2D Human Pose Estimation

LEONG MEI CHEE (PHD STUDENT, IMI & IGS)
ASSOC PROF LEE YONG TSUI (MAE)
ASSOC PROF LIN FENG (SCE)
OUTLINE

• Introduction
• Approaches
• Preliminary work
• Future work
INTRODUCTION

**PROBLEM:** Recover 2D Human Pose from Single Image

- **Applications:** surveillance, entertainment, medical and sports analysis, action recognition, etc.

- **Extension:** 3D human motion recovery.
INTRODUCTION

2D Human Pose Estimation

Body Parts Detection

Pose Estimation

• Difficulties:
  ◦ High variation of human appearance
  ◦ Changes scale and viewpoints
  ◦ Occlusions
  ◦ Cluttered background
  ◦ High degree of freedom in human pose
APPROACHES

1. DEFORMABLE PART MODEL
2. FLEXIBLE MIXTURES-OF-PARTS MODEL
3. HIERARCHICAL POSELETS
1. DEFORMABLE PART MODEL

• Each part model consists of a spatial model and a part filter.
  ◦ Part filter – local appearance template
  ◦ Spatial model – spatial relation between parts

Reference: CVPR 2007 Slides, Part based model by Rob Fergus (MIT)
1. DEFORMABLE PART MODEL


- Use Histogram of Oriented Gradients (HOG) features to create templates at different scales.
- Templates are assigned with specified weights - part filters.
- Spatial model defines placement of a part relative to the detection window, and a deformation cost for each placement.
1. DEFORMABLE PART MODEL

- Scoring function:
  \[ \text{score}(p_1, \ldots, p_n) = \sum_{i=1}^{n} m_i(p_i) - \sum_{(i,j) \in E} d_{ij}(p_i, p_j) \]

- Iteratively train SVM to learn the weights and deformation cost
2. FLEXIBLE MIXTURES-OF-PARTS MODEL


Classic articulated limb model that captures different orientation and foreshortening states of limbs.

Use a mixture of non-oriented structures to approximate limbs’ transformations. Different set of patches and springs are used to capture parts connection.
2. FLEXIBLE MIXTURES-OF-PARTS MODEL

• Parts are grouped into orientations based on their relative location with respect to their parents.

• These clusters are used to generate mixture labels for parts during training.

• All mixtures of parts are set to a fixed HOG cell size.
2. FLEXIBLE MIXTURES-OF-PARTS MODEL

Templates of 14 parts with 4 local mixtures

A tree structure, where parts are placed at their best-scoring location relative to their parent
2. FLEXIBLE MIXTURES-OF-PARTS MODEL

• Scoring function:
  ◦ Appearance model – HOG features.
  ◦ Deformation model – relative connection between two parts.
  ◦ Co-occurrence model – part types and its favour co-occurrence part.

\[
S(I, p, t) = \sum_{i \in V} b_{i}^{t_i} + \sum_{i,j \in E} b_{i,j}^{t_i,t_j} + \\
\sum_{i \in V} w_{i}^{t_i} \cdot \phi(I, p_i) + \sum_{i,j \in E} w_{i,j}^{t_i,t_j} \cdot \psi(p_i - p_j)
\]
3. HIERARCHICAL POSELETS


- Poselets (novel definition of a part) describes a part of one’s pose.
- They are parts that are tightly clustered in both 3D configuration space and 2D image appearance.
- Poselet detection: use HOG features and linear SVM, and train classifier using bootstrapping.
3. HIERARCHICAL POSELETS


- Hierarchical poselets can be rigid parts, parts that cover large portions of human bodies, as well as the whole body.
3. HIERARCHICAL POSELETS

The hierarchical pose representation. Black edges indicate the connectivity among different parts.

- Scoring function:

\[ F(L, I) = \sum_{i \in \mathcal{V}} \phi(l_i; I) + \sum_{(i,j) \in \mathcal{E}} \psi(l_i, l_j) \]

local appearance  
spatial model
Preliminary Work

• Parts Detection

Future Work

• Pose Estimation
Preliminary Work

Body Parts Detection

- Deformable part model with 14 body parts.
- Dataset from CVPR Challenge 2015.
- Limited to frontal and upright position.

2015 Looking At People CVPR Challenge - Human Pose Recovery

[Rob Fergus, CVPR 2007 Slides]
Preliminary Work

- Part filter - use HOG feature as local appearance template.
- Categorize each body part into 8 orientation bins

Torso bin 4  Arm bin 2  Hand bin 5  Leg bin 3
Preliminary Work

- Spatial model – captures relative distance between body parts and their parent.

Body parts’ distribution relative to their parent point:

- Left & Right thighs
- Right arm
- Left forearm
- Right foot
Preliminary Work

• Spatial model – used to constraint detection region.
Preliminary work

- Training methods:
  - Linear SVM
  - Linear SVM + filter
  - Cascade classifier

<table>
<thead>
<tr>
<th>Method</th>
<th>Linear SVM</th>
<th>SVM + filter</th>
<th>Cascade classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td># Left arm configurations</td>
<td>16315</td>
<td>683</td>
<td>1</td>
</tr>
<tr>
<td># Right arm configurations</td>
<td>14670</td>
<td>241</td>
<td>1</td>
</tr>
<tr>
<td># Left leg configurations</td>
<td>1175</td>
<td>68</td>
<td>1 (incomplete)</td>
</tr>
<tr>
<td># Right leg configurations</td>
<td>1003</td>
<td>191</td>
<td>1 (incomplete)</td>
</tr>
</tbody>
</table>
Preliminary work

Body Joints Detection

• Training dataset from Yang and Ramanan [2013]
• Use HOG feature.
• 14 body joints position.
• Training images from arbitrary poses.
• Unconstrained spatial relation.

[Yang and Ramanan 2011]
Preliminary work

Positive images for body joints detection [Yang and Ramanan 2013]
Preliminary work

• Training methods:
  • Linear SVM
  • Cascade classifier

<table>
<thead>
<tr>
<th>Method</th>
<th>Linear SVM</th>
<th>Cascade classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td># Right shoulder joint</td>
<td>11</td>
<td>2</td>
</tr>
</tbody>
</table>
Future work

• Model scoring function to find best configuration.

Body parts: Linear SVM + boosted filter
Body joints: Cascade classifier
Future work

• Extend work to general poses.
• Solve occlusion problem.

Left and right thigh not detected

The hierarchical pose representation. Black edges indicate the connectivity among different parts.
THANK YOU