Beyond Physical Connections: Tree Models in Human Pose Estimation

Fang Wang\textsuperscript{1,2} and Yi Li\textsuperscript{2,3}

1. Nanjing University of Science and Technology, China
2. NICTA, Australia
3. Australian National University

yi.li@nicta.com.au
Models for human body

- Multiple granularity
- Tree structure
- Flexibility
- Interaction
- Latent structure
Models for human body
- Multiple granularity
- Tree structure
- Flexibility
- Interaction
- Latent structure
Models for human body
- Multiple granularity
- Tree structure
- Flexibility
- Interaction
- Latent structure

Felzenszwalb and Huttenlocher, IJCV 2005
Models for human body

- Multiple granularity
- Tree structure
- Flexibility
- Interaction
- Latent structure

 Yang and Ramanan, CVPR 2011
Models for human body
- Multiple granularity
- Tree structure
- Flexibility
- Interaction
- Latent structure

Wang et al, JMLR 12
Models for human body

- Multiple granularity
- Tree structure
- Flexibility
- Interaction
- Latent structure

Tian et al, ECCV 12
Manually defined structure
Learn the structure?
A model of learned structure

- handles compositional parts
- explores latent structure
- is still a tree
- captures dynamics beyond physical connections
A model of **learned** structure

- handles compositional parts
- explores latent structure
- is still a tree
- captures dynamics beyond physical connections
A model of **learned** structure

- handles compositional parts
- explores latent structure
- is still a tree
- captures dynamics beyond physical connections
A model of **learned** structure

- handles compositional parts
- explores latent structure
- is still a tree
- captures dynamics beyond physical connections
A **model** of learned structure

- handles compositional parts
- explores **latent** structure
- is still a **tree**
- captures dynamics beyond physical connections
Latent tree for pose estimation (1)

Latent Tree

To learn tree structured models for approximating joint distribution of observable variables

- Tree building algorithms:
  - [Chow and Liu, 1968]
  - [Choi et al, JMLR 2011]

- Motivations
  - Novel latent models for human, or
  - Discover intrinsic structures

Wikipedia
Latent tree for pose estimation (1)

**Latent Tree**

To learn tree structured models for approximating joint distribution of observable variables

- Tree building algorithms:
  - [Chow and Liu, 1968]
  - [Choi et al, JMLR 2011]

- Motivations
  - Novel latent models for human, or
  - Discover intrinsic structures

NICTA
Latent Tree for Pose Estimation (1)

Latent Tree
To learn tree structured models for approximating joint distribution of observable variables

- Tree building algorithms:
  - [Chow and Liu, 1968]
  - [Choi et al, JMLR 2011]

- Motivations
  - Novel latent models for human, or
  - Discover intrinsic structures
Latent Tree

To learn tree structured models for approximating joint distribution of observable variables

- Tree building algorithms:
  - [Chow and Liu, 1968]
  - [Choi et al, JMLR 2011]

- Motivations
  - Novel latent models for human, or
  - Discover intrinsic structures
Latent tree for pose estimation (2)

**Definition**

Information distance:  
\[ d_{ij} = -\log\left(\frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i)\text{Var}(X_j)}}\right) \]

- **Parent-Child relationship Test**
  - For each triplet \( i, j, k \in V \).
  - Define \( \Phi_{ijk} \triangleq d_{jk} - d_{ik} \), take one of the two actions:
    - If \( \Phi_{ijk} = d_{ij} \), \( j \) is set to be the parent of \( i \).
    - If \( -d_{ij} \leq \Phi_{ijk} = \Phi_{ijk}' \leq d_{ik} \) for all \( k \) and \( k' \in V\setminus\{i, j\} \), add a hidden node as the parent of \( i \) and \( j \).
Recursive Grouping (RG)

- Initialize
- Test parent-child for pairs
- Repeat

[Choi et al, JMLR 2011]
Recursive Grouping (RG)

- Initialize
- Test parent-child for pairs
- Repeat

[Choi et al, JMLR 2011]
Recursive Grouping (RG)

- Initialize
- Test parent-child for pairs
- Repeat

[Choi et al, JMLR 2011]
Chow-Liu Recursive Grouping (CLRG)

- Minimal spanning tree
- Select neighbor of an internal node
- Perform RG and update structure

[Choi et al, JMLR 2011]
Chow-Liu Recursive Grouping (CLRG)

- Minimal spanning tree
- Select neighbor of an internal node
- Perform RG and update structure

[Choi et al, JMLR 2011]
Chow-Liu Recursive Grouping (CLRG)

- Minimal spanning tree
- Select neighbor of an internal node
- Perform RG and update structure

[Choi et al, JMLR 2011]
Leeds Sport Pose from [Johnson and Everingham, BMVC 2010]
Building Trees for Compositional Parts
Our Approach

Learn a tree structured model for human pose estimation that integrates primitive parts and combined parts

- **Primitive parts**
  - **Joints, non-oriented** $\Rightarrow$ geometric clustering
    - [Yang and Ramanan, CVPR 2011]

- **Combined parts**
  - Distinctive $\Rightarrow$ Visual Categorization
    - SVM + HOG [Dalal and Triggs, CVPR 05]

