Maha ELBAYAD

A tree based context model for Object Recognition

Introduction

The probabilistic framework presented in [1] aims to exploit contextual information in addition to local features to detect and localize multiple object categories coexisting in an image.

The model

The model consists of two major components:

Prior model whose role is to capture dependencies between object categories. Learning this model breaks down to the following steps:

• Learning the dependency structure from co-occurrences of object pairs in a set of fully labeled images via Chow-liu's algorithm [2]. A node b_i in the tree is a binary variable indicating the presence of the object i in the image. From the tree structure, the joint probability of $b = (b_i)_i$ is given by:

$$\mathbb{P}(b) = \mathbb{P}(b_{root}) \prod_i \mathbb{P}(b_i|b_{\pi_i})$$

• Learning the location prior: each object's location in an image is encoded with 3 coordinates [3]:

$$(L_x, L_y, L_z) = (l_x, l_y, 1).\frac{H_i}{l_h} \qquad \text{(fig 1a)}$$

The final adopted location variable for object category i would be[1]:

$$L_i = (L_y, \log L_z)$$

Assuming $(L_y^{(i)})_i$, $(\log L_z^{(i)})_i$ are jointly Gaussians and that when $L = (L_i)_i$ is conditionned on the r.v b, it inherits the same dependency tree structure (figure 1b). Thus:

$$\mathbb{P}(L|b) = \mathbb{P}(L_{root}|b_{root}) \prod_i \mathbb{P}(L_i|L_{\pi_i}, b_i, b_{\pi_i})$$

Measurement model which encompasses the global gist descriptor [4] of the image plus the outputs of local detectors for each object category (i), that is a list of K_i candidates $(W_{ik}, s_{ik})_{k=1:K_i}$, $W_{ik} = (L_y, \log L_z)$ parametrizes the bounding box and s_{ik} is a detection score. Those predictions are then assessed on the training set yielding the binary variable $\forall i,k\ c_{ik} =$`is_correct_detection`.

Learning

We first estimate $\mathbb{P}(L_i|L_{\pi_i}, b_i, b_{\pi_i})$ as three gaussians: (1) the case $(b_i = 1, b_{\pi_i} = 1)$ as $L_i|L_{\pi_i}$. (2) the case $(b_i = 1, b_{\pi_i} = 0)$ as $L_i \perp\!\!\!\perp L_{\pi_i}$ and (3) the case $(b_i = 0)$ as $L_i \perp\!\!\!\perp L_j\ \forall j$ and set $L_i = \mathbb{E}_{images}(L_i)$

For the gist decriptor, we use logistic regression to fit $\mathbb{P}(b_i|g)$ and we handle the local detectors similarly to fit $\mathbb{P}(c_{ik}|s_{ik})$.

Alternating inference on trees: Now that we lerned our parameters g, s and W we solve for b, c and L as:

$$\hat{b}, \hat{c}, \hat{L} = \arg\max_{b,c,L} \mathbb{P}(b, c, L|g, s, W)$$

We infer the optimal values iteratively:
(a[2]): $\hat{b}, \hat{c} = \arg\max_{b,c} \mathbb{P}(b, c|g, s, W, \hat{L})$

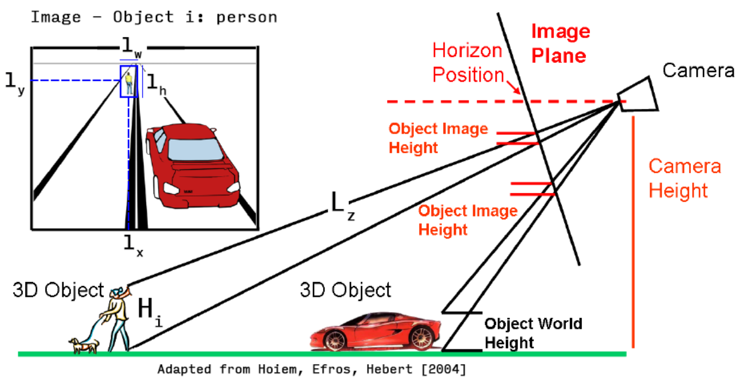
Preliminary results

Currently at the training phase on the $SUN-09$ database (111 categories, 4317 images). A subtree of the inferred dependency tree is shown in figure 2.

