Exploiting Part-Based Models and Edge Boundaries for Object Detection

Josephine Sullivan, Oscar Danielsson and Stefan Carlsson, KTH
http://www.csc.kth.se/cvap
Recent Approaches to Object Detection

**Bag-of-Words** Typically: Histogram occurrence of *visual words* in an image.
Lowe; Schmid et al; Triggs et al

**Constellation Models of Parts** Have a set of parts, learn their relative position wrt one another. Detect an instance of an object by detecting parts in the correct configuration. (Define the spatial relationships in a **restrictive** way.)
Fergus, Perona and Zisserman; Darrell et. al; Feltzenwalb and Huttenlocher

**Match Template Curves** Shape Context matching, Chamfer Matching
Berg and Malik; Feltzenwalb and Huttenlocher
Lessons Learned

Their Successes

- Objects containing distinctive parts whose appearance and shape remain constant across a class can be recognised by constellation and bag-of-words models. Unfortunately, not many such objects - cars, motorbikes, faces.

- Shape matching algorithms based on the matching of image gradient templates can cope with large intra-class variation when no clutter is present. However, in clutter and/or with occlusion they tend to break down.
Our Method overview
Part-Based Model

An object is defined by the coordinates $\{x_i\}_{i=1}^P$ of $P$ parts.
Problem Definition

Have a novel test image. Want to localize the cups existing in this image.
Apply Part Detectors

Each cross is the location of a detected part with colour indicating part number.
Find Plausible Configurations

From these detections want to identify spatially plausible configurations of the cup. The circled detections shows one such configuration.
The consistency of contour information with the detected part locations is used to score each **spatially plausible** configuration. These are then thresholded to determine the presence or not of an object.
Part-Based Model of Shape
Shape’s probability distribution

Assume ∃ a probability distribution such that \( p(x_1, x_2, \ldots, x_P) > 0 \) for instances generated by our object class.

Can decompose \( p(x_1, x_2, \ldots, x_P) \) via the product rule into:

\[
p(x_1, \ldots, x_P) = p(x_{\sigma(1)}) \prod_{j=2}^{P} p(x_{\sigma(j)}|x_{\sigma(1)}, \ldots, x_{\sigma(j-1)})
\]

where \( \sigma : \{1, \ldots, P\} \rightarrow \{1, \ldots, P\} \) is a permutation defining the order of the decomposition.

How can we model these conditional distributions efficiently and expressively?
Exploit Linear Dependencies

Assume there exist linear dependencies between certain part locations and that there is a permutation $\sigma$ and $\beta_j = (\beta_{1,j}, \ldots, \beta_{j-1,j})$’s such that

$$x_{\sigma(j)} = \sum_{k=1}^{j-1} \beta_{k,j} x_{\sigma(k)} + \alpha_j + \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0, \Sigma_j)$$

with $\sum_{k=1}^{j-1} \beta_{k,j} = 1$, $\alpha_j \in \mathbb{R}^2$. 
A learned conditional distribution

An example $p(x_{\sigma(j)}|x_{\sigma(1)}, \ldots, x_{\sigma(j-1)})$ learned using such a linear model.
Learning the Model
The Learning Problems

1. Estimate the parameters $\alpha_j$, $\beta_j$ and $\Sigma_j$ for each of the conditional distributions.

2. Estimate the permutation $\sigma$, the order in which the decomposition of the probability distribution should be performed.
Remove differences between training shapes due to affine transformations. From this data learn each $\alpha_j$, $\beta_j$ and $\Sigma_j$ in the obvious way.
Order of decomposition matters

Not all subsets of parts can estimate the location of another part.
Search for the order of decomposition

Perform a greedy search
Searching for the Learned Model
Generating plausible configurations

Initialize a triplet of parts

Predict part 2 given 1, 4, 6

Predict part 5 given 1, 2, 4, 6

Predict part 3 given 1, 2, 4, 5, 6
Scoring the Plausible Configurations
Exploit Connectedness

A selection of the found spatially plausible configurations

Which is the best one and how do we identify it?
Exploit Connectedness

Must examine the image to find the true configuration

Best configuration

Why ?? Its parts are connected by boundary curves.
Curve representation

Represent the curve joining parts by a spline. Represent the variation in the spline using a PCA basis.

1st mode of variation  2nd mode of variation
Finding the curve

Measurement Process
Measure if hypothesized curve is consistent with the underlying image data. Compare the curve’s normal to the image gradient direction of the nearby pixels.

Best Curve
The red curves are the best fit between the curve normals and the underlying image gradient directions when the PCA curve space is searched.

Final score based on the goodness of fit of best curves.
Experiments & Results
Experiments

The part-based models used.

CalTech 101 database used for learning

ETHZ Shape, TUD Car Sides databases used for evaluation.
Detections by the part detectors

Note: there are many false positives for each part except the car wheels.
Detections shown in red are classified as an object using the threshold at the .4 fppi detection rate. The other detections are classified as non-object.
The **true positive rate** Vs the **false positive rate per image** (fppi) is plotted. At .4 fppi and 50%(20%) overlap the detection rate for cups is .852(.87) and for bottles .85(.92).
Conclusions

Useful findings

- Boundary contours can be used for verification
- Search mechanism via predictive distributions

Improvements to be Made

- Reduce complexity demands
- Deal with missed detections