Problems:

1. Measure: The average error of position of leaf nodes.
2. In the pruning step of Bottom-Up process, $T_1, T_2, ..., T_5$ are not learned from training data. A phenomenon is that most ground truth (about 75%) can not pass the pruning step. If the pruning constrains are relaxed, the time cost will greatly added (one example, 2s$\sim$ 120s).
3. The perception algorithm is used to update parameters. In small training dataset (N=50), overfitting is very obvious. In large training dataset (N=150), the algorithm is slow to converge. After 3000 iterations (20 per image), the performance on training data improved slightly.
4. If the leaf node model is trained respectively, the result using 480D appearance features is similar to result only using magnitude of gradient. In my opinion, the reason is that perception algorithm could not find the features with distinguishing ability. From the distribution of features on ground truth and all candidate pixels, we find that some features (> 10) have similar distinguishing ability to gradient. Maybe there are too much useless features in feature set.

Definition:

1. $GP_\nu^{(t)}$: the ground truth of training image t on node ν
2. T_ν: the threshold of node ν for pruning. The threshold is set to keep the RECALL of node ν is larger than a predefined value.
3. $W_{Appearance}$: w_ν is belong to the class if ν is a LEAF node.
4. W_{Shape}: w_ν is belong to the class if ν is an AND node.
5. $E_{app}(d,Z) = \sum_{\nu \in V^{LEAF}} \sum_{i=0}^{n} w_{\nu,i} \times f_i(d_\nu)$. n is the number of appearance features. $f_i(\cdot)$ is the ith feature.
6. $E_h(z) = \sum_{\nu \in V^{AND}} \sum_{(\mu,\rho,\tau) \in T_\nu} w_{\nu,(\mu,\rho,\tau)} \times g(z_\mu, z_\rho, z_\tau)$
7. $E(z,d) = E_{app}(d,z) + E_h(z)$

Supervised Learning:
Object: Learn $w_{\nu,i}$ and $w_{\nu,(\mu,\rho,\tau)}$ from training data

1. InitializeWeight( )
   (a) $W_{Appearance} = 0$
   (b) $W_{Shape} = 1$
2. **SetThreshold**($model$, $\{GP^{(t)}\}$)
3. for $times$ = 1 : $MaxTimes$, for each image t in training set
   (a) $\{(MP_{root}, CL_{root})\}$ = **BottomUpSearch**($X^{(t)}, root$)
   (b) $MP^*_{root} = argmax_{MP_{root,a}} E_v(MP_{root,a})$
   (c) **UpdateParameters**(MP^*_{root})
   (d) **SetThreshold**($model$, $\{GP^{(t)}\}$)

$\{(MP_\nu, CL_\nu)\}$ = **BottomUpSearch**(X, ν)

1. for $\tau \in T_\nu$, BottomUpSearch(X,τ)
2. $(P_{\nu,a}, z_{\nu,a}) = \oplus_{\tau \in T_\nu} MP_{\tau,a}$
3. Pruning:
   (a) $E_\nu(P_{\nu,a}) > T_\nu$
   (b) $\forall(\mu,\rho,\tau) \in T_\nu, g(z_\mu, z_\rho, z_\tau) > T_1$
   (c) $\forall(\mu,\rho) \in T_\nu, OriDiff(\theta_\mu, \theta_\rho, z_\tau) < T_2$
   (d) $\forall(\mu,\rho) \in T_\nu, T_{3,1} < s_\mu/s_\rho < T_{3,2}$
   (e) $-T_4 < \theta_\nu < T_4$
   (f) $T_{5,1} < s_\nu < T_{5,2}$
4. Surround Suppression: $\{(MP_\nu, CL_\nu)\} = SurroundSuppression(\{P_{\nu,a}\}, \epsilon_W)$, where ϵ_W is the size of window W_ν defined in position, orientation and scale

UpdateParameters($MP_{\nu,a}$)

1. IF ν is a LEAF node
   (a) $w_{\nu,i} = w_{\nu,i} + f_i(GP_\nu) - f_i(MP_\nu)$
2. IF ν is an AND node
   (a) for $\tau \in T_\nu$, and $s.t. MP_{\tau,b} \in MP_{\nu,a}$
      i. UpdateParameters($MP_{\tau,b}$)
   (b) for $(\mu,\rho,\tau) \in T_\nu$
      i. $w_{\nu,(\mu,\rho,\tau)} = w_{\nu,(\mu,\rho,\tau)} + g_{GP_{\nu,a}}(z_\mu, z_\rho, z_\tau) - g_{MP_{\nu,a}}(z_\mu, z_\rho, z_\tau)$

SetThreshold($model$, $\{GP^{(t)}\}$)

1. for each node ν
   (a) for each image t, $E_\nu^{(t)} = E_\nu(GP_\nu^{(t)})$
   (b) $E_{\nu,1}, E_{\nu,2}, ..., E_{\nu,N}$ =sort $E_\nu^{(t)}$ in descending order, N is the number of training images
   (c) $index = N * \alpha\%$, α is RECALL which is predefined
   (d) $T_\nu = E_{\nu,index}$

Figure 1: pseudo-code