Inference by Learning: Speeding-up Graphical Model Optimization via a Coarse-to-Fine Cascade of Pruning Classifiers

Bruno Conejo[a,b], **Nikos Komodakis**[b], **Sebastien Leprince**[a], **Jean Philippe Avouac**[a]:(a) GPS, Caltech, USA; (b) Ecole des Ponts ParisTech, France

Associated materials and code: http://imagine.enpc.fr/~conejob/ibyl/ Contact: bconejo@caltech.edu

Context:

- MAP inference for MRF/CRF models.

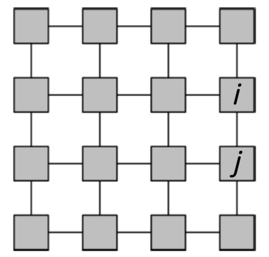
Contributions:

- Novel MAP-estimation framework that uses **learning** to significantly **speed-up** MRF/CRF inference.
- Maintains/improves accuracy of MAP estimation.

Notations

To represent a discrete MRF model $\mathcal{M}$, we use the following notation:

$$\mathcal{M} = (\mathcal{V}, \mathcal{E}, \mathcal{L}, \{\phi_i\}_{i\in\mathcal{V}}, \{\phi_{ij}\}_{(i,j)\in\mathcal{E}}) .$$



A graph $\mathcal{G} = [\mathcal{V}, \mathcal{E}]$

$x_i \in \mathcal{L}$: the configuration of vertex i

$\phi_i \in \mathcal{L} \to \mathbb{R}$: the unary potential of vertex i

$\phi_{ij} \in \mathcal{L} \times \mathcal{L} \to \mathbb{R}$: the pairwise potential of edge ij

$x = (x_i)_{i\in\mathcal{V}}$: the configuration of $\mathcal{M}$.

The energy (the total cost) of a solution x is:

$$E(x|\mathcal{M}) = \sum_{i\in\mathcal{V}} \phi_i(x_i) + \sum_{(i,j)\in\mathcal{E}} \phi_{ij}(x_i, x_j) .$$

The goal of MAP estimation is to find a solution that has minimum energy:

$$x_{\text{MAP}} = \arg\min_{x\in\mathcal{L}^{|\mathcal{V}|}} E(x|\mathcal{M}) .$$

The pruning matrix $A : \mathcal{V} \times \mathcal{L} \to \{0, 1\}$:

$$A(i,l) = \begin{cases} 1 & \text{if label } l \text{ is active at vertex } i \\ 0 & \text{if label } l \text{ is pruned at vertex } i \end{cases}$$

and its associated solution space:

$$\mathcal{S}(\mathcal{M}, A) = \left\{ x \in \mathcal{L}^{|\mathcal{V}|} \mid (\forall i), A(i, x_i) = 1 \right\} .$$

Approach

To obtain an inference speed-up we need to solve:

$$\min_{x\in\mathcal{S}(\mathcal{M},A)} E(x|\mathcal{M}) .$$

with A such that:

(1) $x_{\text{MAP}} \in \mathcal{S}(\mathcal{M}, A)$

(2) Most elements of A are 0 (i.e., the labels are pruned)

The *IbyL* framework iteratively estimates A by:

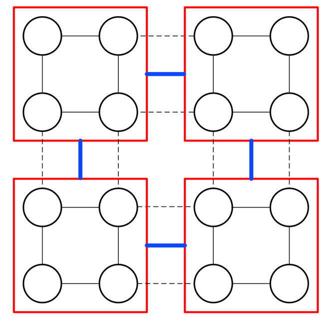
(1) Building a coarse to fine set of MRFs from $\mathcal{M}$.

(2) Learning at each scale pruning classifiers to refine A

Model Coarsening

Given a model $\mathcal{M}$ and a grouping function $g : \mathcal{V} \to \mathbb{N}$, we create a "coarser" version of this model:

$$\mathcal{M}' = \left(\mathcal{V}', \mathcal{E}', \mathcal{L}, \{\phi'_i\}_{i\in\mathcal{V}'}, \{\phi'_{ij}\}_{(i,j)\in\mathcal{E}'}\right)$$



Circles: $\mathcal{V}$, Lines: $\mathcal{E}$, Squares: $\mathcal{V}'$, Lines: $\mathcal{E}'$.