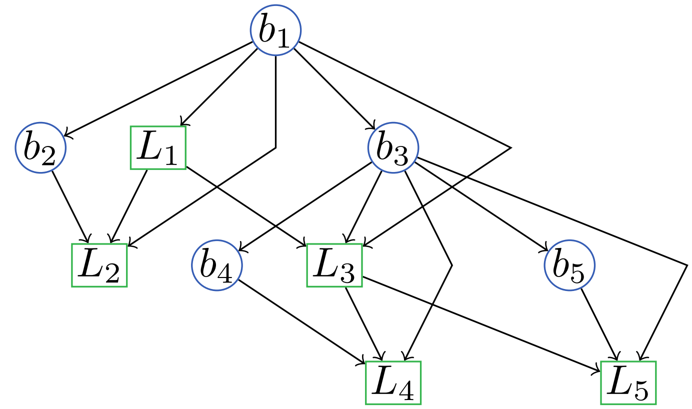
---

[1] L_x dropped given that horizontal locations have weak contextual information

[2] For the first iteration we set $\hat{b}, \hat{c} = \arg\max_{b,c} \mathbb{P}(b, c|g, s)$



(a) 3D world coordinates



(b) Prior model - Tree structure

Figure 1

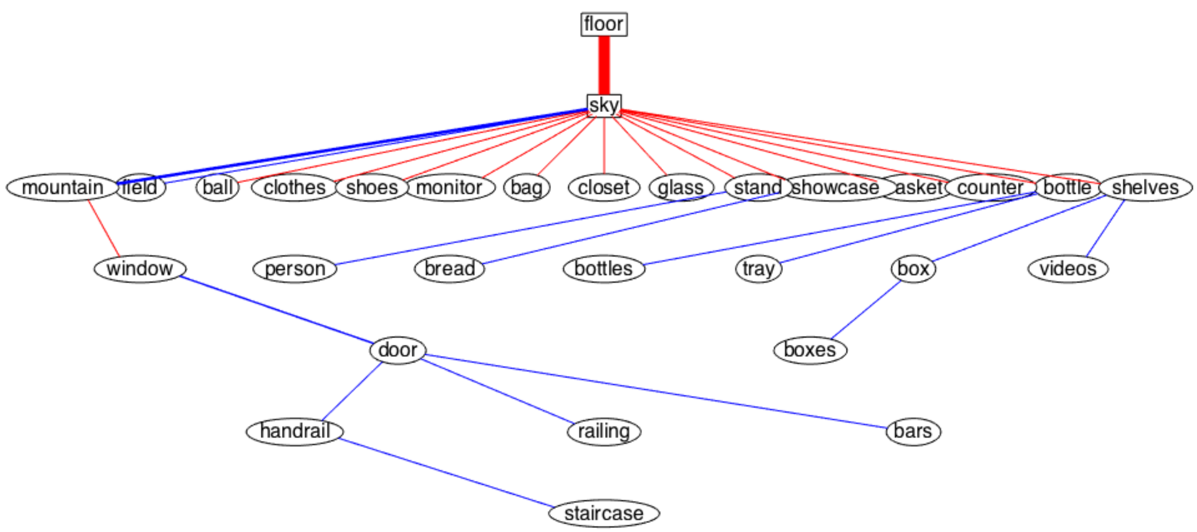


Figure 2: (Sky) subtree considering (Floor) as the root of the tree
Red edges:(-) correlation - Blue edges: (+) correlation
The line width reflects the probability of co-occuring

References

[1] Choi, M. J., Torralba, A., & Willsky, A. S. (2012). A tree-based context model for object recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34(2), 240-252.

[2] Chow, C. K., & Liu, C. N. (1968). Approximating discrete probability distributions with dependence trees. Information Theory, IEEE Transactions on, 14(3), 462-467.

[3] Hoiem, D., Efros, A. A., & Hebert, M. (2008). Putting objects in perspective. International Journal of Computer Vision, 80(1), 3-15.

[4] Torralba, A. (2003). Contextual priming for object detection. International journal of computer vision, 53(2), 169-191.