Siemens Corporate Research | UCSD Jacobs Computer Science and Engineering | IGERT

Supervised Learning of Edges and Object Boundaries

Piotr Dollár
Computer Science and Engineering
University of California, San Diego
pdollar@cs.ucsd.edu

Zhuowen Tu
Lab of Neuro Imaging, School of Medicine
University of California, Los Angles
zhuowen.tu@loni.ucla.edu

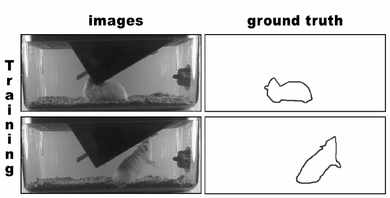
Serge Belongie
Computer Science and Engineering
University of California, San Diego
sjb@cs.ucsd.edu

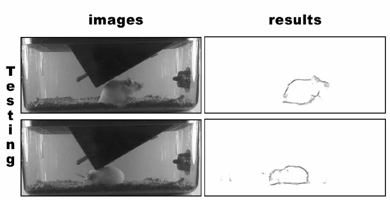
Abstract

Edge detection is one of the most studied problems in computer vision, yet it remains a very challenging task. It is difficult since often the decision for an edge cannot be made purely based on low level cues such as gradient, instead we need to engage all levels of information, low, middle, and high, in order to decide where to put edges. In this paper we propose a novel supervised learning algorithm for edge and object boundary detection which we refer to as Boosted Edge Learning or **BEL** for short. A decision of an edge point is made independently at each location in the image; a very large aperture is used providing significant context for each decision. In the learning stage, the algorithm selects and combines a large number of features across different scales in order to learn a discriminative model using an extended version of the Probabilistic Boosting Tree classification algorithm. The learning based framework is highly adaptive and there are no parameters to tune. We show applications for edge detection in a number of specific image domains as well as on natural images. We test on various datasets including the Berkeley dataset and the results obtained are very good.

Goal

Edges reduce dimensionality of images while preserving information about image content. They can be useful for tasks such as object detection, structure from motion and tracking. *Our goal is to learn to detect edges from images with labeled ground truth.*





Motivations:

- •Make readily adaptable to many domains
- •Avoid explicitly modeling edges; tunable parameters
- •Combine low-level, mid-level and context information
- •Naturally integrate different sources of information

Problem Formulation

$I(i,j)$ image

W scene interpretation that can include spatial location and extent of objects, regions, object boundaries, curves, etc.

$S_W(i,j)$ 0/1 function that encodes spatial extent of component of W

Obtaining optimal or likely W or S_W can be difficult.

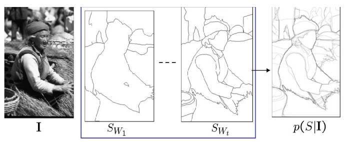
Let: $p(S(i,j)|I) = \sum_{W_t} S_{W_t}(i,j)p(W_t|I)$

We seek to learn this distribution directly from image data. To further reduce complexity, we can discard the absolute coordinates of S: $p(S(c)|I_{N(c)})$, where $N(c)$ is the neighborhood of I centered at c.

•edges from segmentations

W image segmentation

$S_W(i,j)$ 1 on boundaries of segments, 0 elsewhere



•road detection

W location of roads in scene

$S_W(i,j)$ 1 if pixel is on the road, 0 elsewhere

•object boundaries

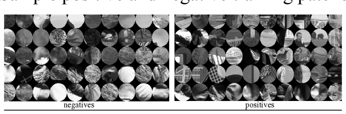
W location and extent of object of interest

$S_W(i,j)$ 1 on boundaries of object, 0 elsewhere

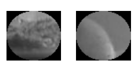
Goal is to learn $p(S(c)|I_{N(c)})$ from human labeled images. Given an image I and n interpretations W obtained by manual annotation, we:

1) Compute: $\hat{p}(S(i,j)|I) = \frac{1}{n}\sum_{W_t} S_{W_t}(i,j)$

2) Sample positive and negative training patches:



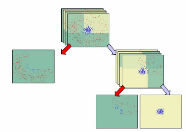
3) Learn: is edge point present in patch center?



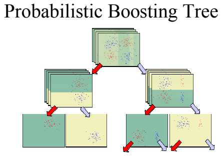
NO YES

Learning Architecture

Viola & Jones Cascade



Probabilistic Boosting Tree

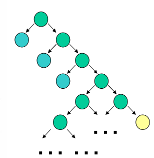


•Why PBT?

