Efficient Inference Algorithms for Scene Understanding Problems

Karteek Alahari
Inria Grenoble – Rhône-Alpes

Joint work with Anoop Cherian, Pushmeet Kohli, Lubor Ladicky, Julien Mairal, Chris Russell, Cordelia Schmid, Paul Sturgess, Philip Torr
What is scene understanding?
Identify scene type and the elements in the scene

Road
Car
Pedestrian
Building
Sky
...

Bosch et al., ECCV’06
Deselaers et al., PR’10, DAGM’06
Localize the elements

Building

Car

Pedestrian

Dalal & Triggs, CVPR’05
Felzenszwalb et al., PAMI’10
Girshick et al., CVPR’14
Segment the elements

Sky

Building

Tree

Car

Sidewalk

Person

Road

Gould et al. NIPS’09
Ladicky et al. ECCV’10
Yao et al., CVPR’12
...
Not all images are the same!
Analyze people in a scene

Inria 3DMovie dataset
Analyze people in a scene

Estimate poses (body-joint locations)

Inria 3DMovie dataset

Yang & Ramanan, PAMI’12
Tompson et al., CVPR’14
Analyze people in a scene

Segment the individuals

Yang et al., PAMI’11
Shotton et al., CVPR’11

Inria 3DMovie dataset
Scene Understanding

Markov Random Fields [Preston '74]
Scene Understanding

Image data nodes

Markov Random Fields [Preston ’74]
Scene Understanding

- Image data nodes
- Random variables

Markov Random Fields [Preston ’74]
Scene Understanding

Markov Random Fields [Preston ’74]

\[ \psi_i(y_i) : \text{Unary cost} \]
e.g., object recognition
Scene Understanding

$\psi_i(y_i)$: Unary cost

- Head
- Shoulder
- Elbow
- Wrist

... e.g., pose estimation

Markov Random Fields [Preston '74]
Scene Understanding

Markov Random Fields [Preston ’74]
Scene Understanding

\[ E(y) = \sum_{i \in V} \psi_i(y_i) + \sum_{(i,j) \in E} \psi_{ij}(y_i, y_j) : \text{Cost of labelling} \]
Scene Understanding

• In the context of such labelling problems
  – Learning parameters of the function
  – Modelling image priors & temporal constraints
  – Performing inference
Scene Understanding

• In the context of such labelling problems
  – Learning parameters of the function
  – Modelling image priors & temporal constraints
  – Performing inference
Parameter learning

• Consider a 2-label problem (for now)

• e.g., binary image segmentation

• Also, consider a pairwise random field
Parameter learning

- Conditional probability of a labelling $\mathbf{y}$:

$$
\Pr(\mathbf{y}|\mathbf{x}, \theta) = \frac{1}{Z(\theta)} \prod_{i \in V} \exp(y_i \theta_u^T h_i(\mathbf{x})) \prod_{(i,j) \in E} \exp(y_i y_j \theta_p^T \nu_{ij}(\mathbf{x}))
$$
Parameter learning

• Conditional probability of a labelling $\mathbf{y}$:

$$
\Pr(\mathbf{y} | \mathbf{x}, \mathbf{\theta}) = \frac{1}{Z(\mathbf{\theta})} \prod_{i \in \mathcal{V}} \exp(y_i \mathbf{\theta}_u^T h_i(\mathbf{x})) \prod_{(i,j) \in \mathcal{E}} \exp(y_i y_j \mathbf{\theta}_p^T \nu_{ij}(\mathbf{x}))
$$

The label taken by pixel $x_i = -1$ or $1$
Parameter learning

- Conditional probability of a labelling $\mathbf{y}$:

$$
\Pr(\mathbf{y}|\mathbf{x}, \mathbf{\theta}) = \frac{1}{Z(\mathbf{\theta})} \prod_{i \in \mathcal{V}} \exp(y_i \mathbf{\theta}_u^\top h_i(\mathbf{x})) \prod_{(i,j) \in \mathcal{E}} \exp(y_i y_j \mathbf{\theta}_p^\top \nu_{ij}(\mathbf{x}))
$$