The vertice and edges of $\mathcal{M}'$ are given by:

$$\mathcal{V}' = \{i' \mid \exists i \in \mathcal{V}, i' = g(i)\} ,$$
$$\mathcal{E}' = \{(i', j') \mid \exists (i,j) \in \mathcal{E}, i' = g(i), j' = g(j), i' \neq j'\} .$$

The unary and pairwise potentials of $\mathcal{M}'$ are given by

$$(\forall i' \in \mathcal{V}'), \quad \phi'_{i'}(l) = \textstyle\sum_{i\in\mathcal{V}|i'=g(i)} \phi_i(l) + \sum_{(i,j)\in\mathcal{E}|i'=g(i)=g(j)} \phi_{ij}(l,l),$$
$$(\forall (i',j') \in \mathcal{E}'), \ \phi'_{i'j'}(l_0,l_1) = \textstyle\sum_{(i,j)\in\mathcal{E}|i'=g(i),j'=g(j)} \phi_{ij}(l_0,l_1) .$$

With a slight abuse of notation, we define the coarsening as:

$$g(\mathcal{M}) = \mathcal{M}'$$

and the up-sampling of a solution of $\mathcal{M}' \to \mathcal{M}$ as:

$$g^{-1}(x') = x$$

Coarse to fine optimization and label pruning

We iteratively apply the previous coarsening to build a coarse to fine set of N+1 progressively coarser MRFs:

$$\mathcal{M}^{(0)} = \mathcal{M} \quad \text{and} \quad (\forall s),\ \mathcal{M}^{(s+1)} = g^{(s)}(\mathcal{M}^{(s)}) \quad \text{with} \quad \mathcal{M}^{(s)} = \left(\mathcal{V}^{(s)}, \mathcal{E}^{(s)}, \mathcal{L}, \{\phi_i^{(s)}\}_{i\in\mathcal{V}^{(s)}}, \{\phi_{ij}^{(s)}\}_{(i,j)\in\mathcal{E}^{(s)}}\right)$$

We set all elements of the coarsest pruning matrix A^N to 1 (no pruning).

At each scale s, we apply the following steps:

Step 1: Optimize the current MRF $\mathcal{M}^{(s)}$:

$$x^{(s)} \approx \arg\min_{x\in\mathcal{S}(\mathcal{M}^{(s)},A^{(s)})} E(x|\mathcal{M}^{(s)})$$

Step 2: Update the next scale pruning matrix $A^{(s-1)}$,

(i):Compute the feature map:

$$z^{(s)} : \mathcal{V}^{(s)} \times \mathcal{L} \to \mathbb{R}^K$$

(ii):Update the pruning matrix from off-line trained classifier $f^{(s)} : \mathbb{R}^K \to \{0,1\}$:

$$(\forall i \in \mathcal{V}^{(s-1)},\ \forall l \in \mathcal{L}), \quad A^{(s-1)}(i,l) = f^{(s)}(z^{(s)}(g^{(s-1)}(i), l)) .$$

Step 3: Up-sample the current solution for next scale:

$$x^{(s-1)} = [g^{(s-1)}]^{-1}(x^{(s)})$$

Inference by learning framework

Data: Model $\mathcal{M}^{(0)} = \mathcal{M}$, grouping functions $(g^{(s)})_{0\leq s<N}$, classifiers $(f^{(s)})_{0<s\leq N}$

Result: $x^{(0)}$

Compute the coarse to fine sequence of MRFs:

for $s = [0 \ldots N-1]$ **do**

  $\mathcal{M}^{(s+1)} \leftarrow g^{(s)}(\mathcal{M}^{(s)})$

Optimize the coarse to fine sequence of MRFs over pruned solution spaces:

Initialization: set $x^{(N)}$ and $(\forall i \in \mathcal{V}^{(N)}, \forall l \in \mathcal{L}),\ A^{(N)}(i,l) \leftarrow 1$

for $s = [N...0]$ **do**

  Update $x^{(s)}$ by iterative minimization: $x^{(s)} \approx \arg\min_{x\in\mathcal{S}(\mathcal{M}^{(s)},A^{(s)})} E(x|\mathcal{M}^{(s)})$

  if $s \neq 0$ **then**

    Compute feature map $z^{(s)}$

    Update pruning matrix for next finer scale:

