Detection and Segmentation of Multiple, Partially Occluded Objects by Grouping, Merging, Assigning Part Responses

Bo Wu   Ram Nevatia

Institute for Robotics and Intelligent System
University of Southern California, LA, CA

Presented by Somchok Sakjiraphong

Occluded Object

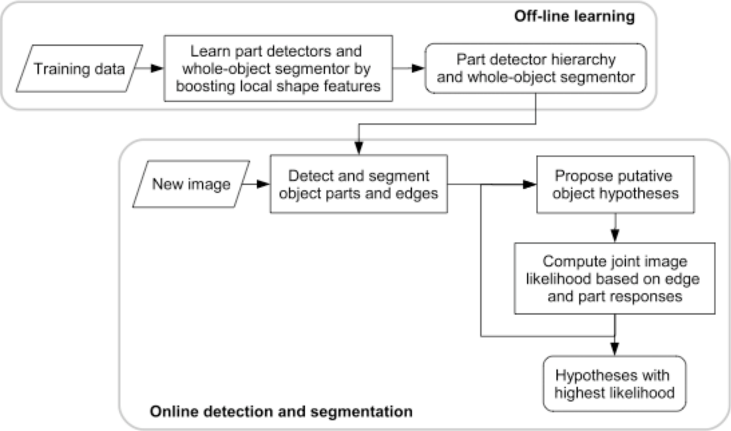
partially occluded object

Outline

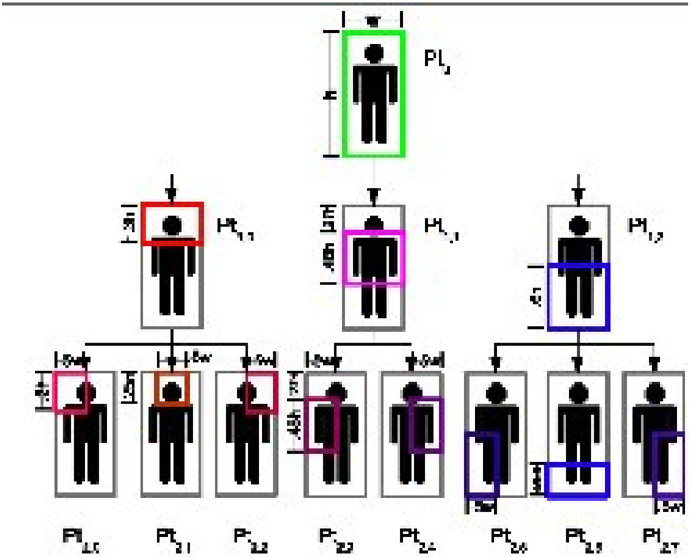
Introduction

- In recent years, methods for direct detection of objects such as (pedestrians, faces, cars) have become popular.
- No prior segmentation was applied rather window of various size was applied so as to check the presence of objects.
- Good performance but not very precise.

Outline of Approach



outline of approach



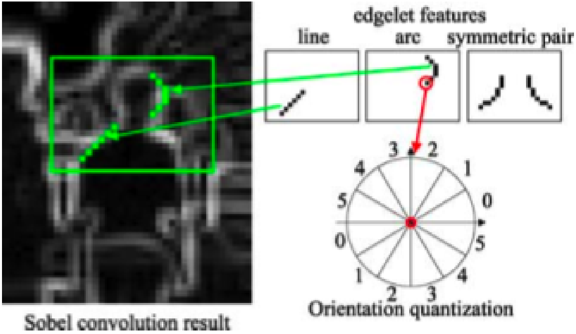
Learning Part Detectors

- A detector is learned for each node.
- Feature sharing is possible between parent and child nodes.
- Boosting algorithm is applied to select features and create a classifier.

Learning Part Detectors

Image Features

- Edgelets features are used to model the appearance of and object.
- An edgelets is just a a line or a curve.

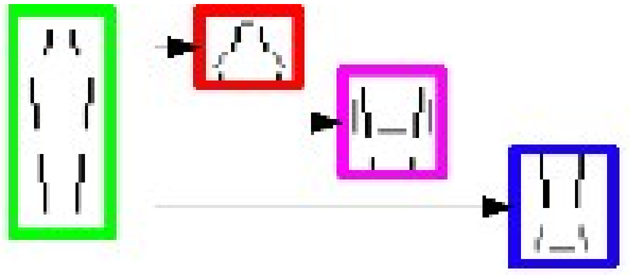


edgelets features

Learning Part Detectors

Feature Sharing

- Inheritance of edgelet features before boosting (except for the whole-object node).
- For each inherited edgelets those points that are outside the node's region are removed.



feature sharing between parent and child nodes

Learning Part Detectors

Learning Algorithm

The learning algorithm used here is called CBT (Cluster Boosting Tree)

Viola and Jones cascade structured classifier is suitable for object-classes with small intra-class variation.
For more diverse patterns such as multi-view faces/human bodies a more powerful classifier model is needed such as tree structured classfier.

Tree structure classifier uses “divide-and-conquer”: divide the object class into several categories and learn a model for each of them.

When the appearance of object changes such as human-body it’s not easy to one dominating property to divide the samples.