Unary features  
Pairwise features
Parameter learning

- Conditional probability of a labelling $\mathbf{y}$:

$$
Pr(y|x, \theta) = \frac{1}{Z(\theta)} \prod_{i \in \mathcal{V}} \exp(y_i \theta_u^T h_i(x)) \prod_{(i,j) \in \mathcal{E}} \exp(y_i y_j \theta_p^T \nu_{ij}(x))
$$

Unary parameters

Pairwise parameters
Learning Approaches

• Maximum likelihood-based

• Large margin based

• Other iterative methods
Maximum likelihood-based

• Compute $\hat{\theta}$ which maximizes likelihood

$$\hat{\theta} = \arg \max_{\theta} \Pr(y|x, \theta)$$

• Hard to solve for vision problems
Maximum likelihood-based

• Why is it hard?

\[ \Pr(y|x, \theta) = \frac{1}{Z(\theta)} \prod_{i \in V} \exp(y_i \theta^T h_i(x)) \prod_{(i,j) \in E} \exp(y_i y_j \theta_p^T \nu_{ij}(x)) \]

Partition function – Infeasible to compute for vision problems \((2^N \text{ values})\)

• Approximate methods: Sampling, mode of the distribution, pseudo-likelihood

e.g., Kumar and Hebert, NIPS’03
Pseudo-likelihood

\[ \Pr(y|x, \theta) = \prod_{i \in V} \Pr(y_i|x, y_j, \theta) \]

Besag, The Statistician, 1975
Pseudo-likelihood

\[
Pr(y|x, \theta) = \prod_{\substack{i \in V \\
(i,j) \in E}} \frac{1}{z_i(\theta)} \exp(y_i \theta_u^T h_i(x)) \cdots
\]

✓ Easy to compute

✗ But, can lead to poor accuracy
Max-Margin Learning

• Consider

$$\log \Pr(y|x, \theta) = \theta X y - \log Z(\theta)$$

• Maximize confidence margin in true label assignment

• In other words, maximize

$$\log \Pr(\hat{y}|x, \theta) - \log \Pr(y|x, \theta) = \theta X (\hat{y} - y)$$

Taskar et al. NIPS’04, Tsochantaridis et al. ICML’04
Max-Margin Learning

• Formulated as SVM learning problem

\[
\min \frac{1}{2} \| \theta \|^2 + C \xi \\
\text{subject to} \quad \theta \mathbf{F} (\hat{y} - y) \geq N - \hat{y}_i^T y_i - \xi, \forall y \in \mathcal{Y}
\]
Max-Margin Learning

• Formulated as SVM learning problem

\[
\min \frac{1}{2} \|\theta\|^2 + C\xi
\]

subject to \( \theta F(\hat{y} - y) \geq N - \hat{y}_i^T y_i - \xi, \forall y \in \mathcal{Y} \)

Exponential no. of constraints!!
\((2^N - 1)\)
Max-Margin Learning

• Formulated as SVM learning problem

\[
\min \quad \frac{1}{2} \| \theta \|^2 + C \xi \\
\text{subject to} \quad \theta F(\hat{y} - y) \geq N \hat{y}_i^T y_i \xi, \forall y \in \mathcal{Y}
\]

\[
\text{subject to} \quad \theta F \hat{y} - N + \xi \geq \max_{y \in \mathcal{Y}} \theta F y - \hat{y}_i^T y_i
\]

Inference problem to find the most violated constraint
e.g., graph cuts
Max-Margin Learning

• Formulated as SVM learning problem

$$\min \frac{1}{2} \| \theta \|^2 + C \xi$$

subject to  $$\theta F\hat{y} - N + \xi \geq \max_{y \in Y} \theta Fy - \hat{y}_i^T y_i$$