    $A^{(s-1)}(i,l) = f^{(s)}(z^{(s)}(g^{(s-1)}(i), l))$

    Upsample $x^{(s)}$ for initializing solution $x^{(s-1)}$ at next scale:

    $x^{(s-1)} \leftarrow [g^{(s-1)}]^{-1}(x^{(s)})$

Feature map

The feature map $z^{(s)}$ is formed of K stacked features defined on $\mathcal{V}^{(s)} \times \mathcal{L}$:

Presence of strong discontinuity:

$$\text{PSD}^{(s)}(i,l) = \begin{cases} 1 & \exists (i,j) \in \mathcal{E}^{(s)} |\ \phi_{ij}(x_i^{(s)}, x_j^{(s)}) > \rho \\ 0 & \text{otherwise} \end{cases}$$

Local energy variation:

$$\text{LEV}^{(s)}(i,l) = \frac{\phi_i^{(s)}(l) - \phi_i^{(s)}(x_i^{(s)})}{N_{\mathcal{V}}^{(s)}(i)} + \sum_{j:(i,j)\in\mathcal{E}^{(s)}} \frac{\phi_{ij}^{(s)}(l, x_j^{(s)}) - \phi_{ij}^{(s)}(x_i^{(s)}, x_j^{(s)})}{N_{\mathcal{E}}^{(s)}(i)}$$

with $N_{\mathcal{V}}^{(s)}(i) = \text{card}\{i' \in \mathcal{V}^{(s-1)} : g^{(s-1)}(i') = i\}$

and $N_{\mathcal{E}}^{(s)}(i) = \text{card}\{(i',j') \in \mathcal{E}^{(s-1)} : g^{(s-1)}(i') = i, g^{(s-1)}(j') = j\}$.

Unary "coarsening":

$$\text{UC}^{(s)}(i,l) = \sum_{i'\in\mathcal{V}^{(s-1)}|g^{(s-1)}(i')=i} \frac{\left|\phi_{i'}^{(s-1)}(l) - \phi_i^{(s)}(l)/N_{\mathcal{V}}^{(s)}(i)\right|}{N_{\mathcal{V}}^{(s)}(i)}$$

Computing the ground truth

On a training set of MRFs, we run the IbyL framework without any pruning, i.e., $A^{(s)} \equiv 1$.

We keep track of features and compute:

$$X_{\text{MAP}}^{(s)} : \mathcal{V}^{(s)} \times \mathcal{L} \to \{0,1\}$$

such that:

$$X_{\text{MAP}}^{(0)}(i,l) = \begin{cases} 1, & \text{if } l \text{ is the ground truth label for node } i \\ 0, & \text{otherwise} \end{cases}$$

$$X_{\text{MAP}}^{(s+1)}(i,l) = \bigvee_{i'\in\mathcal{V}^{(s)}:g^{(s+1)}(i')=i} X_{\text{MAP}}^{(s)}(i',l)$$

where $\bigvee$ denotes the binary OR operator.

Learning pruning classifier

At each scale s, we split the features in two groups w.r.t the PSD feature. For each group:

(1) We have two classes defined from $X_{\text{MAP}}^{(s)}$:

c_0: to prune;

c_1: to remain active.

(2) We also introduce weights for each class:

weight of c_0: 1

weight of c_1: $\lambda \frac{\text{card}(c_0)}{\text{card}(c_1)}$

(3) We train a linear C-SVM classifier.

$\lambda \in \mathbb{R}^+$ is the pruning aggressiveness factor and controls how much pruning happens. During testing, the classifier $f^{(s)}$ applies the trained classifier of the corresponding group (w.r.t. the PSD feature).

Experiments: Setup

We experiment with two problems:

(1) Stereo-matching.

(2) Optical flow.

We evaluate different pruning aggressiveness factor λ, and compute:

(1) The speed-up w.r.t. the direct optimization.

(2) The ratio of active labels.

(3) The energy ratio w.r.t. the direct optimization.

(4 The MAP agreement w.r.t. the best computed solution.

As an optimization subroutine we use Fast-PD.

For all experiments we use five scales and learn the pruning classifiers from only one MRF (Tsukuba for stereo-matching and Army for Optical flow).