- **Tree structured models**
  - Learned directly from data
  - Textbook example of exact inference and parameter learning
Our Approach
Learn a tree structured model for human pose estimation that integrates primitive parts and combined parts

- **Primitive parts**
  - Joints, non-oriented $\Rightarrow$ geometric clustering
    - [Yang and Ramanan, CVPR 2011]

- **Combined parts**
  - Distinctive $\Rightarrow$ Visual Categorization
    - SVM + HOG [Dalal and Triggs, CVPR 05]

- **Tree structured models**
  - Learned directly from data
  - Textbook example of exact inference and parameter learning
Our Approach

Learn a tree structured model for human pose estimation that integrates primitive parts and combined parts

- **Primitive parts**
  - Joints, non-oriented $\Rightarrow$ geometric clustering
    - [Yang and Ramanan, CVPR 2011]

- **Combined parts**
  - Distinctive $\Rightarrow$ Visual Categorization
    - SVM + HOG [Dalal and Triggs, CVPR 05]

- **Tree structured models**
  - Learned directly from data
  - Textbook example of exact inference and parameter learning
Our Approach

Learn a tree structured model for human pose estimation that integrates primitive parts and combined parts

- **Primitive parts**
  - Joints, non-oriented $\Rightarrow$ geometric clustering
  - [Yang and Ramanan, CVPR 2011]

- **Combined parts**
  - Distinctive $\Rightarrow$ Visual Categorization
  - SVM + HOG [Dalal and Triggs, CVPR 05]

- **Tree structured models**
  - Learned directly from data
  - Textbook example of exact inference and parameter learning
Learn visual categories for combined parts

- \( k \)-means algorithm on geometric config to find mean patch sizes
- Latent SVM [Divvala et al, 2012] model for each combined part
- Further info: [Wang and Li, IJCAI 2013]

\[
\arg \min_w \frac{1}{2} \sum_{k=1}^{K} \| w_k \|^2 + C \sum_{i=1}^{N} \epsilon_i, \\
y_i w_{t_i} \phi(x_i) \geq 1 - \epsilon_i, \epsilon_i \geq 0, \\
t_i = \arg \max_k w_k \phi(x_i)
\]
Learn visual categories for combined parts

- *k*-means algorithm on geometric config to find mean patch sizes
- Latent SVM [Divvala et al, 2012] model for each combined part
- Further info: [Wang and Li, IJCAI 2013]

\[
\arg \min_w \frac{1}{2} \sum_{k=1}^{K} \| w_k \|^2 + C \sum_{i=1}^{N} \epsilon_i,
\]

\[
y_i w_{t_i} \phi(x_i) \geq 1 - \epsilon_i, \epsilon_i \geq 0,
\]

\[
t_i = \arg \max_k w_k \phi(x_i)
\]
Visual Categorization for Combined Parts

- Learn visual categories for combined parts
  - $k$-means algorithm on geometric config to find mean patch sizes
  - Latent SVM [Divvala et al, 2012] model for each combined part
  - Further info: [Wang and Li, IJCAI 2013]

$$\arg \min_w \frac{1}{2} \sum_{k=1}^{K} \| w_k \|^2 + C \sum_{i=1}^{N} \epsilon_i,$$

$$y_i w_{t_i} \phi(x_i) \geq 1 - \epsilon_i, \epsilon_i \geq 0,$$

$$t_i = \arg \max_k w_k \phi(x_i)$$
Learn visual categories for combined parts

- $k$-means algorithm on geometric config to find mean patch sizes
- Latent SVM [Divvala et al, 2012] model for each combined part
- Further info: [Wang and Li, IJCAI 2013]

$$\arg\min_w \frac{1}{2} \sum_{k=1}^{K} \| w_k \|^2 + C \sum_{i=1}^{N} \epsilon_i,$$

$$y_i w_{t_i} \phi(x_i) \geq 1 - \epsilon_i, \epsilon_i \geq 0,$$

$$t_i = \arg\max_k w_k \phi(x_i)$$
Results for Categorization

Hand

Elbow

Left arm

Left leg
Objective Function for Inference

Objective function

\[ p = \arg \max_p S(t) + \sum_i S(I, p_i) + \sum_{i,j} S(I, p_i, p_j) \]

- Unary term
- Pairwise term
- Compatibility term

DEFINED AS

\[ S(I, p_i) = \omega_i^{t_i} \phi(I, loc_i) \]
Objective Function for Inference

Objective function

\[ p = \arg \max_p S(t) + \sum_i S(l, p_i) + \sum_{i,j} S(l, p_i, p_j) \]

- Unary term
- Pairwise term
- Compatibility term

Defined as

\[ S(l, p_i, p_j) = \omega_{ij}^{t_i t_j} \psi(p_i, p_j) \]
Objective Function for Inference

Objective function

\[ p = \arg\max_p S(t) + \sum_i S(l, p_i) + \sum_{i,j} S(l, p_i, p_j) \]

- Unary term
- Pairwise term
- Compatibility term

Defined as

\[ S(t) = \sum b_{ti}^{ti} + \sum b_{ij}^{ti \cdot t_j} \]
**Experiments**

PARSE dataset, from [Ramanan, NIPS 2006]

Percentage of Correct Parts (PCP)

Strict evaluation: \( d_1 < D/2, \ d_2 < D/2 \)

Loose evaluation: \( (d_1 + d_2)/2 < D/2 \)

NICTA

Fang Wang\(^1,2\) and Yi Li\(^2,3\)  
Human Pose CVPR 13 (slide 17)  
yi.li@nicta.com.au
### Table: Performance on the LSP dataset.