✓ Eliminates the partition function
× Limited by the inference step
Our Piecewise Model

Training Exemplar 1
Our Piecewise Model

Training Exemplar 2
Our Piecewise Model

Training Exemplar 3
Our Piecewise Model

• Formulate a max-margin problem on these sub-graphs
Our Piecewise Model

- Formulate a max-margin problem on these sub-graphs

\[
\min \quad \frac{1}{2} \| \theta \|^2 + C \xi
\]

subject to \( \theta F_s(\hat{y}_s - y_s) \geq N - \hat{y}_s^T y_{si} - \xi, \forall y_s \in Y_s \)
Our Piecewise Model

- Formulate a max-margin problem on these sub-graphs

\[
\min \frac{1}{2} \|\theta\|^2 + C\xi
\]

subject to \( \theta F_s(\hat{y}_s - y_s) \geq N - \hat{y}_s^T y_i - \xi, \forall y \in Y_s \)

Slack variable
Our Piecewise Model

• Formulate a max-margin problem on these sub-graphs

$$\min \frac{1}{2} \| \theta \|^2 + C \xi_s$$

subject to

$$\theta F_s(\hat{y}_s - y_s) \geq N - \hat{y}_s^T y_i - \xi_s, \forall y_s \in \mathcal{Y}_s$$

Exponential # constraints again!
Our Piecewise Model

• Formulate a max-margin problem on these sub-graphs

\[
\min \quad \frac{1}{2} ||\theta||^2 + C \xi_s \\
\text{subject to} \quad \theta F_s(\hat{y}_s - y_s) \geq N - \hat{y}_s^T y_i - \xi, \forall y \in \mathcal{Y}_s
\]

• Can be solved efficiently using BP messages
Piecewise: Results

• Middlebury Stereo

<table>
<thead>
<tr>
<th>Method</th>
<th>Art</th>
<th>Books</th>
<th>Dolls</th>
<th>Laundry</th>
<th>Moebius</th>
<th>Reindeer</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li and Huttenlocher</td>
<td>14.66</td>
<td>19.12</td>
<td>12.70</td>
<td>19.16</td>
<td>10.88</td>
<td>11.72</td>
<td>14.71</td>
</tr>
<tr>
<td>Our method</td>
<td>12.94</td>
<td>16.24</td>
<td>12.21</td>
<td>16.72</td>
<td>10.82</td>
<td>11.10</td>
<td>13.34</td>
</tr>
</tbody>
</table>

• Man-made Structures

<table>
<thead>
<tr>
<th>Method</th>
<th>FP per image</th>
<th>DR %</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRF</td>
<td>2.36</td>
<td>57.20</td>
</tr>
<tr>
<td>DRF1</td>
<td>1.37</td>
<td>70.50</td>
</tr>
<tr>
<td>DRF2</td>
<td>1.76</td>
<td>72.54</td>
</tr>
<tr>
<td>Our method</td>
<td>1.40</td>
<td>72.60</td>
</tr>
</tbody>
</table>
So far ...

• In the context of labelling problems
  – Learning parameters of the function

Next:
  – Modelling image priors & performing inference
Modelling image priors & inference

• Energy function: $E(y) = \sum_{i \in V} \psi_i(y_i) + \sum_{(i,j) \in E} \psi_{ij}(y_i, y_j)$

• How to minimize this efficiently?

• New image priors, e.g., higher-order term: $\sum_{c \in S} \psi_c(y_c)$
Example: Image Segmentation

\[ E(y) = \sum_{i \in V} \psi_i(y_i) + \sum_{(i,j) \in E} \psi_{ij}(y_i, y_j) \]

\[ E: \{0, 1\}^N \to \mathbb{R} \]
\[ 0 \rightarrow fg \]
\[ 1 \rightarrow bg \]

N: number of pixels

How to minimize \( E(y) \)?

st-mincut algorithm, belief propagation, \( \alpha \)-expansion
The st-Mincut Problem

Graph \((V, E, C)\)

Vertices \(V = \{v_1, v_2 \ldots v_n\}\)
Edges \(E = \{(v_1, v_2) \ldots \}\)
Costs \(C = \{c_{(1, 2)} \ldots \}\)
The st-Mincut Problem

What is an st-cut?