CBT divides the sample space by unsupervised clustering based on the discriminative image features.

Detecting Body Parts and Object Edges

Given an image, the part detector is applied and the output is part responses plus the images edges that correspond to the object.

$$E\{f(\mathbf{x})|\mathbf{x} \in X_{+}\} > E\{f(\mathbf{x})|\mathbf{x} \in X_{-}\}$$

f is the edgelet feature and we call it a positive feature if it's average matching score on the positive object class is more than that of the negative object class.

Positives features with top 5% are retained

Detecting Body Parts and Object Edges

Clustering Algorithm (1)

- One detector usually contains about 1000 positive features. Some of these edgelts correspond to the same edge pixels.
- A clustering algorithm is applied to remove the redundant edgelets.

$$A(E_1, E_2) \triangleq \frac{1}{k} \sum_{i=1}^{k} \left\langle \mathbf{u}_{1,i} - \bar{\mathbf{u}}_1, \mathbf{u}_{2,i} - \bar{\mathbf{u}}_2 \right\rangle \cdot e^{-\frac{1}{2}||\bar{u}_1 - \bar{u}_2||^2} \quad (1)$$

If $k1 \neq k2$ then they are first aligned by their center points and the longer features is shorten by removing points from both ends.

$$Affinity = (1) \times \frac{min\{k_1, k_2\}}{max\{k_1, k_2\}}$$

Detecting Body Parts and Object Edges

Extracted object edgelet pixels

Extracted object edgelet pixels

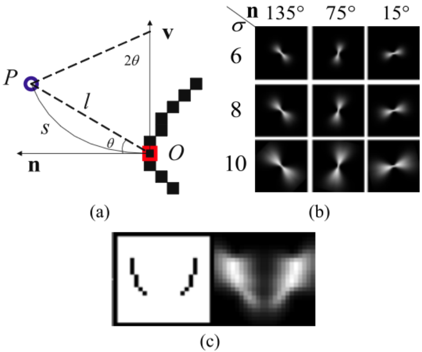
We need to build a weak classifier for segmentation.

Feature sharing is between weak detectors and weak segmentators.

Weak segmentation classifier is a function from the space $X \times U$ to a real value figure-ground classification confidence space where U is the 2D-coordinate space.

Segmentation of Individual Object

Effective Field



O is a point on the edgelet, normal **n** and tangent **v** are known, P is a neighbor of O and OP is the osculating circle at O that goes through P. The effect of O on P is defined by $DF(s, k, \sigma) = exp(-\frac{s^2+ck^2}{\sigma^2})$

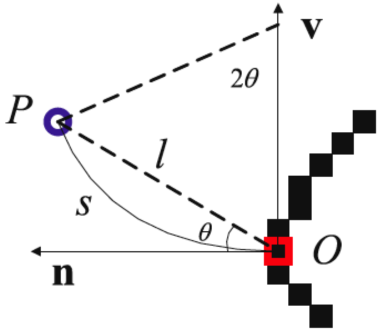
Segmentation of Individual Object

Effective Field

The effect of O on P is defined by $DF(s, k, \sigma) = exp(-\frac{s^2+ck^2}{\sigma^2})$

where l is the Euclidean distance between O and P, θ is the angle between **n** and $\bar{OP}$, $s = \frac{l\theta}{2sin\theta}$ is the length of arc OP,$k = \frac{2sin\theta}{l}$

$\mathbf{F(u)} = max\{F_1(u), ..., F_k(u)\}$



Sample Weight Evolution

For detection problems real value weight is $D^{(d)}$ is assigned to each sample. During boosting weight of misclassified samples are increased while those correctly sampled are decreased.

For segmentation problem, difficulties of different samples vary and different position of the same sample also vary.

$$D_{t+1}^{(d)}(\mathbf{x}) = D_t^{(d)}(\mathbf{x}) \exp\left[-y h_t^{(d)}(\mathbf{x})\right], \quad \forall \mathbf{x} \in S$$

$$\mathbf{D}_{t+1}^{(s)}(\mathbf{x};\mathbf{u}) = \mathbf{D}_t^{(s)}(\mathbf{x};\mathbf{u}) \exp\left[-\mathbf{m}(\mathbf{u}) h_t^{(s)}(\mathbf{x};\mathbf{u})\right],$$

$$\forall \mathbf{u} \in \mathcal{U}$$

Joint Analysis for Multiple Objects

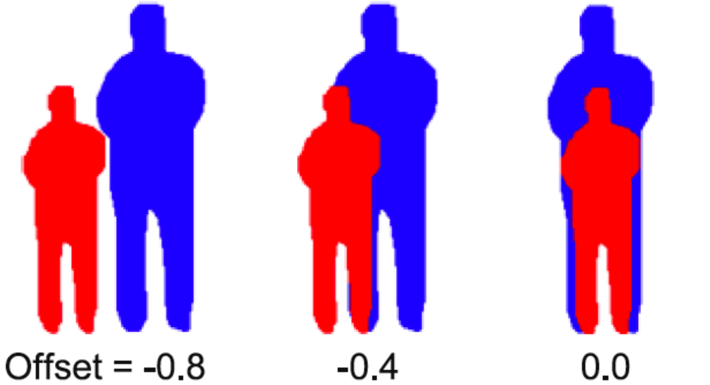
Proposing Object Hypotheses

Propose an initial object hypotheses sorted such that the y-coordinates are in descending order.