<table>
<thead>
<tr>
<th>Exp.</th>
<th>Method</th>
<th>Torso</th>
<th>Head</th>
<th>U.Leg</th>
<th>L.Leg</th>
<th>U.Arm</th>
<th>L.Arm</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>Yang &amp; Ramanan</td>
<td>92.6</td>
<td>87.4</td>
<td>66.4</td>
<td>57.7</td>
<td>50.0</td>
<td>30.4</td>
<td>58.9</td>
</tr>
<tr>
<td>L</td>
<td>Tian et al. (First 200)</td>
<td>93.7</td>
<td>86.5</td>
<td>68.0</td>
<td>57.8</td>
<td>49.0</td>
<td>29.2</td>
<td>58.8</td>
</tr>
<tr>
<td>L</td>
<td>Tian et al. (5 models)</td>
<td><strong>95.8</strong></td>
<td><strong>87.8</strong></td>
<td>69.9</td>
<td>60.0</td>
<td><strong>51.9</strong></td>
<td>32.8</td>
<td>61.3</td>
</tr>
<tr>
<td>L</td>
<td>Ours (First 200)</td>
<td>88.4</td>
<td>80.8</td>
<td>69.1</td>
<td>60.0</td>
<td>50.5</td>
<td>29.2</td>
<td>59.0</td>
</tr>
<tr>
<td>L</td>
<td>Ours</td>
<td>91.9</td>
<td>86.0</td>
<td><strong>74.0</strong></td>
<td><strong>69.8</strong></td>
<td>48.9</td>
<td>32.2</td>
<td><strong>62.8</strong></td>
</tr>
<tr>
<td>S</td>
<td>Johnson &amp; Everingham</td>
<td>78.1</td>
<td>62.9</td>
<td><strong>65.8</strong></td>
<td><strong>58.8</strong></td>
<td>47.4</td>
<td>32.9</td>
<td><strong>55.1</strong></td>
</tr>
<tr>
<td>S</td>
<td>Yang &amp; Ramanan</td>
<td>82.0</td>
<td>75.8</td>
<td>54.4</td>
<td>51.6</td>
<td>41.0</td>
<td>28.4</td>
<td>50.9</td>
</tr>
<tr>
<td>S</td>
<td>Ours (strict eval)</td>
<td><strong>88.3</strong></td>
<td><strong>81.4</strong></td>
<td>55.3</td>
<td>55.3</td>
<td>43.1</td>
<td>30.5</td>
<td>53.8</td>
</tr>
<tr>
<td>PARSE</td>
<td>Yang &amp; Ramanan</td>
<td>78.8</td>
<td>70.0</td>
<td>66.0</td>
<td>61.1</td>
<td><strong>61.0</strong></td>
<td>37.4</td>
<td>60.0</td>
</tr>
<tr>
<td>L</td>
<td>Ours</td>
<td><strong>88.3</strong></td>
<td><strong>78.7</strong></td>
<td><strong>75.2</strong></td>
<td><strong>71.8</strong></td>
<td>60.0</td>
<td>35.9</td>
<td><strong>65.3</strong></td>
</tr>
</tbody>
</table>

**Note:** The performance metrics are likely accuracy scores, typically used to evaluate pose estimation models. Higher values indicate better performance.
## Experiments (2)

<table>
<thead>
<tr>
<th>Method</th>
<th>Head</th>
<th>L.F.Leg</th>
<th>R.F.Leg</th>
<th>Legs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yang &amp; Ramanan, CVPR 2011</td>
<td>56.1</td>
<td>52.8</td>
<td>58.3</td>
<td>55.6</td>
<td>55.7</td>
</tr>
<tr>
<td>Ours</td>
<td>52.8</td>
<td>60.6</td>
<td>63.3</td>
<td>62.0</td>
<td>58.9</td>
</tr>
</tbody>
</table>

---

**NICTA**

Fang Wang\(^{1,2}\) and Yi Li\(^{2,3}\)

Human Pose CVPR 13 (slide 19)

yi.li@nicta.com.au
Tree models for human pose estimation are efficient

Latent tree is an effective tool for recovering intrinsic structure

Learning visual category of combined part
Thank you!

http://users.cecs.anu.edu.au/~yili/

yi.li@nicta.com.au

Funding support:
Bionic Eye (YL) and China Scholarship Council (FW)

Acknowledgement:
Prof. Yiannis Aloimonos and Dr. Cornelia Fermuller (Maryland), and Prof. Luciano Fadiga (IIT Italy)
Dr. Mathieu Salzmann and other NICTA folks.