An st-cut \((S,T)\) divides the nodes between source and sink.

What is the cost of an st-cut?

Sum of cost of all edges going from \(S\) to \(T\)

\[
5 + 2 + 9 = 16
\]

Slide courtesy: Pushmeet Kohli
The st-Mincut Problem

What is an st-cut?
An st-cut \((S,T)\) divides the nodes between source and sink.

What is the cost of an st-cut?
Sum of cost of all edges going from \(S\) to \(T\)

What is the st-mincut?
st-cut with the minimum cost

\[ 2 + 1 + 4 = 7 \]

Slide courtesy: Pushmeet Kohli
How to compute the st-mincut?

Solve the dual **maximum flow** problem

Compute the maximum flow between Source and Sink

**Constraints**
- Edges: Flow < Capacity
- Nodes: Flow in = Flow out

**Min-cut/Maxflow Theorem**
In every network, the maximum flow equals the cost of the st-mincut

Slide courtesy: Pushmeet Kohli

Dantzig, Dinitz, Edmonds & Karp, Karzanov, Ford & Fulkerson, Goldberg & Tarjan, ...
Maxflow Algorithms

Augmenting Path Based Algorithms

1. Find path from source to sink with positive capacity
2. Push maximum possible flow through this path
3. Repeat until no path can be found

Algorithms assume non-negative capacity

Source

\[ v_1 \]

\[ v_2 \]

Sink

Flow = 0
Maxflow Algorithms

Flow = 0

Source

\[ v_1 \]

\[ v_2 \]

Sink

Augmenting Path Based Algorithms

1. Find path from source to sink with positive capacity
2. Push maximum possible flow through this path
3. Repeat until no path can be found

Algorithms assume non-negative capacity

Slide courtesy: Pushmeet Kohli
Maxflow Algorithms

Augmenting Path Based Algorithms

1. Find path from source to sink with positive capacity
2. Push maximum possible flow through this path
3. Repeat until no path can be found

Algorithms assume non-negative capacity

Flow = 0 + 2

Slide courtesy: Pushmeet Kohli
Maxflow Algorithms

Flow = 2

1. Find path from source to sink with positive capacity
2. Push maximum possible flow through this path
3. Repeat until no path can be found

Algorithms assume non-negative capacity

Slide courtesy: Pushmeet Kohli
Maxflow Algorithms

Augmenting Path Based Algorithms

1. Find path from source to sink with positive capacity
2. Push maximum possible flow through this path
3. Repeat until no path can be found

Algorithms assume non-negative capacity

Flow = 7

Source

$\mathbf{v}_1$

$\mathbf{v}_2$

Sink

Slide courtesy: Pushmeet Kohli
Maxflow Algorithms

Augmenting Path Based Algorithms

1. Find path from source to sink with positive capacity
2. Push maximum possible flow through this path
3. Repeat until no path can be found

Algorithms assume non-negative capacity

Slide courtesy: Pushmeet Kohli
Maxflow in Computer Vision

• Specialized algorithms for vision problems
  – Grid graphs
  – Low connectivity ($m \sim O(n)$)

• Dual search tree augmenting path algorithm [Boykov & Kolmogorov PAMI’04]
  – Finds approximate shortest augmenting paths efficiently
  – High worst-case time complexity
  – Empirically outperforms other algorithms on vision problems

Slide courtesy: Pushmeet Kohli
st-mincut  ⇔  Labelling

Construct a graph such that:

1. Any st-cut corresponds to an assignment of $y$
2. The cost of the cut is equal to the energy of $y$ : $E(y)$

Solution

Slide courtesy: Pushmeet Kohli
St-mincut based Move algorithms

\[ E(y) = \sum_{i \in \mathcal{V}} \psi_i(y_i) + \sum_{(i,j) \in \mathcal{E}} \psi_{ij}(y_i, y_j) \]

\[ y \in \text{Labels } \mathcal{L} = \{l_1, l_2, \ldots, l_k\} \]

- Commonly used for solving multi-label problems
- Extremely efficient and produce good solutions
- Not Exact: Produce local optima

Slide courtesy: Pushmeet Kohli
Move Making Algorithms

- Current Solution
- Search Neighbourhood
- Optimal Move

Energy

Solution Space

Slide courtesy: Pushmeet Kohli
Move Making Algorithms

Slide courtesy: Pushmeet Kohli
Move Making Algorithms

Slide courtesy: Pushmeet Kohli
Computing the Optimal Move

Key Property

Move Space

Bigger move space

- Better solutions
- Finding the optimal move hard

Slide courtesy: Pushmeet Kohli
Moves using Graph Cuts

Expansion and Swap move algorithms [Boykov et al., PAMI’01]

• Make a series of changes to the solution (moves)
• Each move results in a solution with smaller energy

Space of Solutions: \( L^N \)
Move Space \((t) : 2^N \)

Current solution
Search neighbourhood

\( N \) Number of variables
\( L \) Number of labels

How to minimize move functions?

Slide courtesy: Pushmeet Kohli
Expansion Move

- Variables take label $\alpha$ or retain current label

Status:

Slide courtesy: Pushmeet Kohli  
Boykov, Veksler, Zabih, PAMI’01
Expansion Move

- Variables take label $\alpha$ or retain current label

Status: Initialize with Tree

Slide courtesy: Pushmeet Kohli

Boykov, Veksler, Zabih, PAMI’01
Expansion Move

- Variables take label $\alpha$ or retain current label

Status: Expand Ground

Slide courtesy: Pushmeet Kohli

Boykov, Veksler, Zabih, PAMI’01
Expansion Move

- Variables take label $\alpha$ or retain current label

Status: Expand House

Boykov, Veksler, Zabih, PAMI’01

Slide courtesy: Pushmeet Kohli
Expansion Move

- Variables take label $\alpha$ or retain current label

Status: Expand Sky

Slide courtesy: Pushmeet Kohli

Boykov, Veksler, Zabih, PAMI’01
Efficient Inference: Recycling

Iteration 1

1-expansion

2-expansion

\[ G_1^1(\mathcal{V}_1^1, \mathcal{E}_1^1) \]

\[ G_2^1(\cdot, \cdot) \]

\[ G_k^1(\mathcal{V}_k^1, \mathcal{E}_k^1) \]

Iteration 2

1-expansion

2-expansion

\[ G_1^2(\mathcal{V}_1^2, \mathcal{E}_1^2) \]

\[ G_2^2(\cdot, \cdot) \]

\[ G_k^2(\mathcal{V}_k^2, \mathcal{E}_k^2) \]

Recycle graphs (flows)

Build a new graph for each expansion move

Alahari et al., PAMI’10
Efficient Inference: Reduction

Original Problem (Large)

Fast partially optimal algorithm

Reduced Problem

Approximate algorithm (TRW, BP, Expansion, Swap) (Slow)

Solved Problem (Global Optima)

Approximate solution

Approximate algorithm (Fast)

Approximate Solution

Alahari et al., PAMI’10
Modelling image priors & inference

• Pairwise energy functions are limited

Kohli et al., IJCV’09, Ladicky et al., ECCV’10
Modelling image priors & inference

• Pairwise energy functions are limited

• Need new energy terms to capture fine contours

Kohli et al., IJCV’09, Ladicky et al., ECCV’10
Modelling image priors & inference