The examined multiple object configuration

Joint Analysis for Multiple Objects
Extracting Silhouettes

Segment object hypotheses and extract their silhouettes.



Occlusions mapping of two humans. (Red one is in front and blue one is at the back)

Joint Analysis for Multiple Objects

Joint Occlusion Map Silhouettes

The visible silhouettes obtained from occlusion reasoning

Joint Analysis for Multiple Objects

Matching Object Edges with Visible Silhouettes

Training Part Detectors

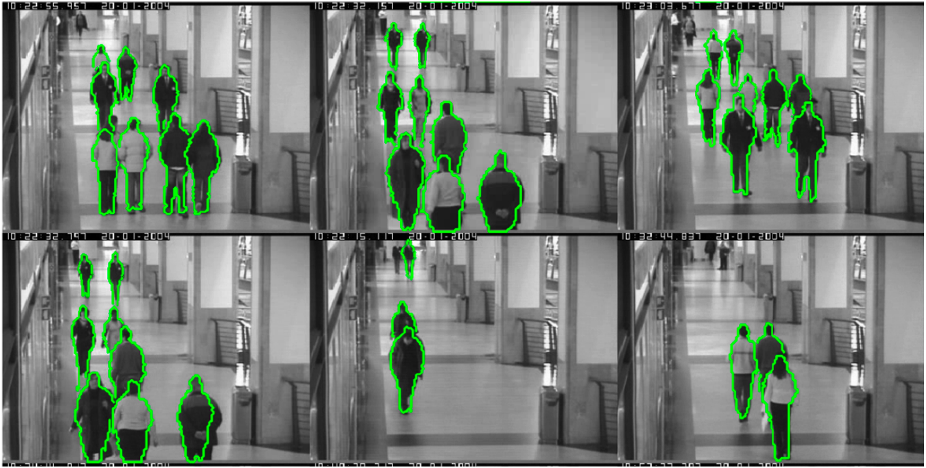
Training Samples

Results on USC Test Set

Part	Recall	Precision
Full-body	0.7638	0.9367
Head-shoulder	0.7269	0.9471
Torso	0.7934	0.9110
Legs	0.5720	0.8470
Head	0.6679	0.7702
Left shoulder	0.6863	0.8857
Left arm	0.7860	0.8694
Left leg	0.5240	0.8208
Feet	0.5092	0.7624

Table 1 Performance of part detectors on the USC pedestrian set B. (The performance of right shoulder/arm/leg is similar to their left counterparts)

Results on USC Test Set



Example detection and segmentation results on the USC pedestrian.

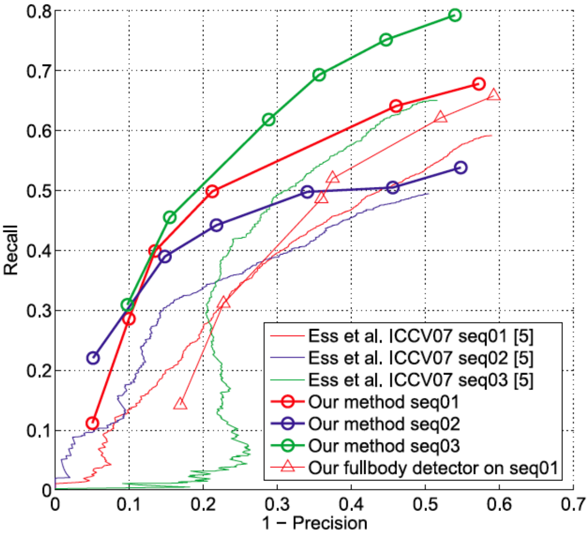
Occlusion degree (%)	$25 \sim 50$	$50 \sim 75$	>75
Human number	34	31	10
Sheet et al. (2007)	87	91.4	92.6
Our previous method (Wu and Nevatia 2005)	91.2	90.3	80
This method	94.12	93.55	90

Table 2 Detection rates (%) on different degrees of occlusions.

Performance

- No scene structure or background subtraction to facilitate detection
- A test image of 384 x 288 pixels was used and humans from 24 to 80 pixels wide were searched.
- 4 threads running the detection of different parts simulataneously on a Intel Xeon 3.0GHz CPU.
- Average speed is about 3.6 seconds per image

Results on Zurich Training Set



Detection precision-recall curves on the Zurich mobile pedestrian sequences.

Results on Zurich Training Set

Zurich Test Result

Conclusion and Discussion

Boosted classifiers are learned for each nodes. For the whole object node the segmentor is learned by boosting local features. For partially occluded object silhouette is extracted and a joint likelihood of multiple objects is maximized to find the best interpretation.

‘’The experimental results show that our method outperforms the previous one.“

The approach of this paper is domain dependent for example the design of part hierarchy only work humans/pedestrians.

The ground plane assumption is not valid for all objects.