Stereo-matching

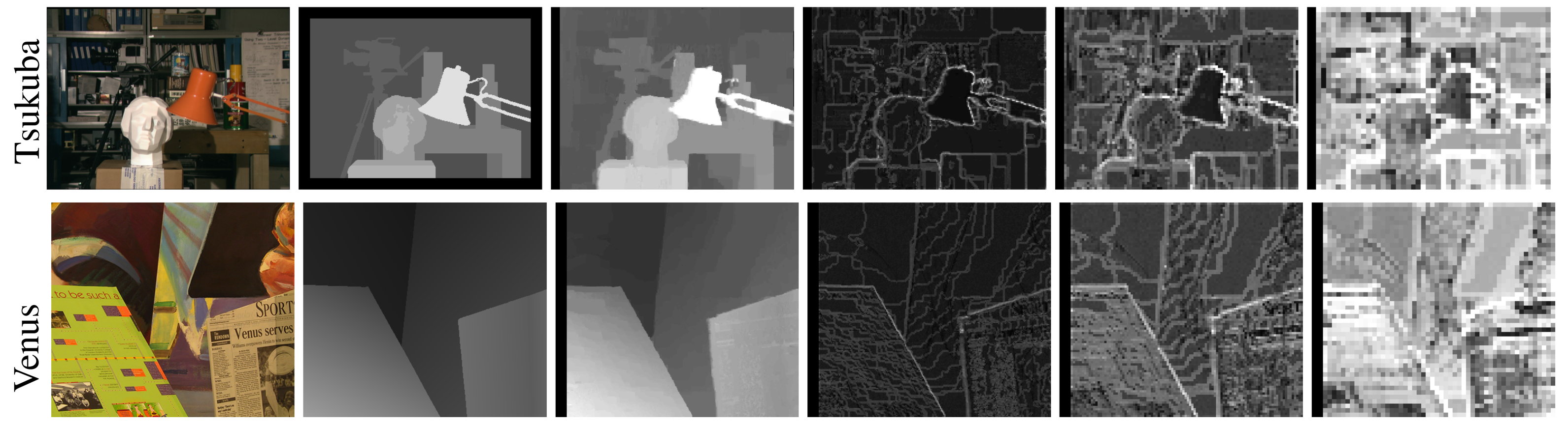
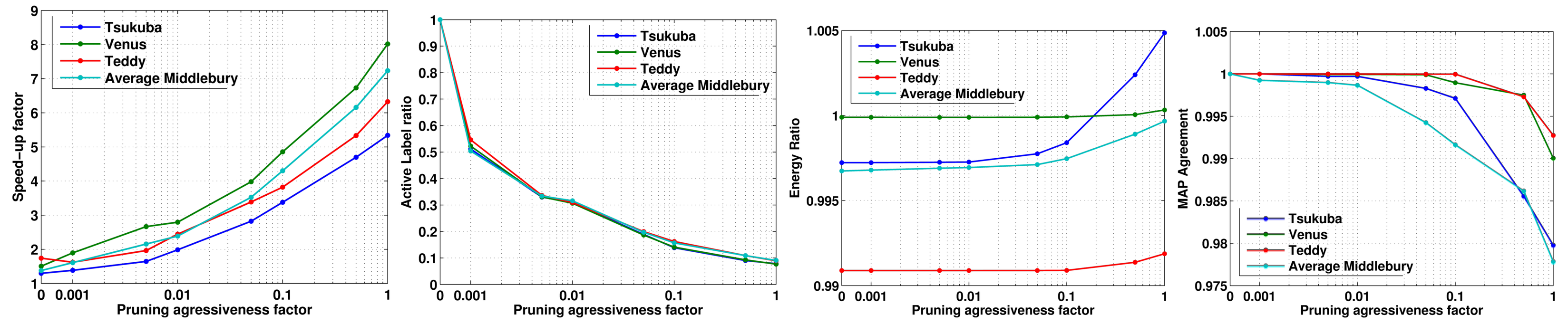


