \sigma = \sqrt{t} \)

Ridge Strength, \( A_t = t^{3/2} \left( (L_{xx} - L_{yy})^2 + 4L_{xy} \right) \)

Edge Strength, \( G_t = t^{1/2} \left( L_x^2 + L_y^2 \right) \)

Blob Strength, \( H_t = t \left( L_{xx} + L_{yy} \right)^2 \)

\( \sigma = 4.1 \quad \sigma = 8.7 \quad \sigma = 13.4 \)
Interest Points from Multiscale Analysis

- Scale - space tracking of strength maxima locations.
  - Ridges
  - Edges

- Approximate continuous curves using straight lines
  - Use of scale-invariant criterion.
  - Ridges
  - Edges
  - Blobs
Construction of Codebook Representations

• Interest point clustering (k-means + EM).

• Selection of codebook subset that is suitable for object detection.
  – Choose clusters based on individual detection performance.

• Codebook representations:
Bottom-Up Object Detection

- Bottom Up Feature Extraction
- Assignment of features to codebook entries

- Likelihood estimation for object hypothesis
Introduction of Top-Down information

- Major problem: missing interest points.
- Approach: use knowledge about object
  - Graphical model formulation, use Belief Propagation.

\[
\begin{align*}
    m_{i,j}(l_j) &= \sum_{l_i} \Phi_i(l_i) \Psi_{i,j}(l_i, l_j) \prod_{k \in \{\mathcal{N}(i) \setminus j\}} m_{k,i}(l_i) \\
    b_i(l_i) &= \sum_{l_i} \Phi_i(l_i) \prod_{j \in \{\mathcal{N}(i)\}} m_{j,i}(l_i) \quad l_i = (x_i, \sigma_i, r_i, \theta_i)
\end{align*}
\]

- However: huge number of terms to be added:
  - Monte Carlo Approximation

\[
\int_{l} \Phi(l) P(l) dX \simeq \frac{1}{N} \sum_{i=1}^{N} \Phi(l_i), \quad l_i \sim P(l) \quad \text{where} \quad P(l_i) = \prod_{j \in \{\mathcal{N}(i)\}} m_{i,j}(l_i)
\]

- Computationally demanding likelihood evaluation:
  - Introduction of efficient method
Efficient Likelihood Estimation

- Use simple templates for detected features

- Generative model: constant value in the interior of each subregion + white Gaussian noise.

- Speedup of parameter and likelihood estimation using Stoke’s rule:

\[
\int \int_S f(x, y) \, dx \, dy = \int_{\partial S} P \, dx + Q \, dy = \int_0^l (P, Q) \cdot T
\]

\[
\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = f(x, y)
\]

\[
Q(x, y) = \frac{1}{2} \int_0^x f(x, y) \, dx
\]

\[
P(x, y) = -\frac{1}{2} \int_0^y f(x, y) \, dy
\]

- Curvilinear instead of area integrals. \(O(N^2) \rightarrow O(N)\)

- Generalization of ‘integral image’ idea by Viola & Jones.

- Typically, instead of 1000-5000 points only 50-100 are used.
Top-down interest point detection

\[ P(l_i) = \prod_{j \in \{N(i)\}} m_{i,j}(l_i) \]
Graphical model construction for part locations

- Major problem:
  - During model construction most codebook pairs, triples are not simultaneously in the image.
  - Using top-down information we can retrieve their locations and thereby estimate statistics of pair locations
- Given the covariance matrix the Minimum Spanning Tree (MST) gives the optimal tree-structured distribution
- Structure learning results
Joint Top-Down and Bottom-Up Object Detection

- Use of top-down information for hypothesis verification
  - Cues: Which keypoints were retrieved
    - Their relative locations
    - Bottom-Up detection strength

- Combination using supra-Bayesian method.
Top-down validation results
Detection Results on Benchmarks

Bottom-Up & Top-Down Car Detection

Speed vs. Accuracy tradeoff

Comparison with prior work

Bottom-Up & Top-Down Face Detection

Bottom-Up & Top-Down Cow Detection

Bottom-Up & Top-Down Horse Detection