Multi-view Body Part Recognition with Random Forests
Vahid Kazemi, Magnus Burenius, Hossein Azizpour, Josephine Sullivan. KTH.
Problem

Multi-view human 3D pose estimation in the wild
The Typical 3D Pose Data Set

HumanEva data set
Our New 3D Pose Data Set

**Challenges:** moving cameras, dynamic backgrounds, motion blur, occlusion.
2D & 3D Pose Estimation using Pictorial Structures / Part-based Models

- Appearance model for each part
- Pose model connecting the parts
Pictorial Structures & Part-based Models

- Position of the parts  \( X = (X_1, \ldots, X_N) \)
- Image evidence for the parts  \( I = (I_1, \ldots, I_N) \)
- Appearance model for each part  \( P(I_n | X_n) \)
- Pose model connecting the parts  \( P(X) \)
- Joint distribution  \( P(X, I) = P(X) \prod_{n=1}^{N} P(I_n | X_n) \)
Pictorial Structures & Part-based Models

\[ P(X, I) = P(X) \prod_{n=1}^{N} P(I_n | X_n) \]

\[ \log P(X, I) = \log P(X) + \sum_{n=1}^{N} \log P(I_n | X_n) \]

\[ \argmax_{X} P(X|I) = \argmax_{X} \log P(X, I) \]

Can use dynamic programming to find global solution:
- For 2D pose estimation see Felzenszwalb et al. CVPR 2000.
- For 3D pose estimation see Burenius et al. CVPR 2013.
Part Appearance Model

1. Single view 2D
2. Multiple view 3D
2D Part Appearance Model

\[ P(I_n|X_n) \]
Body Part Classification as 2D Appearance Model

- Inspired by Kinect approach:
  

- **Input:**
  
  \[ x = (h, p) \]

- **Output:**
  
  \[ y \in \{0, 1, \ldots, N\} \]

Joint-based part representation
Decision Tree for Pixel Classification

- Weak classifier: \( w = (d, n, t) \)

- Decision: \( h(p + d, n) < t \)
Random Forest
2D Pose Estimation Demo Movie

2D part appearance likelihoods and pose estimation using a pose prior. Estimation is done independently for each frame.
3D Part Appearance Model

Assume calibrated cameras and bounding cube of player.
3D Part Appearance Model

Back-project from 2D pixels to a 3D voxel grid (64x64x64) covering the bounding cube:

\[ X_n^c = T^c(X_n) \]

\[ P(I_n | X_n) = \prod_{c=1}^{C} P(I_n^c | X_n^c) \]
The Problem of Symmetric Body Parts

Left and right parts look similar.
The Problem of Symmetric Body Parts

Approach 1:
Classify left/right parts of the person.

Disadvantage:
• Too Difficult
The Problem of Symmetric Body Parts

Approach 2:
Ignore the left/right label of parts.

Disadvantages:
• Double counting
• Correspondences across views
Aggregating Scores Across Views

\[ P(I_n | X_n) \]

**Approach 1:**
Assuming we know the left/right label, relative the person, for each view.

**Approach 2:**
Ignoring left/right label of parts.
Naive Multi-view Pose Estimation
The Problem of Symmetric Body Parts

Approach 3:
Classify the left/right parts of the image.

Disadvantage:
• Correspondences across views
Handle Left-Right Correspondences with a Latent Variable

- Match left and right leg of the image with left and right leg of the person.
- For each view we have 2 choices for the legs and 2 for the arms.
- For C views we have $4^C$ choices.
- Let the latent random variable $S$ describe this unknown mapping.
Multi-view Inference

\[ P(X, I, S) = P(X)P(S)\prod_{n=1}^{N}P(I_n | X_n, S) \]

\[ \max_{X, S} P(X, S | I) = \max_{S} \max_{X} \log P(X, I, S) \]
3D Part Appearance Model

\[ P(I_n|X_n, \tilde{S}) \] & ground truth pose
Multi-view Pose Estimation

- Just using 3D part appearance model.
- No 3D pose model. No motion model.
- Latent variable handles mirror symmetry.
Conclusions

- New data set available at our web-page.
Conclusions

- Random forest classification works well for body part recognition in ordinary images.
Conclusions

• Problem of symmetric body parts, for multi-view part-based models.
• Latent variable solution.
Thank you!