- •Highly varied data; large training set O(10^8)
- •Computational efficiency close to cascade (see below)
- •Adds 'power' when necessary (may overfit)
- •**PBT was necessary to obtain good results.**

•Training:

1. Given a set of images with edges annotated, retrieve a training set $S = \{(x_1, y_1, w_1), ..., (x_m, y_m, w_m); \quad x_i \in \chi, \quad y_i \in \{-1, +1\}, \sum_i w_i = 1$.
2. If the number (or weight) of either positive or negative samples in S is too small, perform bootstrapping to augment S (see below).
3. Compute the empirical distribution of S, $\hat{q}(y) = \sum_i w_i \delta(y_i = y)$. Continue if the depth of the node does not exceed some maximum value and $\theta \le \hat{q}(+1) \le (1-\theta)$, *e.g.* $\theta = 0.99$, else stop.
4. On training set S, train a strong boosted classifier (with a limited number of weak learners).
5. Split the data into two sets S_L and S_R using the decision boundary of the learned classifier and a tolerance ϵ. For each sample (x_i, y_i, w_i) compute $q(+1|x_i)$ and $q(-1|x_i)$, then:
   $(x_i, y_i, w_i * q(+1|x_i)) \to S_R$
   $(x_i, y_i, w_i * q(-1|x_i)) \to S_L$.
   Finally normalize all the weights in S_L and also S_R.
6. Train the left and right children recursively using S_L and S_R respectively (go to step 2).



•Computing probabilities:

If node:

$\tilde{p}(y|x) = \hat{q}(y)$

else:

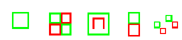
$\tilde{p}(y|x) = q(+1|x)\tilde{p}_R(y|x) + q(-1|x)\tilde{p}_L(y|x)$

where:

$\hat{q}(y)$ -empirical distr. at node

$q(y|x)$ -node classifier posterior

$\tilde{p}_L(y|x)$/$\tilde{p}_R(y|x)$ -recursive definition

•Features: 

Haar features (fast computation using integral images)
Applied to many 'views' of the data:
grayscale / color / Gabor filter outputs / optical flow / etc.

Summary

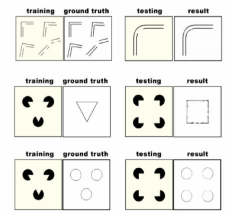
1) We have proposed a learning based algorithm for edge detection which implicitly combines low-level, mid-level and context information across different scales.

2) By learning from ground truth data, we avoided having to explicitly define and model edges.

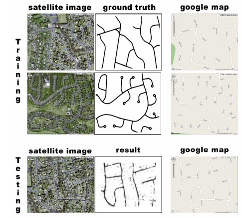
3) The resulting algorithm is highly adaptive and scalable, users need only give images with ground truth data for a given domain.

Results

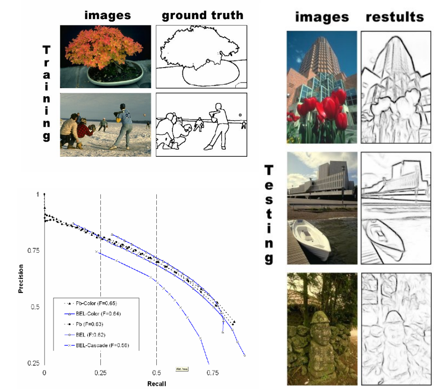
Gestalt factors



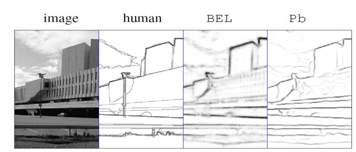
Road detection



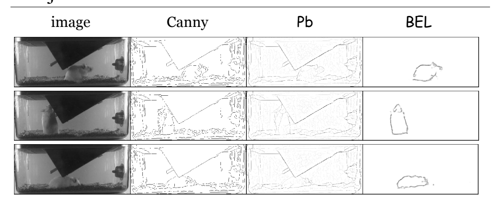
Natural images



- •Berkeley database, 100 training images, 200 testing
- •General edge detection (as opposed to domain specific)



Object boundaries



- •15 training images (not shown)
- •Note: Canny & Pb not designed for this task
- •Significant context information must be used (low level/r
- •Of direct potential use f
tracking, object recogniti