• New energy terms to capture **fine contours**

\[ E(y) = \sum_{i \in V} \psi_i(y_i) + \sum_{(i,j) \in E} \psi_{ij}(y_i, y_j) + \sum_{c \in S} \psi_c(y_c) \]

• Higher-order terms, involving several nodes, e.g.,

Interactions (cliques) of size 6

Kohli et al., IJCV’09, Ladicky et al., ECCV’10, Alahari PAMI’10
Modelling image priors & inference

- Higher-order terms, e.g.,
  - grouping “similar” pixels

\[
\psi_c(y_c) = \min\left\{ \min_{k \in \mathcal{L}} \left( |c| - n_k(y_c) \right) \theta_k + \gamma_k, \gamma_{\text{max}} \right\}
\]
Modelling image priors & inference

• Higher-order terms, e.g.,
  – Using “detected” objects

\[
\psi_h(y) = \min(\gamma_{\text{max}}, \min_l (\gamma_l + k_l \sum_{i \in y} \delta(y_i \neq l)))
\]
Modelling image priors & inference

• Higher-order terms of the form:

$$\psi_h(y) = \min(\gamma_{max}, \min_l (\gamma_l + k_l \sum_{i \in y} \delta(y_i \neq l)))$$

• Can be minimized efficiently with move-making algorithms

Kohli et al., IJCV’09, Ladicky et al., ECCV’10, Alahari PAMI’10
Modelling image priors & inference

- Graph construction with auxiliary nodes

Kohli et al., IJCV’09, Ladicky et al., ECCV’10, Alahari PAMI’10
Modelling image priors & inference

- Evaluation: CamVid dataset [Brostow et al. ’08]

<table>
<thead>
<tr>
<th></th>
<th>Building</th>
<th>Tree</th>
<th>Sky</th>
<th>Car</th>
<th>Sign-Symbol</th>
<th>Road</th>
<th>Pedestrian</th>
<th>Fence</th>
<th>Column-Pole</th>
<th>Sidewalk</th>
<th>Bicyclist</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brostow et al. ’08</td>
<td>46.2</td>
<td>61.9</td>
<td>89.7</td>
<td>68.6</td>
<td>42.9</td>
<td>89.5</td>
<td><strong>53.6</strong></td>
<td>46.6</td>
<td>0.7</td>
<td>60.5</td>
<td>22.5</td>
<td>53.0</td>
</tr>
<tr>
<td>Sturgess et al. ’09</td>
<td>84.5</td>
<td>72.6</td>
<td><strong>97.5</strong></td>
<td>72.7</td>
<td>34.1</td>
<td><strong>95.3</strong></td>
<td>34.2</td>
<td>45.7</td>
<td>8.1</td>
<td>77.6</td>
<td>28.5</td>
<td>59.2</td>
</tr>
<tr>
<td>Without detectors</td>
<td>79.3</td>
<td>76.0</td>
<td>96.2</td>
<td>74.6</td>
<td><strong>43.2</strong></td>
<td>94.0</td>
<td>40.4</td>
<td>47.0</td>
<td><strong>14.6</strong></td>
<td>81.2</td>
<td>31.1</td>
<td>61.6</td>
</tr>
<tr>
<td>Our method</td>
<td>81.5</td>
<td><strong>76.6</strong></td>
<td>96.2</td>
<td><strong>78.7</strong></td>
<td>40.2</td>
<td>93.9</td>
<td>43.0</td>
<td><strong>47.6</strong></td>
<td>14.3</td>
<td><strong>81.5</strong></td>
<td><strong>33.9</strong></td>
<td><strong>62.5</strong></td>
</tr>
<tr>
<td>Intersection vs Union</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sturgess et al. ’09</td>
<td><strong>71.6</strong></td>
<td>60.4</td>
<td><strong>89.5</strong></td>
<td>58.3</td>
<td>19.4</td>
<td>86.6</td>
<td><strong>26.1</strong></td>
<td>35.0</td>
<td>7.2</td>
<td>63.8</td>
<td>22.6</td>
<td>49.2</td>
</tr>
<tr>
<td>Without detectors</td>
<td>70.0</td>
<td><strong>63.7</strong></td>
<td><strong>89.5</strong></td>
<td>58.9</td>
<td>17.1</td>
<td>86.3</td>
<td>20.0</td>
<td><strong>35.8</strong></td>
<td>9.2</td>
<td><strong>64.6</strong></td>
<td>23.1</td>
<td>48.9</td>
</tr>
<tr>
<td>Our method</td>
<td>71.5</td>
<td><strong>63.7</strong></td>
<td>89.4</td>
<td><strong>64.8</strong></td>
<td><strong>19.8</strong></td>
<td><strong>86.8</strong></td>
<td>23.7</td>
<td><strong>35.6</strong></td>
<td><strong>9.3</strong></td>
<td><strong>64.6</strong></td>
<td><strong>26.5</strong></td>
<td><strong>50.5</strong></td>
</tr>
</tbody>
</table>
So far ...