Image GT MAP A scale 0 A scale 1 A scale 2



Speed-up Active label ratio Energy ratio Label agreement

Optical flow

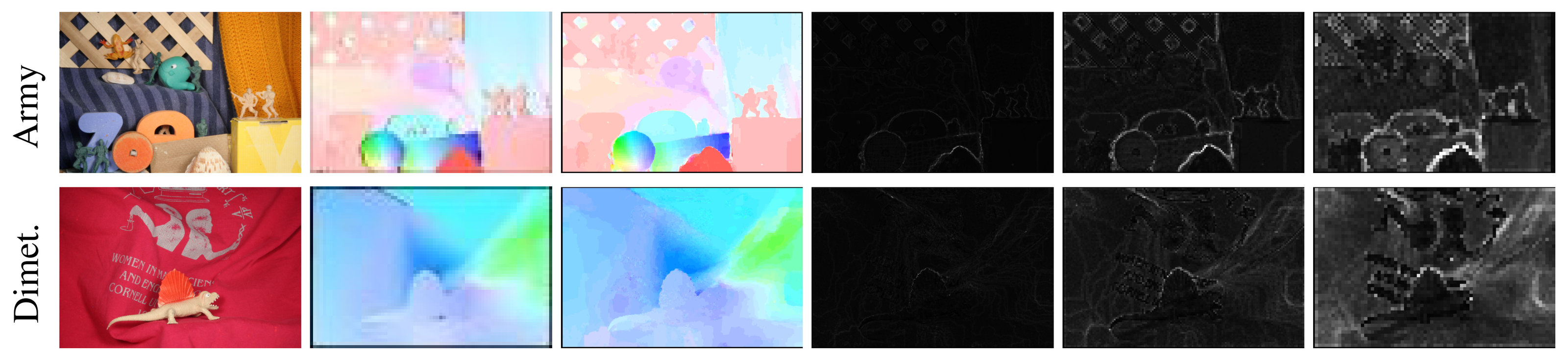
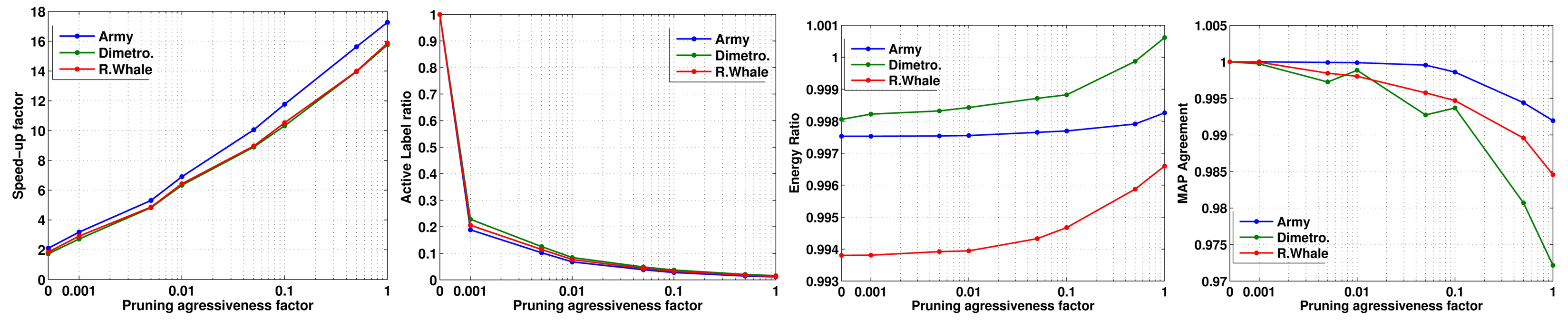


Image GT MAP A scale 0 A scale 1 A scale 2



Speed-up Active label ratio Energy ratio Label agreement

Conclusion

The IbyL framework:

(1) Gives an important speedup.

(2) Maintains excellent accuracy of the solution.

(3) Can be easily adapted by computing task dependent features.

(4) Can be easily adapted to high order MRFs.

(5) Is available to download at:

http://imagine.enpc.fr/~conejob/ibyl/

Acknowledgment

This work was supported by USGS through the Measurements of surface ruptures produced by continental earthquakes from optical imagery and LiDAR project (USGS Award G13AP00037), the Terrestrial Hazard Observation and Reporting Center of Caltech, and the Moore foundation through the Advanced Earth Surface Observation Project (AESOP Grant 2808).