• In the context of labelling problems
  – Learning parameters of the function
  – Modelling image priors & inference

Next:
  – Modelling temporal constraints & inference
Human Pose Estimation

Poses in the Wild dataset, CVPR 2014
Human Pose Estimation (in an image)

• Formulated as a graph optimization problem

For an image $I$, pose model $G = (\mathcal{V}, \mathcal{E})$, and $p = \{p^u = (x^u, y^u) \in \mathbb{R}^2 : \forall u \in \mathcal{V}\}$, the optimization problem is given by:

$$\min C(I, p) := \sum_{u \in \mathcal{V}} \phi_u(I, p^u) + \sum_{(u,v) \in \mathcal{E}} \psi_{u,v}(p^u - p^v)$$

Yang and Ramanan, CVPR 2011
Human Pose Estimation

- Extension to videos: introduce temporal links
- Inference is now computationally expensive – requires approximate methods
Human Pose Estimation

• Extension to videos: introduce temporal links
• Inference is now computationally expensive – requires approximate methods
• e.g.,
  – Change graph structure [Sapp et al. ’11, Weiss et al. ’11]
Human Pose Estimation (Video)

• Approximate the graph as combination of trees

• Several useful interactions are discarded

Sapp, Weiss, Taskar, CVPR 2011
Human Pose Estimation (Video)

• Compute a candidate set of poses in each frame
• Then, track (entire pose or pose-parts) over time

• Limited by the no. of candidates or regularization

Park and Ramanan, ICCV 2011; Ramakrishna, Kanade, Sheikh, CVPR 2013
Our Pose Estimation Approach

• Combines
  1. Candidate pose set
     • Generate better candidates
  2. Decomposition strategy
     • Generate limb sequences and recompose the pose

Cherian, Mairal, Alahari, Schmid, CVPR 2014
Better Candidate Poses

- Stabilize the lower-limb pose estimates

\[ C(I_t, p_t) + \tilde{C}(I_{t+1}, \tilde{p}_{t+1}) + \lambda_1 \sum_{u \in \mathcal{W}} \left\| \tilde{p}_{t+1}^u - p_t^u - f_t(p_t^u) \right\|_2^2 \]
De/Re-composition

• Decompose poses and perform limb-tracking
Evaluation: Poses in the Wild Dataset

Our new public dataset: Poses in the Wild

Available at: http://lear.inrialpes.fr/research/posesinthewild/
Human Pose Estimation: Elbows

Cooking Activities

Poses in the Wild
Human Pose Estimation: Wrists

Cooking Activities

Poses in the Wild
Summary

• In the context of labelling problems

- Learning parameters of the function
- Modelling image priors & inference
- Modelling temporal constraints & inference

Code + Datasets at: http://lear.inrialpes.fr/~alahari